

Evaluating Different Visualization Designs for Multivariate Personal Health Data

**A thesis submitted in partial fulfilment
of the requirement for the degree of Doctor of Philosophy**

Majedah Mohammad Alrehiely

April 2020

**Cardiff University
School of Computer Science & Informatics**

**To the soul of my father,
To my mother, for her endless support
throughout my life.**

Abstract

Over the past decade, self-tracking devices and apps have been increasingly used for personal data collection, particularly in the monitoring of health and physical activity. Data visualization is crucial to helping people understand and make sense of collected data. The literature review carried out on this topic revealed several studies which focus on data visualization, personal informatics systems and people's self-tracking practice and their requirements. However, in terms of personal health data, there is a significant limitation in addressing the knowledge gap between these areas and the visualizations supported by self-tracking apps/dashboards, which people frequently use when viewing and exploring their health data.

The literature review is followed by a user study and a visualization review, which explores the methods used to represent personal health data on popular self-tracking devices, apps and dashboards in order to understand the main limitations of various designs. The work goes on to address two main challenges: the need for combining multiple personal health variables to support a better representation of an individual's health status, which could help in understanding the relationships between different health data; and to find a more suitable layout that is related to the periodic and temporal nature of personal health data.

Therefore, this thesis focuses on the design and evaluation of visualizations to address these challenges by considering different design aspects of the visual layout, methods and visual encoding, while taking into account the multivariate nature of the data. With respect to the visual layout, the design proposes a novel Radial layout that utilizes an analogue clock metaphor to provide cognitive support by representing the data on an hourly basis through a clock-type display. The Radial representation is effective for data that naturally exhibit a periodic pattern. It also supports understanding of patterns with either a circular or seasonal behaviour. The de-

veloped visualizations' evaluation process implements controlled lab-experiments. The evaluation follows a two-step method, starting with a preliminary study that leads to the design of the main evaluation study, which includes quantitative and qualitative measures of participants' performance and preferences in the designs.

The main contributions of the thesis are: (1) A thorough overview of the visualization methods provided by companion apps, dashboards and embedded displays of popular self-tracking devices. It also discusses their limitations and strengths, providing a taxonomy for the applied visualizations. (2) Suggestions and recommendations for addressing the challenges highlighted in personal health visualizations on apps and dashboards. (3) The design and implementation of multiple visualization alternatives to combine multivariate personal health data using various charts, methods and layouts. (4) Insights regarding personal health visualization gained from the results and in-depth analysis of data resulting from a controlled lab experiment. This experiment evaluated the proposed visualizations with respect to the participants' performance regarding visual tasks related to real datasets. (5) Insights on people's preferences on the visualizations and a structured qualitative analysis of their feedback, which demonstrated the effects of each implemented visual element. In addition to structuring design guidelines for this specific data type, the results of the study prove that the traditional Linear layout either outperforms or is comparable to the proposed Radial layout. The study also shows how the applied visual encoding and the visualization method influence the performance according to specific tasks and under different data densities. It provides plausible explanations for the significant differences in the observed performance and preferences patterns, which inform future visualization designs for personal health data.

Acknowledgements

I wish to express my gratitude and appreciation to my supervisor Dr Parisa Eslambolchilar for her continuous guidance, support and valuable feedback. I really appreciate her patience and the positive criticism she has given me during my Ph.D. journey. Furthermore, I would like to thank Dr Rita Borgo, my external supervisor, for her effort and encouragement to improve my research. I appreciate the effort both my supervisors made to continue guiding me throughout the Ph.D. project that I have started at Swansea University.

In addition, I would like to express my appreciation to my sponsors, the Ministry of Education of Saudi Arabia and Taibah University, for their financial support. Without the funding, it would never have been possible to commence the project.

I would like to express my deepest gratitude to my mother for her endless encouragement and support throughout my life. Special thanks and appreciation are due to all my family members who have stood beside me and provided support, either financially or emotionally, throughout my studies despite the long distance. I am incredibly thankful for having them in my life as they have always encouraged me to pursue my goals, even at times when I have doubted myself.

I am grateful to my fellow Ph.D. students in computer science for their support and helpful comments. Special thanks and appreciation go to my extended family and friends for their continued encouragement and support.

Finally, many people contributed, directly and indirectly, to the production of this work, and I apologise to those whom I have inadvertently forgotten to name.

This thesis would not have been possible or complete without the help and support of Allah. I thank Allah for giving me strength and patience and the continuous grace, mercy, and blessings.

Contents

Abstract	iii
Acknowledgements	v
Contents	vi
List of Publications	xiv
List of Figures	xv
List of Tables	xxii
1 Introduction	1
1.1 Introduction	2
1.2 Motivation and Vision	3
1.3 Definitions and Terminology	6
1.4 Research Aims	7
1.5 Research Questions	8
1.6 Research Hypotheses	9
1.7 Research Methodology and Thesis Structure	10

1.8	Research Contributions	12
1.9	Thesis Outline	13
2	Background and Literature Review	16
2.1	Introduction	17
2.2	Personal Health Data	19
2.2.1	Complexity in the Personal Health Data	19
2.2.2	Data Visualization in Personal Informatics	22
2.2.3	Visualization of Personal Health Data	26
2.3	Information Visualization	33
2.3.1	Human Factors in Information Visualization	34
2.3.2	Visualization by Task Taxonomy	42
2.3.3	Visualization of Time-Series Data	48
2.4	Empirical Evaluation Studies	59
2.4.1	Evaluation in Information Visualization	59
2.4.2	Personal Visualization and Evaluation Methods	61
2.5	Summary	62
3	Analysis of the Use and Preferences of Tracker Users	64
3.1	Introduction	65
3.2	Aims and Research Questions	65
3.3	Methodology	66
3.3.1	Survey Design	66
3.3.2	Procedure and Participants Recruitment	67

3.4	Analysis and Results	69
3.4.1	Activity Tracking Data	70
3.4.2	Data Visualization and Sharing	73
3.5	Discussion	78
3.6	Conclusion	79
4	A Review of Visualization Tools for Health Data	80
4.1	Introduction	81
4.2	Aims and Research Questions	82
4.3	Methodology	83
4.3.1	Device Selection	83
4.3.2	Device Review	84
4.4	Results and Analysis	85
4.4.1	Data Visualization Methods Taxonomy	86
4.5	Discussion	92
4.5.1	Visual Layouts	92
4.5.2	Sharing Health Data	99
4.6	Suggestions and Recommendations	99
4.7	Insights for the Study's Design	101
4.7.1	Data Type and Length	101
4.7.2	Selected Visual Choices and Visual Tasks	102
4.7.3	Visualization Layout	103
4.7.4	Visual Encoding	104

4.7.5	Visualization Methods	105
4.8	Updates on the Visualizations on the Apps and Dashboards	106
4.8.1	Apple Watch and its Companion Apps	106
4.8.2	Fitbit Watch and its Companion App and Dashboard	107
4.8.3	The Garmin Watch and its Companion App and Dashboard	107
4.8.4	Samsung Watch and its Companion Apps	108
4.9	Conclusion	108
5	Visualization Design and Preliminary Evaluation	111
5.1	Introduction	112
5.2	Aims and Research Questions	112
5.3	Methodology	114
5.3.1	Data Source and Characteristics	114
5.3.2	Visualization Design and Development	115
5.3.3	Visualization Tasks	122
5.3.4	Experiment Variables	123
5.3.5	Participants and Recruitment	124
5.3.6	Procedure	125
5.4	Analysis and Results	127
5.4.1	Task Completion Time	127
5.4.2	Task Answers and Error Rate	134
5.4.3	Overall Groups' Performance (Time and Accuracy)	142
5.4.4	Subjective Responses	147

5.5	Discussion	153
5.6	Lessons Learned and the Main Evaluation Study Design	155
5.7	Conclusion	158
6	Main Evaluation Study of the Visualizations	159
6.1	Introduction	160
6.2	Aims and Research Questions	161
6.2.1	Hypotheses	162
6.3	Methodology	168
6.3.1	Visualization Design and Development	168
6.3.2	Visualization Task Taxonomy	173
6.3.3	Dataset Source and Characteristics	180
6.3.4	Experiment Variables	192
6.3.5	Participants Recruitment	192
6.3.6	Procedure	199
6.4	Analysis and Results	202
6.4.1	Analysis Method	202
6.4.2	Calculating Time and Accuracy	204
6.4.3	Homogeneity of Variance	204
6.4.4	Tasks Analysis	206
6.4.5	Differences in the Datasets and the Results	224
6.5	Discussion	226
6.5.1	T1 Identify	227

6.5.2	T2 Determine	230
6.5.3	T3 Compare	233
6.5.4	T4 Infer	236
6.5.5	T5 Locate	238
6.5.6	Hypotheses Testing Outcomes	240
6.6	Conclusion and Summary of the Outcomes	244
6.6.1	Summary Tables	246
6.6.2	Conclusion	250
7	Participants' Feedback and Preferences in the Visualizations	252
7.1	Introduction	253
7.2	Research Questions	253
7.3	Methodology	254
7.4	Analysis and Results	255
7.4.1	Visualization Ratings	256
7.4.2	Preferred Visualizations	260
7.4.3	Participants' Justification for their Choices	267
7.5	Discussion	274
7.5.1	An Overview of Visualizations Ratings and Preferences	274
7.5.2	Insights from People's Feedback on the Visualizations	276
7.6	Summary of Findings	284

8 Discussion	287
8.1 Introduction	288
8.2 Visualization of Tracked Health Data	288
8.3 The Developed Visualization Designs	292
8.3.1 Differences from the Visualizations on Apps/Dashboards	293
8.3.2 Differences from the Visualizations in the Literature	294
8.3.3 The Visualizations of the Multivariate Data	294
8.4 Personal Health Visualization Design Implications	295
8.4.1 Task-Related Implications (Based on the Quantitative Data)	295
8.4.2 Insights for Future Designs (Based on the Qualitative Data)	300
8.5 Conclusion	301
9 Conclusion and Future Work	302
9.1 Thesis Summary and Contributions	303
9.2 Limitations	306
9.3 Future Work	307
9.4 Conclusion	310
Bibliography	311
Appendices	336
Preliminary Evaluation Study Surveys	337
.1 Preliminary Evaluation Study: Demographic and Pre-Experiment Questionnaire	338
.2 Preliminary Evaluation Study: Post-Experiment Questionnaire	349

.3	Preliminary Evaluation Study: Training Sessions Samples	357
.4	Preliminary Evaluation Study: Group's Accuracy Graphs	361
Main Evaluation Study		366
.5	Main Evaluation Study: Participation Forms	367
.6	Main Evaluation Study: Demographic and Pre-Experiment Questionnaire . . .	374
.7	Main Evaluation Study: Post-Experiment Questionnaire	390
.8	Groups Demographic Information	402
.9	Groups and Participants' Behaviour Analysis	403

List of Publications

Some of the work introduced in this thesis is based on the following publications.

- Majedah Alrehiely, Parisa Eslambolchilar, and Rita Borgo. 2018. A taxonomy for visualisations of personal physical activity data on self-tracking devices and their applications. In Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI '18). BCS Learning & Development Ltd., Swindon, UK, Article 17, 15 pages. DOI: <https://doi.org/10.14236/ewic/HCI2018.17>
- Majedah Alrehiely, Parisa Eslambolchilar, and Rita Borgo. 2018. Evaluating Different Visualization Designs for Personal Health Data. In Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI '18). BCS Learning & Development Ltd., Swindon, UK. DOI: <http://dx.doi.org/10.14236/ewic/HCI2018.205>

List of Figures

1.1	PhD Research Stages Diagram	12
1.2	PhD Thesis Structure Diagram	14
2.1	Chapter 2 Structure and Basic Section Diagram	17
2.2	Chapter 2 Health Data Section Diagram	19
2.3	Chapter 2 Information Visualization Section Structure Diagram	33
2.4	Chapter 2 Evaluation Studies Section Diagram	59
3.1	User Study: Participants' Demographic Information	70
3.2	User Study: Participants' Used Tracking Devices	71
3.3	User Study: Participants' Preferences in Abstract and the Statistical Visualiza- tions	74
3.4	User Study: Participants' Responses to Data Visualization Questions	75
3.5	User Study: Participants' Preferences for Data Representation and for Sharing .	76
3.6	User Study: Participants' Responses to the Sharing Questions	77
4.1	Screenshots from Fitbit App	87
4.2	Fitbit Circular Rings Chart	87

4.3	Visualizations on the Activity app running on iOS and Apple Watch	90
4.4	Screenshots from Apple Activity App	91
4.5	A Screenshot of Samsung Health App	92
4.6	A Screenshot from Garmin Connect App	97
4.7	Four Screenshots from Garmin Connect App	98
4.8	Two Screenshots of Different sleep Visualization on Garmin App	98
4.9	A Screenshot captured from iOS Activity app. A bar chart represents the move (i.e., total active calories) for the last year	109
4.10	The Visualization of Distance and Step Count on ' Apple iOS Health App . . .	110
5.1	A Sample of the Dataset Used in Developing the Visualizations in the prelim- inary Evaluation	115
5.2	The Background of the Proposed Radial layout in the Preliminary Evaluation .	117
5.3	The Eight Visualization Designs Developed for the Preliminary Evaluation . . .	120
5.4	Screenshots of the Tooltip Presented on the Visualization	121
5.5	A Screenshot of the Preliminary Evaluation Experiment	127
5.6	Task Completion Time for the Eight Visualizations for Group 1	130
5.7	Task Completion Time for the Eight Visualizations for Group 2	131
5.8	Task Completion Time for the Eight Visualizations for Group 3	132
5.9	Task Completion Time for the Eight Visualizations for Group 4	133
5.10	Task Accuracy and Completion Time on the Radial and Linear Layouts for the Four Groups	137
5.11	Group 1 Tasks Accuracy	138
5.12	Group 2 Tasks Accuracy	139

5.13	Group 3 Tasks Accuracy	140
5.14	Group 4 Tasks Accuracy	141
5.15	Overall Task Completion Time for the Visualizations for All Groups	144
5.16	Overall Task Error Rate for the Visualizations for All Groups	145
5.17	Overall Task Error Rate for the Visualizations for All Groups (percentage)	146
5.18	Task Accuracy and Completion Time on the Radial and Linear Layouts	147
5.19	participants' Ratings for the Visualizations' Readability and Attractiveness Before the Experiment	149
5.20	Participants' Preferences in the Visualizations Before the Experiment	151
5.21	Participants' Preferences in the Visualizations After the Experiment	151
5.22	participants' Ratings for the Visualizations' Readability and Attractiveness After the Experiment	152
6.1	The Background of the Proposed Radial layout in the Main Study	169
6.2	The Visualization Designs Developed for the Main Study	172
6.3	Participants' Demographic Information	196
6.4	Participants' Health Tracking Routine/Behaviour	197
6.5	The Participants' Responses of the Chart Familiarity	198
6.6	Two Screenshots of the Main Evaluation Study Experiment's Interface	201
6.7	Task Completion Time and Accuracy for 'Identify' Task on the Eight Visualization Designs (Dataset 0).	209
6.8	Task Completion Time and Accuracy for 'Identify' Task on the Eight Visualization Designs (Dataset 1).	209
6.9	Task Completion Time and Accuracy for 'Determine' Task on the Eight Visualization Designs (Dataset 0).	213

6.10	Task Completion Time and Accuracy for ‘Determine’ Task on the Eight Visualization Designs (Dataset 1).	213
6.11	Task Completion Time and Accuracy for ‘Compare’ Task on the Eight Visualization Designs (Dataset 0).	216
6.12	Task Completion Time and Accuracy for ‘Compare’ Task on the Eight Visualization Designs (Dataset 1).	216
6.13	Task Completion Time and Accuracy for ‘Infer’ Task on the Eight Visualization Designs (Dataset 0).	219
6.14	Task Completion Time and Accuracy for ‘Infer’ Task on the Eight Visualization Designs (Dataset 1).	219
6.15	Task Completion Time and Accuracy for ‘Locate’ Task on the Eight Visualization Designs (Dataset 0).	223
6.16	Task Completion Time and Accuracy for ‘Locate’ Task on the Eight Visualization Designs (Dataset 1).	223
7.1	Ratings for the Visualizations Readability Before and After the Experiment. . .	258
7.2	Ratings for the visualizations Attractiveness Before and After the Experiment. .	259
7.3	The Participants’ Most Preferred Visualization(s) Before and After the Experiment	262
7.4	Participants’ Preferences Justification Model	270
7.5	Task Related Feedback	270
1	Accuracy of All Tasks for the Eight Visualizations for Group 1	362
2	Accuracy of All Tasks for the Eight Visualizations for Group 2	363
3	Accuracy of All Tasks for the Eight Visualizations for Group 3	364
4	Accuracy of All Tasks for the Eight Visualizations for Group 4	365

5	Gender Distribution in Each Group	402
6	Age Distribution in Each Group	403
7	The Participants' Responses of their Education Degree (Distribution Based on Groups).	404
8	The Participants' Responses of their Education Field (Distribution Based on Groups).	405
9	The Participants' Responses of their Employment Status (Distribution Based on Groups).	406
10	The Distribution of the Participants who Track and Do Not Track in Each Group	407
11	The Types of the Health/Physical Activity tracked by the Participants in Each Group	408
12	The Types of the Physical Activity Tracking Devices Used by the Participants in Each Group	409
13	The Participants' Responses of their Familiarity with Different Chart and Statistical Data (Distribution Based on Groups).	410
14	Task completion time for 'Identify' task for the participants in the groups (1, 2, 3 & 4)	411
15	Task completion time for 'Identify' task for the participants in the groups (5, 6, 7 & 8)	412
16	Task Accuracy for the 'Identify' Task on the Eight Visualization Designs Per Dataset (Groups 1, 2, 3 & 4)	413
17	Task Accuracy for the 'Identify' Task on the Eight Visualization Designs Per Dataset (Groups 5, 6, 7 & 8)	414
18	Task Completion Time for the 'Identify' Task on the Eight Visualization Designs Per Dataset (Groups 1, 2, 3 & 4)	415

19	Task Completion Time for the ‘Identify’ Task on the Eight Visualization Designs Per Dataset (Groups 5, 6, 7 & 8)	416
20	Task completion time for ‘Determine’ task for the participants in the groups (1, 2, 3 & 4)	417
21	Task completion time for ‘Determine’ task for the participants in the groups (5, 6, 7 & 8)	418
22	Task Accuracy for the ‘Determine’ Task on the Eight Visualization Designs Per Dataset (Groups 1, 2, 3 & 4)	419
23	Task Accuracy for the ‘Determine’ Task on the Eight Visualization Designs Per Dataset (Groups 5, 6, 7 & 8)	420
24	Task Completion Time for the ‘Determine’ Task on the Eight Visualization Designs Per Dataset (Groups 1, 2, 3 & 4)	421
25	Task Completion Time for the ‘Determine’ Task on the Eight Visualization Designs Per Dataset (Groups 5, 6, 7 & 8)	422
26	Task completion time for ‘Compare’ task for the participants in the groups (1, 2, 3 & 4)	423
27	Task completion time for ‘Compare’ task for the participants in the groups (5, 6, 7 & 8)	424
28	Task Accuracy for the ‘Compare’ Task on the Eight Visualization Designs Per Dataset (Groups 1, 2, 3 & 4)	425
29	Task Accuracy for the ‘Compare’ Task on the Eight Visualization Designs Per Dataset (Groups 5, 6, 7 & 8)	426
30	Task Completion Time for the ‘Compare’ Task on the Eight Visualization Designs Per Dataset (Groups 1, 2, 3 & 4)	427
31	Task Completion Time for the ‘Compare’ Task on the Eight Visualization Designs Per Dataset (Groups 5, 6, 7 & 8)	428

32	Task completion time for ‘Infer’ task for the participants in the groups (1, 2, 3 & 4)	429
33	Task completion time for ‘Infer’ task for the participants in the groups (5, 6, 7 & 8)	430
34	Task Accuracy for ‘Infer’ Task on the Eight Visualization Designs Per Dataset (Groups 1, 2, 3 & 4)	431
35	Task Accuracy for ‘Infer’ Task on the Eight Visualization Designs Per Dataset (Groups 5, 6, 7 & 8)	432
36	Task Completion Time for ‘Infer’ Task on the Eight Visualization Designs Per Dataset (Groups 1, 2, 3 & 4)	433
37	Task Completion Time for ‘Infer’ Task on the Eight Visualization Designs Per Dataset (Groups 5, 6, 7 & 8)	434
38	Task completion time for ‘Locate’ task for the participants in the groups (1, 2, 3 & 4)	435
39	Task completion time for ‘Locate’ task for the participants in the groups (5, 6, 7 & 8)	436

List of Tables

3.1	The Online Survey Sample	68
3.2	User Study: The Data Tracked by the Users of Different Tracking Devices	72
4.1	Visualization Abbreviations Used in Table 4.3	93
4.2	Colour Codings Used in Table 4.3	93
4.3	Main Visualizations Used by the Selected Devices	94
5.1	Tasks/Questions in the Preliminary Evaluation	123
5.2	Data Tracked by the Participants in the Preliminary Evaluation Study and the Used Devices/ Apps	125
5.3	The Order of the Visualizations in the Preliminary Evaluation	126
5.4	Participants' Performance in Terms of Task Completion in Each Group	129
5.5	Task Completion Time on the Radial and the Linear Layouts for the Four Groups	129
5.6	Participants Performance in Terms of Task Accuracy in Each Group	136
5.7	Task Accuracy on the Radial and Linear Layouts for the Four Groups	136
5.8	Task Completion Time on the Radial and Linear Layouts	143
5.9	Task Accuracy on the Radial and Linear Layouts	143

6.1	Explanation of the Abbreviations Used for the Visualization Names	171
6.2	Tasks/Questions in the Main Study	176
6.3	Correct Responses to Tasks/Questions in the Main Evaluation Study	176
6.4	The Cognitive and the Perceptual Components of the Visual Tasks (T1, T2 and T3)	178
6.5	The Cognitive and the Perceptual Components of the Visual Tasks (T4 and T5)	179
6.6	Main Evaluation Study Datasets Characteristics	181
6.7	Images Used for the LSBC Visualizations for Dataset 0 and Dataset 1	182
6.8	Images Used for the LOBC Visualizations for Dataset 0 and Dataset 1	183
6.9	Images Used for the LSLC Visualizations for Dataset 0 and Dataset 1	184
6.10	Images Used for the LOLC Visualizations for Dataset 0 and Dataset 1	185
6.11	Images Used for the RSBC Visualizations for Dataset 0 and Dataset 1	186
6.12	Images Used for the ROBC Visualizations for Dataset 0 and Dataset 1	187
6.13	Images Used for the RSLC Visualizations for Dataset 0 and Dataset 1	188
6.14	Images Used for the ROLC Visualizations for Dataset 0 and Dataset 1	189
6.15	Demographic Information of the Participants	194
6.16	Tracking Apps Used by the Participants	195
6.17	First Viewed Visualizations for the Eight Groups in the Main Study	195
6.18	Results of the Levene Test for Homogeneity of Variance	206
6.19	Task T1 (Identify) Analysis Within the Linear and the Radial Layouts for Dataset 0 and 1	210
6.20	Task T1 (Identify) Analysis Across Both Layouts for Dataset 0 and 1	211
6.21	Task T2 (Determine) Analysis Within the Linear and Radial Layouts for Dataset 0 and 1	212

6.22	Task T2 (Determine) Analysis Across Both Layouts for Dataset 0 and 1	214
6.23	Task T3 (Compare) Analysis Within the Linear and Radial Layouts for Dataset 0 and 1	215
6.24	Task T3 (Compare) Analysis Across Both Layouts for Dataset 0 and 1	217
6.25	Task T4 (Infer) Analysis Within the Linear and Radial Layouts for Dataset 0 and 1	218
6.26	Task T4 (Infer) Analysis Across Both Layouts for Dataset 0 and 1	220
6.27	Task T5 (Locate) Analysis Within the Linear and Radial Layouts for Dataset 0 and 1	221
6.28	Task T5 (Locate) Analysis Across Both Layouts for Dataset 0 and 1	222
6.29	The Hypotheses from the Main Study	240
6.30	Main Evaluation Study Findings Summary T1	247
6.31	Main Evaluation Study Findings Summary T2	248
6.32	Main Evaluation Study Findings Summary T3	248
6.33	Main Evaluation Study Findings Summary 41	249
6.34	Main Evaluation Study Findings Summary T5	249
6.35	Best and Worst Visualizations for the Five Tasks (Dataset 0 and 1)	251
7.1	Participants Feedback Questions	256
7.2	The Results of the analysis of the Visualizations' Ratings Data	257
7.3	The Results the analysis of the Visualizations' Preferences Data	263
7.4	Participants' Preferences Changes.	264
7.7	Major Categories/Themes identified in the Participants' Responses	269
7.8	Task Related feedback	285

Chapter 1

Introduction

1.1 Introduction

Personal tracking has a long history; people have been tracking their activities using different methods, such as diaries, for many centuries and personal computers for many decades [185]. One of the earliest technologies, which was first used for tracking and self-monitoring in the late 19th century, is the weight scale [50]. More recently, self-tracking has increased in popularity with the advancement of smartphones and wearable devices, which have embedded accelerometers and built-in GPS for activity tracking indoors and outdoors [185]. This thesis focuses on a very popular and important aspect of personal data: health data.

Health and physical activity tracking are no longer limited to specialised fitness and activity tracking devices and heart rate monitors. It is becoming highly popular to track personal health and activity using different types of worn accessories such as multi-functional smartwatches, in addition to the broad range of apps on smartphones. Consequently, with more sophisticated embedded sensors and data classification algorithms becoming available on smartphones and smartwatches, the concept of personal tracking is expanding rapidly to capture different types of data about oneself covering different domains. For example, Fitbit ¹ is used to capture sport movement data, calorie intake to managed by software such as MyFitnessPal ² and many others.

According to a study conducted in 2017 by the “Research 2 Guidance” health research group [181], the number of mobile health or mHealth apps available in the major app stores was around 325K published by around 80K app developers, with a total download rate of 3.7 billion in 2017. In addition, the research group reported a significant increase in the growth of sales of health and fitness tracking devices [180]. These health apps and tracking devices often collect different health metrics (e.g., step count, heart rate and pace); therefore, they generate large collections of heterogeneous health data. These large volumes of collected data can potentially increase the difficulty of integration, understanding and sense-making of the data [149].

This increasing interest in personal health and fitness data has not been restricted to the industrial sector and tracking devices and apps developers. It has received increasing interest from academic researchers from different fields, such as digital health and HCI, and several

¹<https://www.fitbit.com>

²<https://www.myfitnesspal.com>

workshops (e.g., [3, 4, 5]) have been organised to discuss various aspects of health data.

The significance of this research lies in the type of data it considers. It focuses on an important aspect of personal data: health and physical activity. However, **physical inactivity** is also a crucial aspect to one's health, as it contributes to the development of various chronic illness and lifestyle-related health issues such as diabetes and heart disease [167]. Therefore, personal health data should not be restricted to physical activity data only or data collected from one device or app, other factors such as diet and nutrition, sleep and family medical history are important to gain a better insight and a comprehensive picture of one's personal health [148].

Visual encoding of information has always played a major role in communication and, when used appropriately, it can become a powerful tool to guarantee access to information. The visualization aims to support people's understanding of data by exploiting a human's visual ability to recognise patterns and spotting trends and outliers [108]. Well-designed visualization of data could improve comprehension, memory and decision-making by replacing cognitive operations with primitive perceptual operations [108]. Data visualization is an important stage in any personal informatics model, and it can help in understanding complex data [65].

1.2 Motivation and Vision

People who track their data, view and explore the collected data using information visualizations. The self-tracking devices and their apps/dashboards support different forms of visualizations. However, these visualizations are simple and provide minimum interpretations; for example, low-level data view of step count every couple of minutes or high-level summary of step count over a day or month [60]. Therefore, they do not always fulfil users' needs in the data [135]. Providing higher-level inference is significant to help people obtain complex correlations from the data, this may result in gaining further value from the collected data [60].

Many research studies have focused on finding new methods that help people in gaining new insights from different types of personal data [60, 66, 149, 218, 222]. Other types of research have explored people needs in personal tracking devices and the limitations in the current methods [44, 45, 134, 135, 177]. Previous studies have shown that there should be visualization methods

that help participants find relationships between different personal data variables; combining multiple personal data might help people develop a better understanding of patterns in the data and how they relate to each other [60]. It has also been found that people are interested in identifying relationships between different types of personal data to identify the factors that impact their behaviour [135]. Although combining data and analysis has been explored in the literature, there remain a number of research questions about the design challenges regarding multivariate data in personal informatics that require further research, such as identifying the best methods for representing correlations [122].

A major challenge in combining the data into one visualization is that personal health data variables are different in their types and are measured using different units. For example, the step count is discrete and refers to the total number of steps taken during the day or at a specific time point (e.g., a specific hour), while Heart rate is time-series data that is recorded on continuous-time points and they also vary in their possible values. These differences in the data types increase the challenge of creating a visualization that combines those different variables due to the difficulty of mapping the data into a unified scale. In addition, there is no evaluated method that addresses the issue of visualizing the multivariate variables. Designing such visualization requires a deep understanding of the data types, users' needs and tracking purposes, and visualization principles.

Personal health and physical activity data are often presented on the dedicated platforms over different time frames (e.g., day or a week). Visualizing time-dependent data is a significant topic in information visualization, and it is challenging due to the complexity of time representation [128]. Known visualization layouts for representing time-dependent data are Linear charts (using line, bar or color encoding) and Cyclical layouts (e.g., star and rose charts) [128].

For time-dependent data, Linear charts are possibly the most popular and traditional method to represent a time series, where the time is represented on a Linear axis and the quantitative data on the other axis [128]. The Radial or Cyclical chart on the other hand, provides a compact representation and people's acceptance of the design might depend on its familiarity to the users and their working habits [128]. The Radial design provides an intuitive understanding of the circular pattern in the data and is aesthetically appealing [36]. Therefore, using a Radial design while imitating an analogue clock metaphor might be easier for people, as it utilises

their skill of reading time on an analogue clock. In addition, the hours in the analogue clock may facilitate referring to the presented data for each hour.

Moreover, the Cyclical timeline has an advantage over the traditional Linear timeline in terms of detecting temporal patterns [92]. Finding patterns in the data is a task that involves identifying and comparing patterns that represent behaviours over multiple reference sets [12]. There are several methods for decomposing time-series data into components (e.g., cyclical, seasonal or irregular components) [12]. With respect to the personal health data, the activity patterns could be depicted within days, weeks or months. For example, a detailed day representation of personal health data shows the person's activity patterns during the day. The detailed day representation could help in identifying and comparing patterns, which represent the person's behaviour over the hours of the day. The hours of the day are the multiple reference sets according to Andrienko and Andrienko [12]'s definition of pattern.

Since the difference between the Radial and Linear layout has been studied in the context of time-series data over a long period of time (i.e., weeks and months) such as [6, 32], few studied the difference in case of hourly distribution within days (e.g., [238]). Evaluating the difference between these layouts is important in representing multivariate personal health data, especially as these data often viewed on a daily basis.

In order to address the highlighted issues, this thesis proposes multiple visualization alternatives which rely on the different options of data encoding, visualization methods and layouts. The design considers the use of the Line and the Bar encoding in representing multivariate health data in the traditional Linear format and the Radial format utilising a clock metaphor.

Therefore, the **overall vision of this thesis** is to investigate the best visualization designs for representing multivariate personal health data through developing and evaluating possible visualization choices in order to provide visualization recommendations and insights for future visualization designs.

To achieve the research vision, some important requirements should be investigated. For example, the research should explore the data to understand aspects such as important data variables for the users and how they are represented. These are discussed in the following sections.

1.3 Definitions and Terminology

This thesis uses various terms. This section defines and explains these terms. The definitions are derived from different resources and were adapted to suit the domain of the thesis.

1. **Personal Health Data:** Personal health data in this thesis refers to the data collected by embedded sensors in smartphones and wearable devices that gather different types of health and physical activity data, such as step count, burned calories and heart rate (HR).
2. **Multivariate Data:** This is defined as “data comprising at least three covariates per observation” [212] (p. 147).
3. **Multivariate Personal Health Data:** In this thesis, the term of ‘multivariate personal health data’ is used to refer to the dataset used in the visualizations. According to the definitions 1&2, it is the personal health data collected from the wearable health-tracking devices and apps, which involves measures of three or more data variables per observation (i.e., time point). In this context, the multivariate personal health dataset used in the visualizations is composed of three variables (i.e., steps, calories and heart rate) plus time.
4. **Dataset Complexity:** This includes the number of objects, attributes and relationships between them [116]. The dataset complexity here refers to how many objects need to be observed to solve the given task and how distinct the target is from the distractors. The dataset description can be found in section 6.3.3.1, where the dataset complexity regarding each task is discussed in detail.
5. **Task complexity:** This refers to the load (perceptual and cognitive) incurred by the task, including how many objects are involved and the interactivity between them [116]. In this thesis, it refers to how many variables are involved in solving the task, how many objects (data items to observe) and how many perceptual and cognitive component tasks are involved and how they interact. Task complexity is not related to how the task is introduced to the participant or how it is constructed.
6. **Visualization Readability:** Readability is defined as being easy and interesting to read

[1]. This is a term used in the evaluation study to describe the clarity and the ability to read the information from the visualization.

7. **Visualization Attractiveness:** According to the Oxford dictionary [1], attractiveness is “the quality of being pleasant to look at”. Although attractiveness in information visualizations could be defined differently according to the context (e.g., aesthetic aspects or engagement), in the context of this thesis, ‘visualization attractiveness’ is used in the evaluation to describe the overall aesthetic aspect of the visualizations [150].

Visualization *readability* and *attractiveness* definitions are used in this thesis to allow the participants to provide their feedback about the developed visualizations.

1.4 Research Aims

The aim of this thesis is to design and evaluate visualizations for multivariate personal health and activity data, all the following aims are sub-aims that are included in the basic aim.

- **RA1:** Understanding users’ behaviour and preferences in health tracking (e.g., the type of data they track) and their preferences in different visualization designs. It includes aspects such as the most tracked data and used devices, how frequently people check the tracked data (e.g., daily or weekly), and identifying people’s interest in the visualization of different types of health.
- **RA2:** Identifying the current visualization methods provided by health tracking devices and their associated apps/dashboards in addition to the related literature, and identifying the limitations in these visualizations considering design aspects and the users’ needs.
- **RA3:** Developing different visualization design alternatives for representing multivariate personal health data using options of visual layout and visual encodings to address the identified limitations of the current visualizations methods.
- **RA4:** Evaluating people performance when using the developed visualizations by implementing a set of visual tasks that are related to the real health dataset to identify the most effective design for the data and the implemented tasks.

- **RA5:** Evaluating people's preferences in the developed visualizations before and after using them to gain better insight into the visualizations and the design elements. This also includes comparing users' preferences in the visualizations with how they perform.
- **RA6:** Providing design implications for the visualization of multivariate personal health data that are presented on desktop displays. These design implications are derived from the analysis of the evaluation study.

1.5 Research Questions

This overview of the thesis and research problem presented earlier highlight a set of concepts that need to be explored in the thesis, including the visualization design principles and evaluation methodologies. This subsequently raised several research questions that need to be addressed through different research stages in order to achieve the research aims. The main research questions are as follows (Chapter indicated as CH):

- **RQ1:** What types of health data do people most track? What is their behaviour in tracking, their needs, requirements and preferences in the visualization of the tracked data? (CH2 + CH3).
- **RQ2:** What are the visualization methods supported by most popular tracking devices and their companion apps and dashboards? What are their limitations? And what needs to be addressed? (CH2 + CH4).
- **RQ3:** What visual layout could be used to support interpreting and identifying daily activity patterns in the data and to provide a relevant context? Would a design informed by data ontological interpretation, e.g., Radial clock layout for daily data, facilitate understanding and interpretation of data and performances of visual tasks? (CH5 + CH6).
- **RQ4:** How can multivariate personal health data be visualized? What are the methods that could be used to integrate the data into one visualization? What could be used as a visual encoding method to represent the data? (CH5 + CH6).

- **RQ5:** What are the differences between visualization methods that could be used to represent multivariate personal health data in terms of users' performance? What are the most effective methods for each of the implemented tasks, and what are the least effective ones? (**CH6**).
- **RQ6:** What are the users' preferences in visualizations? What factors (e.g., performance and familiarity) could impact the preferences? In order to construct design guidelines for personal health visualizations that incorporate both preferences and performance (**CH7**).

1.6 Research Hypotheses

During the research stages, a set of hypotheses were constructed based on previous studies outcomes on the visual perception and evaluation of similar tasks, in addition to the results from the preliminary evaluation study. The main hypotheses are listed below and are based on the selected visual tasks for the evaluation study. The visual tasks are explained and refined in Chapter 5 and 6. The number in each hypothesis name refers to the task, the first part (i.e., the a) is related to a comparison between the two layouts, and the second part (i.e., the b) is related to either the type of the visual encoding or the visualization method. The hypotheses are related to the outcomes of the evaluation study, which are associated with the research question **RQ5**.

- **H1a:** It is hypothesised that using the Radial layout (clock metaphor) to represent activity data during the parts of the day (morning, afternoon, and evening) allows users to compare the data and identify trends faster than when using the Linear layout.
- **H1b:** Using Line chart encodings within the layouts facilitates finding trends in the data over time, therefore, enabling a faster performance of the task than the Bar chart.
- **H2a:** For determining a specific value, it is hypothesised that this task is more accurate and faster in the Linear layout than in the Radial layout.
- **H2b:** In terms of the chart type, it is hypothesised that determining value task is more accurate and faster with the Bar chart than with the Line chart on both layouts.

- **H3a:** The comparison of visualizations over two days is faster and more accurate on the Linear layout than the Radial layout.
- **H3b:** The comparison of visualizations of the multivariate dataset over two days is faster and more accurate on the Stacked design visualizations than the Overlapped ones.
- **H4a:** The Linear layout is faster and more accurate at finding the relationship between two variables from the multivariate dataset than using the Radial layout.
- **H4b:** The Overlapped chart is faster and more accurate at finding the relationship between two data variables from the multivariate dataset, compared to the Stacked chart.
- **H5a:** Locating maximum values occurs faster and more accurately with the Linear layout than the Radial layout.
- **H5b:** Locating maximum and minimum values occurs faster and more accurately with the Bar chart than the Line chart.

1.7 Research Methodology and Thesis Structure

Users are a significant element in designing and evaluating the visualizations. Therefore, to achieve better applications, users and their requirements and tasks should not be ignored; as is often the case in any evaluation study of visualization systems. Tory [227] explained research methods for evaluating information visualization and described their appropriate data collection procedures, which are usually applicable across different methods. These include: observation; performance measurements; log analysis; surveys and questionnaires. Each type of evaluation methods provides different types of results and offers a different perspective. Sole dependence on quantitative methods does not provide a complete understanding of the problem; however, combining quantitative and qualitative methods (e.g., through a mixed-method approach) can help to answer the research question by providing a more exploratory enquiry [227].

In this thesis, each chapter uses a specific methodology that is suitable for the research question the chapter addresses and the type of the outcome. However, the research methodology could be contextualized using Munzner Nested Model [155]. According to the model [155], the

problem characterization is presented in Chapter 2, 3 and 4. The data types and task abstraction is covered in the literature review and discussed in detail in Chapter 5, while the visual encoding design is discussed in Chapter 5 and 6. The main contribution of the thesis falls in the encoding design level in Munzner's Model. The thesis presents a laboratory study to evaluate different design choices that are developed to enhance the interpretation of multivariate personal data. The evaluation is based on previous visualization research, which has studied the effectiveness of different visualization designs through a visual task-based evaluation study such as [28, 107, 211]. It compares the visualizations by measuring task completion time and accuracy (i.e., error rate) when the user performs a set of tasks on each visualization. The visualizations are evaluated in accordance with a taxonomy of tasks for multidimensional visualization [234]. In addition, users' subjective feedback and preferences in the visualizations pre and post the experiment are collected using questionnaires to justify the design choices.

Task completion time, accuracy and preferences are important factors to consider when evaluating different visualizations and when providing guidelines. These factors have been used to evaluate the visualizations by many previous studies such as [32, 186, 211]. Collecting participants opinions and preferences using surveys or semi-structured interviews is a common qualitative approach that usually takes place alongside laboratory experiments [40]. The thesis is divided into three stages, as shown in Figure 1.1 and explained below.

- **Investigation and Gathering Information:** this stage collects the required information for the research. It explores personal health data, the process of health tracking, people needs and preferences in the visualizations of the data, how the data is visualized on the dedicated platforms and what are their limitations. These requirements are addressed through the literature review in Chapter 2 and the studies in Chapter 3 and Chapter 4.
- **Visualization Development and Evaluation:** the proposed visualization designs are developed and evaluated in this stage. The evaluation follows a two-step method, starting with a preliminary study that leads to the design of the main evaluation study. The evaluation includes quantitative and qualitative measures of participants' performance and preferences in the designs, and are presented in Chapters 5, 6 and 7.
- **Interpreting Results and Outcomes:** this stage discusses the results and outcomes of

the conducted research. It puts the results and outcomes in a structured list of design implications, recommendations and insights which direct future visualizations design for personal health data.

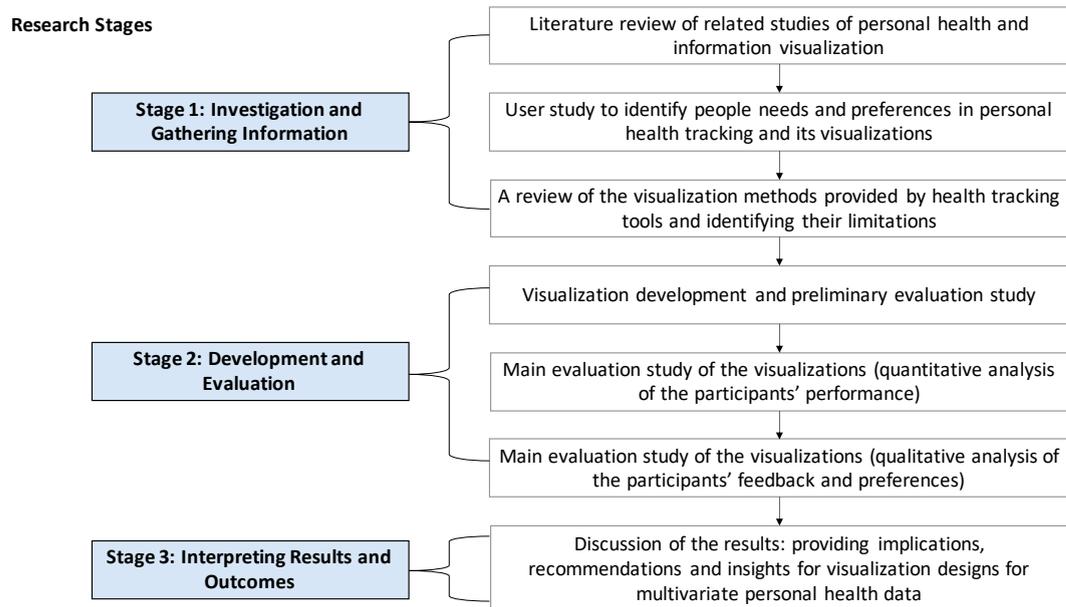


Figure 1.1: PhD Thesis Research Stages and Methodology Diagram.

1.8 Research Contributions

Answers to the research questions are the contributions of the thesis and listed as follows:

- For **RQ1** and **RQ2**, we provide a taxonomy of the visualization methods available on widely used health tracking devices' embedded displays, companion apps and online dashboards (if available). A discussion of the main limitations in the visualizations on the reviewed platforms, in terms of the size of visualized data, visual encoding, visual layout, and sharing from both the information visualization perspective and the personal informatics perspective (i.e., people's needs and requirements). Finally, the study provides suggestions and recommendations to improve the current visualization methods for representing and sharing and offers suggestions for further studies.

- For **RQ3** and **RQ4** is the implementation of eight visualization methods that were designed to address the raised challenges. One of these challenges is the significant differences between the data types and how can they be integrated. The other challenge is the lack of a relevant layout that is suitable for the data. For the visual layout, the implementation uses two visual layouts: the conventional Linear layout and a novel Radial layout that implies the design of an analogue clock and uses coloured sectors to refer to the parts of the day with expressive glyphs (e.g., sun and moon) in order to identify the time of the day faster than the traditional layout.
- For **RQ5**, the analysis of the evaluation results of the developed visualizations provides implications for design according to the tasks and dataset characteristics. The analysis starts by examining the set of hypotheses, which were constructed according to the type of each task. The hypotheses concern the impact of the tested visualizations options for improving the performance of the task, such as the visual layout used, the type of chart and the combining variables methods.
- For **RQ6**, various types of data regarding people's preferences in the visualizations were collected during this research. A significant stage was in the evaluation study, where the participants were allowed to rate and provide subjective feedback about the visualizations after performing real tasks. The results section discusses the relation between the preference and the performance, the tracking devices used and the familiarity with the visualization method. The thematic analysis of the open-ended questions highlights the main features people seek in the visualizations. It also discusses all the evaluated visualizations options such as the visual layout and the encoding. For example, to what extent people may accept a specific layout or visualization method and for what specific tasks they would prefer it to be applied to (e.g., using the Radial layout as a support for the traditional Linear visualization).

1.9 Thesis Outline

The remainder of this thesis incorporates the following structure presented in Figure 1.2:

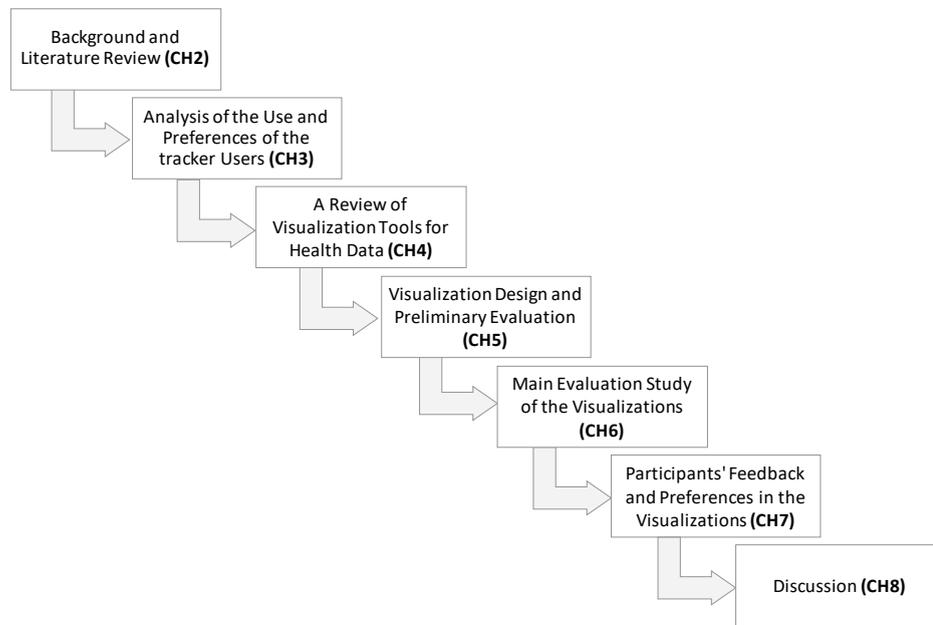


Figure 1.2: PhD Thesis Structure Diagram.

- **Chapter 2:** The background and literature review chapter examines topics from different fields that relate to the scope of this PhD. It begins by describing personal health data, then goes on to discuss related works in personal informatics to understand the visualization's role in the self-tracking process. The chapter also presents an overview of the visualizations methods applied for representing health data, describing the evaluation methods for information visualization in general and personal health visualizations in particular. Other topics related to the evaluation study are also discussed here; such as human factors in information visualization and tasks taxonomies.
- **Chapter 3:** This chapter presents an online questionnaire that is deployed before the visualization implementation phase. It aims to understand more about users' requirements and preferences in different visualizations designs for health data. It also includes questions regarding participants' tracking behaviour and the tools they use.
- **Chapter 4:** Presents a study that was conducted to review popular health and physical activity tracking devices, and their companion apps/dashboards, in the visualization they provide to report on the limitations and research opportunities in personal health data representation. It also provides suggestions for addressing the highlighted limitations.

- **Chapter 5:** Presents the implementation of the visualization designs that are inspired by the studies in Chapter 3 and 4, and validates the evaluation approach. The evaluation compares the visualizations by measuring task completion time and accuracy (i.e., error rate) when the user performs a set of tasks on each visualization. In addition, users' feedback and preferences in the visualizations are collected.
- **Chapter 6:** Following the same methodology set out in Chapter 5, after addressing the issues raised, this chapter describes the main evaluation study used to evaluate the visualizations. It presents the analysis and the results of the quantitative part of the study. It concludes by classifying the visualizations according to their efficiency for each implemented task.
- **Chapter 7:** Presents the qualitative part of the main evaluation study, which is the users' ratings and their subjective feedback of the visualizations. It discusses the relation between visualizations' preferences and performance, and tracking behaviour. The chapter discusses participants' perspectives of the implemented design aspects, and their reflection on the visual tasks.
- **Chapter 8:** This chapter discusses the outcomes of the studies presented throughout the thesis. It reflects on recent updates of the visualizations reviewed in Chapter 4 in light of the raised limitations, our research outcomes and the literature. It discusses the significant differences in the design between the developed visualizations in this research and other similar designs. More importantly, the chapter provides design implications, recommendations and insights for future visualization designs for personal health data.
- **Chapter 9:** Concludes the thesis by providing an overview of the conducted research and its outcomes. It addresses the research questions and highlights the research contributions. Subsequently, it discusses the research limitations, future work, and further directions to be explored based on the presented research.
- **Appendices:** Supplementary materials include additional information about implemented studies in Chapter 5, 6 and 7, including the used questionnaires, users' tasks, participants training and recruitment details in addition to groups and behavioural analysis and charts of the main study discussed in Chapter 6.

Chapter 2

Background and Literature Review

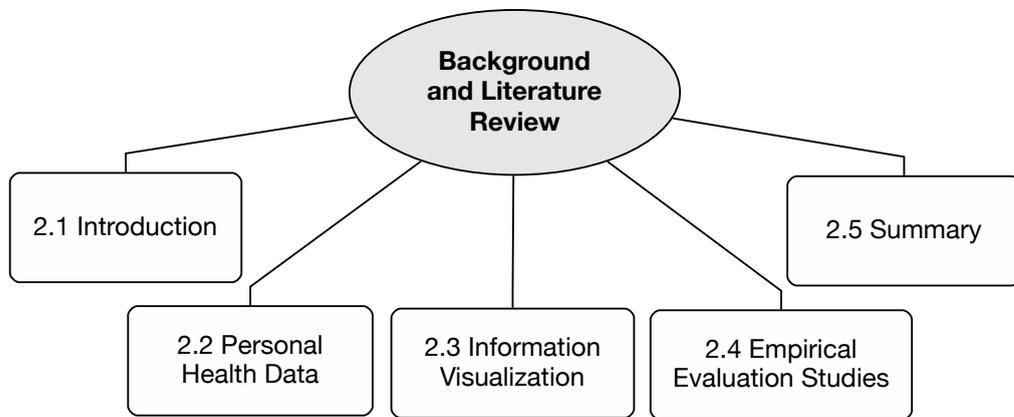


Figure 2.1: Chapter 2 Structure Diagram.

2.1 Introduction

Self-tracking and documenting information about aspects of one's personal and daily life has a long history. It is an effective method which helps us to learn more about ourselves, rather than depending on our limited memory. Recent technology in mobile devices and the advancements in sensors and wearable devices have significantly reduced the effort required to record such data. Although these advancements have contributed to a massive generation of personal data, they have produced several challenges in managing and extracting usable knowledge from the data. Many research papers have focused on these challenges and built models to understand the process. A well-known study in this field is the personal informatics model by Li et al. [134]. The model includes five stages, which are preparation, collection, integration, reflection and action. According to Li et al. [134], data visualizations is one of the essential methods used to help people to understand their collected personal data in the reflection stage.

This thesis focuses on the visualization of one specific type of personal data: health data. Personal visualization is distributed across different research areas such as visualization, human-computer interaction and personal informatics [114]. The scope of this thesis lies in the intersection of different fields and relies on many topics: personal visualization and its evaluation methodologies, personal health visualization, people's preferences and behaviours of personal

health tracking, basic information visualization evaluation research and the used evaluation methodology including visual tasks taxonomies, the perception of different visual encodings used in the primary statistical charts.

Therefore, the topics of this literature review are classified according to the basic elements in the research and divided into three major parts (see Figure 2.1) as follows:

- **Part 1: Personal Health Data:** The first part of this chapter introduces the health data in general and discusses a significant type of health data: *Personal Health Data*, then it discusses *Personal Informatics* and how this relates to personal health visualization.
- **Part 2: Information Visualization:** The second part of the literature review includes three main sections: (1) *Visualization by Task Taxonomies*; (2) *Human Factors in information visualization*, which discusses topics of cognition and perception, and the last section is (3) *Visualization of Time-Series Data*.
- **Part 3: Empirical Evaluation Studies:** This part presents the visualization methods applied in information visualization in general and in personal visualizations specifically, with a reflection on the evaluation methods applied in part 1.

2.2 Personal Health Data

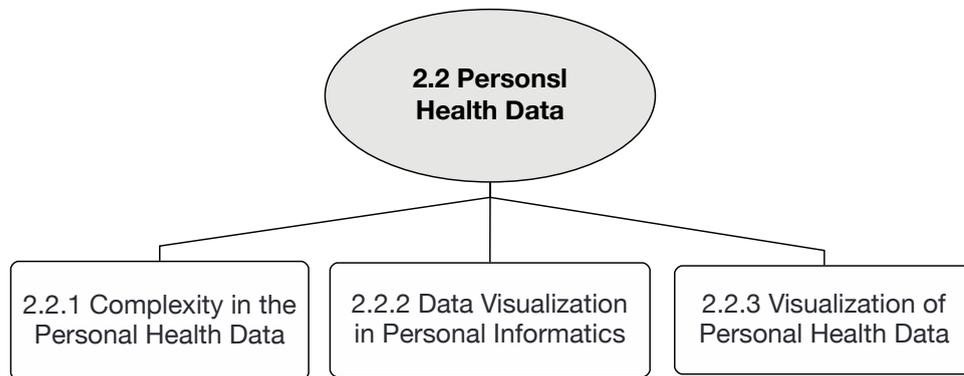


Figure 2.2: Chapter 2 Health Data Section Diagram.

Personal health data is a broad term that incorporates different health-related aspects of individuals. It includes data in different forms and from different resources, such as: patients' health records in clinics, for example summaries of clinical visits, prescriptions and test results recorded by healthcare providers; manually recorded health-related information by people or automatically tracked data using embedded sensors in smartphones and wearable devices. In terms of the data recorded by individuals about themselves, there are numerous aspects of personal data which some people keep, for example: medications; illness-related symptoms; foods and liquid consumptions; alcohol and caffeine intake; mood and logged exercises, as well as many other types of data. The recent advancements in sensing technology, especially those used by phones and wearables, have facilitated the tracking of different health data (e.g., steps, HR, active minutes and calories). Although these devices are fairly affordable, they are also generating a varied collection of valuable health data. This thesis focuses on the latter type: the personal health data that are automatically generated by health tracking devices and apps.

2.2.1 Complexity in the Personal Health Data

Data collected from self-tracking devices is considered to be multi-dimensional data based on how many signals are being tracked [130]. Personal health data is considered multivariate if it

includes measures of three or more data variables (see section 1.3 for definitions and terminology). Several challenges have been identified in personal health data either from the reviewed literature or from dealing with the data in this PhD study, including the lack of accessibility, incompleteness of the tracked data and variability of recording intervals. Each of these challenges is defined and discussed in the following sections.

Accessibility: The type of the data was not easily accessible or found as an open-source or publicly published data; the focus of this thesis is the multivariate dataset. Personal health data is sensitive and could reveal a significant amount of information about oneself, which in turn restricts its availability. The data also required informed consent from the owners to be used in the research. Several studies have highlighted how treasured this type of data is for personal and public health research. For example, Bietz et al. [24], studied the barriers that prevent researchers from using and benefiting from personal health data from different perspectives including the data owners (i.e. the adopters of self-tracking devices) and the researchers and companies that build the tracking-devices. The results of Bietz et al.'s study have identified challenges which have been classified into six areas: ownership of the data; access for research; privacy; informed consent; data quality, reliability and research methods, in addition, there is the challenge of the evolving ecosystem.

Data Variability (Recording Intervals): In some cases, the collected personal health data is recorded using different devices. Even when using the same device, the time interval between recorded one data value and the other usually depends on the physical activity (e.g., it is more frequent during an exercise than when one is sedentary). Data recording interval affects data visualization and analysis (e.g., when analyzing or visualizing the data according to minutes, hours or days interval).

Incompleteness: One of the major challenges in understanding and interpreting personal health data that it is typically incomplete [21, 218, 219]. Many reasons could cause data incompleteness, such as faults in the tracking device or because the user is not wearing it. Tracking success requires that the user wears the tracking device continually. Boll et al. [27] stated that the tracking device needs to be easily integrated into one daily's life to achieve successful personal health tracking. However, some people forget to wear or charge their device, or switch between different devices, or sometimes they do not fit in with the person's lifestyle, match

their personal taste, or in some cases they are not allowed to be worn [27].

Multivariate Data Variables: the tracking devices collect various types of different data, all of which are measured using different units. For example, the step count is discrete and refers to the total number of steps taken during the day or at a specific time point, such as a specific hour. The number of calories burned is considered similar to the steps in terms of the type, but they have different data ranges (e.g., the user that walks the recommended 10,000 steps in one day will burn 500 calories [159]) and this shows how the numbers are different between these two variables. Sleep data is recorded and represented as ranges each day. Heart rate is time-series data that is recorded on continuous time points. What follows are definitions of the major data types recorded by activity tracking devices, smartwatches or phone apps.

- **Calories:** the calorie is defined as a unit of energy used to represent the energy produced from foods and energy consumed by the body [20]. The unit calorie (Cal), is usually abbreviated as (kcal) [20]. The energy is required to keep us alive and for our organs to function normally [162]. Our bodies use the energy we get from food and drinks for every movement from a simple activity (e.g., breathing) to more intense activity such as running [162]. Burning calories depends on how much the person is physically active, and it is affected by other factors such as weight and age [162].
- **Distance:** is the distance travelled by the person during a specific exercise (e.g., running, walking or cycling) or over the course of a day, and is calculated as km or mile per hour.
- **Heart Rate (HR):** HR is defined as “the number of heart beats per minute” [124]. The average heart rate per minutes for an adult at rest is 60-80 per minute; however, other factors such as exercise, illness, and emotion may affect this average rate and elevate it and produce faster rate [2]. The standard unit for HR measurement is (BPM), which is beats per minute [220]. Other HR measurements, such as the maximum HR (i.e. calculated as $220 - \text{user's age}$) and target HR (i.e. a value between 50% and 70% of the maximum HR) are calculated to increase the user's fitness and strength safely [34]. Some devices also track HR variability, which is defined as a measurement of the variability of the intervals between consecutive heartbeats over time; these can indicate stress and its related issues [160].

- **Sleep:** the sleep data is often collected by apps and trackers as the number of sleeping hours during the day, it is usually represented as a range between bedtime and waking time. Some apps add information such as sleep quality (deep or light or times being awake). For longer periods (such as a week or a month), sleep is usually represented as the average of sleeping hours during that period.
- **Step Count:** is the total number of steps the users take during a day. Step count has become more popular since 2010, when use of the activity tracker increased [18]. The step count is a preferred metric for measuring physical activity because of its various advantages, such as it being easy to understand by laypeople, easy to track and its association with other health variables [18].

The list of data variables is much longer than those listed here. It includes: active minutes that calculate the number of minutes the user is active during the day; the pace that calculates how fast the user is moving; the HR zones that are supported by some trackers, such as Fitbit, the calculation of which is based on the maximum HR.

The field of personal health is broad and overlaps with many fields, such as mobile Health (mHealth), HCI and Information Visualization. In recent years, researchers have explored personal health tracking from different perspectives. The following sections reflect on and discuss the related literature from two research areas, beginning with personal informatics and followed by the visualization of personal health data.

2.2.2 Data Visualization in Personal Informatics

Personal informatics is defined as the process of gathering and reflecting on personal data to gain better self-knowledge [135]. Furthermore, personal informatics systems enable people to collect and review personal information for providing data-driven self-insight to promote better behavioural change [235]. Li et al. (2010) [134], developed a five stage-based model of personal informatics: preparation; collection; integration; reflection and action and they listed the barriers people face at each stage. This model was derived from surveys and interviews with users of different personal informatics tools, with the aim of studying their practice, motivations

and issues in each stage. They identified some problems users faced with the visualization of their data during the reflection stage, which affects transition to the action stage, such as failing to use visualizations effectively to extract knowledge from the data. Li et al. [135] noted the importance of referring to different fields (e.g., information visualization) to resolve issues within each stage. This is performed by identifying people's needs and requirements in the data (e.g., why do they collect the data? and what do they want to know?) and what is the size of the data to be presented to the user based on their tracking purposes. As the purpose of the visualization and the size of the data is essential in any visualization design, the discussion of the literature in this section is classified into these two important categories: tracking purposes and requirements and data granularities.

Tracking Purposes and Requirements: As developing effective tools for supporting self-reflection requires sufficient understanding of people's self-reflection needs, Li et al. [135] built on their previous study [134] and interviewed 15 people who used different types of self-informatics tools to explore and identify the questions people have regarding their personal data. They also discussed what tool the participants' used to reflect on their data, how they used it and what issues they encountered. Their study resulted in identifying six types of questions that people usually ask about their personal data during two non-static phases of reflection, i.e. discovery and maintenance. Those six types of questions relate to: status; history; goals; discrepancies; context and factors. Although the authors discussed how the participants used their tools to reflect on the data and to answer their questions, there were no details on the visualization methods used and their limitations.

Choe et al. [44] focused on the reflection stage and they analysed 52 video recordings of Quantified-Selfers talks in Quantified-Self meetings with the goal of understanding the practice of self-trackers for improving the designs of self-informatics tools. The quantified-self is defined as any individual that self-tracks any biological, physical, behavioural or environmental information [112]. Their study explored important aspects of self-reflection by analysing how the Quantified-Selfers analysed their data and what visualization they used [44]. In a follow up research study [45], Choe et al. analysed 30 recorded videos of Quantified-Selfers presentations to identify the types of insight and what visualization method the presenters used to convey their insight. They identified 8 data-driven types of visualization insights (detail, self-

reflection, trend, comparison, correlation data summary, distribution and outlier).

Rooksby et al. [185] interviewed 22 participants, who were users of different physical activity trackers to understand their self-tracking behaviour. They identified five different types of personal tracking styles based on the participants' tracking purposes: directive, documentary, diagnostic, collecting rewards and fetishised tracking (i.e., tracking for technology and data collection purposes).

Other studies focused on the adherence and the abandonment of the tracking to understand people's tracking behaviour. For example, Tang et al. [219] provided a definition for tracking adherence, analysed its impact, and reported it across different datasets. Clawson et al. [46] conducted a study to investigate the reasons for abandoning personal health-tracking technologies. They revealed different reasons, such as upgrading to a new device, the goal of tracking was achieved, or their health status had changed. They also found that the abandonment not always a result of users' dissatisfaction with the technology.

Rapp and Cena [177] implemented a diary study of 14 participants who had no previous experience in personal informatics. The data were analysed with the aim of identifying the issues the participants faced in order to highlight what needs to be solved by the personal informatics tool. The issues fell under four basic categories: issues in tracking; managing; visualizing and using the data. They sought to find relations between various data (e.g., covariation and causation). With respect to data integration, the participants reported a lack of integration between different behavioural data from the same sourced or data collected from multiple sources. This prevented the participants from proceeding further in self-experimentation as time was required to reflect on the data and to identify meaningful connections. In terms of data visualizations, the participants stated that the provided visualizations with self-tracking tools were too abstract and did not reflect their expectations regarding their data. Also, the visualizations did not support navigation and did not help in deeply understanding the data. They found that the participants faced difficulty in data interpretation and visualizations and a lack of suggestions regarding using the data, which resulted in a decreased interest in tracking.

Data Granularities: A significant aspect of designing visualizations for personal data is defining the data size and time-intervals (i.e. visualizing the data based on minutes, hours or days). Visualizations for different time granularities can be used to support different purposes. As ex-

plained by Li et al. [135], people were interested in short-term data to show their current status, while some were also interested in historical or long term-data to see how the data changed over time. Rooksby et al. [185] found from their interview study that the participants were mainly using the trackers to represent short-term data (i.e. data collected on the same day only). This indicates that users were interested in meeting their daily goals. In terms of personal health tracking devices and apps, they considered the data from days', weeks' and months' perspective [27]. However, they should have also considered a lifetime perspective to support people's health throughout their lives [27]. Other studies such as Epstein et al. [63], who explored design opportunities to support people who have stopped practising self-tracking, and Li et al. [136], who demonstrated how contextual information could detect factors that affect behaviour, also showed that people are interested in long-term personal data collected from tracking-devices over a long period of time.

Huang et al. [114] provided an extensive survey for personal visualization systems that support personal data analysis for non-professional users. The survey explored the design space and highlighted key challenges in the field, including the adoption of personal visualization and personal visual analytics technologies, sharing and privacy, and evaluation. They defined the design dimension extracted from the survey (e.g., data, context and insight). They defined two types of insights from the reviewed papers: the first was actionability, which is the insight obtained from the tool to direct actions (e.g., reminding the user when to exercise, reflection on progress and engaging some behaviour). The second was the automated analysis, such as data mining methods. This survey [114] shed the light on other significant aspects of personal visualization such as sharing, which was investigated by Epstein et al. [61] and Thudt et al. [222].

Others research projects focused on users' needs and requirements in personal health apps (such as [167]). Olmsted et al. [167] expanded on previous work [145], which developed a web-based tool to provide individualised physical activity intervention for inactive adults that allows involving other people (e.g., healthcare providers, family and friends) to encourage physical activity. Olmsted et al. [167] described the major themes that classify users' requirements and preferences that are necessary for personal health applications. These themes included an intuitive and user-friendly interface and customisable tools that allow setting preferences and

viewing their activity progress over time.

Finding correlations between different personal data have been found to be beneficial to increase self-understanding [21, 122]. Bentley et al. [21] built the Health Mashups system, which was designed to identify significant correlations between different types of data over time, including (weight, sleep, steps, food intake, mood, location, and calendar data) and presented to the users in natural language. To evaluate the system, they conducted a field study with 60 participants to use the system daily for 90-day. They found that the participants were able to utilise the observation data to make a change in their daily lives. Bentley et al. showed that the interactions between health data and context provided by the system increased self-understanding that guide behaviour changes. Jones and Kelly [122] built on Bentley et al. [21]’s work. They investigated the problem of information overload in personal tracking and informatics. They designed a three-month study with 20 participants who used the commercial tool ‘Exist’¹, which was designed to combine and analyse personal data from different tracking sources, and they applied a filtering mechanism to deal with data overload. Although Jones and Kelly [122] found that the participants were able to predict around the half of the gained correlations correctly from using the tracking app, the rest of the gained correlations either disproved participants’ previous predictions, or presented new information that the participants had not been able to predict earlier.

2.2.3 Visualization of Personal Health Data

The availability of health data generated by self-tracking devices has started to increase patients’ awareness and sense of responsibility regarding their health conditions [207]. These devices (e.g., Fitbit) also support visual representations of the tracked data which can be viewed on their companion apps and/or online dashboards, and in some devices, the visualization could also be displayed on the device’s screen. The visualization methods used to transform the data into an easy-to-understand format are often based on traditional chart formats, such as a Bar chart and a Line chart. Although these charts provide awareness and allow searching for specific information and details [66], they have some limitations and they need to be evaluated

¹<https://exist.io>

within the context of the personal health domain. One example of such an evaluation is the study carried out by Tong et al. [226] who evaluated a Fitbit Bar chart with a circular ringmap and Virtual Pet visualization. The evaluation was conducted taking into account five factors: Readability; Comprehension; Awareness; Attractiveness and Empathy. In their study, Tong et al. showed that the efficiency of the visualization is unrelated to the participants' subjective evaluation and that each visualization method provides a different view of the data. Their results shed light on the significance of having new visualizations that serve these factors [226].

Different research projects have studied the visualization of personal health and physical activity data and developed different visualization designs for the tracked data to support reflection. These studies had different design purposes and goals, some focused on the complexity of the data [60, 63, 149], whilst others designed abstract and metaphorical representation of the data [49, 66, 139, 179, 225]. Others looked at visualization designs for promoting behaviour change [11, 19, 103, 123], whilst some studies were carried out to explore the designs' ability to provide context [115, 130, 218], others looked at visualization designs for specific uses, such as for patients or healthcare providers [132, 176]. In this section, the studies are organised according to this classification.

Visualization of Complex Health Data: Epstein et al. [60] investigated the motivations for self-tracking and examined a new approach to help people (self-trackers) discover valuable and actionable correlations in their data. They defined a set of cuts (i.e. selected subset of tracked data that shared some related features) over location and physical activity data and used different visualizations to represent these cuts. Meyer et al. [149], discussed the challenges of visualizing heterogeneous health data on the small display of mobile devices. They developed and evaluated different designs of both metaphoric and quantitative visualizations to provide an overview and details on demand for different health data. They proposed different types of visualization to support short-term and long-term representation (e.g., tree metaphor, Bar and Pie charts and bubble graphs). Their findings showed that people are interested in finding more complex correlations in their health data that could not be configured using simple methods of data representation [149]. Other research studies focused on personal data incompleteness (e.g., [63]). Epstein et al. [63] focused on investigating how to design visualizations that support people who lapse (i.e. defined as people who stop tracking their data over time)

in tracking. With respect to the data representations, they considered the selection of visual cuts, that transform complex multivariate data into simple and understandable visualizations, and the framing technique. In addition of providing design recommendations for this specific type of users, they found that people who have tracked their data consistently are more likely to understand their habits, and they also may prefer presenting longer-term visualizations.

Abstract Visualization: Other research focused more on developing abstract representations, such as the UbiFit [49], which used a garden metaphor to represent different types of physical activity data and Fish 'n' Steps [139] that linked each user's step count to the growth and activity of a single virtual fish. These living metaphors were developed to encourage the users to be more active, as the users became emotionally attached to these metaphors. Emotional engagement is an important factor in personal visualization that promotes personal motivation toward lifestyle improvement [226]. However, these designs could also allow users develop negative emotions, such as guilt, as reported in Lin et al.'s study [139]. These two aesthetic and non-literal representations could provide high personal data encoding which is essential to preserve users' privacy when sharing personal data [114].

Fan et al. presented Spark [66], a web application that supports different types of visualizations for the Fitbit step count data. The research aimed to find an appropriate method for visualizing physical activity data that overcomes the weakness of statistical charts and living metaphors. They designed four abstract visualizations (Spiral, Rings, Bucket and Pollock) in addition to the Bar chart. The abstract designs were inspired by the informative art provided by Redström et al. [179], because its effective graphical feedback is less discouraging than living metaphors. The study showed that the participants preferred the abstract visualization for its aesthetics and glanceable display features, while the Bar chart was required for gaining a detailed view of the data or looking for specific information. Tong et al. [225] proposed the use of a circular graph as a visualization technique to represent long-term physical activity data to provide users with a better understanding of their data and to allow the comparing and identifying patterns which are limited in the widely used traditional information visualization techniques such as the Bar chart and Pie chart. They also proposed using animated graph approach to express a patient's pain, this used abstracted movement of a Jellyfish to represent six different types of pain. Tong et al. suggest that using these visualizations could provide doctors with valuable

information and make it easier for them to interpret clinical data. In addition, the visualization could encourage patients to track their data to manage the level of their activity and to share the visualization with their doctors and clinic professionals. However, these visualizations were not evaluated.

Visualization for Behavioural Change: Anderson et al. [11] explored the use of the mobile phone application as a tool to promote physical activity. It aimed to design and develop a system that increased users' awareness of their current activity levels and to motivate them toward achieving the daily recommended level of activity. The daily activity calculated could be shared and compared with others. It also provided a historical view of the activity for the last seven days to support activity monitoring. The visualization applied in the application are a Bar chart for representing week activity and an animated character to represent the user's current activity level. They conducted a trial study which included interviews with the participants to explore whether the use of the application and the sharing feature could increase a user's awareness and promote an active lifestyle [11]. Many of the participants reported that seeing their activity level was exciting and motivating, saying it also helped in increasing their awareness and also reported that competing with others was fun. However, the evaluation study was for a short period of time (one week) and did not reflect users' experience for longer use of the system, especially in the case of sharing [11].

Harries et al. [103] designed and implemented a field trial study of a new smartphone application, "bActive", to promote walking in daily life. It was designed to suit users with different activity levels. The app tracked users' step count using the phone's built-in accelerometer and provided users with always available visual feedback about their level of activity, including (step count, distance and calories) of the current day, yesterday, last week and the complete activity history. To conduct the study, Harries et al. used three different versions of the app: one being "null", that is with no visual feedback; the second "partial", that presented the activity data for the user only and the third was "full", a version that showed the activity level for the user and the social group average. The visual display presented an animated walking avatar that represented the users (the grey representing the average activity of the members of the group, while the green represented the user). A feedback message about how the user was performing in comparison to the group's average was displayed on a banner below the avatar. In

addition, the app provided a parallel coordinate view to allow users to compare their progress with their group's average as presented in the past week view. The result of the study showed that using the 'always-on' phone app increased activity level (walking). This study indicated that 'always-on', accelerometer-based smartphone apps could increase walking amongst males by around 64%. However, there was no evidence to show the impact of social comparisons to increase walking.

Bauer et al. developed ShutEye [19], a mobile phone application, that was developed to improve individuals' sleep quality and to provide awareness about how to enhance and maintain healthy sleep habits. The application [19] promoted awareness of major activities known to affect sleep. It advised users when they could engage in these activities based on their desired sleep routine. The app implemented a 24-hour timeline representation as a mobile phone background, and it represented the various activities that affect a user's sleep during the day. Kay et al. [123] implemented and evaluated Lullaby, which is a system that was designed to help people to identify the factors that might affect their sleep quality and cause disruption, such as light, noise and temperature. They used Linear charts (e.g., Line chart) to visualize these factors during sleep and suggested further analysis, such as statistical tests to strengthen inferences for self-tracking and monitoring applications.

Context in Personal Visualization: The calendar layout has been used recently to represent different types of personal data. For example, Mennicken et al. [147] developed Casalendar, a tool that integrated the visualization of data collected from smart home technologies into a widely used calendar interface and Huang et al. [115] integrated a visualization of Fitbit physical activity logs (Line chart or luminance) with a personal calendar that was used as an additional layer. In their previous work, Huang et al. [113] first deployed a prior version of the web application by implementing two pilot studies with two different datasets, one with the home energy conversation data and the other with activity data from Fitbit. Huang et al. [115] used the calendar as a feedback tool to help the users consider their physical activity. They performed an eight-week field study to explore users' experiences with the calendar visualization and to investigate the effect of including additional information for understanding the data. The result of the field study showed that the on-calendar visualization provided easy access information to help in understanding their physical activity data [115]. The calendar layout was

used by Pagno and Nedel [168] to integrate visualization of different types of data, these were the physical activity data and the weather. The calendar showed coloured bars that represented the number of steps taken every day and an icon that represented the weather status (e.g., sun or cloud) [168]. However, their work is incomplete (not all the collected data has been visualized), and no evaluation was performed on the proposed visualization. Another calendar visualization to represent physical activity data was implemented by Tang and Kay [218]. Tang and Kay [218]' study focused on long-term tracking data and incomplete data and their role for reflection. They focused on what people (who track their physical activity for a long time) understand about their behaviour and their activity levels. They implemented *iStuckWithIt*, a calendar visualization tool that integrates hourly and daily adherence (i.e. days and hours with and without data) of people's physical activity (i.e. Fitbit step count) in a calendar layout. The study started with an interview with Fitbit users who used the device for at least six months, followed a think-aloud study with the developed tool and then a post interview. The outcomes provided a new understanding regarding the daily, hourly and goal adherence for long-term tracking. With respect to the calendar layout, 81% of the participants commented on the clarity and intuitiveness of the calendar visualization. They also liked being able to see the overview, as well as the details of the individual days.

Additionally, the calendar layout is widely used by several mobile applications and online dashboards to presents users' physical activities (this is discussed in section 4.4.1). An example of this type of layout is the Garmin calendar that represents the types of activity achieved each day and an iOS Activity app that shows the daily activities on a calendar view.

The visualizations by Larsen et al. [130] show a way of representing personal data in a related context. They proposed the QS Spiral, an interactive visualization technique for time-series data which represents quantified-self data by focusing on its continuity and periodic characteristics. The design of the QS Spiral approach is based on a clock metaphor where circles refer to a time span. Zooming in and out is provided for both an overview and a detailed view of the visualization. Two datasets were used in [130], the first is the geolocation data recorded by the smartphone's sensors over four months. The developed visualization for this data composed of three parts: the QS Spiral that shows visited locations in respect of time and date; a tag cloud to highlight locations visited frequently using colour coding related to the Spiral view and a map

view with a time series to show the places visited. The second dataset is the Fitbit step count data of twelve months that is visualized using Spiral rings, which are colour coded according to the number of steps. These visualizations were implemented on an Android app that benefits from the touch-based interface to allow users to interact with the visualizations.

Kim et al. [125] investigated how people design their own personal visual vocabulary to represent their data. They developed DataSelfie, a web-based system that allows users to collect data about themselves and create visualizations for the data. Their user study showed positive benefits of giving people control over the data visualizations including joyfulness, visualizing the main goal of the data collection and how the data affects their lives.

Visualizations for Specific Users: Some of the personal health visualizations have focused on specific types of users or on managing a specific type of health concern, such as diabetes [176], or on visualizations for older adults [132]. There is also a body of research that focused on the presentation of personal health data to be reviewed by health providers, either for remotely monitoring the patients' health [91, 250], or to reduce the time required to review and understand health data during the clinical visits [176], such as the work by Rajabiyazdi et al. [176] which explored the visualizations of health data for people with chronic conditions and interviewed four diabetes patients. They reviewed a sample of their collected data to understand their tracking and presentation process and their perspectives on the data. In response to the collected data, they designed different personal visualizations based on their needs. Genes et al. [91] also focused on the health context when they studied the technical aspects of the integration of personal health data collected from apps and wearable devices to be accessed by healthcare teams to allow an early health intervention.

Although there have been many research articles discussing the visualization of personal health data, there has not been sufficient discussion or research on the existing visualizations provided by commercial health tracking devices and their companion apps. These issues are presented and discussed in Chapter 4. However, in order to assess these visualizations and to design and evaluate new methods to address the highlighted limitations, an understanding of information visualizations, techniques for the type of the data and their evaluation is essential. Therefore, the following section discusses the basic concepts of the human visual system in cognitive psychology, visualizations task taxonomies and the visualization studies for time-series data.

2.3 Information Visualization

“Graphics is the visual means of resolving logical problems”

Jacques Bertin [23] (p. 16).

The role of the visualization in personal informatics was discussed earlier in this chapter (section 2.2.2); therefore, this section (*Information Visualization*) introduces topics that are related to design of the visualization, including human factors in information visualization, that focus on cognitive psychology, tasks and task taxonomies, visualization methods and the visualization of time-series data. These topics were selected based on their relevance to the scope of the thesis. For example, the first section describes how human factors are involved in the design, use and perception of information visualization; the discussion of the visualization task taxonomies are related to the methodology implemented to evaluate the visualization designs; the visualization of time-series data section is related to the data type considered in the thesis. The structure of the Information Visualization section is presented in Figure 2.3.

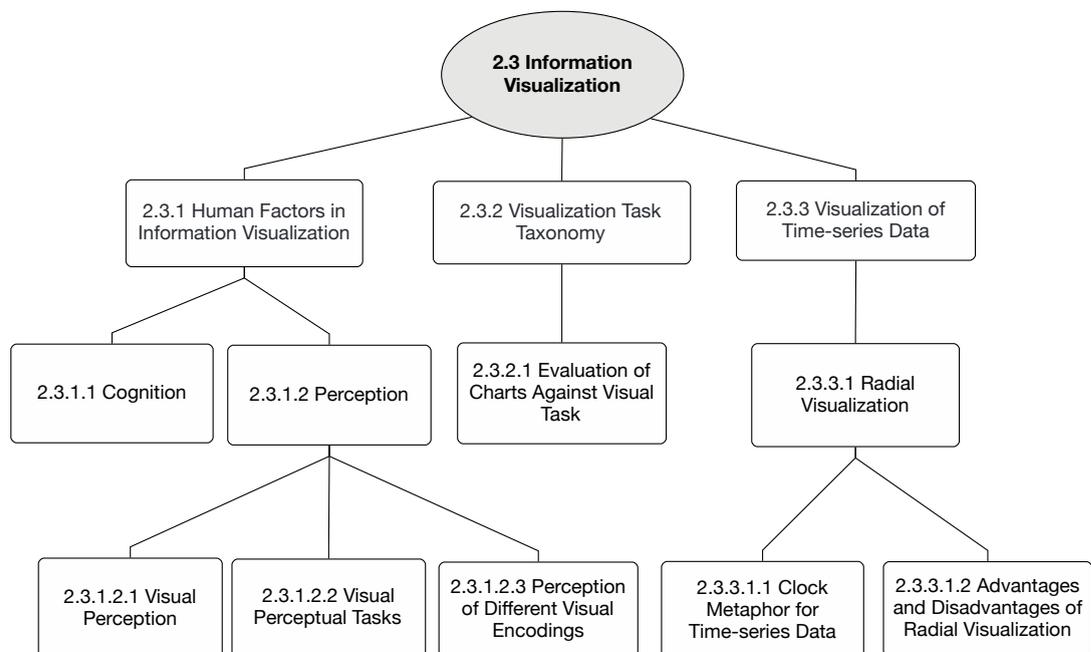


Figure 2.3: Chapter 2 Information Visualization Section Structure Diagram.

2.3.1 Human Factors in Information Visualization

Information visualization can help people with data analysis by visually representing the data, this is called cognitive support [228]. Information visualization uses several mechanisms to provide cognitive support by utilizing human visual perception (e.g., through visual grouping and parallel processing) [228]. For example, appropriate visualization could reduce cognitive effort when performing some visual tasks by utilizing humans' visual perceptual capabilities.

Human factors is a wide-ranging field. This section begins by introducing the field of cognitive psychology to discuss related topics and basic concepts in cognition and visual perception, which the design of the visualizations in this thesis relies on. The section discusses how these concepts are utilised and evaluated in the design of information visualization. In addition, it presents visual perceptual tasks that are usually involved in visual tasks to retrieve information from visualizations, and visual perception studies that evaluated the perception of different visual encodings used to represent information.

Cognitive Psychology is the science concerned with understanding human cognition through the observation of people's behaviour when performing different cognitive tasks [64]. There are several internal processes involved in understanding the environment and deciding on the next action to be taken, these include: attention; perception; thinking; learning; reasoning; memory; language and problem-solving [64]. It is also defined by Sternberg et al. [213] as the area of study that concerns how people perceive, think, learn and remember information. The following is an outline of the background of the main concepts, theories and definitions in the field of cognitive psychology, which this thesis relies on for the design and discussion.

Cognitive psychology is a significant aspect that has been discussed in visualization and HCI research to understand people's capabilities and limitations. A Model Human Processor described by Card et al. [39] is an example that provides a simplified model of the human processing system. The model is composed of three subsystems: the perceptual; the motor and the cognitive system. The perceptual system includes sensors and connected memories buffers, and it manages external stimuli (symbolically coded). The cognitive system includes the processes required to connect the perceptual system with the motor system. It uses the symbolically coded information (by the perceptual system) in its working memory and uses the

information stored in the long-term memory to decide on the action needed to be taken in order to respond. The motor system is responsible for controlling the response (action).

Ware in his books [239] and [240] discussed the fundamental concepts in vision and cognitive psychology that are related to data visualization, including the human visual system, the anatomy of the human eye and how it perceives visual elements. Understanding elementary and secondary perceptual and cognitive tasks is significant for implementing and designing information visualization systems [42]. This section covers the theoretical aspects of visual perception and cognition, which are mainly related to the design of the user studies in this thesis.

Wickens and Carswell [242] defined the characteristics of the major stages of information processing and described processes related to attention, perception, cognition and action selection. They adopted the information-processing model (Wickens and Hollands [243]) as a framework. The model illustrates how the information moves between the human's cognitive subsystems. It starts when stimuli are sensed with the sensory systems (e.g., eyes). The received information then perceived as it interpreted based on stored information in the memory. The perceived information then might be responded to directly after deciding what action to perform to respond. Wickens and Hollands's discussion of the human factors and how the implications of attention (selective and focused) are related to how people perceive the visualization in this thesis and presented in the section on visual perception (section 2.3.1.2.1).

2.3.1.1 Cognition

Cognition refers to the ability of processing the information (e.g., visually perceived information) and applying previously gained knowledge (i.e., from memory) to understand and decide how to initiate responses, it involves attention, learning, language, memory and problem-solving [38]. Examples of cognitive processes are problem-solving, reading, learning and reasoning. Cognitive tasks can be familiar and require only calling information from memory, while other tasks are novel and require instructions and other types of training [184]. Cognitive tasks are usually a component of many visual tasks in the evaluation of different information visualization techniques. The low-level task taxonomy by Amar et al. [9] showed that some tasks' questions involve performing other tasks (i.e., a task question may be a combination of

other tasks). In their analytical task taxonomy [9], they presented some questions that were excluded from the taxonomy because of their overall goals being of a very “low-level” with respect to the taxonomy, as they are cognitive and mathematical tasks rather than analytical. An example of these tasks is the set of mathematical operations (e.g., less than or equal to), which are considered fundamental cognitive actions instead of being “primitive tasks” in the taxonomy [9].

2.3.1.2 Perception

Perception is defined as: “The acquisition and processing of sensory information in order to see, hear, taste, or feel objects in the world and more importantly, it guides an organism’s actions with respect to those objects” ([203] P. 261, [64] P. 35). Perception is also defined as a set of processes that enable recognising, organising and making sense of the information received from stimuli in the environment [213].

2.3.1.2.1 Visual Perception

Human vision is one of the essential sensory channels for a person to receive information. Visual perception is the mental organisation and interpretation of the information received through the visual sensors to achieve awareness of the environment and to understand the involved objects and events [38]. *Stimulus* is defined as an event in the environment (e.g., problem or task) [64]. Visual perception is divided into two stages: the first is receiving external stimuli, whereas the second is the internal processing of the received stimuli [54].

A visual channel is a set of basic visual representations that are used to express various values of a variable and is also referred to as (visual encoding variables or retinal variables) [29]. The visual pathway is the path that the visual information follows in order to be processed in the human perception system [213]. There are two visual pathways, the dorsal pathway (the where) process location information and the ventral pathway (the what) process the identity of the visual stimulus such as shape and colour [213].

The perception system of human has two main kinds of sensory modalities, the identity channels that inform on the shape and the position of something (e.g., identifying the shape of a

circle or a triangle, the colour channel of hue and the position of a region), and magnitude channels which inform on how much is there of something (e.g., length, area, volume and colour channel of saturation and luminance) [156].

2.3.1.2.2 Visual Perceptual Tasks

Cleveland and McGill (1984) [48], aimed to build a scientific foundation of graphical methods for data analysis and representation using the graphical perception, which they defined it as “The visual decoding of information encoded on graphs” [48] (p. 531). Their approach includes a theory that identifies a set of perceptual tasks that are performed when people read graphs. To test their theory, they implemented two experiments. The first is the position-length experiment and the second is the position-angle experiment. As a result, they ranked ten elementary perceptual tasks according to their accuracy when people use them to extract quantitative information from graphs: position on a common scale and nonaligned scales; length; direction; angle; area; volume; curvature; shading and colour saturation. Vandemeulebroecke et al. [236] designed a graphics principles cheat sheet that summarized the most important principles from different sources, and they included Cleveland and McGill [48] ranking of the effectiveness of graphical attributes. Cleveland and McGill’s study [48] was later reimplemented and confirmed by Heer and Bostock using a crowdsourcing experiment. Heer and Bostock [106] replicated the similar design of Cleveland and McGill’s first experiment in a crowdsourcing environment and by adding other types of charts (Pie charts, bubble charts, and treemaps) to the Bar chart.

Mackinlay [143] extended Cleveland and McGill’s [48] ranking of perceptual tasks as it was focused on quantitative information and did not consider non-quantitative information encoding. Mackinlay included nominal and ordinal information and added perceptual tasks to the ranking such as texture and shape.

Simkin and Hastie [209] stated that the effectiveness of a graph depends on the judgment task the viewer performs. Simkin and Hastie claimed that the interaction of the elementary code and the judgement task determined the performance. For example, proportion judgement is outperformed when using angles perceptual judgement task, which is used with a Pie chart. Additionally, they listed five elementary processes: anchoring; scanning; projection; superimposition and detection. Munzner [156] provided a ranking of the effectiveness of visual

channels based on the data type and channel type (magnitude and identity channels). Ordered attributes should be displayed with the magnitude channels while categorical attributes with the identity channels. Munzner's ranking [156] was influenced by many other resources, such as the foundational research on the ranking visual elements by Cleveland and McGill [48], which was presented earlier, and the Cleveland model [47].

Perceptual Organisation:

Gestalt Principles: the Gestaltsits (German Psychologists) studied the perception of complex objects overlapping with others in the environment [64]. Their main principle was the law of Prägnanz, which states that people perceive the simplest organisation of visual fields [64]. The phrase "the whole is more than the sum of its parts" briefly summarises the Gestaltsits approach [213] (p. 13). The following principles are the Gestalt principles of visual perception based [213].

- **Figure-ground:** when people perceive a visual area, some objects (i.e. figures) appear distinctive, and the remainder in the visual area (less important objects) are treated as a background (i.e. the ground).
- **Proximity:** when people perceive a group of objects, they tend to visually group objects which are adjacent to each other.
- **Similarity:** people tend to visually group objects according to their similarity.
- **Continuity:** people tend to perceive objects in continuous forms more than discontinuous ones.
- **Closure:** people tend to perceptually complete objects that are incomplete.
- **Symmetry:** people tend to perceive objects as forming a symmetrical shape around the centre.

These principles are also called **visual grouping** and are associated with the way the visual patterns could be grouped so they can be perceived as groups [249]. People assort visual elements in groups by specific patterns including (proximity, similarity, continuity, and closure) [154].

Proximity grouping means that items are placed within the same spatial region [156]. Similarity grouping refers to the items that are visualized with similar visual appearance, continuity grouping states that items with their continuations are perceived as a whole, Closure grouping relates to the items in a closed shape as people usually perceive them as a whole [249].

The capacity of visual working memory is extremely limited. Evidence from previous studies showed that the limitation of the visual working memory could be enhanced by several factors such as familiarity and perceptual grouping [137]. Although these principles generally improve performance, previous research indicates that these principles could vary in their effectiveness [137]. For example, previous research showed that grouping by proximity was faster than grouping by shape, while shape similarity produced more errors [137].

Visual Search

A visual search is a fundamental task for seeing due to the limited information in our heads, although people are not usually aware that they are performing it [239]. It is defined as a task that involves a rapid detection of a specific target stimulus within a visual environment [64]. According to Trick and Enns [230], a search is an essential task in visual-spatial attention studies. In this task, the user needs to indicate as quickly as possible if a specified target item exists in a display [230]. A significant part of the visual search is related to the way the eyes are moved around in the display to extract information [240]. Ware [240] have classified the visual search types according to eye movement.

Many studies have classified the visual search task based on the patterns of performance into two different types [230]: in the first type, the target has discriminable features which makes it easily distinguishable from the other elements in the visual environment such as colour or size, and the display size has a small effect on search and reaction time (i.e. the time it takes to respond); in the second, the target does not differ by discriminable features from other elements and the display size has a significant effect on search and reaction time [230].

The same classification of search types is also presented by Gramazio et al. [101] as a serial and parallel search as follows:

Parallel search: if a target varies from a homogeneous set of distractors on a single feature

such as the colour, the visual search is considered to be pre-attentive (easily with less effort) as all the elements are scanned in parallel and the target “pops out” [101]. **Feature search** is also known as a parallel search as it is a search task where the target differs from the distractors by one feature such as the colour [245].

Serial search: if a target varies from heterogeneous distractors on more than one feature (e.g., colour and shape), the visual search is considered serial. People need to serially search all elements in the visual environment until the target is found [101]. A serial search is also called a **conjunction search** as it includes searching for a specific combination of features, it is slow and has many implications [229, 240].

Regarding the effect of the number distractors, Treisman and Gelade study [229] showed that the number of distractors had differential impacts on search time that varied based on the difference in the visual features between targets and distractors.

Investigating the effect of different visual encodings on the time required for the visual search task was the subject of many experiments. For example, Haroz and Whitney [102] conducted a set of experiments to evaluate the impact of visual features (colour or motion) and different visual elements on users’ performance. Their results showed that in a visual search, the participants were able to find targets faster when marks were grouped by colour than randomly distributed. The difficulty of the visual search increases if: (1) distractors are closely similar to the targets and (2) distractors highly differ in their visual appearance [59].

The earlier discussion showed that the type of visual search could vary according to the visual features of the target in the visual display. Some visual features such as (colour, brightness, length and position), “pop-out” of a visual display, which make the visual searching for them easily findable.

Visual Saliency

The effectiveness principle states that: “...the importance of the attribute should match the salience of the channel” [156] (p. 101). Therefore, the most important attributes should be encoded using the most effective channels to be most noticeable. Then the other attributes should be decreasingly encoded with less effective channels based on their importance [156].

The following explanations of visual saliency is derived from Ware's book [240]. Each visual map in the visual field is indicated to a particular feature (e.g., specific colour or shape). Ware [240] defined three features that make the item that is being searched for easily findable: (1) *Priori salience*: patterns that attract more neural activity in feature maps than others. (2) *Top-down salience modification*: By enhancing sensitivity, it retunes the feature maps to increase their sensitivity to specific features. (3) *Scene gist* is more related to the experience, and it activates visual search strategies suitable to the scene.

2.3.1.2.3 Perception of Different Visual Encoding

Graphical perception of different statistical data charts has been widely studied. An early work by Eells [58] in 1926 considered comparing circle and Bar chart to identify the best method for representing component parts. Eells [58] found that the circle has the same speed as the Bar chart while it has higher accuracy, especially when the number of the parts within the charts increases. He suggested using the circle diagram to show component parts due to its accuracy in addition to its popularity and psychological appeal. Croxton and Stryker [52] focused on the same question raised by Eells however, they experimented different percentages of parts within the diagrams.

Cleveland and McGill [48] studied different versions of simple Bar charts and found that the accuracy decreases as the distance between the two examined bars increased along the vertical axis. This finding was further investigated and confirmed by Talbot et al. [217] using a series of follow up crowdsourcing experiments, they stated that comparing two non-adjacent bars is difficult as the separation between the bars increases the difficulty, especially if the bars are short. However, these experiments were restricted to judging the height of two bars only. Talbot et al. [217] evaluated designs of adjacent and separated bars, aligned and unaligned stacked bars and divided bars for comparison tasks. Their aim was to understand how the accuracy differs across different Bar charts designs. They concluded that any changes in the design could affect the way that Bar chart is perceived and stated that there is a still a lack in understanding of what could have an impact on perception.

According to the Gestalt principles, Pinker [171] hypothesised that the Bar chart is easier for making comparisons between discrete values, while the Line chart is easier for assessing

trends. Pinker [171] stated that this hypothesis was proven by Simcox [208] whose experiments showed that people were faster with the Bar chart when comparing data point values, while they were faster with the Line chart to judge slope. This hypothesis was also confirmed by Zacks and Tversky [247], who compared the effectiveness of the Bar and the Line charts using two data points datasets. The outcomes of the study found that the participants read the data in the Bar chart as discrete data, while they described the data represented in the Line chart as a trend. The Difference between Pie chart and Bar chart perception has also been evaluated by Simkin and Hastie [209]. They compared Bar charts with Pie charts when performing two judgement tasks: comparison and proportion judgements. The participants in Simkin and Hastie's experiment performed the comparison judgement more accurately with Bar chart (i.e., position judgement), while they performed the proportion judgement more accurately with the Pie chart.

2.3.2 Visualization by Task Taxonomy

Having benchmark datasets and visual tasks are important for improving information visualizations [172] and deciding on the task, or the set of tasks, the user needs to perform is a critical element of any experiment design [240]. Visualization tasks have been discussed broadly in the literature and have been classified according to different taxonomies, where some were developed for specific domains and data types, such as [133], and others were developed to be generic (e.g., [8, 9, 31, 234, 241]). Visualization task taxonomy is defined by Zhou and Feiner [249] as a level of abstraction somewhere in between visualization intents and low-level visual techniques and that visualization intent could be accomplished by using the visualization techniques to achieve specific visual tasks. Casner [41] proposed an analysis of users' goals that are suitable for representations designed to support complex tasks. Casner provided a description of abstract operations required to perform the visual task and selected some presentations that could replace these abstract operations with perceptual operations. The perceptual operators were classified into search and computational operator and they included eleven classes that categorised the perceptual operators (e.g., search, verify, look up and equal) [41].

Wehrend and Lewis [241] addressed users' visualization tasks and operations, they are domain-independent and help in finding the most appropriate visualization techniques for a given prob-

lem. Wehrend and Lewis created a matrix that corresponds to a combination of an object type and a user cognitive task (e.g., distribution, cluster, associate, or locate). Shneiderman [206] proposed task taxonomy by data type, the data were categorized into seven data types: 1D (i.e., linear data); 2D (e.g., planar or map); 3D (i.e., real-world objects); multi-dimensional; temporal; tree and network data. The tasks to analyse and explore the data: overview; zooming; filtering; details-on-demand; relation finding; historical tracking of actions and extraction [206].

Amar and Stasko [8] addressed the analytic gap, which presents the limitations of the visualizations in supporting a high-level analytical task. They proposed a taxonomy for high-level tasks to support complex decision making and learning. Amar et al. [9], have also proposed another taxonomy for low-level tasks that captures and analyses people's activities when using information visualizations to understand data.

Valiati et al. [234] proposed a taxonomy of tasks that support different scenarios for evaluating the visualizations of multidimensional data. Their taxonomy integrates tasks at different levels including analytic, cognitive and operational tasks for supporting exploratory analysis or for statistical analysis. The taxonomy comprises seven tasks: identify; determine; visualize; compare; infer; configure and locate. Five of these tasks, which are (identify, determine, compare, infer and locate), might be goals a user might have when using a visualization method to visually explore or analyse the dataset. These five tasks of the taxonomy are considered analytical, while visualize and configure are considered intermediate-level tasks that support the five analytical tasks presented earlier [234].

Valiati et al.'s taxonomy [234] is based on previous taxonomies in the literature and on the observation of users when they perform exploratory tasks in a multidimensional dataset. They conducted two case-studies using Parallel Coordinates [119] and RadViz [111] multidimensional visualization techniques. In their previous work [244], they studied how to model these tasks using a formal method to benefit from the possibilities of generating different scenarios that included all the implications for each visual task and to overcome the limitations of unstructured evaluations. The tasks were examined using four questions, where each had an estimated scenario to show the potential tasks and subtasks involved in responding to each question. The results showed that the real scenarios included all the estimated tasks with a slight difference

in the order.

Andrienko and Andrienko [12] focused on the dataset and the tokens describe it. They considered the way the dataset characteristics and the visualization requirements for a specific task may be matched to determine patterns in order to predict the most appropriate visualization for the given task. Although their book [12] mainly focused on spatial and temporal data, they provided approaches and definitions for the visualizations and tasks that are not restricted to spatial data only and have more general applicability.

Schulz et al. [202] proposed a design space for information visualization tasks that combines and clarifies existing task taxonomies from previous studies. The developed design space has five dimensions and involves “goal, means, characteristics, target, and cardinality”. Whilst Brehmer and Munzner [31] proposed a multi-level task typology to address the issue of tasks ends-means ambiguity (i.e. distinguishing between why and how a task is performed) resulted from the lack of discrimination between the low-level and high-level tasks presented and was characterised by previous work. They aimed to transform the descriptions of domain-specific tasks into abstract tasks to facilitate comparisons of different visualization techniques designed for different domains. These concepts regarding task abstraction and multi-level typology were further presented and discussed in [156].

Munzner [156] presented a *what-why-how* visualization framework that breaks down the visualization design process into these three questions that have data-task-idioms answers. The author addressed the *what* question with answers about data abstraction which are covered in detail in Brehmer and Munzner’s multi-level task typology [31]. The reason *why* a visualization tool was being used was broken down into action and target. Action has three levels: high; intermediate and low-level actions, which were analysed (consume and produce), search (lookup, browse, locate and explore) and query (identify, compare and summarize) actions respectively. Target is data to be found by the actions and is based on the type of data, for example, for all data it could be trends or outliers [156].

Regarding the health data visualization, Theis et al. [221] constructed a task-data taxonomy by surveying eHealth systems users (older adults) and experts to be used in requirements and evaluation of eHealth data visualization. However, this is not related to personal health apps.

Before moving to the next section, some of the definitions about the tasks and datasets are presented. The dataset will be first defined as the tasks are questions require retrieving information from the data. The definitions are mainly based on Andrienko and Andrienko [12].

Dataset: the components of any datasets are distinguished based on their role. There are *referential components* which refer to the context in which the data was obtained and there are *characteristic components* which means the values measured or observed in the context.

Task: is defined as a question regarding data that can be answered based on the information contained in the data. A task consists of a target (unknown), which refers to the information that needs to be found and the constraints (known), which refers to the conditions this information needs to meet. Andrienko and Andrienko [12]’s task topology idea for exploratory data analysis originated from Bertin (1983) [22] as they distinguished tasks according to the level of data analysis while taking into consideration dataset components (i.e. referers and attributes). However, Bertin classified the questions, based on their type, into three levels of reading: ‘elementary’, that corresponds to a single element from the dataset; ‘intermediate’, that corresponds to a subset or group of data elements and ‘overall’, that considered all dataset elements [22]. Andrienko and Andrienko’s task topology classified the tasks as follows:

- **Elementary Task:** this refers to single elements of the reference set. It is defined as a task that does not include dealing with sets of references or attributes as wholes, but addresses their elements.

The *elementary tasks* consist of look up tasks (direct or inverse look up tasks), comparison tasks (direct or inverse comparison tasks) and relation-seeking tasks.

- **Synoptic Task:** this is the complete reference set or some subsets. It requires dealing with sets as a whole, unlike the elementary task that requires dealing with individual elements.

The *synoptic tasks* consist of patterns and trends in the data.

- **Analytic Activity:** this is user analytic tasks which are part of the analytic activity. They range from the “high-level” goals (e.g., understanding trends over time) to specific “low-level” queries (e.g., determining specific knowledge) [9].

2.3.2.1 Evaluation of Charts Against Visual Tasks

Several studies have examined the effectiveness of different charts for different tasks [41], and for different datasets [142]. Other studies have developed presentation tools the structure of which are based on algebra and the criteria of the graphic design [143].

Mackinlay et al. [142], presented ‘Show Me’, an automatic presentation tool that is composed of a number of user interface commands and is integrated with Tableau. A user experience was performed in this study using computer logs to evaluate the effectiveness of the tool. The tool builds a visualization view for the data based on the highest ranked command. Saket et al. [186] evaluated the effectiveness of different visualizations of tabular data using a crowd-sourced experiment with 180 participants. The aim was to improve visualization recommendation systems. They evaluated two-dimensional visualization types (Table, Line Chart, Bar Chart, Scatterplot, and Pie Chart) using two public datasets (Cars and Movies datasets). Their work examined ten different visual analysis tasks which were derived from [9]. Amar et al.’s task taxonomy presents a set of ten low-level analysis tasks that capture people’s activities when they use information visualizations for understanding data. In addition, they asked the participants to rank the visualizations according to their preference for each task separately. A decision tree was trained on the collected data to drive a recommender. In terms of the visualization preferences, their findings identified a positive correlation between task accuracy for each visualization and user preference. They indicated that people have a choice for visualizations that allows them to complete a task accurately. With respect to the performance, the results showed that people performed faster and more accurately with the Bar chart than the Line chart, while the Line chart was more effective for characterising distributions. The results also showed that the Bar chart is faster than other evaluated visualizations for finding extreme values, including the Line chart. The outcomes of the study contributed design guidelines based on the performance and preferences outcomes.

The findings obtained from Saket et al.’s [186] and Mackinlay et al.’s [142] experiments regarding the participants’ performance with the experimented tasks on the Bar and Line charts, in addition to Zacks and Tversky [247]’s study, contributed in structuring the hypotheses **H1b**, **H2b** and **H5b** that are presented and explained later in the main study (i.e Chapter 6). These developed hypotheses are related to the visual encodings used in the designs.

Comparative Design Visualization: Gleicher et al. [95] have proposed a taxonomy that classified the comparative designs space into three categories: (1) Juxtaposition, which presents different objects separately and relies on the users' memory to recall and connect between the objects, this design is defined by Tufte [232] as small-multiples; (2) Superposition, which visualizes the objects overlaying in the same space and (3) Explicit encoding, which computes the relationships between data objects and produces a new visual encoding that is based on the relationship. In a later work, Gleicher [94] provided consideration of visualization comparison as it is an important task for users. Gleicher provided a framework to help in designing visualizations that involve comparisons. This framework includes four considerations to understand the comparison task, which are identifying the comparative: elements, challenges, strategy and design.

The following section discusses the visualization design for time-series data with a specific focus on the Radial visualization designs and the use of clock metaphor.

2.3.3 Visualization of Time-Series Data

To create an effective visualization for specific data is to transform the data into a perceptually efficient visual format [240]. This thesis focuses on personal health data, which was described in section 2.2 as data collected from the sensor over time. Time-series data is defined as: "... sequences of data points measured at successive times, typically at equal intervals." [248] (p. 1738) or as a set of values that are changing over time [108]. It is a very common type of recorded data in different domains such as financial stock prices, temperature and public policy [108]. The characteristics of time-series data differ significantly and they are extremely dependent on the domain type where the time dimension has multiple aspects [248]. For example, the data could be a set of real numbers (e.g., temperature measures), or events such as records of medical visits [248]. A considerable amount of literature has proposed and evaluated different visualizations methods for representing time-dependent data; however, known visualization methods for representing time-dependent data are Linear charts and Cyclical layouts [128], where some of these studies considered multivariate and multiple time-series such as [121, 214]. Some of the published research concerning the visualizations of multivariate time-series data were focused on supporting or enhancing interactive techniques such as: Lin-spiration tool [100] that proposes a new Focus+Context approach, TimeBench software library [182] that supports interactive visual analytics for time-series data, and StreamStory [214] interactive visualization tool for exploring large multivariate time series. However, this thesis focuses on evaluating the developed visualizations when performing a set of tasks without considering the interactive techniques.

Aigner et al. [7] presented the important criteria that describe different types of time from a visualization perspective. These criteria include the structure of the time, which is the Linear time that has a starting point with data from past to future vs Cyclic time (e.g., seasons) and time points that have no duration vs time intervals which define data as duration (e.g., days). Aigner et al. implemented an event-based visualization approach that focused on representing only relevant information from the large volume dataset. The visualizations developed in this thesis focus on time-series data of personal health and physical activity over a daily basis, which was collected from sensors embedded in phones and wearable devices.

Tominski et al. [224], developed TimeWheel and the MultiComb, which are two Radial visualization methods of interactive axes to enhance multidimensional dataset analysis. The visualization methods were evaluated through testing with different datasets. The techniques are used with big datasets, such as for detecting events in health-related dataset that is composed of data of thousand days. They found that the developed techniques are suitable for the exploration of multiple dimensions.

Javed et al. [121] aimed to provide design guidelines to find a suitable technique for visualizing temporal data. They evaluated the graphical perception for different visual tasks with different visualizations for time series data (simple line graph, small multiples, horizon graphs and braided graph) using a controlled lab experiment. The novel braided graph was designed to solve the problem of identification and comparison of multiple time-series sharing the same space and uses area curves. The used dataset is a generated multiple data series (with a controlled number of series) which were randomised for each new experiment. Their hypothesis stated that both the shared and split space visualization techniques for time series data have strengths and weaknesses according to the different visual tasks. The tasks tested in the study were maximum, slope and discrimination. In a shared-space visualization, visual clutter is generated from the overlapping time-series, and it increases as the number of series increases. This issue increases the difficulty of following the lines as they overlap. On the other hand, following the lines in a split-space visualization require moving gaze between the chart regions. The findings from Javed et al. [121]'s experiment stated that the simple graphs and small multiples are generally most versatile than the other evaluated visualizations for time-series data.

Srinivasan et al. [211] studied the visual comparisons that were performed on a single view using an online experiment. Their research investigated the differences of Bar chart designs to facilitate comparison tasks using two different multi-series datasets, the results implied the design of an information dashboard. To strengthen their understanding of information dashboards before designing the experiment, Srinivasan et al. conducted informal interviews with the managers of Microsoft Power BI and surveyed 68 available information dashboards that showed multiple series data for multiple years. This resulted in three main design constraints: space limitation of Juxtapositions approaches; the users have different levels of visualization literacy and explicit encoding approaches prevent other tasks. They developed four Bar chart

visualizations (grouped Bar chart, grouped Bar chart with difference overlays, Bar chart with difference overlays, and a difference chart). The outcomes indicated that the charts with difference overlays are comparable or better than the other evaluated designs.

Stopar et al. [214] developed, StreamStory, an interactive visualization tool that visualizes long-term time-series data and implement machine-learning algorithm to help with data interpretation. They implemented use cases using three datasets with different complexity and found that StreamStory is good for representing large datasets of medium dimension.

The most related work with respect to the multivariate time-series data was [121] as it included a comparisons between different visualizations to represent multiple time-series and experimented the concept of the split and shared-space. The split-space (i.e. juxtaposed) design requires eye movements between the visualizations to compare the data. However, the shared-space design could benefit from performing the comparison on the same space. Another finding is that in the shared-space design, clutter could be generated between the overlapping time-series, which make it difficult to visually follow a specific line, particularly when the number of time-series increases. These outcomes regarding the efficiency of overlapping and splitting multiple time-series data were used in designing the visualizations in this thesis and structuring the related hypotheses **H3b** and **H4b**.

2.3.3.1 Radial Visualization

The Radial visualization has been referred to in the literature as the visualizations that are based on circular shapes [36]. In the Radial layout, the items are distributed around a circle and it uses the angle channel and other Linear spatial channels [156]. Radial charts were considered to be an aesthetic type of graph and space-efficient [36, 108, 126], and bringing more insight to time-dependent data [141]. Heer et al. [108] stated that using polar coordinates instead of the Cartesian for data type organised as hierarchies is more aesthetically appealing and enables a more efficient use of space. The Radial layout is usually used for its elegance, its support of symmetrical representation, wholeness and closure [36].

The Radial layout has been applied for different visualization purposes and data types such as glyphs [79, 126], hierarchical structures [37], time-dependent [6, 32, 130, 141, 238] and

network traffic monitoring [126]. It was evaluated using different methods such as controlled-lab experiments [79] eye-tracking [37, 96] and crowdsourcing [32, 53, 238], and on different platforms including mobile phones [32] and smartwatches [25].

Lima, in his book [138], presented the context of the circular layout over centuries. Lima's intensive review revealed twenty-one patterns that the author grouped into seven families according to the circular layout's visual configuration. Rings & Spiral, wheels & Pies are examples of these major circular layout families.

Recently, there have been many apps, and dashboards that use the Radial layout to represent different data, for example for time management, the ATracker time tracker [246], Money Manager [97] and iFood Diary [117]. Some of these apps/dashboards use a Radial chart to represent health data based on the user goal (e.g., an iOS Activity and Fitbit apps), others use a clock display to visualize the health and activity data over the day (e.g., Samsung Health app). Moreover, many visualization designers on the Tableau Gallery [216] have presented different Radial layout representations (e.g., [129, 170, 173, 205]). This section of the thesis presents the research on Radial visualization in general, and its use for temporal and time-dependent data specifically and the use of the clock metaphor with the Radial layout view.

Bertin was the first to use the Radial chart to represent the temperature of the months of the year, where the wedges on the Radial graph represent the month, and the colour of the wedges (darkness) was used to encode the temperature (Figure 4.2 in [171]). The Radial chart has been used recently by Brehmer et al. [32] to represent the temperature over a week, a month and a year.

In 2009, Draper et al. [56] presented a survey of Radial visualization that proposed a taxonomy for its design patterns. Their survey also highlighted the typical uses and application domains of Radial visualization. One of the basic uses of Radial visualizations mentioned in this survey [56] is for representing continuous data that has a predictable, repetitive structure (e.g., time-series data). Moreover, Draper et al. motivated other researchers to use the unique characteristics of the Radial visualizations in future designs, as well as emphasising the need for more quantitative evaluation of Radial visualization to provide information about its effectiveness relating to different visualization problems. George Chin et al. [92] evaluated different layouts for representing real-time data, and they found out that a cyclical timeline had higher performance

over the Linear timeline in identifying and detecting temporal and periodic patterns.

Diehl et al. [53] implemented two online experiments to study the strengths and weaknesses of the Radial visualization when compared with the Cartesian visualization. The first was short and with wide-ranging participants (674 participants), while the second was longer and with 21 expert participants from potential users. The design of the visualizations was based on a $n \times n$ matrix for the Cartesian/Linear representation, and n sectors and n rings for the Polar/Radial visualizations. The implemented task was memorising positions of visual objects which are considered essential in different usage scenarios. Their results concluded that in general the Cartesian visualizations outperformed their Radial equivalents especially with respect to the response time. However, they also showed that the Radial visualization appears to be more suitable for focusing on a specific dimension, which could be visualized with circle sectors.

The difference between the Radial and the Linear layout has been evaluated using eye-tracking experiments such as [96] and [37]. Goldberg and Helfman [96] conducted an eye-tracking study to understand how people read and compare values on different information graphics (Line, Area, and Scatter graphs) on Radial and Linear layouts. They also reported value lookup tasks, their results showing that participants completed the tasks with the Linear layout faster than with the Radial layout and they stated that visual scanning of circular graphs along the rings is prone to error. Burch et al. [37] evaluated the use of three different layouts (Orthogonal, Radial and the traditional Linear layout) for a node-link diagram to represent hierarchical structures using an eye-tracking experiment with 38 participants. The metrics for the comparisons were task accuracy, completion times, and exploration behaviour. The layouts were evaluated based on a task related to the type of data. The results showed that the participants performed the task more slowly with the Radial layout than with the other two layouts. In terms of the participants' perspectives, they reported that the Radial layout was difficult to estimate the depth of the node's location with respect to the tree and to compare it with other nodes. It was also the least preferred by the participants when they gave their feedback in terms of motivation, intuitiveness, and suitability.

Zhao et al. [248] proposed KronoMiner, which is a visualization tool for long time-series data that integrates different visualization and interaction techniques and supports a high level of exploration and analytical tasks. The design was based on a Radial layout that views details

when focusing on specific data. The tool was designed to provide the user with a high level of control on the visualization. The user could choose between different layouts to visualize the data (e.g., Linear/Circular and Stacked/Sequential) and different types of plots (e.g., Line chart, Line plots, Scatterplot and Histogram). Expert users from different domains were involved in many phases of developing and evaluating the tool to benefit from their feedback. Although the study showed the possibility to support effective analysis with high interactivity, the participants were confident to use the basic interaction features and familiar visualizations but were not able to begin using the newly provided features and representations.

Fuchs et al. [79] implemented a controlled experiment with 24 participants to compare the effectiveness of different temporal glyphs (Line, Clock, Star, and Stripes) with respect to peak/trend detection and temporal location tasks. They used two different datasets with different density and included the participants feedback in the experiment. Their findings based on the task and the data density showed that the Radial encodings of time for temporal location tasks were more efficient than the Linear ones, while a Linear encoding is generally a good representation for peak and trend detection. These findings were also supported by the participants' subjective feedback.

Brehmer et al. [32] evaluated the difference in performance between Radial and Linear layouts for visualizing ranges over time using a crowdsourced experiment on mobile phones. They evaluated the visualizations of two different datasets (temperature and sleep duration range data) with granularities of a week, a month, and a year across five tasks including value retrieval (locating date, reading value, locating min/max) and comparison tasks that compared values and ranges. Their experiment findings indicated that the participants generally performed faster with the Linear layout, but no difference was found in the accuracy between the two layouts. Regarding a smaller display (i.e. smartwatches), Blascheck et al. [25] implemented two perception studies to evaluate the time people require to perform a simple comparison task for visualizations on the small display of a smartwatch. They aimed to understand what people perceive from the visualizations during quick glances on the smartwatches and to provide design guidelines for the limited-display visualizations. They tested three chart types: Bar; Donut and Radial on three data scales: 7; 12 and 24. Their results showed that the Bar and Donut charts are better options for data comparison tasks on smartwatches.

Maçãs and Machado [141] proposed a Radial-calendar visualization model for representing and analysing the Portuguese's consumption data over two years, to assist the analyst in identifying periodic patterns and variations from normal consumption. They implemented a case-study (n=30) to compare the Radial calendar with the traditional Linear calendar layout. The evaluated tasks were detecting patterns, identifying behaviours within the product's hierarchy, detecting differences within and across months, comparisons, and detecting the deviations (anomalies). They stated that for all the tasks no layout outperformed the other in terms of time and accuracy. The subjective feedback from the participants revealed more about the designs as they considered the Radial to be better for readability and providing insights. They concluded that the Linear layout was more intuitive at first, but when the users learned how to read the Radial layout, it became more preferred and that it could be a better choice for visualizing time-varying data.

In a more recent study, Waldner et al. [238] compared the Radial layout with the Linear layout using a Bar chart encoding for the data within the layouts. The four developed visualizations are 24-hour Radial and Linear designs and two juxtaposed 12-hour Radial and Linear designs. The source of the data for the visualizations was "New York traffic accident data from 2013 to 2016" [238], which was classified into boroughs and divided into months and hours and selected a subset for the experiment. The six evaluated tasks in this experiment are of the following types: one high-level task; four low-analytical tasks and one subjective task. The response type varied according to how they were evaluated. The first task response was evaluated based on the observed variables. The tasks 2 to 5 were evaluated based on the accuracy and completion time and task 6 was evaluated according to the users' ratings (subjective feedback). They recruited 100 participants for a crowdsourced experiment (no specific window size for the browser was required) and stated that the largest visualization was fully visible on a typical mobile phone in portrait mode. Their results showed that for all tasks, the Linear 24-hours Bar chart is the most accurate, most efficient and was also preferred by the users.

2.3.3.1.1 Clock Metaphor for Time-Series Data

Clock metaphor and Radial representation has been used to represent different aspects of temporal data [35, 51, 55, 67, 99, 126, 168, 238]. For example, the SpiraClock [55] that used

the analogue 12-hour clock view with a spiral to represent upcoming events. Additional features were used with the clock such as colours to categorise different events, transparency for the happening and imminence events and tooltip to show more information about the event. The SpiraClock was used in three applications for displaying different events which are: the BusClock; ShuttleClock and the Calendar-Clock. Buono et al. [35] used a circular display to represent the daily work patterns of team members, where each ring in the circular layout was divided into 24hrs and refer to one member of the team. The tool was generally appreciated by the participants and judged to be useful for a work context. Kintzel et al. [126] used a clock style glyph as one of four representations for monitoring big IP spaces. The clock glyphs are composed of 24 segments, which are colour coded based on the traffic of the corresponding hour. They concluded that the clock style is more intuitive, and the natural order of time helped in better perception of the data. The developed tool was verified by a case-study and it distinguished the ClockView from other tools for research purposes. Fischer et al. [67] proposed the ClockMap to visualize hierarchical time-series data of network traffic. The ClockMap combined a circular nested treemap layout with a Radial clock glyph visualization for time-series data. The design was effective for comparative tasks on large hierarchical time-series data and outweighed the traditional treemap for comparison tasks.

Clock Display for Health and Activity Data: Pagno and Nedel [168] used the clock metaphor to represent the hourly step count. The clock was divided into 24 hours slices, where the height of each slice represented the step count. The clock visualization was aligned with a calendar view to show the total number of steps taken every day and the clock visualization was used to represent more details about the user activity (the hourly step count). The inner area of the clock was used to visualize the hours the user spent on each project which was visualized using coloured bubbles [168]. However, no evaluation have been performed on these visualizations. The clock display was also implemented by Verdezoto and Olsen [237] in a tool that aimed to support and manage home-based medication for elderly people.

Gouveia et al. [99] integrated different representations (including Radial) of physical activity levels during the time of the day on a smartwatch interface. However, their study focused on the impact of these representations on the behaviour change, not on the perception and the readability of these visualizations. Croon et al. [51] proposed a 24-hour clock display

to visualize everyday health schedules on a tablet screen to increase the comprehension of personal health schedules. The area inside the clock circle was to present a graph network to help the patient in understanding and exploring each scheduled task (e.g., taking medications). They evaluated the design with 25 participants who were patients and general practitioners using interviews, usability questions and task-based scenarios where the number of mistakes was observed. The participants' feedback and the results were positive, they indicated that the design could be an effective means for interpreting health schedules and no issues raised regarding the clock visualizations; however, it was good for providing useful feedback.

2.3.3.1.2 Advantages and Disadvantages of Radial Visualizations

Turning to the previous work on Radial visualizations, as discussed in the previous sections, the benefits of usage and the advantages of Radial visualizations are presented here, in addition to its disadvantages and the cases where the visualization is not considered suitable or preferable. The points listed are mainly derived from Burch and Weiskopf's discussion about the strengths and limitations of the Radial visualization in 2014, [36] and the other works discussed here in addition to the clock design.

Advantages:

Dataset: the Radial representation makes sense if there is an internal circular behaviour in the data or periodic patterns. The Radial design with respect to the dataset type benefits from its ability to provide an intuitive understanding of the circular pattern and mirror the data perfectly (i.e. symmetry). It also does not have an end point like the Cartesian representations [36].

Performance and readability: in the Radial visualizations, the objects are rendered within the field of view which makes it easier to be reached by the viewer. Unlike the Cartesian representations where the corners are not easily reached by the viewer's eyes. According to Diehl et al. [53], the Radial visualization appears to be more suitable for focusing on a specific dimension.

Look and appearance: in terms of the style and appearance, Radial charts are said to look more aesthetically appealing. Also, the Radial design is usually justified to be elegant due to its support of symmetries and wholeness and closure [36].

Disadvantages:

Application and implementation: different aspects contribute to the complexity in the implementation of the Radial visualization, such as the required geometry and algorithms, the difficulty of implementing interactive features and generating text labels (Radial labels are difficult to apply and hard to read, while non-radial labels are not consistent with the design).

Performance: it has higher perceptual challenges than the Linear representation and has a higher cognitive effort. Comparing visual elements in the Radial form are harder due to the absence of a straight guiding line, which the eyes follow easily in the Linear form of the chart. From a perceptual point of view, according to Cleveland and McGill [48], angle, slope and area judgements which are used when reading from a Radial diagram are less accurate than judging length and position when reading Linear charts. In addition, previous studies [238] and [32] showed a general outperformance of the Linear layout over the Radial layout for some visual tasks such as reading values and comparisons tasks. These findings from previous research contributed in the hypotheses (**H2, H3a, H4a, H5a**), which are explained later in Chapter 6.

In addition to the listed advantages and disadvantages of the Radial layout, there are some benefits of using analogue clock metaphor that follows a Radial layout for representing temporal data as follows.

Advantages and Concept of the Clock Metaphor: Kintzel et al. [126] used a clock metaphor and ordered the 24hrs in the clock to be set out as 00:00 on top, 12:00 at the bottom, 06:00 on the right and 18:00 on the left, as this said to be more intuitive and the natural order of time helps in better perception of the data. They also stated that the use of the circular clock representation is more space-efficient when the data is displayed on a smaller screen.

All the above work mainly used the clock display as a metaphor for representing different daily data as a part of a tool that was developed for a specific domain problem. It was evaluated using case-studies to test the applicability and usability of the developed tools, and to gain subjective feedback from the participants on the design. There is a lack of a controlled user study to evaluate the effectiveness of the Radial clock representation for understanding and interpreting daily periodic data in general and in the personal health context specifically. However, the findings from [67, 126] that showed that the use of the clock-style design is effective for temporal data,

as the natural order of time enhances perception of the data, contributed in constructing the hypothesis (**H1a**) that is presented later in Chapter 6.

The following section of this chapter explores the evaluation methods for information visualization and the evaluation of personal visualization.

2.4 Empirical Evaluation Studies

Good visualization is a key element in personal health technology design space [17]. Evaluation studies of information visualization have become increasingly important to evaluate visualization methods and tools used in several disciplines (e.g., health, education and finance). Different aspects have been evaluated in information visualization, for example: the effectiveness of the visualizations (e.g., [107, 186, 211]); the memorability of the visualizations [30]; users' experience with the visualizations [115] and the usability of the visualizations [251]. This section discusses the evaluation methods used for evaluating the information visualization in general and personal visualization specifically.

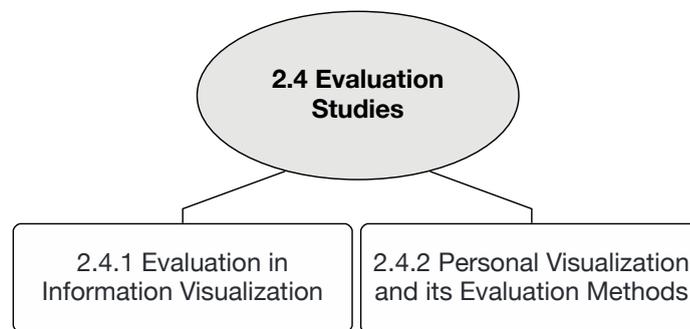


Figure 2.4: Chapter 2 Evaluation Studies Section Diagram.

2.4.1 Evaluation in Information Visualization

A standard method for evaluating the effectiveness of the visualizations is by measuring people's performance, which is a valid approach in several contexts. However, in some cases, relying only on performance measures is not enough to detect differences, this is because the viewer may perform equally on different visualizations, but with different levels of cognitive effort [116]. Therefore, evaluation studies involved additional metrics to evaluate the visualizations, such as: preferences and confidence scales [25, 32, 79, 186]; interviews with expert users [211] and interviews with the participants to gain feedback [79].

Personal trackers are usually looking for insights from their collected data. Previous studies confirmed that insight happens when a visualization method efficiently supports a visual task

[77]. Therefore, the evaluation of the effectiveness of visualizations regarding basic analytical tasks is significant in every evaluation.

In order to ensure that the designed visualizations are interesting for potential users, they should be developed using complex datasets that represent the domain's real data, tested with real users and implement domain-specific tasks [40]. The visualization tasks and the visualization task taxonomy were discussed earlier in section 2.3.2. However, it showed a lack of studies that specifically focused on the visualization tasks in the context of personal health data collected by tracking devices and apps.

Users' preference as an important factor to involve in the evaluation:

Users' preferences in the visualizations should be considered when designing visualization systems. According to Ware [240], some factors, such as the appearance of an interface, could be crucial in its adoption. Therefore, finding the balance between attractiveness and optimality of the display should be a goal of the research [240].

In HCI studies, interviews and questionnaires are two of the most commonly used methods. However, they are usually applied as complementary approaches for data collection along with other methods such as controlled experiments and field studies [77]. User Surveys and questionnaires were used in different stages of the research to collect different data on the tracking and activity in general, and the method for representing the data specifically.

In terms of information visualization, many studies have focused on generalising the results of evaluating visualization techniques. However, there is a trade-off between a desire to develop generalisable visualization tools and the need to ensure that they work in specific domains [227]. Moreover, the suitability of a particular chart depends upon the task and the dataset; different charts are suitable for different tasks and different datasets. Pinker [171] stated that different graphs are said to be easier or more difficult according to the type of information that needs to be extracted. Therefore, this thesis evaluates different visualization techniques that are designed for the personal health domain in general and specifically for data collected from health tracking devices.

2.4.2 Personal Visualization and Evaluation Methods

There are challenging aspects in the evaluation of information visualization in the context of personal data visualization. There are varied metrics which need to be measured to justify design choices in the personal visualization space such as how the design fits into users' daily lives. According to Huang et al. [114] and Thudt et al. [223], depending on conventional metrics (i.e. task completion time and accuracy) for evaluating personal visualization is not enough. Therefore, several methods have been proposed to evaluate personal visualizations [114, 223]. Thudt et al. [223] highlighted the importance of embracing and adapting research methods from other disciplines such as social science to reinforce the evaluation and research methods for personal visualization. Some of the research methods proposed by Thudt et al. are cultural and technology probes, diary studies and analysis of public data. The technology probe method was implemented by Thudt et al. [222] to evaluate the use of a web visualization tool for generating and sharing the visualization of personal movement data to support reminiscing and self-reflection. The method of analysing public data was implemented by Choe et al. in the studies [44, 45] as described in section 2.2.2.

Many recent visualization studies have explored the design for the novice or non-expert users (e.g., [174]). Pousman et al. defined casual infovis as "*the use of computer mediated tools to depict personally meaningful information in visual ways that support everyday users in both everyday work and non-work situations*" (p. 1149) [174]. They presented the casual infovis as more than a combination of ambient and social infovis and artistic visualizations, and that the focus of its design on usefulness, enjoyment and reflection [174].

Several research studies on personal health visualizations were discussed in section 2.2.3. The evaluation methods used for evaluating the visualizations in those studies were different and for various purposes that not necessarily focused on the visualization itself but rather on other purposes such as Consolvo et al. [49]'s study, that evaluated the visual metaphor's role (UbiFit Garden) for promoting physical activity by reporting on people experience. Field studies were one of these methods such as [49, 60, 66, 103, 115, 149]. For example, Huang et al. [115] conducted a field study to evaluate the adoption of calendar layout to represent Fitbit activity data. Their evaluation was focused on people's experiences with the calendar visualizations using a qualitative approach without considering its impact on behaviour change. Fan et al. [66]

evaluated users' experience with the ambient visualizations that represent their Fitbit activity data. Their evaluation focused on what the participants reported regarding the visualizations during field interviews.

Understanding people's visual capabilities and limitations is important for designing and evaluating visualization designs. The effectiveness of a visualization method depends on people's processing abilities (perception and cognition), in addition to potential users' visualization tasks and goals [228]. In Huang et al.'s [115] calendar visualization, they applied two different types of visual encodings: Line chart and luminance. The logs showed that the majority of the participants preferred using the Line chart and they also reported that using luminance for encoding the data requires more cognitive effort to understand, and it interfered with the other data on the calendar layout. The participants in Fan et al.'s [66] study linked the types of visualizations with the information it can support (e.g., Bar chart is good for retrieving specific information). Different studies (e.g., [60, 149, 218]) showed that people valued using different types of visualizations in order to get different viewpoints and perspectives of their personal data.

Implementing an experimental visualization study includes many phases starting from the design, tasks selection, participants recruitment, method of evaluation, analysis and reporting on the outcomes [77]. In the context of personal health visualization, designing experiments and evaluation studies overlap with different fields, as highlighted in the introduction of this chapter. The basic concepts of information visualizations with respect to the perception, cognition, visual tasks and methods were discussed in the previous section, while personal informatics and visualization were discussed in section 2.2.

2.5 Summary

This chapter presented an overview of the research fields related to the scope and vision of the thesis, which has been classified into three primary sections. It started by defining personal health data in general and then went on to describe the specific type of data that this thesis studies: personal health data collected by health tracking devices. The first section also presented the personal informatics models and outlined the role of personal visualization in the

process, before moving on to discuss related work in personal health visualization. It covered the research that focused on the representation of tracked health, which helped in defining the research problem by identifying the limitations in the available methods and understanding what the data visualizations should offer for the user.

The second section, the information visualization, covered basic aspects required to understand human factors such as cognition and perception, task taxonomies and the methods used to visualize time series data. It focused on the Linear and the Radial methods for representing the type of the data and discussed their strengths and weaknesses based on previous evaluation studies. The research reviewed in this section contributed in shaping the structure of evaluation study and the design of the developed visualizations. The last section, an overview of the evaluation of information visualization was presented and then went on to examine the evaluation of personal visualization. This section guided the design of the evaluation study conducted in the thesis and the used metrics in the evaluation.

This chapter has focused on academic research papers related to the work in hand. However, to strengthen the thesis background and highlight existing limitations in visualizing personal health data, it was necessary to identify how the data are currently represented to the users (i.e. how the data is visualized on the devices' screens and companion apps and dashboards). Therefore, a systematic review study was implemented to cover these aspects, this review is presented in Chapter 4 of this thesis in addition to the user study in Chapter 3.

Chapter 3

Analysis of the Use and Preferences of Tracker Users

3.1 Introduction

The literature review chapter presented research studies that considered the visualizations of personal health data and other research papers that studied people behaviour in health tracking. As the scope of this thesis is focused on evaluating the visualization methods for the specified data, further investigation regarding both tracked data and its representation was required.

The user study presented in this chapter aimed to understand users' tracking behaviour and their preferences: what data do they track most? How frequently do they check the tracked data (daily, weekly or monthly)? Another aim was to identify their interest in seeing the visualization of different personal health data and their preferred designs. In addition to the information gathered from the literature review, an online survey was designed to collect more information about potential users' needs and preferences. This is a fundamental step to guide the research through the process of data visualization design choices for the evaluation study.

The chapter is organised as follows: it begins by presenting the purpose and aim of the study in section 3.2, followed by a description of the research methodology and the participants' recruitment in section 3.3. The analysis and results are shown in section 3.4, followed by a discussion of the study outcomes in section 3.5 and finally the conclusion.

3.2 Aims and Research Questions

As highlighted earlier, the aim of the study was to understand users' behaviour in health tracking and their preferences regarding different visualization designs. Achieving this aim is essential for designing and evaluating the visualizations for personal health data, which is the scope and the focus of the thesis.

The study includes aspects such as the most tracked data and used devices, how frequently people check the tracked data (e.g., daily or weekly and identifying people's interest in the visualization of different types of health (this is **(RA1)**).

The research question to be answered by this study is: What types of health data do people most track and what is their behaviour in tracking, their needs, requirements and preferences in

the visualization of the tracked data? (i.e., **(RQ1)** presented in the Chapter 1).

3.3 Methodology

For human-centred design, the designer should clearly understand the problems of the target users through engaging with the target users using several methods such as interviews to determine their requirements [155]. Questionnaires are a common research method that is used to complement other research methods [77]. In this thesis, an online survey was used to collect the required information from the participant before conducting the main evaluation experiment. This information includes necessary details about the data that will be visualized and the visualization methods that should be considered in the evaluation experiment.

3.3.1 Survey Design

The survey consists of three sections: one is related to health and activity tracking behaviour; the second is related to the visualization of this data type and the third consider the sharing of the data. The survey questions are presented in Table 3.1.

The survey started with taking the participants' consent to participate in the study before presenting the questions. As shown in Table 3.1, the next part includes one question about the participant age range, and the other questions about **health tracking**. These questions are (Q3 to Q6) and were aimed to collect data that could be useful for personal health data. Q3 considered the type of tracking device(s)/app(s) the participants use to focus on the most popular ones in the study. Q4 and Q5 asked the participants about the most tracked data and how much they are interested to see them visualized in order to identify what health data to consider in the visualization. As personal health tracking devices/apps collect various health metrics, these two questions were chosen to help in deciding on health variables to include in the visualization in this thesis. Q6 asked about how frequently the participants check the visualization to help in identifying the most suitable length of the data to be visualized and to recognise what level of view the visualizations should provide.

The **visualization part** includes questions (Q7 to Q10), and they were constructed to identify people preferences in the visualization methods. First, for Q7, it asked about the participants' preferences in the traditional charts (e.g., Bar and Line chart) and other abstract visualizations (e.g., using arts and metaphors). The use of traditional charts and abstract representation for personal health data was discussed earlier in section 2.2.3. The questions Q8 and Q9 requested further information about the preferences in other visualization methods that are related to traditional charts and abstract visualization. Q10 aimed to get more details about the participants' opinion and preferences in the selected visualizations.

The **sharing part** includes Q11 and Q12, which are similar to Q7 and Q10 in the visualization part. The sharing part aimed to determine the participants' view of sharing the visualization of personal health data. The sharing feature is critical in personal health due to the sensitivity of the data. Therefore, this part was added to identify the difference in the participants' preferences in the visualizations methods used for personal use and for sharing purposes.

3.3.2 Procedure and Participants Recruitment

The online survey was conducted in the period between November and December 2016 to study users' preferences in physical activity tracking generally and in visualization and sharing data specifically. The survey was designed and constructed using Google Forms survey service, which was chosen because it is free and supports the basic features required for this study. After designing the survey and constructing the questions, the study was approved. Permission to conduct the study and circulate the survey online was given by the ethics committee of Swansea University, Department of Computer Science. The survey was circulated among different mailing lists, and to people who were interested in health and activity self-tracking (such as the Quantified Self forum and the Sports Union mailing list at Swansea University). The study was voluntary, and no personal information was required.

Demographic Information: 84 participants responded to the survey, of whom 89% specified what devices they used for tracking, while around 11% did not use any tracking device. 81 participants out of the total number of 84 participants indicated their age range: 36% of the participants were aged between 20 – 30; 21% were aged between 30 - 40 years old and 28%

Table 3.1: The Online Survey Sample (Questions and Types of Responses)

<i>Section</i>	<i>Questions</i>	<i>Response Type</i>
Consent	Q1: Agreeing to participate in the study?	Checkbox
Demographic & Tracking	Q2: Please select your age range.	Multiple options
	Q3: What activity tracker (brand) are you currently using to track your activity? Check all that apply.	Check list + free text for other options
	Q4: Which of the tracked data are you more interested in?	Check list + free text for other options
	Q5: How much are you interested to see the visualization of each of these data? Mark only one oval per row	Multiple options
	Q6: How frequently do you check your tracked data?	Multiple options
Data Visualizations	Q7: What type of visualization do you prefer, and you think it will be more helpful and effective to represent your data?	Multiple options (statistical charts, Abstract visualizations, both) supported by examples and figures.
	Q8: In case of Traditional Statistical Charts, which one do you prefer?	Check list (Bar chart, Line chart, Pie chart, Radial Pie Chart) supported by figures
	Q9: In case of Abstract Visualization, which one do you prefer?	Check list (Flowers and Garden Metaphor, Living Metaphors (e.g., Fish or other pets), Clock and Calendar Metaphors (Time Related), The Use of Abstract Art (e.g., Spiral Visualization or Rings) supported by figures
	Q10: In a short sentence, could you tell why do you prefer this type of visualizations?	Free text
Sharing	Q11: What type of the following visualizations you think is suitable for sharing and you will share it with family, friends or post in social media?	Multiple options (statistical charts, Abstract visualizations, both) supported by examples and figures.
	Q12: In a short sentence, could you tell why do you prefer this type of visualizations for sharing?	Free text

were equally distributed between the 40-50 and 50-60 age groups. This means about 57% of the participants were adults and young adults, 28% of the participants were middle aged, whilst 10% were seniors and only 1% was under the age of 20. The remaining 4% represents the participants who did not provide their ages. The responses for the survey were collected from different age groups. However, the age data did not affect the general data analysis which considered all participants' responses regardless of their ages (i.e., age is not considered in the general data analysis). For the other portion of the analysis, which is based on age groups, the analysis is restricted to the responses which included age data. Figure 3.1a shows the age distribution for the participants.

3.4 Analysis and Results

The analysis is classified into two main sections: tracking behaviour, data visualization and a sharing section. Each includes the results for the related questions. The discussion of the results does not include the educational and professional backgrounds of the participants as they were not requested in the personal data sections and the questionnaire was anonymously circulated online through different mailing lists.

As shown in Table 3.1, the types of responses to the questions in the survey are of different forms. Therefore, they were analysed differently according to their types. The multiple options and checkbox responses (i.e., Q2-Q9 and Q11) were analysed based on the selections count and the percentage (i.e., how many each category selected with respect to the total number of respondents). For Q10 and Q12, a content analysis was applied [187] to code the textual responses and classify them into basic categories and sub-categories. The analysis performed while focusing on identifying the reasons the participants gave for selecting their preferred visualization(s).

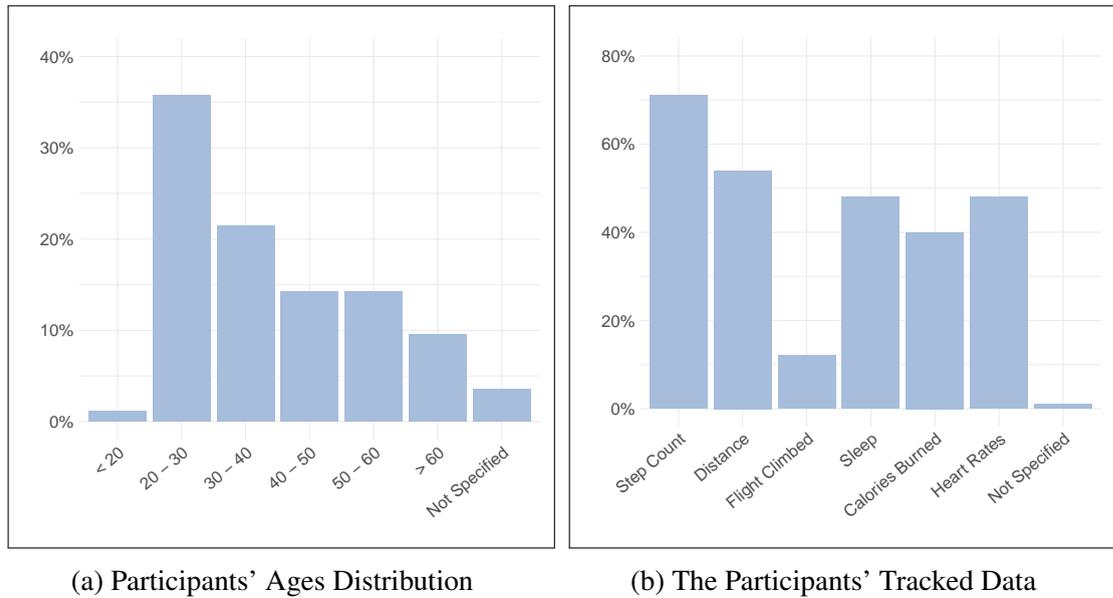


Figure 3.1: Participants' Demographic Information.

3.4.1 Activity Tracking Data

The questions that are related to investigating the participants' tracking activity and behaviour are questions Q3, Q4, Q5 and Q6 in Table 3.1. Out of the 84 participants, 69 specified the trackers they were using and some reported using more than one device, 9 were not using any tracking device and 6 participants were using other devices than the ones listed in the question and did not specify what they were using.

As presented in Figure 3.2, Fitbit was used by 38% of the participants, this is the highest figure followed by the Apple Watch which was used by 13% and then the Garmin tracker which was used by 7% of the participants. Around 14% indicated using a different phone and iPhone apps to track their activity. The Jawbone and Withings were used by 4%. Each of the other devices are presented in Figure 3.2 such as Strava and Misfit, and were used by 1-2%, while 7% of the participants used other unnamed tracking devices. The number of Fitbit users compared to other devices could explain why the majority of the studies in the literature review discussed earlier in Chapter 2 applied Fitbit data to their visualization system such as [66, 115, 130, 149, 226]. The results of this question in the survey confirm the results of the survey done by Epstein et al. [60] which accepted responses from a variety of sources and was completed by 139 respondents.

The type of data tracked by the participants based on the tracking devices are summarized in Table 3.2. The results from Epstein et al.'s survey indicated that 104 of the respondents were using Fitbit trackers to track their physical activity data. However, their survey results were probably biased toward Fitbit as the survey was circulated to the Fitbit community only.

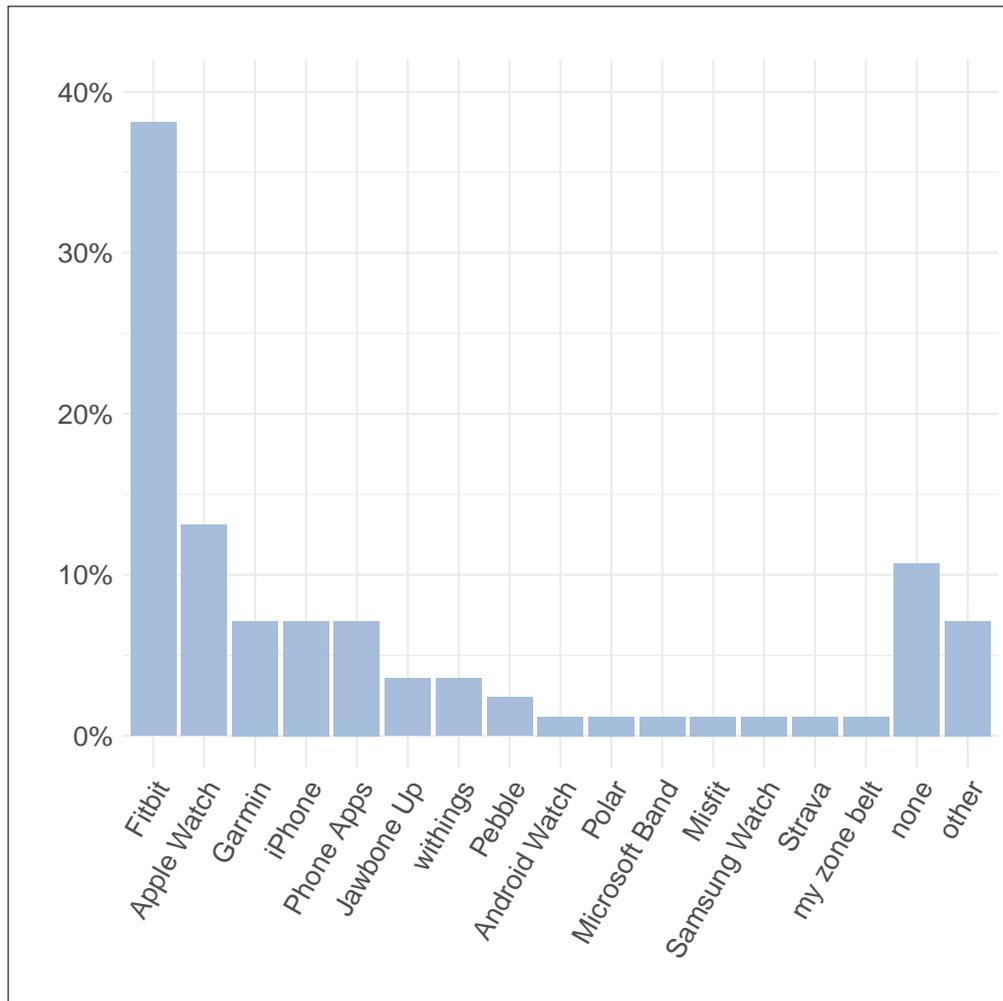


Figure 3.2: Participants' Used Tracking Devices.

The questions (Q4, Q5 and Q6) identify the following: the data that were of most interest to the participants to track; the data they would like to visualize and how frequent each participant checked each of the tracked data. For Q4, the responses showed a high interest in tracking the step count which was tracked by 71% of the respondents, while there was no significant interest in the flight climbed as it was tracked by only 12% of the participants (see Figure 3.1b). It was found that the number of people who tracked sleep was the same as the number of those who

Table 3.2: The Data Tracked by the Users of Different Tracking Devices

Device	Steps	Distance	Flights	Sleep	Calories	Heart Rate
Fitbit	26	11	3	21	10	15
Apple Watch	6	7	1	3	7	7
Garmin	5	3	1	3	3	4
iPhone	6	5	2	2	2	3
Phone apps	2	6	0	2	4	2
Jawbone UP	2	0	1	3	1	3
Withings	3	2	0	2	2	1
Pebble	1	1	1	2	0	2
Android Watch	0	1	0	0	0	0
Polar	0	1	0	0	0	1
Microsoft Band	1	0	1	1	0	0
Misfit	1	1	1	0	1	0
Samsung Watch	1	1	0	0	0	1
Starva	1	0	0	0	0	0
Myzone Belt	0	1	0	0	1	1

tracked heart rates 48% (39 participants), whilst 25 of the 39 participants reported tracking both sleep and heart rates. The devices used to track the data including sleep and heart rates were variant.

Question (Q5) asked the participants to rate their interest in seeing the visualization of their tracked data. As presented in Figure 3.1b, around 45% of the participants indicated a high interest and 31% indicated a moderate interest in the visualization of the step count and distance, while 42% of the participants indicated a high interest and around 25% indicated a moderate interest in the visualization of the heart rate and sleep data. The visualization of calories burned was of a high interest to 33% and of a moderate interest to 31% of the participants. The participants interest in the visualization of flight climbed data was declined when compared to other data, 35% of the participants were not interested at all, 31% indicated a low interest, 24% indicated moderate interest while only 12% of the participants were highly interested in the visualization of the flight climbed.

For Q6, it asked the participants how frequently they checked the tracked data. The results show that 48% of the respondents checked the step count daily, 21% weekly and 7% monthly while 24% of the respondents stated that they did not check the step count at all. Regarding distance, 43% checked it daily, 26% checked it weekly and 6% checked it monthly, while 25% of the respondents never checked the distance. Sleep was checked daily by 32% of the respondents,

11% weekly and 7% monthly, while the other 50% never checked sleep data. Around 35% of the respondents checked their heart rate daily, on the other hand, 48% of them never checked it. The other 18% of the respondents divide between weekly and monthly checking of their heart rate. 31% of the respondents checked the calories burned daily, 20% weekly, 11% monthly and 38% never checked calories burned. The flight climbed was considered to be the least checked data as it was not checked by 67% of the participants. These figures show that the participants who were interested in each of the data were more likely to check the progress daily, thus this highlighting the importance of providing a detailed view for one day of physical activity data.

3.4.2 Data Visualization and Sharing

The main question in the visualization section is (Q7) in Table 3.1, which aimed to identify the best method for the participants to visualize their data. The responses to this question show the majority (63% of the participants) preferred to visualize their data using traditional statistical charts. For all the 84 participants, 53 chose traditional statistical charts, only 1 participant chose the abstract visualization, while the rest (around 36%) suggested that both would be helpful. Regarding which type of chart is preferred (Q8), the highest preferred was the line chart (57%) and then the bar chart (52%), while the least preferred were the pie chart (18%) and the Radial pie chart (12%), as presented in Figure 3.3a. 6% of the participants did not indicate any preference.

Less participants showed an interest in the presented types of abstract visualization, as around 63% preferred to visualize their data on traditional statistical charts rather than using abstract or both methods. For Q9, 41 participants preferred the clock and calendar metaphor, which was the highest preferred type of abstract visualization (49%), 27 participants (32%) choose abstract art, 14 (17%) preferred living metaphors and only 9 participants (11%) preferred flowers and garden metaphors. 17 participants (20%) had no preference amongst any of the presented abstract visualizations (see Figure 3.3b).

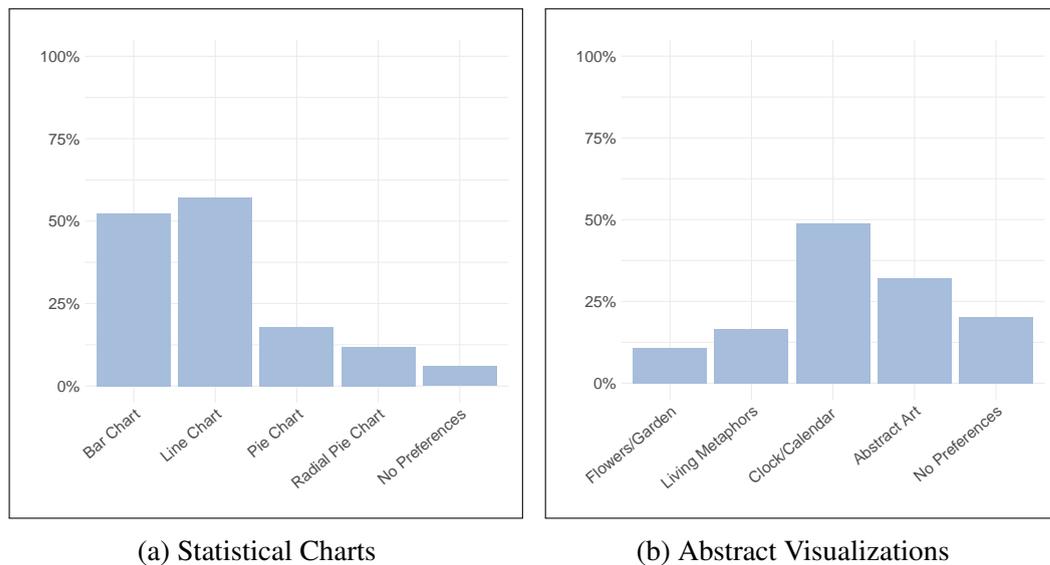


Figure 3.3: Participants' Preferences in Abstract and the Statistical Visualizations.

The responses for the data visualization open-ended Question (Q10) were analysed and coded. As can be seen from the answers to this question (summarised in Figure 3.4), there was a very high interest from the participants in having a good level of comprehension of their data, indeed, this may explain why the majority of the participants preferred to see their data visualized using the traditional statistical charts.

When looking at the results in detail, 47 participants out of the 84 gave their reasons for the chosen type of visualization. The short answers given included various topics, the three main trends that were identified being: (1) Data Comprehension and Gain Knowledge; (2) Aesthetic View and Entertainment and (3) Personal Preferences with sub-topics in the first two categories. Each response could include more than one of the listed categories. In general, the provided responses mentioned 49 different reasons related to the data comprehension and the knowledge gained from the visualization in either a positive or a negative manner. 14 responses discussed the aesthetic view and the entertainment of the visualization and only 3 responses provided some reason for their personal preference. For the topics data comprehension and gain knowledge, 20 participants gave a “clear and easy to understand” description for their chosen visualization, 10 wanted to “extract knowledge” from the visualization, this included (progress,

comparisons and finding patterns in the data), 7 participants mentioned the “familiarity” of the traditional visualization so they would not be required to interpret new visualization symbols, 6 wanted to get a “quick overview” while 3 wanted to have a “detailed view” of the data, 3 participants gave their reasons for not choosing an abstract visualization as it is “confusing and ambiguous”.

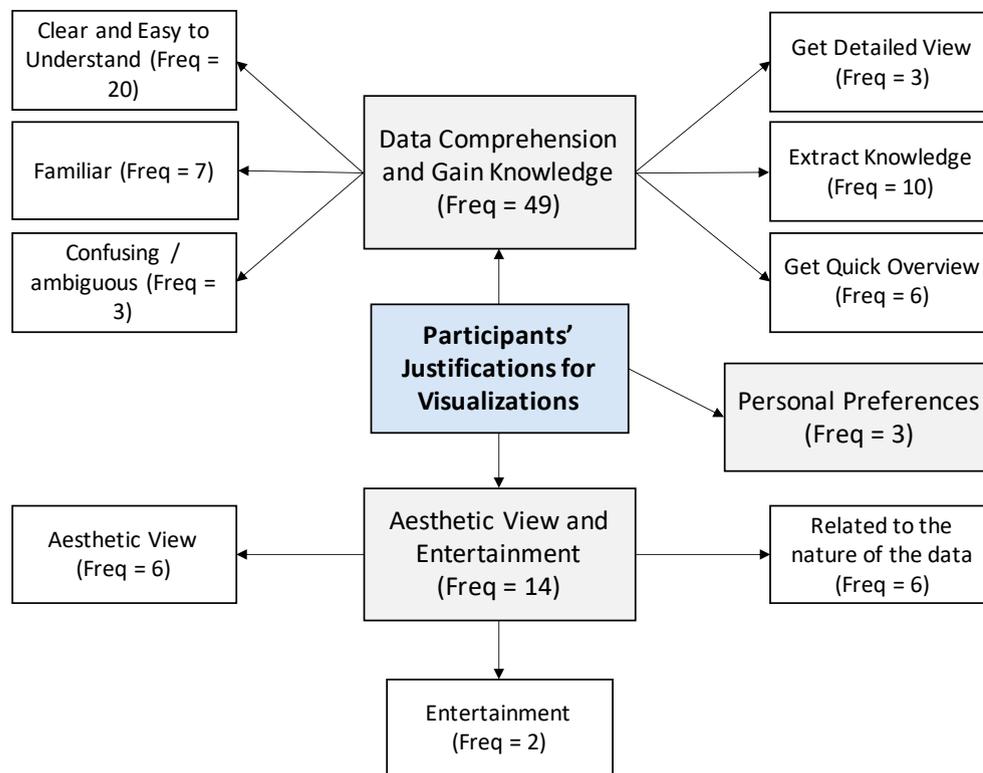


Figure 3.4: Participants’ Responses to (Q10) in the Data Visualization Section of the Online Questionnaire. The diagram shows the basic themes and the sub-themes identified in the open-ended responses.

The question presented in the survey related to sharing (Q11) was designed to identify what type of visualization the participants preferred when sharing their personal physical activity data. 38 of the 83 participants chose the statistical charts, 9 preferred the abstract visualization and 27 thought both would be helpful. We can see from the chart (Figure 3.5) that the visualization preference for some of the participants was different than for sharing, some participants were not interested in sharing (11%) and few participants preferred statistical charts, which significantly decreased from 63% to 46%. However, as only 1 participant chose the abstract

visualization for the data representation, the number who preferred an abstract visualization for sharing the data increased to 10 participants.

When comparing each participant's answers to both questions (Figure 3.5), 49% had the same preferences for both personal use and sharing, about 40% had different preferences and the remaining 11% were the participants who did not want to share their data.

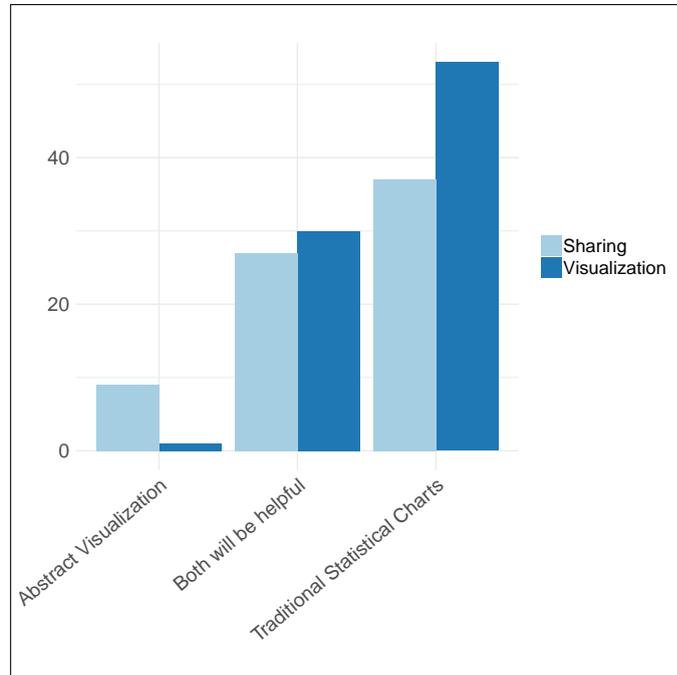


Figure 3.5: A Comparison Between Participants' Preferences for Data Representation and for Sharing.

For question Q12, similar to the result from Q10 in the data visualization section, the majority of the responses had a high interest in data comprehension and gaining knowledge from the visualization (a total of 35 different terms from 49 responses). The responses have been classified into four basic categories, the first three are the same for data visualization, which are: (1) Data Comprehension and Gain Knowledge; (2) Aesthetic View and Entertainment; (3) Personal Preferences with sub-topics in the first two categories and (4) Not Interested in sharing (the responses classification are shown in Figure 3.6).

All sub-categories for the first two categories are similar, except for an additional sub-category in (1) Data Comprehension and Gain Knowledge, which was "Accurate" and an additional "So-

cial” sub-category in the second one. The social responses were in the form of “Discussion”, “Comments” and “Starts conversation” and they all were related to an abstract visualization. The traditional and statistical charts were selected by some participants because they are easy to understand and familiar, so there was no need to explain to others what they meant, while on the other hand participants who chose an abstract visualization thought that it was more social, as showed earlier in their responses. However, one response preferred an abstract visualization in sharing physical activity data because “there is no need for details”.

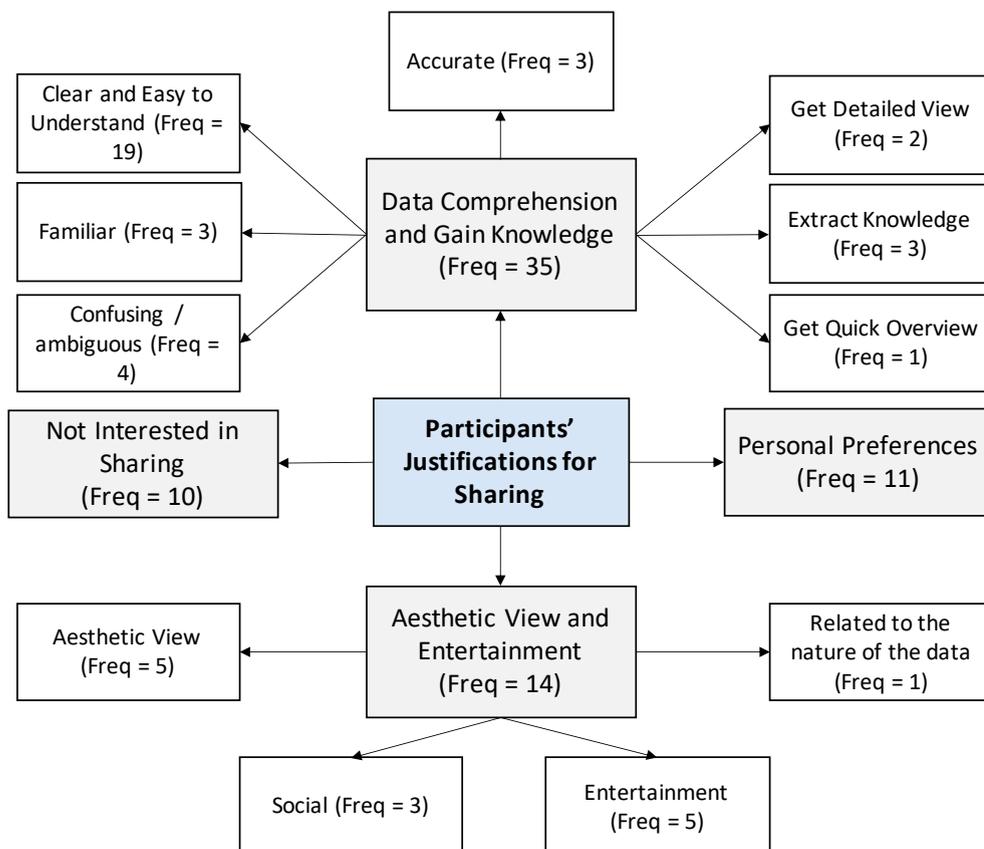


Figure 3.6: Participants’ Responses to (Q12) in the Sharing Section of the Online Questionnaire. The diagram shows the basic themes and the sub-themes identified in the open-ended responses.

3.5 Discussion

The results are discussed with respect to the type of data that needs to be visualized, the visualization methods preferred by the participants in light of related studies.

Types of the Data to be Visualized: Most of the related literature discussed in the literature review in 2.2.3 focuses on the visualization of one sort of physical activity data, which is the step count as [66], or merge it in a visualization with other types of data from different sources (such as locations, weather, appointments, etc.) and discard other forms of physical activity data such as heart rates and calories burned. Huang et al. [115] created an on-calendar visualization that adds the visualization of the Fitbit step count data as an additional layer on a personal calendar. The research by Larsen et al. [130] and Epstein et al. [60] both created a visualization system for the step count and location data. On the other hand, Tong et al. [226] included calories burned as a form of activity data in their research which compares three different types of visualization for physical activity data. Meyer et al. [149] considered more than one physical activity data in their visualizations, they included the step count, calories burned, sleep and weight. The responses from the data section in the questionnaire and the lack of literature covering the visualization methods of heterogeneous health data encouraged the consideration of different types of physical activity data in this research.

Types of the Visualization Methods: The result of the first question in the data visualization section showed that about 63% of the participants preferred to use traditional statistical charts to represent their physical activity data. This result is reasonable, as these charts are familiar and most of the tracking devices apps use these charts to represent the user's progress (e.g., the Fitbit Bar chart and Apple Activity Radial chart). Choe et al. [45] in their research studied the insight the Q-Selfers could obtain from visualizing their data. They identified different types of visualization used by the Q-Selfers in their talks. The result showed that the "Line chart" is the most frequently used visualization (frequency = 71), followed by the "Bar Chart" with frequency = 33 and then the scatterplot with frequency = 23 [45]. Looking at the results of our survey, the line chart is the most preferred type of chart (57%) followed by the bar chart (52%). The fact of the "Line Chart" being the most frequently used visualization is

not surprising as the self-tracking data considered to be a time-series and the line chart was used to find the trend in these data [45].

In Fan et al.'s [66] study that evaluated the abstract art visualizations (the Spiral, Rings, Bucket, Pollock) and Bar chart. The participants preferred the abstract visualization for an aesthetic and glanceable display, while the bar chart was required to gain a detailed view of the data or looking for specific information. Tong et al. in their research [226] evaluated three visualizations for physical activity data: the Circular Ringmap; Virtual Pet visualization and the Fitbit bar chart. Although their participants ranked the Fitbit bar chart at the top in terms of readability, awareness and comprehension, they reported that it was difficult to use it for finding a pattern in the data while the Circular Ringmap facilitated the process of finding patterns in long-period data with a detailed view. Therefore, this showed that different types of visualization provide different ways of understanding the data.

3.6 Conclusion

This chapter presented a survey for understanding people behaviour in tracking and viewing personal health data in addition to their preferences in the visualization methods. The study sought answer to the first research question (**RQ1**), which was identified through the outcomes of the study as shown in section 3.4 and section 3.5. The following chapter presents a follow up study that explores the visualizations methods of personal health data and highlights their limitations with respect to people needs.

A Review of Visualization Tools for Health Data

4.1 Introduction

The review of relevant literature in Chapter 2 of this work set out the different types of research related to the topic of this thesis. It revealed that although there have been many research articles discussing information visualization in general and the visualization of personal health data specifically, there has not been sufficient discussion or research regarding the existing visualizations provided by commercial health tracking devices and their connected platforms, their types and their limitations.

This chapter aims to build a bridge between published research, exploring users' experiences of self-tracking devices and the information collected from manufacturers' websites, user manuals and online communities and Weblogs with respect to personal health data visualization. In addition, it links the outcomes with the results from the online survey that was presented in the previous chapter.

The contribution of this chapter differs from the previous data visualization studies in personal informatics (discussed in section 2.2.3) in two ways: firstly, it explores the visualization methods that are available on the most popular tracking devices and applications with a particular focus on health and physical activity data; secondly, it examines whether the current visualizations supported by such devices meet users' needs derived from the previous studies and the user study in this chapter, it also reports on what is missing in personal health visualization design space. This was a preliminary step that could contribute to the structuring of the design space of personal health visualization and the improvement of the current visualization designs. It also helped with the design choices in the evaluation study in this thesis.

The chapter is organised as follows: the study aims and research questions are presented in section 4.2; the methodology is presented in section 4.3. Then the results and analysis in section 4.4, and followed by a discussion of the findings in section 4.5. After the presentation of the study, the chapter goes on to discuss potential suggestions to address the limitations in the visualization methods for personal health data. These were based on the outcomes of the studies in Chapter 3 and 4 and related literature in section 4.6. After the discussion, section 4.7 presents an insight into designing visualizations to address some of the highlighted limitations and, therefore, guiding the evaluation design of the study in this thesis. As the review study

was an early step in this PhD project, an update was performed later and summarised based on the reviewed devices in section 4.8. Finally, the conclusion of this chapter is presented in section 4.9.

4.2 Aims and Research Questions

The study presented in this chapter addresses related questions concerning people's needs in the visualization of data collected from the tracking devices and explores what these devices already offer. This was a follow up step done to understand the gaps with respect to the visualizations, especially as several studies on people who track their data (such as [44, 134, 135]) reported that people were unable to fully benefit and reflect on their data due to limitations in the visualization methods supported by the tracking tools they used.

Further to the related literature on the topic, the aim of the review study is to identify the current visualization methods provided by health tracking devices and their associated apps/dashboards, as well as to identify the limitations in these visualizations: both their design aspects and users' needs and requirements: **RA2**, which was presented in the introduction (see section 1.4).

Therefore, the research question answered by this study **RQ2** is: What are the visualization methods supported by most popular tracking devices and their companion apps and dashboards, what are their limitations and what needs to be addressed?

The review builds on previous studies that focus on the reflection stage of the personal informatics model [134], which are [135] and [44], to discuss the visualizations. It explores the visualization methods provided by four popular self-tracking devices and their companion apps in the context of personal health tracking. It discusses the issues and limitations of such devices with regards to data visualization, interpretation, reflection and sharing. It provides an updated view of the typology of the automatically tracked data, how these data are presented on the different platforms and what level of view each visualization provides. This research may provide some insight regarding how to represent health and physical activity data. As the term 'personal health' is broad and includes different types of data, the review only considered health data that are automatically tracked by self-tracking devices (i.e., physical activity, sleep

and heart rate) in discussing the visualization layouts; health data manually entered, such as nutrition, smoking and alcohol intake are not included in this review.

4.3 Methodology

The first step in the visualizations review process was to select the devices from which the health data is collected, this is described in section 4.3.1. The second step was to decide which methods to adopt and which resource to use in order to review the visualizations, this is discussed in section 4.3.2.

4.3.1 Device Selection

The self-tracking devices shortlisted and selected, as reviewed in this study, were based on two sources: the IDC report on unit shipment and market share of wearable vendors in Q4 2016 [118] and the results from the online survey that was conducted to understand physical activity tracking and the devices used in addition to trackers' preferences in visualizing their data (discussed in section 3).

The results presented earlier showed that the most used tracking devices in order of use were: Fitbit 38%; Apple watch 13%; Garmin 7%; iPhone 7%; Android phones 7% and Jawbone and Withings at 4% each.

The "IDC Worldwide Quarterly Wearable Device Tracker" report published on March 2, 2017 [118], provided valuable information on the worldwide popularity and sales numbers of the devices. In this report, the top 5 wearable devices in Q4 of 2016 based on unit shipments and market share were ordered as: Fitbit; Xiaomi; Apple; Garmin and Samsung [118].

Based on our Online Survey and Wearable Device Tracker's market [118], the Fitbit, Apple watch and Garmin were shortlisted as they had the highest popularity rating. In addition to these three trackers, Samsung was shortlisted as it is the vendor of the fifth top wearable devices. Mobile phones were excluded as the focus of the research was on wearable trackers, as was

Xiaomi because the brand was not widely recognised outside of the country in which it is manufactured [118].

The intention was to review the visualization displayed on the device screen in addition to the visualizations displayed on their companion app interfaces and online dashboards; therefore, the latest version of each of the devices that have larger displays, i.e., smart watches, from our shortlisted vendors were chosen, they were selected to ensure that only devices with identical features were reviewed. Therefore, the devices selection criteria are:

1. Choosing 4 of the most popular activity tracking devices based on IDC report [118] and the result of our Online survey.
2. From these vendors, we are selecting the most recent trackers that has large display and track more than two physical activity data.

According to these criteria, the chosen devices were: the Fitbit Blaze; Garmin Forerunner 935; Samsung Gear 3 and the Apple Watch series 2.

4.3.2 Device Review

Given the rapid development of tracking technology and the human-centred focus of this work, it was decided the traditional revision of scientific and technical literature should be enriched by including contributions from devices' official websites and user manuals, users' online communities and forums, in addition to some online users' reviews of some devices, apps and dashboards. When reviewing the data visualization features of the Fitbit Blaze, its User's Manual [68], Introductory Guide [74] and videos from the Fitbit YouTube Channel [76] (e.g., [69]) were relied upon. When reviewing the visualization on the app and on the dashboard, the search was extended to include more general sources on the Fitbit, which are related to the Fitbit trackers in general; for example, all Fitbit trackers synchronise to the same companion app and the dashboard webpage (Fitbit, 2017). To review the visualization on the Garmin Forerunner 935 display and its companion app and dashboard, its User's Manual [83] and the related videos on Garmin YouTube channel [87] were relied upon. Due to a lack of screenshots of the device display and the graphs of the data, additional sources that reviewed the watch e.g., [144, 200] were

included. Moreover, to obtain more details about the visualization on the Garmin Connect app and dashboard, the reviews of other versions of the Garmin forerunner e.g., [57, 80, 81, 110] were included. When reviewing the Samsung Gear S3 and Samsung Health app [193], several sources were referred to: the Gear S3 User's Manual [189]; related videos on the Samsung YouTube Channel [192] (e.g., [188]); various articles on the Samsung Health app and Gear S3 posted on the Samsung newsroom [194, 195, 196, 197, 198] and other sources that reviewed the watch such as [233] and [10], in addition to the researcher's personal experience using the Samsung health app itself without the watch.

In addition to these listed resources, there were cases where the researcher owned a device or had access to an app or a device's companion dashboard. In these cases, the researcher collected the required information through using and reviewing the visualizations on these platforms. This method was used to strengthen reviews of the following: the Apple Watch; iOS Health App; iOS Activity App; Fitbit app on an iOS platform and the Fitbit Online dashboard (without the Fitbit device), the Garmin Connect app and dashboard and the Samsung Health App (the basic app without using the watch), in addition to the sources listed on each device review that is presented in the following sections. Visual layout choices within the Fitbit Community are highly dependent on the type of data collected by the specific device. To ensure an unbiased review, different details were extracted from the varied selected data sources (including screenshots and videos) regarding how the same tracked data were actually presented across the devices and their platforms, personal opinions or experiences were not considered.

4.4 Results and Analysis

After selecting the devices and collecting the data sources to review the visualizations, a profile for each device was created by capturing different information about the data and how they are presented. Firstly, the visualization methods were based on the platform(s) that support any sort of visualization for the data tracked by the device (e.g., the device's screen, Online dashboard). Although all shortlisted devices in this study track similar activity data (e.g., step count, distance, flights climbed, heart rate (HR) and calories burned, in addition to exercise details such as running and cycling), they vary significantly in the way they visually represent

the activity data. Therefore, for each platform, the following information was captured:

1. What data types are visualized?
2. How different physical activity and health data are presented? And what visualization method(s) are used?
3. The length of data being visualized: current/recent data visualization versus history/past data visualization.

Lastly, the collected information was summarised and categorised using a multi-dimensional table, which is presented in Table 4.3. The different dimensions present different aspects of the data visualization methods.

For each device, the visualizations were reviewed for each companion platform separately. For example, for the Fitbit Blaze, the review explored the visualizations on the app/dashboard and the device's screen separately. The review process for each platform started with the tracked data. All the data types are presented in the top row of Table 4.3. Then, each row in the table presents the visualization methods related to a specific platform for each type of tracked data. The visualization methods in Table 4.3 were coded according to their initials (e.g., Bar chart was abbreviated as BC). Table 4.1 presents all the visualization methods and the abbreviations used. The colour of the cells represents the length of the data being visualized. The darker the colour of the cell, the longer period of data being visualized. The used colour codings are explained in Table 4.2.

The following section discusses visualization methods used by each device including the visualizations on the tracker's screen, its companion app and/or on the dashboard when provided and each visualization method is categorised according to its dominant layout.

4.4.1 Data Visualization Methods Taxonomy

Table 4.3 summarises the data visualization methods used in representing data collected by each device. The abbreviations for the visualization methods are defined in Table 4.1. The

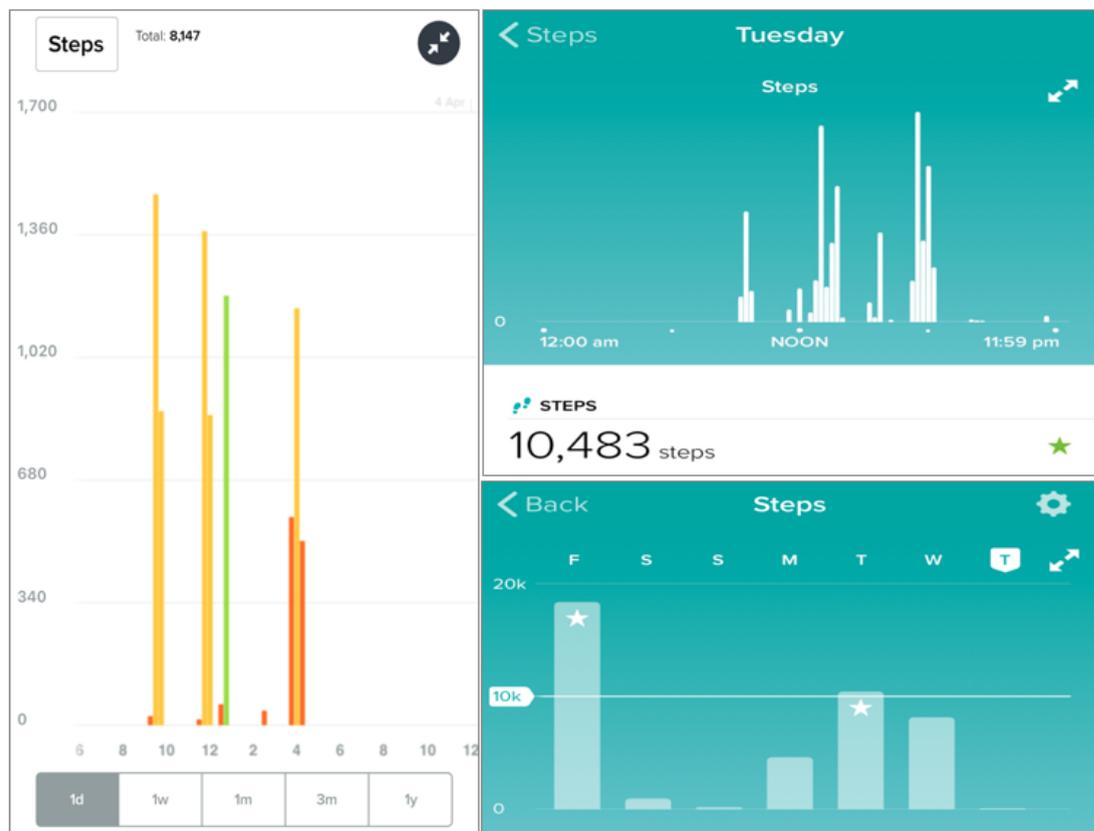


Figure 4.1: Three screenshots from Fitbit app. Left and top: show detailed Bar chart for steps during a day. Bottom screenshot shows Bar chart for steps of seven days. The star on the Bars means that the user met the steps goal for that day.



Figure 4.2: Fitbit Circular Rings for Steps, Distance and Calories Burned for a Day.

colour coding used in Table 4.3 are defined in Table 4.3 and they refer to the length of time of the data being visualized.

Several common representation methods were identified across the different platforms, such as the widely used traditional Bar and Line charts. The visualization methods can be grouped into five major categories: chart based; radial based; calendar based; metaphorical and high-dimensional.

4.4.1.1 Chart

Chart based layouts are common across all devices and companion apps that were reviewed and differences appear mostly at the level of display size and resolution. As shown in Table 4.3, the Fitbit app and dashboard provide different options for representing the data for the current day versus historical data. The “overview” of the data is mainly presented via Bar charts based on days and a detailed Bar chart can be viewed by selecting a specific day. Figure 4.1 and Figure 4.1 are screenshots of the Fitbit app viewing different visualizations for different data. Fitbit provides a coloured Line chart for HR data for a single exercise where the colours refer to the HR zones to show how the user is performing during the exercise, in addition to a Bar chart to represent the time in HR Zones (as seen in [69]).

The Bar chart and Line chart are the preferred visual layouts of Apple Health app [13] when dealing with daily data at high resolution e.g., hourly based, or to provide a historical overview e.g., week vs. month vs. year. The graphs and statistical charts provided by Garmin are more focused on exercises rather than viewing a historical overview of users’ activity levels over time. It offers a number of detailed graphs for physical exercises that are not available on the companion apps of other devices. For example, Line chart representing HR information for a single day of physical activity (see Figure 4.6) has colour coding that represents time in HR zones, in addition to the glyphs on the time axis that show the type of exercise that has been performed and the sleep time. It also represents HR data in different graphs (Bar Chart/Line Chart) that show time in HR zones, even for a single exercise.

The Samsung Gear S3 employs a Bar chart to display the step count history where data collected in the past seven days can be viewed at one time. On the Samsung Health app, the historical

data can be viewed on the trends section. The Bar chart visualizes one-week of data (e.g., active minutes). Although the user can view past as well as current data, it displays only seven days' worth of data at one time on the screen.

4.4.1.2 Radial Projection

A Circular Ring that moves towards a daily activity goal provides its users with quick feedback on their activity progress throughout the day. On the Fitbit app, the user is also able to view the visualization of different data from one day up to one year.

On the Fitbit Blaze display, the user can see 'Circular Rings' of the current day's data, such as the step count and distance and these Circular Rings are also viewed on the Fitbit app/dashboard (see Figure 4.2). The visualization of activity data tracked by the Apple watch is mainly provided by two apps: Health app [13] and Activity app [14]. The latter is related to the watch itself; there is no online dashboard or desktop application, so the view is limited to the phone and the screen of the watch. The visualization on the watch display is a Circular Ring that combines three different variables with different colours (see Figure 4.3a). The red ring represents active calories burned, the green represents exercise duration in minutes and the blue represents standing hours.

4.4.1.3 Calendar

The Activity app on the iPhone shows Circular Rings on a calendar layout (see Figure 4.3b). The user can interact with the calendar to view the detailed Bar chart for each day. The Health app also uses the calendar as the main layout to view the information, but there is no visualization combined with the calendar view; the user can view the details by selecting a specific day on the calendar. The Garmin and Fitbit apps use a calendar view to show logged exercises. These calendars are interactive, and additional details on logged exercises can be viewed on demand.



(a) Four screenshots captured from Activity app running on Apple watch. (b) A screenshot from iOS Activity app

Figure 4.3: Visualizations on the Activity app running on iOS and Apple Watch. (a) Top left shows the Circular Ring that represents active calories, exercise minutes and standing hours followed by three screenshots providing more details on the activities (b) A Circular Ring visualization of one month's worth of activity data.

4.4.1.4 Metaphorical

Rich visual metaphors, such as icons, are a very powerful tool to convey multidimensional information; moreover, when the metaphor features a link that is familiar to the user, cognition and information retention are increased [28]. Samsung offers a distinctive visualization on the display of Gear S3, which uses a clock metaphor to represent the minutes of the user's activity during the day, along with sleep hours (as presented in [10, 195]). Colour coding is used to distinguish between sleep hours, minutes of healthy and light activity and the time when the user was not active. The user can interact with the visualization to view further details, such as calories burned. This visualization could provide the user with a quick view of the level of activity and sleep for the current day. The data visualized on the watch is not restricted to the current day's data only; the user can access the data from previous days.



Figure 4.4: A Bar Chart Visualization of One month’s Activity Data - A Screenshot from the iOS Health app.

4.4.1.5 High-dimensional

The Activity app provides details about different workouts, but the details are mainly presented in textual format. However, it presents a route on a map to visualize outdoor workouts, such as an outdoor run. Furthermore, the Activity app is more focused on physical activity data while the Health app includes the details and the visualization of other health data, such as sleep and nutrition. Although there are some useful visualizations for the data tracked by the Apple watch, such as the calendar view provided by the Activity app, there are some limitations such as the absence of HR visualization. The Garmin Connect dashboard provides an animated visualization for exercises (e.g., running or cycling) that simulates user activity. It presents the route on the map along with a Line chart for representing the speed and elevation during the exercise and coloured Bars for representing all data metrics that are changed based on the movement on the map, such as the total time and HR, as presented in [84].

Text and numerical representation of data is common to all devices. On the Fitbit Blaze, the user is able to view details of different workouts in text and number format, as well return to either app or dashboard for more detailed charts integrating textual and numerical information. The Activity app also provides details on different workouts, but the details are mainly presented in textual formats.

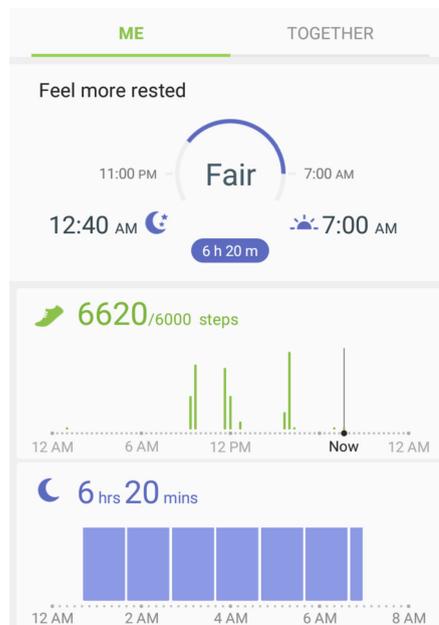


Figure 4.5: A Screenshot of Samsung Health App. On the app interface, a Bar chart represents step count of the current day's sleep.

4.5 Discussion

Visual representation methods play a significant role in reducing the cognitive load of data interpretation for users and enabling users to easily understand large data [131]. Personal health data could be of interest not only to the person himself but to others such as health providers, patient groups and family members [131]. The following sections discuss the limitations of the data representation methods on the reviewed devices and their connected displays.

4.5.1 Visual Layouts

To facilitate the discussion of the visualization layouts, the visualizations are classified into categories based on their designs, this is followed by a discussion of the main challenges in visualizing personal health data within each category.

Table 4.1: Abbreviations Used in Table 4.3

Visualization Technique	Abbreviation Used	visualization Technique	Abbreviation Used	visualization Technique	Abbreviation Used
Bar Chart	BC	Line Chart	LC	Icon	I
Circular Ring	CR	Route on Map	RoM	Animated Icon	AI
Text & Numbers	T&N	Calendar View	CV	Donut Chart	DC
Heat Map	HM	Clock Metaphor	CM	Pyramid	PY

Table 4.2: Colour Codings Used in Table 4.3

Colour Coding				
Daily/Current Day	Week	Per Exercise	Data > Week (visualization per day)	Data > Week (Day/week/ month/year)
The visualization represents data of one day only	The visualization represents data up to one week.	The visualization represents data of a single exercise.	The visualization represents data of more than one week, and it represents a single visualization for each day in the view (e.g., Figure 4.3b shows Circular Ring visualization for each day for one month of data)	The visualization represents data of more than one week, and it represents all the data in one visualization (e.g., Figure 4.4 shows Bar Chart that represent step count for one month)

Table 4.3: Main Visualizations Used by the Selected Devices on the Devices' Display, Dashboards and Companion Phone Apps. See Table 4.1 for the used abbreviations, and Table 4.2 for the colour codings.

Tracker Name	Platform	Step Count	Distance	Flight Climbed	Heart Rates	Time in HR Zones	Active Minutes	Calories/ Active Calories	Exercise/ Exercise Minutes	Walking/ running/ cycle	Pace	Elevation Gain	Stand/ Activity Hours	Sleep	
Fitbit Blaze (No GPS)	Phone App	CR/BC	CR/BC/T&N	CR/BC/T&N	BC/LC/T&N	LC/BC/T&N	CR/BC/T&N	LC/T&N	CV	RoM/T&N	BC/T&N	LC/T&N	CM/T&N	BC/HM/T&N	
		BC/T&N	BC/T&N	BC/T&N	BC/LC/T&N	LC/BC/T&N	BC/T&N	CR/BC/T&N						BC/T&N	CM/HM/T&N
	CR/BC/T&N	CR/BC/T&N	CR/BC/T&N	BC/LC/T&N	LC/BC/T&N	CR/BC/T&N	LC/T&N	CV						RoM/T&N	BC/T&N
	BC/T&N	BC/T&N	BC/T&N	BC/LC/T&N	LC/BC/T&N	BC/T&N	CR/BC/T&N		BC/T&N	CM/HM/T&N					
	Device Screen	CR/T&N	CR/T&N	CR/T&N	T&N/AI	T&N/AI	CR/T&N	CR/T&N	T&N	T&N	T&N	—	T&N	T&N	CR
		T&N	T&N	T&N	T&N		T&N	T&N							
Apple Watch (Activity App)	Phone App	T&N	T&N	—	T&N	—	T&N	BC/CR+T&N	BC/CR/T&N	RoM/T&N	T&N	T&N	BC/CR/T&N	—	
		T&N	T&N		T&N			CV+CR	CV+CR						CV+CR
	Device Screen	T&N	T&N	—	T&N/AI	—	T&N	CR+BC+T&N	T&N	T&N	T&N	T&N	BC/CR/T&N	—	
Health app (iOS)	Phone	BC	BC	BC	BC	—	—	BC	BC	BC	—	—	BC	BC	
		T&N	T&N	T&N	T&N			T&N	T&N	T&N			T&N	T&N	
Garmin	App/ Dashboard	CR/BC/T&N	LC/T&N	T&N	LC/T&N/I	LC/T&N/I	BC/T&N	T&N	T&N	RoM/T&N	LC/T&N	LC/T&N	—	CR/LC/HM/T&N	
		BC/T&N		BC/T&N	LC/T&N	LC/T&N	BC/T&N	T&N	T&N					CV/T&N	BC/HM/T&N
	Device Screen	CR/T&N	T&N	T&N	BC/T&N	BC/T&N	T&N	T&N	T&N	T&N	RoM/T&N	T&N	LC/T&N	—	—
		BC/T&N	BC/T&N		T&N		LC/T&N								
Samsung (Gear S3)	App/ Dashboard	BC/T&N	T&N	BC/T&N	T&N	—	CR/BC/T&N	BC/T&N	CR/BC/T&N	T&N	LC/T&N	T&N	T&N	CM/HM/T&N	
		BC/T&N	T&N	BC/T&N	LC/T&N			—	T&N					RoM/T&N	BC/T&N
	Device Screen	DC/BC/CR/PY/T&N	CR/T&N	DC/T&N	LC/T&N	BC/T&N	T&N	CM/DC/T&N	DC/T&N	RoM/T&N	LC/T&N	LC/T&N	T&N	T&N	CM/HM/T&N
		T&N	T&N												

4.5.1.1 Detailed View of Short-Term Data

Reflection on short-term data is valuable for making users aware of their status [134], to see if they are meeting their goal and whether they should take action toward their behaviour [135]. The reviewed displays in this research support short-term data visualization using different methods such as the Circular Ring and text. A detailed view of the data is provided for a day or an individual physical activity (e.g., a Bar chart representing step counts throughout the day). In addition, there is an overview that provides the user with instant feedback on the progress made towards a goal (e.g., a Circular Ring representing the progress made on being active).

For exercise visualizations, Fitbit app for example, provides a detailed Line chart for visualizing HR data during an exercise as (see [69]). Garmin supports a detailed representation of different variables of exercises using statistical charts (see Figure 4.7). The Garmin dashboard provides an animated visualization for exercises where each graph represents one data variable only (see [84]). Combining the different graphs into one visualization could be more efficient for comparison across visual representations since it would not require the user's gaze to move across different graphs to assess differences or similarities [121].

Another limitation in short-term data visualization appears in the lack of ability to personalise the visualizations to suit users' tracking purposes. For example, the main visualization for current data on Apple Activity app is designed for goal driven tracking. This is not an optimal visualization design for all users as they vary significantly in their tracking purposes [135, 185]. Moreover, users generally have no control over what variables to present in the visualization.

4.5.1.2 Overview of Long-Term Data

A lack of interest in personal historical data has been identified by many researchers [98, 185]. A potential reason for people being less interested in their historical data could be the difficulty in comprehending and making sense of the collected data, especially as tracked data grows massively and becomes complex over time. Data interpretation difficulties were reported by many Quantified-Selfers in Choe et al. [44]'s study which also observed that many people tried to simplify their tracking strategy after they failed in reflecting on their data.

Reflection on long-term data can reveal valuable information and finding patterns and trends [134]. In reviewed devices, statistical charts (especially Bar and Line charts) are the most common methods used to represent historical data, as presented in Table 4.3. A study by Asimakopoulos et al. [16] conducted with 34 Jawbone and Fitbit trackers users showed that the data visualization designed for these trackers' companion apps appears to be useful for the viewing of short-term data and tends to be less useful for the viewing of historical data.

Apple Activity app offers a different form of representation for long-term data using Circular Ring visualization on a calendar view (presented in Figure 4.3b). A main disadvantage of this method in visualizing long-term data is that full rings do not always reflect how the users are progressing in their activity as the users could change their goal settings over time. However, the user could identify each day's view by clicking on the Circular Ring for details.

4.5.1.3 Supporting Multivariate Data visualization

As discussed earlier, two out of five questions people ask about their tracked data are related to context and factors [135]. The context is to seek what could affect the present data, while factors seek what influence personal behaviour has over the long-term [135]. Finding answers to these two questions requires identifying correlations between different data and this depends on a users' ability to access additional data [135]. For example, a person with a weight loss goal may need to see how other factors, such as physical activity levels, affect weight.

Effective multivariate data visualization requires users to be engaged with the visualization to customise data selection. In the study by Meyer et al. [149], different forms of visualizations were evaluated for heterogeneous health data (step count, calories burned, weight and sleep) on a mobile application and the users showed interest in the development of integrated visualizations. The participants requested supporting analytical features in the visualization, such as data filtering, for more effective health reflection that could identify complex relations in the data. During the review carried out in this work it was found that long-term data, generally visualized using charts, mostly represents data for one variable only (e.g., step count) and those charts do not visualize combined multivariate activity variables. Multivariate data visualization is found on the Apple Activity app and on the Apple watch screen (Figure 4.3).

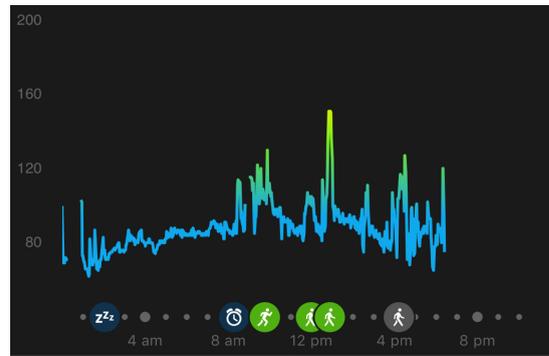


Figure 4.6: A Screenshot of Garmin Connect HR Details for a Single Day.

Multivariate data visualization for short-term data (i.e., context) was more supported by reviewed devices and apps than long-term data. A good example of multivariate data visualization for short-term data could be found in the Garmin HR Details (Figure 4.6), which combines HR levels with the user's activity modes e.g., running or sleeping. Another example would be the Samsung watch display that uses a clock metaphor to represent sleeping hours combined with minutes of activity throughout the day. The limitations in these reviewed visualizations appear in supporting appropriate visualization designs for long-term multivariate data and in supporting additional interaction features, such as allowing users to choose the length of data and variables to be visualized. Providing such features may help the users in generating and testing hypotheses about their personal behaviour [44].

4.5.1.4 Supporting Multiple Types of visualizations

Providing different choices of visualizations is important to enable the users to customise their experience [66]). Moreover, different visualization designs afford different perspectives and understanding of the data, therefore it may address the challenge of the diversity of users' requirements and needs [226]. Health data, such as sleep, are represented differently to the activity data on our reviewed devices. Fitbit, Garmin and Samsung use heat maps to encode sleep quality by showing the minutes of deep and light sleep against minutes of wakefulness (e.g., Figure 4.8), and heat maps of several days could be viewed as Bar charts to compare one night's sleep with a past night's sleep. The HR visualization on the Fitbit app for a single exercise (presented in [69]) gives two types of presentations for HR details: a coloured Bar



Figure 4.7: Four Screenshots from Garmin Connect App Showing Different Data Charts for a Running Exercise.

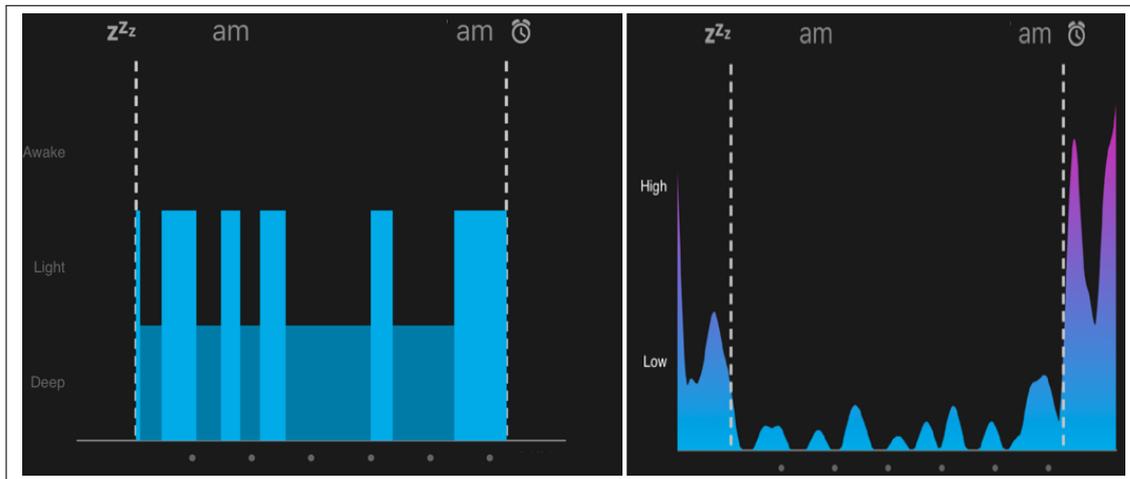


Figure 4.8: Two Screenshots of Different sleep Visualization on Garmin App.

chart and a coloured Line chart. The different charts convey different information about one data variable (i.e., HR). As illustrated in Table 4.3, long-term data are mainly presented in one type of visualization and, mostly, it is either through the Bar or Line chart with the aid of textual/numerical annotation. However, there is a significant limitation in supporting different methods of visualization for the different data types.

4.5.2 Sharing Health Data

The devices reviewed in this study have limitations in sharing visualized health data. They either allow sharing the data in textual form (e.g., sharing activity summary with non-Apple watch users) or restricting visualization sharing to users added in the social platform on the dashboard (e.g., Fitbit and Garmin Dashboards). However, the Fitbit app allows the sharing of the visualization of a specific day (as presented in Figure 4.2). However, the user has limited control over resolution and dimensionality of the data to be shared.

Sharing features provided by the shortlisted apps and devices are generally for social purposes, for example, for competing with friends. Sharing activity and health data representations could have broader benefits, such as for medical purposes (if shared with a GP or other healthcare providers), or research related to public health, or friends and family (who are carers and/or live far away). The reviewed platforms do not provide abstract visual encoding of the data. The supported visualizations, such as charts with textual/numerical details, are clear and could reveal high levels of information about oneself, especially when sharing a significant amount of data (e.g., publicly posting daily physical activity details).

4.6 Suggestions and Recommendations

This review presented an up-to-date review (according to the details presented in section 4.3.1) of the visualizations of health and physical activity data tracked by four of the most recent and popular tracking devices on the market. The review suggests that future personal health systems need to address the following issues:

- Personal health data are multivariate time-series in nature as the sensors collect different variables over time. Presenting each variable data individually does not help people in understanding health context (i.e., the status) and the factors that could affect their health. For example, it could be helpful for some users to see how their sleep (total hours and quality) affects their daily physical activity or vice versa. Multivariate visualizations could facilitate finding correlations between the tracked data, testing hypotheses

or investigating cause and effect. Therefore, there is a need for addressing the concept of combining multivariate personal and health data and identifying design guidelines for the data.

- Discussed visualization methods were essentially designed for data tracked by one device. There are many users who have access to multiple self-tracking devices and, as a result, they face significant limitations in visualizing all tracked data on one platform (app/dashboard). Ideally, all devices and apps should integrate with each other and across multiple platforms. Moreover, users could be allowed to select what data to integrate and for how long. This would provide users with better engagement and management of their health data.
- The review showed that the basic visualization methods used are statistical charts. Although these methods provide good interpretation of the data, combining them with more advanced methods could provide users with a better understanding.
- There is still a need for adding analytical features and interactive visualization. For example, the user could filter data being visualized to view only the days where they met their daily goal, to view the level of activity at a specific time on a specific day, or to view the HR details for a specific exercise.
- Visualization methods that offer high personal data encodings are required for social sharing purposes. The devices and apps should offer some abstract representations as an alternative when details of health data are not required.

This review aimed to help in identifying the limitations in available visualizations and the research gaps in the literature and guided the direction of this thesis. It also aimed to direct future research and to inspire some novel ideas for visualization and data representation of health and physical activity within the research of designers who work with emerging and interactive technologies.

4.7 Insights for the Study's Design

The literature review in Chapter 2, the survey presented in Chapter 3 and the review of the available visualizations discussed in this chapter inspired the study's design presented in Chapter 5 by highlighting the important data to consider and the visualization methods. The earlier section presented suggestions and recommendations derived from the studies in this chapter. The thesis focuses on the first point of these suggestions and recommendations (in section 4.6) that suggested addressing limitations of the design for the multivariate health data., as outlined in the following sections. The review in this chapter covered the visualizations on different platforms, and there was no significant difference between the visualization methods used on the apps and dashboards. However, in the design of the visualization, we focus on the visualizations that are displayed on the dashboards.

4.7.1 Data Type and Length

The responses to the questions in study presented in Chapter 3 showed the importance of involving different types of data and visualizations, as the responses clearly indicated that users of the devices have different preferences. They also have different tracking purposes and behaviour [134, 135]. Therefore, the visualization designs should mainly look for a proper combination of multivariate personal health data variables, in addition to considering different designs.

Data Types: the data that should be considered in this research are: the step count, distance, sleep and heart rates data as their figures in tracking were relatively high (ranging between 48% to 72%) and as between 42% - 46% of participants were highly interested in the visualization of these data. The burned calories data is also important as it was tracked by 40% of the participants and had a high and moderate interest in its visualization: 33% and 31% of the participants, respectively. The flight climbed data has been excluded as only 12% (10 out of 84) indicated they tracked this data and the high interest of its visualization was only 12% as presented in Figure 3.1b. The main issue with the sleep data that it is not easily accessed. Furthermore, as the distance and steps could be relevant and the aim of the developed visualization is to facilitate finding relationships between the variables, only one of these two variables have been selected.

Data Length: according to the responses to the survey shown in Chapter 3 (see section 3.4.1), most of the participants who track the selected data types reported they check the charts of the data daily. 48% of the respondents check the step count daily. For HR, around 35% of the respondents check the HR daily, while 48% of them never check it. For calories data, 31% of the respondents check the calories burned daily, 20% weekly, 11% monthly and 38% never check calories burned.

4.7.2 Selected Visual Choices and Visual Tasks

Chen et al. [43] gathered a collection of visualization definitions from the literature that define what visualization is and what it is used for. Chen et al. defined the visualization as the study concerning transforming data into visual representation to facilitate cognitive processes when conducting tasks involving data, where the measure for effectiveness is correctness and for efficiency it is the time required for accomplishing a task; where effectiveness and efficiency of the visualization are measured by the correctness and the time needed for accomplishing a task [43].

Deciding what tasks people perform when using these visualizations is difficult, as they could vary according to the tracking purposes (e.g., reaching goals) and what knowledge they want to retrieve from the chart. For example, some users might be interested in meeting their daily step count goal, others might be interested to see how many steps they have taken during a specific time, others might only check the graph for exploration and curiosity.

Although the literature search uncovered no specific research that studied the types of the visual tasks that should be supported by these visualizations, it did find studies in personal informatics that investigated the types of questions people have about their data in order to use the visualizations [135], also, what insight self-trackers gained from their collected data and what methods they used to convey these insights [45]. Although one of the visualization's benefits is to gain insight, Chen et al. [43] stated that insight in visualization is what the user or the creator of the visualization intended to achieve and, as such, it is hard to measure and evaluate.

Choe et al. [45] presented eight different data-driven insight types for personal data and their sub-types. These insight types are reported by quantified-selfers during their presentation, they

included: identifying extremes; identifying value; different types of comparisons and correlations between data. Although the 'insight' is not the focus of this thesis, the study inspired the choice of tasks used in our studies presented in Chapters 5 and 6. It helped in identifying which information is more relevant to the user, which therefore informed the design of the visualizations in this thesis. The selection of the tasks appropriate to this thesis is discussed in the next chapter.

4.7.3 Visualization Layout

This thesis focuses on the visualizations of personal health data, which are often used by people who use the charts daily or weekly to check their health and activity progress. Therefore, it was necessary to consider the visualization methods commonly used by the reviewed apps and dashboards. The derived taxonomy (in section 4.4.1) that classifies the visualization methods on the reviewed platforms in this chapter included the Radial layout, which could be combined with a clock metaphor to support better context. Although the Radial chart is not a traditional layout, it is not new to users as it has been applied in different domains. The visualizations review in section 4.4.1 showed some types of Radial representation (e.g., Rings on Activity and Fitbit apps). A quick survey of the Tableau Public website showed a variety of uses for the Radial charts, especially when related to time-series data [129, 170, 173].

Offering different visualizations could support different perspectives and allow a better understanding of the data; furthermore, a relevant context in personal visualization could also help people to make sense of the tracked data [114]. Huang et al. [115] showed positive results of using a calendar layout to understand the Fitbit physical activity data. Hence, the data is time-dependent and mainly reviewed and checked on a daily or hourly basis, the use of a clock metaphor as a layout could offer an additional layer of information to facilitate the interpretation of daily data.

Section 2.3.3 in the previous chapter discussed visualization methods for time dependent data with a specific focus on the radial design. According to the advantages of the radial design outlined in section 2.3.3.1.2, it is effective for data that naturally exhibits a periodic pattern and also supports the understanding of data with a circular or seasonal behaviour. On the other

hand, it has higher perceptual challenges than the Linear layout and has a higher cognitive effort [36]. With respect to specific visual tasks, section 2.3.3.1.2 also highlighted the need for evaluating the effectiveness of using the radial clock design for representing daily periodic data in general and in the personal health context specifically.

The effectiveness of this layout should be evaluated regarding its efficiency for retrieving data from the charts, and from the participants' perspective and their feedback. Therefore, a comparison between the traditional Linear layout and the Radial layout could justify choices for visualization designs of this and similar data types.

4.7.4 Visual Encoding

Although different types of visualizations were identified by the review (as presented in section 4.4.1), there are some sort of similarity regarding which visualization type is the most common design. As shown in Table 4.3, the most frequent visualization types are the Bar chart (BC), Circular Ring (CR) and Line chart (LC) respectively, along with the use of text and number (T&N) to provide a summary of the information.

From the users' point of view, as shown in section 3.5, more people preferred the Line chart to represent their health data as it was selected by 57%, followed by the Bar chart, which was selected by 52%. This supports Choe et al.'s [45] study which also identified a preference for the use of the Bar and Line chart for visualizing personal data, as set out earlier in this work.

With respect to the use of these charts in the literature, it is quite dependent on the type of data and the information that is needed to be obtained from the data. Previous research evaluated the Bar and Line charts across different tasks and datasets. For example, the Line chart has been identified in the literature as being effective for finding trends tasks [142] and characterising distributions [186]. The Line chart was found to be easier for perceiving trends, while the Bar chart was found to be easier for comparing two discrete data values [247]. According to Saket et al.'s study [186] discussed in the earlier chapter, people performed faster and more accurately with the Bar chart than the Line chart. Further, experiments on graphical perception showed that people performed more accurately when perceiving information that encoded using length than with other encoding elements such as angle [48, 217].

The Bar charts were found to be effective for comparing values because the human visual system is good at comparing Bar lengths and are recommended for finding extremes tasks [142]. Saket et al. [186] also found that the Bar chart is significantly faster than other evaluated visualizations for similar tasks, including the Line chart. Therefore, these two types of visual encoding (Line and Bar) are used for representing the data due to their familiarity, suitability for the data and potential visual tasks, as presented in next chapter.

4.7.5 Visualization Methods

One of the visualization design's aims in this thesis is to find a suitable method to combine more than one data variable into one visualization space to reflect on the relationships between collected metrics. The significance of such visualization design was discussed in section 1.2 and section 2.2.2, whereas the literature review (see section 2.3.2.1) discussed how the data could be visualized for this specific problem. Based on the reviewed research, it can be hypothesised that using an overlapping design to visualize the data might be effective for comparing the data variables as they are plotted on the same place. As only three variables were considered in this study, it is expected that overlapping that data could be more effective (faster and more accurate) than when visualizing each variable separately. This aspect led to develop the hypothesis **H4b** presented later in Chapter 6. Javed et al. [121] showed that superimposed design outperforms the juxtaposed design at comparisons tasks as it benefits from the shared space, while in the Juxtaposed design, performing comparisons requires eye movements between the data charts. Comparing these two methods for conducting comparisons and for other different tasks is important in order to understand their efficiency and whether they increase or decrease the accuracy and time.

4.8 Updates on the Visualizations on the Apps and Dashboards

In this chapter, a review study on the methods of visualizations implemented on health tracking devices and apps were presented. The study was performed at an early stage of this PhD project and led to the design of the evaluation studies in Chapters 5, 6 and 7. This section presents updates on what was changed on the reviewed devices platforms, in terms of the visualization methods used and the data. The section is organised based on the surveyed tracking devices and their companion apps and dashboards.

4.8.1 Apple Watch and its Companion Apps

The two companion apps with the Apple watch are the Activity app (which is associated with the watch) and the Health app, which is running on all iOS phones.

Apple Activity App (iOS): Apple produces a new method in their latest update of Watch OS and iOS supporting trends that are provided after 180 days of tracking of different types of activity measures including: movement; exercise; standing hours and minutes; distance and pace. The trend calculations are presented as a text summary for each category, but a Bar chart for the past year distributed over the months appears upon user interaction with each category (see Figure 4.9). In addition to the observation and the use of the Watch and the app, online resources regarding the new changes in the Activity app were reviewed, as in [15, 166].

The Activity app in the previous version was showing the historical data using a calendar layout, where the Circular Ring visualization is used to represent each day's data. The trends feature in the app provides a better representation of data over time with more detail regarding changes in activity level, peaks and trough patterns that were not supported in the calendar view. Although steps and distance visualizations are represented over a long period of time in the trends tab, there are no detailed daily charts for them in the Activity app.

Apple Watch Display: The change in the Watch app was mainly in the visualization of the HR data. The Heart app presents different charts of the HR data, including resting HR levels and

walking average where the data is plotted as ranges Bars.

Apple Health App (iOS): In addition to the standard Bar chart used in the previous version of the app, it uses a Line chart for representing the daily step count that overlaps the current steps for the day with the user average. The same is also applied to distance data (see Figure 4.10).

4.8.2 Fitbit Watch and its Companion App and Dashboard

The Fitbit app and dashboard still support the same type of representations. The main update identified was in sleep tracking and its visualization. The sleep visualization includes coloured Line chart to represent different sleep stages (Figures and details can be seen in [73, 127]). No changes were identified in the other data including the HR as presented in [165] and other recent posts on [75], except the new visualization that combines calories in and out over the past two weeks (i.e., multiple data visualization). However, consumed calories (i.e., food intake) are not included in this review as they were manually entered by the participants.

Other features have been added to the sleep, such as a sleep score, but this requires a premium subscription (as their website informs [70]), whereas this review discusses the basic free features only. In addition, new visualizations were added to represent sleep data. Fitbit implemented a clock style display to view the sleeping hours during a 12 hours night (see Fitbit Website [72]). Moreover, a calendar layout was used to represent the hourly activity goal. The layout presents the hours the user is active/inactive for each of day of a given week (see [71]). The app also offers a cardio fitness score representation as a coloured stacked bar that shows how fit the users are with respect to their age and gender.

4.8.3 The Garmin Watch and its Companion App and Dashboard

In addition to the Garmin official resources such as [87, 88, 90], [105] was used to retrieve additional information. The visualization of sleep data in the recent version of the app includes a coloured Donut chart to represent minutes within each sleep stage. The app/dashboard also includes a feature for comparing calories in and out, as in the Fitbit app/dashboard; however, it requires connecting with MyFitnessPAL app [157] to import recorded calorie intake.

Garmin supports combining different data on one chart, especially for training data. In addition to the HR visualization discussed earlier (presented in Figure 4.6) and the representation of HR and pace shown in Figure 4.7, other combinations of data are also supported, such as: the pace; HR and running power (see Figure in [86]); the blood oxygen saturation levels during sleep that shows the oxygen levels within sleep stages (see Figures in [169]) and the training load with VO2 Max [85].

The calendar view offers additional information in its latest version. The calendar shows the types of the activities that have been performed in each day as coloured Bars that refer to different types of activity (multiple data visualization), more details are presented when clicking on each day (similar to the HR view shown in Figure 4.6) with additional textual information at the bottom. Garmin provides ‘insight’ feature that allows users to compare some types of the current activity (e.g., steps) to other Garmin users (all users, male or female users) of similar age (see [105]). The additional information about the training visualizations was derived from different videos that explain training data, status effect and their charts Garmin YouTube Channel [87].

4.8.4 Samsung Watch and its Companion Apps

Recently, the Samsung Health app added multiple data visualizations for representing HR data during walking and running along with a speed (see link in [190] for these visualizations). The new version of the Samsung app does not offer as many new changes in the visualizations as the interface does. The interface offers additional features such as adding friends, customising running exercises and a pacer and the discover tab that gathers different functions such as programs and accessories. Information was retrieved by using the app [191] and reviewing the videos on the Samsung Health app found on the Samsung official YouTube Channel [192].

4.9 Conclusion

This chapter presented a review and the classifications of the visualizations methods provided in most popular health tracking devices and their companion platforms to answer the second

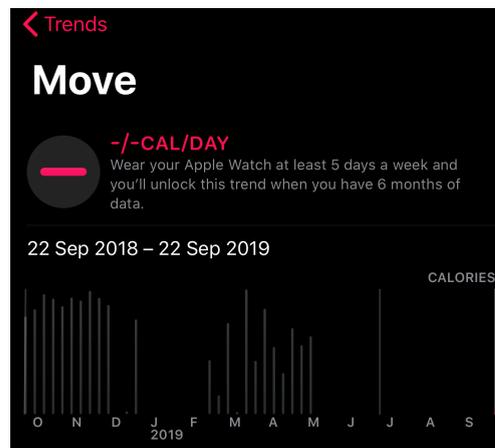


Figure 4.9: A Screenshot captured from iOS Activity app. A bar chart represents the move (i.e., total active calories) for the last year.

research questions (**RQ2**). The outcomes of this study along with the literature review and the user study in the previous chapters, contributed in identifying the main gaps between people needs and purposes and health tracking and what most popular health apps and dashboards support. The results from this chapter helped in defining several limitations that need to be addressed, however, the scope of the thesis is focused on the visualization designs for multivariate health data. Moreover, it helped in shaping the design of the visualizations and the data used for the study that are presented and evaluated in the following chapters.



(a) Daily Step Count



(b) Daily Distance

Figure 4.10: The Visualization of Daily Walking/Running Distance and Step Count on the Apple iOS Health App Overlaps with their Averages.

Visualization Design and Preliminary Evaluation

5.1 Introduction

In Chapter 3 and Chapter 4, after analysing and discussing the results from the user study and the visualizations review, several limitations were identified and a number of recommendations were suggested to address these limitations. This led to insights into the data types, size and visualization methods that could be applied to overcome these limitations, these are presented in section 4.7. This chapter aims to address the challenges raised in the previous chapter by developing and evaluating the various visualization alternatives for multivariate personal health data. In this chapter, the design and the implementation of a preliminary evaluation study are described.

5.2 Aims and Research Questions

The definition of the research problem was explained in section 1.2, the literature review regarding personal informatics and people requirements in Chapter 2 and the visualizations review in Chapter 4 discussed the limitations in the available visualizations and their limitations in addressing people's needs. In this chapter, the aim is to develop different visualization designs in order to understand the impact of the used visual elements on data perception and data people's preferences.

Different studies on personal informatics have raised the importance of finding relationships between collected personal data. For example, Rooksby et al. [185] interview-based study (see Chapter 2 for further details) which stated that users relied on diagnostic tracking as a form of tracking that looks for a relationship between one data variable and another. For example, one participant in the study used the tracking data to try to find the reason behind their stomach pain (foods and medication tracking), and another used the data to help him sleep better.

Similarly, Li et al. [135], reported several types of questions people ask of their data including finding relationships between their personal data (both short and long-term data) and finding patterns and trends. For example, using the personal tracked data to work out the relationship between physical activity, sleep and the blood sugar level. Although the necessity of combining variables has been highlighted in the literature (e.g., [134, 135]). Self-tracking device

companies (as discussed in Chapter 4) provide some visualizations on the apps/dashboards that represent more than one variable, especially for exercise and training data, where HR is an essential element (e.g., training and exercise visualizations by Garmin in Figure 4.7 and Samsung in [190]). Improving and evaluating the combination methods for this kind of data is important.

This stage of the research aimed to validate the selected approach and the implementation of the lab experiments for evaluating the proposed visualization alternatives to represent the chosen personal health data. It aimed to test whether the outcomes would answer the research questions (presented in Chapter 1), and to promptly detect and address any issues before conducting the main evaluation study.

Regarding the visualization designs, this chapter considers the design aspects that could contribute to answering the research questions that are related to the design of the visualizations (which are the third and the fourth questions presented in Chapter 1).

The validation of the evaluation sought answers to some questions that would help in the design of the main evaluation study of the visualizations. These questions are regarding the design and the implementation of the evaluation study, which are addressed in the results section of this chapter (section 5.4). These questions are:

1. Do the participants understand the instructions/tasks and the visualizations correctly?
2. Does the order of the visualizations have an effect on the overall performance of the participants (time and accuracy)?
3. Does the order in which charts are presented have an effect on the interpretation of the Radial layout? (e.g., when the first stimuli are Radial layout).

This chapter presents the implementation of the evaluation study. It discusses the experiment design (including the visual tasks, questionnaires, participants and the procedure), the utilised dataset, the technical details of implementing the visualization techniques and concludes with a discussion of the results gained from the study.

5.3 Methodology

The study presented in this chapter included designing visualizations for representing multivariate personal health data and evaluated the proposed designs. The evaluation is based on previous visualization evaluation studies, which examine the effectiveness of visualization designs alternatives by applying a visual task-based evaluation study similar to the methodology used in previous work [28, 107, 211]. The experiment compares the visualization choices by comparing users' performance when solving a set of tasks on each visualization, using the metrics of accuracy (i.e. error rate) and task completion time. This section describes the health dataset used in the visualization, the visualization designs, the visualization tasks and the experiment design.

5.3.1 Data Source and Characteristics

Data Source: To generate the visualizations for the evaluation study, a real dataset generated by Apple Watch Series 2 was used. The dataset was exported to a CSV file that stores the data on an hourly basis. The QS Access app [175] was used to export the dataset to the required format. The implementation of the visualizations focused on three main variables of the physical activity data: the step count; active calories and heart rate (HR). Also, data covering two days was selected to use in the visualization; the selection of the data variables is discussed in section 4.7.

Data Characteristics: The selected variables are different in their types and possible values. For example, HR data is time-series data and most commonly visualized using a Line chart. However, step count data is discrete and on the majority of physical activity tracking apps are widely visualized with a Bar chart (more details about these data variables and definitions are presented in section 2.2).

When considering the distribution of possible values for each of the three variables, there is a significant difference. According to the NHS [161], the resting heart rate for most adults is between (60 - 100 bpm), and for athletes it could be (40 - 60 bpm). Hence, zero values for heart rate data exported from the wearable device are considered to be missing values (i.e. gaps

or no data). On the other hand, the step count could have a zero value at certain times and are not necessarily treated as no data recorded (i.e. the person may still have the tracker on but is inactive, and no steps were recorded). The maximum possible value for the step count could reach thousands, while for calories, it could be in the hundreds. In the utilised dataset, the highest step count over an hour was relatively high (3370 in an hour). Although these three variables vary according to the physical activity level of the person, the total step count per day for a moderately active person is between 7500 and 9999 steps, according to Tudor-Locke and Bassett's classification of step count [231]. Some tracking devices and apps distinguish between the calories users typically burn just by being alive and the calories they burn while being active (i.e. referred to as active calories and used in this study), this, in turn, reduces the range of the active calories data. Since the data in this thesis has been visualized on an hourly basis, the value of the heart rate data is calculated for each hour as the average of the heart rate values recorded during that hour. Figure 5.1 shows a sample of the used dataset in the visualization.

Start	Finish	Active Calories (kcal)	Distance (mi)	Flights Climbed (count)	Heart Rate (count/min)	Resting Calories (kcal)	Steps (count)
01/07/2017 9:00	01/07/2017 10:00	0	0	0	0	53.594	0
01/07/2017 10:00	01/07/2017 11:00	2.56539556	0.02093025	0	0	53.63506733	44.1907994
01/07/2017 11:00	01/07/2017 12:00	35.5375816	0.660429836	0	133	59.7749532	1275.96389
01/07/2017 12:00	01/07/2017 13:00	66.30133002	0.986104428	2	151	68.92126173	2457.023057
01/07/2017 13:00	01/07/2017 14:00	63.58144988	1.389838813	1	130	68.94182531	3370.822254
01/07/2017 14:00	01/07/2017 15:00	28.37128483	0.490167748	1	102	64.44135957	1204.070362
01/07/2017 15:00	01/07/2017 16:00	18.20515394	0.210470981	2	94	62.89582082	495.6501466
01/07/2017 16:00	01/07/2017 17:00	11.1405768	0.051337573	0	120	62.29406195	117.279491
01/07/2017 17:00	01/07/2017 18:00	73.79066963	0.283744842	0	180	60.66729167	501
01/07/2017 18:00	01/07/2017 19:00	8.850197631	0.061157994	0	115	58.52645508	128
01/07/2017 19:00	01/07/2017 20:00	8.189360118	0.033767255	0	90	58.55777747	69
01/07/2017 20:00	01/07/2017 21:00	10.44529646	0.126168967	0	112	61.12235227	282
01/07/2017 21:00	01/07/2017 22:00	6.542703545	0.075123624	0	90	58.53064195	166
01/07/2017 22:00	01/07/2017 23:00	6.83	0.032426631	0	90	56.43635401	71
01/07/2017 23:00	02/07/2017 0:00	8.085325801	0.034383063	0	88	56.53755675	78

Figure 5.1: A Sample of the Dataset Used in Developing the Visualizations in the preliminary Evaluation. The Start and Finish columns show the dates and start and the end of the hour, other columns such as the Steps and the Heart Rate show the recorded data in that hour.

5.3.2 Visualization Design and Development

Further to the discussion of the visualization insight in section 4.7, the visualization design relies on the possible options for representing the three selected variables (Steps, Calories and HR). The insight also discussed the data types and length that needs to be considered according to the outcomes of the study in Chapter 3.

The following sections present what are the data visualization options that were considered in the designs and how they were implemented.

5.3.2.1 Data Visualization Options

The study aims to compare users' performance with the graphs in the three cases: the visual layout, the visual encoding and the methods used to integrate the variables.

Visual Layout: The designing of the visualizations began with the layout, with the aim of identifying an appropriate layout to represent the data in a relevant context. The proposed Radial design was inspired by a clock metaphor, which was suggested to provide an additional layer of information by depicting the hourly health data in the natural order of the day's hours. Health data is usually presented on dedicated health apps and dashboards as activity levels and progress made during the day. Therefore, a metaphor that specifically relates to time was chosen. The perimeter of the circle and its numbers represent the 24 hours of the day. The four coloured sectors refer to different parts of the day (morning, afternoon, evening, and night), which allows the user to perceive each part as a whole and so facilitate the estimation of the total data (e.g., steps or calories) within each part according to the visual grouping principles. The symbols (glyphs) are used as extra information to refer to the time. Glyph-based designs can provide efficient information encoding and can attract more attention [29]. The Radial layout is presented in Figure 5.2. The data was mapped from the centre toward the corresponding hours on the perimeter of the circle. The Radial visualizations are shown in Figures 5.3d, 5.3h, 5.3b and 5.3f, while the visualizations with the traditional Linear layout are shown in Figures 5.3c, 5.3g, 5.3a and 5.3e. The abbreviations used for the visualizations are presented in Figure 5.3. In both layouts, the visual encoding method (i.e. Bar or Line) was used to encode the quantitative attributes with the length channel. The implementation of both layouts used the same scaling method; the difference is in the orientation of the axes.

Visual Encoding: Several encoding methods could have been used to visualize the data on Cartesian and Polar coordinates. For example, the Cartesian Bar chart, coloured Stacked Bar chart, multivariate Parallel Coordinates; and their Radial equivalents Rose chart, Pie chart and Star plot [36]. In this experiment, Bar and Line chart encoding were selected to be used for

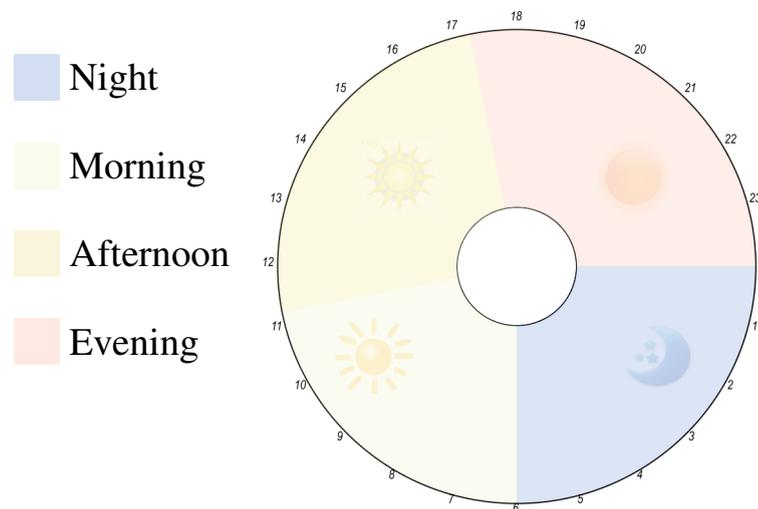


Figure 5.2: The Background of the Proposed Radial layout in the Preliminary Evaluation Showing the Four Coloured Sectors and the Glyphs.

visualizing the data. According to the design insight (section 4.7) derived from the study in Chapter 4, Bar and Line charts were the basic methods used in the reviewed dashboards as presented in Table 4.3. The Bar chart is shown in Figures 5.3c, 5.3g, 5.3d and 5.3h), while the Line chart is shown in Figures (5.3a, 5.3e, 5.3b and 5.3f). For the colour selection, ColorBrewer was referred to for choosing [33, 104] colours from their colour palettes. The colour selection was focused on selecting three categorical blind-friendly colours to represent the three variables.

Visualization Method: The visualizations of multivariate-time dependent data and the comparative designs for multiple data were discussed in sections 2.3.2 and 2.3.3 in Chapter 2. Gleicher [94] proposed the Juxtaposition design for visualizing multiple variables, which presents different data variables separately and the Superposition design, which visualizes the objects overlaying in the same space. In this thesis, the concept of juxtaposition is applied in the Stacked design and Superposition is the Overlapped design; the aim was to identify the most suitable method for finding relations between personal health data. When representing the selected health data variables (steps, calories, and HR), the visualizations for when the data is Stacked in separate charts (shown in Figures 5.3c, 5.3a, 5.3d and 5.3b), versus representing the data variables overlapping on the same chart area (shown in Figures 5.3g, 5.3e, 5.3h and 5.3f) were compared. The Linear Stacked and Overlapped Line design were also presented in Figure

2 of Zhao et al. [248]'s study as design space examples for multivariate time-series layouts.

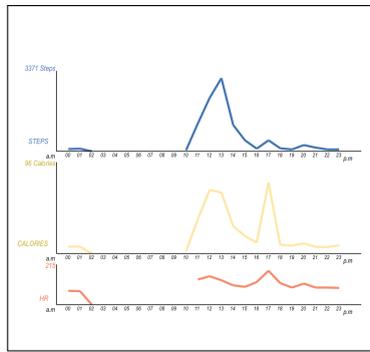
These cases generate eight different designs of visualizations, which are presented in Figure 5.3. The Figure also shows the abbreviations that are used to refer to the different visualizations in the following sections of the thesis.

As discussed in the section on data characteristics, the selected variables differ in their value ranges; this makes mapping the three variables on the same chart area challenging, especially for the overlapped chart type. For the *Linear* layout, each variable was mapped on a linear scale based on its minimum and maximum value, and the interaction feature was depended upon to inform the user of the values when interacting with the visualizations. The X-axis represents the time, and the Y-axis represents the value of the mapped variables. For the stacked visualizations (which are LSLC and LSBC) each variable was mapped based on its minimum value (i.e., zero) and the maximum value in the current dataset using the built-in `map()` method to map the data to the specified chart area for each data type, which has its axes. For the overlapped visualizations (LOLC and LOBC), all the three variables were mapped to the same chart area where each variable (data type) was mapped within the specified space based on its maximum value. The design of the *Radial* layout was based on a clock metaphor, the perimeter of the circle represents the time where the numbers are the 24 hours of the day. The four sectors in the circle represent the parts of the day (morning, afternoon, evening and night). The data were mapped from the centre towards the corresponding hours on the perimeter of the circle. The mapping process is similar to the mapping in the Linear Layout, but with the use of the angles to determine the position of the points. The eight visualizations are presented in Figure 5.3.

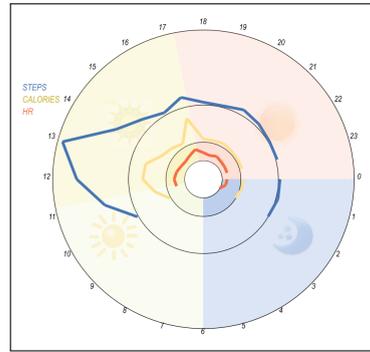
The visualization choices considered in the final designs does not mean that other visualization options are not suitable for the design purpose. The selected layouts, visual encodings and methods were suggested based on the literature review and the outcomes of the survey and review in Chapter 3 and 4. This thesis evaluated these visualization choices to demonstrate their effectiveness.

5.3.2.2 Visualization Development

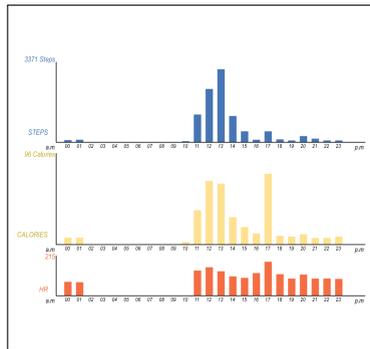
To develop the visualization, Processing [78] was used: a free, open source software. The Processing project was specifically designed to simplify the process of creating graphical applications [163]. The core language and the various additional libraries were based on Java programming language. Hence, this makes Processing easy to understand and familiar to many people with a programming background [178]. The ControlP5 [201] library was used to build the experiment's interface on top of the sketch.



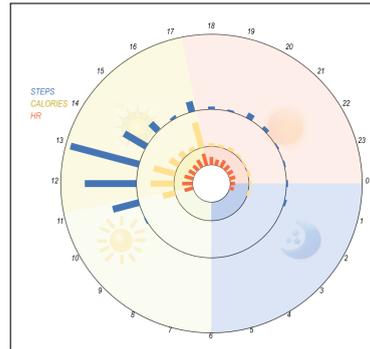
(a) Linear Stacked Line Chart (LSLC)



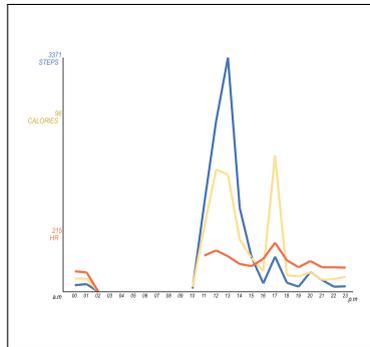
(b) Radial Stacked Line Chart (RSLC)



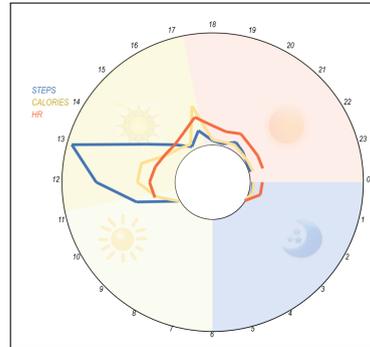
(c) Linear Stacked Bar Chart (LSBC)



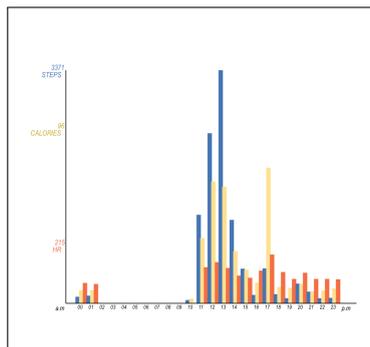
(d) Radial Stacked Bar Chart (RSBC)



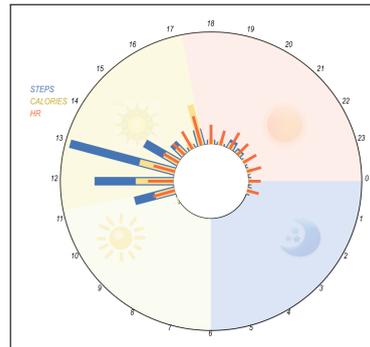
(e) Linear Overlapped Line Chart (LOLC)



(f) Radial Overlapped Line Chart (ROLC)



(g) Linear Overlapped Bar Chart (LOBC)



(h) Radial Overlapped Bar Chart (ROBC)

Figure 5.3: The Eight Visualization Designs Developed for the Preliminary Evaluation..

5.3.2.3 The Colour Selection

When designing visualizations with colours, designers should take into consideration the common problem of colour sensitivity, especially sensitivity to the red-green spectrum as it affects 8% of males and 0.5% of females [156]. Color-Brewer [33, 104] was referred to for colour selection, and [140] to choose from their colour palettes. The colour selection process focused on selecting three categorical colour blind-friendly colours to represent the three variables. The suitability of the selected colour to the variable and what colour was usually used on apps and dashboards to represent it was also taken into consideration. The final selection was blue for the step count, yellow for the calories and orange for the HR, as presented in the visualization in Figure 5.5.

5.3.2.4 Interaction Features

To interact with the visualization, the user is allowed to hover the mouse over specific areas on the graphs (e.g., moving the mouse over the bars or the lines) to view their corresponding values. The aim of supporting this interaction feature is to help in identifying and estimating the values of the visualized variables (a screenshot is presented in Figure 5.4).

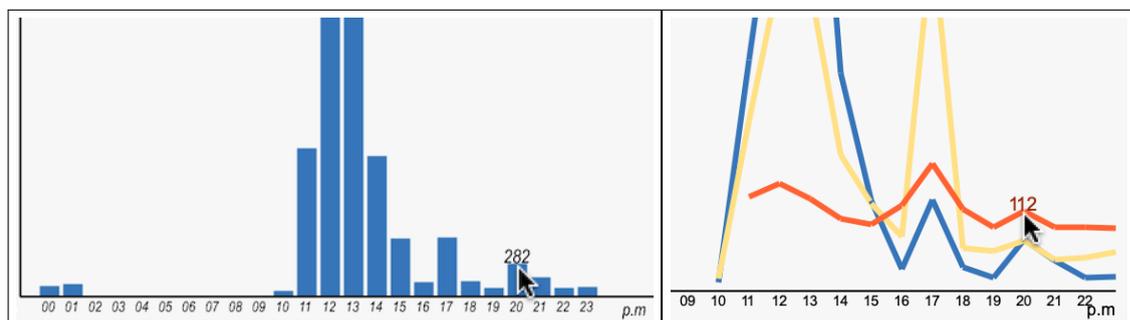


Figure 5.4: Screenshots of the Tooltip Presented on the Visualization. In the Overlapped design, the number presented by the tooltip is colour coded according to the colour of the associated data variable.

5.3.3 Visualization Tasks

Task selection is challenging, it should be selected based on evaluation goals, and it should be simple and easy to implement in the experiment [199]. The tasks in the experiment were derived from [234], which was discussed earlier in this work in the background and literature review chapter (see section 2.3.2). This taxonomy was selected for the tasks in this experiment because it is domain-independent, it covers the basic tasks the user may need to perform when using visualizations for exploratory analysis or to support more advanced statistical analysis and the tasks are supported by many other generic task taxonomies (such as in: [31, 241]).

The tasks in Valiati et al.'s [234] taxonomy could be easily applied to the personal health dataset in this study and could be related to the information people seek when looking at their data. Although this thesis has not directly recorded potential users' tasks (e.g., by interviewing real users of personal health trackers to understand and investigate the type of the tasks they perform on the visualizations), the design of the tasks and questions in the experiment were directed by previous refined visual task taxonomies, as discussed earlier, and previous research that studied the requirements and needs of the users of personal tracking devices and tools (such as [44, 45, 134, 135]). One of the major purposes of tracking, or the question people have about their data, is the relationship between different variables to identify factors for either long-term or short-term tracking, or for specifying the direct relationship between two variables (as shown in [44, 45, 135]), which is similar to the target in the 'Infer' task. According to Choe et al. [45], people may seek different information from the visualization of their data. An example is identifying the measured value (similar to the Determine value). Another example is comparing measured values by time segmentation, which is similar to the Compare task that requires comparing data over two days. In addition to the importance of these selected tasks for potential users, the tasks of inferring relationships and identifying daily patterns in the way presented in this thesis were not well studied in the literature and are specifically not well studied on both layouts.

Five of the seven tasks from Valiati et al.'s taxonomy [234] were selected for this study, which are the five analytical tasks (Identify, Determine, Compare, Infer and Locate). The other two tasks (Visualize and Configure) were not applied as they require interaction and the study focuses only on the visualizations without considering the interaction features. The five tasks

formulated for the study are presented in Table 5.1, which also shows how the user provides the answers to the questions.

Table 5.1: Tasks/Questions Performed in the Preliminary Evaluation.

Task/Question	Answer (Response Type)
T1: (Identify): What part of the day the user is more active?	Multiple Choice. (Morning, Afternoon, Evening, Night)
T2: (Determine): Find the number of step count at 01:00 P.m. at the shown day visualization	Text field.
T3: (Compare): Compare the active calories for the two days presented, which day has higher active calories?	Multiple Choice. (First day, Second day, Both are the same)
T4: (Infer) What is the relation between HR and step count that you can identify in the presented visualization?	Multiple Choice. (There is no relation, Higher step count has peak HR, Higher step count has resting HR)
T5: (Locate) What hour(s) of the day the user has the minimum HR average?	Text field.

5.3.4 Experiment Variables

Various metrics need to be measured to justify design choices in personal visualization space, such as how the design fits in with a user's daily life. According to previous authors [223] and [114], depending on conventional metrics (i.e. task completion time and accuracy) for the evaluation of personal visualization is not enough.

In this research, the focus is on the effectiveness of different visualization layouts for physical activity data; therefore, the accuracy and task completion time for comparing the proposed visualizations has been measured. This restricts the evaluation to use a fixed dataset for the visualizations and evaluation was not conducted in a personal context, such as using participants' personal health data with the visualizations. However, to strengthen the evaluation, other metrics related to users' personal preferences in the evaluation of the visualization designs were investigated.

Performance Measures: The metrics of the task accuracy and completion time were used to evaluate the effectiveness of the developed visualizations.

Subjective Feedback: Participants were asked to complete questionnaires before and after performing the experiment on the visualizations in order to evaluate the design choices and to compare users' preferences in the visualizations with their performance. The pre-experiment questionnaire includes both demographic questions and Likert scale questions to collect participants' ratings regarding the developed visualization designs. The demographical section of the questionnaire collects some personal information about the participants, including age, gender, educational level, in addition to some other information regarding their tracking behaviour, for example whether they track any health or physical activity data, e.g., what data do they track? What tool do they use? The other section consists of 8 Likert scale questions that request rating each visualization in terms of the readability of the graph and in terms of the attractiveness of the design, together with one multiple-choice question that asks which visualization (s) from the eight presented they preferred to use to visualize their health data. The post-experiment questionnaire is similar to the second part of the pre-experiment questionnaire, it aims to capture users' ratings regarding the visualizations after interacting with it. The complete questionnaires are attached in Appendices .1 and .2.

5.3.5 Participants and Recruitment

Sixteen participants from Cardiff University were recruited to take part in the experiment. The participants were recruited among students and staff in the school of computer science and informatics: eight identified themselves as females; seven as males and one did not specify a gender. Permission to conduct the study was given by the ethics committee of Cardiff University, School of Computer Sciences and Informatics, the ethics application approval code is COMSC/Ethics/2018/001. The maximum time required for completing the study including the briefing and the debriefing sessions was around 50 minutes. For their time, the participants were rewarded with a £10 Amazon voucher upon the completion of the study. The participants in this experiment were instructed to complete the pre- and post-experiment questionnaires

Table 5.2: Data Tracked by the Participants in the Preliminary Evaluation Study and the Devices/Apps they Use.

Data Type	No of Participants	Tracking Device/App	No of Participants
Step Count	9	iPhone Apps	7
Walking/Running Distance	8	Android Apps	3
Active/Burned Calories	7	Fitbit Tracker	2
Heart Rate	4	Android Watch	1
Sleep	4	Garmin Tracker	1
Flight Climbed	3	Apple Watch	1

and were guided to explore the eight visualization methods and answer the five tasks/questions presented earlier.

5.3.5.1 Participants' Characteristics

The sixteen participants who were recruited to complete the study reported attaining the following level of education: 11 with a post-graduate degree; 4 with an undergraduate degree and 1 with a high school degree. 12 out of 16 participants tracked different types of physical activity data (e.g., step count, walking distance or heart rate) using phone apps or wearable tracking devices. Table 5.2 shows the data the participants in the study track and the devices/apps they used for tracking.

5.3.6 Procedure

The study consisted of three sessions. In the first session, the study, its purpose and the data collected were explained to the participants. After receiving their consent to participate in the study, they were asked to complete the pre-experiment questionnaires. This session took around 10 to 15 minutes. The main visualization experiment took place in the second session, and it required about 20 to 30 minutes to complete. First, the experiment was explained to them, i.e. how they should respond to the tasks, then they were shown some examples of the visualizations and the tasks (see Appendix .3 for examples). Following this, the experiment was carried out, it was composed of eight trials. Each trial presented one type of visualization and the five tasks (presented in Table 5.1). When the participants had answered one question

Table 5.3: The Order of the Visualizations Presented for Each of the Four Groups in the Study.

Group	Visualizations Order
Group 1	Linear layout first and Radial layout second: (LSLC, LOLC, LSBC, LOBC, RSLC, ROLC, RSBC, ROBC)
Group 2	Radial layout first and Linear layout second: (RSLC, ROLC, RSBC, ROBC, LSLC, LOLC, LSBC, LOBC)
Group 3	Bar Chart first and Line chart second for each layout: (LSBC, LOBC, LSLC, LOLC, RSBC, ROBC, RSLC, ROLC)
Group 4	Overlapped type chart first and followed by Stacked type chart for each chart and layout type: (LOBC, LSBC, LOLC, LSLC, ROBC, RSBC, ROLC, RSLC)

they were able to move to the next question by clicking on the "Next Task" button, a screenshot of the experiment's software is presented in Figure 5.5. In the last session, the post-experiment questionnaire was presented to the participants to rate the visualization designs. Finally, they were provided with a debriefing sheet that included information about the study and whom they could contact for further information, this session took around 5 minutes. The questionnaires provided the information required to understand the users' preferences in the visualizations and the ability to compare the preferences with their performance when solving the visual tasks.

The sixteen participants were randomly assigned to four groups; each group went through the same experiment (i.e. the same visualizations and tasks), but the visualizations were presented in a different order. The aim of using the groups was to see if the order of the presented visualizations would affect the results of the experiment (e.g., to see if the tasks on the last viewed visualization were easier to solve than the first viewed visualizations). The four groups and the order of the visualizations are shown in Table 5.3.

5.3.6.1 Experiment Setup

The experiment was conducted in a study room that was isolated from outside distractions. The experiment and the online surveys were run on a MacBook Pro laptop with a 13-inch display, 2.9 GHz Intel Core i7, 8GB memory and Intel HD Graphics 4000 with 1536 MB. The participants interacted with the software using the built-in touchpad and keyboard, no additional devices were required.

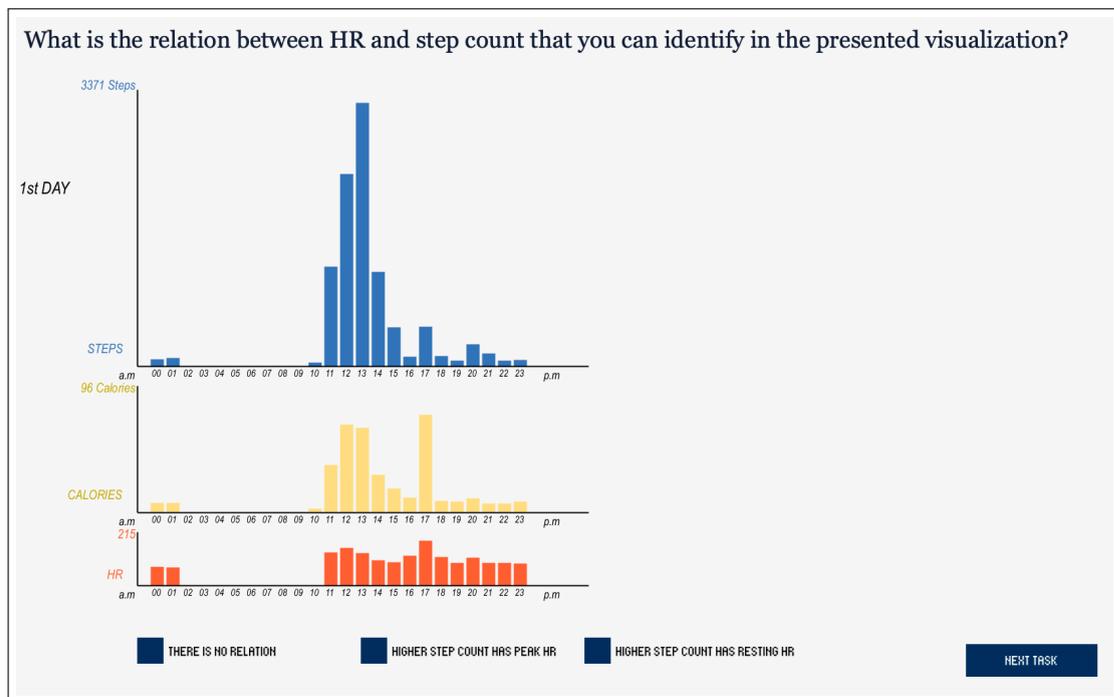


Figure 5.5: A Screenshot of the Preliminary Evaluation Experiment.

5.4 Analysis and Results

During the experiment, the task completion time and task answer accuracy were measured. The task completion time was calculated as the period from the time the user clicked on the button to show the task and the visualization to the time the user clicked on the next button to submit the answer. The accuracy relates to the answer entered or selected by the user for each task. The following sections present the results obtained from the study based on observation, the participants' performance with the visualizations, their feedback and the questionnaire data.

5.4.1 Task Completion Time

By observing the participants during the study it was noticed that some participants spent more time on the first viewed visualizations than the last viewed. Some participants reported that they were able to recall their answers to the tasks as they knew that the visualizations represented the same data: this might have affected their speed of response. However, to check whether the order affected the speed of response to the tasks on the last viewed visualizations, the task

completion time, which was measured from the experiment software, was referred to.

To analyse the results for the task completion time, the mean and the standard deviation (SD) was calculated for all four groups, beginning with the task completion time for all the tasks (i.e. the average time spent by each group of participants on each visualization to complete all five tasks) based on the order of the visualizations for each group. The graphs of the average time spent by each group's participants to complete all the tasks on each visualization are presented at the top in Figures 5.6, 5.7, 5.8, 5.9. Then, the mean and the SD for the completion time of each task on the eight visualizations were calculated, as presented in each task graph in the figures in each group category (see Figures 5.6, 5.7, 5.8 and 5.9).

The results from the experiment confirm the observation: the average time spent on the second viewed visualization was less than the average time spent on the first viewed visualization: by 46% in group 1; 40% in group 2; 35% in group 3 and 34% in group 4 (of the average time spent on the first viewed visualization).

When comparing the average time on the first viewed visualization with the average time on the other visualizations, the difference can be seen on each time graph presented in each group's figures. For example, in group 1 (see Figure 5.6), the participants spent around 250 seconds on the first visualization, while for the other visualizations they spent 66-135 seconds on average. This is also the case for the average completion time for each task for all four groups, as presented in the following figures.

On the task level, Table 5.4 shows the average time spent to complete each task for the participants in the four groups. The groups' average time for the 'Infer' and 'Locate' tasks is higher by 5-8 seconds than the average time for the 'Identify', 'Determine' and 'Compare' tasks, as their averages are 16-18 seconds.

With respect to the layouts, Figure 5.10b shows the average time spent by each group's participants to complete the tasks on the Radial and Linear layouts. The participants in groups 1, 3 and 4 performed faster by 47%, 44% and 39% respectively (of the average time on the Linear layout) with the Radial layout, while the participants in group 2 (who started with the Radial layout) performed faster with the Linear layout by only 18%. The average time spent on the Radial layout visualizations for all groups is around 78 seconds, while the average time

Table 5.4: Participants' Performance in Terms of Task Completion in Each Group (on all the Visualizations in General). This table shows the average time (in seconds) spent by the groups' participants to complete each task.

	Group 1	Group 2	Group 3	Group 4	Average
T1: <i>Identify</i>	18.83	10.98	15.33	22.94	17.02
T2: <i>Determine</i>	18.89	11.49	14.93	20.61	16.48
T3: <i>Compare</i>	17.85	13.62	19.42	22.41	18.32
T4: <i>Infer</i>	26.05	16.64	27.95	22.83	23.37
T5: <i>Locate</i>	22.26	13.31	34.49	28.40	24.61

Table 5.5: Task Completion Time on the Radial and the Linear Layouts for the Four Groups.

	Group1	Group2	Group3	Group4	Average
Radial	71.84	71.56	80.88	88.41	78.71
Linear	135.93	60.51	143.35	145.97	121.44

on the Linear layout visualizations is around 121 seconds, which shows that the participants performed faster by around 36% (of the average time on the Linear layout) with the Radial layout than the Linear layout. However, the Radial layout visualizations were presented first for group 2, while they were presented second for groups 1, 3 and 4.

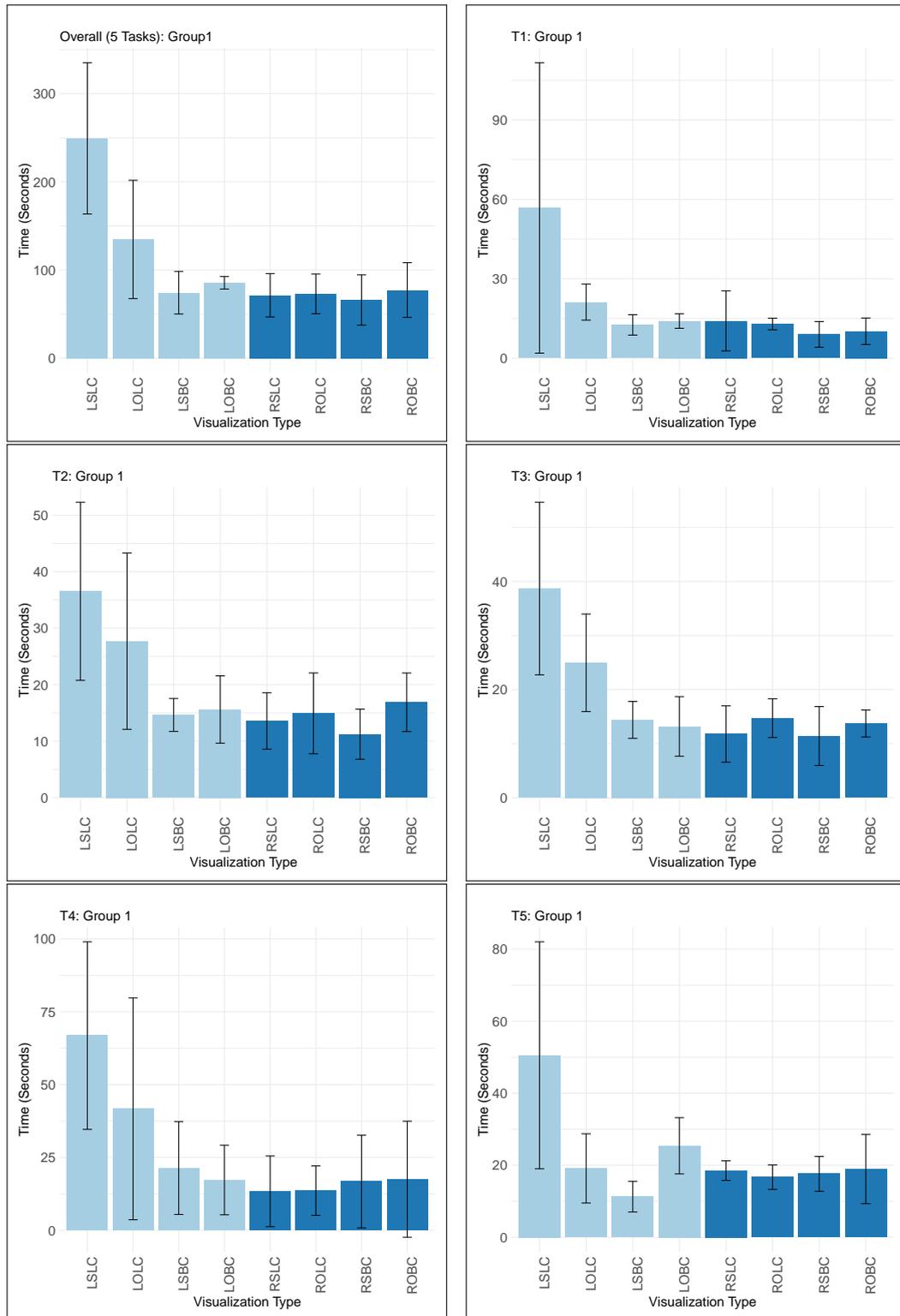


Figure 5.6: Task Completion Time for the Eight Visualizations for Group 1 (in the order presented). The top-left graph shows the mean and the SD of the overall task completion time and the others show the average time for each task individually. The Y-axis represents the average time, while the X-axis shows the visualizations based on their presentation order. (layout: Linear ■ | Radial ■).

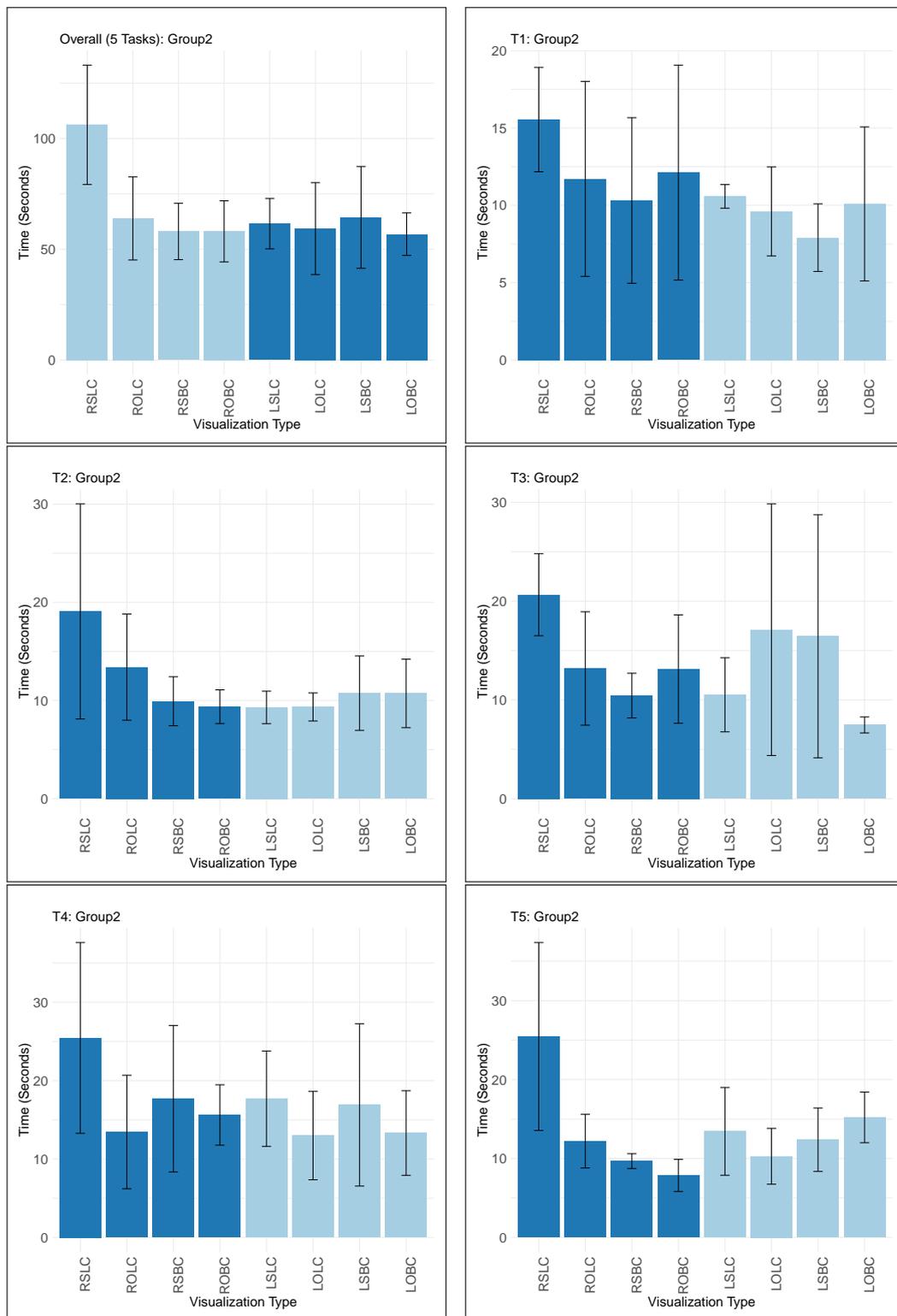


Figure 5.7: Task Completion Time for the Eight Visualizations for Group 2 (in the order presented). The top-left graph shows the mean and the SD of the overall task completion time and the others show the average time for each task individually. The Y-axis represents the average time, while the X-axis shows the visualizations based on their presentation order. (layout: Linear ■ | Radial ■).

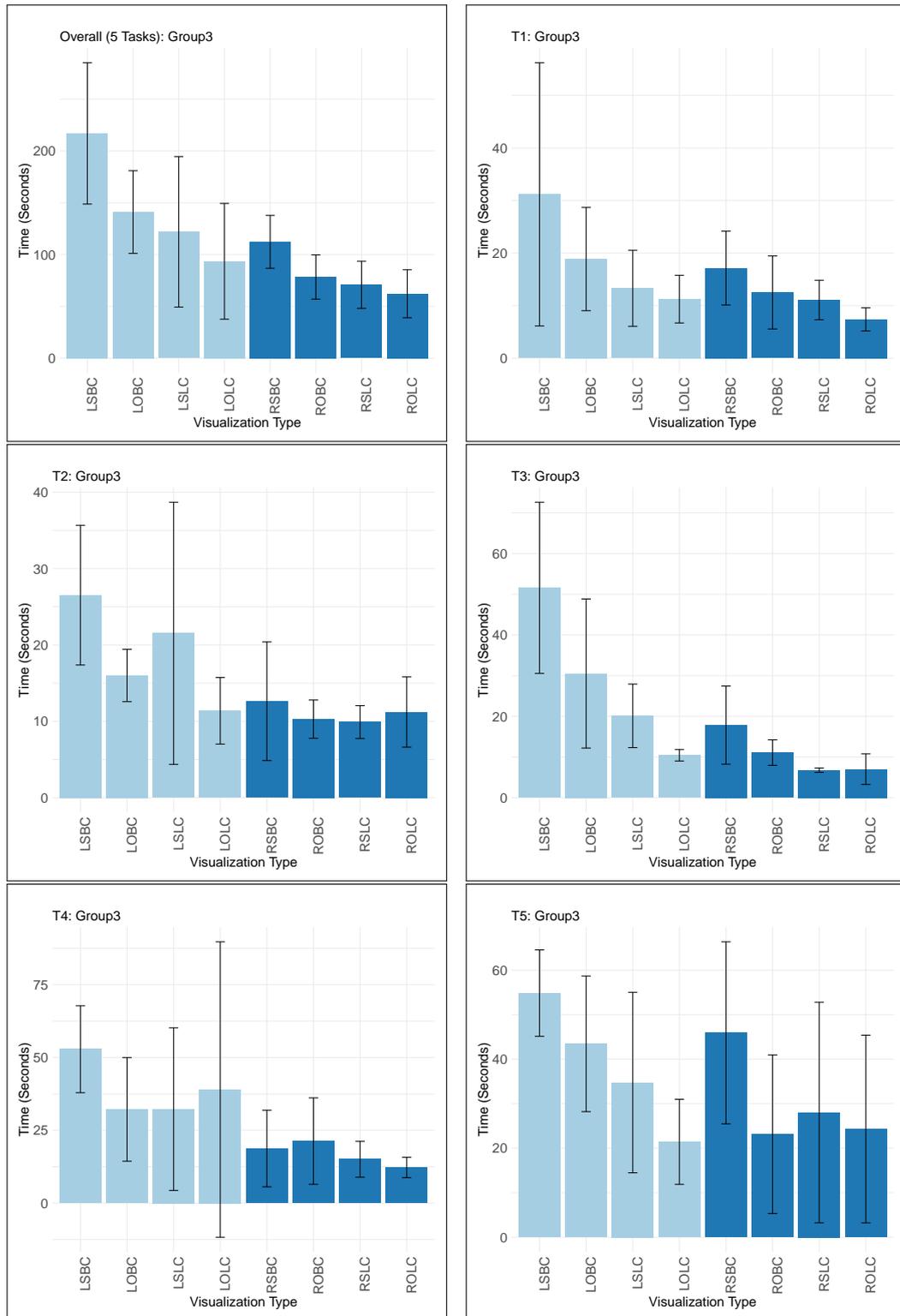


Figure 5.8: Task Completion Time for the Eight Visualizations for Group 3 (in the order presented). The top-left graph shows the mean and the SD of the overall task completion time and the others show the average time for each task individually. The Y-axis represents the average time, while the X-axis shows the visualizations based on their presentation order. (layout: Linear ■ | Radial ■).

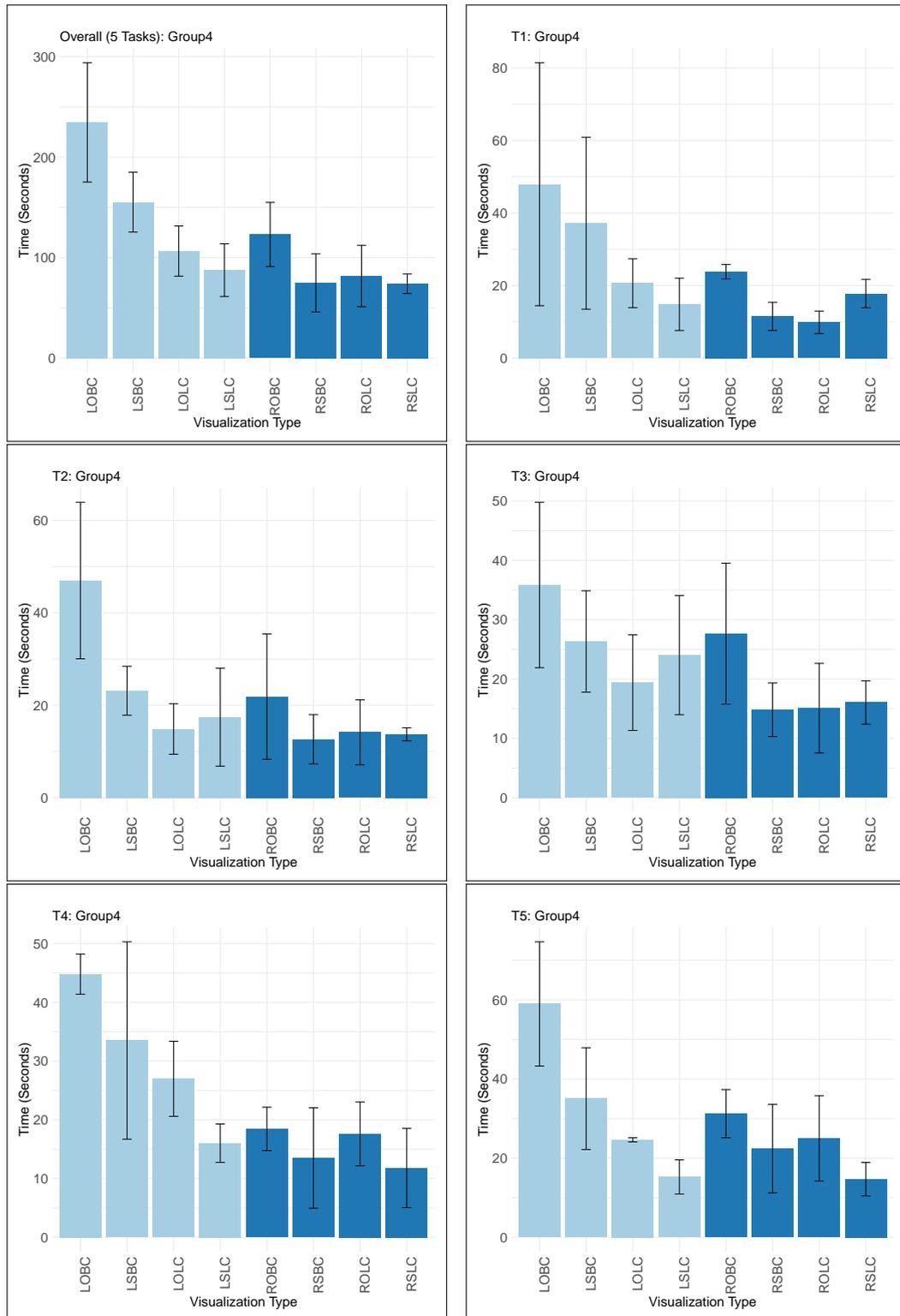


Figure 5.9: Task Completion Time for the Eight Visualizations for Group 4 (in the order presented). The top-left graph shows the mean and the SD of the overall task completion time and the others show the average time for each task individually. The Y-axis represents the average time, while the X-axis shows the visualizations based on their presentation order. (layout: Linear ■ | Radial ■).

5.4.2 Task Answers and Error Rate

To measure the accuracy of the tasks, the the answers entered or selected by the users were coded with '1' for the correct answer and '0' for the wrong answer for the tasks. The mean and the SD for overall accuracy (all five tasks) was calculated for each visualization based on the presentation order for each group. The graphs of the overall accuracy for each group are presented in Appendix .4 (at the top in Figures 1, 2, 3 and 4). The mean of overall task accuracy was calculated thus:

$$\text{OverallAverageAccuracy}_{(alltasks)} = \frac{\sum(\text{ParticipantsResponses})_{(alltasks)}}{\text{Number-of-Participants}}$$

The mean and the SD for participants' accuracy for each task in each group was also computed (see Appendix .4 Figure 1, 2, 3 and 4).

$$\text{Per-TaskAverageAccuracy} = \frac{\sum(\text{ParticipantsResponses})}{\text{Number-of-Participants}}$$

Furthermore, we calculate the percentage of task accuracy for each group to provide an easier representation of how accurate the participants were when responding to each task on each visualization and to facilitate the process of comparing the accuracy across all the visualizations and tasks. The percentage of task accuracy is shown in Figures (5.11, 5.12, 5.13 and 5.14), and it is calculated by the following equation:

Furthermore, the percentage of task accuracy for each group was calculated to provide an easier representation of how accurate the participants were when responding to each task on each visualization and to facilitate the process of comparing the accuracy across all the visualizations and tasks. The percentage of task accuracy is shown in Figures (5.11, 5.12, 5.13 and 5.14), was calculated using the following equation:

$$\text{Accuracy}\% = \frac{\text{ParticipantsResponses}}{\text{CorrectResponses}} \times 100$$

For each group, it was identified that participants' responses to T1 and T2, which are the 'Identify' and 'Determine' tasks, were more accurate than those responses to the other three tasks, as shown in Figure (5.11, 5.12, 5.13 and 5.14). According to the groups' performance (in terms of accuracy) for all five tasks, which is presented in Table 4.7, the accuracy for the 'Identify' and 'Determine' tasks ranges between 97% and 100 % for groups 1, 2 and 3, while for group 4 the accuracy for these two tasks is 78% and 91% respectively, whereas the average group accuracy for the other three tasks is between 63% and 67%. For the 'Compare' task, the participants in group 1 were the least accurate among the groups, with 28% average accuracy on all the visualizations. For the other three groups, it can be seen from Figures 5.13 and 5.14 that the participants in groups 3 and 4 performed the 'Compare' task more accurately on the Radial layout (with an average accuracy of 100% versus 88% for group 3 and 81% versus 44% for group 4), while as shown in Figure 5.12, group 2 performed more accurately on the Linear layout (94% average accuracy with the Linear layout versus 75% average accuracy with the Radial layout).

For the 'Infer' task, the participants in group 1 were also the least accurate with an average accuracy of 47%, while the other groups had an average accuracy of 59%, 78% and 84% respectively. The accuracy varied among the groups for this task for each visualization. For example, the 'ROBC' visualization has 100% accuracy for the 'Infer' task in groups 2, 3 and 4, while there was only 25% average accuracy in group 1. For the 'Locate' task, the participants in groups 2, 3 and 4 were least accurate when responding to this task on the 'LSLC' and 'LOLC' visualizations (the stacked and the overlapped linear line chart), as their average accuracy was 25 %, while the average accuracy on the other visualizations ranged between 50 % and 100 %. Group 3 had 100 % accuracy for the 'Locate' task on all the visualizations except for two visualizations (see Figure 5.13). For group 1, the average accuracy on 'LSLC' is 25%, while none of the four participants in this group responded correctly to the 'Locate' task on 'LOLC'. This could be related to the identification of the low accuracy rate for the 'Locate' task (which requires locating the hour when the user has minimum average HR) on the 'LSLC' and 'LOLC' visualizations. The reason for the difficulty in solving the task was probably because the values are close (71 and 74) and the two hours are adjacent on the Linear layout.

With respect to the order of the visualizations, no clear relationship has been identified between

Table 5.6: Participants Performance in Terms of Task accuracy in Each Group (on all the Visualizations in General).

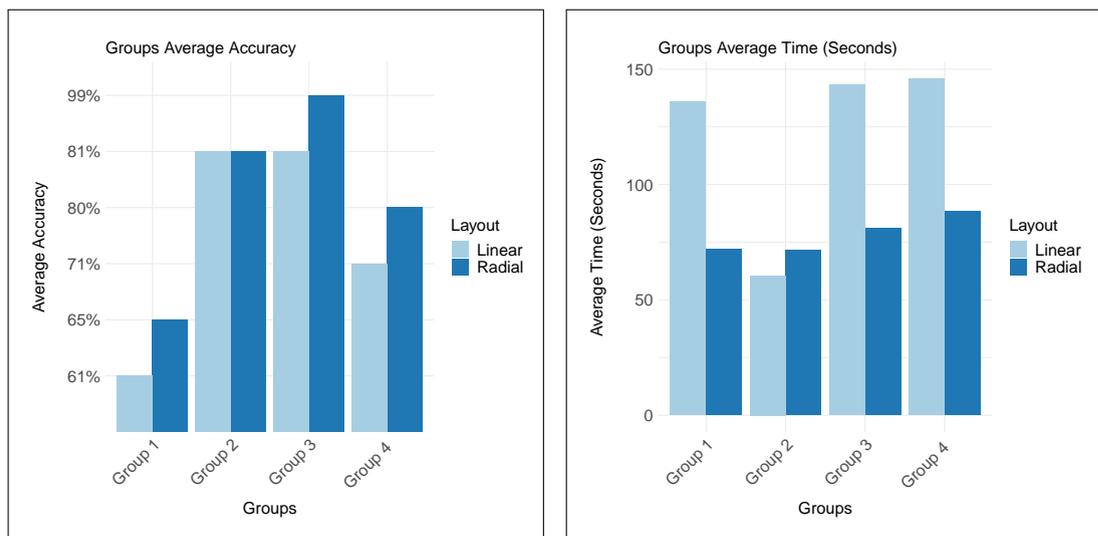
	Group 1	Group 2	Group 3	Group 4	Average
T1: <i>Identify</i>	97%	100%	100%	78%	94%
T2: <i>Determine</i>	97%	100%	97%	91%	96%
T3: <i>Compare</i>	28%	84%	94%	63%	67%
T4: <i>Infer</i>	47%	59%	78%	84%	67%
T5: <i>Locate</i>	47%	63%	81%	63%	63%

Table 5.7: Task Accuracy on the Radial and Linear Layouts for the Four Groups

	Group1	Group2	Group3	Group4	Average
Radial	65%	81%	99%	80%	81%
Linear	61%	81%	81%	71%	74%

the accuracy and the order of the visualizations similar to the relationship identified between task completion time and the order of the visualizations discussed in section 5.4.1. For example, the average accuracy for the ‘Determine’ task in groups 1 and 3 is 75% on the first viewed visualization then it increased to 100% on the other visualizations, while for group 4, the average accuracy for the ‘Determine’ task is 100% on the first four viewed visualizations and then it decreased to 75% on the next three visualizations, then increased to 100% on the last viewed visualization. This is the case for all tasks, as presented in Figures 5.11, 5.12, 5.13 and 5.14. The responses provided by each participant for each task were not the same on all the visualizations and the task accuracy was changed according to the visualization used. In the following section, all participants’ responses as a group are discussed with the aim of detecting any relationship between task accuracy and the type of the visualizations used.

In terms of the visualization layouts, Figure 5.10a shows participants’ accuracy in each group on both layouts. The participants in groups 1 and 4 performed more accurately with the Radial layout with a small difference of 4% and 9% respectively, while in group 3, the participants were more accurate by 18%. The participants in group 2 performed the same on both layouts with 81% overall accuracy. The average accuracy for all the groups showed that the participants were more accurate with the Radial layout with 8% only overall.



(a) Task Accuracy on the Radial and Linear Layouts for the Four Groups. The graph shows the average accuracy for the participants in each group with both layouts

(b) Task Completion Time on the Radial and the Linear Layouts for the Four Groups. The graph shows the average time for each group to complete the tasks on both layouts

Figure 5.10: Task Accuracy and Completion Time on the Radial and Linear Layouts for the Four Groups.

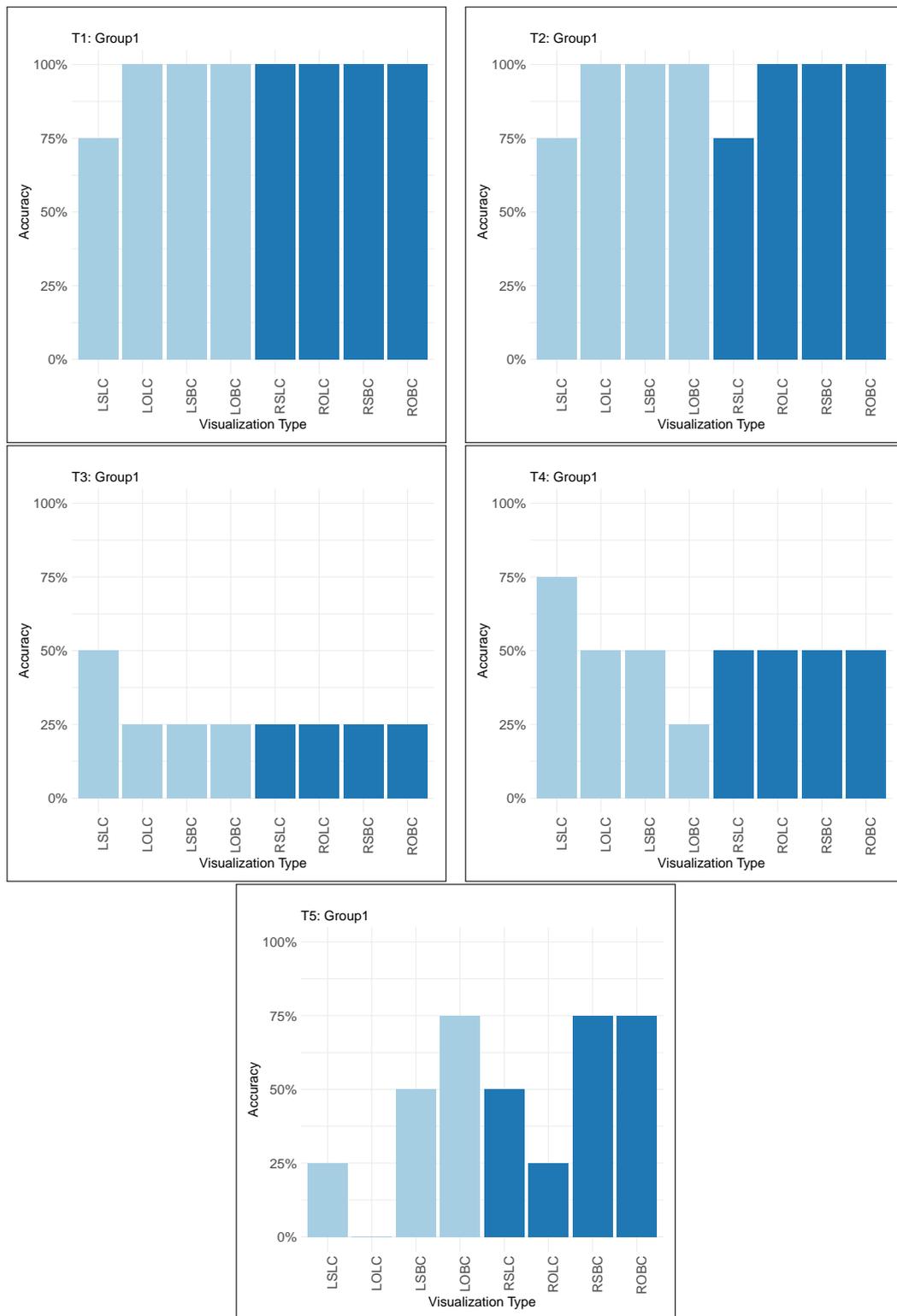


Figure 5.11: Group 1 Tasks Accuracy. The Y-axis represents the average accuracy (Percentage), while the X-axis shows the visualizations based on their presentation order. (layout: Linear ■ | Radial ■).

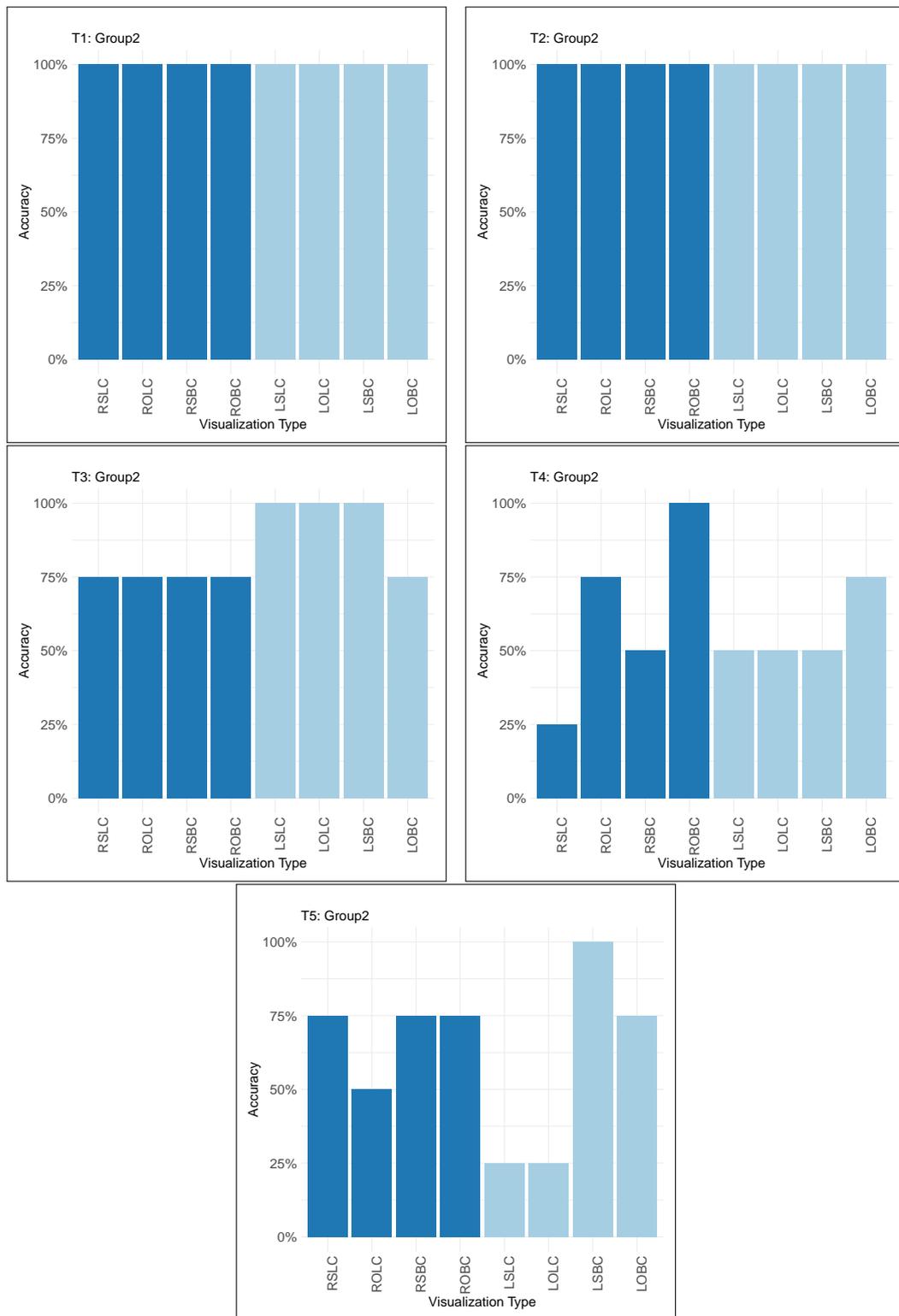


Figure 5.12: Group 2 Tasks Accuracy. The Y-axis represents the average accuracy (Percentage), while the X-axis shows the visualizations based on their presentation order. (layout: Linear ■ | Radial ■).

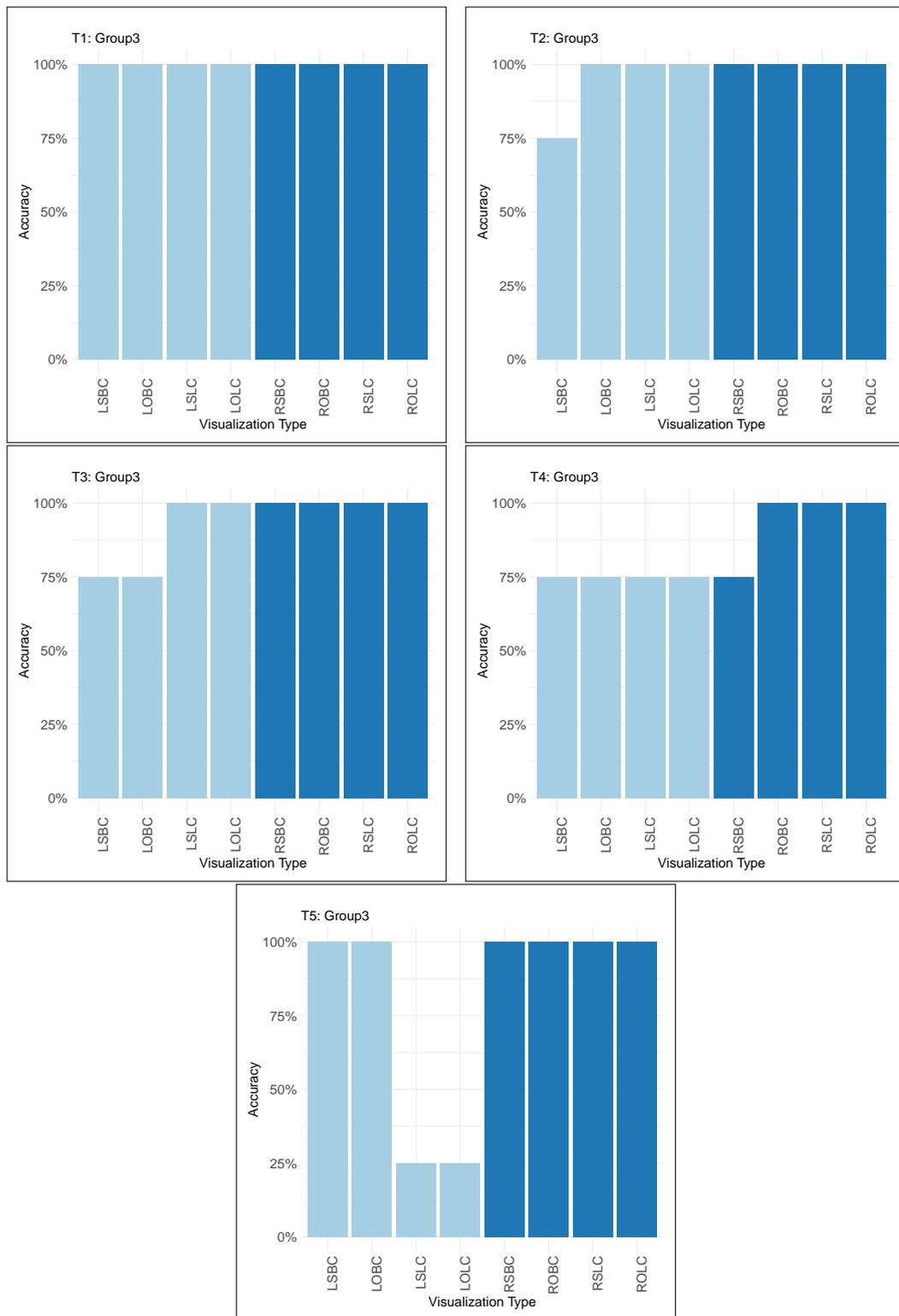


Figure 5.13: Group 3 Tasks Accuracy. The Y-axis represents the average accuracy (Percentage), while the X-axis shows the visualizations based on their presentation order. (layout: Linear ■ | Radial ■).

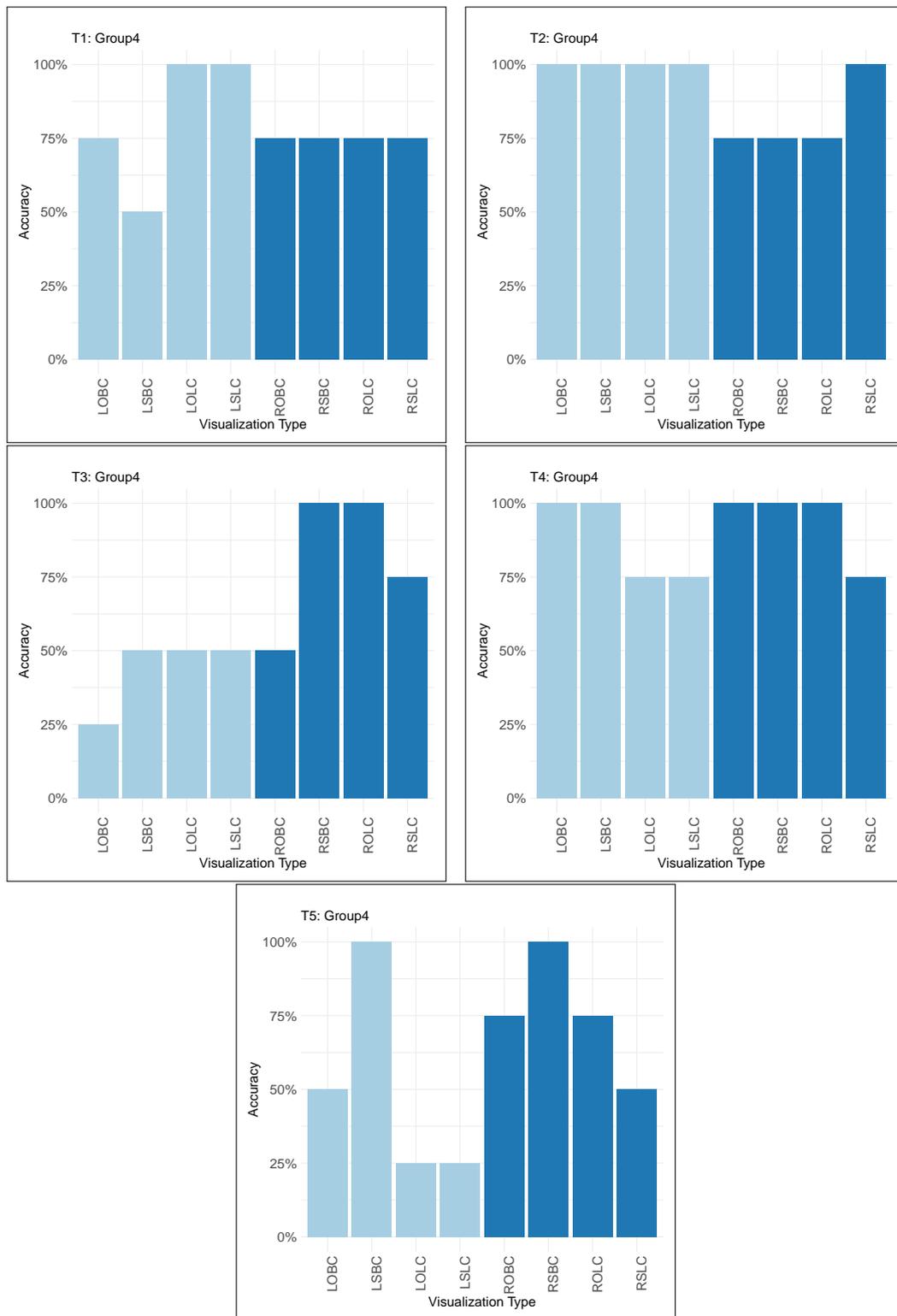


Figure 5.14: Group 4 Tasks Accuracy. The Y-axis represents the average accuracy (Percentage), while the X-axis shows the visualizations based on their presentation order. (layout: Linear ■ | Radial ■).

5.4.3 Overall Groups' Performance (Time and Accuracy)

In this section, the same analysis method performed previously on the data for each group is used. However, the performance data of all the participants is combined into one group and the order of the visualizations in the experiment are discarded. The mean and the SD for the task completion time and task accuracy for all the visualizations are calculated. The results of this analysis are presented in Figures 5.15 and 5.16. Moreover, the percentage of task accuracy for all the participants on all the visualizations are calculated to provide an overview of the accuracy on each task, on each design of visualization and to facilitate the comparison of them. The percentage of task accuracy is presented in Figure 5.17. This analysis is carried out to compare the participants' performance on all the visualizations and to determine whether there is a relation between the layouts and the task completion time and error rate.

With respect to the layouts, the analysis showed that the participants performed faster with the Radial layout by 36% on average for all five tasks. Figure 4.21 displays the difference in the average response time for the five tasks between the two layouts. The participants performed faster with the Radial layout by 39% for the 'Identify' task, 31% for the 'Determine' task, 38% for the 'Compare' task, 46% for the 'Infer' task and 23% for the 'Locate' task. Conversely, the accuracy for the 'Identify' and 'Determine' tasks was relatively high (85% and above) on each visualization with both layouts, as shown in Figure 5.17. According to Figure 5.18a and Table 5.9, which shows the accuracy on both layouts for each task, there is no difference in the accuracy for the 'Identify' task (94% with both layouts) and the 'Determine' task (94% with the Radial and 97% with the Linear) between the layouts. For the 'Compare' and 'Infer' tasks, the accuracy is slightly higher (by 6% and 9% respectively) with the Radial layout and for the 'Locate' task the accuracy with the Radial layout is higher by 13%. As reported previously in the analysis of the groups, the accuracy for the 'Locate' task accuracy on the 'LSLC' and 'LOLC' visualizations was low (25% and 19% respectively), while the accuracy for this task on the other six visualizations ranged between 63% and 88% (as presented on the 'Locate' task graph on Figure 5.17).

For task completion time ¹, the results presented in Figures 5.15, 5.18b and Table 5.8 show that the participants generally performed faster in all the tasks with the Radial layout (i.e. around

¹The differences in percentage were calculated of the Linear layout average time

Table 5.8: Task Completion Time on the Radial and Linear Layouts (average per task).

	T1:Identify	T2:Determine	T3:Compare	T4:Infer	T5:Locate	Average
Radial	12.93	13.43	14.08	16.38	21.35	15.63
Linear	21.12	19.35	22.56	30.35	27.88	24.29

Table 5.9: Task Accuracy on the Radial and Linear Layouts (average per task).

	T1:Identify	T2:Determine	T3:Compare	T4:Infer	T5:Locate	Average
Radial	94%	94%	70%	72%	75%	81%
Linear	94%	97%	64%	63%	52%	74%

36% based on the average time for all tasks). This is the case when discarding the order of the visualizations. However, looking back to the results for group 2 which started with the Radial layout (see 5.6 in section 5.4.1) to check whether this is the case in this group as well or whether it only resulted from the effect of the order of the visualizations, revealed that only the first viewed visualization had a significantly higher response time than the other visualizations (as discussed in section 5.4.1). To investigate the difference between the task completion time on the two layouts in relation to the presentation order of the visualizations, the performance of the participants in group 1 and group 2 were compared. These two groups were selected for comparison as group 1 started with the Linear layout, while group 2 started with the Radial layout with the same order of visualization types in each layout (see Table 5.3). The results show that the performance on the four visualizations with the Linear layout improved by 75% for 'LSLC', 56% for 'LOLC', 13% for 'LSBC', and 34% for 'LOBC' when presented after the Radial layout for group 2. Conversely, the task completion time on the other four visualizations with the Radial layout showed no improvement in terms of time when presented second in group 1 than when they were presented first as in group 2. It can be seen from the overall task completion time for groups 1 and 2 (the graphs at the top in Figures 5.6 and 5.7) that the task completion time improved only for 'RSLC' (by 33%), which was presented first for group 2, while it was presented fourth for group 1. However, for the other three visualizations (i.e. 'ROLC', 'RSBC' and 'ROBC'), the participants in group 2 who saw these visualizations as second, third and fourth performed slightly faster, by 14%, 14% and 33% respectively, than the participants in group 1 who saw these three visualizations as sixth, seventh and eighth.

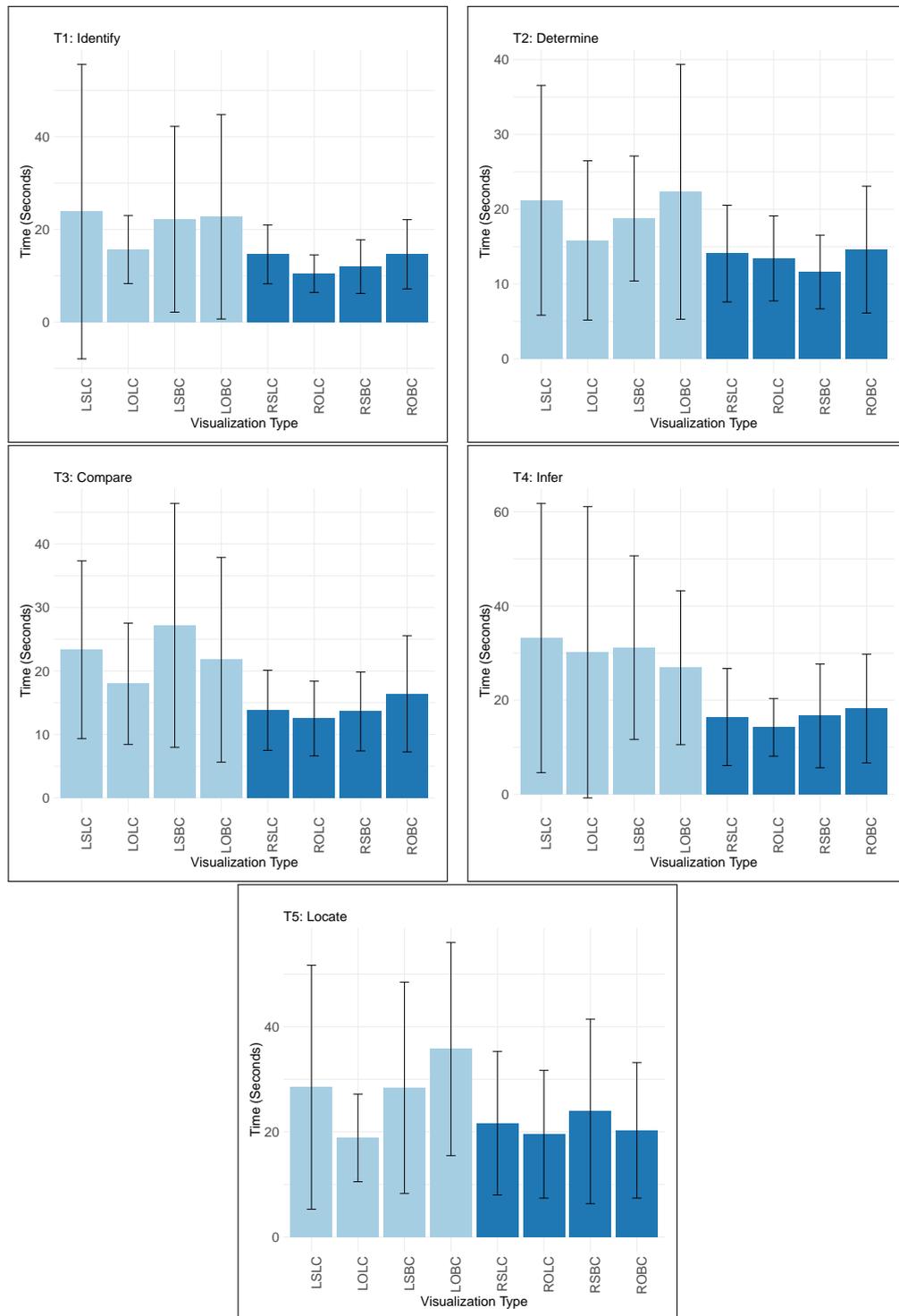


Figure 5.15: Overall Task Completion Time for the Eight Visualizations for All Four Groups (Mean and SD) (layout: Linear | Radial).

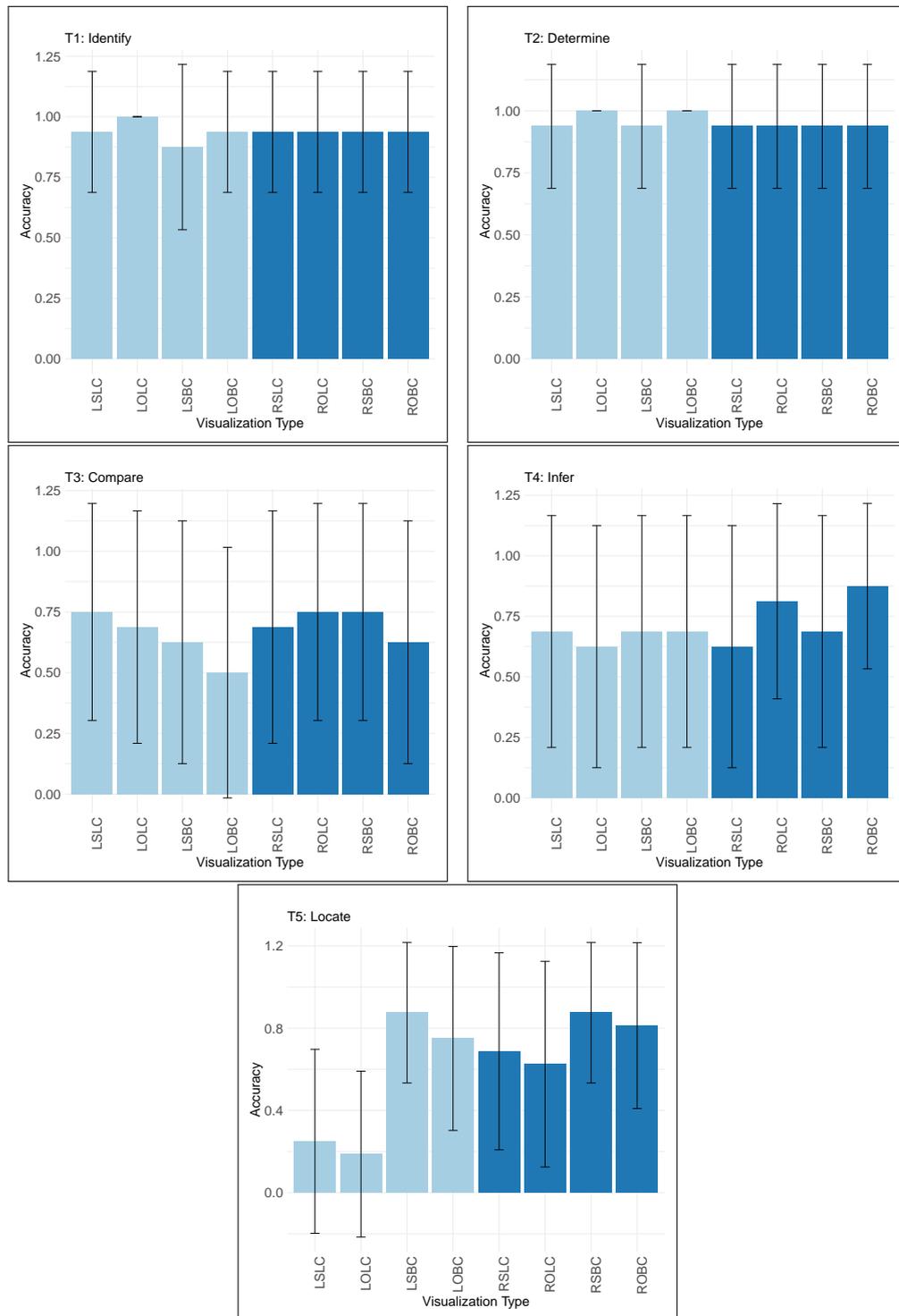


Figure 5.16: Overall Task Error Rate for the Eight Visualizations for All Four Groups (Mean and SD) (layout: Linear ■ | Radial ■).

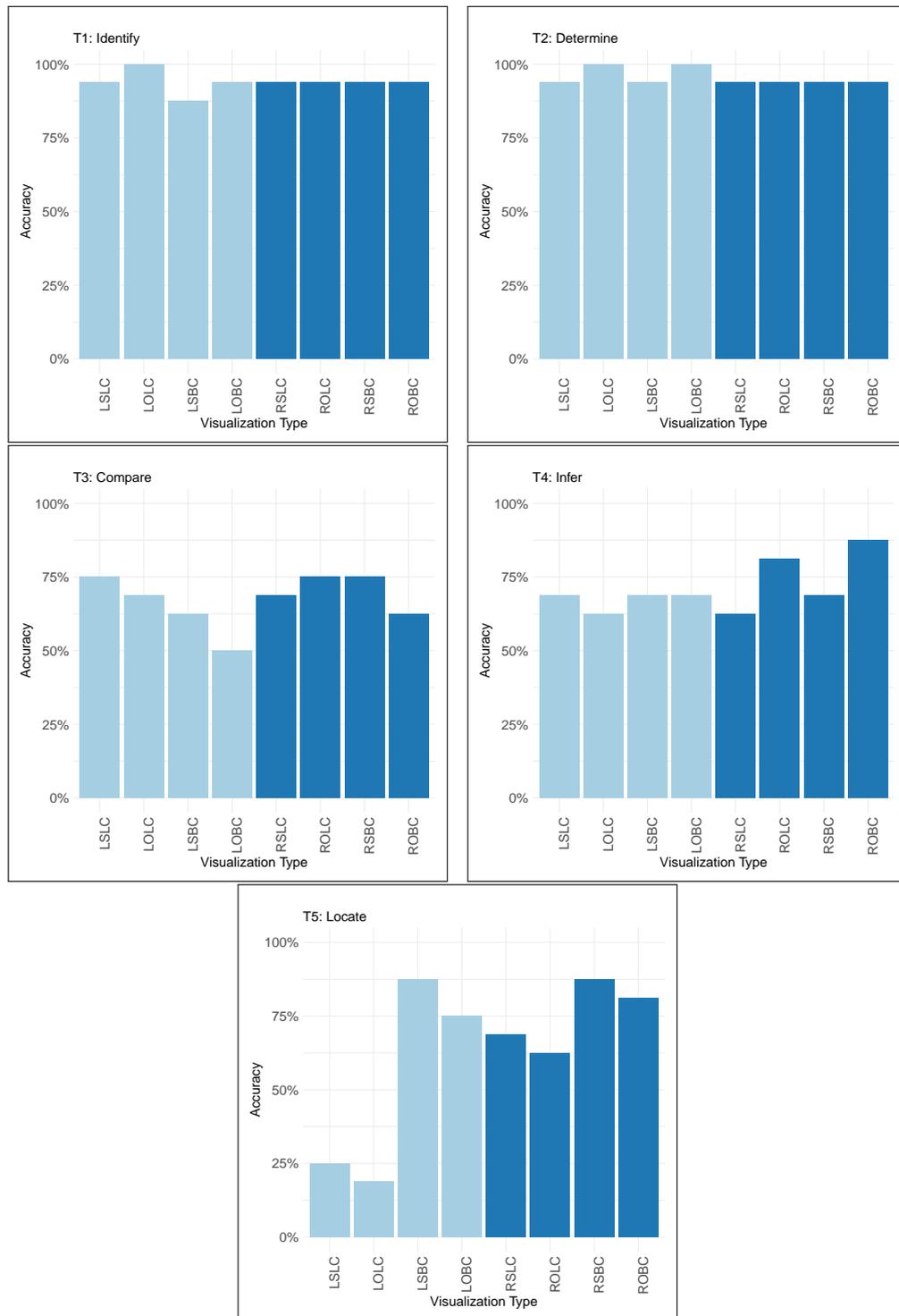
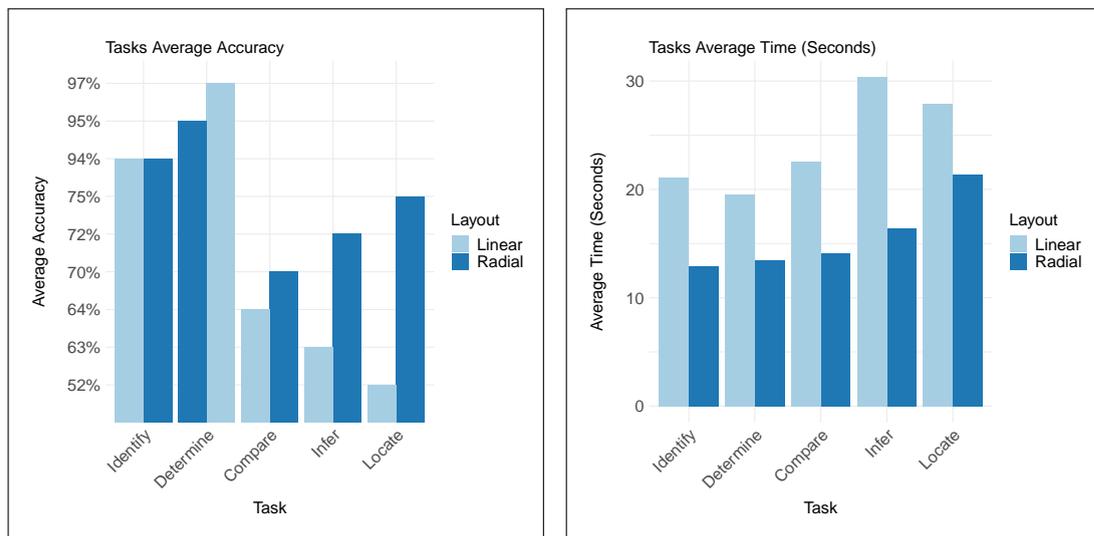


Figure 5.17: Overall Task Error Rate for the Eight Visualizations for All Four Groups (percentage) (layout: Linear | Radial).



(a) Task Accuracy on the Radial and Linear Layouts (average per task). The graph at the top shows the average accuracy for each task with both layouts

(b) Task Completion Time on the Radial and Linear Layouts (average per task). The graph shows the average time spent to complete each task by all the participants

Figure 5.18: Task Accuracy and Completion Time on the Radial and Linear Layouts (average per task).

5.4.4 Subjective Responses

The main criteria tested from the questionnaire used in this study are: (1) Is there an interest in the Radial layout from the participants? (2) Do the users' preferences differ after conducting the experiment than before conducting the experiment? The following subsections discuss the results from both questionnaires (both are attached in appendices .1 and .2).

5.4.4.1 Pre-experiment Questionnaire

The structure of this questionnaire consists of two sections. The first section includes basic demographic questions (i.e. age group, gender, level of education, if they track any health/physical activity data, if so, what data do they track? What devices or apps do they use for tracking?). The second section is about the visualization designs and it includes 8 Likert scale questions to rate the visualizations in terms of its readability and attractiveness of the design from 'Very Poor' to 'Very Good' (the readability and attractiveness are defined in section 1.3). It is followed by a multiple-choice question that asks the participants about their most preferred visu-

alization to represent their physical activity data. The complete questionnaire is attached in Appendix .1. Sixteen participants completed the study and responded to the questionnaires. Their demographic information is described in section 5.3.5.1.

The data from the pre-experiment questionnaire showed limited preferences in the presented visualizations. Only four out of the eight presented visualizations were preferred by the participants: 8 out of 16 participants liked the Linear Stacked Bar chart 'LSBC'; 6 out of 16 participants preferred the Linear overlapped line chart 'LOLC'; 4 out of 16 participants preferred the Linear Stacked Line chart 'LSLC' and 4 out of 16 participants preferred the Radial Stacked Bar chart 'RSBC'. However, the other four types of visualizations were chosen by any participants (see Figure 5.20).

For Likert scale questions (see Figure 5.22), the participants' ratings for the readability of the Linear visualizations are more distributed across the Likert scale with more higher ratings for the Line chart (9 and 7 out of 16 participants rated 'LSLC' and 'LOLC' respectively as 'Good' or 'Very Good'). While 'LSBC' had the lowest rating, being rated by 12 out of 16 participants as 'Poor'/'Very Poor'. For the Radial visualizations, the participants mostly rated its readability as 'poor' or 'very poor' as 10, 12, 14 and 15 out of 16 participants rated 'RSBC', 'ROBC', 'ROLC' and 'RSLC' respectively as 'poor' or 'very poor' in terms of the readability. Figure 5.22 shows that some participants gave a better rating for the Radial layout in terms of the attractiveness of the design than the readability, as the 'RSBC', 'ROBC', 'ROLC' and 'RSLC' are respectively rated as 'Good' or 'Very Good' by 10, 6, 2 and 2 out of sixteen participants. However, this shows that Radial Bar charts have a higher rating than the Radial line charts in terms of attractiveness and readability.

5.4.4.2 Post-experiment Questionnaire

The structure of the post-experiment questionnaire is the same as section 2 (i.e. the visualization designs section) of the pre-experiment questionnaire, see Appendix .2). This questionnaire aimed to compare users' preferences before and after conducting the experiment and to identify what the participants thought was the most appropriate visualization after performing the tasks. Participants' preferences after conducting the experiment changed. The data shows that the

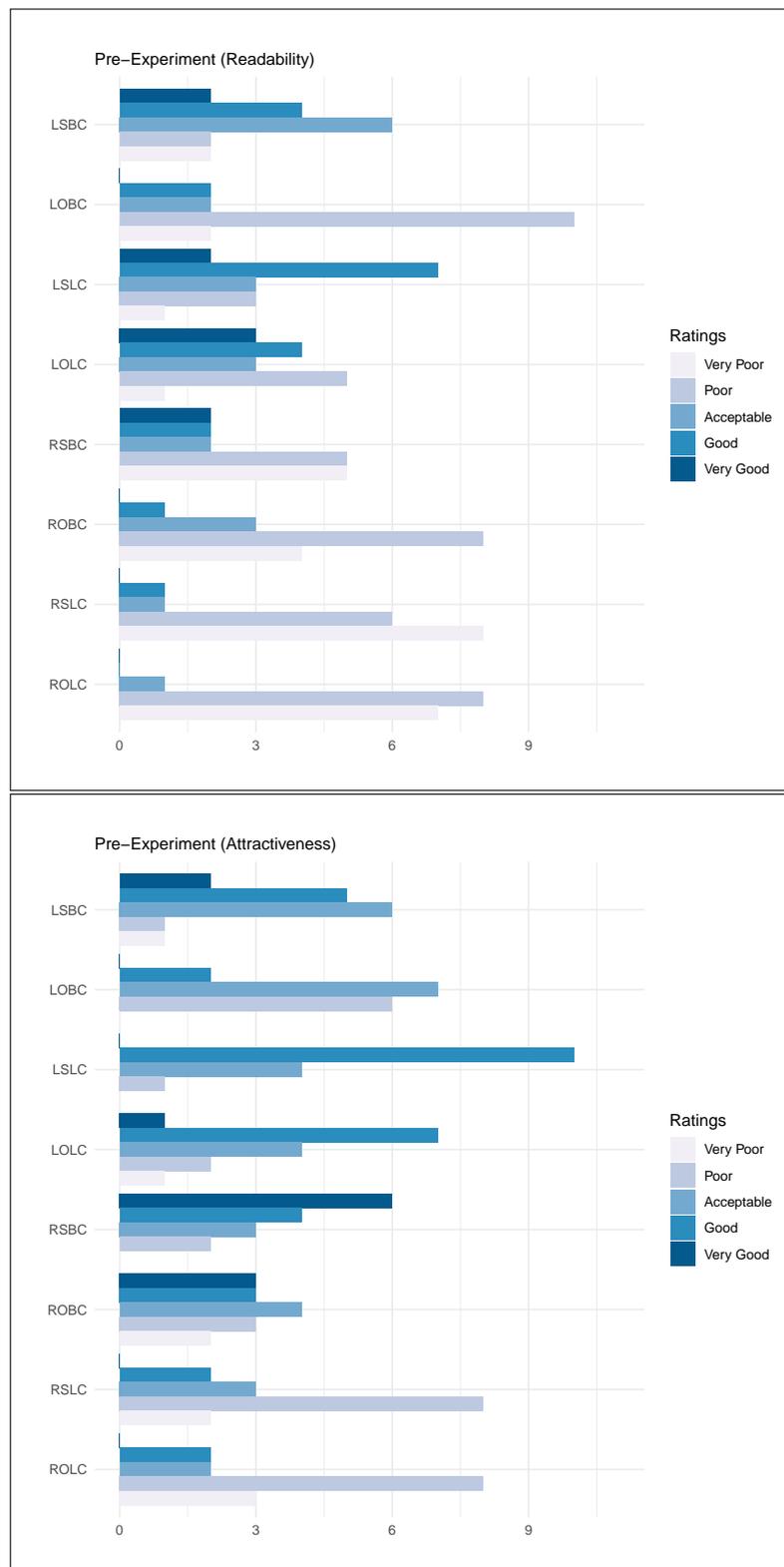


Figure 5.19: participants' Ratings for the Visualizations' Readability and Attractiveness Before the Experiment.

most preferred type of the visualizations is the Linear Stacked Bar chart 'LSBC' as 10 out of 16 participants preferred it, followed by the Linear overlapped line chart 'LOLC' which was preferred by 6 out of 16 participants, this was followed by the Linear Stacked Line chart 'LSLC' as it was selected by 5 out of 16 participants. The least preferred visualization in the Linear layout is 'LOBC' which was preferred by only 2 out of 16 participants.

The Radial Stacked Bar chart 'RSBC' was the most preferred visualization among the Radial layouts, it was preferred by 4 out of the 16 participants while the least preferred one was the Overlapped Bar chart 'ROBC' which was chosen by one participant only and the rest two Radial Line chart visualizations was preferred by 2 out of 16 participants for each one.

For the Likert scale data (see Figure 5.19), it can be seen that the Radial visualizations have a slightly higher rating (by one and two participants) in the 'Good' category than the rating before conducting the experiment for 'RSLC' and 'ROLC' visualizations. Similar to the results from the previous questionnaire, 'RSBC' had a better rating than the other Radial visualizations in terms of the readability, as it is rated by 4 out of 16 participants as 'Good'/'Very Good' in terms of readability, while the other three Radial visualizations were rated by only 2 out of 16 participants as 'Good'. Furthermore, in terms of the attractiveness, 'RSBC' was rated by 8 out of 16 participants as 'Good'/'Very Good', while for the other Radial visualizations it was rated as 'Good' by 3 to 5 out of the 16 participants.

The results from both questionnaires show that the participants had less interest in the Radial layout, but after performing the tasks on the all the eight visualizations, it is clear that some participants considered the Radial visualizations in the preferences question.

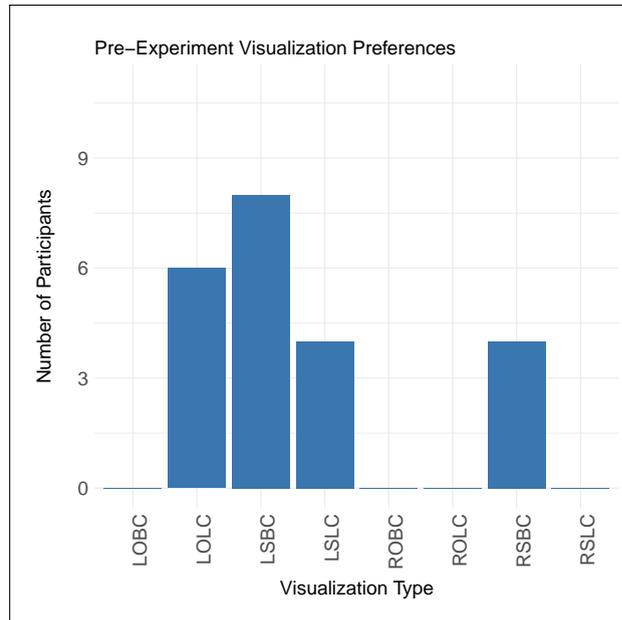


Figure 5.20: Participants' Preferences in the Visualizations Before the Experiment.

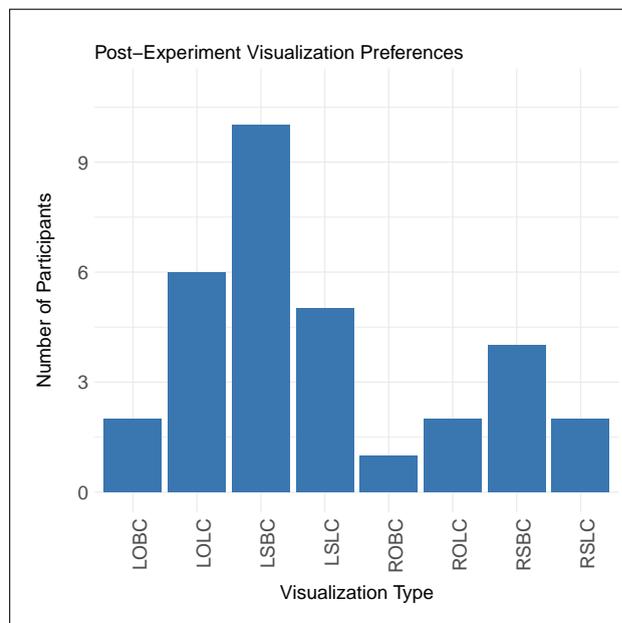


Figure 5.21: Participants' Preferences in the Visualizations After the Experiment.

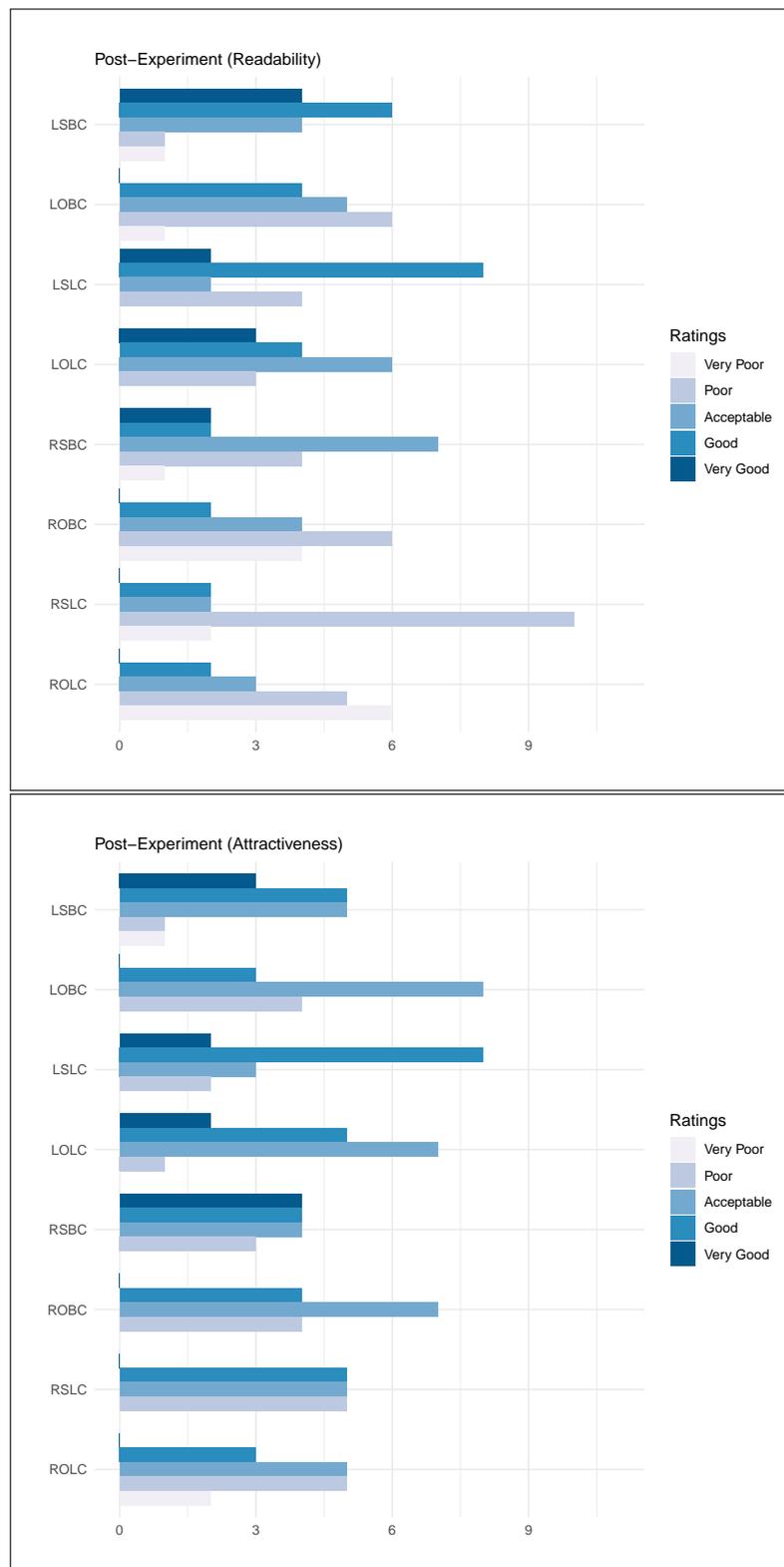


Figure 5.22: participants' Ratings for the Visualizations' Readability and Attractiveness After the Experiment.

5.5 Discussion

The quantitative results in the previous section were presented based on each group (in sections 5.4.1 and 5.4.2) and overall groups (in section 5.4.3). The analysis was based on the visualization design and tasks; in addition, an overall analysis was based on the layout. Although there was a noticeable difference between the task completion time on the first viewed visualization and the time on the visualizations presented afterwards, the task accuracy differed according to the visualization design (only on group 3 the accuracy of the tasks did not decrease except for LOLC and LSLC with the ‘Locate’ task as shown in Figure 5.13). These outcomes showed that whilst the participants became faster with the tasks, they were not necessarily learning as the accuracy did not increase (or remained the same). The following section considers the overall groups’ performance and preferences results based on the three basic visualization choices explained in the visualization design section (see section 5.3.2.1).

Layout: The analysis of the data from the evaluation study has revealed some interesting outcomes regarding users’ preferences and performance on the Radial and Linear layouts. With respect to the performance differences between the two layouts, the participants performed well with the Radial layout, even when the visualizations with the Radial layout were the first presented stimuli as discussed in section 5.4.1 and 5.4.2 (which is the answer to research question 3 in section 5.2). This outcome was encouraging and was the motivation for the next experiment: to investigate the efficiency of the developed Radial design, and overcome the effect of the order of the visualizations on the task completion time. In terms of the visualization preferences, only six of the 16 participants included Radial layout visualizations in their preferences selection after conducting the experiments, while only four of the 16 participants preferred these visualizations before the experiment. However, the participants’ performance (as discussed in section 5.4.3) on the Linear layout was not better than on the Radial layout. In terms of accuracy, there was a limited difference between the two layouts, as the analysis in Figure 5.10a shows that the average accuracy on the Radial layout was higher, by 7% overall. The participants also performed faster on the Radial layout, by around 36% (of the Linear layout average time). The methodology utilised for the collection of participants’ feedback needed to be expanded as the justifications for their preferences which helped in gaining better understanding of the data.

Visual Encoding: The most interesting outcomes regarding the type of visual encoding were in the ‘Compare’ and the ‘Locate’ tasks. In the ‘Compare’ task, in the overall groups’ task accuracy (see Figure 5.16 and 5.17), the Line chart within the Linear layout (i.e. LSLC and LOLC) encoding was more accurate than the Bar encoding (i.e. LSBC and LOBC). On the other hand, the Line chart encoding within the Linear encoding was less accurate than the Bar encoding for the ‘Locate’ task. However, within the Radial layout, only the Bar encoding was more accurate than the Line for the ‘Locate’ task. The Bar chart is recommended for finding extremes tasks as it is known to be effective for tasks that require the comparing of values as the human visual system excels at comparing Bar lengths [142].

Visualization Method: As explained earlier, the aim of evaluating the Overlapped and the Stacked design was to find a suitable method for representing the multivariate data variables. Within the Radial layout, there was no difference between the Overlapped and the Stacked design identified, with the exception of the ‘Infer’ task. The Overlapped designs (LOLC and LOBC) were more accurate than the Stacked designs (LSLC and LSBC) as shown in Figure 5.17. The ‘Infer’ task required observation of two variables in the multivariate dataset to infer the relationship between them. This was the aim of the Overlapped design, which facilitated the task. However, within the Linear layout, the Overlapped design was not more accurate than the Stacked design (LSLC, LSBC and LOBC had the same accuracy as they were 75% and the accuracy of LOLC was around 62%). For the ‘Identify’ and the ‘Determine’ tasks, the Overlapped method within the Linear layout (LOLC and LOBC) was more accurate than the Stacked method (LSLC and LSBC). However, the responses for those tasks were generally accurate (93% - 100% for the ‘Determine’ task and 88% - 100%).

The preferences also showed that the Linear Stacked Bar chart (LSBC) remained the most preferred design and preference increased from being preferred by 8 participants to 10 participants. The LOLC and RSBC remained the same, while the LSLC increased from 4 to 6. The other visualizations, which were not selected as preferred visualizations before the experiment, were selected after the experiment as presented in Figure 5.20 and 5.21. This showed that performing the visual tasks on the visualizations allowed the participants to consider other visualization options that were not preferred.

5.6 Lessons Learned and the Main Evaluation Study Design

According to the results discussed in section 5.4.1, the main issue identified from the experiment data is the significant difference between the task completion time for the first viewed visualization and the task completion time for the other visualizations presented afterwards. Figures 5.6, 5.7, 5.8 and 5.9 present the mean of task completion time for each group, and clearly show the significant drop (by 34 - 46% of the average time spent on the first viewed visualization) in time between the first viewed visualization and the other visualizations. In the overall tasks graph (top left graph in Figures 5.6, 5.7, 5.8 and 5.9), the average time for completing the tasks on the last viewed visualization is less than half of the time spent on the first viewed visualization for groups 1, 3 and 4 and almost half of the time in group 2.

Based on these results and the observations from the experiment, one possible reason for this difference in time could be that the participants may have required some time to understand the experiment/tasks and how to answer the tasks, this could have resulted in the high difference (34 - 46%) in the task completion time between the first and the next viewed visualization. However, for task accuracy, the results showed that the order has no effect on accuracy as they were more related to the type of visualization (see Figures 1, 2, 3 and 4). To address the issue of task completion time, and to guarantee unbiased outcomes in the main evaluation study, changes to the design included several aspects, including the datasets, the number of the groups and the visualization design.

Datasets: To overcome any learning effect during the study and to prevent the users from recalling the answers for the tasks, more than one dataset needed to be used for creating the visualizations, and they had to be presented to the users in random order. Using different datasets that have different data distributions and peaks allowed the evaluation of the visualizations under different data conditions to produce more reliable results.

Training Session: The participants should be given real tasks to complete in the training session for the study before running the main experiment. The tasks in the training session should be running in a similar way to the main evaluation study and the participants should be able to

see the correct answers for the tasks to compare their answers to ensure that they understand how the software works and to avoid possible bias in the analysis. Implementing real tasks and enabling the participants to assess their performance by seeing the correct answers to the tasks in the training session helped to ensure that the participants understood the tasks correctly. This also helped to avoid the error rate for some tasks that could result from misunderstanding the question/instruction (e.g., T3 for comparison in group 1, as shown in Figure 5.11), as reported in section 5.4.2.

Interactions with the Visualizations: It was assumed that the participants were able to recognise the answers easily as the values of the represented data were shown when they were interacting with the visualization by moving the mouse over it, which helped them to memorise the answers and simplified the tasks for the participants. Diehl et al. [53] in their evaluation study of Radial and Cartesian visualizations, stated that the interactive features in their experiment might have influenced the participants' performance.

Order of the Visualizations and the Number of Groups: The means of task completion time for the participants in the four groups show that they spent more time on the first viewed visualization than on the others. Since the preliminary evaluation study had only four groups, there were four visualizations that were not presented first (Table 5.3 shows the order of the presented visualizations in each group). To address this issue in the main evaluation study and to determine whether the order of the visualization affects the task completion time (as discussed in section 5.4.1), the participants were assigned to eight groups, where each group started with a different visualization from the eight designs.

However, according to the tasks' accuracy presented in section 5.4.2, the results showed that there was no clear relationship between the accuracy and the order of the visualizations, as in task completion time (discussed in section 5.4.1). The responses provided by each participant for each task were not the same on all the visualizations. The task accuracy changed according to the visualization used.

Subjective Feedback: For the pre- and post-experiment questionnaires, an open-ended question appeared in order to allow the participants to justify their preferences among the visualizations. This will provide a better overview and an explanation of why people like some charts and do not like others.

Designing the Visualization and the Implementation In developing the visualizations and mapping the data into the proposed chart types, a number of challenges were identified and that needed to be addressed when creating the visualizations for the main evaluation study.

- **No-data Values or Gaps:** For the used dataset, there was confusion between ‘no-data’ values and ‘no activity’ data, as both were recorded as zero values. The ‘no-data’ values in this experiment means that there was no recorded data from the device/app either due to a fault in the sensors or simply due to the fact that the user was not wearing/using the device. On the other hand, ‘no activity’ data means that the user was using the device, but no activity data was recorded (e.g., sedentary or sleeping). If the tracking device captures heart rate data, the difference can be identified by looking at the heart rate for the corresponding hour/minute. In the dataset used in this study, ‘no-data’ heart rate values were recorded as zero. For example, if the number of steps for a specific hour was equal to zero, it was not known whether this meant that the user was not active and did not take any steps at that hour or whether the user was not using the device without looking at the heart rate data for the corresponding hour.

In terms of visualizing this type of data, as two types of charts are used (i.e. lines and bars), ‘no data’ was not an issue in the Bar chart as it depends on the height of the Bar and this shows no data. However, in the case of the Line chart, as the data should be continuous to draw the lines, the ‘no-data’ heart rate was represented as zero, which caused confusion as this should mean ‘no data recorded’ and not zero.

- **Presenting the Visualizations:** In this study, the visualizations were rendered directly every time a new task was presented to the participants. To reduce the processing time in the main evaluation study (especially with the use of multiple datasets), the visualizations were presented as static images which were created earlier with Processing for all the used datasets and were saved with high resolution.

5.7 Conclusion

This chapter presented the design, implementation, and the results of the preliminary evaluation study that was performed to validate the used approach and the implementation of the lab experiments for evaluating the proposed visualization alternatives for representing the multivariate personal health data. It aimed to test whether the outcomes of the evaluation study could answer the research questions (as presented in Chapter 1) and detect and address any possible issues before conducting the main evaluation study. The outcomes of this study were beneficial to guide the design of the main evaluation study and for constructing the research hypotheses.

Main Evaluation Study of the Visualizations

6.1 Introduction

The outcomes from the preliminary evaluation study have revealed some interesting points regarding users' preferences and performance on the developed visualizations. For example, the overall results from performing the tasks on the visualizations showed that the participants perceived the Radial designs well. These outcomes were encouraging and were the motivation for investigating the efficiency of the visualizations by a follow-up experimental study to answer the research questions. The implementation of the preliminary evaluation study and its results directed the design of the main evaluation study in both the visualizations and the implementation of the study, which were discussed in section 5.6. This main evaluation study aimed to address the research questions that are related to the design of the visualizations, including the layout and the visualization methods, which led to a greater understanding of the efficiency of the evaluated visualization elements.

The main evaluation study followed similar methodology to that used in the preliminary evaluation study in the previous chapter and it included quantitative measures of the performance and qualitative measures of the subjective feedback. These were considered to assess the visualization from different perspectives. Each type of evaluation requires different methods of analysis, discussion and answering different research questions. Therefore, the data analysis and results of the main evaluation study is divided into two chapters: the quantitative aspect which is covered in this chapter, while the qualitative aspect is covered in Chapter 7.

This chapter starts with presenting the aims and research questions in section 6.2 followed by the examined hypotheses, which are explained in section 6.2.1. Section 6.3 presents the research methodology, which includes different aspects of the experiment design as follows: the visualization implementation is discussed in section 6.3.1; the experiment's visual tasks are presented section 6.3.2; the datasets are described in section 6.3.3; the participants recruitment in section 6.3.5 and the procedure, including the sessions during the experiments and the experiment setup, are explained in section 6.3.6. The analysis and results of the task performance part of the experiment are explained in section 6.4, (the qualitative part of the study is analysed and discussed in Chapter 7), the discussion of the task results and the hypotheses testing outcomes are discussed in section 6.5 and 6.5.6. The chapter concludes by a summary of the

outcomes based on the tasks and the data characteristics in section 6.6.

6.2 Aims and Research Questions

As explained earlier, this thesis aims to design and evaluate visualizations for multivariate personal health. The visualization designs and evaluation procedure were examined in the previous chapter, the outcomes of which showed an ability to answer the research questions. However, this chapter develops the revised version of the visualizations and the study that was designed based on the following research aims (i.e., **RA3** and **RA4**):

1. Developing different visualization design alternatives for representing multivariate personal health data using options of visual layout, visual encodings and methods to address the identified limitations of the current visualizations methods.
2. Evaluating participants' performance when using the developed visualizations by implementing a set of visual tasks that are related to the real health dataset to identify the most effective design for the data and the implemented tasks.

The evaluation in the main study sought answers to some of the research questions presented in Chapter 1. The research questions addressed in this chapter are of two types: the first type is related to the design of the evaluation visualizations (which are the third and the fourth research questions), these are:

- **RQ3:** What visual layout could be used to support interpreting and identifying daily patterns in the data and to provide a relevant context? Would a design informed by data ontological interpretation, in this context the Radial clock layout for daily data, facilitate understanding and interpretation of data and performances of visual tasks?
- **RQ4:** How can multivariate personal health data be visualized? What are the methods that could be used to integrate the data into one visualization? What could be used as a visual encoding method to represent the data?

The initial design of the visualizations was outlined in the previous chapter; however, further details and enhancements to the design are addressed in section 6.3.1 of this chapter. The second type of question is related to the performance with these visualizations, this is the fifth research question. The participants' performance with visualizations and the most and least effective design are discussed in the results section. The fifth research question is:

- **RQ5:** What are the differences between visualization methods that could be used to represent multivariate personal health data in terms of users' performance? What are the most effective methods for each of the implemented tasks, and what are the least effective ones?

6.2.1 Hypotheses

Throughout the thesis, a set of hypotheses were developed, starting from the literature review to the preliminary evaluation study. The hypotheses were constructed based on previous studies outcomes on the visual perception and evaluation of similar tasks, in addition to the results from the evaluation study conducted in Chapter 5 if applicable. These hypotheses were summarized earlier in the introduction of the thesis in Chapter 1. The hypotheses, with their justifications, are listed in the following sections and are based on the tasks (the list is also presented in Table 6.29). These visual tasks were experimented and explained in section 5.3.3 in the previous chapter. The number in the hypotheses refers to the task, the first part (i.e., the a) is related to a comparison between the two layouts, and the second part (i.e., the b) is related to either the type of the visual encoding or the visualization method.

6.2.1.1 Identify

H1a: It is hypothesised that using the Radial layout (clock metaphor) to represent activity data during the parts of the day (morning, afternoon, and evening) allows users to compare the data and identify trends faster than when using the Linear layout.

The preliminary evaluation showed that the participants performed the 'Identify' task faster with the Radial layout with higher accuracy that was equal to their accuracy with the Linear

layout for all groups (see Figure 5.15 and 5.16).

From the literature: the literature search revealed no evidence that showed the outperformance of one layout over the other one for a similar task. However, from the design perspective, Radial visualizations benefit from the clock metaphor to represent time data. The clock-style visualization was found to be effective for temporal data, as the natural order of time enhances perception of the data [67, 126]. The coloured sectors refer to different parts of the day, which may allow perceiving each part as a whole and facilitate estimating the total calories within each part according to the visual grouping principles. The symbols (glyphs) are used as extra information to refer to the time. Glyph-based designs can provide efficient information encoding and can attract more attention [29].

H1b: This hypothesis states that the Line chart encodings within the layouts facilitates finding trends in the data over time, therefore, enabling a faster performance of the task than the Bar chart.

The preliminary evaluation: this showed that the task performed with the Line and Bar chart was generally performed with similar accuracy (except the LSBC had less accuracy than others).

From the literature: previous evaluation studies on the Bar and Line chart gave an overview of how the task could be performed with these charts (e.g., [142, 186, 247]). The Line chart has been identified in the literature as being effective for finding trends tasks [142] and characterising distributions [186]. The Line chart has been found to be easier for perceiving trends, while the Bar chart has been found to be easier for comparing two discrete data values [247].

6.2.1.2 Determine

H2a: For determining the value, it is hypothesised that this task is more accurate and faster in the Linear layout than in the Radial layout.

The preliminary evaluation showed that the total accuracy for all the groups (see Figure 5.16) for the visualizations with the Linear layout was higher than the Radial layout (LOBC and LOLC were the most accurate).

From the literature: several types of research that have compared the Radial with the Linear layout have shown that the Linear layout outperformed the Radial layout in tasks that require reading and retrieving values. For example, Brehmer et al. experiment on mobile phones [32] and Goldberg and Helfman eye-tracking experiment with quantitative information [96]. The results from Waldner et al.'s experiment [238] with time-dependent data also showed that the linear layout is less error-prone than the Radial layout for the reading value task and it is also more time-efficient. Moreover, from perception point of view, reading values on the Radial layout was expected to be more difficult than on the Linear layout because the position of the value could possibly be affected by the angle on the Radial design. According to Cleveland and Cleveland and McGill [48] reading a value from a circular line is less accurate than estimating the position on a common scale.

H2b: In terms of the chart type, it is hypothesised that this task is more accurate and faster with the Bar chart than with the Line chart on both layouts.

From the literature: according to Saket et al.'s study [186], people performed faster and more accurately with the Bar chart than the Line chart. Experiments on graphical perception showed that people performed more accurately when perceiving information that encoded using length than other encoding elements such as angle [48, 217]. Moreover, from the design point of view, for the Bar chart, the link between the specified hour on the x-axis and the height of the Bar could lead to faster and more accurate value location and then value estimation.

6.2.1.3 Compare

H3a: The comparison of visualizations over two days is faster and more accurate on the Linear layout than the Radial layout.

Although the results from the *preliminary evaluation study* showed that the same visualization design was more accurate on the Radial layout than the Linear layout, it was expected that the comparing values of two visualizations on the same the horizontal level could be easier. This was also expected because the interactive features (i.e., showing the values when moving the mouse over the charts) were removed in this study as that was believed to facilitate the tasks.

From the Literature: previous studies on Radial and Linear layouts implemented comparison tasks but they are different from the ‘Compare’ task used in this study. Waldner et al. [238] implemented the compare task as comparing AM/PM interval values, while Brehmer et al. [32] required comparing one value with others and comparing two ranges (two subsets of the data). Therefore, their results are not entirely related to the comparison task here. From the design perspective, this hypothesis relied on the fact that perceiving values on a common scale is more accurate than angles. According to Cleveland and McGill [48], the slope and angle are less accurate than position and length judgement. Also, Fuchs et al. [79] hypothesised that mental rotations might be required when performing comparisons and this could make the tasks that require value comparisons harder on the Radial representations.

H3b: The comparison of visualizations of the multivariate dataset over two days is faster and more accurate on the Stacked design visualizations than the Overlapped ones.

The preliminary evaluation: showed that within the Linear layout, the Stacked design of the same chart was more accurate than the Overlapped design (e.g., LSLC was more accurate than the LOBC).

From the literature: Based on the task description, in this hypothesis, it is assumed that it is easier to compare the variables separately rather than being Overlapped with other variables. As the task requires visually searching for the target (i.e., calories) and visually exploring the data distribution to estimate its total values, it could be affected by other variables. In the Stacked design, the effect of the distractors could be reduced, the participants are expected to find the value and estimate the distribution and the total calories faster than when the variable is Overlapped with the others. According to Javed et al. [121]’s study on the perception of multiple time-series charts, the Overlapped (i.e., shared-space) representation could increase the difficulty of following a single line over a distance.

6.2.1.4 Infer

H4a: The Linear layout is faster and more accurate at finding the relationship between two variables from the multivariate dataset than using the Radial layout.

From the literature: according to Goldberg and Goldberg and Helfman's [96] eye-tracking experiment, the values labelled in the Linear layout are aligned horizontally and vertically, which better support organised searches for values, unlike the Radial layout where the value labels are aligned on a circle. This task requires searching and reading values to examine the condition in the task question. Reading value of step count is similar to the 'Determine' task, and it is expected to be more accurate and faster with the Linear layout based on [48].

H4b: The Overlapped chart is faster and more accurate at finding the relationship between two different data variables from the multivariate dataset, compared to the Stacked chart.

The preliminary evaluation study: showed a difference between the Overlapped and the Stacked design according to the layout. Within the Radial layout the Overlapped designs are more accurate than the Stacked, while within the Linear layout, the Stacked designs are more accurate than the Overlapped.

From the literature: the 'Infer' task implies examining a relationship between two variables, which could include comparisons. According to Gleicher et al.'s survey [95] of visual comparisons, the visualizations in this thesis apply the juxtaposition design (i.e., Stacked) and the superposition design (i.e., Overlapped) as discussed in section 5.3.2 in the previous chapter. The juxtaposition is also known as small multiples or side-by-side display according to the presentation. The challenge in this design is in helping the user with identifying the relationships between the objects visualized separately. It relies on people's ability to see patterns in the separated objects which may be required to be adjacent in the design. On the other hand, the Superposition design that overlays objects on the same design space, but which are encoded differently (e.g., different colours). This overlaying of the variables could be challenging when data are dense, as it may produce issues of visual clutter.

As there are only three variables Overlapped in this study, it is expected that the Overlapped design outperforms the stacked design. Javed et al. [121] found that superimposed design (i.e., the Overlapped) outperforms the juxtaposed (i.e., the Stacked) design at comparisons with a local visual span as it benefits from the shared space when performing the comparison. The Juxtaposed design requires eye movements between the visualizations to perform the comparisons. From the design perspective, the Overlapped design was developed to facilitate the task

of finding relationships between different health variables. It was expected to help the participants to infer the relationships between the variables when they share the same visualization space.

6.2.1.5 Locate

H5a: Locating maximum values occurs faster and more accurately with the Linear layout than the Radial layout.

From the literature: this hypothesis follows the same reasoning as H4a for task T4.

Also, the comparison between the Radial and the Linear visualizations by Brehmer et al. [32] showed the Linear outperforms the Radial layout. Waldner et al. [238] experiment with time-dependent data showed that the Linear layout was faster than the Radial layout for the locating maximum values task.

H5b: Locating maximum and minimum values occurs faster and more accurately with the Bar chart than the Line chart.

The preliminary evaluation study showed that the visualizations with the Bar chart are more accurate than the same visualization with the Line chart. Within the Linear layout, the LSBC and LOBC are more accurate than the LSLC and LOLC and within the Radial layout the RSBC and ROBC are more accurate than the RSLC and ROLC.

From the literature: the Bar charts are effective for comparing values because the human visual system is good at comparing Bar lengths and it is recommended for finding extremes tasks [142]. Saket et al. [186] also found that the Bar chart is significantly faster than other evaluated visualizations for similar tasks, including the Line chart.

6.3 Methodology

The methodology and experiment design of the main evaluation study is similar to the design of the preliminary evaluation study in terms of the developed visualizations, the visualization tasks, data collection method and the procedure. The following sections describe the methodology while highlighting the differences from the design of the evaluation study in Chapter 5 and refers back to the related sections.

6.3.1 Visualization Design and Development

The implementation of the visualizations follows the same designs as those presented in Chapter 5, using Processing [78] and ControlP5 GUI library [201]. However, some amendments were performed to enhance the layout and visual mapping process of the three data variables, this is expanded upon in the following subsections.

6.3.1.1 Visual Layout

The used layouts are the same as those used in the previous chapter: the Linear layout and the Radial clock layout (see section 5.3.2 for further details). The Linear layout is the traditional design where the x-axis represents the time, while the Radial layout represents the time on the perimeter of the circle showing the 24 hours of the day. The four coloured sectors refer to different parts of the day: morning; afternoon; evening and night (see Figure 6.1). The first hour of the day (0 hours) was moved to the top right sector of the clock metaphor instead of the bottom right sector to refer to the beginning of the day. Moreover, the way the day's sectors divided was changed to distribute the hours of the day across the a.m. and p.m. evenly.

This separation was only used on the Radial design as it was supposed to show a clock metaphor, while the Linear design was the traditional design used to represent this type of data on the dedicated dashboards. The day's sector was divided into an a.m. and p.m. separation on the 24-hour clock (12 a.m. & 12 p.m.), where the afternoon and evening are usually referred to as pm and night and morning as am. They were also divided to include an equal number of hours in each sector. The Radial visualizations are shown in Figures 6.2d, 6.2h, 6.2b and 6.2f,

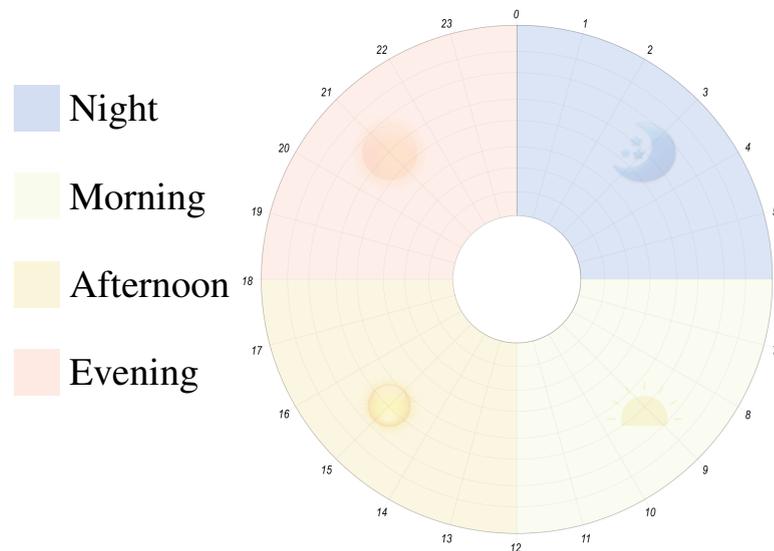


Figure 6.1: The Background of the Proposed Radial layout Showing the Four Coloured Sectors, the Glyphs and the Order of the 24 Hours.

while the visualizations with the traditional Linear layout are shown in Figures 6.2c, 6.2g, 6.2a. The abbreviations used for the visualizations are presented in Figure 6.2. The implementation of both layouts used the same scaling method, the difference being in the orientation of the axes. However, it differed from the scaling method applied in the first evaluation study and is discussed in section 6.3.1.4.

6.3.1.2 Visual Encoding

The visualizations used the same visual encodings implemented in the study in the previous chapter (i.e., Bar and Line chart encoding). The results from the preliminary evaluation showed the visual encodings vary according to the layout and the task, which required further investigation in this study. The Bar chart is shown in Figures (6.2c, 6.2g, 6.2d and 6.2h), while the Line chart is shown in Figures (6.2a, 6.2e, 6.2b and 6.2f).

Colour Selection: The visualizations in this experiment follow the same colour selection of the visualizations in the preliminary evaluation, which is discussed in section 5.3.2.3.

6.3.1.3 Visualization Method

The visualization design in the study followed the same two methods as those described in section 5.3.2 when representing the selected health data variables (steps, calories, and HR). The study compared the visualizations for when the data is Stacked (i.e., juxtaposed) in separate charts (shown in Figures 6.2c, 6.2a, 6.2d and 6.2b) versus representing the data variables in an Overlapped design (i.e., Superposed), which represent the data on the same chart area (shown in Figures 6.2g, 6.2e, 6.2h and 6.2f).

6.3.1.4 Visual Mapping

Visual mapping is the process of visually representing the data, it includes three structures [146]: (1) The spatial dimensions of the data in the space where the visualization created (can be defined in terms of axes) (2) The graphical elements that are used to represent the data (e.g., point or line) (3) The graphical properties of the used graphical elements (e.g., size, orientation or colour) [146].

The difficulty of visually mapping the data is in selecting the graph scale for the axes. For the values of the same variable (e.g., the step count), there are a number of extreme values (e.g., 3000) and at the same time there are many zero values. Also, in the case of the Overlapped design, HR data range between 70 to 200 and it needs to be Overlapped with the step count that could reach thousands of steps at one time point. The following methods from the literature were used to deal with the challenges of the different scaling options.

- **Standard (Linear) Scale:** Using the standard Linear scale with our dataset produce visualizations that have too many small bars compared with some extreme bars due to the huge variance between the values of the visualized variables, which will affect performing some visual tasks such as a quantitative comparison.
- **Logarithmic Scale:** The main problem with using the log-scale method for mapping is in the difficulty of performing estimating values task. As stated by Hlawatsch et al. [109] the log-scale is suitable for qualitative comparisons such as finding extreme tasks, while

Table 6.1: Explanation of the Abbreviations Used for the Visualization Names in this study.

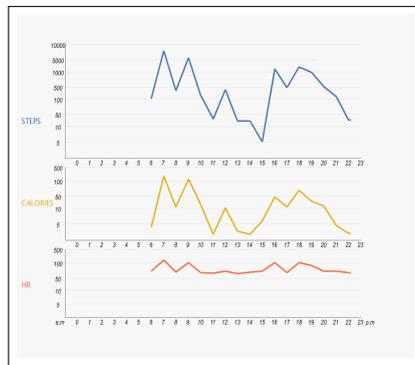
Abbreviation	Visualization Type
LSLC0/ LSLC1	Linear Stacked Line Chart (0 for Dataset 0 and 1 Dataset 1)
LOLC0/ LOLC1	Linear Overlapped Line Chart (0 for Dataset 0 and 1 Dataset 1)
LSBC0/ LSBC1	Linear Stacked Bar Chart (0 for Dataset 0 and 1 Dataset 1)
LOBC0/ LOBC1	Linear Overlapped Bar Chart (0 for Dataset 0 and 1 Dataset 1)
RSLC0/ RSLC1	Radial Stacked Line Chart (0 for Dataset 0 and 1 Dataset 1)
ROLC0/ ROLC1	Radial Overlapped Line Chart (0 for Dataset 0 and 1 Dataset 1)
RSBC0/ RSBC1	Radial Stacked Bar Chart (0 for Dataset 0 and 1 Dataset 1)
ROBC0/ ROBC1	Radial Overlapped Bar Chart (0 for Dataset 0 and 1 Dataset 1)

difficulties could arise when performing other quantitative tasks such as estimating or quantitative comparisons.

- **Dual Scale:** Another option was considered is the use of a Dual Scale to map the data in the step count differently in the scale used for the calories and HR. The Dual-Scale method was derived from Isenberg et al.'s [120] and Hlawatsch et al.'s [109] studies. Although this method was suitable for applying on the visualizations on the Linear layout, it was not appropriate for the visualizations on the Radial layout.

Neither are any of above methods suitable because of the difficulty in them being applied with some of the proposed visualization choices, or because of their difficulty with some of the visual tasks such as comparisons or estimation. However, as the log-scale seems to be suitable for representing data for the majority of tasks (except for value estimation, as it will be difficult for both expert and people in general) and as it works well with all our proposed visualizations, it has been applied with all the charts but using linear incremental values in the mapping process to facilitate the task of determining and estimating the values.

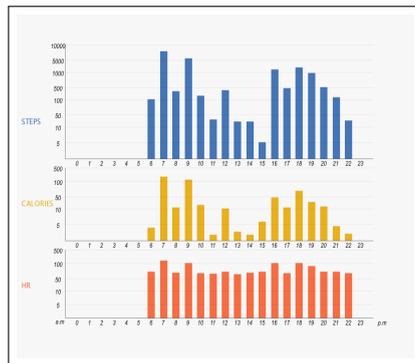
The Mixed Linear-log Mapping: In this method, the visualization space is divided in the same way for the logarithmic scale (e.g., 5, 10, 50, 100 etc.). However, the way that the values are mapped to these spaces (i.e., data points) is based on a linear mapping and not exponentially. This was proposed to facilitate value estimation based on its location between the adjacent lines.



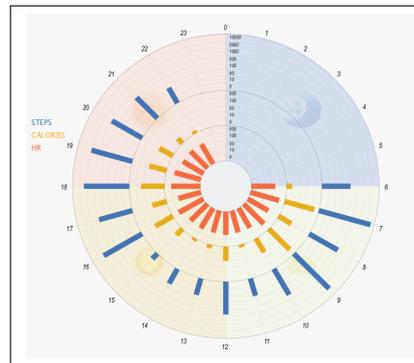
(a) LSLC



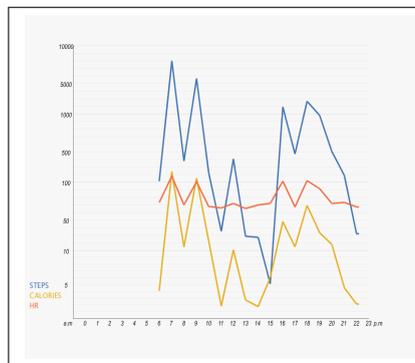
(b) RSLC



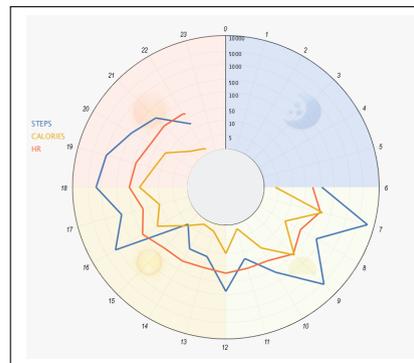
(c) LSBC



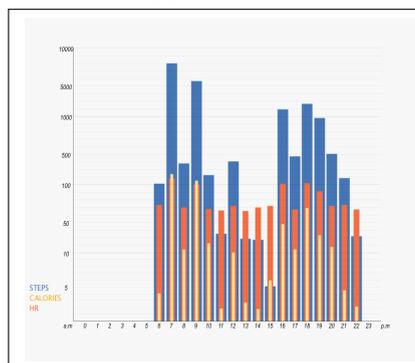
(d) RSBC



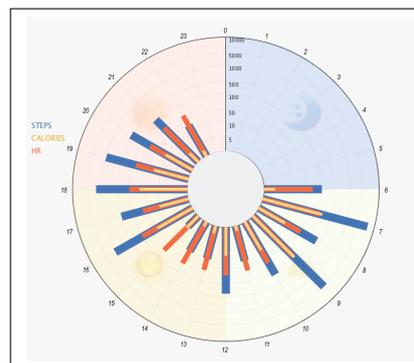
(e) LOLC



(f) ROLC



(g) LOBC



(h) ROBC

Figure 6.2: The Eight Visualization Designs Developed for the Main Study

6.3.2 Visualization Task Taxonomy

The method of task selection is as described in section 5.3.3 since this experiment uses the same task taxonomy. However, some changes were performed on the question/task instruction. The five tasks formulated for the study are presented in Table 6.2. The correct responses for these tasks are shown in Table 6.3. The comparison between the visualizations is performed on the task level to validate the hypotheses in section 6.2.1.

Several criteria need to be considered to ensure the validity of the experiment design when evaluating visualization techniques. For example, Huang et al. [116] considered some criteria to control possible factors that may have an impact on the experiment's results as follows:

1. Tasks should not be too simple to accomplish (i.e., not completed by simple perceptual operations).
2. Tasks should not be too complex to understand and process.
3. Task instructions should be clear, short and not hard to remember (to avoid affecting memory resources required for performing the task).

As there is a need to distinguish between the way the task is performed and the target needs to be fulfilled by the task, as highlighted by Brehmer and Munzner [31], the tasks used in the experiment are described in detail based on their description in the used taxonomy [234] and other related work that discusses similar tasks. As a user's goal may need only one task to be fulfilled and may also require performing other tasks and involve perceptual tasks to accomplish [234], the tasks' descriptions include possible tasks that may be involved. The selected tasks are described as follows:

- **Identify:** The 'Identify' task starts with the goal of finding or visually estimating new information about the data [234]. In the task description in the main study, the participant needs to visually estimate the part of the day that has a higher number of total calories by looking at the distribution of calories along the day. Calories data is already presented (visualized) when the participant starts the task.

According to Munzner's [156] tasks framework, in this task the participant has to find

new knowledge which is not previously known (i.e., discover as a high-level goal). This high-level case requires performing a search as a mid-level goal. Since both the location and the target are unknown, the type of search is to explore. When the target for the search has been found, the goal is to query this target. Explore returns a target of matching characteristics specified in the task question (i.e., has highest total calories). This includes visually exploring the data and identifying the distribution of the calories and then locating the part with the highest distribution.

- **Determine:** The ‘Determine’ task is related to an action of precisely indicating values such as mean or median or any other statistics, like testing a hypothesis [234]. The design of this task in the experiment requires the participant to precisely estimate the value of the step count at a given hour. The visual search involved in this task requires the participants to find the position of the value and estimate the value according to the labels on the axis. The reference is given in this task (i.e., the hour) and the value of the step count needs to be determined. The challenging aspect in this task is estimating the value which is not explicitly presented.
- **Compare:** Comparison is a common task performed by users and visualization can often help in performing this task [94]. It is an analytic task which is performed to compare data items, values and clusters presented in the visualizations and may be preceded by some subtasks (e.g., identify, determine or locate) in order to accomplish the task [234]. The ‘Compare’ task has a scope of multiple targets and is typically harder than the identify task [156]. The difficulty of the comparison tasks grows according to three factors: the number of items that should be compared, the size/complexity of both items and their relationships [94]. In the task design in this experiment, the user needs to compare the calories of two different days to decide which day has the higher number of active calories. Therefore, the number of items being compared are not single items and depend on how many hours have calorie data over the day, which increases the level of the task’s difficulty. The design of the compare task shares some similarities with the ‘Identify’ task, as the distribution of the data needs to be identified. Therefore, like the ‘Identify’ task, the user has to find the trends and visually estimate the total number of calories within a range in the data. In the case of this task, the comparison is between

two days and not within the same day.

- **Infer:** This task is about inferring knowledge from the information presented. The ‘Infer’ task requires the user to infer knowledge from the visualization, such as defining hypotheses and identifying cause and effect relationships and it is performed as part of the data analysis [234]. In this study, it is considered as a seeking relation task. According to Andrienko and Andrienko [12], a seeking relation task is a task that includes a search for occurrences of specific relations between dataset characteristics or references. This task was designed to find a relation between two variables in the dataset (i.e., step count and HR). The target of the task the participant needs to achieve is confirming or disconfirming the hypothesis highlighted in the task question regarding the impact of the step count on the HR. The task involves searching for data items (i.e., hours that have step count values over or equal to a hundred) and then looking for the HR value for these hours to see if the condition has been met.
- **Locate:** Refers to the action of searching and precisely finding values, data items, clusters or properties in the data already visualized [234]. The ‘Locate’ task is one of the search tasks where the user has to find a known target while not knowing its location [156]. When responding to the task question in the experiment, the participant is required to locate the hour with highest step count. This task requires visually searching for the position of the item that has the highest step count value and locating the corresponding hour. The ‘Locate’ task in the this study is similar to the ‘Filter’ task in Amar et al.’s low-level tasks taxonomy [9], they clarified that accomplishing this task relies on the attributes of other data items in the data set in order to find the item that has the highest value.

6.3.2.1 Cognitive Components of the Study Tasks

In this section, the estimated cognitive components required to solve each task are presented. However, the response strategy of the participants when responding to the tasks is not presented here, as related data were not collected (e.g., eye-tracking or subjective responses by the participants to describe how they respond to the tasks). As the results of evaluation of the used

Table 6.2: Tasks/Questions Performed in the Main Study

Task/Question	Answer (Response Type)
T1: (Identify): Which part of the day is the user most active (burned the most calories)?	Multiple Choice. (Morning, Afternoon, Evening, Night)
T2: (Determine): How many steps has the user taken at (6pm /18:00) for the presented day? (estimate the number)	Text field (to enter an estimation for the value).
T3: (Compare): Compare the active calories for the two days. Which day has the higher total active calories?	Multiple Choice. (First day, Second day, Both are the same)
T4: (Infer): (Relationship between Steps and HR) Does step count over 1000 in an hour increase HR to reach 100 and more?	Multiple Choice. (There is no relation, Higher step count has peak HR, Higher step count has resting HR)
T5: (Locate): At what time (at what hour) did the user perform the maximum number of steps during the day	Text field (to enter an exact hour e.g., 9).

Table 6.3: The Correct Responses to the Tasks/Questions Performed in the Main Study.

Task/Question	Dataset 0	Dataset 1
T1: Identify	Afternoon	Morning
T2: Determine (est.)	426	2622
T3: Compare	Second Day	First Day
T4: Infer	There is no relation	Yes, it increases HR to 100 or more
T5: Locate	16:00	07:00

tasks taxonomy by Valiati et al. showed that the participants followed the same expected scenarios and the subtasks with a slight difference in the order, this section presents the expected subtasks, the cognitive operations and the perceptual tasks that are required to achieve the task instructions/questions without discussing their order.

The time and the accuracy of the task may differ according to the visual encoding (i.e., the chart type) and the layout design; for example, people can easily identify the difference between two positions but cannot easily identify the difference between two shades [41]. These factors can also affect the performance of the involved perceptual tasks. An example of these perceptual tasks is the conjunction search, which is essential for completing the tasks in the experiment. Most of the stimuli include a representation of more than one data variable (i.e., step count,

calories and HR). Therefore, the participant needs to search for the variable of interest that has the specified feature in the task question (e.g., the highest step count). The description of each visual task's components is based on the definition of the cognitive and perceptual tasks in section 2.3.1. The task that automatically receives information through visual sensors is perceptual, while the task that requires processing and thinking is cognitive. The cognitive and perceptual aspects and the factors that may affect the results are presented for each task in Tables 6.4 and 6.5. A discussion of these components follows in the next section of this chapter.

Table 6.4: The Cognitive and the Perceptual Components of the Visual Tasks (T1, T2 and T3). The numbering of the perceptual and cognitive tasks are overlapping based on the expected order of the performance.

T1 Identify	
Variable	Calories, clusters (multiple data items, 6 hours)
Perceptual	(1) Conjunctive search (features: yellow Bars/Lines, then visually exploring the distribution to find the part with the highest-level calories (2) Identifying the parts of the day (i.e., visual grouping) (5) Locate the position of the part. (6) Retrieve the value (i.e., the name of the part)
Cognitive	(3) Clustering (finding the data in each segment, and estimating the total calories in each segment) (4) Compare the total (based on the data distribution), then decide which part has the highest value
Factors	Dataset: The distribution of calories within each segment (see section 6.3.3.2) Layout design: Could impact the visual grouping. Vis Method: A possible effect of the overlapping variables on the time required to search for the data item especially in the Bar chart design. Visual Encoding: Could impact the visual grouping to estimate the total in each part
T2 Determine	
Variable	Step Count (Single data item: 1 Data value)
Perceptual	(1) Conjunctive search (features: blue Bars/Lines, then the given attribute i.e., at specific hour). (2) Locating the position of the hour on the time axis. (3) Locating the position of the step count at the located hour (e.g., the tip of the located Bar)
Cognitive	(4) Estimate the value according to the position of the values that falls between two horizontal lines
Factors	Visual Encoding: Could impact upon locating the position. The Bars could be more accurate or faster due to the saliency and the link between the data value (top of the Bar) and the hour value on the time axis. Layout: the angle may affect locating the position and retrieving the value
T3 Compare	
Variable	Calories (Clusters: multiple data items, all day data)
Perceptual	(1) Feature visual search (feature: yellow Bars/Lines). (2) Visually exploring and searching for trends (high distribution) through visually grouping the calories data in each day
Cognitive	(3) Estimate the value of the total calories in each day based on the data distribution. (4) locate the day with the highest calories (5) Retrieve the value (1st or 2nd day)
Factors	Dataset: the total and the distribution of calories in each day could affect the difficulty of the task (see section 6.3.3.2). Visual Encoding: could impact the visual grouping to estimate the total in each day (e.g., the line chart could facilitate grouping as a continuous data)

Table 6.5: The Cognitive and the Perceptual Components of the Visual Tasks (T4 and T5). The numbering of the perceptual and cognitive tasks are overlapping based on the expected order of the performance.

T4 Infer	
Variable	Two variables: the step count and HR (multiple data items)
Perceptual	(1) Conjunctive visual search (features: blue Bars/Lines, and value of any hour over a thousand steps). Finding the hours that have step count over a thousand do not require cognitive processing as the value explicitly encoded on the axis. (2) When the value has been found and the position of the hour has been located, locate the value of HR at that hour. (3) Retrieve the value of the HR
Cognitive	(4) Check if the value of the HR is equal to or over a hundred. (5) Check the hypothesis in the task question to see if the goal has been met or not. The steps from 1 to 5 will be repeated
Factors	Dataset: the number of the step count value that is over a thousand (see section 6.3.3.2). Visual Encoding: could impact locating the position. The bars could be more accurate or faster due to the saliency and the link between the data value (top of the bar) and the hour value on the time axis. Layout: the position of the value may be affected by the angle and its distance from the axis when reading the values. Vis method: in the Overlapped design, the participant can check the two variables on the same chart area, unlike the stacked that represent each variable separately
T5 Locate	
Variable	Step Count (Single data item: 1 Data value)
Perceptual	(1) Conjunctive visual search (features: blue Bars/Lines and has the highest value). (2) In case of the existence of more than one variable that could be the highest, the task of comparing the length in the Bar chart and the position in the Line chart could be involved. (4) Locate and retrieve the value of the hour
Cognitive	(3) Compare (if required)
Factors	The layout: the position of the value may be affected by the angle. Visual Encoding: (Bar or Line) may have an impact on the time or the accuracy for finding the highest value. Vis method: the space in the Overlapped design may facilitate spotting the highest value more than the Stacked design

6.3.3 Dataset Source and Characteristics

Two different datasets were used to create the visualizations. The first is referred to as “dataset 0” and the second as “dataset 1”. The datasets are also encoded with the visualization names that are used in the chapter (see Table 6.1).

Similar to the datasets used to generate the visualizations in Chapter 5, a real dataset that belongs to the author was used; the same export method and the same data types (steps, HR and calories) but different days were used. According to the dataset description by Andrienko and Andrienko [12] presented in Chapter 2, the datasets used to create the visualizations in the study are composed of time data as the referrer (i.e., hours of the day), and the measurements or the values of the step count, HR and calories as attributes.

The exported dataset has been used to randomly generate two different datasets to be used in the visualizations. The exported dataset has been modified by scaling the data points randomly within a specific range. The data of the step count and calories were scaled for each hour by a random number generated between 0.1 and 2.5 and assigned the corresponding HR value, a random number selected from within a possible range based on the newly generated data for the step count and active calories. The data was scaled up and scaled down to ensure a noticeable difference within the visualization channels and to ensure fair comparison across the different visualizations, especially between the Overlapped and non-Overlapped ones.

As the ranges of HR vary according to different factors such as age, gender and activity status, the normal heart rate table was used to keep the heart rate data within the possible maximum and normal values. The maximum heart rate for a person during exercise can be calculated by subtracting the person’s age from 220 [153]. The created random data was for two levels of HR, assuming that the two datasets could belong to different people with different ages. It considered that the normal resting heart rate for adults over the age of 10 years, including older adults, is between 60 and 100 beats per minute (bpm) according to the American Heart Association (AHA) [220]. In the following section, the main differences between the two datasets overall and based on the correct responses to the tasks are discussed.

6.3.3.1 Overall Differences Between the Datasets

All visualizations with both datasets are presented in Tables 6.7 to 6.14. The activity distribution (i.e., the hours that have vital sign recorded during the day) is different in the two datasets; dataset 0 has activity hours between 10:00 - 22:00, while dataset 1 has activity hours between 06:00 - 22:00. They also differ in the step count and the total burned calories as shown in Table 6.6.

In addition, the distribution of high activity levels (for both step count and calories) is clearly different in the two datasets when looking at the visualizations (in the Tables 6.7 to 6.14). In dataset 0, the higher activity data falls in the middle of the Afternoon area, while in dataset 1 it is distributed between the Morning area and Evening area and there is a slope in the data curve in the Afternoon area.

This difference shows that the two datasets have a different level of difficulty that may have influenced the participants' performance when completing those tasks that require looking for trends and patterns within the presented day data: the 'Identify' and the 'Compare' tasks. These two tasks have a similar response strategy, as the participants visually estimate clusters within a specified range in the visualized dataset.

Table 6.6: Main Evaluation Study Datasets Characteristics. This table presents basic information about the two datasets used with the visualizations in the experiment. It starts with the activity period, which refers to the first and last hours the activity begins and ends during the day, then the total step count for the day, the total active calories, the average HR of the day and then the number of peaks (i.e., >1000) in the step count over the day.

	Activity Period	Step Count	Calories	Avg HR	Peaks
Dataset 0	10:00 - 22:00	7426	217	84	2
Dataset 1	06:00 - 22:00	21540	641	84	4

Table 6.7: Images Used for the LSBC Visualizations for Dataset 0 and Dataset 1

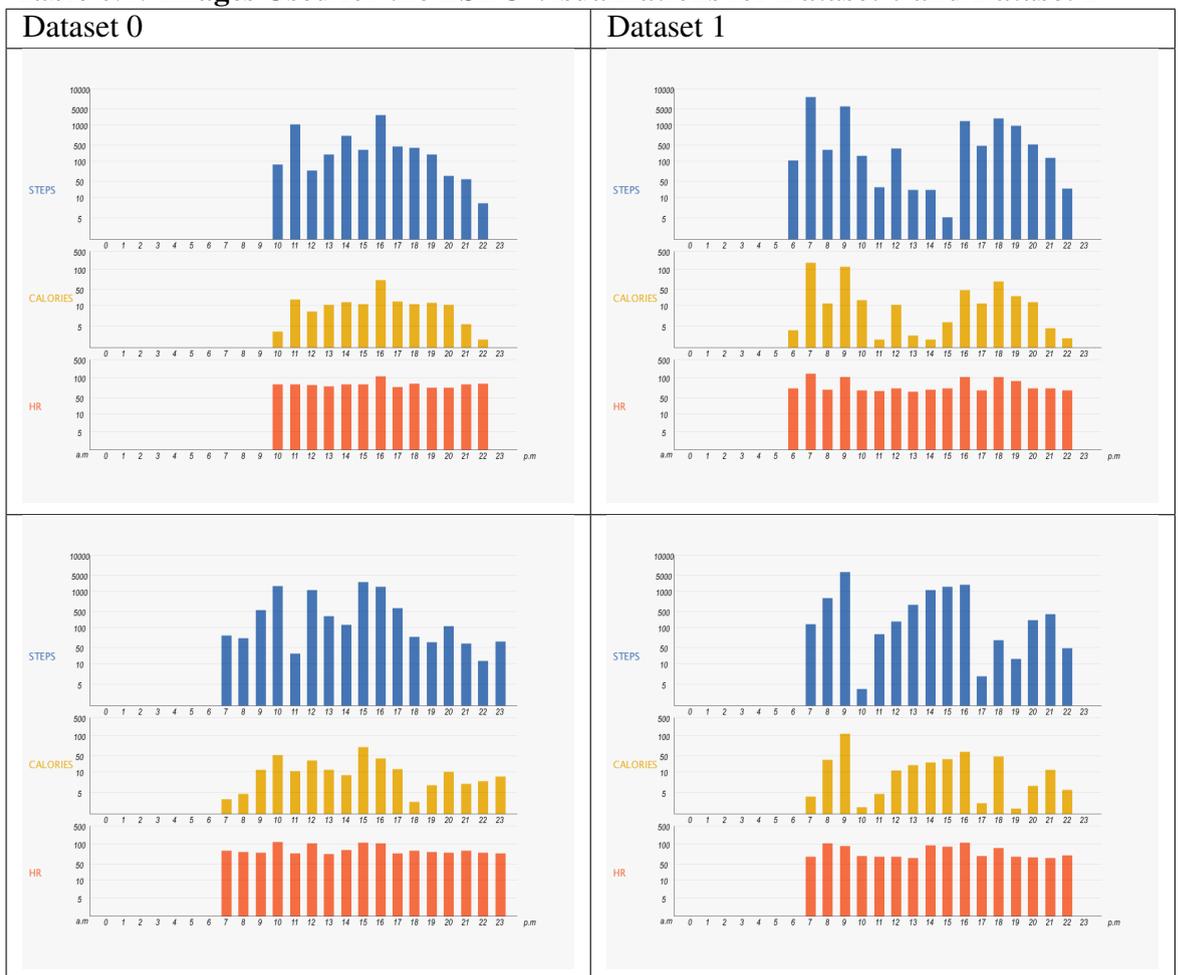


Table 6.8: Images Used for the LOBC Visualizations for Dataset 0 and Dataset 1

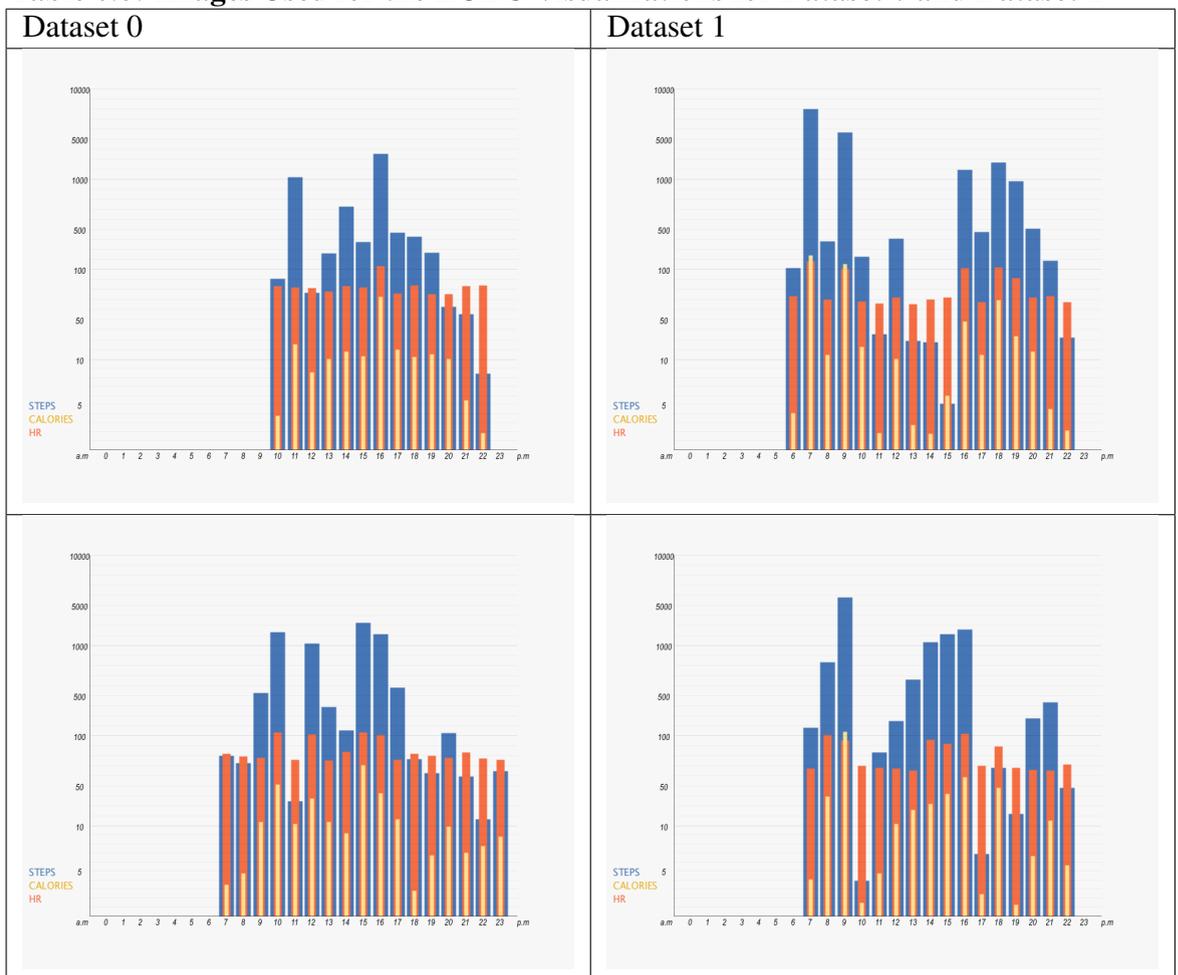


Table 6.9: Images Used for the LSLC Visualizations for Dataset 0 and Dataset 1

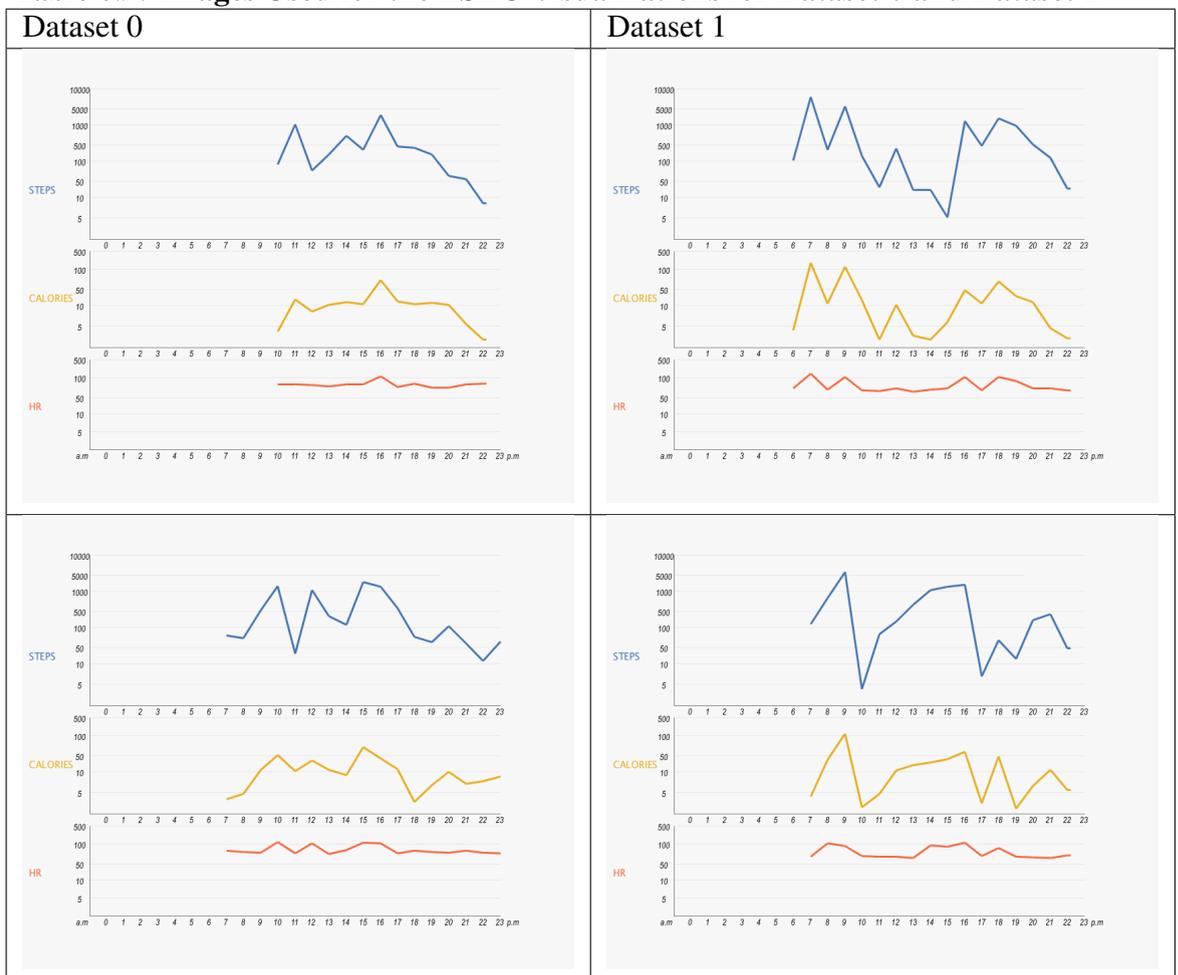


Table 6.10: Images Used for the LOLC Visualizations for Dataset 0 and Dataset 1.

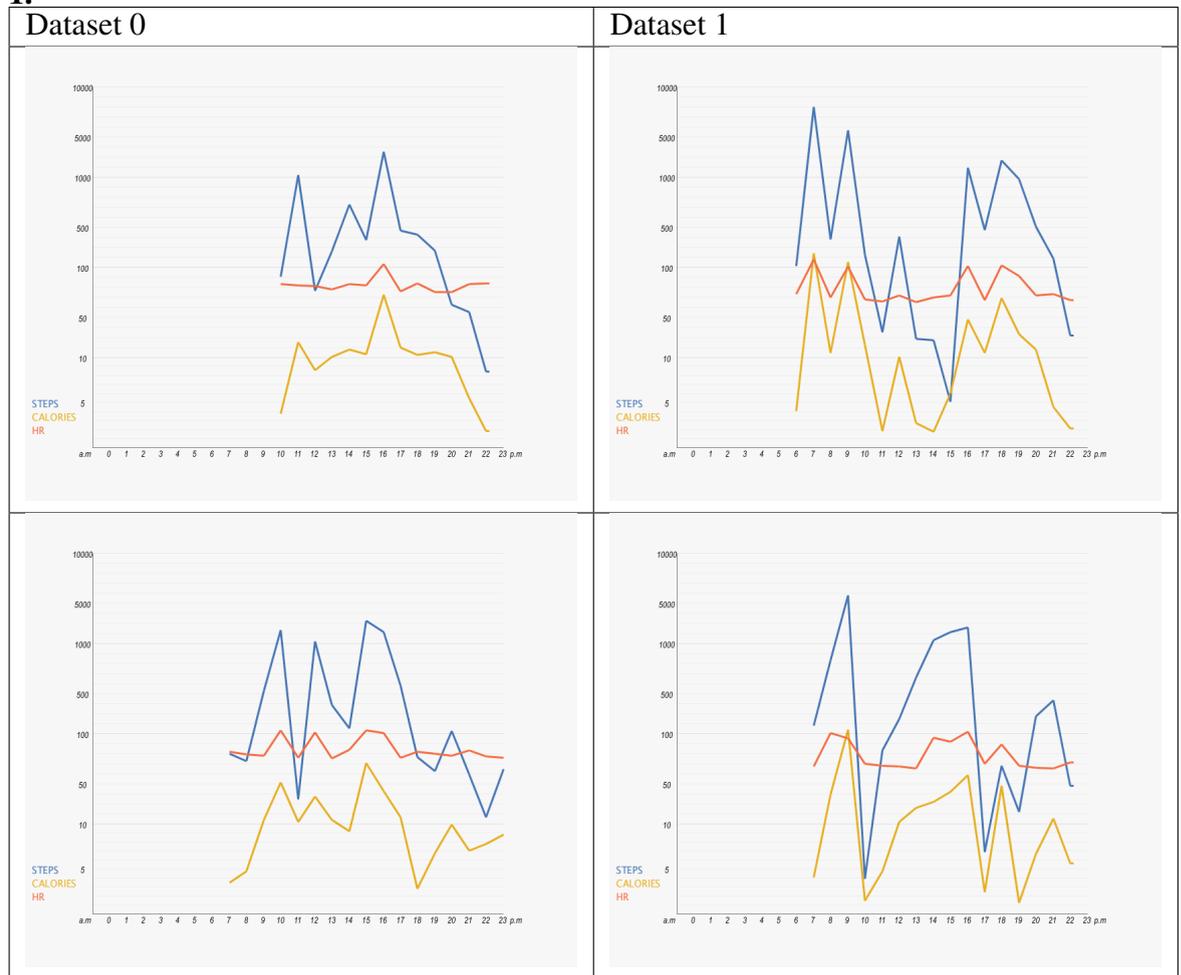


Table 6.11: Images Used for the RSBC Visualizations for Dataset 0 and Dataset 1

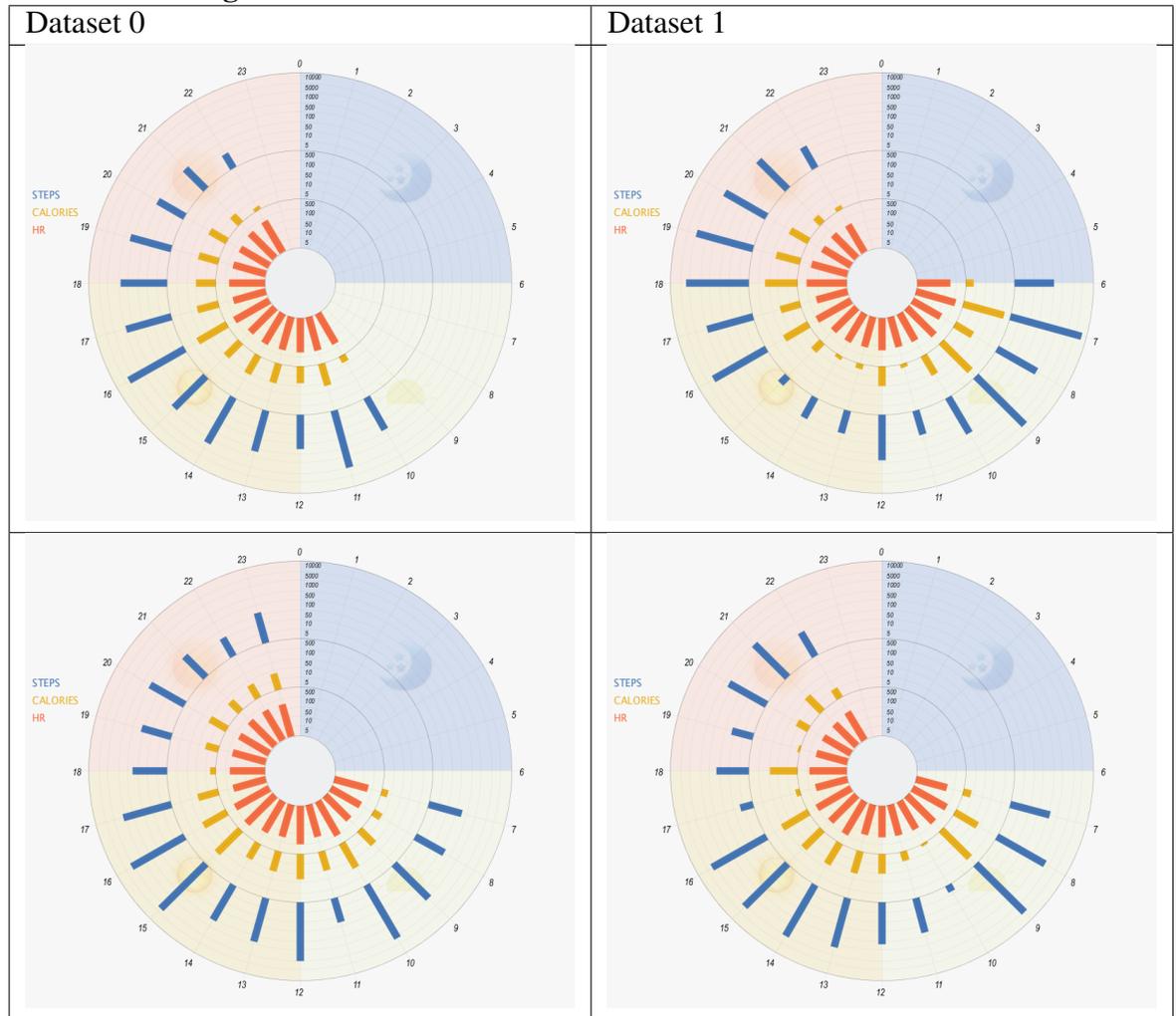


Table 6.12: Images Used for the ROBC Visualizations for Dataset 0 and Dataset 1.

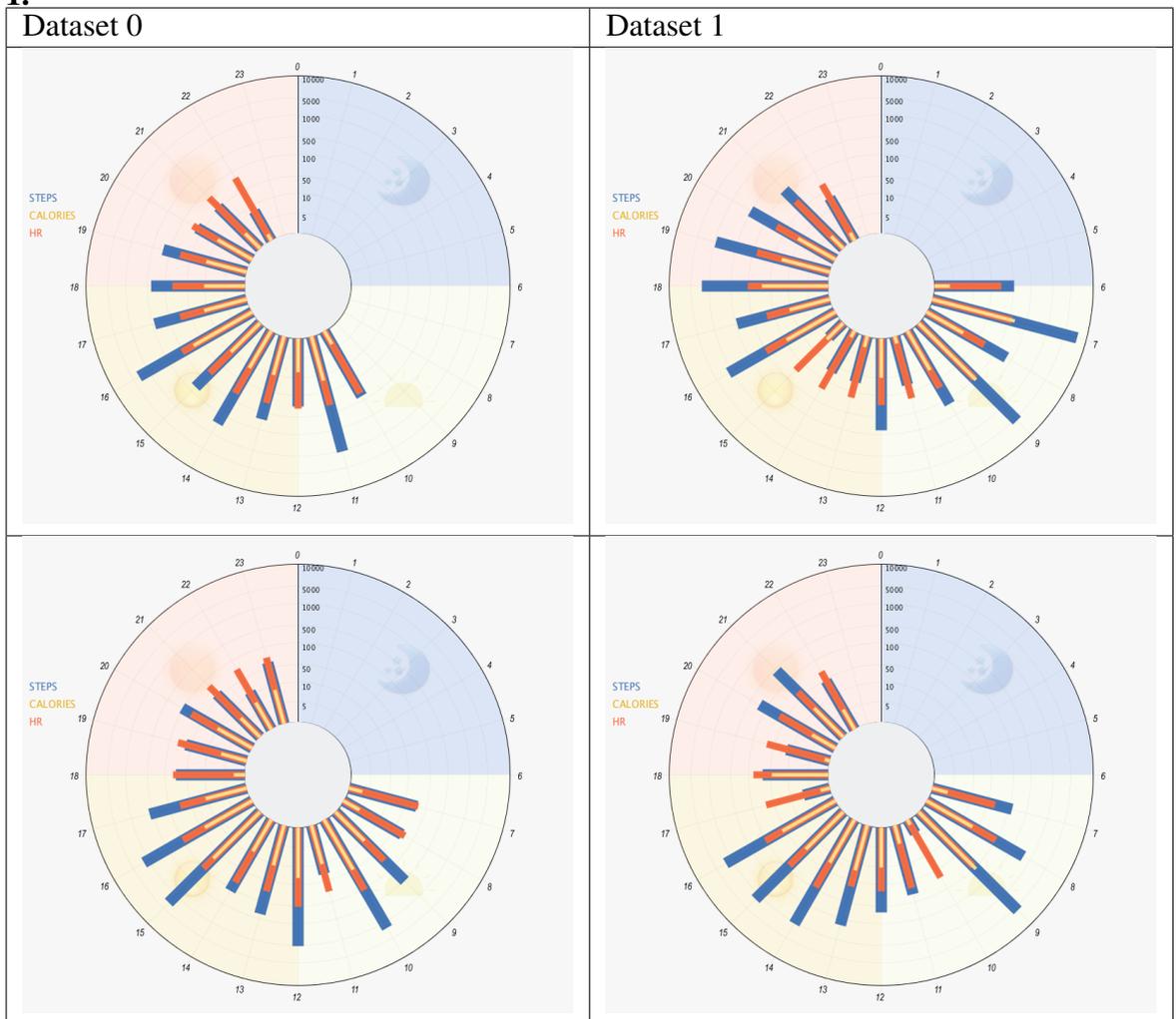


Table 6.13: Images Used for the RSLC Visualizations for Dataset 0 and Dataset 1

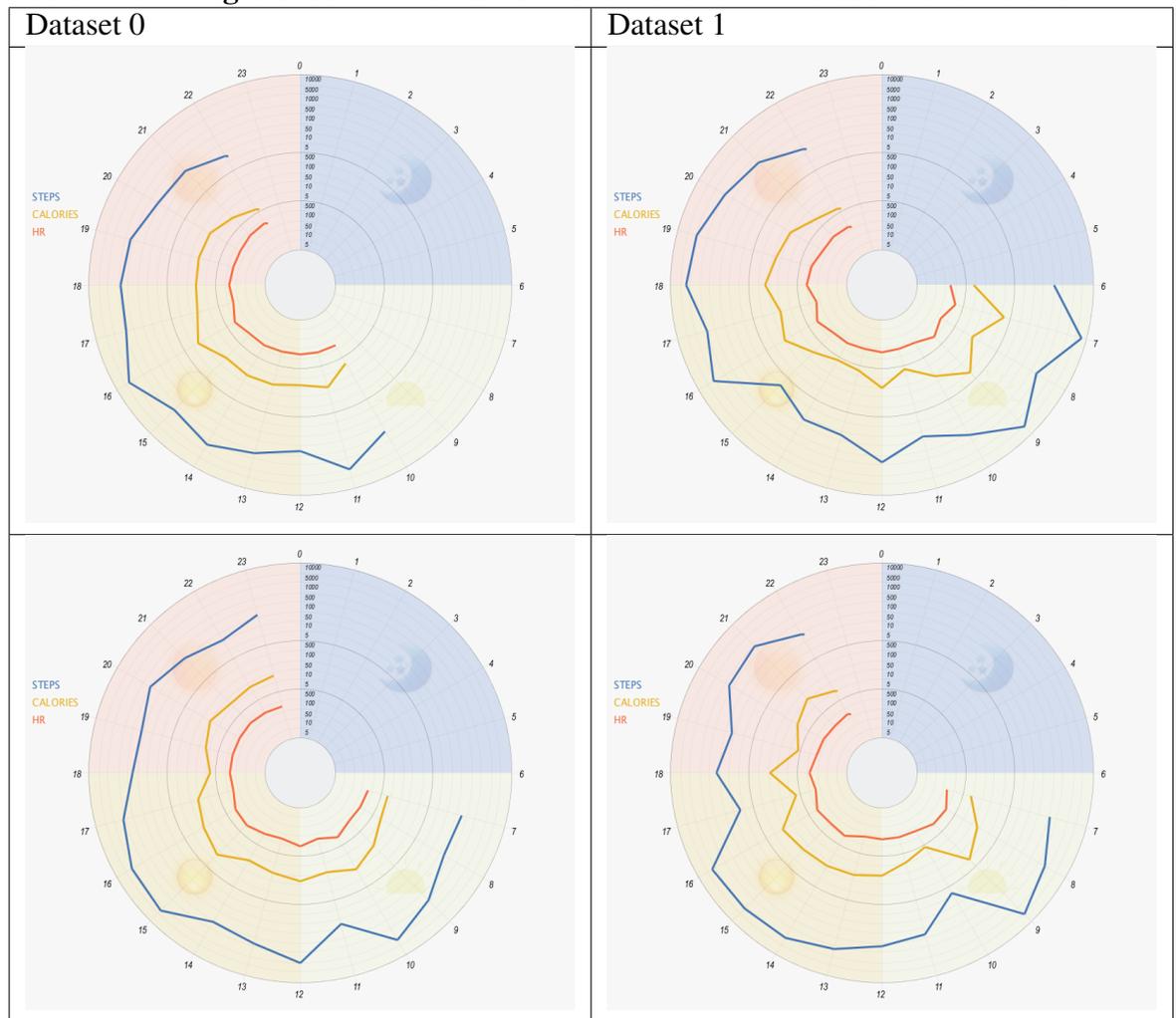
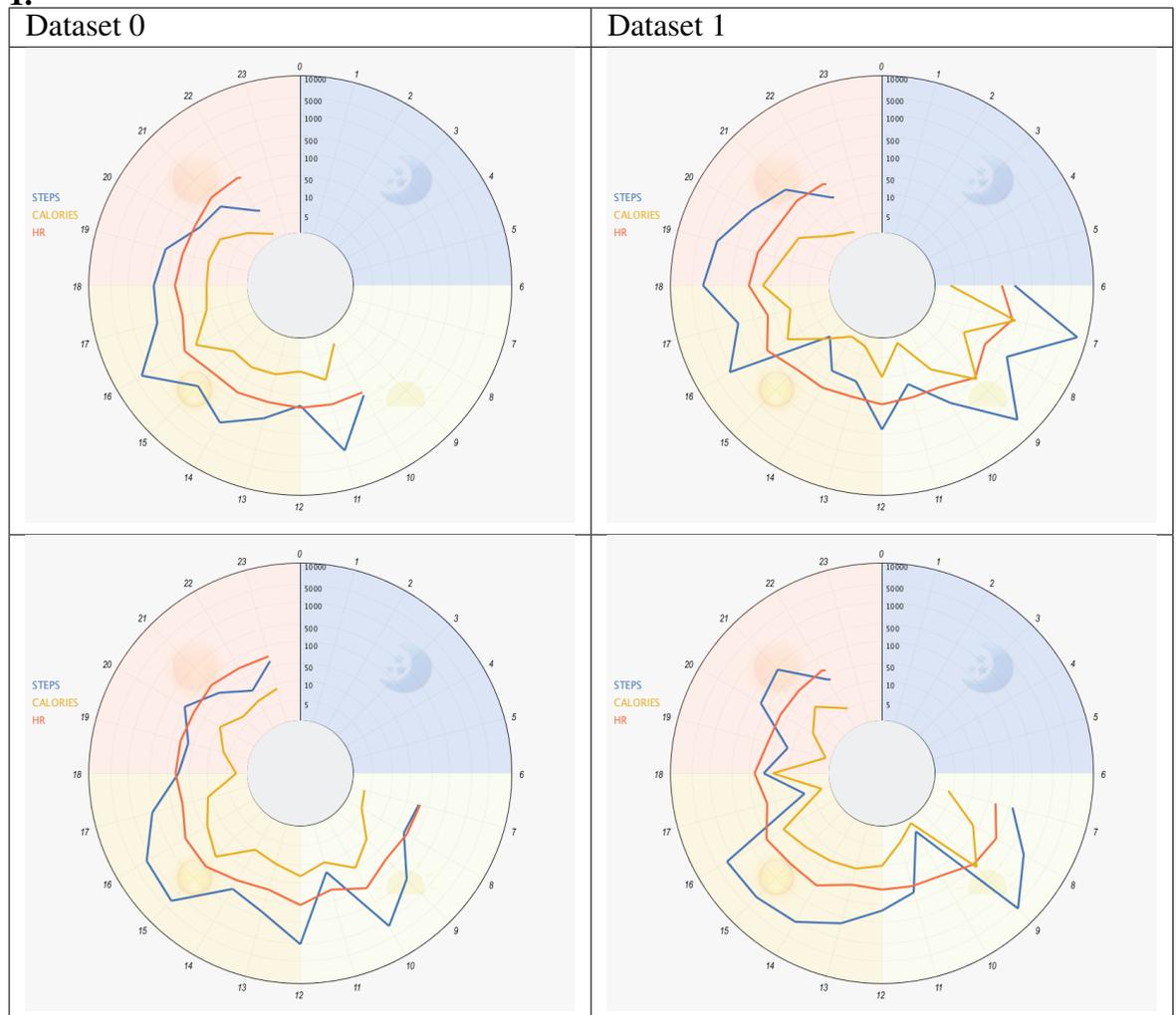


Table 6.14: Images Used for the ROLC Visualizations for Dataset 0 and Dataset 1.



6.3.3.2 Differences Based on the Tasks

The tasks' instructions used in the experiment and responses are presented in Table 6.2. In the following, how the responses to the tasks differ on the two datasets and the strategies for completing each task are explained.

- **T1 (Identify Task):** according to the definition of the task in section 6.3.2, the numbers the user needs to compare are as the following:

Dataset 0: Morning (28), Afternoon (142), Evening (45), Night (0) calories. Dataset 1: Morning (432), Afternoon (82), Evening (125), Night (0) calories.

Despite the difference between the correct response and other responses in dataset 1 being more significant than the difference in dataset 0, dataset 1 could have been more difficult for this task. Dataset 1 is assumed to be more difficult because of the distribution and the length of the activity hours of the data, as discussed earlier. For dataset 0, activity data began at 10:00 and ended at 22:00, while for dataset 1 it started at 06:00 and ended at 22:00, which made it easier for dataset 0 to find the most active part of the day, which is centred in the middle (i.e., Afternoon). Also, the participant can easily eliminate the 'Morning' part from the comparison, because it has only two activity hours, and they do not have high values for the active calories, while for dataset 1 the participant may require more time to compare the activity on the three parts of the day (Morning, Afternoon and Evening).

- **T2 (Determine Task):** the value of step counts for the requested hour for dataset 0 = 426 steps (which falls between the two scales 100 and 500 in the visualizations) and dataset 1 = 2622 steps (which falls between 1000 and 5000 scale points in the visualizations). The difficulty of the estimation task may depend heavily on the visualization designs rather than the datasets. Estimating on the overlapped visualization could be more accurate than the stacked visualization due to the space and the lines between the scales in the overlapped charts. In addition to the effect of the overlapped and the stacked designs, the visualization Layout could also influence the performance of this task. Estimating the value on the Linear layout requires the user to look horizontally from the point that needs to be determined towards the Y-axis that shows the numbers to estimate the correct

response. While on the Radial layout the user needs to follow the circular lines from the point towards the Y-axis that is located in the centre, which could make the task more difficult to complete on the Radial layout.

- **T3 (Compare Task):** the difference between the total active calories between the two days on the two datasets is: Dataset 0: Day 1 (217), Day 2 (315) calories. Dataset 1: Day 1 (640), Day 2 (434) calories. Although the difference between the two presented days is bigger on Dataset 1 than the difference between the two days on dataset 0, the task could have been easier on dataset 0, this is because of the distribution of the active hours during the day. For dataset 0, the used two days data have a different distribution of active hours, which could have made it easier for the participants to find that the 'Second Day' had more burned calories than the 'First Day'. As the first day had active hours between 10:00 and 22:00 while the activity on the second day was between 07:00 and 23:00 and it had more peaks than the second day.
- **T4 (Infer Task):** as presented in Table 6.6, dataset 0 has only two hours with a step count equal to or greater than a thousand steps, while dataset 1 has four hours that have a step count which is equal to or greater than a thousand steps. This difference in the number of peaks could have led to a variance in the difficulty in the datasets when performing the task.
- **T5 (Locate Task):** for this task, dataset 0 and dataset 1 are different in the number of peaks and the step count for each hour as shown earlier. When looking at the visualizations with both datasets, there are two hours that have a high step count and are close in values (height in the graphs). The user has to visually compare between these hours to decide which one has the maximum step count to complete the task. The two hours in dataset 0, are 11:00 and 16:00 which have the values 1140 and 3468 steps respectively, while in Dataset 1 the two hours are 07:00 and 09:00 which shows 8000 and 5650 steps. Despite the fact that two hours in both datasets fall in the same scale-space (i.e., 1000 - 5000), the difference between the two hours values is more significant in dataset 1 which may have helped the participant in estimating the correct response faster.

From the previous discussion, it can be seen that some of the tasks have some similarity in terms

of the response strategy. For example, ‘Identify’ and ‘Compare’ tasks are similar in terms of identifying trends and visually estimating the total calories within ranges (in a day or parts of the day) in the visualized dataset. There is also a similarity between the ‘Determine’ and ‘Locate’ tasks as they both require a visual search for completing the tasks. As regards to the ‘Infer’ task, it also requires the participants to search for peaks in the step count as a part of the process for solving the task and, thus, it shares similar responding behaviour with the ‘Locate’ task.

6.3.4 Experiment Variables

This study follows the same methodology of the study in the previous chapter, using the same experiment variables.

Performance Measures: The metrics of the task accuracy and task completion time were recorded during the experiment. These metrics were used to evaluate the effectiveness of the developed visualizations for the selected tasks.

Subjective Feedback: As a part of the experiment, participants’ subjective feedback regarding the visualization designs was collected before and after conducting the experiment. The pre- and post-experiment questionnaires were similar to the one used in the preliminary evaluation study (see section 5.3.4), as they all included a Likert-scale to rate the visualization in terms of the readability and attractiveness. However, in this study, open-ended questions were added to allow the participants to justify their visualizations’ preferences. The used questionnaires are attached in the appendices .6 and .7. The data analysis and the results of the subjective feedback are presented in Chapter 7.

6.3.5 Participants Recruitment

80 participants were recruited via email from a Cardiff University mailing list (mainly Computer Science and Engineering students and staff). The study was also advertised on the university’s social network site to find participants from different fields and backgrounds. Participation in the study was restricted to people from Cardiff University (students and staff),

who had no known learning difficulties and had not participated in the preliminary study of this research. Permission to conduct the study was given by the ethics committee of Cardiff University, School of Computer Sciences and Informatics, the ethics application approval code is COMSC/Ethics/2018/028.

6.3.5.1 Demographic Information and Groups Structure

The participants' demographic information is presented in Table 6.15. The analysis and the graphs used to present the demographic information about the participants, both overall and in each group, is based on the count (i.e., the total number of the participants). The Figures (6.3, 6.4 and 6.5) present the overall demographic information of the participants, while the Figures in Appendix .8. present the same information but are distributed based on the groups.

The participants were: 47 females; 31 males and two who preferred not to specify their gender. The 80 participants provided us with their ages, these were multiple age ranges and the distribution was as follows: 48 (60%) were aged 25-34; 19 (24%) were aged 18-24; ten (13%) were aged 35-44; one (1%) was aged 45-54 and two participants (3%) preferred not to reveal their age range. Based on the midpoints of age-ranges categories, the average age is 29 years, and the SD is 6 years. Participants' familiarity with tracked health data and their level of familiarity with different chart types was also included in the questionnaire. The data showed that 52 of the 80 participants (65%) tracked different types of health data (including steps, HR, calories, distance, and sleep), 27 (34%) did not track any health data and one participant preferred not to provide any information. The participants who completed the study received a £10 Amazon voucher for their time.

For the question that asks the participants about the tracking device/app they use, the majority of the participants (44 out of 80 participants, i.e. 55%) reported using iPhone and/or Android apps to keep track of their health data. Those participants to knew the type of the app they had been using were followed up, as this information was thought useful when reflecting on what representation method the participants were already familiar with based on the used app/device. 24 of those 44 participants gave details of their tracking apps. Most of the reported apps are the apps already installed on the phones by the OS, such as: the iOS Health app; Google Fit;

the Samsung Health app and the Huawei Health app. In addition to the apps that are connected to the used wearable devices such as the iOS Activity app that tracks the health data from the Apple watch, the Fitbit app and the Mi-Fit app that connects the Xiaomi trackers with the apps. Table 6.16 shows all the apps the participants had previously been using to track their health and activity data.

Table 6.15: Demographic Information of the Participants

Demographic	Responses
<i>Gender</i>	Female (47), male (31), prefer not to say (2)
<i>Age</i>	18-24 (19), 25-34 (48), 35-44 (10), 45-54 (1), Prefer not to say (2)
<i>Degree</i>	Secondary school or equivalent (13), College degree or equivalent (25), Master's degree or equivalent (35), Doctoral degree (PhD) (5), Other (Chartered Accountant) (1), Prefer not to say (1)
<i>Education</i>	Physical Science and Engineering (e.g., Computer Science, Chemistry and Engineering) (54), Biomedical and Life Science (e.g., Bioscience, Psychology and Healthcare Science) (15), Art, Humanities and Social Science (e.g., Business, Languages, Law and Music) (9), Other (Data Science and analytics) (1), Other (Science communication) (1)
<i>Employment Status</i>	Student (64), Administrative staff (7), Academic staff (3), Researcher staff (3), Other (Professional Services staff) (1), Other (research technician) (1), Prefer not to say (1)
<i>Tracking Health/ Physical Activity</i>	Yes (52), No (27), Prefer Not to Say (1)
<i>Tracked Data</i>	Walking/Running/Cycling distance (33), Flight climbed (5), Active/Burned calories (22), Step count (35), Heart rate (21), sleep (16), Other (3), No (27)
<i>Tracking Devices</i>	Fitbit Tracker (9), Apple Watch (7), Garmin Watch/Tracker (2), Android Watch (4), Jawbone Up (1), iPhone apps (25), Android apps (23), Other (5)

6.3.5.2 Groups

In the evaluation study (Chapter 5), the participants were divided into four groups that present the visualizations in a different order, and an *order effect* on the task completion time was identified as it took the participants a noticeably longer time to complete the tasks on the first

Table 6.16: Tracking apps used by the participants on iOS and Android and the number of participants that use these apps.

Device	N	Device	N
<i>iOS Health app</i>	8	<i>Mi Fit app (Xiaomi)</i>	2
<i>iOS Activity app</i>	1	<i>MyFitnessPal app</i>	1
<i>Samsung Health app</i>	3	<i>stepz app</i>	1
<i>Fitbit app</i>	2	<i>Misfit app</i>	1
<i>Huawei Health app</i>	2	<i>LifeLog app</i>	1
<i>Google Fit app</i>	2	<i>8Fit app</i>	1
<i>Strava app</i>	2	<i>Keep Trainig</i>	1

presented visualization than the others. There also could have been a possible *carryover effects* as they became more familiar with the tasks over time.

The design of the experiment was counterbalanced with respect to the order of the presented visualizations to control for potential order effects. Counterbalancing design helps in reducing carryover and order effects by controlling the presentation order of conditions across participants [93]. The number of participants in each group was equally balanced to avoid any possible bias in the analysis. For each group, only the type of the first presented visualization was controlled, as presented in Table 6.17; the rest of the visualizations were randomised.

Table 6.17: First Viewed Visualizations for the Eight Groups in the Main Study

Group	First Viewed Visualization
Group 1	Radial Overlapped Bar Chart (ROBC0/ROBC1)
Group 2	Radial Stacked Bar Chart (RSBC0/RSBC1)
Group 3	Radial Overlapped Line Chart (ROLC0/ROLC1)
Group 4	Radial Stacked Line Chart (RSLC0/RSLC1)
Group 5	Linear Overlapped Bar Chart (LOBC0/LOBC1)
Group 6	Linear Stacked Bar Chart (LSBC0/LSBC1)
Group 7	Linear Overlapped Line Chart (LOLC0/LOLC1)
Group 8	Linear Stacked Line Chart (LSLC0/LSLC1)

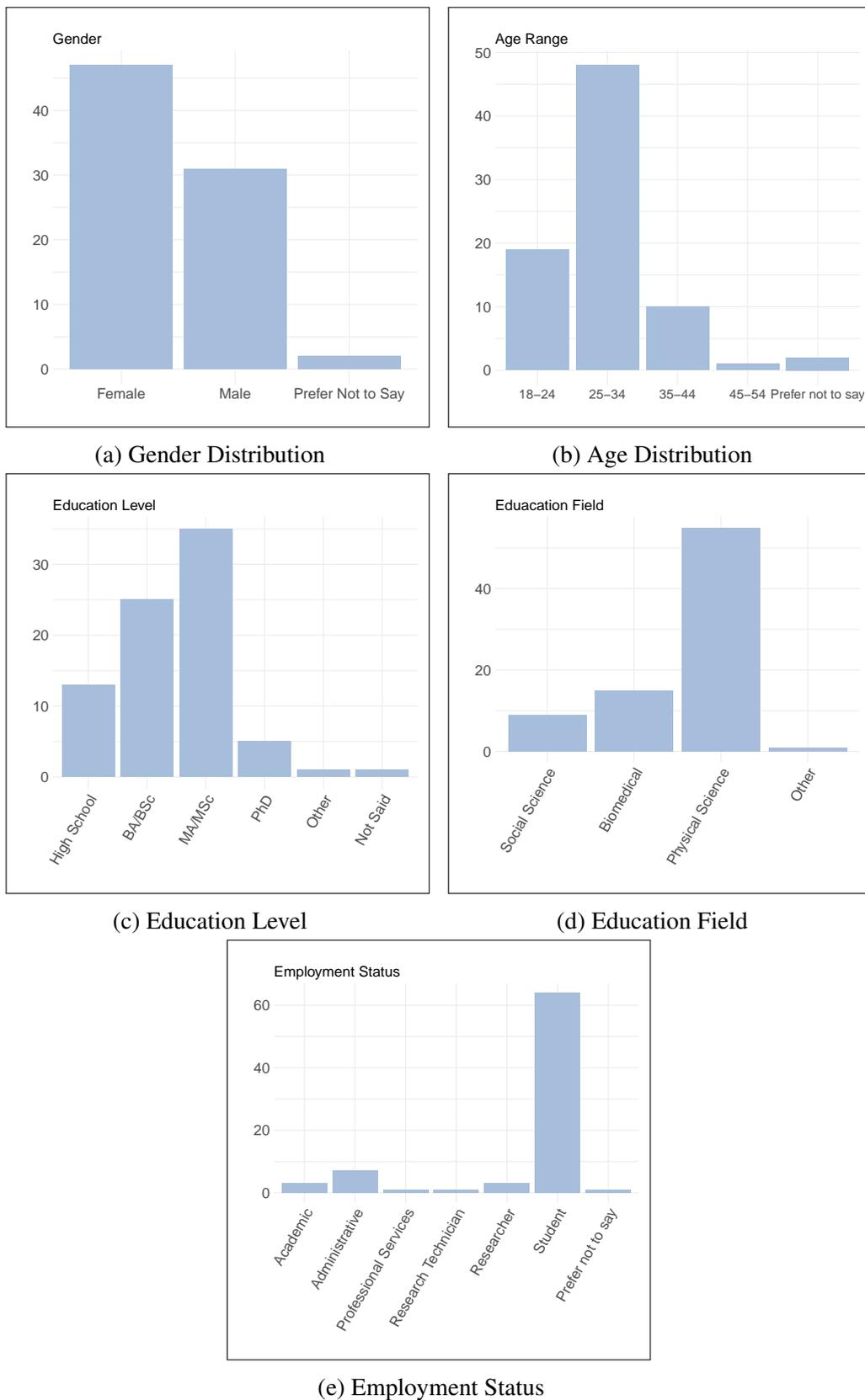
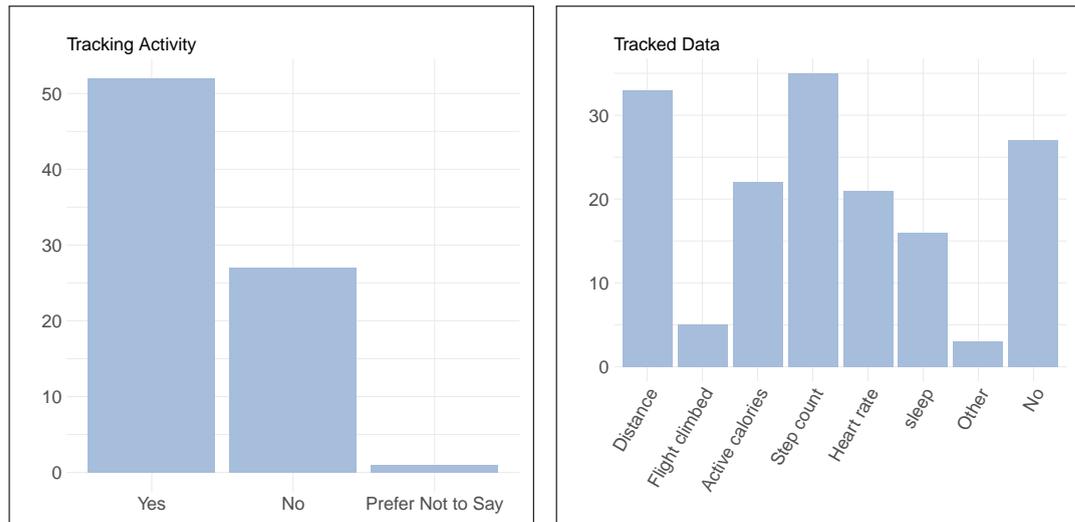
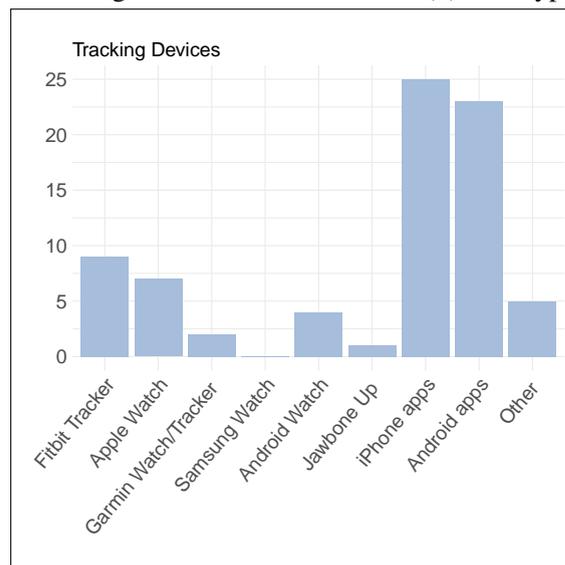


Figure 6.3: Participants' Demographic Information (Age-Range, Gender, Education Degree, Education Field and Employment Status). The X-axis present the categories and the Y-axis shows the number of the participants.



(a) Health Tracking

(b) The Types the Data



(c) The Tracking Devices and Apps

Figure 6.4: Participants' Health Tracking Routine/Behaviour (Tracked Data and Used Devices/Apps). The X-axis present the categories and the Y-axis shows the number of the participants.

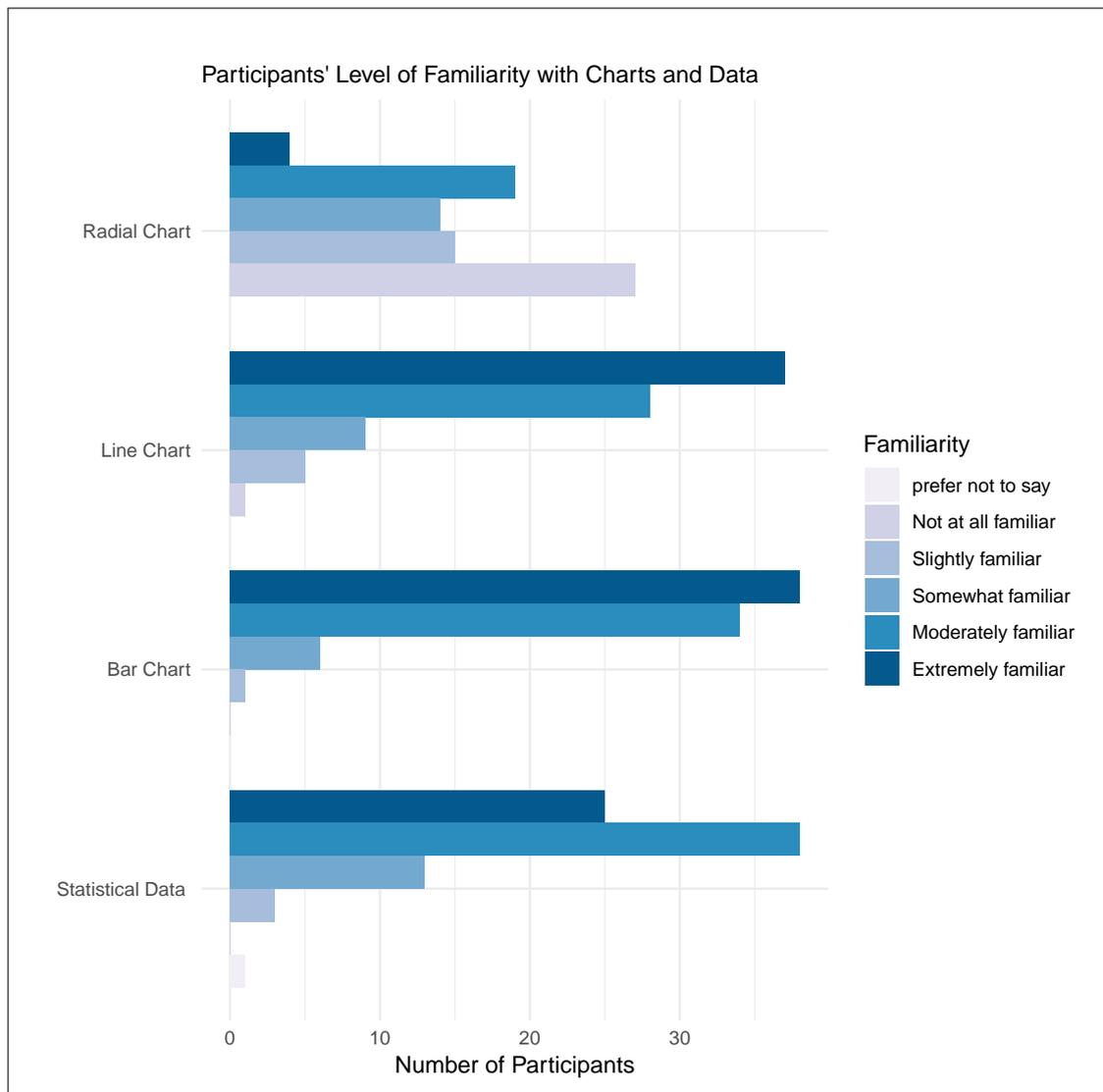


Figure 6.5: The Participants' Responses of their Familiarity with Different Chart and Statistical Data.

6.3.6 Procedure

After recruiting the participants, the participants were randomly assigned into eight groups. The participants in each group conducted the same experiment; the only difference was that the first presented visualization was different for each group. The aim of establishing these groups was to ensure that each visualization design had the chance to be first presented an equal number of times across the participants and to avoid bias in the results. The study required 50 to 60 minutes to complete. Each participant conducted the experiment individually and with the researcher present. In the following section, the stages of the study and the estimated time required to complete each stage are presented.

1. **Introduction and Consent:** The study was introduced to the participants by presenting its aims, procedure, and compensation. Then, before asking for their consent, they were informed of the data collection to be made during the experiment. They were also introduced to the charts to be used in the experiment and the type of data presented (three minutes). The consent form and information sheet used in the experiment are presented in the appendices (see appendix .5).
2. **Pre-Experiment Survey:** The participants were asked to fill in a survey to collect their demographic information (e.g., age, gender, and education) and to estimate their familiarity with different chart types. In the second part of the survey, they were asked to rate the proposed visualizations in terms of readability and the attractiveness of the design, to identify their preferred visualizations and to justify their preferences (five minutes).
3. **Training Session:** In the training session, the participants performed all the tasks on a trial dataset. The correct answers were presented after submitting the answer for each task. They were allowed to repeat the training session if needed (three minutes).
4. **The Main Experiment:** The experiment consisted of five tasks on eight visualizations with two datasets (i.e., $5 \times 8 \times 2$) resulting in 80 stimuli presented and 80 responses for each participant. The order of the visualizations was randomised; only the first presented visualization (with both datasets) was fixed for each group. As the aim was to evaluate the visualizations only with no interaction, the visualizations were presented as static

images (40 minutes). Two screenshots from the experiment software are presented in Figure 6.6.

5. **Post-Experiment Survey:** The participants were asked to rate the visualizations after completing the experiment, choose the preferred visualizations, and provide their feedback (five minutes).

The design of the main experiment, its implementation, participants recruitment and demographics, together with the results are presented in this chapter. Steps (2) and (5) are discussed in detail in Chapter 7.

Experimental Setup:

The experiment was conducted in a study room that was isolated from outside distractions. The experiment and the online surveys were run on a MacBook Pro laptop with a 13-inch display, 2.9 GHz Intel Core i7, 8GB memory, and Intel HD Graphics 4000 with 1536 MB. The experiment application was running in full-screen mode (see Figure 6.6 for application screenshots). The participants interacted with the software using the built-in touchpad and keyboard, no additional devices were required.

Visualization Resolutions: the visualizations for each dataset were generated previously with Processing, then uploaded to the experiment software as static images. These visualizations were exported as ‘TIFF’ file format in a high resolution. The Linear layout visualizations were 820×680 , and the Radial layout visualization images were 780×680 (72 dpi). The produced images were then imported to experiment.

Please Enter Your Participation ID

YOUR ID:

START

0 of 95

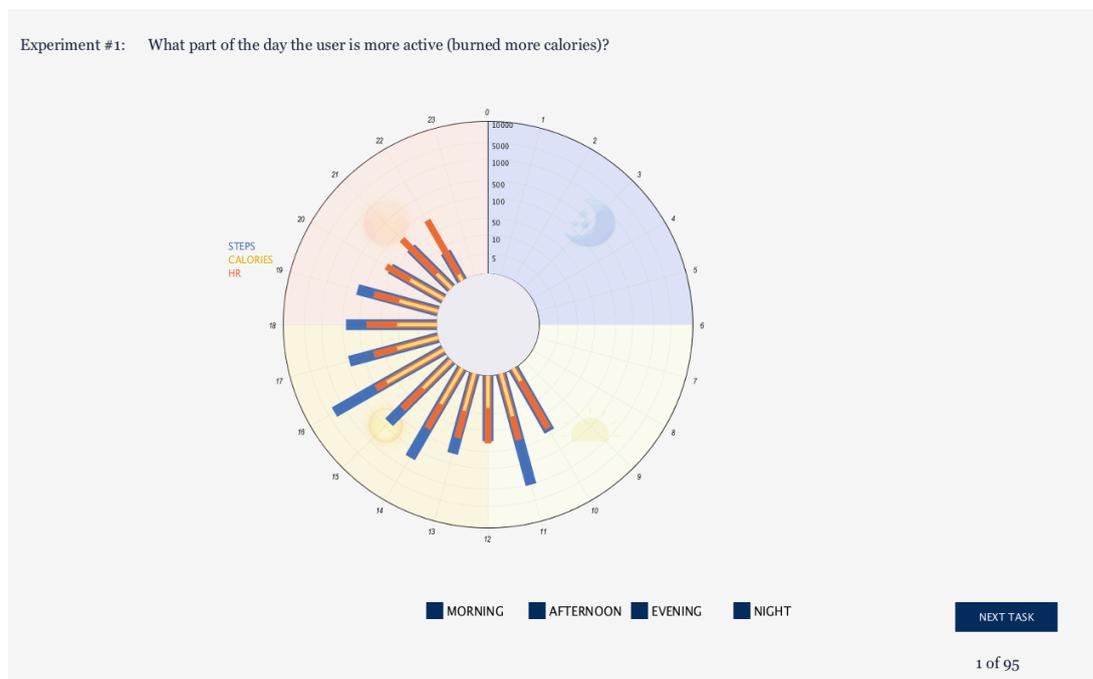


Figure 6.6: Two Screenshots of the Main Evaluation Study Experiment's Interface.

6.4 Analysis and Results

This section presents the analysis methods and the results of the experiment. It begins with describing the analysis methods applied; then it continues with showing how the task completion time and accuracy were calculated, the homogeneity of variance test between the groups, the analysis based on the tasks and finally discusses how the tasks' results differed according to the dataset and how it could have affected the performance of the tasks.

6.4.1 Analysis Method

In the experiment, the variable of interest is the type of visualization and how it affects the time and accuracy of the participants when responding to each task. Each visualization design has three main characteristics: the visual layout (i.e., Radial or Linear); the type of the chart (i.e., Line chart or Bar chart) and the method for representing the three variables (i.e., Overlapped or non-Overlapped data representation). The factors in the experiment are the eight visualization techniques that have two levels (i.e., the two datasets used in the visualizations) and the independent variables are time and accuracy.

The analysis of the data is performed based on the task and the dataset level. It began with an exploratory analysis of the mean and the standard errors of task completion time and accuracy, and then statistical tests were performed to examine the hypotheses. The analysis reports on 0.05 significance level and 95% confidence interval.

Before deciding on the statistical test that would be used, the distribution of the task completion time data for each task was examined to check for normality using the Shapiro-Wilk test. The test showed that not all the data follow a normal distribution even after applying log-transformation. The data of the task accuracy for four of the experiment's tasks are binomial data as they were measured on a pass/fail scale, only the 'Determine' task has a response accuracy scale as a number between 0 and 1. Also, the responses of the 'Determine' task do not confirm the normality test. According to Robertson and Kaptein [183], task completion time and error rate are common non-normal data distributions in HCI studies. Consequently, it was decided that a nonparametric analysis should be applied, considering a pairwise comparison

between the visualizations and designs. The analysis in the following sections was performed on the data collected from participants from all groups. The groups analysis of task times and accuracy and the participants' behaviour during the experiment are in Appendix .9.

Hypothesis Testing: the analysis is performed on the task level and then the datasets level (all groups). Pairwise comparisons between the visualizations based on task completion time and accuracy are performed. The pairwise comparisons are implemented on two levels: (1) across layout (Radial vs Linear), and (2) within each layout. This implemented analysis compared the participants' performance on all the visualizations, which allowed comparing the implemented visual elements:

1. The *visual layout*, which is a comparison between Radial with Linear designs.
2. The *visual encoding*, which is a comparison between the Bar and Line encoding within both layouts.
3. The *visualization method*, which is related to the **multivariate** aspect of the data through comparing the Overlapped with the Stacked design within each layout.

These three aspects are linked in the hypotheses with each task as presented in section 6.2.1.

For the task completion time, the Wilcoxon Signed-Rank nonparametric test was applied, which is suitable when participants try each of two alternatives [183] (in this study they are comparing the paired visualization designs). The study design met the test assumptions: the study data have a continuous dependent variable (i.e., time), and the independent variable is categorical with two related groups (e.g., Linear and Radial). The third assumption is that the difference between the time values on both layouts is in a symmetrical shape and this is examined when performing each analysis test. For comparing the accuracy of a task on two alternatives (e.g., Radial and Linear layout) when the accuracy data are of binomial type data (either 0 or 1), McNemar's nonparametric test was applied and according to [215], McNemar's test has been used in this case. The basic assumptions of McNemar's test meet with the accuracy data of this study as it has "one dichotomous dependent variable with two mutually exclusive groups" which is either 1 for the correct response and 0 otherwise, has "one independent variable that consists of two categorical or matched pairs" which is (the layout Radial or Linear, chart type

Bars or Lines). In the following, the results of this test on the task completion time for each dataset separately are presented.

6.4.2 Calculating Time and Accuracy

Based on the design of the experiment, the time was calculated as the period when visualization and task appear on the screen until the user pressing on the 'Next Task' button to submit the response. The accuracy was calculated according to the type of the task/question.

For the tasks (T1, T3 and T4), the responses were coded as either 1 for the correct response or 0 otherwise based on the selected answer from the provided options as presented in Table 6.2. Task T5, which requires the user to enter a number, was treated similarly to three previous tasks (T1, T3 and T4) as the responses were coded as 1 for the correct answer and 0 for the wrong response, because there is only one acceptable answer for this task. The responses for task T2 were given an accuracy rate between 0 and 1 based on how close the entered answer is to the accepted correct response using the following equation:

$$Accuracy = 1 - \frac{|SelectedResponse - CorrectResponse|}{CorrectResponse}$$

$$Error = \frac{|SelectedResponse - CorrectResponse|}{CorrectResponse}$$

The results of time and accuracy are presented for each task separately in the remainder of this section.

6.4.3 Homogeneity of Variance

To examine the homogeneity of the variance between the groups, Levene's Test for Equality of Variances has been used. Levene's test is used to identify the likelihood of whether the samples are pulled from populations that have equal variance [183]. Levene's test was applied to the task accuracy and the task completion time of each dataset separately to check for homogeneity. The test was applied to the eight different visualizations methods in each dataset to check for the variance across the groups. Table 6.18 shows the results of the test for all tasks.

T1 (Identify): The results of the Levene test declare the equality of the variance between the visualizations for dataset 0 in terms of task completion time ($P=.747$), but fail to declare the equality of variance in the task accuracy ($P=.020$). For dataset 1, the Levene test declares the homogeneity of the variance in the participants' responses for the eight visualizations in terms of the task completion time ($P=.918$) and the task accuracy ($P=.082$).

T2 (Determine): The results of the Levene test declare the equality of the variance between the visualizations for dataset 0 and 1 in terms of the task completion time ($P=.390$ and $.909$ respectively). However, the test fails to declare the equality of variance in the task accuracy for both datasets as the $P<.001$.

T3 (Compare): The results of the Levene test show that the variance between the visualizations for dataset 0 and 1 are equal in terms of the task accuracy and the completion time as the P value for all the tested cases are above $.05$ as presented in Table 6.18.

T4 (Infer): The results of the Levene test declare the equality of the variance between the visualizations for dataset 0 in terms of task completion time ($P=.222$) and the task accuracy ($P=.094$). However, for dataset 1, the Levene test failed to declare the homogeneity of the variance in the participants' responses between the eight visualizations in terms of the accuracy as the result shows a statistically significant difference ($P=.002$). On the other hand, there is an equality in the variance for the visualizations in terms of the task completion time ($P=.109$).

T5 (Locate): For the responses on dataset 0, the Levene test shows that the variance between the visualizations is equal in terms of the task accuracy ($P=.084$) while the test fails to declare the equality of the variance in terms of the task completion time ($P=.032$). On the other hand, the responses on dataset 1 have homogeneity in the variance between the visualizations in terms of the accuracy and the completion time ($P=.406$ and $.254$ respectively).

Table 6.18: Results of the Levene Test for Homogeneity of Variance. P-values of Levene’s statistical test to check the homogeneity of the variance between the responses for the eight visualizations with each dataset. The cells that show a significant difference in the variance are coloured in orange and therefore fail to declare homogeneity.

		T1	T2	T3	T4	T5
Dataset 0	Accuracy	.020	<.001	.530	.094	.084
	Time	.747	.390	.426	.222	.032
Dataset 1	Accuracy	.082	<.001	.054	.002	.406
	Time	.918	.909	.540	.109	.254

6.4.4 Tasks Analysis

The results of the analysis (described in section 6.4.1) are presented in the Tables (6.19-6.28) and Figures (6.7-6.16).

Tables Structure: the tables show the results of the analysis, they present the p-values of the comparisons between the visualization designs in terms of the task completion time and accuracy. The cells that show a significant difference between the paired visualization designs are coloured in **blue** if the left side of the pair is significantly faster or more accurate than the right side of the pair, and in **yellow** otherwise. They are organised as the following:

T1: Identify: the within layout analysis is shown in Table 6.19 and the across layout analysis is shown in Table 6.20. The graphs of means and SE are presented in Figure 6.7 and 6.8.

The results of the analysis show that no significant difference between the visualizations in dataset 0 was identified. There was only one exception, in dataset 0 the task completion time for the LSLC was faster than the LOBC ($P=0.026$) as shown in Table 6.19. Results in dataset 1 showed more differences between the visualizations, especially in terms of the task completion time, due to the distribution of the data. For example, within the Linear layout, the LSLC and LOLC were significantly faster than LSBC (as $P=0.001$) and LOBC (as $P= 0.042$ and 0.012 respectively). However, within the Radial layout, the RSBC was found to be significantly faster than ROBC ($P<0.001$) and faster and more accurate than ROLC ($P= 0.01$ and 0.039 respectively), while the RSLC was significantly faster than ROBC ($P=0.001$). While there was no significant difference identified across the both layouts analysis in dataset 0 (see Table 6.20), the Linear visualizations were generally faster than Radial visualizations. The Linear

Stacked Line visualization (LSLC) is significantly faster than all other Radial visualizations, and only significantly more accurate than the Radial visualizations with Line charts. The Linear Overlapped Line visualization (LOLC) is significantly faster than all other Radial visualizations (except the RSBC) and is only significantly more accurate than the Radial Overlapped Line chart visualization (ROLC) (the results are presented in Table 6.20).

T2: Determine: the within layout analysis is shown in Table 6.21 and the across layout analysis is shown in Table 6.22. The graphs of means and SE are presented in Figure 6.9 and 6.10.

Within the Linear layout, LOBC is significantly more accurate than LSBC, LSLC and LOLC in both datasets. However, regarding the task completion time between the Linear Stacked visualizations (LSBC and LSLC) they are significantly faster than the LOBC although more accurate, while there was no difference in task completion time between the LOLC and LOBC on both datasets. While within the Radial layout, the ROBC is significantly more accurate than RSBC, RSLC and ROLC on both datasets while it is faster than RSLC and RSBC only on dataset 0 (see Table 6.21 for test results). Other significant results were also identified, the ROLC is more accurate than on both datasets ($p < 0.001$). Across both layout, the results generally show that the visualizations with the Linear layout is either significantly more accurate, or faster, or both, as compared with the Radial layout on both datasets (see Table 6.22).

T3: Compare: the within layout analysis is shown in Table 6.23 and the across layout analysis is shown in Table 6.24. The graphs of means and SE are presented in Figure 6.11 and 6.12.

Similar to the results from the 'Identify' task (T1), on dataset 0, the Radial layout was comparable to the Linear layout (i.e., there was no significant difference with respect to performance in either accuracy or time in both analyses). However, on dataset 1, some Linear visualizations were significantly more accurate and/or faster than the Radial counterparts. Results showed Linear Line chart visualizations (LOLC and LSLC) outperformed over Radial visualizations. LOLC was significantly more accurate than all Radial visualizations (except with ROLC where $P = 0.112$), and only significantly faster than ROBC ($P < 0.001$). LSLC was more accurate than all the Radial visualizations (except with ROLC where $P = 0.499$), and significantly faster than ROBC only ($P < 0.001$). Within the Linear layout, LOLC was more accurate than LSBC and LOBC ($P = 0.012$ and < 0.001 respectively) and LOLC was more accurate than LOBC ($P = 0.001$). While within the Radial, ROLC was significantly more accurate than all other Radial design

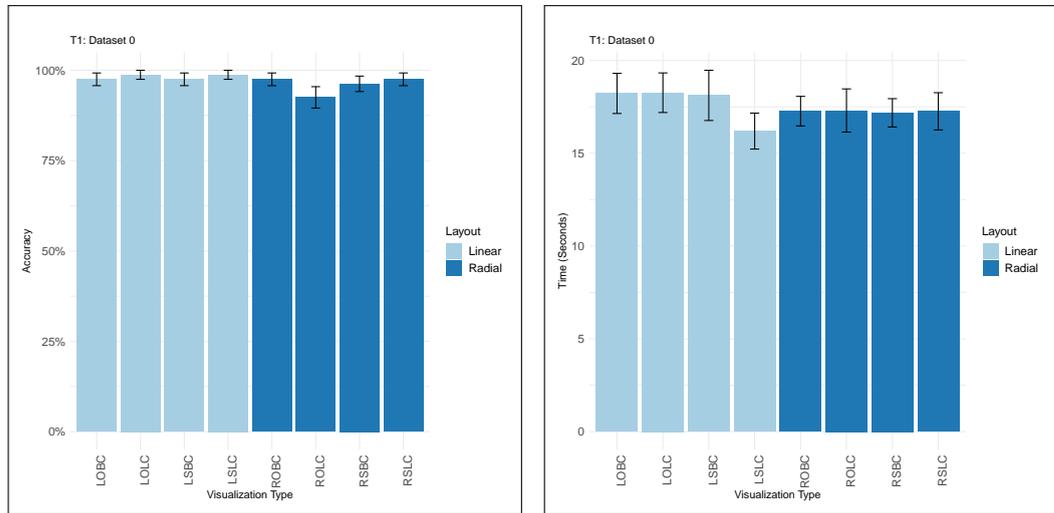
and faster than ROBC, which was also slower than RSBC and RSLC (see Table 6.23 for the P values).

T4: Infer: the within layout analysis is shown in Table 6.25 and the across layout analysis is shown in Table 6.26. The graphs of means and SE are presented in Figure 6.13 and 6.14.

The results of the analysis in Table 6.26 showed that in all the cases, the Linear visualizations were significantly faster on both datasets (one exception was in comparing LSBC with RSBC as $P=0.2$). In terms of accuracy, LOLC and LOBC were more accurate than most Radial visualizations on dataset 0 (one exception in comparing LOLC with ROLC $p=0.189$). On dataset 1, the Linear Stacked visualizations were more accurate than the Radial Overlapped designs ROBC and ROLC. Within both layouts, the results varied according the visualizations as presented in Table 6.25 and discussed in detail in the discussion section.

T5: Locate: the within layout analysis is shown in Table 6.27 and the across layout analysis is shown in Table 6.28. The graphs of means and SE are presented in Figure 6.15 and 6.16.

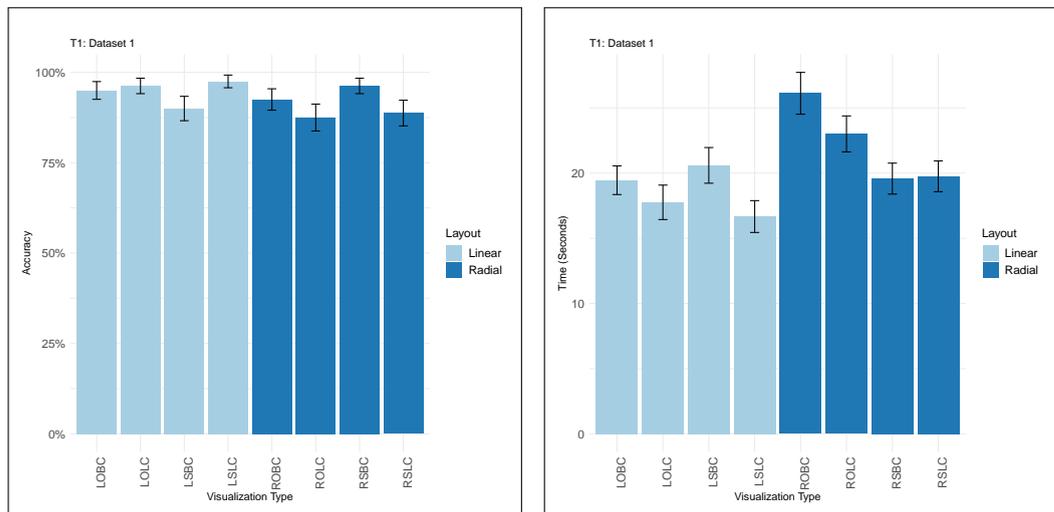
The results in Table 6.26 showed that the difference was more related to the type of the visualization within the Linear layout. The Linear Bar charts LOBC and LSBC were significantly faster than most of the Radial visualizations on both datasets with three exceptions with dataset 0 and one with dataset 1 which exhibited no difference. The results also showed that LOLC was significantly slower than all the Radial visualizations on dataset 0, while no significant difference was identified on dataset 1. No significant differences were identified within the Radial layout on both datasets.



(a) Task Accuracy for 'Identify' Task on the Eight Visualization Designs (Dataset 0). The graph shows the average accuracy for the task and the standard errors.

(b) Completion Time for 'Identify' Task on the Eight Visualization Designs (Dataset 0). The graph shows the means of the completion time and the standard errors.

Figure 6.7: Task Completion Time and Accuracy for 'Identify' Task on the Eight Visualization Designs (Dataset 0).



(a) Task Accuracy for 'Identify' Task on the Eight Visualization Designs (Dataset 1). The graph shows the average accuracy for the task and the standard errors.

(b) Completion Time for 'Identify' Task on the Eight Visualization Designs (Dataset 1). The graph shows the means of the completion time and the standard errors.

Figure 6.8: Task Completion Time and Accuracy for 'Identify' Task on the Eight Visualization Designs (Dataset 1).

Table 6.19: Task T1 (Identify) Analysis Within the Linear and the Radial Layouts for Dataset 0 and 1.

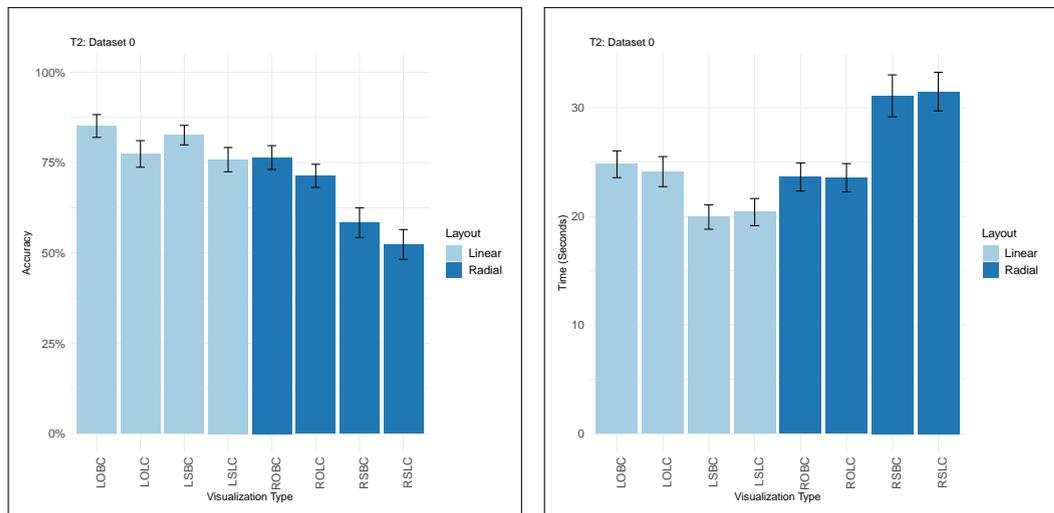
T1: Identify Within Linear layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	LSBC0 VS LOBC0	1.000	0.391
	2	LSBC0 VS LSLC0	1.000	0.342
	3	LSBC0 VS LOLC0	1.000	0.377
	4	LOBC0 VS LSLC0	1.000	0.026
	5	LOBC0 VS LOLC0	1.000	0.705
	6	LSLC0 VS LOLC0	1.000	0.094
T1: Identify Within Linear layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	LSBC1 VS LOBC1	0.344	0.969
	2	LSBC1 VS LSLC1	0.070	0.001
	3	LSBC1 VS LOLC1	0.180	0.042
	4	LOBC1 VS LSLC1	0.500	0.001
	5	LOBC1 VS LOLC1	1.000	0.012
	6	LSLC1 VS LOLC1	1	0.437
T1: Identify Within Radial layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	RSBC0 VS ROBC0	1.000	0.666
	2	RSBC0 VS RSLC0	1.000	0.558
	3	RSBC0 VS ROLC0	0.453	0.375
	4	ROBC0 VS RSLC0	1.000	0.585
	5	ROBC0 VS ROLC0	0.219	0.385
	6	RSLC0 VS ROLC0	0.219	0.656
T1: Identify Within Radial layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	RSBC1 VS ROBC1	0.453	< .001
	2	RSBC1 VS RSLC1	0.109	0.645
	3	RSBC1 VS ROLC1	0.039	0.01
	4	ROBC1 VS RSLC1	0.549	0.001
	5	ROBC1 VS ROLC1	0.344	0.216
	6	RSLC1 VS ROLC1	1.000	0.037

Table 6.20: Task T1 (Identify) Analysis Across Both Layouts for Dataset 0 and 1

T1: Identifyl Across layouts				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	LSBC0 VS RSBC0	1.000	0.262
	2	LSBC0 VS ROBC0	1.000	0.517
	3	LSBC0 VS RSLC0	1.000	0.935
	4	LSBC0 VS ROLC0	0.125	0.825
	5	LOBC0 VS RSBC0	1.000	0.694
	6	LOBC0 VS ROBC0	1.000	0.418
	7	LOBC0 VS RSLC0	1.000	0.305
	8	LOBC0 VS ROLC0	0.219	0.125
	9	LSLC0 VS RSBC0	0.500	0.062
	10	LSLC0 VS ROBC0	1.000	0.202
	11	LSLC0 VS RSLC0	1.000	0.314
	12	LSLC0 VS ROLC0	0.063	0.325
	13	LOLC0 VS RSBC0	0.625	0.863
	14	LOLC0 VS ROBC0	1.000	0.814
	15	LOLC0 VS RSLC0	1.000	0.42
	16	LOLC0 VS ROLC0	0.125	0.585
T1: Identifyl Across layouts				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	LSBC1 VS RSBC1	0.125	0.752
	2	LSBC1 VS ROBC1	0.774	0.004
	3	LSBC1 VS RSLC1	1.000	0.95
	4	LSBC1 VS ROLC1	0.774	0.047
	5	LOBC1 VS RSBC1	1.000	0.870
	6	LOBC1 VS ROBC1	0.688	<.001
	7	LOBC1 VS RSLC1	0.180	0.962
	8	LOBC1 VS ROLC1	0.109	0.008
	9	LSLC1 VS RSBC1	1.000	0.004
	10	LSLC1 VS ROBC1	0.125	<.001
	11	LSLC1 VS RSLC1	0.039	<.001
	12	LSLC1 VS ROLC1	0.008	<.001
	13	LOLC1 VS RSBC1	1.000	0.058
	14	LOLC1 VS ROBC1	0.375	<.001
	15	LOLC1 VS RSLC1	0.109	0.025
	16	LOLC1 VS ROLC1	0.016	<.001

Table 6.21: Task T2 (Determine) Analysis Within the Linear and Radial Layouts for Dataset 0 and 1.

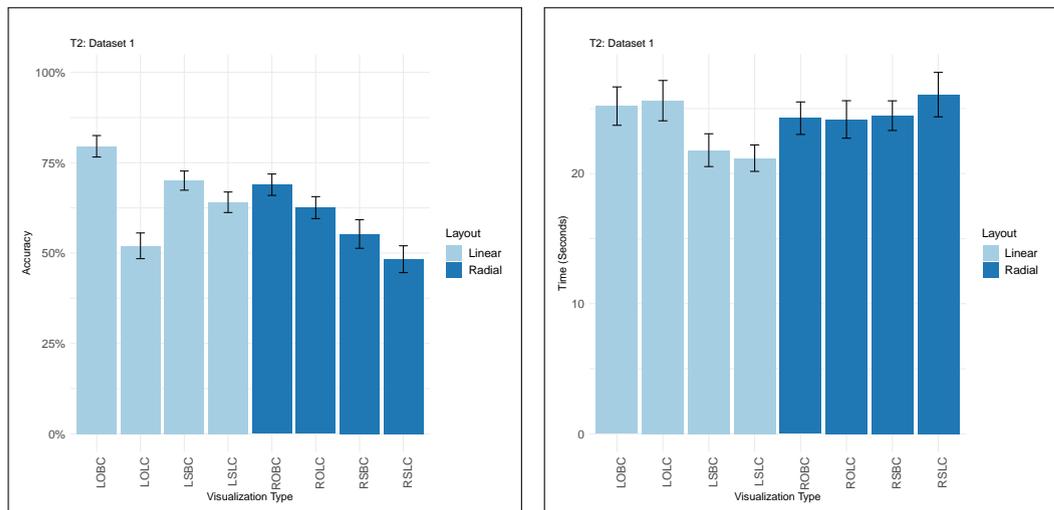
T2: Determine Within Linear layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	LSBC0 VS LOBC0	< .001	< .001
	2	LSBC0 VS LSLC0	0.004	0.542
	3	LSBC0 VS LOLC0	0.557	< .001
	4	LOBC0 VS LSLC0	0.001	< .001
	5	LOBC0 VS LOLC0	0.003	0.287
	6	LSLC0 VS LOLC0	0.053	0.002
T2: Determine Within Linear layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	LSBC1 VS LOBC1	< .001	0.040
	2	LSBC1 VS LSLC1	0.031	0.591
	3	LSBC1 VS LOLC1	< .001	0.018
	4	LOBC1 VS LSLC1	< .001	0.007
	5	LOBC1 VS LOLC1	< .001	1.0
	6	LSLC1 VS LOLC1	0.012	0.008
T2: Determine Within Radial layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	RSBC0 VS ROBC0	< .001	< .001
	2	RSBC0 VS RSLC0	0.29	0.684
	3	RSBC0 VS ROLC0	0.012	< .001
	4	ROBC0 VS RSLC0	< .001	< .001
	5	ROBC0 VS ROLC0	0.005	0.837
	6	RSLC0 VS ROLC0	< .001	< .001
T2: Determine Within Radial layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	RSBC1 VS ROBC1	0.003	0.719
	2	RSBC1 VS RSLC1	0.050	0.774
	3	RSBC1 VS ROLC1	0.102	0.269
	4	ROBC1 VS RSLC1	<.001	0.726
	5	ROBC1 VS ROLC1	0.023	0.588
	6	RSLC1 VS ROLC1	0.001	0.214



(a) Task Accuracy for 'Determine' Task on the Eight Visualization Designs (Dataset 0). The graph shows the average accuracy for the task and the standard errors.

(b) Completion Time for 'Determine' Task on the Eight Visualization Designs (Dataset 0). The graph shows the means of the completion time and the standard errors.

Figure 6.9: Task Completion Time and Accuracy for 'Determine' Task on the Eight Visualization Designs (Dataset 0).



(a) Task Accuracy for 'Determine' Task on the Eight Visualization Designs (Dataset 1). The graph shows the average accuracy for the task and the standard errors.

(b) Completion Time for 'Determine' Task on the Eight Visualization Designs (Dataset 1). The graph shows the means of the completion time and the standard errors.

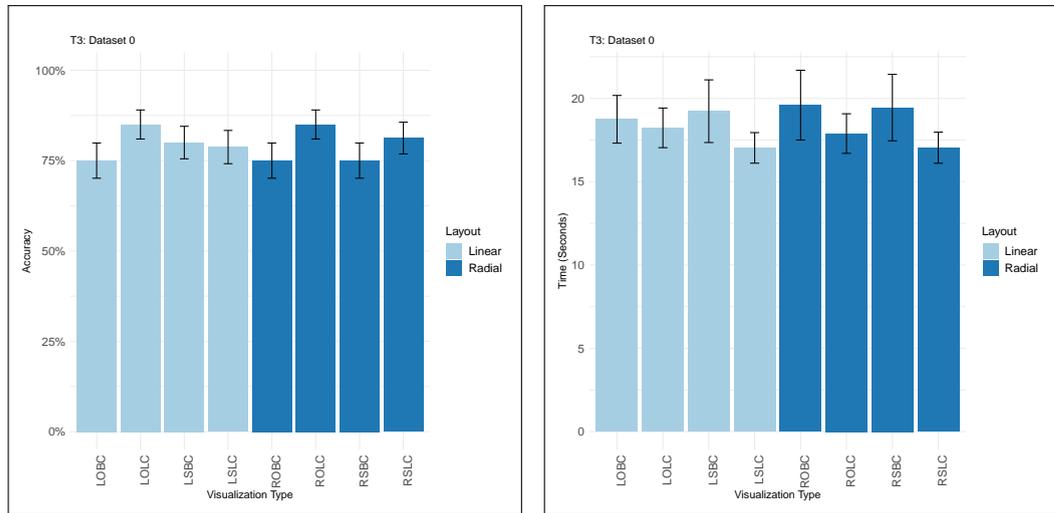
Figure 6.10: Task Completion Time and Accuracy for 'Determine' Task on the Eight Visualization Designs (Dataset 1).

Table 6.22: Task T2 (Determine) Analysis Across Both Layouts for Dataset 0 and 1.

T2: Determinel Across layouts				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	LSBC0 VS RSBC0	< .001	< .001
	2	LSBC0 VS ROBC0	0.098	0.011
	3	LSBC0 VS RSLC0	< .001	< .001
	4	LSBC0 VS ROLC0	< .001	0.002
	5	LOBC0 VS RSBC0	< .001	0.022
	6	LOBC0 VS ROBC0	0.001	0.211
	7	LOBC0 VS RSLC0	< .001	< .001
	8	LOBC0 VS ROLC0	< .001	0.264
	9	LSLC0 VS RSBC0	< .001	< .001
	10	LSLC0 VS ROBC0	0.246	0.029
	11	LSLC0 VS RSLC0	< .001	0.001
	12	LSLC0 VS ROLC0	< .011	0.009
	13	LOLC0 VS RSBC0	< .001	< .001
	14	LOLC0 VS ROBC0	0.138	0.472
	15	LOLC0 VS RSLC0	< .001	< .001
	16	LOLC0 VS ROLC0	< .001	0.924
T2: Determinel Across layouts				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	LSBC1 VS RSBC1	0.002	< .001
	2	LSBC1 VS ROBC1	0.98	0.012
	3	LSBC1 VS RSLC1	<.001	0.005
	4	LSBC1 VS ROLC1	<.009	0.115
	5	LOBC1 VS RSBC1	<.001	0.822
	6	LOBC1 VS ROBC1	< .001	0.788
	7	LOBC1 VS RSLC1	<.001	0.741
	8	LOBC1 VS ROLC1	<.001	0.454
	9	LSLC1 VS RSBC1	0.025	0.005
	10	LSLC1 VS ROBC1	0.203	0.042
	11	LSLC1 VS RSLC1	< .001	0.004
	12	LSLC1 VS ROLC1	0.729	0.142
	13	LOLC1 VS RSBC1	0.482	0.935
	14	LOLC1 VS ROBC1	<.001	0.496
	15	LOLC1 VS RSLC1	0.527	0.748
	16	LOLC1 VS ROLC1	0.02	0.294

Table 6.23: Task T3 (Compare) Analysis Within the Linear and Radial Layouts for Dataset 0 and 1.

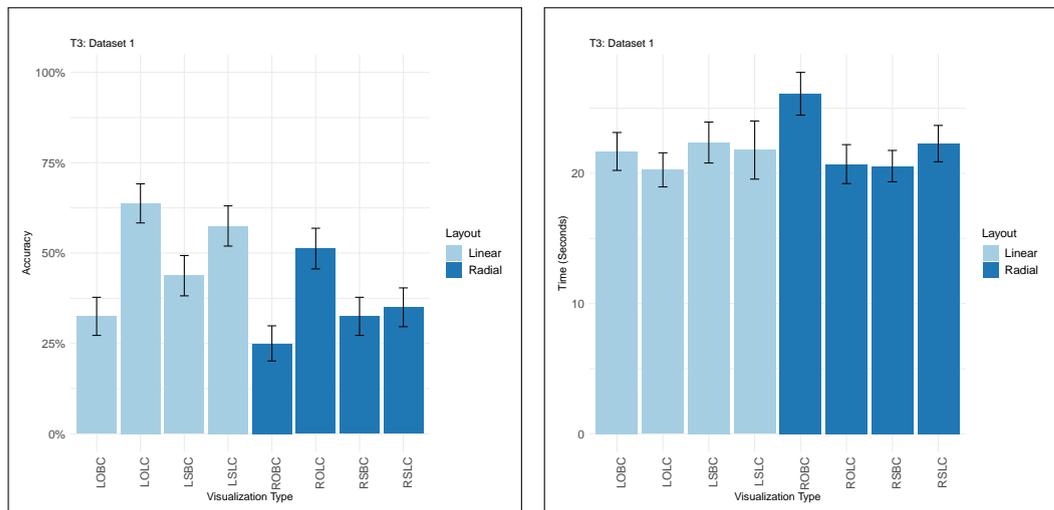
T3: Compare Within Linear layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	LSBC0 VS LOBC0	0.481	0.565
	2	LSBC0 VS LSLC0	1.000	0.774
	3	LSBC0 VS LOLC0	0.503	0.511
	4	LOBC0 VS LSLC0	0.644	0.741
	5	LOBC0 VS LOLC0	0.134	0.818
	6	LSLC0 VS LOLC0	0.267	0.638
T3: Compare Within Linear layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	LSBC1 VS LOBC1	0.164	0.950
	2	LSBC1 VS LSLC1	0.063	0.360
	3	LSBC1 VS LOLC1	0.012	0.272
	4	LOBC1 VS LSLC1	0.001	0.365
	5	LOBC1 VS LOLC1	< .001	0.432
	6	LSLC1 VS LOLC1	0.405	0.939
T3: Compare Within Radial layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	RSBC0 VS ROBC0	1.000	0.649
	2	RSBC0 VS RSLC0	0.424	0.946
	3	RSBC0 VS ROLC0	0.096	0.837
	4	ROBC0 VS RSLC0	0.424	0.555
	5	ROBC0 VS ROLC0	0.134	0.418
	6	RSLC0 VS ROLC0	0.629	0.763
T3: Compare Within Radial layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	RSBC1 VS ROBC1	0.345	0.004
	2	RSBC1 VS RSLC1	0.855	0.38
	3	RSBC1 VS ROLC1	0.029	0.833
	4	ROBC1 VS RSLC1	0.186	0.022
	5	ROBC1 VS ROLC1	0.001	0.001
	6	RSLC1 VS ROLC1	0.031	0.131



(a) Task Accuracy for 'Compare' Task on the Eight Visualization Designs (Dataset 0). The graph shows the average accuracy for the task and the standard errors.

(b) Completion Time for 'Compare' Task on the Eight Visualization Designs (Dataset 0). The graph shows the means of the completion time and the standard errors.

Figure 6.11: Task Completion Time and Accuracy for 'Compare' Task on the Eight Visualization Designs (Dataset 0).



(a) Task Accuracy for 'Compare' Task on the Eight Visualization Designs (Dataset 1). The graph shows the average accuracy for the task and the standard errors.

(b) Completion Time for 'Compare' Task on the Eight Visualization Designs (Dataset 1). The graph shows the means of the completion time and the standard errors.

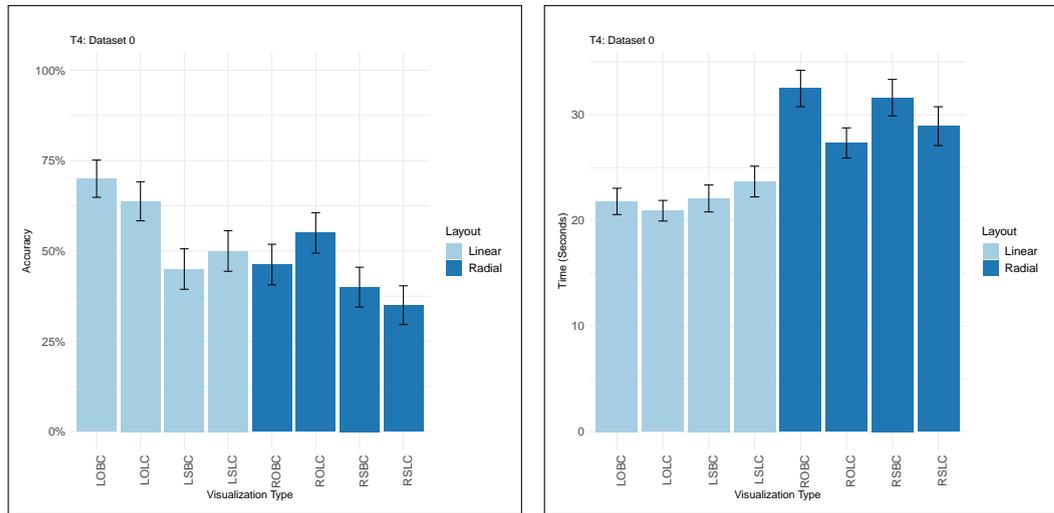
Figure 6.12: Task Completion Time and Accuracy for 'Compare' Task on the Eight Visualization Designs (Dataset 1).

Table 6.24: Task T3 (Compare) Analysis Across Both Layouts for Dataset 0 and 1.

T3: Comparel Across layouts				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	LSBC0 VS RSBC0	0.503	0.893
	2	LSBC0 VS ROBC0	0.541	0.152
	3	LSBC0 VS RSLC0	1.000	0.755
	4	LSBC0 VS ROLC0	0.523	0.642
	5	LOBC0 VS RSBC0	1.000	0.777
	6	LOBC0 VS ROBC0	1.000	0.581
	7	LOBC0 VS RSLC0	0.405	0.785
	8	LOBC0 VS ROLC0	0.115	0.796
	9	LSLC0 VS RSBC0	0.629	0.781
	10	LSLC0 VS ROBC0	0.664	0.591
	11	LSLC0 VS RSLC0	0.815	0.977
	12	LSLC0 VS ROLC0	0.359	0.837
	13	LOLC0 VS RSBC0	0.115	0.542
	14	LOLC0 VS ROBC0	0.152	0.581
	15	LOLC0 VS RSLC0	0.581	0.451
	16	LOLC0 VS ROLC0	1.000	0.751
T3: Comparel Across layouts				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	LSBC1 VS RSBC1	0.137	0.698
	2	LSBC1 VS ROBC1	0.009	0.008
	3	LSBC1 VS RSLC1	0.281	0.578
	4	LSBC1 VS ROLC1	0.405	0.391
	5	LOBC1 VS RSBC1	1.000	0.905
	6	LOBC1 VS ROBC1	0.377	0.005
	7	LOBC1 VS RSLC1	0.864	0.463
	8	LOBC1 VS ROLC1	0.025	0.568
	9	LSLC1 VS RSBC1	0.001	0.511
	10	LSLC1 VS ROBC1	< .001	< .001
	11	LSLC1 VS RSLC1	0.005	0.277
	12	LSLC1 VS ROLC1	0.499	0.810
	13	LOLC1 VS RSBC1	< .001	0.466
	14	LOLC1 VS ROBC1	< .001	< .001
	15	LOLC1 VS RSLC1	< .001	0.615
	16	LOLC1 VS ROLC1	0.112	0.615

Table 6.25: Task T4 (Infer) Analysis Within the Linear and Radial Layouts for Dataset 0 and 1.

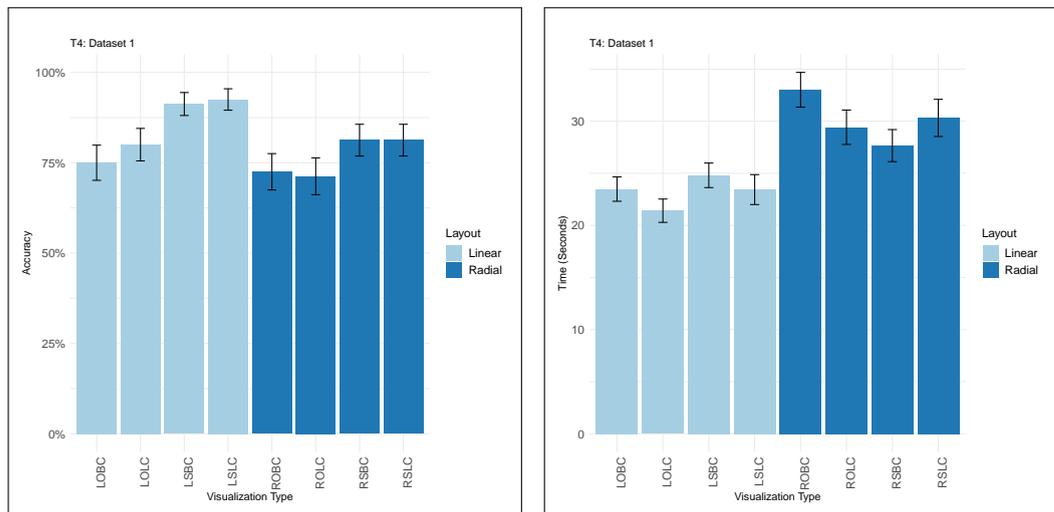
T4: Infer Within Linear layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	LSBC0 VS LOBC0	< .001	0.882
	2	LSBC0 VS LSLC0	0.523	0.283
	3	LSBC0 VS LOLC0	0.001	0.733
	4	LOBC0 VS LSLC0	< .001	0.463
	5	LOBC0 VS LOLC0	0.227	0.601
	6	LSLC0 VS LOLC0	0.019	0.289
T4: Infer Within Linear layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	LSBC1 VS LOBC1	0.011	0.347
	2	LSBC1 VS LSLC1	1.000	0.042
	3	LSBC1 VS LOLC1	0.064	0.041
	4	LOBC1 VS LSLC1	0.004	0.527
	5	LOBC1 VS LOLC1	0.503	0.149
	6	LSLC1 VS LOLC1	0.013	0.517
T4: Infer Within Radial layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	RSBC0 VS ROBC0	0.424	0.323
	2	RSBC0 VS RSLC0	0.584	0.108
	3	RSBC0 VS ROLC0	0.031	0.102
	4	ROBC0 VS RSLC0	0.108	0.011
	5	ROBC0 VS ROLC0	0.118	0.023
	6	RSLC0 VS ROLC0	0.005	0.833
T4: Infer Within Radial layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	RSBC1 VS ROBC1	0.248	0.005
	2	RSBC1 VS RSLC1	1	0.125
	3	RSBC1 VS ROLC1	0.201	0.287
	4	ROBC1 VS RSLC1	0.265	0.048
	5	ROBC1 VS ROLC1	1	0.115
	6	RSLC1 VS ROLC1	0.17	0.844



(a) Task Accuracy for 'Infer' Task on the Eight Visualization Designs (Dataset 0). The graph shows the average accuracy for the task and the standard errors.

(b) Completion Time for 'Infer' Task on the Eight Visualization Designs (Dataset 0). The graph shows the means of the completion time and the standard errors.

Figure 6.13: Task Completion Time and Accuracy for 'Infer' Task on the Eight Visualization Designs (Dataset 0).



(a) Task Accuracy for 'Infer' Task on the Eight Visualization Designs (Dataset 1). The graph shows the average accuracy for the task and the standard errors.

(b) Completion Time for 'Infer' Task on the Eight Visualization Designs (Dataset 1). The graph shows the means of the completion time and the standard errors.

Figure 6.14: Task Completion Time and Accuracy for 'Infer' Task on the Eight Visualization Designs (Dataset 1).

Table 6.26: Task T4 (Infer) Analysis Across Both Layouts for Dataset 0 and 1

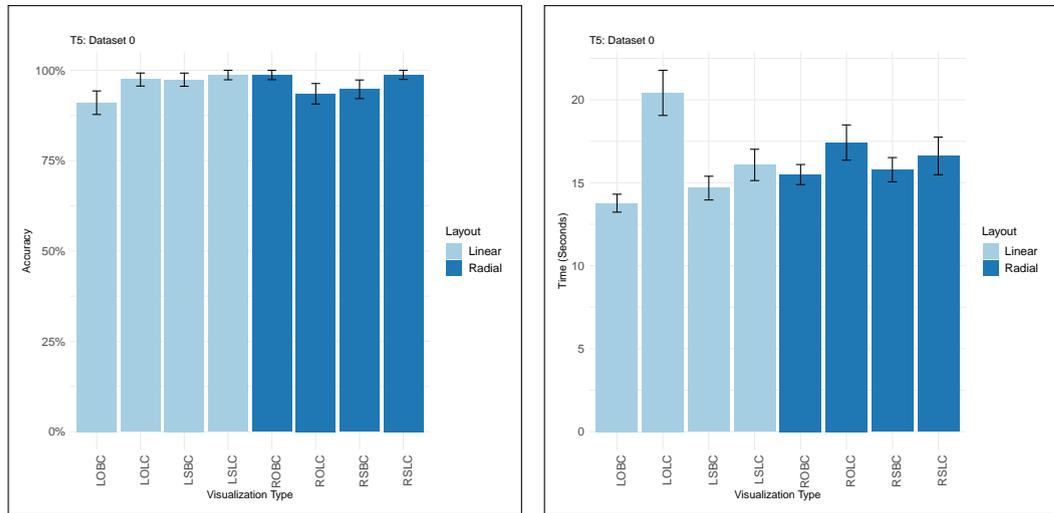
T4: Inferl Across layouts				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	LSBC0 VS RSBC0	0.571	< .001
	2	LSBC0 VS ROBC0	1.000	< .001
	3	LSBC0 VS RSLC0	0.17	0.006
	4	LSBC0 VS ROLC0	0.134	0.001
	5	LOBC0 VS RSBC0	< .001	< .001
	6	LOBC0 VS ROBC0	< .001	< .001
	7	LOBC0 VS RSLC0	< .001	0.003
	8	LOBC0 VS ROLC0	0.008	0.001
	9	LSLC0 VS RSBC0	0.201	0.001
	10	LSLC0 VS ROBC0	0.69	< .001
	11	LSLC0 VS RSLC0	0.052	0.028
	12	LSLC0 VS ROLC0	0.523	0.019
	13	LOLC0 VS RSBC0	< .001	< .001
	14	LOLC0 VS ROBC0	0.007	< .001
	15	LOLC0 VS RSLC0	< .001	0.005
	16	LOLC0 VS ROLC0	0.189	< .001
T4: Inferl Across layouts				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	LSBC1 VS RSBC1	0.134	0.200
	2	LSBC1 VS ROBC1	0.003	< .001
	3	LSBC1 VS RSLC1	0.115	0.016
	4	LSBC1 VS ROLC1	0.002	0.02
	5	LOBC1 VS RSBC1	0.424	0.008
	6	LOBC1 VS ROBC1	0.832	< .001
	7	LOBC1 VS RSLC1	0.441	0.001
	8	LOBC1 VS ROLC1	0.719	0.001
	9	LSLC1 VS RSBC1	0.078	0.005
	10	LSLC1 VS ROBC1	0.002	< .001
	11	LSLC1 VS RSLC1	0.049	< .001
	12	LSLC1 VS ROLC1	0.002	0.001
	13	LOLC1 VS RSBC1	1.000	0.001
	14	LOLC1 VS ROBC1	0.286	< .001
	15	LOLC1 VS RSLC1	1.000	< .001
	16	LOLC1 VS ROLC1	0.248	< .001

Table 6.27: Task T5 (Locate) Analysis Within the Linear and Radial Layouts for Dataset 0 and 1.

T5: Locatel Within Linear layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	LSBC0 VS LOBC0	0.180	0.237
	2	LSBC0 VS LSLC0	1.000	0.326
	3	LSBC0 VS LOLC0	1.000	< .001
	4	LOBC0 VS LSLC0	0.070	0.02
	5	LOBC0 VS LOLC0	0.180	< .001
	6	LSLC0 VS LOLC0	1.000	< .001
T5: Locatel Within Linear layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	LSBC1 VS LOBC1	1.000	0.666
	2	LSBC1 VS LSLC1	0.250	0.097
	3	LSBC1 VS LOLC1	0.125	0.021
	4	LOBC1 VS LSLC1	1.000	0.034
	5	LOBC1 VS LOLC1	1.000	0.01
	6	LSLC1 VS LOLC1	1.000	0.651
T5: Locatel Within Radial layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	RSBC0 VS ROBC0	0.375	0.936
	2	RSBC0 VS RSLC0	0.375	0.551
	3	RSBC0 VS ROLC0	0.250	0.307
	4	ROBC0 VS RSLC0	1.000	0.772
	5	ROBC0 VS ROLC0	0.25	0.307
	6	RSLC0 VS ROLC0	0.219	0.083
T5: Locatel Within Radial layout				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	RSBC1 VS ROBC1	1.000	0.837
	2	RSBC1 VS RSLC1	1.000	0.214
	3	RSBC1 VS ROLC1	1.000	0.689
	4	ROBC1 VS RSLC1	1.000	0.905
	5	ROBC1 VS ROLC1	0.453	0.985
	6	RSLC1 VS ROLC1	1.000	0.832

Table 6.28: Task T5 (Locate) Analysis Across Both Layouts for Dataset 0 and 1

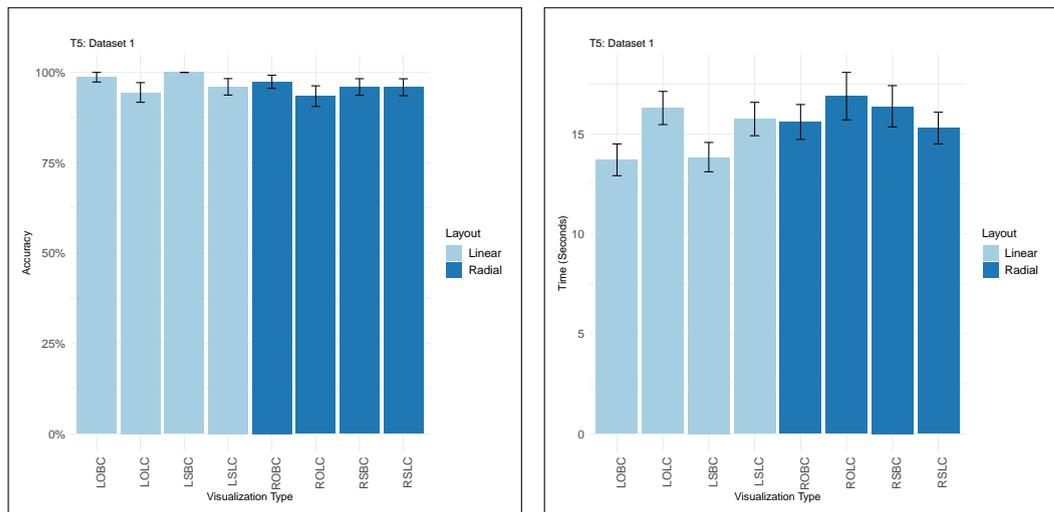
T5: Locatel Across layouts				
D	N	Pair Comparisons	Accuracy	Time
Dataset 0	1	LSBC0 VS RSBC0	1.000	0.044
	2	LSBC0 VS ROBC0	1.000	0.094
	3	LSBC0 VS RSLC0	1.000	0.197
	4	LSBC0 VS ROLC0	0.453	0.005
	5	LOBC0 VS RSBC0	1.000	0.006
	6	LOBC0 VS ROBC0	0.07	0.01
	7	LOBC0 VS RSLC0	0.031	0.094
	8	LOBC0 VS ROLC0	0.754	< .001
	9	LSLC0 VS RSBC0	0.375	0.625
	10	LSLC0 VS ROBC0	1.000	0.875
	11	LSLC0 VS RSLC0	1.000	0.921
	12	LSLC0 VS ROLC0	0.375	0.207
	13	LOLC0 VS RSBC0	0.688	0.008
	14	LOLC0 VS ROBC0	1.000	< .001
	15	LOLC0 VS RSLC0	1.000	0.001
	16	LOLC0 VS ROLC0	0.453	0.006
T5: Locatel Across layouts				
D	N	Pair Comparisons	Accuracy	Time
Dataset 1	1	LSBC1 VS RSBC1	0.250	0.011
	2	LSBC1 VS ROBC1	0.500	0.028
	3	LSBC1 VS RSLC1	0.25	0.03
	4	LSBC1 VS ROLC1	0.063	0.051
	5	LOBC1 VS RSBC1	0.500	0.02
	6	LOBC1 VS ROBC1	1.000	0.019
	7	LOBC1 VS RSLC1	0.500	0.036
	8	LOBC1 VS ROLC1	0.219	0.012
	9	LSLC1 VS RSBC1	1.000	0.786
	10	LSLC1 VS ROBC1	1.000	0.928
	11	LSLC1 VS RSLC1	1.000	0.912
	12	LSLC1 VS ROLC1	0.727	0.918
	13	LOLC1 VS RSBC1	1.000	0.314
	14	LOLC1 VS ROBC1	0.688	0.58
	15	LOLC1 VS RSLC1	1.000	0.129
	16	LOLC1 VS ROLC1	1.000	0.222



(a) Task Accuracy for 'Locate' Task on the Eight Visualization Designs (Dataset 0). The graph shows the average accuracy for the task and the standard errors.

(b) Completion Time for 'Locate' Task on the Eight Visualization Designs (Dataset 0). The graph shows the means of the completion time and the standard errors.

Figure 6.15: Task Completion Time and Accuracy for 'Locate' Task on the Eight Visualization Designs (Dataset 0).



(a) Task Accuracy for 'Locate' Task on the Eight Visualization Designs (Dataset 1). The graph shows the average accuracy for the task and the standard errors.

(b) Completion Time for 'Locate' Task on the Eight Visualization Designs (Dataset 1). The graph shows the means of the completion time and the standard errors.

Figure 6.16: Task Completion Time and Accuracy for 'Locate' Task on the Eight Visualization Designs (Dataset 0).

6.4.5 Differences in the Datasets and the Results

According to the explanation of the datasets in section 6.3.3, the datasets are different in the data distribution, their peaks and active hours during the day. This difference showed an impact on the participants' performance in different tasks, as shown in the results. Tables (6.7, 6.8, 6.9, 6.10, 6.11, 6.12, 6.13 and 6.14) show all the visualizations for the two datasets. For each dataset, two-days data were used, the first in each table is the primary visualization and the second is the one used with the second visualization for the 'Compare' task (i.e., T3). In the following, the differences between the two datasets, and how they might have affected the performance are discussed.

Identify: For the 'Identify' task, the results show some statistically significant differences between the visualizations in the task completion time on dataset 1 only. There are also some statistically significant differences between the visualizations in the task accuracy on dataset 1. The possible reason for this difference between the two datasets is that the length of the active hours for the two datasets. For dataset 0, activity data begins at 10:00 and ends at 22:00, while for dataset 1 it starts at 06:00 and ends at 22:00, which could have made it easier for dataset 0 to find the most active part of the day, centred in the middle (i.e., Afternoon) based on the distribution. As the participant can easily eliminate the 'Morning' part from the comparison because it has only two activity hours and they do not have high values for the active calories, while for dataset 1 the participant may require more time to compare the activity on the three parts of the day (Morning, Afternoon and Evening).

Determine: In the responses for 'Determine' task, the results are similar in the two datasets, and the results of the statistical tests are the same for both datasets as shown in Tables 6.21 and 6.22.

Compare: The difference in the total calories between the two days the two datasets are different, which could have affected the difficulty of the task (see the Tables 6.7 to 6.14). For dataset 0, the used two days data have different distribution of the active hours, which could have made it easier for the participants to find that the 'Second Day' had more burned calories than the 'First Day'. As the first day had active hours between 10:00 and 22:00 while the activity on the second day was between 07:00 and 23:00 and it had more peaks than the second day.

Infer: The ‘Infer’ task has similar statistical test results for both datasets as presented in Tables 6.25 and 6.26. However, there is a contradiction between which visualization method (i.e., Overlapped/Stacked) was more accurate on the two datasets. The participants were more accurate with the Overlapped visualizations with dataset 0, while they were more accurate with the ‘Stacked’ visualization with dataset 1. Looking at the visualizations of both datasets, some explanations were assumed for the contradiction in the results between which visualization method Stacked or Overlapped is more accurate for this task.

- Dataset 1 has four hours that have a step count greater than a thousand steps and require the participants to look at the corresponding HR value to examine the condition and as the variables are rendered above each other in the Overlapped design, this may have led to the participants confusing them with the other variables’ values. Unlike with the stacked visualizations that show each visualization separately. However, in dataset 0 there were only two hours that their step count was greater than a thousand, so the condition is not met in this dataset and it may have been inferred immediately when looking at the Overlapped visualization (the second hour) and helped the participant to eliminate the "Yes" answer faster.
- In dataset 0, the Overlapped data values (i.e., non-target) are not very close to the targets which could support a better perception of the information (i.e., reduced distraction). In dataset 1, the Overlapping data variables are very close to the targets (e.g., in the LOBC in dataset 1, the HR value at some time points are very close to the calories values, which is drawn on top of the HR value). This caused the value to become less salient when compared with the Stacked designs. On the other hand, with the Overlapped Line charts design, in dataset 1 the variables are interfere with each other more than dataset in 0, where they appear more separated from each other.

Locate: The difference in the results of the statistical tests for the ‘Locate’ task appears between the Bar and the Line chart and between the Overlapped and the Stacked visualizations. The results in Table 6.27 and 6.28 shows that the there is a statistically significant median difference in the task completion time between some of the visualizations within the Linear layout only on both datasets, and no differences within the Radial layout. These results

are consistence across both datasets. Dataset 0 and dataset 1 are different in the number of peak hours and the step count for each hour. As presented in the Tables 6.7 to 6.14, dataset 0 has two hours that have a high step count that are relatively close in their values when looking in the visualization and requires the participant to compare between their values (1140 and 3468 steps). On the other hand, dataset 1, the two hours that have a high step count have a more significant difference between their values (8000 and 5650 steps) than in dataset 0 which might have allowed the participant to estimate the correct response faster.

6.5 Discussion

The discussion of the results is organised based on the experiment's tasks. Within each task, the design aspects such as the layout and the visual encoding (and the possible factors listed in Table 6.4 and 6.5) are discussed. The main results of the measurements of the experiment showed that the dataset had an impact on the performance. Many information visualization studies that used more than one dataset (i.e., the same data type but with different granularities and values), showed that the dataset had an impact on the performance (e.g., [32, 158]). In Brehmer et al. [32]'s experiment, the two tested data sets showed different outcomes across the tasks in both accuracy and completion time for all data granularities. Mylavarapu et al. [158] evaluated the use of six different types of ranked-list visualizations with three dataset sizes and three tasks, their study showed that there were no effects on the results. However, it did provide evidence for identifying strengths and weaknesses of each chart with respect to the dataset, task and user. Therefore, this section discusses the results with a consideration of the dataset differences and other indicators that also impacted on the results, such as layout, visual encoding and the visualization method used to integrate the data.

According to the research problem (presented in section 1.2) and the results of the study, the discussion is based on each task, which discusses the following factors:

- **Dataset:** the study evaluated two different datasets which have multiple differences as discussed in sections 6.3.3.1 and 6.3.3.2. These differences between the datasets showed a significant differences in the results for some tasks and are discussed for each task.

- **Visual Layout:** the fundamental element of the visualization designs. The study evaluates whether the proposed Radial clock layout help in interpreting the day's data. Therefore, a set of the hypotheses considering the tasks were constructed, which are **H1a-H5a**.
- **Visual Encoding:** the Bar and Line encodings used to represent the data. Evidence from the literature and the preliminary evaluation showed different performance according to the task, which helped in constructing the hypotheses **H1b, H2b** and **H5b**. The use of the two encodings also helped in characterizing best visualizations in the case of the multivariate dataset.
- **Visualization Method:** the Overlapped and the Stacked design concerning the multivariate aspect of the dataset. These two designs were developed to address the limitation of representing only one data variable at a time. The pairwise comparisons between the visualizations for each task showed when one design outperforms the other. The hypotheses related to this factor and examine which design is suitable for the multivariate data are **H3b** and **H4b**.

6.5.1 T1 Identify

Dataset effect: In this task, the difference between the target (the correct answer) and the other answers (i.e., other parts of the day) relies on data distribution (i.e., bars and lines in each part). As presented in section 6.3.3.2, the difference between the two datasets showed that the distribution in the dataset 0 might enable the task to be more easily solved. However, no significant difference between the visualizations in dataset 0 was identified. There was only one exception, in dataset 0 the LSLC was faster for completing this task than the LOBC ($P=0.026$) as shown in Table 6.19; this could have resulted from the impact of the overlapping bars (discussed in the visual encoding section). Results in dataset 1 showed more differences between the visualizations, especially in terms of the task completion time, due to the distribution of the data.

By looking at the results for dataset 0 and dataset 1, it was assumed that in dataset 0 the participants were more likely to compare the bars in the afternoon section with the bars in the evening section, as these were considered to be adjacent and had no distractors (other bars that

were not included in the comparison) between them. However, for dataset 1 the participants were more likely to compare the bars in the morning section with the bars in the evening section, as they were considered not adjacent and had other bars (afternoon bars) between them. This may explain the lack of significant differences in the results for dataset 0, as the task might have been easier for the aforementioned reason. The significance appears in dataset 1 where the distance and the distractor between the judged bars increased, which in turn increased the difficulty, as claimed by Talbot et al. [217].

Layout (Linear vs Radial): One of the main purposes for designing the Radial layout is to facilitate identifying patterns during different parts of the day. However, the results of the across layout analysis presented in Table 6.20, show that none of the Radial visualizations outperformed any of the Linear visualizations in either of the two layouts, neither in time nor accuracy. On the Linear layout, the task completion time was affected by the visual encoding, while, on the Radial layout, the time was more affected by the visualization method rather than the chart type.

Visual Encoding (Bar vs Line): The results show that the visualizations that use the Line chart on the Linear layout helped the participants to perform the task faster than the visualizations that use the Bar chart on the same layout. However, there was no significant difference identified in terms of task accuracy. The Linear Line chart also outperformed when compared with the Radial visualizations, according to the results of the task presented in section 6.4.4. The Line chart usually refers to continuity and could facilitate perceiving data as a coherent object as identified by Zacks and Tversky [247] when compared the effectiveness of Bar and Line charts.

Visualization Method (Stacked vs. Overlapped): The outcome of the results concerning the Overlapped Bar chart (both Radial and Linear) is significant; the results show that they had a slower response time when compared with the other visualizations. A possible reason for the low performance with the Radial/Linear Overlapped Bar chart design is that the bars which indicate the calories (i.e., yellow bars) do not stand out because of the other overlapping bars (distractors) in the visual environment that may reduce the saliency of the searched pattern, which could affect the response time. In terms of the Radial layout, the comparisons between visualizations within the Radial layout show that the Stacked designs excel over the Overlapped ones in terms of task completion time. A possible explanation is that the choice of the back-

ground colours of the sectors in the Radial design may interfere with the other overlapping variables, making the target less salient and more difficult to identify. Therefore, more time was required to complete the task, unlike the Linear layout that has a white background (Wild Sand (245, 245, 245)).

In the Radial layout, the Stacked visualizations within the Radial layout are faster than the Overlapped visualizations, with no difference regarding the chart type. No significant difference was identified in terms of task accuracy (only one stacked visualization RSBC is significantly more accurate than the Overlapped ROLC). One of the possible factors in Table 6.4 for this task is the impact of the overlapping variables on the time required for the search. This impact is greater with the Radial layout than the Linear layout (which only appears in the case of the Bar chart), perhaps because of the background colour that could interfere with the target, in addition to the other variables.

Major Findings:

- The results show an outperformance of the Line chart encoding in terms of time over the Bar chart when comparing the different visualization designs within the Linear layout. This outcome is consistent with the results from a similar task (i.e., T3 Compare task), as they share similar underlying perceptual and cognitive components. Both tasks concerning the same variable (i.e., calories), require the visual grouping of many data items, estimating the total of the data within the perceived groups and comparing the totals to identify the highest total. These findings support the hypothesis H1b that the Line chart may facilitate the task of identifying patterns and trends in the data. However, this only applies to the Line chart within the Linear layout.
- With both layouts, the use of the Overlapped Bar chart was the least effective for the aforementioned possible reasons. As the bars are mapped above each other and the yellow calories bars are drawn directly on top of the orange HR bars, it is unclear whether the colour of the distractors has an impact.
- The Radial layout did not help the participants to perform this task faster or more accurately. Estimating the values of each part of the data requires referring to the axis, which is affected by the position and angle of the data.

- The Linear Line chart performed best in the identify task, for the previously discussed reasons.

6.5.2 T2 Determine

Dataset effect: In the Determine task, both datasets show significant differences in the participants' performance between the visualizations. However, the results vary based on the dataset.

Layout (Linear vs Radial): The type of layout had a significant impact on the results (both time and accuracy) as presented in section 6.4.4.

The design of the Linear layout allowed the participants to read the value of the encoded data (in Bar or Line) from the axis, which is located on the same horizontal level. However, reading the value from the Radial layout is more difficult than from the Linear layout because the position of the value is affected by the angle in the Radial design. The value on the Radial layout is located on a circular curved line, which is less accurate than estimating the position on a common scale according to [48]. This perceptual point of view could explain the long response time and the low accuracy for the Radial visualizations when compared with the Linear visualizations. The slow response time for reading values from the Radial layout when compared with the Linear layout has also been identified in previous work [32, 96] and the lower accuracy was identified in [32]. The results from Waldner et al.'s experiment [238] for evaluating both the Radial and Linear layouts with time-dependent data showed that the Linear layout is less error-prone for the reading value task and less time-efficient than the Radial layout. In addition to these reasons, searching for direction could have an impact on the time. According to the eye-tracking study performed by Goldberg and Helfman [96], the participants started searching for the item on both directions on the Radial layout (clockwise and counterclockwise) while in the Linear layout the search was performed as a rightward search.

Visualization Method (Stacked vs. Overlapped): What could be derived from the results of the comparisons within the Linear layout, is that the Stacked design is significantly faster than the Overlapped design in both datasets while it is not significantly accurate (i.e., in dataset 1 all the

Overlapped designs are more accurate than the Stacked design, and in dataset 1 the Overlapped Bar chart is significantly more accurate than the others). This tradeoff between time and accuracy could be related to the perceptual components of the task. The time is generally faster with the Stacked design as it possibly reduces the visual search time. The Stacked design clearly separates the variables from each other, which could decrease the impact of the distractors of being presented in the same search area as the Overlapped design. In terms of the accuracy of the Overlapped visualizations: (1) the design benefits from the whole visualization space, which increases the area between the two horizontal lines where the participants need to estimate the position of the value; moreover, (2) the Overlapped design has more horizontal lines between the major labelled lines, which could contribute to guiding the participants toward a more accurate estimation position and then the estimation of the value. It was not possible to add these Lines in the Stacked design because of the limited space.

With respect to the Radial layout, the Overlapped design is more accurate, but in the Linear layout the Stacked design is faster. The fact that the Stacked design is not faster in the Radial layout supports the anticipated reason for the fast response on the Linear layout. The visual search for the position of the value could be affected by the circular lines which are very close in the Stacked design. Moreover, the participants need to follow the closet line to the target toward the axis to estimate the value, which may interfere with adjacent lines. Regarding the accuracy of the Overlapped design, the first reason for the Linear layout applies here. The second reason is not related, because there are no additional lines between the labelled lines, which were avoided in the design due to the space and the background colour. Only one exception for these findings is in dataset 1 which shows no difference in the accuracy between RSBC and ROLC.

Visual Encoding (*Bar vs Line*): For both datasets, the Overlapped Bar chart design is significantly more accurate than all other visualizations within each layout (as presented in section 6.4.4). The effect of the Overlapped and the Stacked design is discussed in the visualization method section.

Another finding regarding the visual encoding is that the Bar chart is significantly more accurate than the visualizations with the Line charts within the Linear layout on both datasets. This outcome is in line with previous research that the Bar chart has better accuracy for reading or retrieving value tasks such as [186] and graphical perception experiment [48, 217]. From

the time the position of the hour in the task question is located (i.e., task 2 in Table 6.4), the participant moves to task 3 in Table 6.4 to determine the value. This step could be more straightforward with the Bar chart (as following the Line of the Bar between the two points) while in the Line chart, there is no vertical line to guide estimating the position. The results of the comparisons within the Radial layout show that in the case of both the Bar and the Line charts each performs better than the other only when it is Overlapped.

An interesting finding is that regarding the performance of the Linear Overlapped Line chart (LOLC): it declined according to the dataset. For dataset 1, this visualization had the lowest accuracy and the longest completion time between all other visualizations on the Linear layout and it was not faster or more accurate than any of the visualizations on the Radial layout (it was also significantly less accurate than ROBC and ROLC). For dataset 0, LOLC was not significantly less accurate or slower than other Linear visualizations, and when compared with the Radial visualizations it was faster and had higher accuracy than the Radial Stacked visualizations (RSBC and RSLC). What could explain this decrease in LOLC with dataset 1 is the increased distance between the chart height and axis, which could affect the time and accuracy. Heer et al. in their evaluation of time-series visualizations [107] hypothesised that chart height might have a primary impact on the estimation performance, as the increased distance vertically or horizontally could negatively affect the time and the accuracy of the estimation.

Major Findings:

- Using the Radial layout for determining value task is not effective, neither in time nor accuracy. An evaluation of a similar task in the previous works [32, 96] showed similar findings.
- The Overlapped Bar chart (LOBC) was significantly more accurate than other visualizations (Radial and Linear), while the Stacked Bar chart (LSBC) was significantly faster than all other visualizations (only one exception showed no significance between LOBC and ROLC on dataset 1). The estimated reasons for these findings are that space and the additional horizontal lines in the LOBC could help in more accurate estimation of the position and the value. While in the LSBC, the separated representation of the variables could reduce the impact of the distractors and decrease the visual search time.

- The performance of the Linear Overlapped Line chart showed different results according to the dataset. The main difference between the two datasets in this task is the distance between the target value (the height of the chart) and the x axis that show the hours. A possible explanation provided is that the distance between the two points could impact the accuracy and the time, and this is also hypothesized by Heer et al. [107]. However, this has not been evaluated in this study.

6.5.3 T3 Compare

As explained earlier in section 6.3.2, the compare task requires comparing two groups of items that relate to two days, which increases the complexity of the task, as there are many data items to compare. According to Gleicher [94], the difficulty of the comparison increases as the number of items being compared increases. The visual grouping aspects and estimating the distribution of the data share some similarities with the Identify task, they also relate to the same data variable (calories). The data in the Identify task is divided into four segments on the same day, while in the Compare task, the participant needs to compare the total and the distribution on two different days which are located on different sides on the screen.

Dataset effect: For dataset 0, there was no significant difference between all the visualizations within each layout and across both layouts (see Table 6.23 and 6.24). The level of the difficulty of the two datasets varies for this task (see datasets descriptions in Table 6.6 and section 6.3.3.1). The images used for the two days are presented in Tables (5.21 - 5.28). For dataset 0, the utilised two days' data have different active hours distribution which could makes it easier for the participants to find that the 'Second Day' has more burned calories than the 'First Day'. As the first day has active hours between 10:00 and 22:00 while the activity on the second day is between 07:00 and 23:00 and it has more peaks than the second day. The number of items to be compared, their relationships, their size and complexity could increase the difficulty of the task [94]. The following is only related to dataset 1, as no significant differences were identified in dataset 0.

Layout (Linear vs Radial): In dataset 1, there were some significant differences between the layouts, but they also were related to the type of visual encoding used within the layout. The

results showed that the Linear Line charts were significantly more accurate than other Radial designs (except ROLC), while in case of the Linear Bars, only LSBC was more accurate than ROBC. Additionally, the results showed that ROBC was significantly slower than all other Linear design. These differences are explained further in the remaining text.

Visualization Method (Overlapped vs Stacked): A major observation from the results is not only related to the visualization method, it is also related to the visual encoding. The Radial Overlapped Bar chart has ROBC the worst performance between the Radial visualizations and across both layouts. It has a significantly longer response time and less accuracy than all other Radial visualizations (except with RSLC as no difference in the accuracy identified) and all other Linear visualizations (except with the LOBC as no difference in the accuracy was identified). This result is in line with the results from the Identify task, as the ROBC has the longest response time between the visualizations. The same explanation given for its long response time in the Identify task is related here. In the Radial Overlapped Bar chart design there is a background colour and the variables are rendered on top of each other. The yellow Bars (i.e., the calories) do not pop out with the overlapping variables, which may reduce the saliency of the searched data pattern and which could affect the response time. This also may affect the visual grouping that is required to estimate each day's total data. On the other hand, the Radial Overlapped Line chart is significantly more accurate than other Radial visualizations and it is not less accurate or slower than the Linear visualizations (significantly more accurate than the LOBC).

Visual Encoding (Bar vs Line): Within the Linear layout, there was no difference in the task completion time between all the visualizations. The Linear Overlapped Line chart visualization was significantly more accurate than the Linear Bar chart visualizations (LSBC and LOBC), it was also more accurate than all the Radial visualizations (except with the ROLC which showed no difference), and only significantly faster than the ROBC. The Linear Stacked Line chart visualization was significantly more accurate than the LOBC and also more accurate than all the Radial visualizations (except with the ROLC which showed no difference) and was only significantly faster than the ROBC.

This recurring pattern in the result shows that the Overlapped Line design (on both layouts) helped the participants to perform this task more accurately than any of the alternatives. It also

shows that the Radial Overlapped Line chart is the best choice when considering comparing on a Radial layout. For the Linear layout, the results show that the Line encoding is more accurate than the Bars, this could have resulted from the continuity of the Line chart that allows it to be perceived as one object (visual grouping), as it only requires comparing the height of it instead of looking at the item separately in the Bar chart. It is also appears that the Bar chart is more suitable for comparing discrete data items rather than looking for trends [247], something which is required in this task in order to find the day with the highest total calories.

Major Findings:

- The Radial Overlapped Bar chart had the worst performance between the Radial visualizations and across both layouts. It had a significantly longer response time and less accuracy than all other Radial visualizations (except with the RSLC as no difference in the accuracy was identified) and longer than all other Linear visualizations (except with the LOBC, as no difference in the accuracy was identified). This result is in line with the results from the Identify task, as the ROBC had the longest response time.
- The Radial Overlapped Line chart is generally faster than other Radial visualizations, with no difference in the accuracy (only more accurate than the ROBC). The Linear visualizations were not faster or more accurate than this Radial design in any of the cases. As explained earlier, the Linear Overlapped Line chart does not have the same issue as the Overlapped Bars have and there are fewer circular lines in the background than the Radial Stacked design, which could affect the visual search.
- Similar to the results from the Identify task T1, the Radial layout in dataset 0 is not less accurate or slower. However, in dataset 1, the Linear layout is significantly more accurate and/or faster in many visualization designs. It is assumed that as the difficulty of the task increases (e.g., dataset 1 has more elements to compare than dataset 0), the designs with the Linear layout are more likely to perform better. As discussed earlier, there are many factors that are expected to impact on the perception of the data encoded within the Radial visualizations. For example, the position of the axis with respect to the data (i.e., Bars) that are distributed around the centre, the possible interference of the circular lines with the background and the data. In addition, many of the participants

declared low familiarity with the use of the Radial designs, as shown in Figure 6.5.

- The results show no significant difference in the performance between the two layouts for dataset 0, while for dataset 1, the Radial layout was slower and less accurate in some cases according to the visual encoding and visualization method. Previous studies that have evaluated proportional comparison tasks between the Pie and the Bar charts showed that the Pie chart performed more accurately with a similar completion time [52, 58].
- The results show that the Radial Overlapped Line chart is the best choice when considering comparison tasks on a Radial layout

6.5.4 T4 Infer

The Infer task is considered more complicated in comparison with the other tasks as it is composed of many perceptual and cognitive tasks, these are presented in Table 6.5. According to the possible other factors, the dataset was estimated to have an impact on the results, in addition to the layout and the visualization method which are discussed in the following sections.

Dataset effect: Section 6.3.3.1 presents the major differences between the two datasets. What relates to the Infer task is that dataset 1 has four hours that have a step count greater than a thousand steps and requires the participants to look at the corresponding HR value to examine the condition, while in dataset 0 there are only two hours for which their step count is greater than a thousand. Therefore, the condition could be inferred as being faster than dataset 1, as the subtasks listed in Table 6.5 had to be repeated fewer times than in dataset 1.

Layout (Radial vs Linear): A significant finding in the Infer task is the impact of the layout on the participants' task completion time. The participants performed significantly faster with the Linear layout. In terms of the accuracy, the results vary according to the dataset. In dataset 0, the Overlapped design (i.e., the LOBC and LOLC) are more accurate than all other Radial visualizations (except no significant difference identified between the LOLC and ROLC). In dataset 1, the Linear Stacked (i.e., the LSBC and LSLC) is significantly more accurate than the Radial Overlapped designs (i.e., the ROBC and ROLC). Searching for the target here (i.e., steps over a thousand) on the Linear layout requires searching for data that has a height that

reaches a specific level (the point on or height of the Bar). This task was expected to be faster in cases where the value and axis are on the same level (i.e., no effect of angle) as the results confirm. In the Radial layout, the position of the axis and the values on the circular lines may have an impact on searching and retrieving value, as discussed in Task T2 Determine, as the results of the task show a significant slow response time for the Radial layout. In the Linear layout, the values are aligned vertically and horizontally, this provides a more effective, more organised search as compared to the Radial layout that has the values labelled on a circle.

Visualization Method (Overlapped vs Stacked): In terms of the Overlapped and the Stacked design, the results show some contradictions between the two datasets. The Overlapped design tend to be more accurate with dataset 0, while the Stacked design is more accurate than other visualizations with dataset 1. The main difference between the two datasets is the number of peaks. In dataset 0 there are only two peak values and the distribution of the data does not seem to produce highly cluttered representation such as on dataset 1 (see the figures in the Tables 6.7 to 6.14) when the data is Overlapped. In dataset 0 the Overlapped data (i.e., non-target) are not very close to the targets, this could support better perception of the information. On dataset 1, the Overlapping variables are very close to the targets (e.g., in the LOBC on dataset 1, the HR value at 7 and 9 are very close to the calories value at that hour which is drawn on top of the HR value). In this case, the value became less salient when compared with the Stacked designs. Looking at the Overlapped Line charts on both datasets, in dataset 1 the variables interfere with each other more than they do in dataset 0 which appear as being more separated from each other.

Visual Encoding (Bar vs Line): As discussed earlier, choosing between the Overlapped and the Stacked design for inferring a relationship between multivariate data were more affected by the dataset. Within both layouts and for both datasets, no evidence showed an outperformance of one encoding over the other. The outcomes were more related to the layout and the visualization method, as explained earlier.

Major Findings:

- The Linear layout is significantly faster than the Radial layout (and more accurate in some cases according to the design). These results are supported by tasks that have

similar perceptual components (i.e., T2) as presented earlier and from previous work, such as [96] which stated that the participants were slower with the Radial layout when reading values because they visually searched in different directions unlike the with the Linear layout where the participants only searched in a rightward direction.

- The effectiveness of the Overlapped design for the multivariate data could be affected by the dataset. A possible reason for the difference of the accuracy performance in the Overlapped design is the distribution of the values within the dataset. The closer the data is to each other, the greater is the chance of a decrease in accuracy. This outcome did not fully support the Overlapped design that was proposed to facilitate observing multivariate variables in the dataset.

6.5.5 T5 Locate

The difference between the results according to the dataset was generally consistent across both layouts in the previously discussed tasks (e.g., in T1 and T3, there is no significant difference between the visualizations within both layouts). In this task, the within layout analysis varies according to the layout. Within the Radial layout, the results show no single significant difference between the visualizations on both datasets. Within the Linear layout analysis, the Linear Overlapped Bar chart (the LOBC) is significantly faster than the Linear Line charts (the LSLC and LOLC) on both datasets. The Linear Overlapped Line chart (the LOLC) is slower than the Bar charts (the LSBC and LOBC) on both datasets. No difference in accuracy was identified in any of the cases.

A major perceptual component of this task that was expected to affect the task completion time was the visual search for the target. The task instruction asked the participants to find the maximum step count hour. As presented in Table 6.5, the type of visual search is a conjunction and requires looking for targets with two features: blue colour and highest value. Visual search (either parallel or serial), is considered efficient if the target pops-out from other non-target elements in the display and inefficient otherwise [245]. According to this, a conjunction search for a target with a combination of features (e.g., colour and size) could be efficient [245]. How salient the target is from other distractors, in this case, depends on the design and the datasets,

as discussed in the following paragraphs.

Layout (Radial vs Linear): No difference was identified between the visualizations within the Radial layout. Also, when comparing the visualizations across both layouts the results are more related to the design of the Linear visualizations (i.e., the fast response time of the LSBC and LOBC on both datasets and the slow response time for the LOLC that was also consistent when compared with the other Linear visualizations).

Visual Encoding and Visualization Method: For both datasets, the results of within the Linear layout analysis show that the Linear Overlapped Bar chart LOBC is faster than the Line charts, but not faster than the Stacked Bar LSBC. The LSBC is only faster than the LOLC and there is no difference when compared with the Stacked Line chart LSLC. These results show that a combination of both visual encoding and the visualization method may affect the task completion time. Across both layouts, the Linear Bar charts LOBC and LSBC are significantly faster than most of the other Radial visualizations on both datasets (3 exceptions with dataset 0 and one with dataset 1 show no difference).

A significant outcome of the results is related to the Linear Overlapped Line chart LOLC. It has the longest completion time than other visualizations on both datasets and it is significantly slower than all other visualizations on dataset 0 (there were no significant differences in time on dataset 1). A possible reason for this slow response time is that the distance between the searched target (i.e., the highest point on the Line) and the corresponding value on the X-axis (i.e., the hour that the participant needs to retrieve), which is the time required to locate the hour after locating the highest value (step 4 in Table 6.5). This reason is the most related as the LOLC has the greatest distance between these two points and this does not happen with the Radial layout as the values are located around the outer circle at a much closer distance to the peaks of the step count than the LOLC. With respect to why the LOLC is significantly slower than Radial visualizations only with dataset 0, this could be because the target in dataset 0 is more salient from the distractors which could reduce the time for locating the peak (i.e., task 1 and 2 in Table 6.5).

Major Findings:

- The Bar chart encoding within the Linear layout is faster than the Line chart for locating

maximum tasks. The visual encoding within the Radial layout had no significant impact on the results.

- The Linear Overlapped Line chart could be the worse option within the Linear layout because of the possible reasons previously provided.

Table 6.29: The Hypotheses from the Main Study

Task	Code	Hypothesis
T1	H1a	The Radial layout (clock metaphor) to represent daily activity (during morning, afternoon, and evening) allows users to compare the data and identify trends faster than the Linear layout
	H1b	The Line chart encodings within the layouts facilitates finding trends in the data over time, and therefore performing the task faster than the Bar chart
T2	H2a	This task is more accurate and faster in the Linear layout than in the Radial layout
	H2b	This task is more accurate and faster with the Bar chart than with the Line chart on both layouts
T3	H3a	The comparison of visualizations over two days is faster and more accurate on the Linear layout than the Radial layout
	H3b	The comparison of visualizations of the multivariate dataset over two days is faster and more accurate on the Stacked design visualizations than the Overlapped ones
T4	H4a	The Linear layout is faster and more accurate at finding the relationship between two different variables from the multivariate dataset than using the Radial layout
	H4b	The Overlapped chart is faster and more accurate at finding the relationship between two different data variables from the multivariate dataset, compared to the Stacked chart
T5	H5a	Locating maximum and minimum values occurs faster and more accurately with the Linear layout than the Radial layout
	H5b	Locating maximum and minimum values occurs faster and more accurately with the Bar chart than the Line chart

6.5.6 Hypotheses Testing Outcomes

This section discusses the hypotheses listed at the beginning of the chapter and is detailed in section 6.2.1 and listed in Table 6.29.

- **H1a:** The results did not support the hypotheses; therefore, it is rejected. The Radial layout did not help the participants to perform this task faster or more accurately. On dataset 0, the time and accuracy of the Radial layout was comparable to the Linear layout. On dataset 1, the Linear layout was generally faster (with an exception when comparing the Radial Stacked charts with any of the Linear Bar chart), and only more accurately in few cases of the comparisons. Estimating the values of each part of the data requires referring to the axis, which is affected by the position and angle of the data.
- **H1b:** For the visualizations within the *Linear* layout, the findings supported this hypothesis and showed that the Line chart facilitated the task of identifying patterns and trends in the data. For the *Radial* layout, the results did not confirm the hypothesis, there was no significant difference identified in terms of the time or accuracy regarding the chart type. The results only showed that Stacked visualizations are significantly faster than the Overlapped visualizations, with no difference in terms of task accuracy. These results are related to dataset 1, as all the visualizations were comparable on dataset 0.
- **H2a:** The Linear layout generally outperformed the Radial layout. The result showed that the type of layout had a significant impact on the results (both time and accuracy). The comparisons generally showed that the visualizations with the Linear layout are either significantly more accurate, or faster, or both, as compared with the Radial layout on both datasets. Related previous studies (e.g., Brehmer et al. [32] and Goldberg and Helfman [96]) discussed earlier in section 6.5.2 also support this finding.
- **H2b:** Within the *Linear* layout, the Bar chart was significantly more accurate than the visualizations with the Line charts on both datasets. The LOBC showed the best performance of all the Radial visualizations on both datasets. This outcome is in line with previous research that the Bar chart has greater accuracy for reading or retrieving value tasks when applied on a Linear layout, such as [186]. Within the *Radial* layout, the results were more affected by the visualization method than the chart type. When comparing the Overlapped design with the Stacked designs, the Overlapped design was significantly more accurate on both datasets and faster on dataset 0 only. There is only one exception on dataset 1, which showed no difference in accuracy between the RSBC

and ROLC. The ROBC had the best performance of all the Radial visualizations on both datasets.

- **H3a:** Similar to the results from the Identify task T1, the Radial layout was comparable to the Linear layout (i.e., no significant difference between the visualizations). However, on dataset 1 some Linear visualizations were significantly more accurate and/or faster than the Radial visualizations. The results were more related to the type of design within each layout (see section 6.5). However, it could be identified that Linear layout is either has the same effectiveness or better for the comparison task.
- **H3b:** On both layouts, no difference was identified in dataset 0. The hypothesis could not be confirmed as the results varied according to the layout and the dataset. Within the Linear layout, the results were more affected by the chart type than the visualization method. The Linear Overlapped Line chart (LOLC) was significantly more accurate than the Linear Bar charts (LSBC and LOBC). The Linear Stacked Line chart (LSLC) was only significantly more accurate than the LOBC. No significant difference was identified in time.

Within the *Radial* layout, the results were more related to the visualization design (i.e., chart and visualization method). The Overlapped design (the ROLC) performed more accurately than all Radial designs with no difference in time, while the ROBC was significantly slower than the other Radial visualizations. The ROLC is the best choice when considering the comparison of two multivariate datasets on a Radial layout. In summary, the findings reject the hypothesis and suggest that within the Linear layout the Line chart is either comparable or more accurate than the Bar chart based on the dataset. This is similar to the results for task T1 (H1b). Within the Radial layout, the results showed that the ROLC is more accurate while the ROBC is the slowest between the Radial visualizations.

Calling a task “comparison” is not sufficient as it may include a range of different actions and it can be framed as a relationship between a target and other objects [94]. The design of the comparison tasks implemented in previous studies that evaluated Radial and Linear layouts [32, 238] are different from the comparison task in this study. The Compare task T3 requires

the participant to compare the total calories data on two separate days and decide which day has a higher total or decide whether they are equal. Moreover, it includes multivariate datasets that were visualized on top of each other or represented separately. Waldner et al. [238] implemented the compare task, comparing AM/PM interval values, while Brehmer et al. [32] required comparing one value with others and comparing two ranges (two subsets of the data).

- **H4a:** This hypothesis is confirmed by the results. Similar to the T2 results (T4 shares some components with T2 as they both require reading values of the step count). The Linear layout generally outperformed the Radial layout; the Linear visualizations were significantly faster than the Radial ones on both datasets (one exception showed no difference between the LSBC and RSBC on dataset 1). They were also more accurate in some cases, the reasons, related to the design type and the dataset, for which were discussed earlier.
- **H4b:** According to the results and discussion, the hypothesis could not be confirmed because of the changing performance between the Overlapped and Stacked designs between the two datasets. An estimated conclusion from the experiment is that the Overlapped design could only outperform if the effect of the clutter is eliminated. In the case of the experiment's datasets, it is the density of data items and how the targets overlap with the no-target objects.
- **H5a:** The results did not fully support this hypothesis. The difference is more related to the design of the visualizations. The result showed a fast response time of the LSBC and LOBC on both datasets and the slow response time for the LOLC that was also consistent when compared with the other Linear visualizations.
- **H5b:** Within the Linear layout, the result supported the hypothesis. The Linear Bar visualizations are either faster than the Lines or comparable. There is also evidence from the literature such as [142], which stated that the bars are effective for comparing values because the human visual system is good at comparing Bar lengths. The results from [186] also stated that the Bar chart is significantly faster than other alternatives. Within the Radial layout, the hypothesis is rejected as the results showed no significant difference between the visualizations within the Radial layout.

6.6 Conclusion and Summary of the Outcomes

In section 6.5, the results were discussed according to the tasks. In each task, the discussion was carried out according to the factors that affected the results (e.g., dataset, layout and visual encoding). The aim was to provide an interpretation for the results of the pairwise analysis performed between the visualizations.

This section finalises the study findings as a summary to address the research hypothesis and the novelty of the findings concerning the literature. The hypotheses and the discussion of the visualization alternatives are classified into two levels based on the analysis: between both layouts and within each layout.

1. The Radial layout did not significantly facilitate the performance of any of the tasks. However, in some cases, such as in dataset 0 of T1 Identify and T3 Compare tasks, the difference is not significant. In Diehl et al.'s [53] experiment, one of their observations regarding the difference between the Radial and the Linear visualizations is that it is easier to memorise a single cell on a Radial chart, while it is easier to memorise three cells on the Linear chart [53]. Although the implemented task in their study is different and the visual encoding within each layout is based on colour coding which was not implemented in this thesis, this observation might explain the drop in the participants' performance on the Radial layout when the data density increases.
2. The Radial and Linear Overlapped Bar charts (LOBC and ROBC) performed the worst with respect to the task completion time in the Identify task T1 within their layouts, while they performed the best in terms of the task accuracy in the Determine task T2.
3. Some visualizations reduced the time users required for searching for information by grouping related information using particular visual encoding (e.g., colour or spatial grouping) to support a serial and parallel search [41]. This concept was attempted to achieve with the Radial design by employing the coloured segments to reduce search time and for perceiving data more quickly within each segment as a whole. However, the results showed no positive performance either in the task completion time or accuracy. The possible explanations for these results are that the colours of the background and

the observed data items might interfere and reduce the saliency, which could impact on search time and that the participants had no significant experience with the Radial design, unlike the Linear designs that might need to be learned. Future work may investigate the effect of these coloured segment on both Radial and Linear layout.

4. Visual metaphors could amplify users' cognition and information retention when they feature a familiar link with the user [29]. Although they were not specifically evaluated, the used glyphs (i.e., sun and moon icons) that refer to the parts of the day were assumed to be a supportive factor in perceiving the daily patterns in the Radial clock layout. However, the data from the analysis for the 'Identify' task did not confirm the effectiveness of the design.
5. The results show that as the difficulty of the visual tasks increases, either due to dataset complexity or the difficulty of the involved cognitive tasks, the efficiency of the Radial layout significantly decreases. This is clear with a difficult task such as T4 Infer where the Linear layout was significantly faster than the Radial layout and most accurate in some of the cases. Moreover, for some tasks such as T1 and T3, as the complexity of the dataset increase (as discussed in each task separately), the significant low performance with the Radial layout start to appear in the results.
6. Bar chart is a better choice for encoding the data when the tasks require reading, retrieving and locating a single data item. The Bar chart is known to be a common chart type and easy to interpret [142], as it is a familiar, simple and uncluttered representation [158], which makes it recommended for many cases. Uncluttered representations with respect to multivariate dataset evaluated here only apply with the Stacked design.
7. The results and components of tasks T1 and T3, explained in the discussion section, show similar results for the tasks and involve similar cognitive and perceptual components. The two tasks can be considered as comparing within a single day (i.e., the identify task T1) and comparing across two days (i.e., the compare task T3).

6.6.1 Summary Tables

In order to have a better view of participants' performance with each visualization, the analysis results have been summarised in several tables. The first is the group (Tables 6.30 to 6.34), which highlights the main findings obtained from the main study. Each task's table presents the best and the worst visualizations in terms of the performance when comparing (1) the Radial with Linear visualizations (i.e., the across layout analysis), (2) the visualizations within the Linear layout and (3) the visualizations within the Radial layout. As the effect of the datasets had a major impact on the results, the table states in which dataset each finding applied. These tables also show the best and the worst visualization based on the performance of each task. In the across layout analysis (L vs R): the visualization is classified to be the worst of the visualizations if it is significantly slower or less accurate than three or more of the compared visualizations and the same for the best visualization. In the within layout analysis (L for within Linear and R within Radial): the visualization is classified to be of the worst visualization if it is significantly slower or less accurate than two or more of the compared visualizations, the same for the best visualization.

The second summary is presented in Table 6.35, showing how each visualization performed overall based on the task and dataset.

Table 6.30: Main Evaluation Study Findings Summary T1 (ns refers to no significant).

Task T1: Identify				
D	Layout	Findings	Best Vis	Worst Vis
0	R vs L	ns	-	-
	L	ns (one exception LSLC is faster than LOBC)	-	-
	R	ns	-	-
1	R vs L	LSLC is faster than all R charts.	LSLC, LOLC	ROBC, ROLC
		LSLC is more accurate than Radial Lines (O+S)		
		LOLC is faster than all Radial charts (except RSBC)		
		LOLC is more accurate than ROLC.		
	L	The Line chart perform faster than the Bars	LSLC, LOLC	LSBC, LOBC
R		The Stacked designs perform faster than the Overlapped	RSBC, RSLC	ROBC, ROLC
		The Overlapped Bar chart (R and L) has a slower response time compared to other visualizations.		

Table 6.31: Main Evaluation Study Findings Summary T2 (ns refers to no significant)..

Task T2: Determine				
D	Layout	Findings	Best Vis	Worst Vis
0	R vs L	Linear layout is either significantly more accurate or faster or both than the Radial layout. One exception when compared LOLC vs ROBC (ns in time and accuracy)	LSLC, LOLC, LSBC, LOBC	RSLC, ROLC, RSBC, ROBC
	L	LOBC is more accurate than LSBC and LSLC but they are significantly faster	LSBC	LOLC
		LSLC is faster than Overlapped designs LOBC and LOLC but not more accurate		
R	The Overlapped (ROBC and ROLC) are more accurate and faster than the others. ROBC is more accurate than ROLC.	ROBC, ROLC	RSBC, RSLC	
1	R vs L	Linear layout is either significantly more accurate or faster or both than the Radial layout. Except LOLC is not faster or more accurate than any of Radial visualizations	LSLC, LSBC, LOBC	RSLC, ROLC, RSBC, ROBC
	L	LOBC is more accurate than LSBC and LSLC but they are significantly faster	LSBC	LOLC
		LSLC is faster than Overlapped designs LOBC and LOLC but not more accurate		
R	ROBC is more accurate than the other visualizations. RSLC the least accurate	ROBC	RSLC	

Table 6.32: Main Evaluation Study Findings Summary T3 (ns refers to no significant)..

Task T3: Compare				
D	Layout	Findings	Best Vis	Worst Vis
0	R vs L	ns	-	-
	L	ns	-	-
	R	ns	-	-
1	R vs L	LOLC is more accurate than ROBC, RSBC and RSLC and faster than ROBC. LSLC is faster ROBC, RSBC and RSLC and faster than ROBC	LSLC, LOLC	ROBC
	L	LOLC is more accurate than LOBC and LSBC. LSLC is more accurate than LOBC	LOLC	LOBC
	R	ROLC is more accurate than RSBC and RSLC, and more accurate and faster than ROBC. RSLC is more accurate than ROBC.	ROLC	ROBC

Table 6.33: Main Evaluation Study Findings Summary T4 (ns refers to no significant)..

Task T4: Infer				
D	Layout	Findings	Best Vis	Worst Vis
0	R vs L	All Linear visualizations are faster or faster and more accurate than the Radial.	LSLC, LOLC, LSBC, LOBC	RSLC, ROLC, RSBC, ROBC
	L	LOLC is more accurate than LSBC and LSLC. LOBC is more accurate than LSBC and LSLC	LOLC, LOBC	LSBC, LSLC
	R	ROLC is more accurate than the Stacked RSLC and RSBC, and faster than ROBC. ROBC is slower than RSLC	ROLC	ROBC
1	R vs L	All Linear is faster or faster or more accurate than the Radial (except for LSBC VS RSBC ns identified)	LSLC, LOLC, LSBC, LOBC	RSLC, ROLC, RSBC, ROBC
	L	LSLC is more accurate than LOBC and LSLC and faster than LSBC. LSBC is more accurate than LOBC. LOLC is faster than LSBC.	LSLC	LSBC, LOBC
	R	ROBC is slower than the stacked RSBC and RSLC		ROBC
		On both datasets ROBC is the slowest		

Table 6.34: Main Evaluation Study Findings Summary T5 (ns refers to no significant)..

Task T5: Locate				
D	Layout	Findings	Best Vis	Worst Vis
0	R vs L	LOBC is faster than ROBC, RSBC and ROLC. LOLC is significantly slower than all other visualizations	LOBC	RSBC, ROLC, LOLC
	L	LOBC is faster than LSLC and LOLC. LOLC is slower than all other visualizations.	LOBC	LOLC
	R	ns	-	-
1	R vs L	The Bar chart is faster. LOBC is faster than all other Radial visualizations. LSBC is faster than all except ROLC.	LSBC, LOBC	RSBC, ROBC, RSLC, ROLC
	L	LOBC is faster than LSLC and LOLC. LOLC is slower than LOBC other LSBC	LOBC	LOLC
	R	ns	-	-

6.6.2 Conclusion

This chapter provided a quantitative evaluation of different visualization choices for multivariate personal health data. The study focused on (1) identifying the effectiveness of overlapping and separating different variables into one visualization design and how such method may affect the performance of the implemented visual tasks (2) using a traditional Linear layout and a Radial 24-hour clock layout (3) with different visual encoding.

To summarise the main contributions in this chapter, the differences between the visualizations evaluated in the experiment, the visualizations currently available on personal health platforms and designs from related literature were highlighted in section 8.3. Then, the findings in the results regarding the design and the tasks were listed in section 6.6. Finally, a set of best and worst visualizations according to the task and the dataset characteristics were presented in Tables 6.30 to 6.34. The following chapter is focused on the qualitative evaluation of the visualization and the subjective feedback from the participants.

Table 6.35: Best and Worst Visualizations for the Five Tasks (Dataset 0 and 1). This table shows how each visualization performs for each of the five tasks (for cross layout analysis and within layout analysis) for Dataset 0 and 1. It shows patterns of best and worst visualizations across tasks based on the Tables (6.30 - 6.34). The blue cells mean that the visualizations are significantly faster or more accurate and yellow otherwise.

		Dataset 0							
Task	Analysis	LSBC	LOBC	LSLC	LOLC	RSBC	ROBC	RSLC	ROLC
T1	Across								
	Within								
T2	Across	Blue	Blue	Blue	Blue	Yellow	Yellow	Yellow	Yellow
	Within	Blue			Yellow	Yellow	Blue	Yellow	Blue
T3	Across								
	Within								
T4	Across	Blue	Blue	Blue	Blue	Yellow	Yellow	Yellow	Yellow
	Within	Yellow	Blue	Yellow	Blue				Blue
T5	Across		Blue		Yellow	Yellow			Yellow
	Within		Blue		Yellow				
		Dataset 1							
Task	Analysis	LSBC	LOBC	LSLC	LOLC	RSBC	ROBC	RSLC	ROLC
T1	Across			Blue	Blue		Yellow		Yellow
	Within	Yellow	Yellow	Blue	Blue	Blue	Yellow	Blue	Yellow
T2	Across	Blue	Blue	Blue		Yellow	Yellow	Yellow	Yellow
	Within	Blue			Yellow		Blue	Yellow	
T3	Across			Blue	Blue		Yellow		
	Within		Yellow						Blue
T4	Across	Blue	Blue	Blue	Blue	Yellow	Yellow	Yellow	Yellow
	Within	Yellow	Yellow	Blue					
T5	Across	Blue	Blue			Yellow	Yellow	Yellow	Yellow
	Within		Blue		Yellow				

Chapter 7

Participants' Feedback and Preferences in the Visualizations

7.1 Introduction

During the main study, several different types of subjective feedback data were collected from the participants. The objective of analysing this data in this chapter is to gain further insight from the participants regarding the visualizations to support the quantitative outcomes presented in Chapter 6. Users' questionnaires are usually used to support and complement other implemented methods, such as performance measures of the task accuracy and completion time [77]. Although subjective feedback data are referred to as qualitative, some are recorded on quantitative scales (e.g., preferences and visualizations ratings). In some cases, quantifying qualitative data is possible, for example by measuring them onto a scale, such as user preferences [54].

The chapter starts by introducing the research objectives in section 7.2, it then outlines the methodology that describes the data collection methods in section 7.3 and goes on to detail the applied analysis and the results in section 7.4. Then, the discussion in section 7.5 provides an overview of the insight gained from the analysis in terms of the major design factors discussed in Chapter 6, such as the layout and the visual encoding. Finally, the chapter concludes with an overview of the findings in section 7.6.

7.2 Research Questions

From the research questions presented in Chapter 1, this chapter concentrates on addressing the following research questions (**RQ6**): What are the users' preferences in visualizations? What factors (e.g., performance and familiarity) could impact the preferences?

The aim of collecting the data during this part of the study was to seek an answer for the highlighted research question, which aimed to identify the preferred visualizations and their relation with the performance. In addition to understand how reading data from the visualizations through the implemented visual tasks affected the participants' perspective to the visualizations. The answers to the research questions set out below, contribute to measuring the experiment effect on the participants' view of the developed visualizations.

1. Is there a statistically significant difference in the participants' ratings of the visualization in terms of the readability and attractiveness before and after the experiment?
2. Is there a statistically significant difference in the participants' preferences in the visualization before and after the experiment?
3. What are the participants justifications for their choices? How do the participants' preferences and comments relate to their performance with the visualizations?

The outcomes of this chapter are used later in Chapter 8 to structure implications for visualization design and recommendations along with the quantitative data for personal health visualizations:

7.3 Methodology

As a part of the main evaluation study presented in Chapter 6, participants' feedback regarding the visualizations was collected to obtain more information about designs' pros and cons from the user point of view. The first part of the pre-experiment survey was designed to collect different demographic information about the participants (see section 6.3.5.1 for the participants' demographic information). The second part was designed to collect information about the visualization design, which is the same design as that of the post-experiment survey (see Appendices .6 and .7 for both surveys). The questions and the response types are presented in Table 7.1. The subjective feedback data collected from these surveys fall into the following three categories:

- **Likert-Scale Data:** Visualizations' ratings in terms of the readability and the attractiveness (visualization readability and attractiveness are defined in section 1.3) of the design, they are ordinal data on a scale from 'Very Poor' to 'Very Good'. The analysis and the results of this category is described in section 7.4.1.
- **Multiple-Choice Data:** The most preferred visualization(s) to use to visualize their health/physical activity data, and the results are described in section 7.4.2.

- **Open-Ended Question Data:** Short answers written by the participants to express their opinion in the visualizations, explaining why they preferred/did not prefer certain visualization(s). This analysis of the open-ended questions is presented in section 7.4.3.

As was noted in the preliminary evaluation study, the outcomes of the visualizations ratings and preferences required more explanation and justification from the participants in order to draw a conclusion for prospective personal visualization designs. Therefore, open-ended questions were added to encourage the participants to provide some notes or details on their personal opinion in the visualizations. There are many methodological uses of open-ended questions [210]; however, here these were mainly used to promote more truthful responses and to allow an opportunity for more meaningful feedback [210]. Thus, they will be analysed and used with the other data collected in this study to gain an insight.

These three categories of the data were collected during two sessions, before and after the experiment (i.e., participants conducted the tasks with the visualizations). This was performed to see how the participants' perspective of the visualization might change.

For its analysis, the data was initially explored by calculating the sum of the total ratings for each visualization in terms of the readability as presented in Figure 7.1 and in terms of attractiveness which is shown in Figure 7.2 for both pre and post-study data. Moreover, the total number of the participants who selected each of the visualizations was calculated (see Figure 7.3).

7.4 Analysis and Results

This section describes the analysis of all data collected and described in the methodology (section 7.3) and presents the results. The analysis is organised according to the type of the data as: visualization ratings, preferred visualizations and the participants' justifications for their feedback.

Table 7.1: Participants Feedback Questions. Major Parts of the Subjective Feedback and its Questions and Response Types.

Type	Question	Response type
Visualizations Ratings	Rate the Following Visualization Designs in Terms of its Readability and its Attractiveness	Likert-Scale. Ordinal data on a scale from 'Very Poor' to 'Very Good' for the Readability and Attractiveness of each visualization
Visualizations Preferences	Which of the Previous Visualization Designs Would You Prefer to Visualize your Daily Physical Activity Data?	Multiple-Choice Data. Selections from the visualizations
Preferences Justifications	Please justify your choices and clarify why did you not like other visualizations	Open-Ended Answers (Text)

7.4.1 Visualization Ratings

For the first data category (i.e., the Likert-Scale data), a nonparametric analysis was applied due to the fact that this (ordinal scale) data is not normally distributed. According to Robertson and Kaptein [183], ordinal scales, such as Likert-scale data are common non-normal data distributions in HCI studies. Since the aim was to compare the visualizations ratings' before and after the study, the data was arranged in pairs with two observed measures for each participant for their ratings of each visualization before and after the experiment. This data arrangement led to applying 16 statistical tests to identify which of the visualization(s) had a statistically significant difference in their ratings before and after the experiment in both rating terms.

The Wilcoxon-Signed Rank was selected for use, which is the equivalent nonparametric test for the paired-samples t-test [183]. The test was applied on each visualization ratings. It was selected for the following reasons: the data is ordinal-scale data and not normally distributed and the data is a paired-sample (i.e., pre- and post- study).

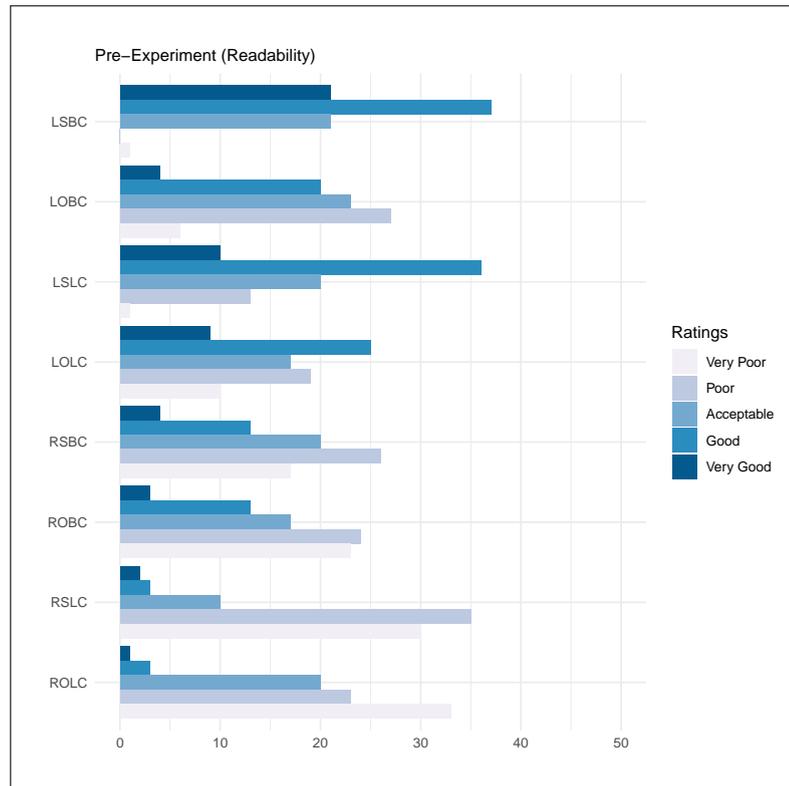
To run the test on the visualizations' rating (Likert-scale) data, the textual data needed to be transformed into numbers (a scale type-data) ranging from 1 to 5, where 'very poor' transformed to 1 and 'Very Good' transformed to 5. According to [152], this Wilcoxon test requires the variables to be in scale type, even if it is ordinal, by reassigning the variables to scale where appropriate. Table 7.2 shows the p -values for the 16 tests applied. The analysis reports on 0.05 significance level and 95% confidence interval.

The ranks start from 1 ‘very poor’ to 5 ‘very good’. The difference between the two variables was calculated by subtracting the rating after the experiment from the rating before the experiment. A negative difference means that the rating has decreased, a positive difference means that the rating for the visualization has improved and if the difference is equal to zero the ratings have not changed.

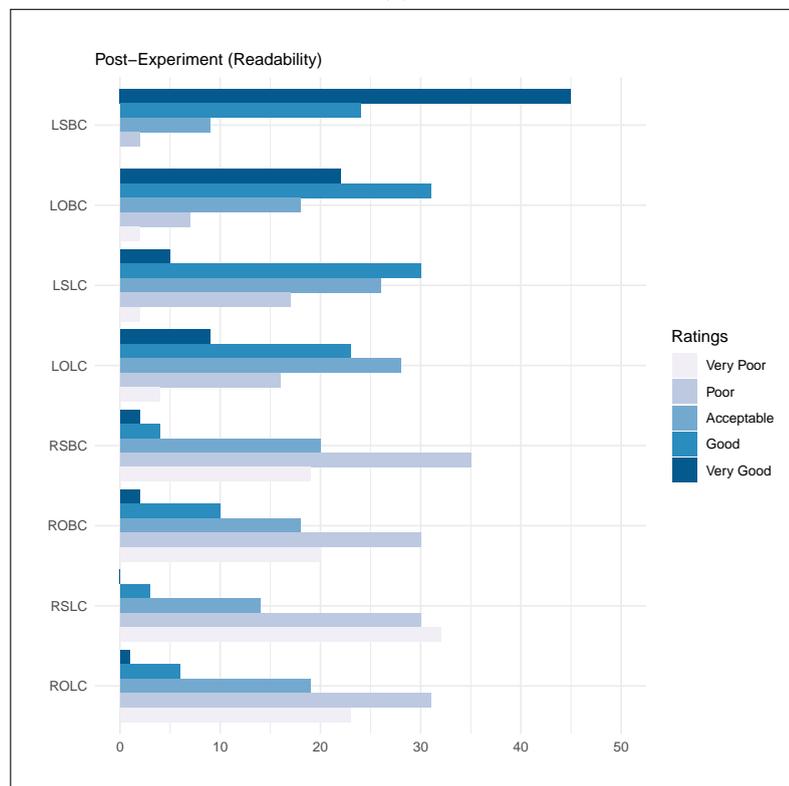
In terms of the design readability, the participants’ rating for the Linear Bar charts (LSBC and LOBC) had significantly increased after conducting the experiment ($P < .001$) in both cases. On the other hand, LSLC and RSBC rating for the readability had significantly decreased ($P = 0.15$ and 0.14 , respectively) as shown in Table 7.2. No other significant differences were identified with other visualizations. Furthermore, in terms of the design attractiveness, only LSBC and LOBC had significantly increased ($P < .001$).

Table 7.2: The Results of the Wilcoxon Signed-Rank Test on the Visualizations’ Ratings Data. The table presents the outcomes of the statistical tests that are applied to compare the participants’ rating for the eight visualizations pre- and post the study.

Readability	Test Result Summary	Attractiveness	Test Result Summary
LSBC	$P < .001$ 36 Positive, 9 negative and 35 Ties	LSBC	$P < .001$ 39 Positive, 7 negative and 34 Ties
LOBC	$P < .001$ 55 Positive, 3 negative and 22 Ties	LOBC	$P < .001$ 45 Positive, 5 negative and 30 Ties
LSLC	$P = .015$ 13 Positive, 30 negative and 37 Ties	LSLC	$P = .178$ 18 Positive, 26 negative and 36 Ties
LOLC	$P = .258$ 31 Positive, 22 negative and 27 Ties	LOLC	$P = .124$ 26 Positive, 14 negative and 39 Ties
RSBC	$P = .014$ 17 Positive, 34 negative and 29 Ties	RSBC	$P = .097$ 18 Positive, 34 negative and 28 Ties
ROBC	$P = .599$ 20 Positive, 26 negative and 34 Ties	ROBC	$P = .101$ 31 Positive, 19 negative and 30 Ties
RSLC	$P = .977$ 23 Positive, 24 negative and 32 Ties	RSLC	$P = .116$ 28 Positive, 18 negative and 31 Ties
ROLC	$P = .087$ 25 Positive, 16 negative and 39 Ties	ROLC	$P = .126$ 24 Positive, 13 negative and 34 Ties

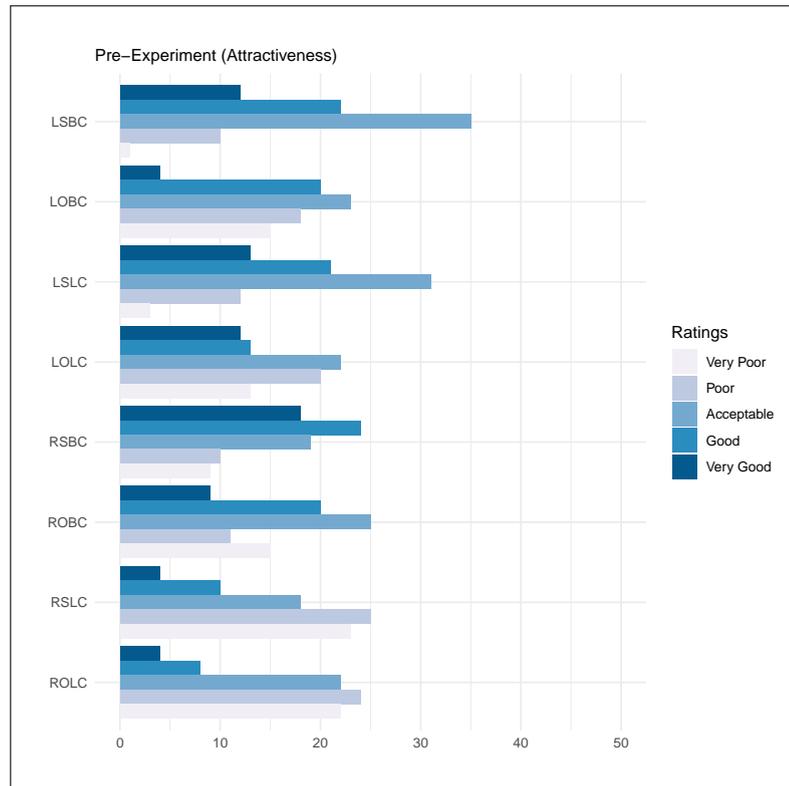


(a)

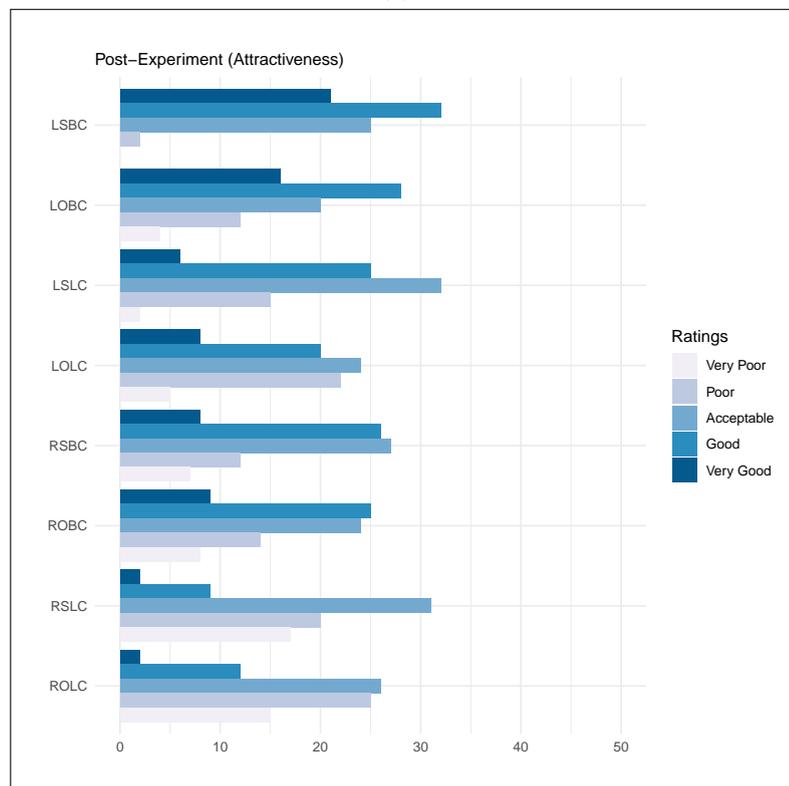


(b)

Figure 7.1: The participants' Ratings for the Readability of the Visualizations Before and After the Experiment.



(a)



(b)

Figure 7.2: The participants' Ratings for the Attractiveness of the Visualizations Before and After the Experiment.

7.4.2 Preferred Visualizations

In this question (see the Appendices .6 and .7), the participants were asked to choose the visualizations they would prefer to use to represent their physical activity data. The participants were allowed to select more than one visualization. The aim was to understand which visualization people considered as a good choice to represent the physical activity data and if there was a difference between the participants' selection after the experiment.

The participants were allowed to select more than one visualization, and the responses were collected at two different time points. The analysis which was applied was a paired comparison test for each visualization to discover if there was a statistically significant difference between the participants' preferences for each visualization before and after interacting with the visualizations.

The participants' selection was coded as zero for non-selected and one for the selected visualizations. Then, the McNemar statistical test was applied; this a test used to analyse paired dichotomous data (as the same subjects have been measured twice) and is commonly used to analyse pre and post studies ¹. The Results of the test are presented in Table 7.3 and the participants' preferences are presented in Figure 7.3.

LSBC: In the pre-experiment survey, 37 out of the 80 participants (i.e., 46% of the participants) selected the LSBC. The number of participants who selected this design increased to 44 participants out of 80 (55%) after the experiment. However, there was no statistically significant difference in the result as shown in Table 7.3. Of the 37 participants who selected the LSBC before the experiment, 6 participants did not select it in the post-survey, while 13 of the 43 participants who did not prefer this design selected it after the experiment.

LOBC: The McNemar test shows a statistically significant difference ($P=.001$) between the proportion of the participants who selected the LOBC as one of the preferred visualizations before and after the experiment. In the pre-survey, only 15 out of 80 participants (19%) included this visualization in their selection. The number of participants increased to 34 out of the 80 participants (43%) in the post-experiment survey as shown in Figure 7.3. The test shows that 10 of the 15 participants who chose the LOBC as one of the preferred visualizations had not

¹<https://statistics.laerd.com/spss-tutorials/mcnemars-test-using-spss-statistics.php>

changed their opinion, while the other 5 participants did not include it in their selection in the post-experiment survey. On the other hand, 24 participants out of the 65 who had not chosen the LOBC considered it as one of the preferred visualizations to represent physical activity data.

LSLC/LOLC: The results for both the Stacked and Overlapped Linear line charts were similar. The total number of participants who preferred them decreased after the experiment (but not significantly). Out of the 20 participants who selected the LSLC, 9 participants deselected it after the experiment and only 3 participants of the 60 participants who had not preferred it before the experiment included it in their selection afterwards. Likewise, for LOLC, 8 out of 20 participants deselected it after the experiment and 5 participants of the 61 participants who had not preferred it in the pre-survey included after completing the experiment.

RSBC: The results of the McNemar test showed a statistically significant difference ($P = .001$) between the proportion of the participants who selected the RSBC as one of the preferred visualizations before and after the experiment. In the pre-survey, 16 out of 80 participants (20%) included this visualization in their selection. The number of participants significantly decreased to 6 out of the 80 participants (43%) in the post-experiment survey. The test showed that only 4 of the 16 participants who chose the RSBC as one of the preferred visualizations had not changed their opinion on it, while the other 12 participants did not include it in their selection in the post-experiment survey. On the other hand, only 2 participants out of the 64 who had not chosen the RSBC considered it as one of the preferred visualizations after the experiment.

ROBC: Despite the results showing a large difference (as shown in Figure 7.3), further inspection showed no significant difference. The test showed no statistically significant difference between the proportion of the participants who selected the ROBC before and after the experiment. 5 out of 80 participants had selected it before the experiment and the number of participants increased to 9 in the post-experiment survey. Only one of the 5 participants who had chosen it before the experiment kept as their preferred choice afterwards and 8 of the 75 participants who had not preferred it in the pre-survey included it their preferred visualization.

RSLC/ROLC: These two visualizations were the least preferred visualizations of the eight proposed designs. As presented in Table 7.3, both visualizations were preferred by only one

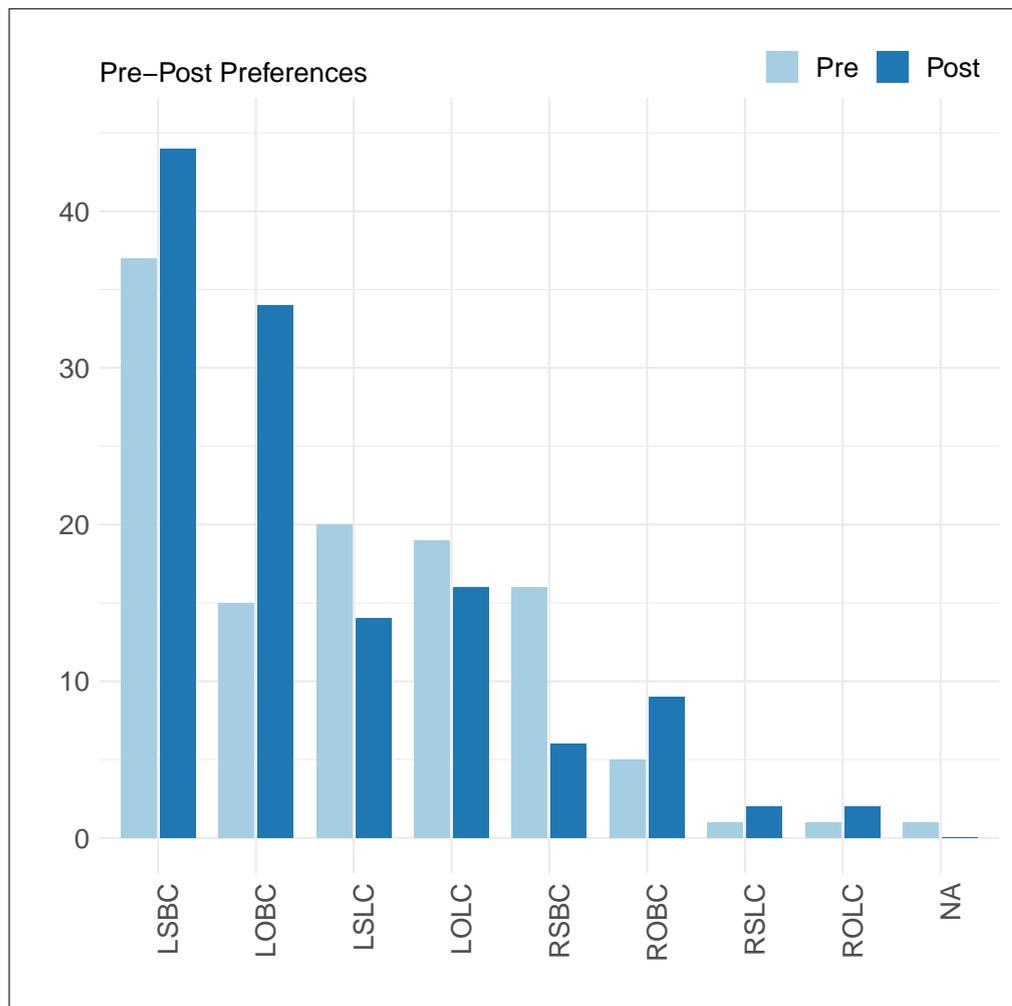


Figure 7.3: The Participants' Most Preferred Visualization(s) Before and After the Experiment. In both figures, the X-axis shows the visualization names and the Y-axis present the number of participants who selected this visualization as their preferred design to represent their physical activity data. The participants were able to select more than one visualization.

participant (two different participants for each visualization), who then changed their minds after the experiment regarding the visualizations and did not include them. The two participants who selected these after the experiment had not preferred them at first.

Change in the Preferences:

See Table 7.4 for the report on the change in the participants' preferences: how many participants kept the same choices before and after the experiment; how many kept some of their selections and how many entirely changed their preferences. The responses to this ques-

Table 7.3: The Results of McNemar Test on the Visualizations' Preferences Data. The table presents the outcomes of the statistical tests that are applied to compare the participants' most preferred visualization(s) for the eight visualizations pre- and post the study. The third column in the table provides an additional information about the number of the participants who selected each visualization in both surveys. The word 'Pr' is used in the table to refer to the word "Participant".

Visualization	P-Value	The Number of Selections in Pre-& Post-Surveys
LSBC	0.167	Pre-: 37 Prs (46%) Post-: 44 Prs (55%)
LOBC	0.001	Pre-: 15 Prs (19%) Post-: 34 Prs (43%)
LSLC	0.146	Pre-: 20 Prs (25%) Post-: 14 Prs (18%)
LOLC	0.581	Pre-: 19 Prs (24%) Post-: 16 Prs (20%)
RSBC	0.013	Pre-: 16 Prs (20%) Post-: 6 Prs (8%)
ROBC	0.338	Pre-: 5 Prs (6%) Post-: 9 Prs (11%)
RSLC	1.00	Pre-: 1 Pr (1%) Post-: 2 Prs (3%)
ROLC	1.00	Pre-: 1 Pr (1%) Post-: 2 Prs (3%)

tion showed that 53% of the participants entirely changed their preferences, 48% made some changes to the selection (added or eliminated some of the visualizations) and 36% of the participants did not make any changes to their preferences.

Based on the previous discussion regarding which visualization was more or less preferred after the experiment, it was concluded that the visualizations that used the Radial layout were not favoured to use to represent personal activity. Moreover, a visualization that is not widely used in the health/physical activity app, that is the Linear Overlapped Bar Chart visualization, was more preferred after the participants used it in the experiment to solve the requested tasks.

The Preferences Based on the Layout:

In general, regarding the layout, the results showed that visualizations with a Radial layout were the least preferred by the participants. In the pre-experiment survey, Radial visualizations were selected by 21 of the 80 participants (26%), and the Linear visualizations were preferred by 68 of the 80 participants (85%) and one participant did not choose any of them.

In the post-experiment survey, the number of participants who included a Radial visualization in their preferences decreased to 16 participants (20%), while those who included Linear visualizations in their preferences increased to 77 participants (96%).

Table 7.4: Participants' Preferences Changes. The number of participants who kept the same, partially changed and entirely changed their preferences in the visualizations after the experiment.

Participants	Number	Percent
Kept the same preferences	29 Prs	36%
Partially changed their preferences	38 Prs	48%
Entirely changed their preferences	42 Prs	53%

Table 7.5: The Relationship Between the Participants Preferred Visualizations and their Gender. The columns show the number of participants selected each type of visualizations (for each female and male participant) and the percentage of selection for each group.

Visualization	Pre		Post	
	Female	Male	Female	Male
LSBC	21 (45%)	16 (52%)	26 (55%)	17 (55%)
LOBC	11 (23%)	3 (10%)	22 (47%)	11 (35%)
LSLC	10 (21%)	10 (32%)	5 (11%)	9 (29%)
LOLC	11 (23%)	7 (23%)	11 (23%)	5 (16%)
RSBC	9 (19%)	7 (23%)	3 (6%)	3 (10%)
ROBC	4 (9%)	1 (3%)	5 (11%)	4 (13%)
RSLC	0 (0%)	1 (3%)	2 (4%)	0 (0%)
ROLC	0 (0%)	1 (3%)	2 (4%)	0 (0%)

7.4.2.1 Demographic Data and the Preferences

The demographic data presented in section 6.3.5.1 discussed the participants' characteristics, which are of two types: (1) demographic data that includes gender, age, education, employment and experience. (2) tracking data, which includes all the data regarding the participants with health tracking activity, type of the data collected and the used tracking devices.

The summary of the demographic data presented in Table 6.15, show how these data divided the participants into subgroups according to gender, age, education, tracking activity and other metrics. However, the participants within the generated groups are not equally distributed. For example, with respect to the employment status, the majority of the participants were students 64 out of the 80 participants and the rest were academic, administrative or research staff.

Due to the unbalanced distribution across demographics, no statistical test was applied on the preferences for the different demographic groups. However, the preferences of each demo-

Table 7.6: The Relationship Between the Participants Preferred Visualizations and their Tracking Activity. The columns show the number of participants selected each type of visualizations (for each tracker and not-tracker participant) and the percentage of selection for each group.

Visualization	Pre		Post	
	Trackers	Non-Trackers	Trackers	Non-Trackers
LSBC	28 (60%)	9 (19%)	31 (66%)	13 (28%)
LOBC	6 (13%)	8 (17%)	21 (45%)	12 (26%)
LSLC	13 (28%)	7 (15%)	8 (17%)	6 (13%)
LOLC	8 (17%)	11 (23%)	8 (17%)	8 (17%)
RSBC	11 (23%)	5 (11%)	5 (11%)	1 (2%)
ROBC	4 (9%)	1 (2%)	6 (13%)	3 (6%)
RSLC	1 (2%)	0 (0%)	1 (2%)	1 (2%)
ROLC	1 (2%)	0 (0%)	2 (4%)	0 (0%)

graphic group were analyzed based on the frequency (i.e., count) and percentage.

Gender: according to the data presented in Table 7.5, the preferences in the visualizations are similar across the gender. For example, for both genders, LSBC was the most preferred visualization in pre-post surveys, and the preference of LOBC had noticeably increased after the experiment.

Age: no noticeable differences were identified between the preferences across the different age groups. According to the data presented in section 6.3.5.1, the majority of the participants (around 60%) were aged between 25-34 years old, and the rest 40% was distributed to three other age groups.

Education, Degree and Employment status: according to information presented in section 6.3.5.1, the participants' education degrees could be classified into undergraduate (48%) and postgraduate (52%) degrees, which allowed a fair distribution. However, the general overview of the preferences followed the same overall preferences discussed earlier in section 7.4.2. The education field for the majority of the participants was 'Physical Science and Engineering' (69%), then 'Biomedical and Life Science' only (20%) and then 'Social Science' (11%) only. All those three groups had similar preferences patterns with the overall results with the LSBC the most preferred design prior and post the experiment, the LOBC increased after the experiment, the Radial Lines charts were the least preferred ones. Regarding the employment status,

the majority of the participants (80%) were students, which made it hard to compare with other groups.

Tracking (yes/no): of the 80 participants, 52 used tracking devices, while 27 were not using any and 1 participant preferred not to say. As shown in Table 7.6, which presents the preferences in the visualizations for each group, the majority of the tracker participants (60%) preferred LSBC and only few (13%) preferred the LOBC prior the experiment. However, for the non-tracker participants, the preferences were distributed across the Linear designs. The preferences after the experiment changed for both groups to be similar as the overall results (i.e., LSBC is the most preferred visualization followed by LOBC).

Tracking devices and tracked data: the participants who were tracking their health data (65%) were interested in tracking different types of data as Figure 6.4b shows. The majority of the participants used tracking apps (iOS and android) as presented in Figure 6.4c, further information about these apps are shown in Table 6.16. These collected data suggested that the participants used different tracking methods. However, no sufficient data were collected regarding the used tracking methods, the visualizations provided and their opinion in these visualizations.

The familiarity with charts: the participants' familiarity with the charts are shown in Figure 6.5. The results showed lower familiarity ratings with the Radial charts in comparison with Bar and Line charts, which was estimated the reason for the very low preferences in all the Radial designs. However, the participants who rated their familiarity with the Radial charts as 'Extremely Familiar' 4/80 (5%) and 'Moderately Familiar' 19/80 (24%) did not necessarily included the Radial visualizations in their preferences before the experiment (as of those 23 participants, only three selected RSBC and one selected RSBC and ROBC). The familiarity with the Bar and Line charts were considerably high as the majority of the participants were 'Extremely Familiar' (46% and 49%) and 'Moderately Familiar' (35% and 43%) respectively, and the results showed earlier the Linear Stacked Bar and Line charts were the most preferred before the experiment.

7.4.3 Participants' Justification for their Choices

The analysis of the first two parts of the participants' subjective responses from the surveys showed an interesting change in both the visualizations' ratings and the personal preferences in the visualization methods. The participants' justifications for their choices might be useful to understand the reason for the change in opinion regarding the visualizations.

When analysing the data, there was no preconceived idea or hypothesis regarding the responses, and the intention was to capture the participants' experiences with the visualizations, to interpret them and investigate their impact on their performance. A content analysis was applied [187] in order to code the responses and classify them into categories and sub-categories while focusing on what feature(s)/reason(s) they gave for selecting their preferred visualization(s). The coding procedure was aided by NVivo (Version: 12.3.0) [164].

The textual feedback was rich and exceeded the primary purpose of only giving reasons for preferring a specific graph(s), as some respondents gave recommendations or notes about specific details of the visualizations. This feedback was utilised to strengthen the discussion of what is essential to consider when designing a visualization for personal use. Structural Coding was followed, which uses a content-based phrase that represents a topic to a part of data that relates to a research question; in addition to the 'descriptive coding', that summarises the responses into a word or a short phrase [187]. After that, quantitative analysis was applied to count the number of occurrences (i.e., how many times the category/sub-category appeared in the responses). In the following, the main categories and sub-categories mentioned in the text and their frequencies are presented. Some of the identified codes such as "clear" were mentioned positively and negatively. Words that relate directly to the visualization's name (e.g., the Radial Bars) were not used in the coding process. However, they were used to link the comments to the visualization designs as discussed in section 7.5.2.1, 7.5.2.2 and 7.5.2.3.

Pre-Experiment Survey: Firstly, the script was skimmed through to identify the major codes in each response. This process generated around 42 codes that were grouped into five major categories based on the relationship between the codes. The comments were also classified in the responses as either: related to the design, which the previous codes related to; recommendations/suggestions to enhance the visualization designs and others that related to personal

opinion and preferences.

As a result, seven major themes have been identified from analysing the text, these are as follows:

1. Readability and Graph Comprehension (e.g., Easy to Read/Compare/ Understand/ Reading Values, Clear, Finding Trends, Easy/ Hard to Follow, Easy to Interpret)
2. Aesthetic Aspects of the Design (e.g., Attractive, Aesthetically Appealing, Innovative, Colours, Bright, Symbols, Metaphor, Clock Face, Look)
3. Design Features (e.g., Concise, Intuitive, User-Friendly, Familiar, Gridlines/ Grid Background, Separated, All-in-One Graph)
4. Effectiveness (e.g., Effective, Less/More Effort, Less/More Time, Quick)
5. Dataset Related (e.g., Appropriate, Suitable, Not Suitable to the Dataset)
6. Suggestions and Recommendations (for enhancing the designs)
7. Personal Preferences in the Visualization (that is not related to the design)

After grouping the codes based on their relationship into the corresponding theme/category, the total frequency of the codes was calculated in each theme as presented in Table 7.7 to reflect its importance. The frequencies of the codes in each theme identified in the participants' responses show the importance of its related aspects for the design of personal visualization. For example, the theme "Readability and Graph Comprehension" involves different types of features that are important for some people to enhance the readability of the charts, such as the ability to compare the data and identifying trends. Each theme and its underlying codes are discussed in detail in the following sections.

Post-Experiment Survey: To analyse participants' justifications in the post-experiment survey, the same procedure used in the pre-experiment analysis was applied. Starting from the previously identified themes, then exploring the text further to find new themes or codes in the responses. When analysing the text for post-experiment responses, it was found that the

Table 7.7: Major Categories/Themes identified in the Participants' Responses in the Pre- and the Post-Experiment Surveys. The table presents the major themes, the number of participants provided comments related to each theme (n) and the themes frequencies (Freq), including all the codes in each theme..

Themes	Pre-Study		Post-Study	
	n	Freq	n	Freq
Readability and Graph Comprehension	73	155	67	180
Aesthetic Aspects of the Design	20	29	16	29
Design Features	23	34	16	20
Effectiveness	6	8	5	5
Dataset Related	9	9	6	6
Suggestions and Recommendations	4	4	3	4
Personal Preferences	3	4	3	5

responses identified the same themes as were identified in the pre-experiment responses. However, the underlying codes in each theme showed some differences (i.e., some of the participants referred to different aspects that fall under the specified themes). For example, after performing the tasks in the experiment, two out of the 80 participants commented on the scale of the three variables on one chart and how that affected their ability to respond to some tasks accurately.

7.4.3.1 Theme Definitions

Table 7.7 shows the frequencies of the major categories in the responses both pre- and post-experiment. Figure 7.4 shows the major themes and examples of underlying codes within each theme. In the following sections each theme is discussed in detail with a link to the relevant responses.

Readability and Graph Comprehension: the majority of the participants' justifications for their choices (73/80, i.e., 91% in the pre-study and 67/80, i.e., 84% in the post-study survey) included comments related to the readability of the graphs. Based on the participants' responses, the readability and comprehension refer to the clarity, the ability to read and understand the graph easily, extracting information and finding trends and anomalies. The participants' justifications included different terms that were used in coding such as: 'clear'; 'easy to read'; 'easy to understand'; 'easy to follow trends'; 'read exact values'; 'easy to compare' and 'glanceable or reading the data at a glance'. The readability was also mentioned negatively in some responses such as: 'difficult/hard to read'; 'hard to understand'; 'confusing'; 'misinterpreted'

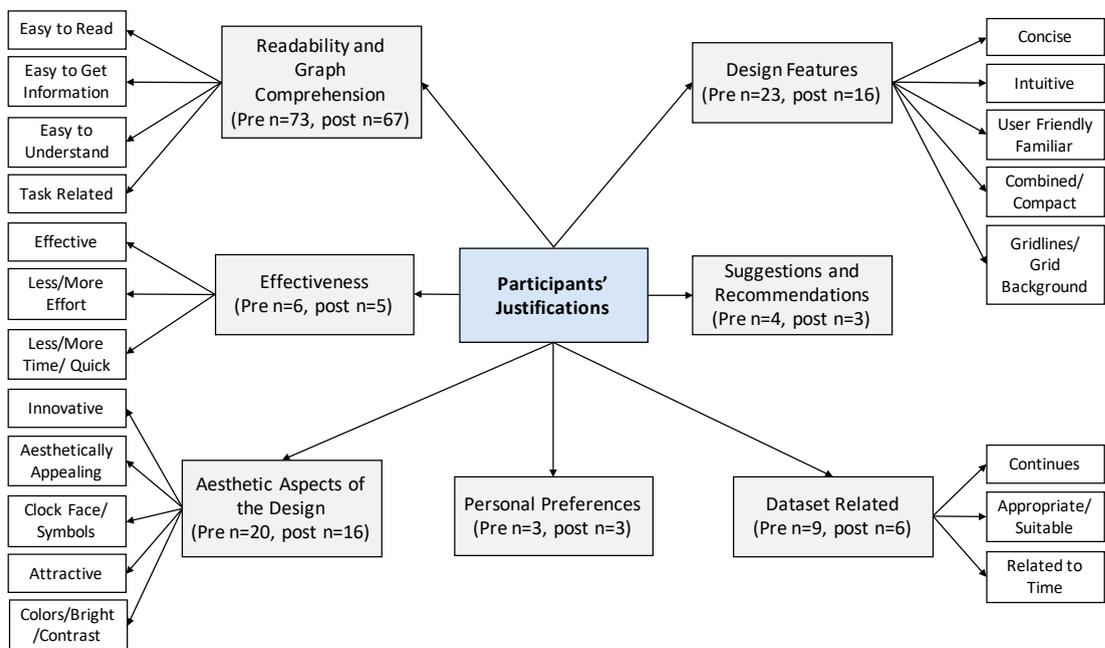


Figure 7.4: Participants' Preferences Justification Model. The grey squares show the basic themes, while the white squares represent examples of codes for each theme. The 'Task Related' is sub-theme that is described in section 7.5.2.4 and Figure 7.5.

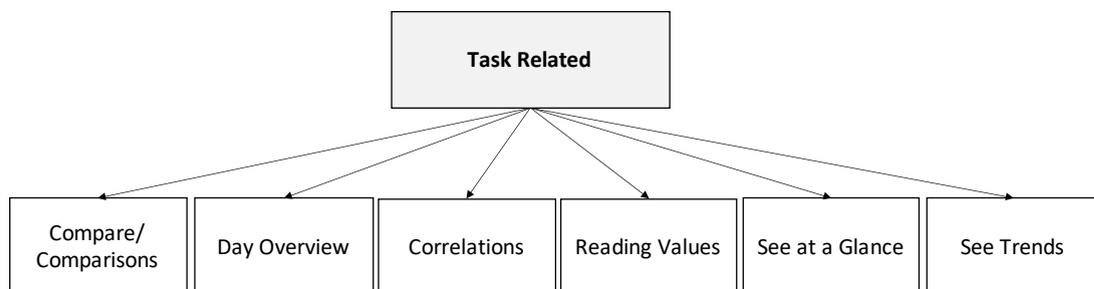


Figure 7.5: Task Related Feedback. It is a sub-theme of 'Readability and Graph Comprehension'.

and 'not simple'. For example, the response by participant P208 included both positive and negative comments regarding the readability of different visualizations. P208's comment was: "Linear and bar overlapped are very clear and readable and show it all in one graph while radial is very hard to read and understand".

Some participants' responses were more related to the tasks after the experiment such as 'finding correlations between the data', 'provides day overview' and 'precision and accuracy'. The

types of responses that were related to the visual tasks are summarised in Figure 7.5. Moreover, the task related responses are discussed in detail later in section 7.5.2.4.

Aesthetic Aspects of the Design: some of the participants considered the aesthetic features of the visualizations (20/80, i.e., 25% in the pre-study and 16/80, i.e., 20% in the post-study survey) as shown in Table 7.7. All of the following terms have been included in the coding in this theme: ‘attractive’; ‘aesthetically/visually appealing’; ‘looks good/nice’ and ‘tidy’. For example, P105* commented on the Radial charts: "they look nice, and are maybe nice to use as infographics or to glance at for trends". Other comments mentioned the clock face, the symbols, the colours, the parts of the day or the quarters (i.e., morning, afternoon, evening and night). For example, P710 said: "I like the idea of radial chart (looks like a clock)". With respect to the glyphs used in the Radial design, P501 commented on the Radial Bar chart by. "Those charts are quite easy to read and the design tell me directly the time of the day. I can easily imply what they mean just by seeing the symbols on the charts". The negative responses mentioned for example, the brightness of the colours or that it was hard for their eyes to adapt to the used colours.

These highlighted terms, which were used by the participants, used in different context and for different visualizations. For example, the term ‘attractive’ was used by P105 to describe RSBC visualization: "That one was most attractive because the round shape made it easier to instantly look at and see where on a clock face the activity fell", while P507* used it to refer to the Linear Line chart: "simple, easy to read the overall result, the line is attractive and in term of monitoring your physical activity is more motivation".

Design Features: in this theme, all the responses that discussed aspects related to the basics of the visualization designs were included, such as the fact that some visualizations were compact and combined the data variables together while others were separated into different charts, the scales and the axes, as well as other features such as familiarity and user-friendliness. P302 justified the selection of LOLC by saying: "has a background grid, all graphs combined and therefore easy to compare", and P606 commented on the Radial Bar charts "enough clear and user-friendly". The theme was identified by 23/80 participants, i.e., 29% prior the study and 16/80, i.e., 20% after the study.

Effectiveness: the effectiveness theme included all the responses that discussed the efficiency

in the visualizations, such as ‘inefficient’; ‘effective’; ‘quick’; ‘long time’; ‘less/more time’ and ‘less/more effort’. The efficiency was discussed by only 6/80 (8%) and 5/80 (6%) of the 80 participants in the pre- and the post-survey respectively, as presented in Table 7.7.

Related to the Dataset: this theme involves all the justifications that are more related to the dataset such as: appropriateness; suitable or not suitable for the dataset; time reminders; change over time and providing an overview of a day. Moreover, two participants were more specific and commented on the chart used for the type of data. For example, Participant (P504) stated: “The line charts suggest continuous data rather than hourly data” while (P710) pointed out: “The bar plots helped getting a better visualization of quantities”. This theme was discussed by 9/80 (11%) and 6/80 (8%) participants in the pre- and post surveys.

Suggestions and Recommendations: few participants (4/80 and 3/80 participants in the pre- and the post surveys respectively, i.e., 5% and 4% of the participants) proposed some modifications to the used visualizations, for example, in the pre-experiment survey, participant P109 suggested having a combination of the bar and the line chart in the visualization, the bar chart is to represent the steps and the calories and a line chart to represent the HR data. Whilst participant P101 proposed representing the data accumulatively (i.e., representing the progress as total steps taken and calories burned from the start of the day until the current hour) rather than on an hourly basis. Regarding the Radial visualizations, participant P409 suggested splitting the layout into two separate graphs: one for the a.m. and the other for the p.m. to improve the readability by reducing the number of the data points and P602 recommended increasing the colour distinction between different categories.

In the post-experiment survey, participant P409*² stated that he still liked the Radial visualizations because it was attractive but reading the data from it was not easy. Therefore, the participant provided two suggestions to improve the readability of the Radial visualizations: firstly, “making more reference points along the clock”, i.e., adding another axis at the bottom of clock equivalent to the one at the top; secondly, to make the “Radial reference lines thicker making it easier to track the reference points” or using “different coloured lines for certain reference lines,” i.e., using different colours for the lines presenting the numbers (5, 10, 50,

²The symbol * is used after the participant ID to refer to the post-experiment comment, and pre-experiment comment otherwise.

100), something which the participant also proposed to be used for the Linear visualizations to enhance the readability. Participant P505* asked if it was possible to use both bar and line chart as each one provided a different perspective: “Line charts give you a better idea of the trend during the day, while bar charts help to get a more precise value at a specific hour. Then, why do you not overlap these two diagrams in one?”. Another participant, P710*, proposed to have the same coloured sectors as in the Radial layout as well as the symbols that indicate the parts of the day in the Linear layout visualizations.

Personal Preferences: in the pre-experiment survey, 3 of the 80 participants refer their choices to their personal preferences and what they usually liked in the visualizations, such as: P102 “I generally prefer the look of line charts”; P302 “I like to compare graphs and see them combined”, “I don’t like to have the data separated” and P104 “I prefer each variable to be on a different graph”. Other personal preferences were also identified in the post-experiment (by 3/80 participants) as follows: P205* “just prefer the line chart”; P806* “I think I much prefer the bar charts”; P710* “I much prefer the design of the line chart” and “I really like the design of the radial chart”. The total number of participants who mentioned comments specifically about the Radial visualizations were 30/80 (38%) in the pre-experiment and 31/80 (39%) in the post-experiment and they were divided into positive, negative, and suggestions.

A Reflection on the Online Survey in Chapter 3: in the study outlined in Chapter 4 people were asked about their justifications in their visualization preferences. The identified codes (see Figure 3.4) share some similarities with the codes here in this chapter (see Figure 7.4 and Table 7.7). These codes are: “easy to read and understand”; “Familiar”; “Aesthetic” and there were also some comments that related to the dataset i.e., “related to the nature of the data”. The outcomes of the survey in Chapter 3 also showed that “easy to read and understand” was the most frequently mentioned code in the responses. This is also conforms with the outcomes of the surveys in this chapter, as the “Readability and Graph Comprehension” code had the highest frequency in both surveys as shown in Table 7.7.

7.5 Discussion

The discussion is mainly organised to answer the research questions presented in section 7.2. The first two questions are related to visualizations ratings and preferences which were analysed in section 7.4.1 and 7.4.2, the results of which are discussed in section 7.5.1.

The third question in section 7.2, the results of which were analysed in section 7.4.2, is related to the information the participants provided regarding their preferences, which is discussed in section 7.5.2. After identifying the main themes in the responses that were considered as the main reasons for preferring the selected visualization(s), the analysis led to the investigation of and an understanding of the implications of the participants' responses regarding the main design aspects in the developed visualizations, specifically the layout, the visual encoding and the visualization method in addition to their relations to the visual tasks. Each of these aspects are addressed in sections 7.5.2.1, 7.5.2.2, 7.5.2.3 and 7.5.2.4 respectively. The results of the participants' subjective feedback were used in a discussion of broader suggestions for designing visualizations for personal health data.

7.5.1 An Overview of Visualizations Ratings and Preferences

7.5.1.1 Visualizations Ratings

The analysis of the visualization ratings in section 7.4.1 presented the differences in each visualization ratings. This analysis was performed to understand how the experiment's tasks impacted on the overall rating of the developed designs. The main research question was to investigate if there was a statistically significant difference in the participants' ratings of the readability and attractiveness of the visualizations before and after the experiment. The results of the analysis (presented in Table 7.2) showed that only the rating of the Linear Bar charts significantly increased in terms of the readability and attractiveness. The LSLC and the RSBC rating significantly decreased in terms of the readability only.

Although only few cases showed significant differences in the overall ratings of the visualizations, Figures 7.1 and 7.2 demonstrate interesting changes in how the degree of ratings for

each visualization differed after the experiment. For example, the LOLC had better readability ratings after the experiment, but were generally between ‘good’ and ‘acceptable’. The readability of the Radial visualizations was generally ‘poor’ and ‘very poor’ after and before the experiment, while they received better ratings with respect to the attractiveness of the design as shown in Table 7.2. The outcomes showed that performing the tasks on the visualizations affected the overall ratings; however, the change was only significant in some cases, as reported earlier.

7.5.1.2 Visualizations Preferences

Regarding the research question related to the visualization preferences (i.e., the second question in section 7.2), the participants’ preferences significantly changed after the experiment with only two designs. These two visualizations are the LOBC, which was significantly more preferred after the experiment and the RSBC, which was significantly less preferred after the experiment. The outcomes also showed a relation between the tracking behaviour and the preferred visualizations as discussed in the following section.

Preferences and the Used Tracking Devices: the participants responded to the preferences question in the pre-experiment survey based on their knowledge and prior experience of working with graphs. Although the number of participants who track their health data was higher than the participants who do not (52 and 27 respectively), it can be seen that the Linear Overlapped charts: LOBC and LOLC, were more preferred by the non-tracker participants (see Table 7.6). LOBC and LOLC were favoured by 8 and 11 out of 27 non-tracker participants respectively (that constitutes 30% and 41%) while they were only preferred by 6 and eight out 52 tracker participants respectively, being around 12% and 15% of the total tracker participants. On the other hand, the stacked visualizations LSBC and LSLC were more preferred by tracker participants (28 and 13 out of 52 participants respectively, i.e., 54% and 25%), while they were less preferred by participants who tracked their data (9 and 7 participants out of 27 respectively, i.e., 33% and 26%).

It can be argued that the participants who track their data might be more familiar with the Linear stacked visualization, more specifically the LSBC (as the figures in Table 7.6 confirm),

as it is one of the most popular visualization designs that is supported by most health tracking apps/dashboards, see the review in Chapter 4, unlike the overlapping visualizations presented in the study.

7.5.2 Insights from People’s Feedback on the Visualizations

The third research question outlined in section 7.2 aimed at gaining a deeper understanding of the participants’ preferences and their feedback analysed earlier in section 7.4.3. This structured qualitative analysis of the responses helped in demonstrating the effects of each implemented visual element. Therefore, the following sections are focused on the main visual elements evaluated in the performance analysis in the previous chapter (e.g., layout and visual encoding) and also reflect on the responses with respect to the visual tasks.

7.5.2.1 An Overview of the Visual Layout

The type of Layout is a major factor in this thesis, this was specifically discussed in Chapters 5 and 6. The participants’ subjective feedback related to the type of layout is discussed in the following sections.

7.5.2.1.1 Linear Layout

The Linear layout was generally more preferred by the participants than the Radial layout according to the discussion in section 7.4.2. In the pre-experiment survey, the Linear visualizations were preferred by 68 of the 80 participants (85%), increasing to 77 participants (96%) in the post-experiment survey. The comments on the Linear layout were mainly regarding a specific type of design. For example, P806* said: “I prefer the linear bar chart because I can understand it better than the rest of the charts” and P608* commented: “Line stacked Line chart - not precise”. These factors are discussed in sections 7.5.2.2 and 7.5.2.3.

7.5.2.1.2 Radial Layout

In the pre-experiment survey, Radial visualizations were selected by 21 of the 80 participants (26%), while in the post-experiment survey, the number of participants who included a Radial visualization in their preferences decreased to 16 participants (20%).

As the analysis in the previous sections showed that the Radial design was more preferred by the participants who track their data, the Radial design comments here are classified according to the 'trackers' and the 'non-trackers' participants.

Readability: the readability of the charts included different aspects, as were discussed in section 7.4.3.1. For example, P205 stated: "I don't understand the radial visualization". Although some participants found the Radial layout readable (7/80 pre and 1/80 post), e.g., P501: "Those charts (RSBC and ROBC) are quite easy to read and the design tell me directly the time of the day". However, most of the readability aspects were mentioned negatively (20/80 pre and 30/80 post) such as by P207: "I found the radial charts hardest of all"; P303*: "very hard to read" and P306: "difficult to read".

Attractiveness: many participants mentioned the attractiveness of the Radial design and in some cases, these comments appeared alongside comments regarding the difficulty of reading the graph. For example, P301 commented: "the radial overlapped line chart is a bit difficult to read but is attractive and innovative"; P102 said: "I can see that the radial charts may be considered attractive but they appear a bit gimmicky rather than presenting useful data". Another participant, P105*, considered the Radial design good as an infographics or to read at a glance saying: "they look nice, and are maybe nice to use as infographics or to glance at for trends, but not to actually get numbers from".

Clock Face/ Relates to Time/ Coloured Sectors/ Symbols: one of the major advantages mentioned about the Radial layout is that it relates to time, e.g., P401*: "The radial charts were hard to read and determine values but gave good indication of change over time". Also that it links to a clock and that it reminds on a watch such as P105: "the round shape made it easier to instantly look at and see where on a clock face the activity fell"; P302: "reminds on a watch which gives a relation to the time"; P710: "I like the idea of radial chart (looks like a clock)". Some participants (1/80 pre and 2/80 post) also commented on the glyphs and the coloured sectors that were implemented to refer to the parts of the day, P501: "Those charts (RSBC and ROBC) are quite easy to read and the design tells me directly the time of the day. I can easily

imply what they mean just by seeing the symbols on the charts”. P302*: “the differentiation into quarters makes it easier to see in relation to time”. P710*: “I really liked the design used in the radial chart and it helped clearly divide the day”.

Implementing these concepts of the day’s parts into segments and the use of glyphs in the visualizations with a Linear layout was suggested by P710*: “Maybe it would have been good to have in the background of bar and line charts the same as in the radial charts, that is the indication of morning, afternoon, evening,... I really liked the design used in the radial chart and it helped clearly divide the day”.

Glyphs are found to be used by most health tracking apps/dashboards to support additional information or to refer to specific health variable. For example, the small fire icon used to refer to the burned calories in the Fitbit app and dashboard [72], the step foot icon used for step count and walking and running icons to reflect the level of activity in the Garmin app [89].

Having only one axis: one of the comments mentioned by the participants is that it has only one axis which made it difficult to follow the reference lines to the values on the axes. Especially if the target was far from the axis (e.g., at the bottom part of the circle), P408 commented: “the radial chart is even more confusing since the reference points are not clear and you have to follow the reference line (for example 500) all around the circle to know the actual value”; P510*: “radial is tiresome to read because the x axis is potentially quite far away from the measure you’re trying to read” and P302: “values are hard to read as there is only one axis for data”. The effect of the angle on the accuracy with the Radial chart in comparison with the Linear chart was also identified by one of the participants, P609* said: “having to follow a curve around to the scale on the radial graph instead of gazing horizontally the scale on a line or bar graph was more difficult”.

Reflection on the Quantitative Results: the participants were aware of the difficulty of reading values from the Radial layout or being accurate with it such as P501* who said: “The radial charts are more attractive but it’s a bit hard to read the number” (more details are presented in section 7.5.2.4). It is assumed that one of the reasons could be that it affected the accuracy of the Radial Line charts because the Lines were not salient with the background colours. This was raised by P707*: “Radial Overlapped Line Chart they are attractive however they are so hard to read I think maybe because the lines are not clear with the background colours”. An-

other issue that might have affected the performance of the tasks that required reading values or estimations is following the lines from the value position toward the axis (which is related to the one axis issue discussed earlier). An example is what was mentioned by P104: “more confusing especially the radial diagrams I found it hard to follow the lines to get the correct answers”.

Do People Like Using the Radial visualizations? Some of the previous studies that evaluated the Radial layout (see section 2.3.3.1) collected different subjective data from the participants about the evaluated Radial designs. For example, Fuchs et al. [79] reported that their participants liked the design of the 24-hour clock glyph design. On the other hand, Brehmer et al. [32] reported that in general the participants preferred the Linear layout over the Radial one, likewise, Waldner et al. [238]. However, it is hard to compare their participants’ feedback about the Radial layout with this study due to the significant differences in the design and the applied task that may have affected people’s opinions.

In Fuchs et al.’s study [79], the participants argued that they faced some difficulties when comparing lengths on the Radial Star glyph (i.e., similar in the design of the Radial Line chart in this thesis) because of the different orientation. They stated that this feedback from the participants supported their hypothesis that mental rotations might be required when performing comparisons and this could make the tasks that require value comparisons harder on these Radial representations. However, in this current study the participants reported that they liked the clock metaphor for temporal location tasks and some suggested using the 12-hour rather than the 24-hour encoding for intuitive visualization [79]. The suggestion of splitting the 24-hour display into 12-hour displays for a.m. and p.m. was suggested by participant P409. However, Waldner et al. [238] implemented both the 24 and the 12-hour on the Radial and Linear layouts. Their results showed that the 12-hour Radial design was an ineffective for several tasks such as the reading values, locating maximum, value comparisons. In addition, the participants reported it was hard to read, and some even did not find a connection between the clock layout and the 12-hour display.

The difference in collecting the participants’ preferences in those studies and the study this thesis presents is that the participants were allowed to choose and rate the visualizations before conducting the experiment to avoid possible bias judgement that could be as a result of imple-

menting a specific set of tasks on a fixed dataset. The participants were then asked the same questions after completing the experiment, this was done to obtain more results that which would focus attention on the efficiency of the graphs and enable a greater understanding of how the preferences changed after the use of the visualizations.

Familiarity with the Visualizations: Familiarity is an important factor that could significantly affect the participants' choices. People tend to be more capable and comfortable when the data is visualized in the way they are used to seeing it, even if the new design is better [82]. In a study by Zhao et al. [248], the expert participants were confident enough to use the basic interaction features and familiar visualizations, but were not able to begin using the new provided features and representations. Zhao et al. stated that breaking users' habits is a significant challenge in visualization design as people feel confident with the visualizations that they are familiar with, which, in turn, affects whether the visualization is judged to be effective or not. However, in [248], the Radial layout received positive feedback from the participants after they learned it. Maçãs and Machado [141]'s case-study for comparing the Linear and the Radial calendar layout stated that the participants were more familiar with the traditional Linear layout, but after some interaction they preferred the Radial layout, saying it was better in summarising and presenting the information, and brought more insights. Maçãs and Machado also explained that the fact that the Linear calendar representation in their study is more familiar to people might result in a tendency to select it as most useful and easier.

In the participants' justifications for their preferences, the familiarity with the visualization was mentioned by 4/80 participants prior the experiment (three of them are trackers). In all these four responses, they were related to LSBC design. Representing the data individually using a Bar chart on a Linear time axis is a common form of representation on health tracking platforms as the review in Chapter 4 shows. After the experiment, one participant stated that the Radial design is new to her.

According to the previous discussion, the following findings regarding the first question in the discussion were identified. The participants' respective comments regarding the Linear layout were mainly focused on specific designs. The major issues identified with the Radial design were that it has only one axis, which made it difficult to be accurate with values retrieval or estimation for the items that were far from the axis.

7.5.2.2 An Overview of the Visual Encoding

Participants preferences in the Linear Bar charts encodings increased after the experiment (significantly for LOBC $P=.001$). However, for the Linear Line charts, the preferences decreased (as shown in Table 7.3 and Figure 7.3). After scanning through the textual feedback, looking for comments regarding the Bar/Line charts, the comments were generally identified to be related to the encoding uses (i.e., for what task), or related to whether the encoding was suitable for the dataset, as set out below.

In relation to the type of data being visualized, some participants considered the Bar chart more suitable for the dataset as P109 stated: “I thought bars are the most appropriate form of visualization for the variables”, or for representing quantities P710*: “The bar plots helped getting a better visualization of quantities”. Others thought the Line chart was more suitable for continuous data, e.g., P504: “line charts suggest continuous data rather than hourly data” and P709: “I prefer a line graph over line graphs generally as data over time can be visualized easier”.

With respect to the tasks (more details are discussed in section 7.5.2.4), the Bar chart was generally described before and after the experiment as good for reading values, while the Line chart was considered more suitable for trends and change overtime, as stated by P102: “The bars were much clearer for me to read exact values than the lines or the radial charts. Whilst I generally prefer the look of line charts and they are easier to follow trends”. More details regarding the visualizations with a reflection on the tasks are presented in Table 7.8. The Line chart was also considered as motivating in its representation of physical activity data because it showed progress and changes over time, e.g., P507* stated: “the line is attractive and in terms of monitoring your physical activity is more motivation”.

These comments from the participants reflect their performance with the tasks discussed in Chapter 6; these are the outperformance of the Bar encoding within the Linear layout for T2 that required estimating values and the outperformance of the Line encoding for T1 that required identifying trends. Preferences and performance were discussed in previous studies. For example, in Saket et al. [186]’s study, that implemented different visualizations for tabular data including Bar and Line charts. The participants were asked to select their preferences for each

task. They identified a positive correlation between the tasks' accuracy and the participants' preferences (i.e., the participants preferred the visualizations that helped them to perform the task accurately).

7.5.2.3 An Overview of the Visualization Method

As mentioned previously throughout the thesis, the aim of implementing two visualization methods was to assess the effectiveness of overlapping and stacking the variables in the multivariate dataset to understand the relationships between them. The visualizations preferences in section 7.4.2 showed a significant increase in the preferences of the Linear Overlapped design (LOBC) after the experiment $P=.001$ (preferred by 15/80 participants pre- and 34/80 post- experiment). The feedback regarding the Overlapped and the Stacked designs focused on the ability to compare the data with the graphs such as P401* who selected LOBC and LOLC as their preferred visualizations and suggested that those designs “helps to critically analyse and compare the results”. Table 7.2 shows that the Overlapped design was suggested by the participants for comparing data. However, some of the participants did not like the idea of the Overlapped design because they found it confusing and it increased the difficulty, e.g., P104 stated: “It looks the clearest and I prefer each variable to be on a different graph. The other visualizations were more difficult to read and looked more confusing”. This confusion aspect was also reported by the participants in Li et al.'s [135] work, who collected multivariate data and used paper graphs to combine the data. Those participants encountered other difficulties as a result of multiple data, including confusing and understanding how data influences behaviour. However, participants experienced other setbacks: exploring multiple types of data was confusing and understanding the influences of the factors on their behavior was difficult.

7.5.2.4 Reflection on the Visual Tasks

Whilst coding the responses, some of the comments were found to be related to different types of tasks, e.g., to reading and retrieving values and comparisons. These were classified under the “Readability and Graph Comprehension” theme. This section presents a detailed discussion of the specific tasks mentioned by the participants, and which visualization was recommended for

each type. These tasks are classified as the following: (1) Comparisons (2) Reading/ Retrieving values (3) Finding trends/fluctuations (4) Finding anomalies (5) Analysing data (6) Correlations (7) Read at a glance/Overview.

- **Reading Values:** This task was mentioned by several participants before and after the experiment (Pre: $n=3/80$, Freq=5, Post: $n=22/80$, Freq=25). In addition to ‘reading values’, the participants referred to this task using different phrases including: ‘read exact values’, ‘extract data points’, ‘read and determine values’, ‘read and determine values’, ‘pinpoint the value of a point’ and ‘to read particularly if you want to the particular time’. The participants also provided some comments on the best visual encoding for this task. For example, the difficulty of reading values on the Radial visualizations and the Linear with the Line encodings. As P207 stated: ‘I found this one’ (i.e., LOBC) ‘the easiest and clearest to read and crucially compare the data. This was harder with the linear line charts and nearly impossible with the radial charts’. In Table 7.8, the participants’ feedback for what visualization could be suitable for the reading value task is presented. Only the visualization/encoding that was specifically mentioned to be good/not good for reading values were counted (comments such as easy to read or clear were not counted).
- **Comparisons:** Using the visualizations for data comparisons was mentioned by (Pre: $n=13/80$ Freq=13 Post: $n=11/80$ Freq=11). The compare task could differ according to the number of items being compared. The Compare task implemented in this thesis is a comparison of overall data between two different days. The participants in their feedback referred to different forms of comparisons such as comparing data items: “to read and crucially compare the data”, and general such as: “easy to compare” or: “easy for comparison”. The participants generally appreciated the Overlapped design for comparing the data variables such as: “prefer overlapping charts independent of style as they allow an easy comparison between the data”, further details on the data are shown in Table 7.8.
- **Correlation:** Finding correlations between data was mentioned by few participants before and after the experiment (Pre: $n=1/80$ Freq=1 Post: $n=2/80$ Freq = 2). This shares some similarity with the compare task as it referred to looking at more than one data items, and it relates to the Overlapping designs: “overlapping allows you to relate one

variable to the others”. The term correlation appeared in the post-experiment comments as P107*: “easier to make correlations between the steps, calories and HR”, and the other as: “easy also to relate with other variables”.

- **Trends:** It is related to seeing trends in the data and was mentioned by (Pre: n=2/80 Freq=2 Post: n=3/80 Freq=3). The comments regarding the trends generally preferred the Lines for this task, and only one participant suggested the Radial design.

7.6 Summary of Findings

The following points summarise the major findings identified from the analysis of the participants’ subjective feedback. Some of the findings reflect on the quantitative results presented in Chapter 6.

- The Radial visualizations were not highly preferred in comparison with the visualizations with the Linear layout before and after the experiment. The preferences in the Radial visualizations decreased after the experiment.
- The participants who track their data might be more familiar with the Linear stacked visualization, more specifically, the LSBC. Hence, it is one of the most popular visualization designs that are supported by most health tracking apps/dashboards, unlike the overlapping visualizations presented in the survey
- The change in the participants’ preferences reflect their performance with visualizations. For example, after the experiment, the preferences in the LOBC has increased. LOBC was found in Chapter 6 to be a good visualization choice for T5 the ‘Locate’ task. Also, for T2 the ‘Determine task and T4 the ‘Infer’ task when performing the analysis across both layouts.
- The total number of participants who preferred the Linear Bar chart designs (LSBC and LOBC) increased after the experiment.

Table 7.8: Task Related feedback. This table shows the tasks extracted from the participants' feedback. The first column presents the task name, the second presents some feedback examples with the selected preferred visualization, the third and the fourth columns show what participants considered good and not a good design for the specified task (the numbers for how many times appeared in this context [pre, post]), the last column reflects on the findings from Chapter 6 regarding identified best design for a similar task (in all the analysis cases).

Task	Textual Feedback examples	Good	Not Good	Best Performance
Reading Values Pre: n=3/80 Freq=5 Post: n=22/80 Freq=25	P102 (LSBC): "The bars were much clearer for me to read exact values than the lines or the radial charts. Whilst I generally prefer the look of line charts and they are easier to follow trends, I think in this case it was more difficult to read values as specific times from the graphs presented here", P109* (LOBC): "The line chart is acceptable, but still problematic when it is necessary to get a correct estimate of data", P409: (LSBC) "The Radial Stacked bar chart looks amazing at first, but it is a fair bit harder to extract data points from such a graph"	Bars ^{1,4} Lines ^{0,1} LSBC ^{0,3}	Lines ^{1,3} Radial ^{1,11} RSBC ^{1,1}	LSBC
Compare Pre: n=13/80 Freq=13 Post: n=11/80 Freq=11	P306 (LOBC): "Easy to read and to compare data", P703* (LOLC): "when you compare more than one physical activity it is better to use V(4)", P608* (ROLC): "To compare the data is also easy on radial overlapped line chart as it gives you an idea throughout the day"	LOLC ^{6,3} LOBC ^{3,4} LSLC ^{1,1} LSBC ^{0,1} ROLC ^{0,1} ROBC ^{1,0} Lines ^{1,0} Overlap ^{0,1}	Radial ^{2,0} Lines ^{1,0} LSLC ^{0,1}	LOLC
Correlation Post: Pre: n=1/80 Fre=1 n=2/80 Freq=2	P107* (LOBC) "easier to make correlations between the steps, calories and HR", and P410* (LOLC) "easy also to relate with other variables"	LOBC, LOLC	-	-
Trends Pre: n=2/80 Freq=2 Post: n=3/80 Freq=3.	P505* "Line charts give you a better idea of the trend during the day", P309: "The radial overlapped line chart was significantly easier to see any fluctuations and/or anomalies"	LOLC ^{0,1} ROLC ^{1,0} Lines ^{1,1} Radial ^{0,1}	-	Linear Line charts

- The total number of participants who preferred the Linear Line chart designs (LSLC and LOLC) decreased after the experiment. The major comment identified was that the Line chart is not good for reading values.
- The total number of participants who preferred the RSBC decreased after the experiment, while the number of participants who preferred the ROBC increased after they took part in the experiment.
- The Radial Line charts (RSLC and ROLC) were the least preferred visualizations between the visualizations. As they were selected by one participant before the experiment and 2 participants after the experiment. These two visualizations were not good in general, but in a few cases were found to be better than other Radial alternatives (as presented in Table 6.30 to Table 6.34 and Table 6.35).
- The majority of the participants agreed on the difficulty of reading the data and solving the tasks with the Radial layout; therefore, the Radial visualizations were less preferred than the Linear ones and had a lower readability rating. This is in accordance with the performance results. However, some participants liked the concept of the clock layout in terms of its relation to time. Some participants suggested the Radial design should be used for providing a day overview or used as an infographic, but not for retrieving specific data.
- Overlapping the data into one graph should be an option, but not a basic design feature. In some cases, it helped in the task analysis and the participants also reported that it was a good design for comparing the physical activity data. However, it could also be confusing and hard to read, especially with a high-density dataset and could generate visual clutters which was reported by one of the participants as 'heavy', regarding the Overlapped Bars.

Chapter 8

Discussion

8.1 Introduction

In this chapter, we discuss the outcomes of the research implemented throughout the thesis. The chapter is organised according to the research stages and starts by discussing the visualization review presented in Chapter 4. It reflects on recent updates of the visualization methods on the devices' apps and dashboards reviewed in light of the limitations raised earlier in the thesis, the outcomes of our evaluation if applicable and the literature. Then, with respect to the development of the visualizations and the evaluation stage (presented in Chapter 5, 6 and 7), it discusses the significant differences between the implemented designs and similar design in the literature and what supported by the tracking devices. The chapter ends by highlighting the research significance and providing design implications, recommendations and insights for the visualizations personal health data that were derived from the evaluation study.

8.2 Visualization of Tracked Health Data

The visualizations updates show some enhancements towards addressing the challenges such as the reflection on past data and overlapping data for comparisons. However, the challenges were not fully addressed. Although Fitbit and Garmin support online dashboards, they still do not support advanced filtering features for the visualizations. The limitations and the challenges discussed in both Chapter 4 and in this section are focused on the data visualization only. Other challenges with respect to the accuracy, security and privacy aspects of the data, data ownership, users' access and control over the data and data integration from multiple sources are not considered. This section discusses the recent updates of the visualizations on the devices' apps and dashboards reviewed in Chapter 4 in light of the raised challenges and limitations, the results of our study and the guidelines and implications if applicable and related literature.

Multivariate Data Visualizations: In section 4.5.1.3, the visualizations that support representation of more than one data variables were discussed. The visualization review in Chapter 4 concluded by listing several limitations that need to be addressed in terms of data visualizations (see section 4.6), including the need for more options to combine different aspects of health data in the visualizations. In addition to the visualizations discussed in section 4.5.1.3,

the following visualizations were identified: The Apple Health app offers an overlapping of one data type (e.g., steps) with its average which allows the user to compare the day activity level with the average, as presented in Figure 4.10, which appears as two Overlapped Linear Lines. According to our design guidelines (in section 8.4), the Linear Line chart encoding is a good visualization choice for comparing data and identifying trends.

Our quantitative and qualitative results showed some efficiency and benefits of using the Overlapped representation. The findings showed that Overlapping the data into one graph should be an option but not to be used as the basic visualization design. It helped in some cases in the task analysis and the participants also reported that it is a good design to compare the physical activity data. The preferences in the LOBC design has also significantly increased after the experiment. However, several limitations still exist in providing users with options in multiple data visualizations.

Moreover, the visualized data must not be restricted to health data variables only as many other non-health-related factors could affect personal health. For example, Huang et al. [115]’s calendar visualization that integrates a visualization of physical activity data with a personal calendar provided the users with helpful contextual information about their activity. Although some of the reviewed visualizations support calendar view of physical activity data (e.g., the calendar view on Garmin dashboard that presents training logs), these calendars are standalone and do not integrate with personal calendars to show related data, such as events or work activities.

Overview of Long-Term Data: One of the challenges mentioned in section 4.6 is the lack of visualizations that support a comparison with past data or the reviewing of the data over a long period of time (see section 4.5.1.2). There are different understandings and consideration for long-term and short-term personal data in the literature. For example, Li et al. [135] considered short-term data as the data of “an hour to a day”, and other than this is long-term data. Huang et al. [115] stated that some participants used their feedback tools for short-term goals, which are “daily and weekly” and others are long-term goals. Meyer et al. [149] found that for visualizing the past data, the maximum length of data to be visualized is between 2 and 6 months. However, the study in this thesis included only short-term data representing hourly data for one day and a comparison of two days. Long-term data is considered as the data that is longer than

one week.

The recent update on the Activity app (the companion app with the Apple Watch) provides trends that allow the user to compare the current activity level with past data. However, this comparison is based on an average of the past data as shown in Figure 4.10 and requires the collection of data for at least six months in order to calculate averages. This presentation is suitable for people who want to compare their current activity level with their past data. Epstein et al.'s [63] study showed that people preferences regarding their physical activity feedback are correlated with the period of using the activity tracking tool. People who used their tracked tool for a long period preferred presentations that show timelines of their tracked data aggregated by month or year. Moreover, in the study by Meyer et al. [149], it was indicated that users are interested in discovering more complex correlations in their health data that could not be identified using simple visualization methods over both short and long time frames. This reveals a necessity for more advanced methods for representing and analysing long-term data as they are currently mainly visualized using statistical charts.

Data Sharing: The Samsung Health app allows the adding of friends to the app to compare their activity levels during the day and to compete with each other. The app also allows for comparing the average of all users, or with users in the same age group. Adding friends to the dashboard remains an available feature on both the Fitbit and Garmin. No changes in sharing options were identified in the Apple Health and Activity app (other than that which was discussed in Chapter 4). The reviewed platforms do not yet support high personal encoding (e.g., by using abstract representation as with the UbiFit [49]) for social sharing purposes.

Using abstract visualization for sharing purposes is important as it makes the data difficult to read by other people, especially when privacy is a concern; as is the case with personal data [114]. The data from the survey carried out in Chapter 3 showed that people could accept abstract visualization which could be used for sharing purposes more than for their own personal use. However, the control over the data for sharing is limited. An important consideration that needs to be considered when designing for sharing is the need to be aware of the possible negative consequences that could result from sharing health and activity data, such as stress, guilt or perceived failure [114].

In Rapp and Cena's [177] study, the participants raised limitations related to their control over

the data. Although the majority of the participants considered the data to be extremely private and not appropriate to share on social media, some participants showed interest in sharing with specific people (family and friends) for comparisons or sharing a subset of the data. However, social interaction for health data has been applied in the literature to increase motivation and behaviour change, as in Fish'n' Steps [139] which allows setting activity goals for groups.

As people have different preferences and have different purposes and as the physical activity data can be shared with different audiences such as family, friends, publicly on social media or sharing for medical purposes, the visualization should not be restricted to one or two types of visualizations. The question here is how to provide adequate visualization with an adequate level of interaction for the required type of sharing with respect to privacy.

Personalisation and Customisation: The reviewed visualizations have limited options in terms of customising the data and how it is being visualized. For example, the Fitbit app and dashboard allow its users to customise the interface by adding and removing whichever data is to be presented, but there is no control on how the data is visualized. Kim et al.'s [125] study, which investigated how people design personal visualizations for their data, highlighted the benefits of giving the users the ability to personalise their visualizations. They found that the process allowed the participants to visualize the main goal of the data collection and how the data affected their lives. In Rapp and Cena's [177] study, the participants reported the personalisation of the visualizations to match their needs and that the tracking goal and purposes were desired features.

The lack of customisation is also present in the multiple data visualizations, as there are no options to allow the user to select which health or activity data is to be visualized. For example, the Apple activity app provides the visualizations of three Radial Rings (presented in Figure 4.3), but the user cannot customise exactly which data the rings visualize (e.g., choosing to view the step count goal instead of standing hours).

The results in Chapters 5, 6 and 7 showed significant differences between the performance with the visualizations with respect to the different tasks, and the provided design guidelines suggest different forms of representation for different tasks and purposes, and there is no one optimal visualization for all the tasks. The participants' subjective feedback also showed significant differences in the preferences and designs ratings before and after using the visualization and

between the participants as well. This finding is considered critical in self-tracking since the users are often restricted to the limited supported visualizations in case the user does not prefer them. Many previous studies which have investigated people's experience with the self-tracking (such as [44, 134, 218]) showed that that people faced different difficulties in benefiting from the data due to limitations in its representation and exploration methods. The dedicated trackers surveyed in [44] mostly used a spreadsheet or built their own visualizations to represent the data. In Li et al.'s [134] personal informatics model, they found that problems in the visualizations or difficulties in using them effectively during the reflection stage prevented users from moving on to the next stage: the action stage.

Tracking Goals: Although some people start tracking their data purely for curiosity without having specific goals, others track their data for a specific reason such as behaviour change [62], or to identify factors that might affect their health [135]. The reviewed devices allow the setting of some tracking goals, such as a daily step count and a sleep schedule with Fitbit or setting daily active minutes goals with the Apple watch. However, the user experience with the tracking remains limited to the provided charts, which may not be useful for inferring relationships between the data. An exception could be the Garmin connect app and dashboard as they provide different multiple data visualizations, but they mostly represent short-term data such a single day or an exercise.

8.3 The Developed Visualization Designs

This thesis evaluated the effectiveness of eight different visualization choices for multivariate personal health data. The process of developing the visualizations started from the early stages of the research that inspired the designs discussed in Chapter 5. As there are some similarities with similar designs in the literature (that were discussed earlier in Chapter 2) and the visualizations of personal health data on apps/dashboards (that were discussed in Chapter 4), this section discusses the main differences.

8.3.1 Differences from the Visualizations on Apps/Dashboards

One of the limitations concerning the visualizations on dashboards and apps is the lack of visualizations that combine multiple personal data (see section 4.5.1.3). It is helpful to provide a better context in short-term data representations and to help in understanding the impact of factors in long-term data representation.

The main layout used to represent the data is the Linear layout. The Radial charts and a clock metaphor have been adopted by several apps and dashboards, but these are different from the developed Radial visualizations. The differences between the implemented visualizations and the designs of multiple data visualizations reviewed in Chapter 4 (see the Figures 4.6, 4.7 and 4.3a) are as follows:

1. The Apple Activity app (Figure 4.3a), which is connected with the Apple Watch, represent three variables (active calories, exercise minutes and standing hours). These three variables are represented in a Radial layout (daily goal completion rings and are not based on time) and in a detailed linear layout on an hourly basis, which is similar to the LSBC design in this study. Besides the difference in the visualized data types and the design of the Radial layout, this study investigates the effectiveness of the overlapped design (shared space) on the two layouts.
2. The design of the Garmin multiple data visualizations presented in Figure 4.7 and the Samsung HR data representation on the Samsung Health app (see link in [190]) both use the dual scale (discussed in the implementation section 6.3.1), which could not be applied in this study because it is not suitable to be implemented in the Radial layout.
3. The Garmin HR detailed view presented in Figure 4.6 is different, since it only combines the users' status of activity as glyphs (e.g., running or cycling icons) with the HR levels.
4. The Samsung Gear S3 Clock face (see Figures in [195]) also represents multiple data, but it is a representation for the user's activity mode during the day and not a detailed representation such as the visualizations in this study.

8.3.2 Differences from the Visualizations in the Literature

What is novel in the Radial layout? In section 2.3.3.1, many Radial visualizations and clock design metaphors have been discussed. The Radial layout in this thesis differs from the Radial layout in previous work as it considers a particular case of a personal health dataset on an hourly basis that may be more suitable to be represented on a clock design layout and benefits from the proposed segments to identify the parts of the day. In addition, the design of the visualizations includes a representation of multivariate data to evaluate their effectiveness for tasks that require observation of more than one variable.

The most related work to this study is Waldner et al.'s recent work [238] in terms of comparing the Radial and Linear layouts with time-dependent data with daily periodical patterns. However, the main difference between the visualization designs in this thesis and the visualization designs in Waldner et al.'s study is that they only used the Bar chart for encoding the data, while in this implementation both the Bar and Line charts were implemented and compared on both layouts. Moreover, this study provides design considerations for the efficiency of integrating three variables in daily visualizations, while [238] represents data for one variable (number of accidents in hours of a day). The type of the experiment also may have an impact on the results, as they did not require a specific window size in their crowdsourced experiment, which may have affected the size of the visualization presented to the participants and, therefore, the performance. Their discussion of the results is more focused on the Radial vs Linear layout and the difference between the 24-hour view and the two juxtaposed 12-hour views, while in this thesis the discussion is focused on the Radial vs Linear layout and considers visual encoding and the visualization methods of multivariate data on a 24-hour day representation. In terms of the reviewed visualizations in Chapter 4, the differences are discussed in section 8.3.1. Therefore, this thesis presents design considerations for representing multivariate personal health data within the traditional Linear layout and the Radial layout and across both layouts.

8.3.3 The Visualizations of the Multivariate Data

Although not all the tasks involved more than one data variable dataset, the multivariate design of the Stacked and the Overlapped methods were included in the discussion of all the tasks. This

is due to their impact on the results, along with the other factors. The observations regarding the multivariate design are outlined below:

The effect of the Overlapped and Stacked design varied across the two layouts based on the task. For example, for the Compare task, within the Radial layout, the Overlapped design is more accurate, but in the Linear layout the Stacked design is faster. The Overlapped design was also affected by the visual encoding. For example, the results showed that the Radial and Linear Overlapped Bar charts were the slowest in the Identify task T1 within the layouts. On the other hand, they performed they were more accurate in reading values in task T2 as the overlapped data did not interfere and affected the accuracy. Another example is the Linear Overlapped Line chart that had a long task completion time comparing to other designs for locating the maximum. Therefore, the recommendations in Tables 6.30 to 6.34 are given based on the designs.

Regarding the task that involved two data variables (i.e., Infer), the results also showed that the effectiveness of the Overlapped design for the multivariate data could be highly affected by the dataset. With respect to the hypotheses, the results failed to fully support the hypotheses related to the multivariate design (**H3b** and **H4b**), which were explained in section 6.5.6.

8.4 Personal Health Visualization Design Implications

A large number of previous visualization evaluation studies have provided design guidelines and implications which have directed the design of data visualization such as [48, 52, 58, 247]. Based on the results from the study in the thesis and the data collected, we were able to devise a set of implications for designing visualizations for personal health data. The data is time-dependent dataset which is composed of three different data variables that are different in their types and possible values and ranges.

8.4.1 Task-Related Implications (Based on the Quantitative Data)

In section 2.3.2 in Chapter 2, many visualization evaluation studies based on task taxonomies were presented. However, it was found that the structure of visualization design guidelines

and implications are usually based on the visual tasks to be implemented (e.g., [186]). So, in this work the implications for design focus on the most effective visualizations to represent multivariate personal health data on dashboard displays, according to the implemented visual tasks. Therefore, the implications are mainly derived from the results set out in Chapters 6 and 7 and are discussed in light of the related literature.

- **I1:** *Use Line chart encoding with a Linear layout for identifying trends and comparing the data changes either within the same day, or for comparing the data between two days.*

Discussion: This principle is related to the ‘Identify’ task (T1) and the ‘Compare’ task (T3). The similarity between T1 and T3 were explained in Chapter 6 (see the major findings in section 6.5.1), the two tasks share some similarity in terms of the cognitive and perceptual components, as shown in Table 6.4. The Line chart has been identified in the literature for its effectiveness for finding trends tasks [142] and characterising distributions [186].

The Line chart usually refers to continuity, which may allow the participants to perceive data in each part as a coherent object more than the Bar chart. In addition, in the previous study by Zacks and Tversky [247] that compared the effectiveness of the Bar and the Line chart, the participants read the data in the Bar chart as discrete data; yet, they described the data represented in the Line chart as a trend.

Avoid: Bar chart encoding should be avoided as it does not facilitate observing multiple data values for seeing trends and estimating the total for comparisons. Based on the outcome of the experiment, the Overlapped Bars within both layouts should be avoided. The Bar chart is identified as being more suitable for comparing discrete data items rather than looking for trends [247].

- **I2:** *For tasks that require reading values or estimating a value of a specific data item (i.e., T2) the Bar chart encoding within the Linear layout should be used.*

Discussion:

The discussion of the experiment in section 6.5.2 showed that the Bar chart is more accurate than the visualizations with the Line charts within the Linear layout on both

experiment datasets. Previous evaluation studies have shown that the Bar chart is more accurate when reading or retrieving value tasks [186], and graphical perception experiments [48, 217]. Furthermore, the results showed that the visualizations with the Linear layout outperformed the Radial layout (either significantly more accurate or faster or both than the Radial layout on both datasets. The inefficiency of the Radial layout of reading and retrieving value tasks was identified in different evaluation studies that compared the Radial with the Linear layout on different platforms. For example, Brehmer et al. [32]’s evaluation on mobile phones, Waldner et al. [238]’s study on web browsers, and Goldberg and Helfman [96]’s eye-tracking experiment as discussed in detail in section 6.5.2.

Avoid: Line chart encoding and the Radial layout should be avoided for such a task.

- **I3:** *For tasks that require inferring relationships between two different data variables, the Linear layout should be used with either Bar or Line encoding.*

Discussion: The results of ‘Infer’ task T4 in section 6.5.4, showed an outperformance of the Linear layout over the Radial layout. With respect to the visualization method, the Overlapping Bar chart (with both layouts) should be avoided when the dataset is dense or the values of the overlapping variables are close to each other. Although the Overlapped design helped with finding relationships more accurately between the data variables (as identified in the ‘Infer’ task T4 with dataset 0), it was less accurate than the Stacked design in the cases where the dataset had many values that were close to each other (as with dataset 1). According to Munzner [156], there is a trade-off between the advantages of presenting as much information as possible, to reduce the need for exploration; and the cost of presenting too much information that generates visual clutter. This, in fact, corresponds with the participants’ feedback in Chapter 7 that highlighted the efficiency of the Overlapped design for comparing the data on the same chart and also commented on its difficulty and confusion due to too much detail, one participant described it as being ‘heavy’. As explained in the discussion, the Overlapped design was more accurate with datasets with the specified characteristics.

Avoid: The Radial layout is not suitable for inferring relationships between two data variables in any of the designs.

- **I4:** *For tasks that require locating maximum values, the Bar chart encoding within a Linear layout should be used.*

Discussion: Section 6.5.5 discussed the outperformance of the Linear Bar chart over other evaluated visualizations and other studies that showed that the Bar chart is effective for locating maximum values. It also showed that the Linear Overlapped Line chart (LOLC) is the least effective between all other visualizations, including the Radial designs. The Bar chart is more salient than the Line chart, which may facilitate the process of searching for the peaks. These findings are also confirmed by previous research that considered both chart types: the results from Mackinlay et al. [142] stated that the Bars are effective for comparing values because the human visual system is good at comparing Bar lengths. Additionally, results from Saket et al.'s [186] experiment also stated that the Bar chart is significantly faster for detecting maximum than other evaluated alternatives, including the Line chart. With respect to the use of the Radial layout,

Avoid: The Linear Overlapped Line chart should be avoided for locating maximum tasks.

In addition to the previous listed implications for design, the quantitative analysis of the experiment's results and the performance, differences explanations led to devise another set of recommendations and considerations for future personal health visualization designs with respect to the novel concept of the Radial clock layout and the Overlapped design.

- **R1 (Radial Layout):** In case that the Radial layout will be used in the visualization, from the analysis of the data, the following observations of the performance were identified.
 - The Radial layout could only be used (as it is comparable to the Linear layout) for finding and identifying trends task T1 and comparing two days' data T3 if the difference between the compared data (i.e., between the two days and within the same day) is big and can be easily perceived (similar to dataset 0). In addition, the Stacked design seems to be more suitable than the Overlapped design when using the Radial layout.

- The Radial layout is not suitable for representing the data if the implemented tasks require observation of multiple data items or include multiple cognitive processing, and could be used otherwise. This was clear from the experiment results, for example, T4 that required multiple data observation and more cognitive processing, and the result showed that the Radial visualizations were significantly slower than the Linear visualizations and less accurate in some of the cases. Moreover, in task T1 and T3, as the density of the dataset increases (as discussed in each task separately), the significant low performance with the Radial layout start to appear in the results.
- **R2 (Overlapped Design):** The Overlapped design for multivariate data is effective, but, not in all cases as discussed in section 8.4.
 - As was revealed from the main study, for the hypothesis H4b, the results showed that the Overlapped design could outperform the Stacked design if the effect of the clutter is eliminated. For this reason, it might be useful to change the encoding method for representing HR to Line chart to reduce the clutter and the interference with the other variables. The line chart is also more suitable for HR data than the Bar chart due to the continuity of the data.
 - In ‘Infer’ task in the main study, there was a changing performance between Overlapped and Stacked designs between the two datasets. An estimated conclusion from the experiment is that the Overlapped design could only outperform if the effect of the clutter is eliminated. In the case of the experiment’s datasets, it is the density of data items and how the targets overlap with the no-target objects. Implementing the tasks with different chart options (e.g., bars with lines) may support a view of the combination of the variables.
 - Since the Radial visualizations was not effective for this task, further study on this task may consider the Linear layout only. The Linear visualizations may include an additional design that represents the step count and calories in Bar chart and HR in Line chart to reduce the effect of visual clutter from Overlapping Bars.

These guidelines and implications were derived from a controlled experiment that implemented

specific visual tasks on specific datasets. Although the experiment evaluated different visual tasks, other evaluations, such as a field study with people and their real data with the visualizations, could show further results.

8.4.2 Insights for Future Designs (Based on the Qualitative Data)

The analysis of the data collected from either the core quantitative data regarding performance or from the qualitative subjective feedback allows to provide some suggestions or to be considered when designing visualizations for personal health.

- **Radial Clock Layout:** the designs in this thesis are focused on the 24-hour representation of the clock layout. Although one participant in the main evaluation study reported an interest in the 12-hour data representation, Waldner et al. [238] in their experiment showed that the design of 12-hour was not more effective than the 24-hour representation. They found that the participants committed most errors when conducting the tasks on the two juxtaposed 12-hours Radial visualizations and most errors were caused by swapping a.m. and p.m. The suggestion of using 12-hour rather than the 24-hour encoding for intuitive visualization was also suggested by some participants in Fuchs et al. [79]’s study. Visualizing one day’s data into a design two juxtaposed 12-hour clock displays might be more intuitive to the reader but it also requires moving eye gaze between the two circles. Therefore, if the clock designs are used for representing daily data, they should either use the 24-hour representation or make an enhancement on the two juxtaposed 12-hour representation to .
- **Preferences:** the Linear Stacked Bar chart (LSBC) visualization was the most preferred visualizations by the participants either before or after the experiment in both the preliminary evaluation study and the main evaluation study.
- **Familiarity and Tracking Experience:** familiarity and previous experience with the visualizations is an important factor that needs to be considered when designing the visualizations. The study showed that the participants, who track their data, preferred the Linear Stacked designs (especially LSBC). The Linear Line and Bar charts are a

common representation on health tracking platforms as the review in Chapter 4 presents. All the responses that use the term ‘familiar’ for their selection was related to the LSBC visualization.

- **Aesthetic Aspects of the Design:** using some aesthetic features could be interesting for the users at first, however, using the visualizations for retrieving information could impact their perspectives to the design towards more efficient design and losing interest in the aesthetic features. Our evaluation showed that the aesthetical aspects of the visualizations did not affect the participants’ preferences. Several comments mentioned some aesthetic features of the visualizations, but this did not specifically lead to them being preferred.

8.5 Conclusion

In this chapter, we discussed the outcomes of the studies conducted in the thesis. We first discussed the updates of visualization methods of personal health data on dedicated displays in light of the limitations and challenges presented earlier in the thesis, literature and in light of the outcomes of our evaluation study. We also discussed the differences between our visualization designs and other similar designs, to highlight the significance of the outcomes and to show its novelty. Finally, we provided a set of implications for designs that were derived from the quantitative and the qualitative analysis of the evaluation outcomes.

Chapter 9

Conclusion and Future Work

9.1 Thesis Summary and Contributions

Modern sensing technology facilitates the generating of high volumes of multivariate health-related and physical activity data, making them a potential data source which is valuable for many purposes: such as improving or maintaining personal and public health. Data visualizations can help people understand and interpret these complex health data. The dedicated health and fitness tracking devices and apps provide different forms of visual representations and summaries. However, relevant literature has reported that current visualizations limit people's view of the data to a simple data representation that does not reflect or exploit the value of the gathered data. Moreover, these visualizations do not always satisfy people's main health tracking purposes. Motivated by the significance of personal health and physical activity, this research has focused on the visualization methods for multivariate personal health data.

According to the research stages presented in Figure 1.1, the first stage is the investigation and gathering information. The research problem, defined in section 1.2, led to a set of research questions that have been answered in this thesis. A review of the literature showed that many research papers on personal health and personal informatics discussed aspects such as people's behaviour and requirements, issues people face in the used tracking tools, or how they use the tools to reflect on their data; however, there were no specific details on or discussions of the visualization methods used, their types and their limitations. Therefore, this research started by answering the research questions **RQ1** and **RQ2**, aiming to establish a connection between published literature in personal health and personal visualizations and examining what most popular health tracking devices provide with respect to tracked health data representation. The first two questions were both addressed through Chapters 2, 3 and 4. Beginning with a survey of related literature, then, using an online survey to collect information about the used tracking devices, type of tracked data and interest in different forms of data visualization, the outcomes of which were used to guide the next study: a review of the visualization methods on health tracking devices and their companion apps and dashboards. The answer to these two research questions has contributed to the field by providing a taxonomy to classify the visualization methods used to represent health and physical activity data on the most popular tracking devices and their platforms. The taxonomy also provides information regarding the data being visualized with each method. In addition, the visualization review examines whether

the current visualizations supported by such devices meet users' needs. This derived from the previous studies and reports on what is missing in personal health visualization design space and provides suggestions for addressing the highlighted limitations. The answers to research questions **RQ1** and **RQ2** led to the following stage, which is the visualization development and evaluation stage. These helped with the design choices in the implementation (i.e., **RQ3** and **RQ4**), as described in section 4.7 and the evaluation studies in the following chapters.

Personal health data is time-dependent, as it is continuously measured by embedded sensors in the tracking devices while they are in use (e.g., worn by the user). A time series is a form of data where the observed values change continuously over time and are measured at discrete intervals (e.g., HR measures) [151]. One of the methods used to visualize time-series data is the Radial chart, which was suggested for its effectiveness in demonstrating internal circular behaviour in the data or periodic patterns [36]. It was also chosen because it is space-efficient, has an intuitive design [126] and is aesthetically appealing [36]. Relying on the periodic nature of personal health data and the advantages of the Radial chart, to represent this data a Radial analogue clock layout was proposed. The layout design was suggested to imply an additional layer of information by depicting the hourly health data in the natural order of the 24-hour day. The Radial designs included four coloured sectors to represent the parts of the day: morning, afternoon, evening, and night, which may allow the user to perceive each part as a whole and so facilitate the estimation of the total calories within each part according to the visual grouping principles. The glyphs are used as extra information to refer to the time. Another important feature of personal health data is that it composed of different types of health variables, which is difficult to be combined into one visualization due to the different scales each variable uses. Two methods were used to combine the data (Overlapped and Stacked), where the first visualizes the data on separate charts and in the other design the data variables overlap on the same chart area. Several encoding methods could be used to visualize the data on Cartesian and Polar coordinates. However, Bar and Line chart encoding were considered to be used for visualizing the data. Consequently, eight different visualization designs were developed to evaluate the effectiveness of the applied visual elements to represent the data. The visualization development and the implementation details in Chapter 5 and 6 provide the answers to research question **RQ3** and **RQ4**.

After developing the visualizations, a preliminary evaluation study was designed to validate the evaluation procedure, the implementation of the lab experiments and to evaluate the developed designs before conducting the main evaluation study. The outcomes of the preliminary evaluation study (which is presented in Chapter 5) delivered a set of lessons learned that guided the design of the visualizations and the implementation for the main evaluation study. Also, at this stage of the research a set of hypotheses regarding the visualizations and tasks were constructed to be tested in the main evaluation study. The evaluation of the visualizations in the main study (presented in Chapter 6) conducted a controlled user experiment with 80 participants to compare the effectiveness of Radial and Linear designs, using measures of task accuracy and completion time with regard to five visual tasks related to the real datasets. The analysis performed pairwise comparisons between the visualizations across the implemented tasks regarding participants' performance (which is the answer to **RQ5**). The results proved that the traditional Linear layout either outperforms, or is comparable to the proposed Radial layout. Regarding the other evaluated visual encodings and visual methods, these change in performance according to the task and the dataset. They also provided plausible explanations for the significant differences in the observed performance patterns, which inform future visualization designs for personal health data and led to classify the visualization based on its effectiveness for each task. The study's main experiment also included qualitative measures to evaluate the subjective feedback in order to answer **RQ6**, the analysis and results of which were described in Chapter 7. The results showed how the preferences and the ratings of the visualizations changed after the experiment, which helped in justifying the visualization features that affected the performance and what needs to be considered in future designs. Moreover, the thematic analysis of the open-ended question provided an insight to the role of the visual layout, the visual encoding and the visualization method in interpreting the data from the participants' points of view. The study and the analysis of the results have contributed to the field by providing implications for visualization designs to the data type and suggestions for designing visualizations for personal health data, these are discussed in Section 8.4 of this chapter.

In section 8.3.2, the differences between the study in the thesis and Waldner et al.'s [238] study, which is the most related work, were highlighted. Although there are some similarities in terms of the concepts of comparing Radial and Linear layout for representing daily data, the visualizations in this thesis expanded and included more visualization alternatives within

both layouts. The Radial layout was also designed with specific sectors to facilitate identifying the parts of the day in the clock-style design with glyphs of moon and sun. Unlike Waldner et al.'s [238] study that encoded the data using Bar chart only, the visualizations in the thesis experimented both Bar and Line encodings. In addition, the study evaluated two alternative methods for combining multivariate data. Therefore, the outcomes from the studies in the thesis contributed implications for visualization design and recommendations that discuss the visual layout, the visual encoding and the method for combining multivariate data.

9.2 Limitations

Although this research has achieved many successes, it has faced some challenges. Therefore, there are some limitations with the research that are discussed below.

The review of the visualizations in Chapter 4, which directed the design of the visualizations in Chapter 5 and 6 reviewed four of the most popular tracking devices. Although the review targeted most popular health tracking devices and included different platforms (i.e., device screen and companion app(s)/dashboard(s)), the review was restricted to four devices. We have faced some difficulties in accessing more tracking devices to include in the review study. However, the study highlighted the main limitations that need to be addressed to fulfil potential users' requirements.

The thesis has conducted a successful evaluation study for eight visualizations that represent multivariate personal health data with 80 participants. The outcomes of the evaluation provided a set of implications of design and insights that could direct future visualization designs. However, there are some adjustments and improvements that could be made. In both evaluation studies (in 5 and 6), we recruited a good number of participants for the goal and the purpose of the study, who were from different fields and backgrounds. However, the recruitment was restricted to student and staff from Cardiff University due to research ethics restrictions and access permissions. Recruiting more diverse participants might have provided more diverse feedback.

With respect to the evaluated visualization tasks, there was no a domain-specific task taxonomy

available in the literature to be used for evaluating the visualizations for representing personal health data. This research has not directly recorded potential users' tasks (e.g., by conducting an interview study with real users of personal health trackers to understand and investigate the type of the tasks they perform on the visualizations) as this was not the main focus of the thesis. However, to conduct a good evaluation with more realistic tasks, the design of the tasks and questions in the evaluation studies were directed by previous refined visual task taxonomies for similar data type and previous research that studied potential user' needs and requirements (e.g., Li et al.'s [134, 135] and Choe et al.'s studies [44, 45]).

Regarding the methodology applied to evaluate the developed visualizations, this is restricted to provide an understanding of how people perform with basic visual tasks. Further insight into the personal visualization design might be gained by other methodologies, such as a field study with the visualizations to examine the usability of each visualization and which visualization is most used. With respect to the evaluation platform, the implementation and the evaluation of the visualizations in this thesis were conducted on desktop applications where the results apply only to online dashboards that usually be viewed on large screens. Therefore, the results, guidelines and implications are more applicable to visualizations on desktop displays.

9.3 Future Work

There are many ways in which this study can be taken further to solve various open questions in the personal health visualization domain. This thesis has the potential to be expanded in future work, taking many directions in various areas, these are discussed in the following section.

Evaluation on a Mobile Phone Platform:

The significance of implementing the same study on a mobile platform is in determining the best design option on different platforms to structure design guidelines, especially in terms of multivariate data. Moreover, many of the tracking devices that support viewing the data on an online dashboard (e.g., Fitbit) use some similar visualizations to those that are used in the apps. Evaluating the developed visualizations on mobile platforms is anticipated to incorporate findings regarding visualizing personal health data across different platforms. The

survey on information visualization on mobile devices [26] highlights the need for more studies to evaluate visualization designs for mobile displays. In terms of personal health visualization, the research studies, that were discussed in Section 2.2.3, only focused on a specific platform such as a web browser [115] or a mobile app [149], and no research was found which focused on personal health visualization across different platforms.

Adjustments on the Visualization Designs:

Changes to the developed visualizations could be applied to investigate the efficiency of different visual elements, for example implementing the Radial and Linear visualizations with and without the background colours and glyphs (parts of the day). One of the limitations of the Linear visualization that was highlighted by Diehl et al. [53] when compared with the Radial layout was the lack of natural reference points. Therefore, Diehl et al. emphasised the importance of providing a visual context for the Linear visualization. The visual context is even more important with respect to personal visualization where multiple factors could provide a better understanding of the data.

Dataset:

As this thesis focused on short-term data (i.e., an hourly data of one-day), future studies could expand the study to include different granularities of time such as a week, a month or a year. Different types of data, rather than that used in this thesis, could be used in testing the multivariate data. For example, representing sleep data, which is usually represented as ranges, as described in Section 2.2, with other physical activity data in order to support the user by testing hypotheses regarding the relationships between their activities and sleeping patterns (as reported in [177]). Another example is combining personal physical activity and locations, which was found to be helpful in identifying correlations [60].

Field Study:

Evaluating the visualization in a personal context is important, as personal knowledge plays a significant role in data interpretation [114]. An example of such evaluation is the study by Huang et al. [115] that implemented a field study to evaluate the effectiveness of the calendar visualization to represent physical activity data. The visualizations could also be evaluated using different methodologies, such as elicitation interviews that guide the participants' attention

toward an individual experience, as suggested by Thudt et al. [223].

Glanceable Displays:

A significant finding of the study was that it showed that the Radial layout was not as effective as the Linear layout for retrieving specific information. However, the Radial layout could be evaluated for its effectiveness for glanceable representations. This requires less information to be represented, such as limiting the visualization to only one variable, removing the hours and relying on the coloured sectors and glyphs to infer the time. An evaluation of such a task involves quick perceptual judgements of the information visualized. The outcomes may contribute to understanding the efficiency of the developed clock metaphor in perceiving activity progress data over a day.

New Domain Tasks Taxonomy:

Designing information visualization requires taking into consideration the tasks users want to pursue with the visualizations [202]. In Section 2.3.2, visualization tasks taxonomies were presented, it showed that some taxonomies were domain-specific and others were developed to be generic. Evaluating the visualizations with real users who implement domain-specific tasks is essential to ensure that the designed visualizations are interesting to potential users [40]. There was some research, such as that carried out by Serrano et al. [204], that studied users' engagement with data from a health app (a weight loss app). However, there is no study that proposes a task taxonomy for basic tasks that are implemented by people when viewing and exploring their health and physical activity data on the dedicated displays, which could be displayed on different screen sizes as discussed in Chapter 4. There are several methods for constructing visualizations task-taxonomies, for example, using surveys to ask prospective users about the required tasks [221] and specific content analysis [9].

Interactivity:

Another aspect that needs to be considered is the level of interactivity that should be supported by each platform. The type of interactivity supported by wearable devices, apps and online dashboards is different and requires different designs and evaluations. Advanced interactive features, such as data filtering, has been found to be required by the participants in previous studies (e.g., [149]). Analytical interactive features should be accessible, as the majority of

people are not experts in data analytic or visualization [114].

Future studies on the visualizations of personal health data are not restricted to those mentioned here; others, such as considering the visualizations for sharing for social and medical purposes, for promoting behaviour change, or for addressing data integration from multiple sources (e.g., from using multiple tracking devices) could be considered in future work on this topic.

9.4 Conclusion

This thesis started with the goal of designing and evaluating visualization designs to represent multivariate personal health data. The visualization designs aimed to address the need for combining multiple personal health variables to support a better representation of an individual's health status, and to find a more suitable layout that is related to the periodic and temporal nature of personal health data. Extensive work has been performed to achieve this goal, starting from the investigation and gathering information stage to the visualization design and the evaluation stage. The studies presented throughout this thesis have successfully addressed the research questions highlighted in the introduction chapter. The answers to the research questions and the contributions were summarised at the beginning of this chapter (see section 9.1).

The outcomes of the evaluation study provided insights regarding personal health visualization. It demonstrated how the applied visual encoding and the visualization method influence the performance according to specific tasks and under different data densities. In addition, it contributed insights on people's preferences on the visualizations, which showed the effects of each implemented visual element. And as with any research, this study was constrained by some limitations that were discussed in section 9.2. We also pointed out some future directions where this research can proceed in section 9.3.

Bibliography

- [1] Oxford English Dictionary (Online), 1989. URL <https://www.oed.com>.
- [2] *Concise medical dictionary*. Oxford University Press, Oxford, 8th ed. edition, 2010. ISBN 0199557144.
- [3] *MMHealth '16: Proceedings of the 2016 ACM Workshop on Multimedia for Personal Health and Health Care*, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4518-7.
- [4] *MMHealth '17: Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care*, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-5504-9.
- [5] *HealthMedia'18: Proceedings of the 3rd International Workshop on Multimedia for Personal Health and Health Care*, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5982-5.
- [6] M. Adnan, M. Just, and L. Baillie. Investigating time series visualisations to improve the user experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5444–5455, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858300.
- [7] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):47–60, Jan. 2008. ISSN 1077-2626. doi: 10.1109/TVCG.2007.70415.
- [8] R. Amar and J. Stasko. Best paper: A knowledge task-based framework for design and evaluation of information visualizations. In *IEEE Symposium on Information Visualization*, pages 143–150, Oct 2004. doi: 10.1109/INFVIS.2004.10.
- [9] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Proceedings of the 2005 IEEE Symposium on Information*

- Visualization*, INFOVIS '05, pages 15–, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7803-9464-x. doi: 10.1109/INFOVIS.2005.24.
- [10] M. Analytical. Samsung gear s3 more tips & tricks - customise the menus etc, December 2016. URL <https://www.youtube.com/watch?v=hDtbyfrezxQ>. [last accessed July 2017].
- [11] I. Anderson, J. Maitland, S. Sherwood, L. Barkhuus, M. Chalmers, M. Hall, B. Brown, and H. Muller. Shakra: Tracking and sharing daily activity levels with unaugmented mobile phones. *Mobile Networks and Applications*, 12(2):185–199, 2007. ISSN 1572-8153. doi: 10.1007/s11036-007-0011-7.
- [12] N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer-Verlag, Berlin, Heidelberg, 2005. ISBN 3540259945.
- [13] Apple Inc. A bold new way to look at your health. <https://www.apple.com/uk/ios/health/>, 2017. [last accessed July 2017].
- [14] Apple Support. Use the activity app on your apple watch. <https://support.apple.com/en-us/HT204517>, April 2017. [last accessed July 2017].
- [15] Apple Support. Track your trends in the activity app. <https://support.apple.com/en-us/HT210343>, September 2019. [last accessed September 2019].
- [16] S. Asimakopoulos, G. Asimakopoulos, and F. Spillers. Motivation and user engagement in fitness tracking: Heuristics for mobile healthcare wearables. *Informatics*, 4(1), 2017. ISSN 2227-9709. doi: 10.3390/informatics4010005.
- [17] J. E. Bardram and M. Frost. The personal health technology design space. *IEEE Pervasive Computing*, 15(2):70–78, Apr 2016. ISSN 1536-1268. doi: 10.1109/MPRV.2016.37.
- [18] J. Bassett, David R, L. P. Toth, S. R. LaMunion, and S. E. Crouter. Step counting: A review of measurement considerations and health-related applications. *Sports medicine (Auckland, N.Z.)*, 47(7):1303–1315, 07 2017. doi: 10.1007/s40279-016-0663-1.
- [19] J. S. Bauer, S. Consolvo, B. Greenstein, J. Schooler, E. Wu, N. F. Watson, and J. Kientz. Shuteye: Encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1401–1410, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2208600.
- [20] D. A. Bender. *A dictionary of food and nutrition*. Oxford University Press, 4th ed. edition, 2014. ISBN 9780191752391.

- [21] F. Bentley, K. Tollmar, P. Stephenson, L. Levy, B. Jones, S. Robertson, E. Price, R. Catrambone, and J. Wilson. Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Trans. Comput.-Hum. Interact.*, 20(5):30:1–30:27, Nov. 2013. ISSN 1073-0516. doi: 10.1145/2503823.
- [22] J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. UMI Research Press, 1983. ISBN 9780835735322.
- [23] J. Bertin. *Graphics and Graphic Information Processing*, page 62–65. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999. ISBN 1558605339.
- [24] M. J. Bietz, C. S. Bloss, S. Calvert, J. G. Godino, J. Gregory, M. P. Claffey, J. Sheehan, and K. Patrick. Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association : JAMIA*, 23(e1): e42–e48, 04 2016. doi: 10.1093/jamia/ocv118.
- [25] T. Blascheck, L. Besançon, A. Bezerianos, B. Lee, and P. Isenberg. Glanceable visualization: Studies of data comparison performance on smartwatches. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):630–640, Jan 2019. ISSN 1077-2626. doi: 10.1109/TVCG.2018.2865142.
- [26] K. Blumenstein, C. Niederer, M. Wagner, G. Schmiedl, A. Rind, and W. Aigner. Evaluating information visualization on mobile devices: Gaps and challenges in the empirical evaluation design space. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, BELIV '16, pages 125–132, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4818-8. doi: 10.1145/2993901.2993906.
- [27] S. Boll, W. Heuten, and J. Meyer. From tracking to personal health. *interactions*, 23(1): 72–75, December 2015. ISSN 1072-5520. doi: 10.1145/2843586.
- [28] R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Reppa, L. Floridi, and M. Chen. An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2759–2768, Dec 2012. ISSN 1077-2626. doi: 10.1109/TVCG.2012.197.
- [29] R. Borgo, J. Kehrler, D. H. S. Chung, E. Maguire, R. S. Laramée, H. Hauser, M. Ward, and M. Chen. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In M. Sbert and L. Szirmay-Kalos, editors, *Eurographics 2013 - State of the Art Reports*. The Eurographics Association, 2013. doi: 10.2312/conf/EG2013/stars/039-063.

- [30] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, Dec 2013. doi: 10.1109/TVCG.2013.234.
- [31] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, Dec. 2013. ISSN 1077-2626. doi: 10.1109/TVCG.2013.124.
- [32] M. Brehmer, B. Lee, P. Isenberg, and E. K. Choe. Visualizing ranges over time on mobile phones: A task-based crowdsourced evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):619–629, 2019. ISSN 1077-2626.
- [33] C. Brewer and M. Harrower. ColorBrewer 2.0, color advice for cartography. <http://colorbrewer2.org/>, 2018. [last accessed October 2018].
- [34] British Heart Foundation. Your heart rate. <https://www.bhf.org.uk/>. [last accessed October 2019].
- [35] P. Buono, M. F. Costabile, and R. Lanzilotti. A circular visualization of people’s activities in distributed teams. *Journal of Visual Languages and Computing*, 25(6):903–911, Dec. 2014. ISSN 1045-926X. doi: 10.1016/j.jvlc.2014.10.025.
- [36] M. Burch and D. Weiskopf. *On the Benefits and Drawbacks of Radial Diagrams*, pages 429–451. Springer New York, New York, NY, 2014. ISBN 978-1-4614-7485-2. doi: 10.1007/978-1-4614-7485-2_17.
- [37] M. Burch, N. Konevtsova, J. Heinrich, M. Hoferlin, and D. Weiskopf. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2440–2448, Dec. 2011. ISSN 1077-2626. doi: 10.1109/TVCG.2011.193.
- [38] J. E. Capó-Aponte, L. A. Temme, H. L. Task, A. R. Pinkus, M. E. Kalich, A. J. Pantle, and C. E. Rash. Visual perception and cognitive performance. 2009.
- [39] S. K. Card, A. Newell, and T. P. Moran. *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1983. ISBN 0898592437.
- [40] S. Carpendale. Information visualization. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, chapter Evaluating Information Visualizations, pages 19–45. Springer-Verlag, Berlin, Heidelberg, 2008. ISBN 978-3-540-70955-8. doi: 10.1007/978-3-540-70956-5_2.

- [41] S. M. Casner. Task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics*, 10(2):111–151, Apr. 1991. ISSN 0730-0301. doi: 10.1145/108360.108361.
- [42] C. Chen. Top 10 unsolved information visualization problems. *IEEE Comput. Graph. Appl.*, 25(4):12–16, July 2005. ISSN 0272-1716. doi: 10.1109/MCG.2005.91.
- [43] M. Chen, L. Floridi, and R. Borgo. *What Is Visualization Really For?*, pages 75–93. Springer International Publishing, Cham, 2014. ISBN 978-3-319-07121-3. doi: 10.1007/978-3-319-07121-3_5.
- [44] E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz. Understanding quantified-selfers’ practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’14, pages 1143–1152, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557372.
- [45] E. K. Choe, B. Lee, and m. c. schraefel. Characterizing visualization insights from quantified selfers’ personal data presentations. *IEEE Computer Graphics and Applications*, 35(4):28–37, July 2015. ISSN 0272-1716. doi: 10.1109/MCG.2015.51.
- [46] J. Clawson, J. A. Pater, A. D. Miller, E. D. Mynatt, and L. Mamykina. No longer wearing: Investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’15, pages 647–658, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3574-4. doi: 10.1145/2750858.2807554.
- [47] W. S. Cleveland. A model for studying display methods of statistical graphics. *Journal of Computational and Graphical Statistics*, 2(4):323–343, 1993. doi: 10.1080/10618600.1993.10474616.
- [48] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984. doi: 10.1080/01621459.1984.10478080.
- [49] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. A. Landay. Activity sensing in the wild: A field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’08, pages 1797–1806, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-011-1. doi: 10.1145/1357054.1357335.

- [50] K. Crawford, J. Lingel, and T. Karppi. Our metrics, ourselves: A hundred years of self-tracking from the weight scale to the wrist wearable device. *European Journal of Cultural Studies*, 18(4-5):479–496, 2015. doi: 10.1177/1367549415584857.
- [51] R. D. Croon, B. D. L. R. P. Cardoso, and K. Verbert. Myhealthtoday: Helping patients with their healthschedule using a 24-hour clock visualization. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 415–420, Aug 2017. doi: 10.1109/ICHI.2017.32.
- [52] F. E. Croxton and R. E. Stryker. Bar charts versus circle diagrams. *Journal of the American Statistical Association*, 22(160):473–482, 1927. ISSN 01621459.
- [53] S. Diehl, F. Beck, and M. Burch. Uncovering strengths and weaknesses of radial visualizations—an empirical approach. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):935–942, Nov 2010. ISSN 1077-2626. doi: 10.1109/TVCG.2010.209.
- [54] A. Dix, J. E. Finlay, G. D. Abowd, and R. Beale. *Human-Computer Interaction (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003. ISBN 0130461091.
- [55] P. Dragicevic and S. Huot. Spiraclock: A continuous and non-intrusive display for upcoming events. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '02, pages 604–605, New York, NY, USA, 2002. ACM. ISBN 1-58113-454-1. doi: 10.1145/506443.506505.
- [56] G. M. Draper, Y. Livnat, and R. F. Riesenfeld. A survey of radial methods for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15(5): 759–776, Sep. 2009. ISSN 1077-2626. doi: 10.1109/TVCG.2009.23.
- [57] L. Edwards. Garmin forerunner 735xt review. <https://www.wearable.com/garmin/garmin-forerunner-735xt-review>, September 2016. [last accessed July 2017].
- [58] W. C. Eells. The relative merits of circles and bars for representing component parts. *Journal of the American Statistical Association*, 21(154):119–132, 1926. ISSN 01621459.
- [59] S. G. Eick and A. F. Karr. Visual scalability. *Journal of Computational and Graphical Statistics*, 11(1):22–43, 2002. ISSN 10618600.
- [60] D. Epstein, F. Cordeiro, E. Bales, J. Fogarty, and S. Munson. Taming data complexity in lifelogs: Exploring visual cuts of personal informatics data. In *Proceedings of the 2014*

- Conference on Designing Interactive Systems, DIS '14*, pages 667–676, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2902-6. doi: 10.1145/2598510.2598558.
- [61] D. A. Epstein, A. Borning, and J. Fogarty. Fine-grained sharing of sensed physical activity: A value sensitive approach. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '13*, page 489–498, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450317702. doi: 10.1145/2493432.2493433.
- [62] D. A. Epstein, A. Ping, J. Fogarty, and S. A. Munson. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '15*, page 731–742, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450335744. doi: 10.1145/2750858.2804250.
- [63] D. A. Epstein, J. H. Kang, L. R. Pina, J. Fogarty, and S. A. Munson. Reconsidering the device in the drawer: Lapses as a design opportunity in personal informatics. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '16*, page 829–840, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450344616. doi: 10.1145/2971648.2971656.
- [64] M. Eysenck and M. Keane. *Cognitive Psychology: A Student's Handbook*. Online access with purchase via Coutts: EBSCO. Taylor & Francis, 2015. ISBN 9781317695172.
- [65] S. Faisal, A. Blandford, and H. W. Potts. Making sense of personal health information: Challenges for information visualization. *Health Informatics Journal*, 19(3):198–217, 2013. doi: 10.1177/1460458212465213. PMID: 23981395.
- [66] C. Fan, J. Forlizzi, and A. K. Dey. A spark of activity: Exploring informative art as visualization for physical activity. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing, UbiComp '12*, pages 81–84, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1224-0. doi: 10.1145/2370216.2370229.
- [67] F. Fischer, J. Fuchs, and F. Mansmann. Clockmap: Enhancing circular treemaps with temporal glyphs for time-series data. In *EuroVis*, 2012.
- [68] Fitbit. *Fitbit Blaze User Manual*. URL https://staticcs.fitbit.com/content/assets/help/manuals/manual_blaze_en_US.pdf. [last accessed July 2017].
- [69] Fitbit. Introducing fitbit blaze. <https://www.youtube.com/watch?v=3k3DNT54NkA&feature=youtu.be>, January 2016. [last accessed July 2017].

- [70] Fitbit. What's sleep score in the fitbit app? https://help.fitbit.com/articles/en_US/Help_article/2439#what, September 2018. [last accessed September 2019].
- [71] Fitbit. What's my fitbit hourly activity goal? https://help.fitbit.com/articles/en_US/Help_article/1878/?l=en_US&c=Topics%3ADashboard&fs=Search&pn=1, November 2019. [last accessed December 2019].
- [72] Fitbit. Fitbit App: The fitness app for everyone. <https://www.apple.com/uk/ios/health/>, 2019. [last accessed November 2019].
- [73] Fitbit. Get a good night's sleep with fitbit's sleep tools. <https://blog.fitbit.com/get-a-good-nights-sleep-with-fitbits-new-sleep-tools/>, January 2019. [last accessed September 2019].
- [74] Fitbit. Fitbit blaze introductory guide. <https://www.fitbit.com/uk/blaze/blaze-101>, n.d. [last accessed July 2017].
- [75] Fitbit. Fitbit community. <https://community.fitbit.com/t5/Community/ctp/EN>, n.d. [last accessed July 2017].
- [76] Fitbit. Fitbit youtube channel. <https://www.youtube.com/fitbit/videos?pbjreload=101>, n.d. [last accessed July 2017].
- [77] C. M. D. S. Freitas, M. S. Pimenta, and D. L. Scapin. *User-Centered Evaluation of Information Visualization Techniques: Making the HCI-InfoVis Connection Explicit*, pages 315–336. Springer New York, New York, NY, 2014. ISBN 978-1-4614-7485-2. doi: 10.1007/978-1-4614-7485-2_12.
- [78] B. Fry and C. Reas. Processing. <https://processing.org>, 2017. [last accessed February 2020].
- [79] J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 3237–3246, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1899-0. doi: 10.1145/2470654.2466443.
- [80] GadFit. Garmin forerunner 235 user review | GadFit, December 2015. URL <http://www.gadfit.com/garmin-forerunner-235-user-review/>. [last accessed July 2017].

- [81] GadFit. Elevate your training with garmin vivosmart hr - user review | gadfit, November 2015. URL <http://www.gadfit.com/garmin-vivosmart-hr-user-review/>. [last accessed July 2017].
- [82] E. Gansner, Y. Hu, and S. Kobourov. *Viewing abstract data as maps*, pages 63–89. Springer New York, United States, 1 2014. ISBN 9781461474845. doi: 10.1007/978-1-4614-7485-2_3.
- [83] Garmin. *Garmin Forerunner 935® Owner’s Manual*. Garmin. [last accessed July 2017].
- [84] Garmin. Share, store and explore at garmin connect. <https://www.youtube.com/watch?v=Wsoug4FJwoU>, October 2009. [last accessed July 2017].
- [85] Garmin. What every cyclist should know - v02 max and aerobic performance on your garmin. , May 2019. URL <https://www.garmin.com/en-US/blog/fitness/v02-max-and-aerobic-performance/>. [last accessed September 2019].
- [86] Garmin. Running Science. , September 2019. URL <https://www.garmin.com/en-US/performance-data/running/>. [last accessed September 2019].
- [87] Garmin. Garmin youtube channel. <https://www.youtube.com/user/garminblog/videos>, n.d. [last accessed July 2017].
- [88] Garmin Ltd. Get Started with Your Forerunner. <https://connect.garmin.com/en-US/start/forerunner/>, 2017. [last accessed July 2017].
- [89] Garmin Ltd. GarminConnectApp. <https://buy.garmin.com/en-GB/GB/p/125677>, 2019. [last accessed November 2019].
- [90] Garmin.com. Garmin Ltd. <http://www.garmin.com/en-US/>, 2017. URL <http://www.garmin.com/en-US/>. [last accessed July 2017].
- [91] N. Genes, S. Violante, C. Cetrangol, L. Rogers, E. E. Schadt, and Y.-F. Y. Chan. From smartphone to ehr: a case report on integrating patient-generated health data. *npj Digital Medicine*, 1(1):23, 2018. doi: 10.1038/s41746-018-0030-8.
- [92] J. George Chin, M. Singhal, G. Nakamura, V. Gurumoorthi, and N. Freeman-Cadoret. Visual analysis of dynamic data streams. *Information Visualization*, 8(3):212–229, 2009. doi: 10.1057/ivs.2009.18.
- [93] D. Gergle and D. Tan. *Experimental research in HCI*, pages 191–227. Springer New York, 1 2014. ISBN 9781493903771. doi: 10.1007/978-1-4939-0378-8_9.

- [94] M. Gleicher. Considerations for visualizing comparison. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):413–423, Jan 2018. ISSN 1077-2626. doi: 10.1109/TVCG.2017.2744199.
- [95] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, Oct. 2011. ISSN 1473-8716. doi: 10.1177/1473871611416549.
- [96] J. Goldberg and J. Helfman. Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization*, 10(3):182–195, 2011. doi: 10.1177/1473871611406623. URL <https://doi.org/10.1177/1473871611406623>.
- [97] Google Commerce Ltd. Money manager: Expense tracker, free budgeting app. , September 2019. URL https://play.google.com/store/apps/details?id=money.expense.budget.wallet.manager.track.finance.tracker&hl=en_GB. [last accessed September 2019].
- [98] R. Gouveia, E. Karapanos, and M. Hassenzahl. How do we engage with activity trackers?: A longitudinal study of habito. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 1305–1316, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3574-4. doi: 10.1145/2750858.2804290.
- [99] R. Gouveia, F. Pereira, E. Karapanos, S. A. Munson, and M. Hassenzahl. Exploring the design space of glanceable feedback for physical activity trackers. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, pages 144–155, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4461-6. doi: 10.1145/2971648.2971754.
- [100] E. Graells and A. Jaimes. Lin-spiration: Using a mixture of spiral and linear visualization layouts to explore time series. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces*, IUI '12, page 237–240, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450310482. doi: 10.1145/2166966.2167006.
- [101] C. C. Gramazio, K. B. Schloss, and D. H. Laidlaw. The relation between visualization size, grouping, and user performance. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1953–1962, 2014.

- [102] S. Haroz and D. Whitney. How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2402–2410, Dec 2012. ISSN 1077-2626. doi: 10.1109/TVCG.2012.233.
- [103] T. Harries, P. Eslambolchilar, C. Stride, R. Rettie, and S. Walton. *Walking in the Wild – Using an Always-On Smartphone Application to Increase Physical Activity*, pages 19–36. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-40498-6. doi: 10.1007/978-3-642-40498-6_2.
- [104] M. Harrower and C. A. Brewer. Colorbrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003. doi: 10.1179/000870403235002042.
- [105] Hayward Fitness. Garmin Connect Mobile App Overview @YouTube. <https://www.youtube.com/watch?v=akgMw5ivNuY>, July 2019. [last accessed September 2019].
- [106] J. Heer and M. Bostock. Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 203–212, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-929-9. doi: 10.1145/1753326.1753357.
- [107] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1303–1312, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7. doi: 10.1145/1518701.1518897.
- [108] J. Heer, M. Bostock, and V. Ogievetsky. A tour through the visualization zoo. *Commun. ACM*, 53(6):59–67, June 2010. ISSN 0001-0782. doi: 10.1145/1743546.1743567.
- [109] M. Hlawatsch, F. Sadlo, M. Burch, and D. Weiskopf. Scale-stack bar charts. *Computer Graphics Forum*, 32(3pt2):181–190. doi: 10.1111/cgf.12105.
- [110] J. Hoff. Garmin updates connect mobile app, improved interface and customization. <https://androidcommunity.com/garmin-updates-connect-mobile-app-improved-interface-and-customization-20151022/>, October 2015. [last accessed July 2017].
- [111] P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley. Dna visual and analytic data mining. In *Proceedings. Visualization '97 (Cat. No. 97CB36155)*, pages 437–441, Oct 1997. doi: 10.1109/VISUAL.1997.663916.

- [112] M. Hoogendoorn and B. Funk. *Machine Learning for the Quantified Self: On the Art of Learning from Sensory Data*. Springer Publishing Company, Incorporated, 1st edition, 2017. ISBN 3319663070.
- [113] D. Huang, M. Tory, and L. Bartram. Data in everyday life: Visualizing time-varying data on a calendar. In *Poster at IEEE VIS 2014*, Paris, France, 2014.
- [114] D. Huang, M. Tory, B. A. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 21(3):420–433, March 2015. ISSN 1077-2626.
- [115] D. Huang, M. Tory, and L. Bartram. A field study of on-calendar visualizations. In *Proceedings of Graphics Interface 2016*, GI 2016, pages 13–20. Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, 2016. ISBN 978-0-9947868-1-4. doi: 10.20380/GI2016.03.
- [116] W. Huang, P. Eades, and S.-H. Hong. Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 8(3):139–152, June 2009. ISSN 1473-8716. doi: 10.1057/ivs.2009.10.
- [117] iBuild Lab. ifood diary - simple food tracker/reminder/journal. https://play.google.com/store/apps/details?id=com.ibuild.fooddiary&hl=en_GB, September 2019. [last accessed September 2019].
- [118] IDC Research, Inc. Wearables aren't dead, they're just shifting focus as the market grows 16.9% in the fourth quarter, according to idc. , March 2017. URL <http://www.idc.com/getdoc.jsp?containerId=prUS42342317>. [last accessed June 2017].
- [119] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, Aug 1985. ISSN 1432-2315. doi: 10.1007/BF01898350.
- [120] P. Isenberg, A. Bezerianos, P. Dragicevic, and J. Fekete. A study on dual-scale data charts. *IEEE Transactions on Visualization & Computer Graphics*, 17(12):2469–2478, dec 2011. ISSN 1077-2626. doi: 10.1109/TVCG.2011.160.
- [121] W. Javed, B. McDonnel, and N. Elmqvist. Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):927–934, Nov. 2010. ISSN 1077-2626. doi: 10.1109/TVCG.2010.162.

- [122] S. L. Jones and R. Kelly. Dealing with information overload in multifaceted personal informatics systems. *Human-Computer Interaction*, 33(1):1–48, 2018. doi: 10.1080/07370024.2017.1302334.
- [123] M. Kay, E. K. Choe, J. Shepherd, B. Greenstein, N. Watson, S. Consolvo, and J. A. Kientz. Lullaby: A capture & access system for understanding the sleep environment. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, page 226–234, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450312240. doi: 10.1145/2370216.2370253.
- [124] M. Kent. *The Oxford Dictionary of Sports Science & Medicine*. Oxford University Press, 3 edition, 2006. ISBN 9780198568506. URL <http://www.oxfordreference.com/view/10.1093/acref/9780198568506.001.0001/acref-9780198568506>.
- [125] N. W. Kim, H. Im, N. Henry Riche, A. Wang, K. Gajos, and H. Pfister. Dataselfie: Empowering people to design personalized visuals to represent their data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 79:1–79:12, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300309.
- [126] C. Kintzel, J. Fuchs, and F. Mansmann. Monitoring large ip spaces with clockview. In *Proceedings of the 8th International Symposium on Visualization for Cyber Security*, VizSec '11, pages 2:1–2:10, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0679-9. doi: 10.1145/2016904.2016906.
- [127] D. Kosecki. Rem, light, deep: How much of each stage of sleep are you getting? <https://blog.fitbit.com/sleep-stages-explained/>, September 2018. [last accessed September 2019].
- [128] S. Kriglstein, M. Pohl, and M. Smuc. *Pep Up Your Time Machine: Recommendations for the Design of Information Visualizations of Time-Dependent Data*, pages 203–225. Springer New York, New York, NY, 2014. ISBN 978-1-4614-7485-2. doi: 10.1007/978-1-4614-7485-2_8.
- [129] P. Lam. How to move on? <https://public.tableau.com/en-gb/gallery/how-move?tab=viz-of-the-day&type=viz-of-the-day>, July 2019. [last accessed September 2019].
- [130] J. Larsen, A. Cuttone, and S. Jørgensen. Qs spiral: Visualizing periodic quantified self data. In *Proceedings of CHI 2013 Workshop on Personal Informatics in the Wild: Hacking Habits for Health & Happiness*, 2013.

- [131] T. Le, B. Reeder, H. J. Thompson, and G. Demiris. Health providers' perceptions of novel approaches to visualizing integrated health information. *Methods of information in medicine*, 52(3):250–258, 2013. ISSN 2227-9709. doi: 10.3414/ME12-01-0073.
- [132] T. Le, H. J. Thompson, and G. Demiris. A comparison of health visualization evaluation techniques with older adults. *IEEE Computer Graphics and Applications*, 36(4):67–77, July 2016. doi: 10.1109/MCG.2015.93.
- [133] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '06, pages 1–5, New York, NY, USA, 2006. ACM. ISBN 1-59593-562-2. doi: 10.1145/1168149.1168168.
- [134] I. Li, A. Dey, and J. Forlizzi. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 557–566, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-929-9. doi: 10.1145/1753326.1753409.
- [135] I. Li, A. K. Dey, and J. Forlizzi. Understanding my data, myself: Supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 405–414, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0630-0. doi: 10.1145/2030112.2030166.
- [136] I. Li, A. K. Dey, and J. Forlizzi. Using context to reveal factors that affect physical activity. *ACM Trans. Comput.-Hum. Interact.*, 19(1):7:1–7:21, May 2012. ISSN 1073-0516. doi: 10.1145/2147783.2147790.
- [137] J. Li, J. Qian, and F. Liang. Evidence for the beneficial effect of perceptual grouping on visual working memory: an empirical study on illusory contour and a meta-analytic study. *Scientific reports*, 8(1):13864–13864, 09 2018. doi: 10.1038/s41598-018-32039-4.
- [138] M. Lima. *The book of circles: Visualizing spheres of knowledge*. Princeton Architectural Press, The address, 1 edition, May 2017. ISBN 9781616895280.
- [139] J. J. Lin, L. Mamykina, S. Lindtner, G. Delajoux, and H. B. Strub. Fish'n'steps: Encouraging physical activity with an interactive computer game. In *Proceedings of the 8th International Conference on Ubiquitous Computing*, UbiComp'06, pages 261–278, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 978-3-540-39634-5. doi: 10.1007/11853565_16.

- [140] S. Lu and E. Meeks. VIZ PALETTE. <https://projects.susielu.com/viz-palette>, 2019. [last accessed June 2019].
- [141] C. Maçãs and P. Machado. Radial calendar of consumption. *2018 22nd International Conference Information Visualisation (IV)*, pages 96–102, 2018.
- [142] J. Mackinlay, P. Hanrahan, and C. Stolte. Show me: Automatic presentation for visual analysis. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1137–1144, Nov 2007. ISSN 1077-2626. doi: 10.1109/TVCG.2007.70594.
- [143] J. D. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5:110–141, 1986.
- [144] R. Maker. Garmin forerunner 935 in-depth review. <https://www.dcrainmaker.com/2017/03/garmin-forerunner-935-depth-review.html>, March 2017. [last accessed July 2017].
- [145] B. L. Massoudi, M. G. Olmsted, Y. Zhang, R. A. Carpenter, C. E. Barlow, and R. Huber. A web-based intervention to support increased physical activity among at-risk adults. *Journal of Biomedical Informatics*, 43(5):S41–S45, Oct. 2010. ISSN 1532-0464. doi: 10.1016/j.jbi.2010.07.012.
- [146] R. Mazza. *Introduction to Information Visualization*. Springer Publishing Company, Incorporated, 1 edition, 2009. ISBN 1848002181, 9781848002180.
- [147] S. Mennicken, J. Hofer, A. Dey, and E. M. Huang. Casalendar: A temporal interface for automated homes. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 2161–2166, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2474-8. doi: 10.1145/2559206.2581321.
- [148] J. Meyer and P. Eslambolchilar. Research challenges of emerging technologies supporting life-long health and wellbeing. In *MMHealth@MM*, 2017.
- [149] J. Meyer, A. Kazakova, M. Büsing, and S. Boll. Visualization of complex health data on mobile devices. In *Proceedings of the 2016 ACM Workshop on Multimedia for Personal Health and Health Care*, MMHealth '16, pages 31–34, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4518-7. doi: 10.1145/2985766.2985774.
- [150] A. V. Moere and H. Purchase. On the role of design in information visualization. *Information Visualization*, 10(4):356–371, 2011. doi: 10.1177/1473871611415996.
- [151] D. Moritz and D. Fisher. Visualizing a million time series with the density line chart. *ArXiv*, abs/1808.06019, 2018.

- [152] A. Morris. A guide to spss for information science, March 2012. URL <http://www.statstutor.ac.uk/resources/uploaded/spssguide1mar2012.pdf>. [last accessed February 2019].
- [153] A. Moutzoglou. *Mobile Health Applications for Quality Healthcare Delivery*. Advances in Healthcare Information Systems and Administration. IGI Global, 2019. ISBN 9781522580225.
- [154] K. Mullet and D. Sano. *Designing Visual Interfaces: Communication Oriented Techniques*. A Prentice Hall title. SunSoft Press, 1995. ISBN 9780133033892.
- [155] T. Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, Nov. 2009. ISSN 1077-2626. doi: 10.1109/TVCG.2009.111.
- [156] T. Munzner. *Visualization Analysis and Design*. A.K. Peters visualization series. A K Peters, 2014. ISBN 978-1-466-50891-0.
- [157] MyFitnessPal.com. MyFitnessPal. <https://www.myfitnesspal.com>, September 2019. [last accessed September 2019].
- [158] P. Mylavarapu, A. Yalcin, X. Gregg, and N. Elmqvist. Ranked-list visualization: A graphical perception study. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 192:1–192:12, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300422.
- [159] NHS. Pedometers make those steps count. <https://www.nhs.uk/news/lifestyle-and-exercise/pedometers-make-those-steps-count/>, November 2007. [last accessed October 2019].
- [160] NHS. Worrying about work out-of-hours may be bad for the heart. <https://www.nhs.uk/news/mental-health/worrying-about-work-out-of-hours-may-be-bad-for-the-heart/>, February 2017. [last accessed October 2019].
- [161] NHS. How do i check my pulse? <https://www.nhs.uk/common-health-questions/accidents-first-aid-and-treatments/how-do-i-check-my-pulse/>, February 2018. [last accessed September 2018].
- [162] NHS. Understanding calories. <https://www.nhs.uk/live-well/healthy-weight/understanding-calories/>, August 2019. [last accessed October 2019].

- [163] J. Noble and N. Joshua. *Programming Interactivity: A Designer's Guide to Processing, Arduino, and Openframeworks*. O'Reilly Media, Inc., 1st edition, 2009. ISBN 0596154143, 9780596154141.
- [164] NVivo. NVivo. <https://www.qsrinternational.com/nvivo/home>, April 2019. [last accessed April 2019].
- [165] A. Oglethorpe. 6 Ways Continuous Heart Rate Tracking Gets You Closer to Your Goals. , September 2018. URL <https://blog.fitbit.com/sleep-stages-explained/>. [last accessed September 2019].
- [166] A. O'Hara. All the new health & fitness features coming to ios 13 & watchos 6. <https://appleinsider.com/articles/19/09/12/all-the-new-health-fitness-features-coming-to-ios-13-watchos-6>, September 2019. [last accessed September 2019].
- [167] M. G. Olmsted, B. L. Massoudi, and Y. Zhang. What consumers want in personal health applications: findings from project healthdesign. *Personal and Ubiquitous Computing*, 19(1):79–83, Jan 2015. ISSN 1617-4917. doi: 10.1007/s00779-014-0811-2.
- [168] B. Pagno and L. Nedel. Everyday visualization. In *Electronic proceedings of the IEEE VIS 2015 workshop Personal Visualization: Exploring Data in Everyday Life*, Chicago, October 2015.
- [169] V. Palladino. Garmin Vivosmart 4 review: An affordable band for easy fitness, SpO2 data. <https://arstechnica.com/gadgets/2018/10/garmin-vivosmart-4-review-the-attractive-new-face-of-the-vivosmart-line/>, October 2018. [last accessed September 2019].
- [170] R. Pandey. Rule of law around the world. <https://public.tableau.com/en-gb/gallery/?tab=viz-of-the-day&type=viz-of-the-day>, September 2019. [last accessed September 2019].
- [171] S. Pinker. A theory of graph comprehension. In *Artificial intelligence and the future of testing.*, pages 73–126. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1990. ISBN 0-8058-0117-0 (Hardcover).
- [172] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '04*, pages 109–116, New York, NY, USA, 2004. ACM. ISBN 1-58113-867-9. doi: 10.1145/989863.989880.
- [173] B. Plantinga. Naomi osaka vs. serena williams. <https://public.tableau.com/en-gb/gallery/naomi-osaka-vs-serena-williams-0?tab=viz-of-the-day&type=viz-of-the-day>, September 2019. [last accessed September 2019].

- [174] Z. Pousman, J. Stasko, and M. Mateas. Casual information visualization: Depictions of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1145–1152, Nov. 2007. ISSN 1077-2626. doi: 10.1109/TVCG.2007.70541.
- [175] Quantified Self Labs. Qs access. <https://itunes.apple.com/gb/app/qs-access/id920297614?mt=8>, 2015. [Mobile application software (Version 1.3)], [last accessed July 2017].
- [176] F. Rajabiyazdi, C. Perin, L. Oehlberg, and S. Carpendale. Personal Patient-Generated Data Visualizations for Diabetes Patients. In *IEEE VIS 2018 Electronic Conference Proceedings*, Berlin, Germany, Oct. 2018.
- [177] A. Rapp and F. Cena. Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *International Journal of Human-Computer Studies*, 94:1 – 17, 2016. ISSN 1071-5819. doi: <https://doi.org/10.1016/j.ijhcs.2016.05.006>.
- [178] C. Reas, B. Fry, and J. Maeda. *Processing: A Programming Handbook for Visual Designers and Artists*. Putty-clay and investment: a business cycle analysis. National Bureau of Economic Research, 2007. ISBN 9780262182621.
- [179] J. Redström, T. Skog, and L. Hallnäs. Informative art: Using amplified artworks as information displays. In *Proceedings of DARE 2000 on Designing Augmented Reality Environments*, DARE '00, pages 103–114, New York, NY, USA, 2000. ACM. doi: 10.1145/354666.354677.
- [180] Research2Guidance. mHealth App Developer Economics 2016. Technical Report 6th edition, Research2Guidance. The Mobile App Strategy Company, October 2016. URL <https://research2guidance.com/product/mhealth-app-developer-economics-2016/>. [Last Accessed June 2017].
- [181] Research2Guidance. mhealth app economics 2017/2018 current status and future trends in mobile health. Technical Report 7th edition, Research2Guidance. The Mobile App Strategy Company, November 2017. URL <https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/>. [Last Accessed October 2019].
- [182] A. Rind, T. Lammarsch, W. Aigner, B. Alsallakh, and S. Miksch. Timebench: A data model and software library for visual analytics of time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2247–2256, Dec. 2013. ISSN 1077-2626. doi: 10.1109/TVCG.2013.206.

- [183] J. Robertson and M. Kaptein. *Modern statistical methods for HCI*. Human-computer interaction series. Springer, Cham, 2016. URL <https://cds.cern.ch/record/2144810>.
- [184] R. D. Rogers and S. Monsell. Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124(2):207–231, 1995. ISSN 1939-2222(Electronic),0096-3445(Print). doi: 10.1037/0096-3445.124.2.207.
- [185] J. Rooksby, M. Rost, A. Morrison, and M. C. Chalmers. Personal tracking as lived informatics. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14*, pages 1163–1172, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557039.
- [186] B. Saket, A. Endert, and ç. Demiralp. Task-based effectiveness of basic visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(7):2505–2512, July 2019. ISSN 1077-2626. doi: 10.1109/TVCG.2018.2829750.
- [187] J. Saldana. *The Coding Manual for Qualitative Researchers*. SAGE Publications, 2012. ISBN 9781446271421.
- [188] Samsung. Samsung gear s3: Official launch film, August 2016. URL <https://www.youtube.com/watch?v=ynTB211p1PQ>. [last accessed July 2017].
- [189] Samsung. *Samsung User Manual*. Samsung, April 2017. URL http://downloadcenter.samsung.com/content/UM/201701/20170117210121872/SM-R760_R765_R770_UM_EU_Tizen_Eng_Rev.1.1_170111.pdf. [last accessed July 2017].
- [190] Samsung. Samsung Community/ Gear Sport Not Tracking Heart Rate When Exercising, March 2018. URL <https://eu.community.samsung.com/t5/Wearables/Gear-Sport-Not-Tracking-Heart-Rate-When-Exercising/td-p/461604>. [last accessed September 2019].
- [191] Samsung. Your health in your hands. <https://www.samsung.com/global/galaxy/apps/samsung-health/>, September 2019. [last accessed September 2019].
- [192] Samsung. Samsung mobile youtube channel. <https://www.youtube.com/user/SamsungMobile/videos>, n.d. [last accessed July 2017].
- [193] Samsung Electronics Co., Ltd. Samsung Health. <https://play.google.com/store/apps/details?id=com.sec.android.app.shealth>, 06 2017. [Mobile application software (Version 5.9.1.003)], [last accessed July 2017].

- [194] Samsung Newsroom. Track, manage, improve: Better health with s health app. <https://news.samsung.com/global/track-manage-improve-better-health-with-s-health-app>, May 2015. [last accessed July 2017].
- [195] Samsung Newsroom. [Hands-On] The Samsung Gear S3. <https://news.samsung.com/global/hands-on-the-samsung-gear-s3>, September 2016. [last accessed July 2017].
- [196] Samsung Newsroom. New S Health Features Add Fun to Fitness. <http://www.samsung.com/us/support/answer/ANS00061417/>, July 2016. [last accessed July 2017].
- [197] Samsung Newsroom. New s health lets users team up to tone up. <https://news.samsung.com/global/new-s-health-lets-users-team-up-to-tone-up>, August 2016. [last accessed July 2017].
- [198] Samsung Newsroom. Samsung releases comprehensive gear s3 value pack update. <https://news.samsung.com/global/samsung-releases-comprehensive-gear-s3-value-pack-update>, April 2017. [last accessed July 2017].
- [199] B. S. Santos and P. Dias. Evaluation in visualization: some issues and best practices. In *Visualization and Data Analysis*, 2014.
- [200] M. Sawh. Garmin forerunner 935 review. <https://www.wearable.com/garmin/garmin-forerunner-935-review>, May 2017. [last accessed July 2017].
- [201] A. Schlegel. controlp5: A gui library for processing. , July 2015. URL <http://www.sojamo.de/libraries/controlP5>. [last accessed February 2020].
- [202] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2366–2375, Dec. 2013. ISSN 1077-2626. doi: 10.1109/TVCG.2013.120.
- [203] R. Sekuler and R. Blake. *Perception*. McGraw-Hill, 2002. ISBN 9780071121590.
- [204] K. J. Serrano, K. I. Coa, M. Yu, D. L. Wolff-Hughes, and A. A. Atienza. Characterizing user engagement with health app data: a data mining approach. *Translational Behavioral Medicine*, 7:277–285, 2017.
- [205] W. Shijia. 50 years of china’s space journey. <https://public.tableau.com/en-gb/gallery/50-years-chinas-space-journey?tab=viz-of-the-day&type=viz-of-the-day>, August 2019. [last accessed September 2019].

- [206] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, pages 336–343, Sep. 1996. doi: 10.1109/VL.1996.545307.
- [207] B. Shneiderman, C. Plaisant, and B. W. Hesse. Improving healthcare with interactive visualization. *IEEE Computer*, 46(5):58–66, May 2013. ISSN 0018-9162. doi: 10.1109/MC.2013.38.
- [208] W. A. Simcox. A design method for graphic communication. *ABCA Bulletin*, 47(1):3–7, 1984. doi: 10.1177/108056998404700101.
- [209] D. Simkin and R. Hastie. An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82(398):454–465, 1987. doi: 10.1080/01621459.1987.10478448.
- [210] E. Singer and M. P. Couper. Some methodological uses of responses to open questions and other verbatim comments in quantitative surveys. *methods, data, analyses*, 11(2): 20, 2017. ISSN 2190-4936. doi: 10.12758/mda.2017.01.
- [211] A. Srinivasan, M. Brehmer, B. Lee, and S. M. Drucker. What’s the difference?: Evaluating variations of multi-series bar charts for visual comparison tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI ’18, pages 304:1–304:12, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5620-6. doi: 10.1145/3173574.3173878.
- [212] B. Steele, J. Chandler, and S. Reddy. *Data Visualization*, pages 133–159. Springer International Publishing, Cham, 2016. ISBN 978-3-319-45797-0. doi: 10.1007/978-3-319-45797-0_5.
- [213] R. Sternberg, K. Sternberg, and J. Mio. *Cognitive Psychology*. Wadsworth/Cengage Learning, 2011. ISBN 9781133313915.
- [214] L. Stopar, P. Skraba, M. Grobelnik, and D. Mladenic. Streamstory: Exploring multivariate time series on multiple scales. *IEEE Transactions on Visualization and Computer Graphics*, 25(4):1788–1802, April 2019. ISSN 2160-9306. doi: 10.1109/TVCG.2018.2825424.
- [215] Su-In Lee. Lecture notes in statistics for genome scientists. <https://homes.cs.washington.edu/~suinlee/genome560/handouts.html>, May 2012. [last accessed January 2019].

- [216] Tableau. Tableau Gallery. <https://public.tableau.com/en-gb/gallery/>, September 2019. [last accessed September 2019].
- [217] J. Talbot, V. Setlur, and A. Anand. Four experiments on the perception of bar charts. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2152–2160, Dec 2014. ISSN 1077-2626. doi: 10.1109/TVCG.2014.2346320.
- [218] L. M. Tang and J. Kay. Harnessing long term physical activity data-how long-term trackers use data and how an adherence-based interface supports new insights. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(2), June 2017. doi: 10.1145/3090091.
- [219] L. M. Tang, J. Meyer, D. A. Epstein, K. Bragg, L. Engelen, A. Bauman, and J. Kay. Defining adherence: Making sense of physical activity tracker data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1):37:1–37:22, Mar. 2018. ISSN 2474-9567. doi: 10.1145/3191769.
- [220] The American Heart Association (AHA). Target Heart Rates. ONLINE, October 2016. URL http://www.heart.org/HEARTORG/HealthyLiving/PhysicalActivity/FitnessBasics/Target-HeartRates_UCM_434341_Article.jsp. [last accessed May 2019].
- [221] S. Theis, P. W. V. Rasche, C. Bröhl, M. Wille, and A. Mertens. Task-data taxonomy for health data visualizations: Web-based survey with experts and older adults. *JMIR medical informatics*, 6(3):e39–e39, 07 2018. doi: 10.2196/medinform.9394.
- [222] A. Thudt, D. Baur, S. Huron, and S. Carpendale. Visual mementos: Reflecting memories with personal data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1): 369–378, Jan 2016. ISSN 1077-2626. doi: 10.1109/TVCG.2015.2467831.
- [223] A. Thudt, B. Lee, E. K. Choe, and S. Carpendale. Expanding research methods for a realistic understanding of personal visualization. *IEEE Computer Graphics and Applications*, 37(2):12–18, Mar 2017. ISSN 0272-1716. doi: 10.1109/MCG.2017.23.
- [224] C. Tominski, J. Abello, and H. Schumann. Axes-based visualizations with radial layouts. In *Proceedings of the 2004 ACM Symposium on Applied Computing, SAC '04*, page 1242–1247, New York, NY, USA, 2004. Association for Computing Machinery. ISBN 1581138121. doi: 10.1145/967900.968153.
- [225] X. Tong, D. Gromala, C. Shaw, and P. Clarke. Transformation between electronic arts and chronic pain: Long-term body activity data visualisation and pain animation expression. In *Proceedings of the EVA London 2014 on Electronic Visualisation and the*

- Arts, EVA London 2014, pages 75–81, UK, 2014. BCS. ISBN 978-1-78017-285-9. doi: 10.14236/ewic/eva2014.11.
- [226] X. Tong, D. Gromala, L. Bartram, F. Rajabiyazdi, and S. Carpendale. Evaluating the effectiveness of three physical activity visualizations ? how people perform vs. perceive. In *Electronic proceedings of the IEEE VIS 2015 workshop Personal Visualization: Exploring Data in Everyday Life*, Chicago, October 2015.
- [227] M. Tory. *User Studies in Visualization: A Reflection on Methods*, pages 411–426. Springer New York, New York, NY, 2014. ISBN 978-1-4614-7485-2. doi: 10.1007/978-1-4614-7485-2_16.
- [228] M. Tory and T. Moller. Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 10(1):72–84, Jan 2004. doi: 10.1109/TVCG.2004.1260759.
- [229] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97 – 136, 1980. ISSN 0010-0285. doi: [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5).
- [230] L. M. Trick and J. T. Enns. Lifespan changes in attention: The visual search task. *Cognitive Development*, 13(3):369 – 386, 1998. ISSN 0885-2014. doi: [https://doi.org/10.1016/S0885-2014\(98\)90016-8](https://doi.org/10.1016/S0885-2014(98)90016-8).
- [231] C. Tudor-Locke and D. Bassett. How many steps are enough? pedometer-determined physical activity indices. *Sports Med*, 34(1):1–8, 2004.
- [232] E. Tufte. *Envisioning Information*. Graphics Press, Cheshire, CT, USA, 1990. ISBN 0-9613921-1-8.
- [233] T. TV. Every samsung gear s3 app, in depth review, April 2017. URL <https://www.youtube.com/watch?v=O--f2mmOFp0>. [last accessed July 2017].
- [234] E. R. A. Valiati, M. S. Pimenta, and C. M. D. S. Freitas. A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '06, pages 1–6, New York, NY, USA, 2006. ACM. ISBN 1-59593-562-2. doi: 10.1145/1168149.1168169.
- [235] E. T. K. van Dijk, J. H. Westerink, F. Beute, and W. A. IJsselsteijn. Personal informatics, self-insight, and behavior change: A critical review of current literature. *Human-Computer Interaction*, 32(5-6):268–296, 2017. doi: 10.1080/07370024.2016.1276456.

- [236] M. Vandemeulebroecke, M. Baillie, D. Carr, L. Kanitra, A. Margolskee, A. Wright, and B. Magnusson. How can we make better graphs? an initiative to increase the graphical expertise and productivity of quantitative scientists. *Pharmaceutical Statistics*, 18(1): 106–114, 2019. doi: 10.1002/pst.1912.
- [237] N. X. Verdezoto and J. W. Olsen. Personalized medication management: Towards a design of individualized support for elderly citizens at home. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, IHI '12*, page 813–818, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450307819. doi: 10.1145/2110363.2110463.
- [238] M. Waldner, A. Diehl, D. Gračanin, R. Splechtna, C. Delrieux, and K. Matković. A comparison of radial and linear charts for visualizing daily patterns. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2019. ISSN 1077-2626. doi: 10.1109/TVCG.2019.2934784.
- [239] C. Ware. *Visual Thinking: For Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2008. ISBN 0123708966, 9780123708960.
- [240] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3 edition, 2012. ISBN 9780123814647, 9780123814654.
- [241] S. Wehrend and C. Lewis. A problem-oriented classification of visualization techniques. In *Proceedings of the First IEEE Conference on Visualization: Visualization '90*, pages 139–143, Oct 1990. doi: 10.1109/VISUAL.1990.146375.
- [242] C. D. Wickens and C. M. Carswell. *Information Processing*, chapter 5, pages 111–149. John Wiley & Sons, Ltd, 2006. ISBN 9780470048207. doi: 10.1002/0470048204.ch5.
- [243] C. D. Wickens and J. G. Hollands. *Engineering Psychology and Human Performance*. Upper Saddle River, NJ : Prentice Hall, 3rd ed edition, 2000. ISBN 0321047117 (hbk.). Includes bibliographical references and index.
- [244] M. Winckler, P. Palanque, and C. Freitas. Tasks and scenario-based evaluation of information visualization techniques. In *Proceedings of the 3rd annual conference on task models and diagrams*, volume 86 of *TAMODIA '04*, pages 165–172. ACM, 2004. ISBN 1595930000.
- [245] J. M. Wolfe. Visual search. In *Attention.*, pages 13–73. Psychology Press/Erlbaum (UK) Taylor & Francis, Hove, England, 1998. ISBN 0-86377-812-7 (Hardcover); 0-86377-813-5 (Paperback).

- [246] WonderApps AB. Atracker pro time tracker. <https://apps.apple.com/gb/app/atracker-pro-time-tracker/id519561306>, September 2019. [last accessed September 2019].
- [247] J. M. Zacks and B. Tversky. Bars and lines: A study of graphic communication. *Memory & Cognition*, 27:1073–1079, 1999.
- [248] J. Zhao, F. Chevalier, and R. Balakrishnan. KronoMiner: Using multi-foci navigation for the visual exploration of time-series data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1737–1746, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979195.
- [249] M. X. Zhou and S. K. Feiner. Visual task characterization for automated visual discourse synthesis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, pages 392–399, New York, NY, USA, 1998. ACM Press/Addison-Wesley Publishing Co. ISBN 0-201-30987-4. doi: 10.1145/274644.274698.
- [250] H. Zhu, J. Colgan, M. Reddy, and E. K. Choe. Sharing Patient-Generated Data in Clinical Practices: An Interview Study, 2016. ISSN 1942-597X (Electronic).
- [251] T. Zuk, L. Schlesier, P. Neumann, M. S. Hancock, and S. Carpendale. Heuristics for information visualization evaluation. In *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '06, pages 1–6, New York, NY, USA, 2006. ACM. ISBN 1-59593-562-2. doi: 10.1145/1168149.1168162.

Appendices

Appendix

Preliminary Evaluation Study Surveys

.1 Preliminary Evaluation Study: Demographic and Pre-Experiment Questionnaire



Demographic_and_Pre-Experiment_Questionnaire_Final_

Page 1: Research Purpose and Consents

Evaluating Different Visualization Designs for Multivariate Personal Health Data: A Pilot Study

What is the purpose of the study?

The aim of this study is to identify the effectiveness of different visualization designs of multivariate personal activity data by comparing users' performance when performing a set of tasks on the proposed designs. The study also aims to compare users' performance to their preferences in the visualizations by collecting qualitative data from the questionnaires. Filling this questionnaire is the first part of the study, it will be followed by Performing an experiment which consists of 5 tasks' questions for 8 visualization designs and then followed by another questionnaire to provide your feedback about the visualizations.

Page 2: Participants' Information

1. Please Enter your participation ID * *Required*

2. Please Select Your Age Group

- < 20
- 20 - 29
- 30 - 39
- 40 - 49
- 50 - 59
- >= 60
- Prefer Not to Say

3. Please Select Your Gender

- Male
- Female
- Prefer Not to Say

4. What is the Highest Degree or Level of School You Have Completed?

- High School Graduate
- Undergraduate Degree
- Postgraduate Degree
- Prefer Not to Say
- Other

4.a. If you selected Other, please specify:

5. Are you Tracking or Have Ever Tracked your Health or Physical Activity Data?

Yes

No

Prefer Not to Say

5.a. If (Yes), What Type of Data Do (or Did) you Track?

- Step Count
- Walking/Running/Cycling Distance
- Flight Climbed
- Active/Burned Calories
- Heart Rate
- Sleep
- Other

5.a.i. If you selected Other, please specify:

5.a.ii. If (Yes), What are you Using for Tracking your Health/ Physical Activity Data?

- Fitbit Tracker
- Apple Watch
- Garmin Watch/Tracker
- Samsung Watch

- Android Watch
- Jawbone Up
- iPhone Apps
- Android Apps
- Other

5.a.ii.a. If you selected Other, please specify:

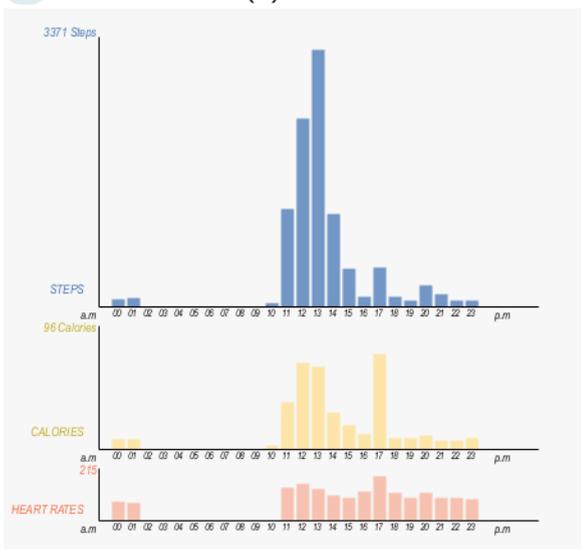
Page 3: Visualization Designs

Visualization Designs

This Section displays eight different designs of visualization that represent 3 different variables of physical activity data, which are step count, active calories and heart rate. The aim of these questions is to rate each design in term of readability (easy to read and understand) and in term of its Attractiveness.

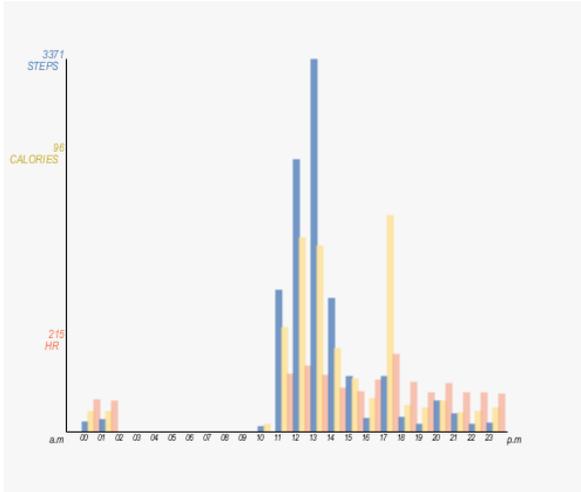
Rate the Following Visualization Designs in Term of its Readability and its Attractiveness

6. Visualization (1)



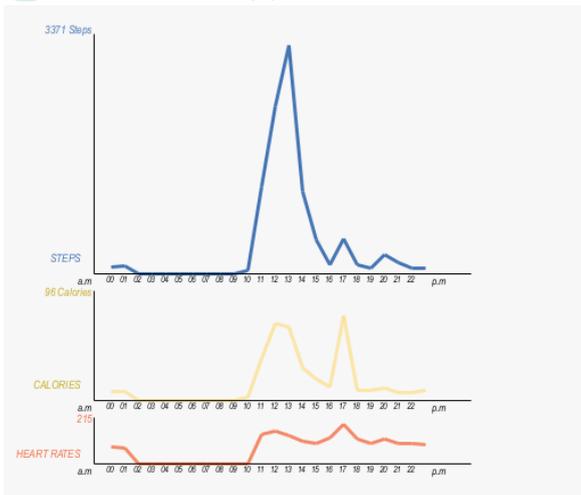
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

7. Visualization (2)



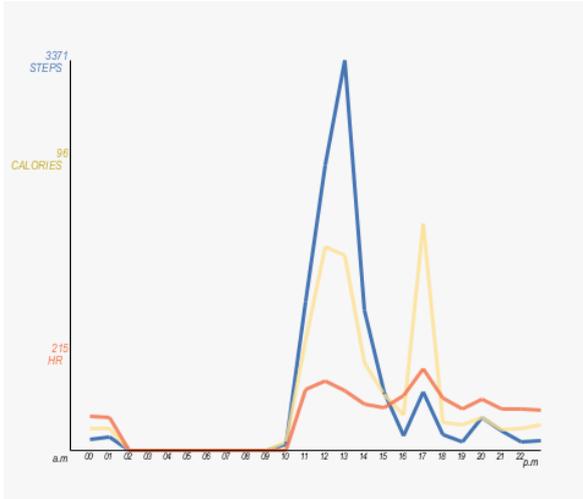
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

8. Visualization (3)



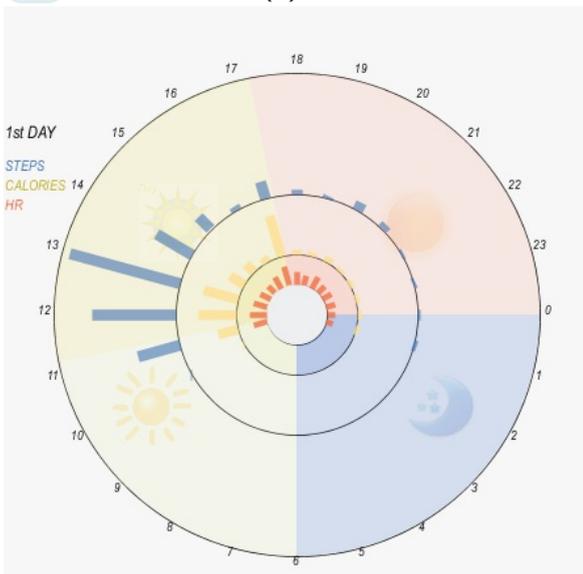
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

9. Visualization (4)



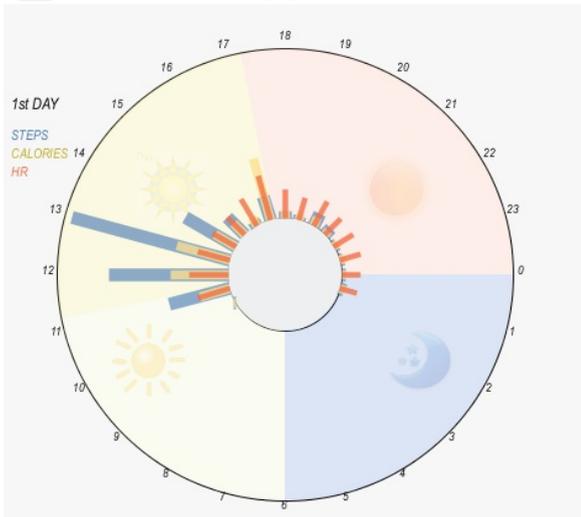
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

10. Visualization (5)



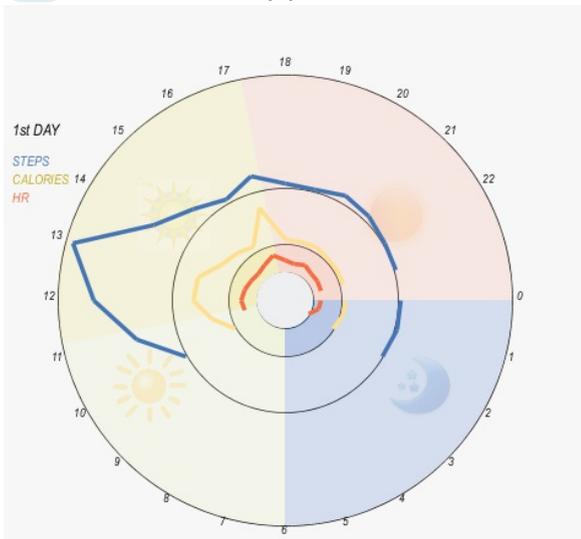
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

11. Visualization (6)



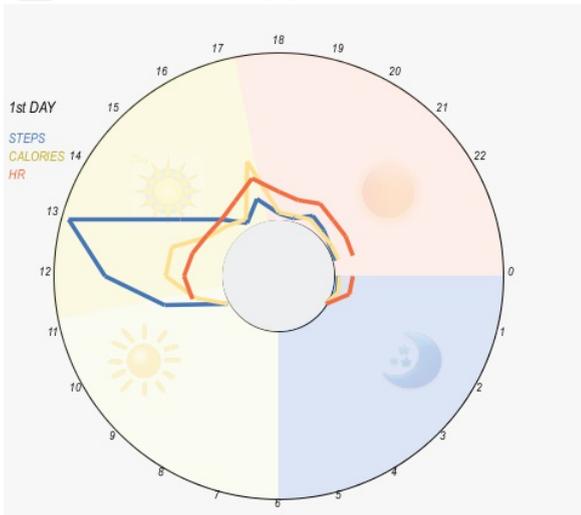
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

12. Visualization (7)



	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

13. Visualization (8)



	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

14. Which of the Previous Visualization Designs Would you Prefer to Visualize your Daily Physical Activity Data?

- Visualization (1)
- Visualization (2)
- Visualization (3)
- Visualization (4)
- Visualization (5)
- Visualization (6)
- Visualization (7)
- Visualization (8)
- I don't Prefer any of the visualizations

.2 Preliminary Evaluation Study: Post-Experiment Questionnaire



Post-Experiment_ Questionnaire

Page 1: Visualization Designs

Participant's Information

Please Enter your Participation ID * *Required*

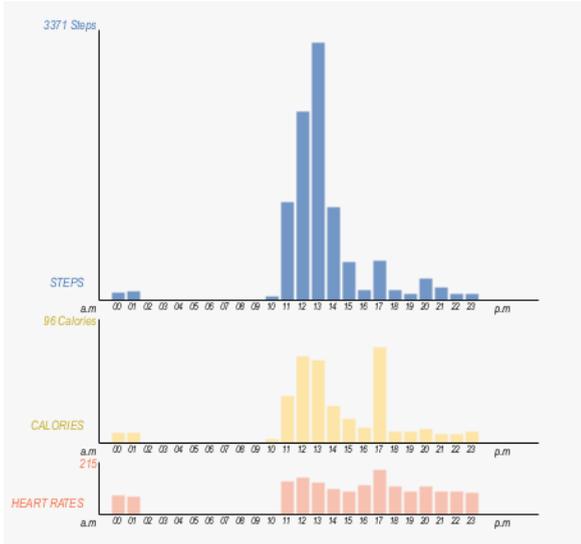
Visualization Designs

This Section displays eight different designs of visualization that represent 3 different variables of physical activity data, which are step count, active calories and heart rate. The aim of these questions is to rate each design in term of readability (easy to read and understand) and in term of its Attractiveness.

Rate the Following Visualization Designs in Term of its Readability and its Attractiveness

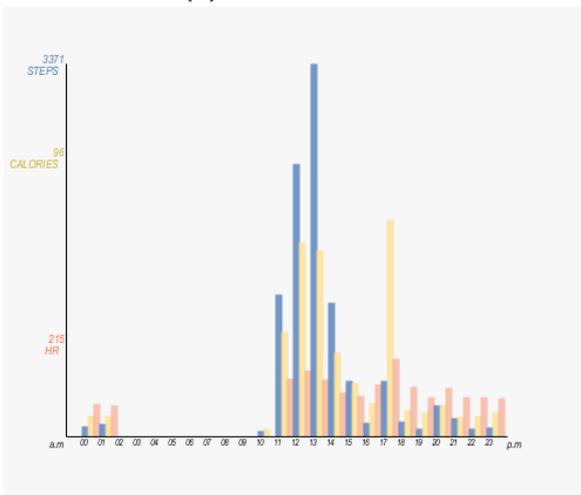
Visualization (1)

Please don't select more than 1 answer(s) per row.



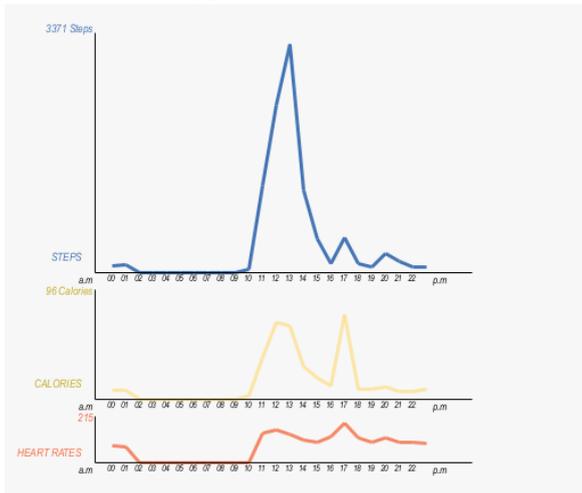
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (2)



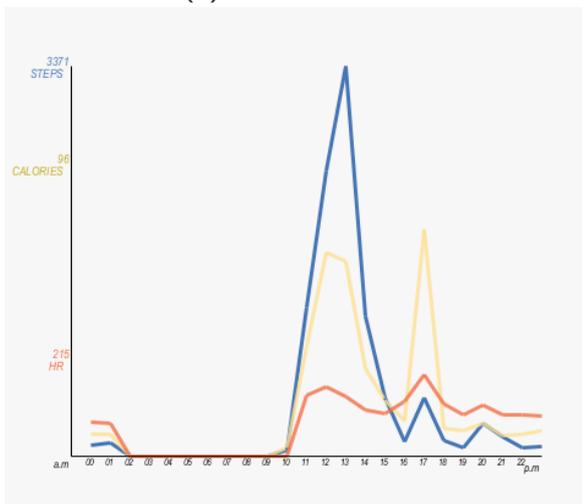
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (3)



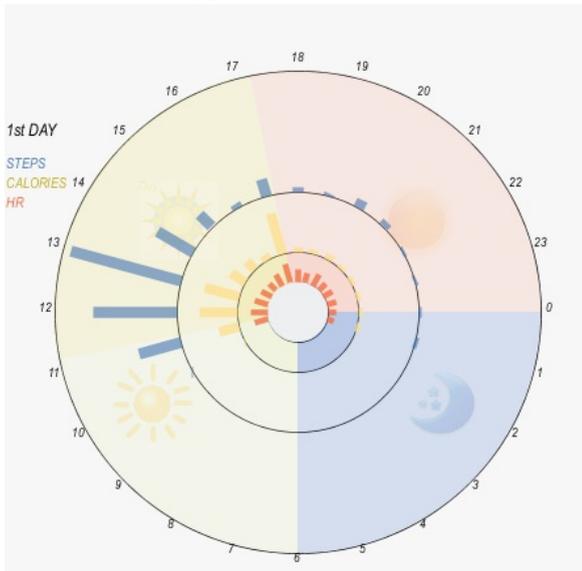
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (4)



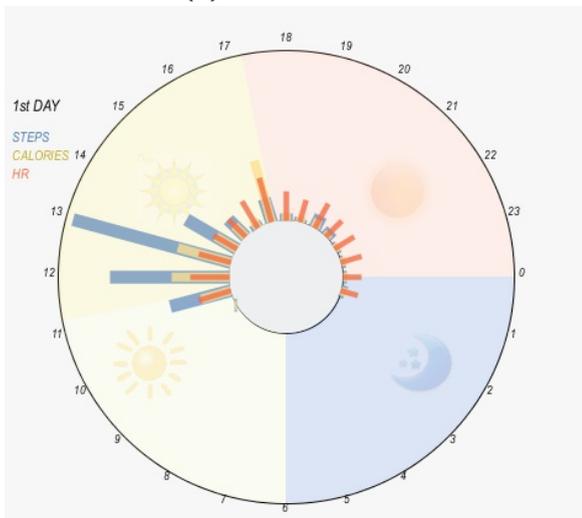
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (5)



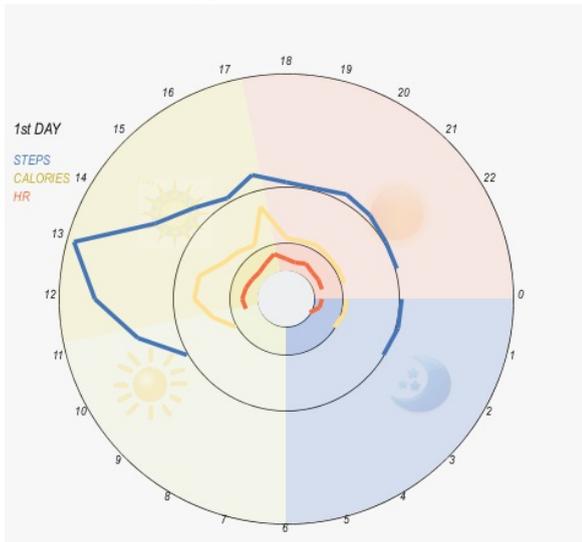
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (6)



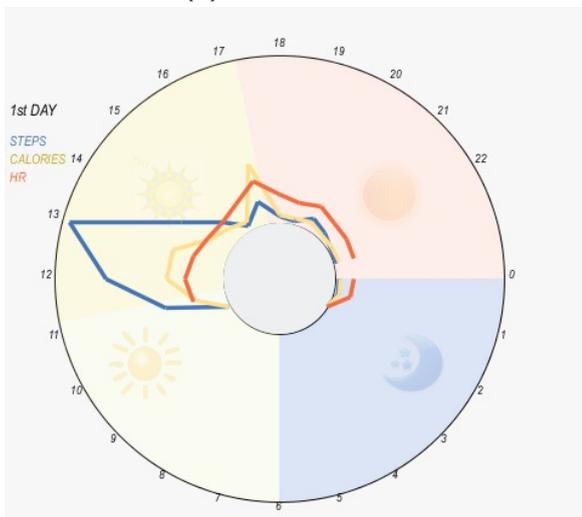
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (7)



	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (8)

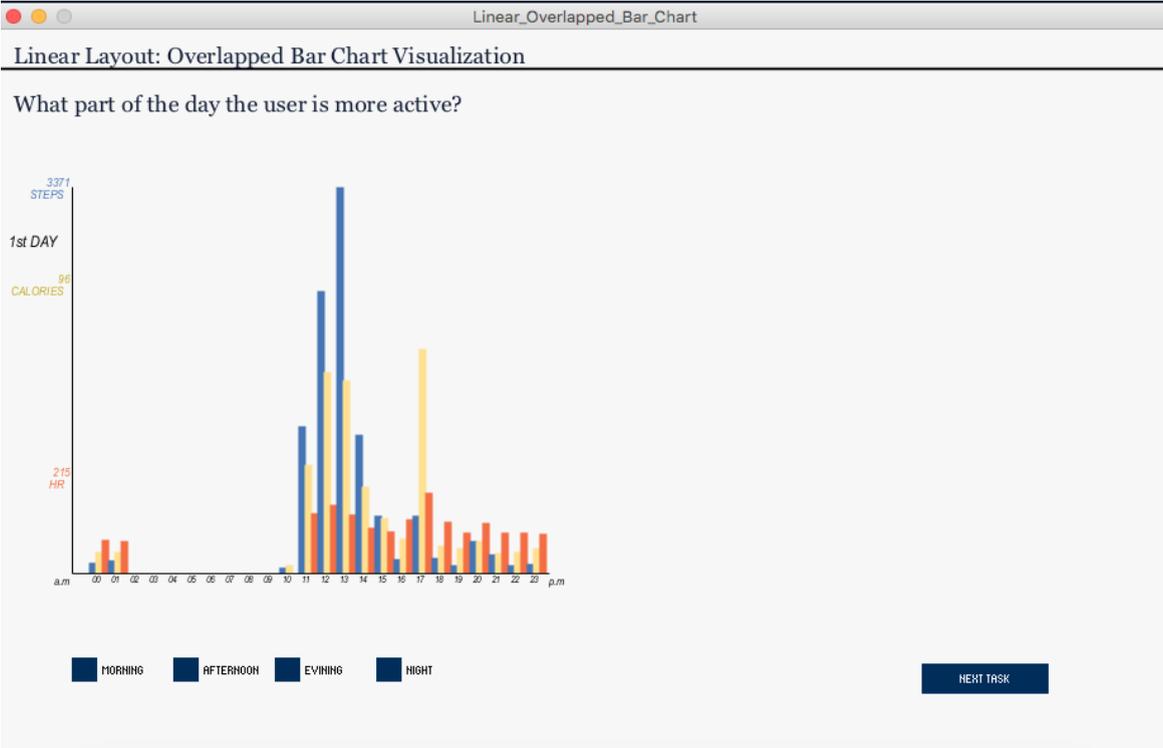
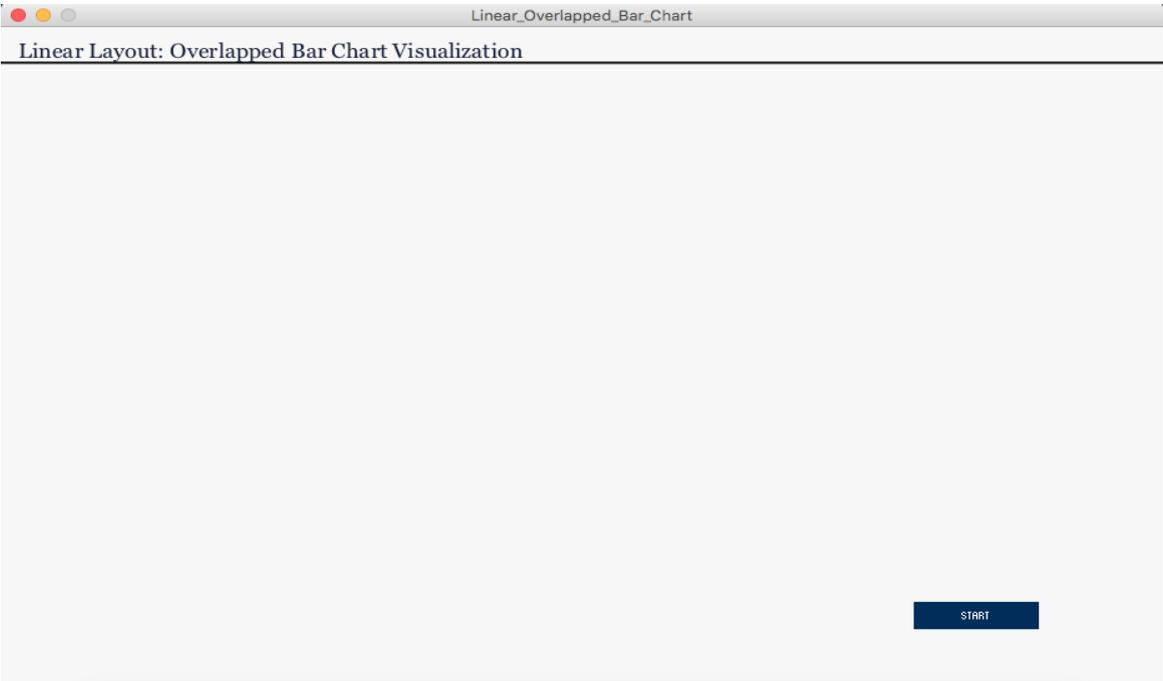


	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Which of the Previous Visualization Designs Would you Prefer to Visualize your Daily Physical Activity Data?

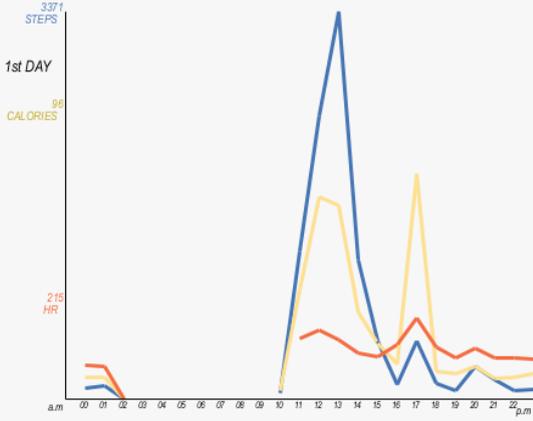
- Visualization (1)
- Visualization (2)
- Visualization (3)
- Visualization (4)
- Visualization (5)
- Visualization (6)
- Visualization (7)
- Visualization (8)
- I don't Prefer any of the visualizations

.3 Preliminary Evaluation Study: Training Sessions Samples



Linear Layout: Overlapped Line Chart Visualization

Find the number of step count at 01:00 P.m. at the shown day visualization

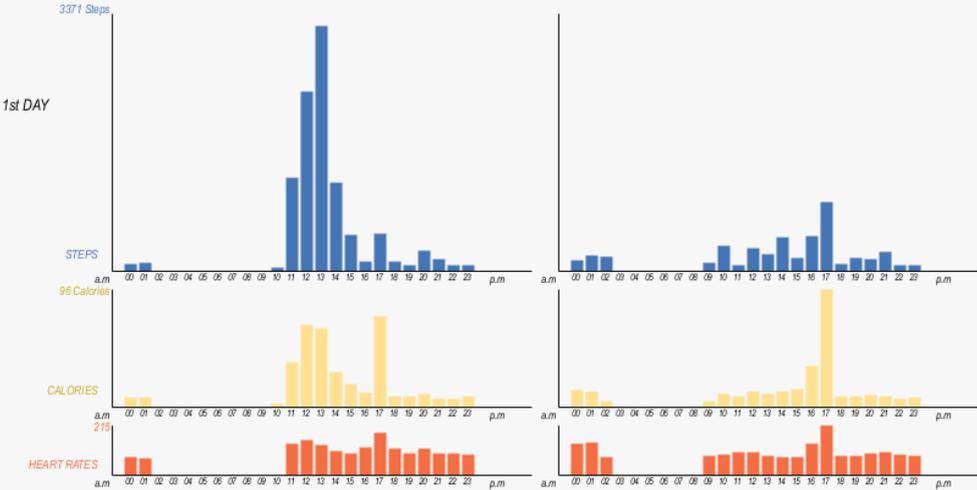


ANSWER:

NEXT TASK

Linear Layout: Stacked Bar Chart Visualization

Compare the active calories for the two days presented, which day has higher active calories?

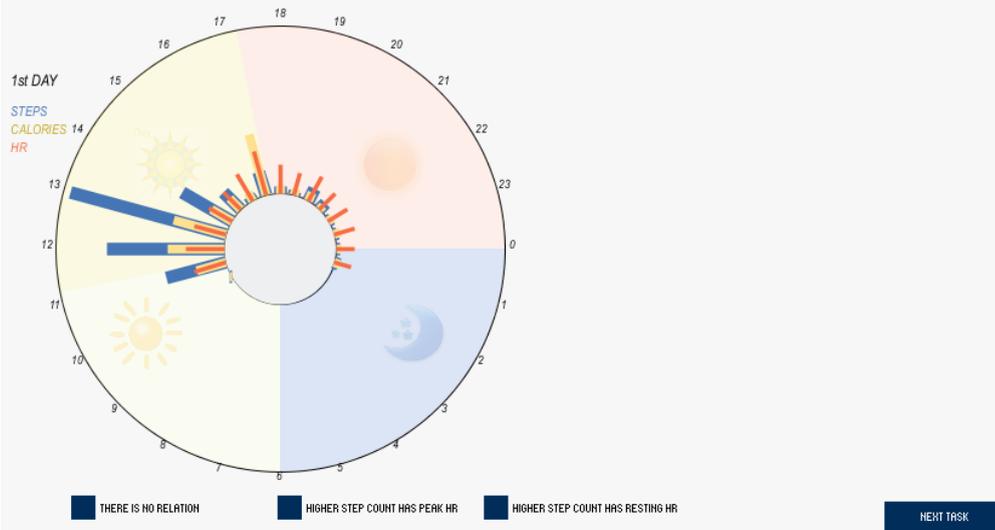


FIRST DAY SECOND DAY BOTH ARE THE SAME

NEXT TASK

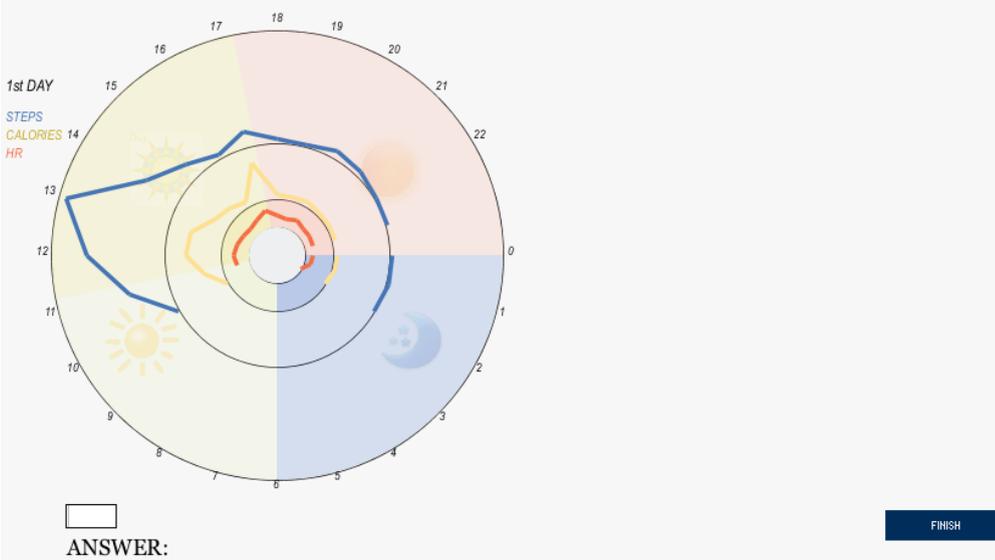
Radial Layout: Overlapped Bar Chart Visualization

What is the relation between HR and step count that you can identify in the presented visualization?



Radial Layout: Stacked Line Chart Visualization

What hour(s) of the day the user has the minimum HR average?



.4 Preliminary Evaluation Study: Group's Accuracy Graphs

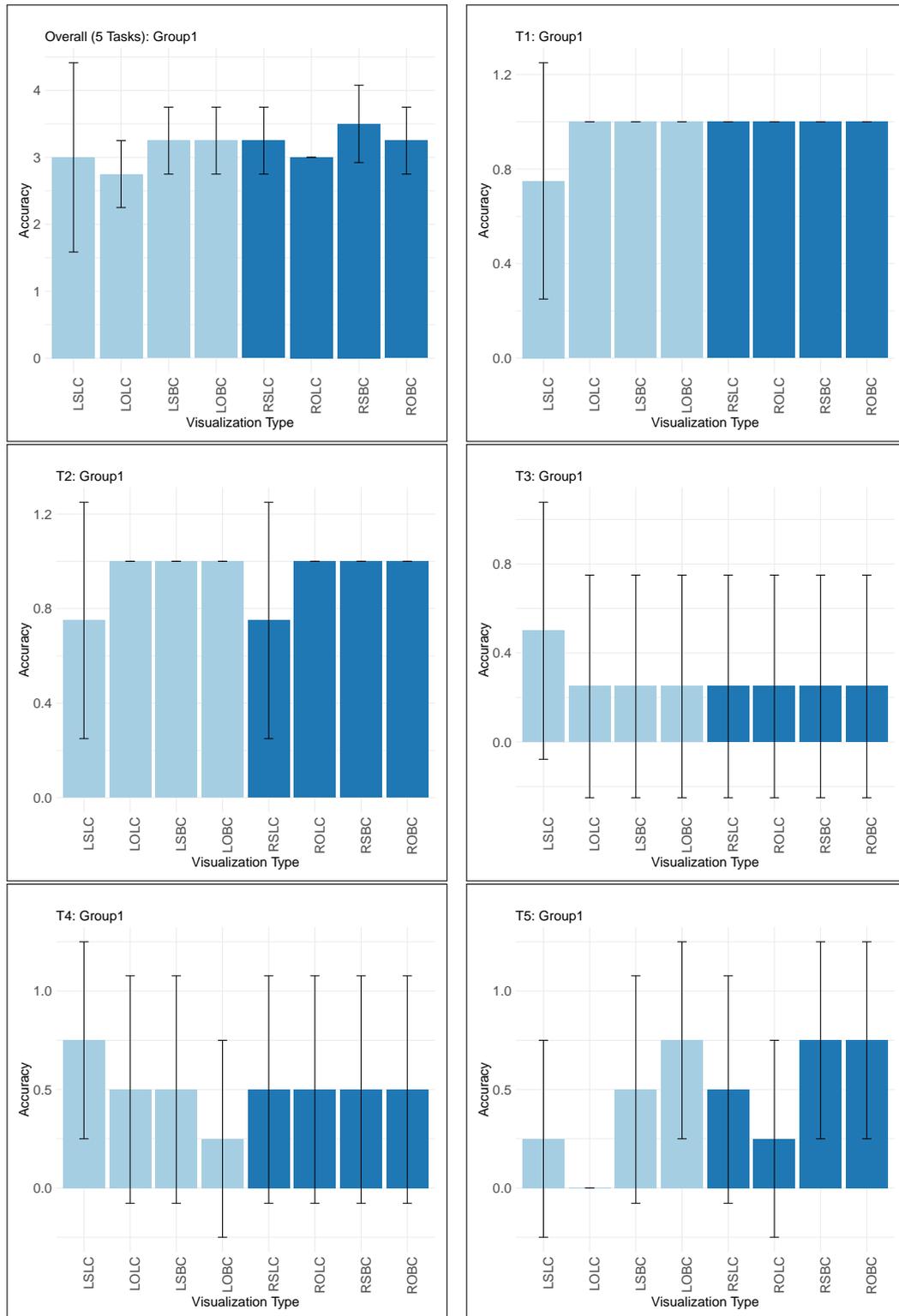


Figure 1: Accuracy of All Tasks for the Eight Visualizations for Group 1 (in the order presented). The top left figure shows the mean and the SD of the overall task accuracy for the eight visualizations. The Y axis represents the average accuracy of the five tasks, while the X axis shows the visualizations based on their presentation order for group 1. The other five subfigures show the average accuracy for each task individually (layout: Linear ■ | Radial ■).

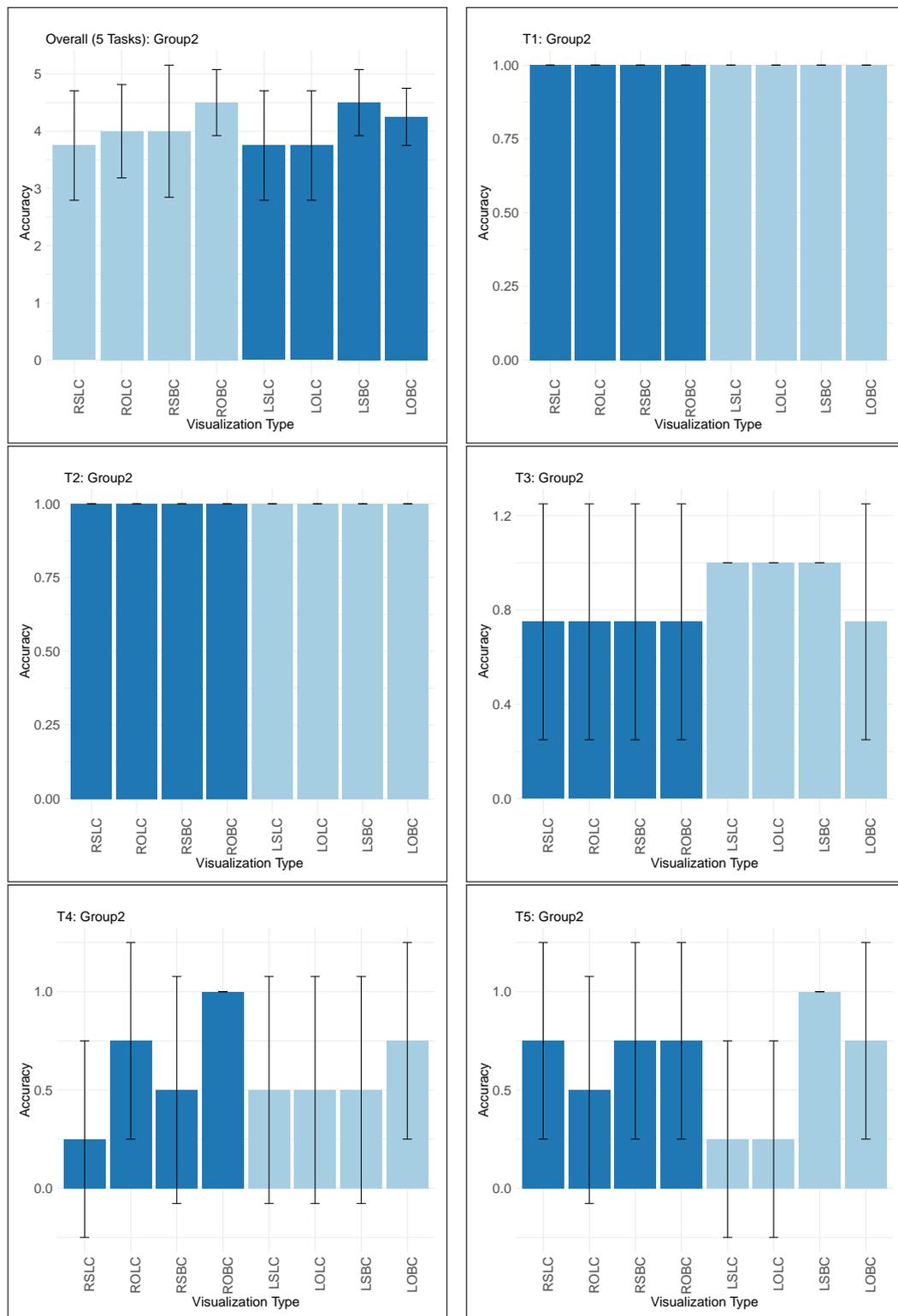


Figure 2: Accuracy of All Tasks for the Eight Visualizations for Group 2 (in the order presented). The top left figure shows the mean and the SD of the overall task accuracy for the eight visualizations. The Y axis represents the average accuracy of the five tasks, while the X axis shows the visualizations based on their presentation order for group 2. The other five subfigures show the average accuracy for each task individually (layout: Linear ■ | Radial ■).

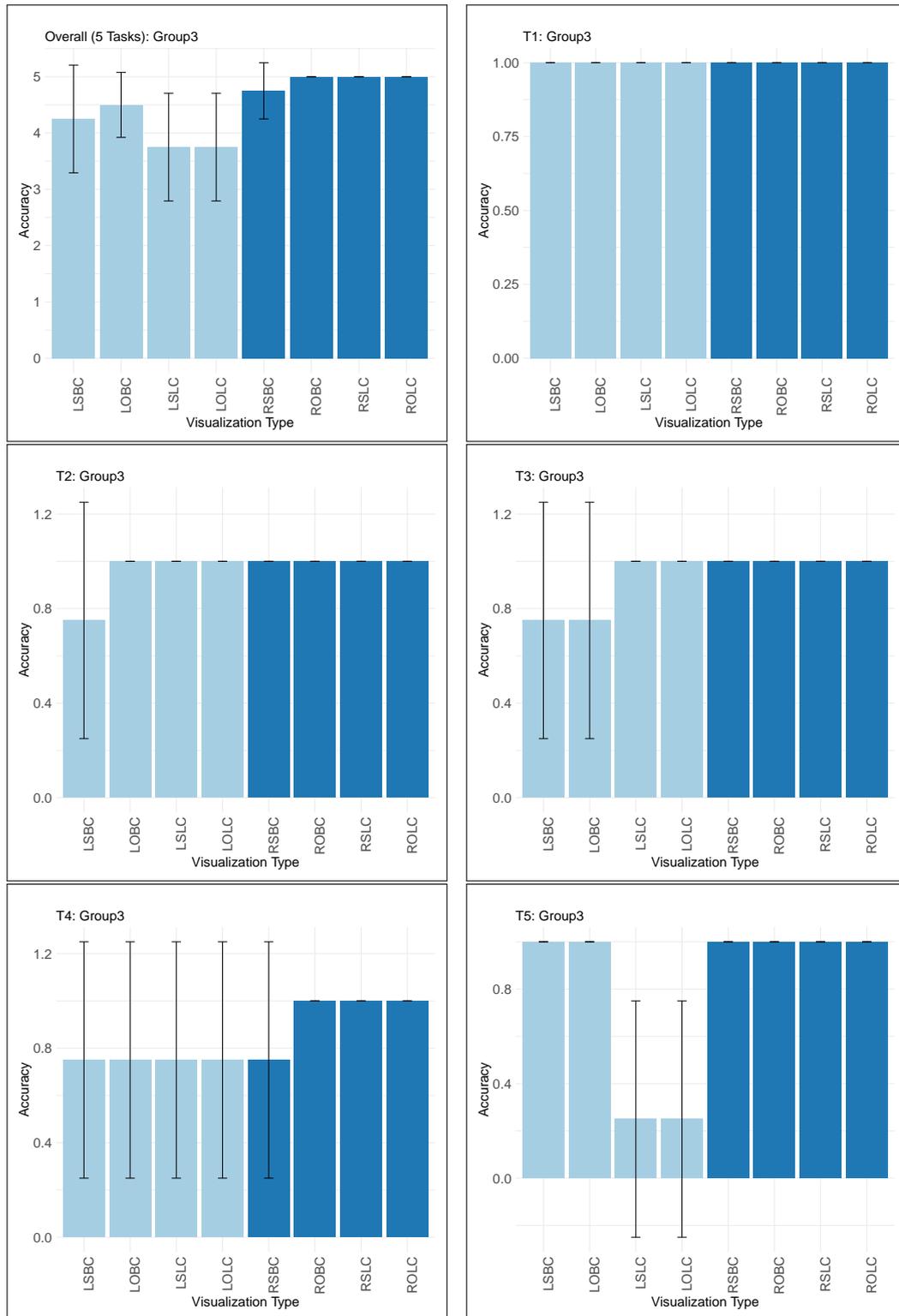


Figure 3: Accuracy of All Tasks for the Eight Visualizations for Group 3 (in the order presented). The top left figure shows the mean and the SD of the overall task accuracy for the eight visualizations. The Y axis represents the average accuracy of the five tasks, while the X axis shows the visualizations based on their presentation order for group 3. The other five subfigures show the average accuracy for each task individually (layout: Linear ■ | Radial ■).

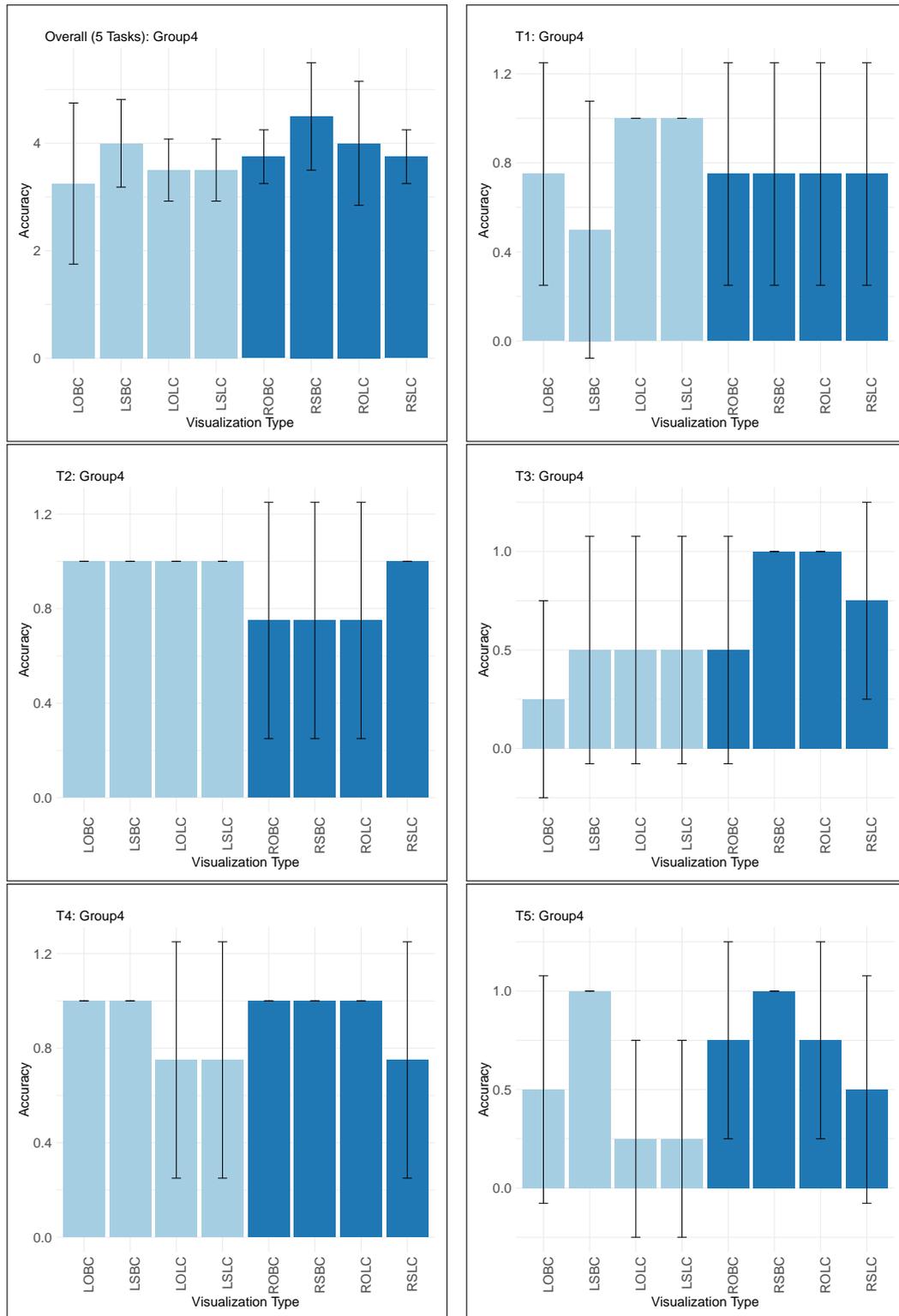


Figure 4: Accuracy of All Tasks for the Eight Visualizations for Group 4 (in the order presented). The top left figure shows the mean and the SD of the overall task accuracy for the eight visualizations. The Y axis represents the average accuracy of the five tasks, while the X axis shows the visualizations based on their presentation order for group 4. The other five subfigures show the average accuracy for each task individually (layout: Linear ■ | Radial ■).

Appendix

Main Evaluation Study

.5 Main Evaluation Study: Participation Forms

This section presents the forms used in the main evaluation study, which are the information sheet, consent form and debriefing sheet.

[Evaluating Different Visualization Designs for Multivariate Personal Health Data]

Information Sheet

Introduction

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information carefully and discuss it with others if you wish. Ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

What is the purpose of the study?

The aim of this study is to identify the effectiveness of different visualization designs of multivariate personal activity data by comparing users' performance when performing a set of tasks on the proposed designs. The study also aims to compare users' performance to their preferences in the visualizations by collecting qualitative data from the questionnaires.

How is the study structured?

Once you have agreed to take part in the study and unless you choose to withdraw at any point, you will be asked to fill in two questionnaires and perform a set of tasks using our software.

In the briefing session, the purpose and workings of this study will be explained to you, and you will be asked to:

- Filling in a demographic questionnaire (*3 minutes*).
- Filling in a pre-experiment questionnaire (*5 minutes*).
- Watching a short tutorial describing the experiment and what you are expected to do (*3 minutes*).

During the experiment, you will be asked to:

- Performing the experiment which consists of 5 tasks' questions for 8 visualization designs for 2 different data (*40 minutes*).

After the experiment, you will be asked to:

- Filling in post-experiment questionnaire to provide feedback about the visualization designs (*5 minutes*).

During the previous sessions, you will be provided with the opportunity to ask questions regarding this study.

In the debriefing session, you will be provided with debriefing sheet that includes information about the study and whom you can contact. You will again be provided with the opportunity to ask questions regarding this study.

Who is organising and funding the research?

The study is organised by a research student from Cardiff University, School of Computer Science and Informatics (Majedah Alrehiely, a PhD Student), and is funded by Cardiff University, School of Computer Science and Informatics.

Where are the briefing and debriefing sessions taking place?

The briefing and debriefing sessions of this study are taking place at Cardiff University, the school of Computer Science and Informatics, Queens Building, Second Floor. A room will be booked for the study.
Address:
Queens Building, Cardiff University, 5 The Parade, Roath, Cardiff CF24 3AA.

Why have I been chosen?

As a volunteer, you have responded to our request for participants to take part via email (Cardiff University mailing list or Yammer).

Do I have to take part?

It is up to you to decide whether or not you want to take part in the study. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

How will the data be collected and stored?

All information that is collected about you during the course of this research will be kept strictly confidential. The information will be collected by (the researcher/ the used software and online questionnaires), and stored securely through the researcher's Cardiff University OneDrive and a backup will be stored on password protected internal storage, for a period of 6 years. The data will be accessible by [the researcher "Majedah Alrehiely", the supervisor "Dr. Parisa Eslambolchilar" and the external supervisor "Dr. Rita Borgo, a senior lecturer in computer science at King's College London"]

We may share the data we collect with researchers at other institutions, but any information that leaves Cardiff University will have your personal details removed. In any sort of output we might publish, we will not include information that will make it possible for other people to know your name or identify you in any way.

What data will be collected?

Type	Data
Consent Forms	The consent forms need to be signed by the participant before taking a part in this study. You will be assigned a number and your personal data will be anonymized.

Demographic Data	Gender, age Range, highest degree of education, employment status, education field, what type of personal physical activity data you track (e.g. step count, active minutes etc.) and the used tracking tool (if applies) and familiarity with different charts.
Answers to the questionnaire (pre-and post-experiment design rating and feedback)	Multiple choices and open-ended questions
Performance data during the experiment	Mouse movements and clicks, selected answers to tasks' questions, accomplishment time for each task and the experiment time.

How can I request access to my data?

By emailing the researcher, Majedah Alrehiely at AlrehielyMM@cariff.ac.uk

What if I want to end my participation in this study?

If you want to stop your participation for any reason, you may do so by informing the researcher of your withdrawal without the need for giving any reason.

What will happen to the results of the research study?

Where appropriate, the results of this study will be confidentially stored and analysed for the research purpose. You will not be identified in any report or publication. The results of this study will help to understand users' performance with different visualization designs for personal physical activity data. We will inform you of the results of the study if you wish to have the information.

Who has reviewed the study?

This study has been reviewed and approved by the Ethics Committee of the School of Computer Science and Informatics, Cardiff University. Ethics application number: [COMSC/Ethics/2018/028]. Should you have any concerns about this study, please contact comsc-ethics@cardiff.ac.uk

Contact for Further Information

We welcome the opportunity to answer any question you may have about any aspect of this study or your participation in it.

Majedah Alrehiely, PhD student in the School of Computer Science and Informatics, Cardiff University

AlrehielyMM@Cardiff.ac.uk

[Evaluating Different Visualization Designs for Multivariate Personal Health Data]

Consent Form

Name of participant/identifier: _____

Staff/ Student: _____ College: _____

Email: _____

Please read the Information Sheet and then read the following statements carefully and then add your signature. If you have any questions, please ask the person who gave you this form. You are under no pressure to give your consent and you are free to withdraw from this study at any time. By signing the form you are agreeing to the following:

- I understand that I am to take part in a briefing session, filling in questionnaires, performing a software experiment that includes visual tasks and a debriefing session.
- I confirm that I have all of the prerequisites for participation as outlined in the Information Sheet, and consent that, where appropriate, any personal equipment I provide can be used for the purposes of this study.
- I confirm that I have read and understand the Information Sheet and have had the opportunity to ask questions about it.
- I understand that upon completion of the study, I will receive *£10 Amazon Voucher*. In the case that I withdraw from the study, I understand that I will receive *Nothing*.
- I understand that participation in this study is entirely voluntary and that I can withdraw from the study at any time without giving a reason, and without loss of credit / payment (as applicable).
- I understand that I am free to ask any questions at any time and that I am free to withdraw or discuss my concerns with the lead researcher *Majedah Alrehiely*.
- I also understand that at the end of the study I will be provided with additional information and feedback about the purpose of the study.
- I understand that the information provided by me will be held confidentially, such that only the researchers can trace this information back to me individually. The information will be retained for up to *6 years* when it will be deleted/destroyed. I understand that I can ask for the information I provide to be deleted/destroyed at any time.
- I agree to data generated from my participation can be shared with other researchers.

I, _____ (NAME) consent to participate in the study conducted by *Majedah Alrehiely*, School of Computer Science, Cardiff University.

Signed:

Date:

[Evaluating Different Visualization Designs for Multivariate Personal Health Data]

Debriefing Sheet

Thanks

Many thanks for taking part in this study. We hope it was interesting. Your participation was an important contribution to research on *evaluating different visualization designs for multivariate personal health data*. Please feel free to ask the researcher any questions you have about what happened. Please note that some of the information contained on this form is a repeat of what might be found on the Information Sheet, which you should already have and can keep.

What is the purpose of the study?

The aim of this study is to identify the effectiveness of different visualization designs of multivariate personal activity data by comparing users' performance when performing a set of tasks on the proposed designs. The study also aims to compare users' performance to their preferences in the visualizations by collecting qualitative data from the questionnaires.

How was the data collected and how will it be stored?

All information that is collected about you during the course of this research will be kept strictly confidential. The information will be collected by (the researcher/ the used software and online questionnaires), and stored securely through the researcher's Cardiff University OneDrive and a backup will be stored on password protected storage, for a period of 6 years. The data will be accessible by [the researcher "Majedah Alrehiely", the supervisor "Dr. Parisa Eslambolchilar" and the external supervisor "Dr. Rita Borgo, a senior lecturer in computer science at King's College London"]

We may share the data we collect with researchers at other institutions, but any information that leaves Cardiff University will have your personal details removed. In any sort of output we might publish, we will not include information that will make it possible for other people to know your name or identify you in any way.

What will happen to the results of the research study?

Where appropriate, the results of this study will be confidentially stored and analysed for the research purpose. You will not be identified in any report or publication. The results of this study will help to understand users' performance with different visualization designs for personal physical activity data. We will inform you of the results of the study if you wish to have the information.

What do I do if I am unhappy with the way I was treated or with something that happened to me?

In the first instance, you should contact the leader of the Research Project:

Majedah Alrehiely, AlrehielyMM@Cardiff.ac.uk

If you are still unhappy, please should contact the relevant Ethics Committee:

Cardiff School of Computer Science & Informatics Ethics Committee
Email: comsc-ethics@cardiff.ac.uk

Who has reviewed the study?

This study has been reviewed and approved by the Ethics Committee of the School of Computer Science and Informatics, Cardiff University. Ethics application number: [COMSC/Ethics/2018/028].

Contact for Further Information

We welcome the opportunity to answer any question you may have about any aspect of this study or your participation in it.

Majedah Alrehiely, AlrehielyMM@Cardiff.ac.uk

.6 Main Evaluation Study: Demographic and Pre-Experiment Questionnaire



Main_Study_Demographic_and_Pre-Experiment_Questionnaire_1

Page 1: Research Purpose

Evaluating Different Visualization Designs for Multivariate Personal Health Data: A Research Study

What is the purpose of the study?

The aim of this study is to identify the effectiveness of different visualization designs of multivariate personal activity data by comparing users' performance when performing a set of tasks on the proposed designs. The study also aims to compare users' performance to their preferences in the visualizations by collecting qualitative data from the questionnaires. Filling this questionnaire is the first part of the study, it will be followed by Performing an experiment which consists of 5 tasks' questions for 8 visualization designs and then followed by another questionnaire to provide your feedback about the visualizations.

Page 2: Participant Information

1. Please enter your participation ID * *Required*

2. Please select your age group:

- 18–24
- 25–34
- 35–44
- 45–54
- 55–64
- > 64
- Prefer not to say

3. Please select your gender

- Male
- Female
- Prefer Not to Say
- Other

3.a. If you selected Other, please specify:

4. What is the highest degree or level of education you have completed?

- Secondary school or equivalent
- College degree or equivalent
- Master's degree or equivalent
- Doctoral degree (PhD)
- Prefer not to say

Other

4.a. If you selected Other, please specify:

4.b. What is your educational field?

- Art, Humanities and Social Science (e.g. Business, Languages, Law and Music)
- Biomedical and Life Science (e.g. Bioscience, Psychology and Healthcare Science)
- Physical Science and Engineering (e.g. Computer Science, Chemistry and Engineering)
- Other

4.b.i. If you selected Other, please specify:

4.c. Employment status:

- Student
- Researcher staff
- Academic staff
- Administrative staff
- Prefer not to say
- Other

4.c.i. If you selected Other, please specify:

4.d. Please state your familiarity with the following charts

Please don't select more than 1 answer(s) per row.

	Not at all familiar	Slightly familiar	Somewhat familiar	Moderately familiar	Extremely familiar
Line Chart	<input type="checkbox"/>				

Bar Chart	<input type="checkbox"/>				
Radial Chart	<input type="checkbox"/>				

4.e. How familiar you are with reading statistical data and charts (e.g. line and bar chart)?

- Not at all familiar
- Slightly familiar
- Somewhat familiar
- Moderately familiar
- Extremely familiar
- prefer not to say

5. Are you tracking, or have you ever tracked, your health or physical activity data?

- Yes
- No
- Prefer Not to Say

5.a. If (Yes), what type of data do (or did) you track?

- Step count
- Walking/Running/Cycling distance
- Flight climbed
- Active/Burned calories
- Heart rate
- Sleep
- Other

5.a.i. If you selected Other, please specify:

5.a.ii. If (Yes), what type of data do (or did) you track?

- Fitbit Tracker
- Apple Watch
- Garmin Watch/Tracker

- Samsung Watch
- Android Watch
- Jawbone Up
- iPhone apps
- Android apps
- Other

5.a.ii.a. If you selected Other, please specify:

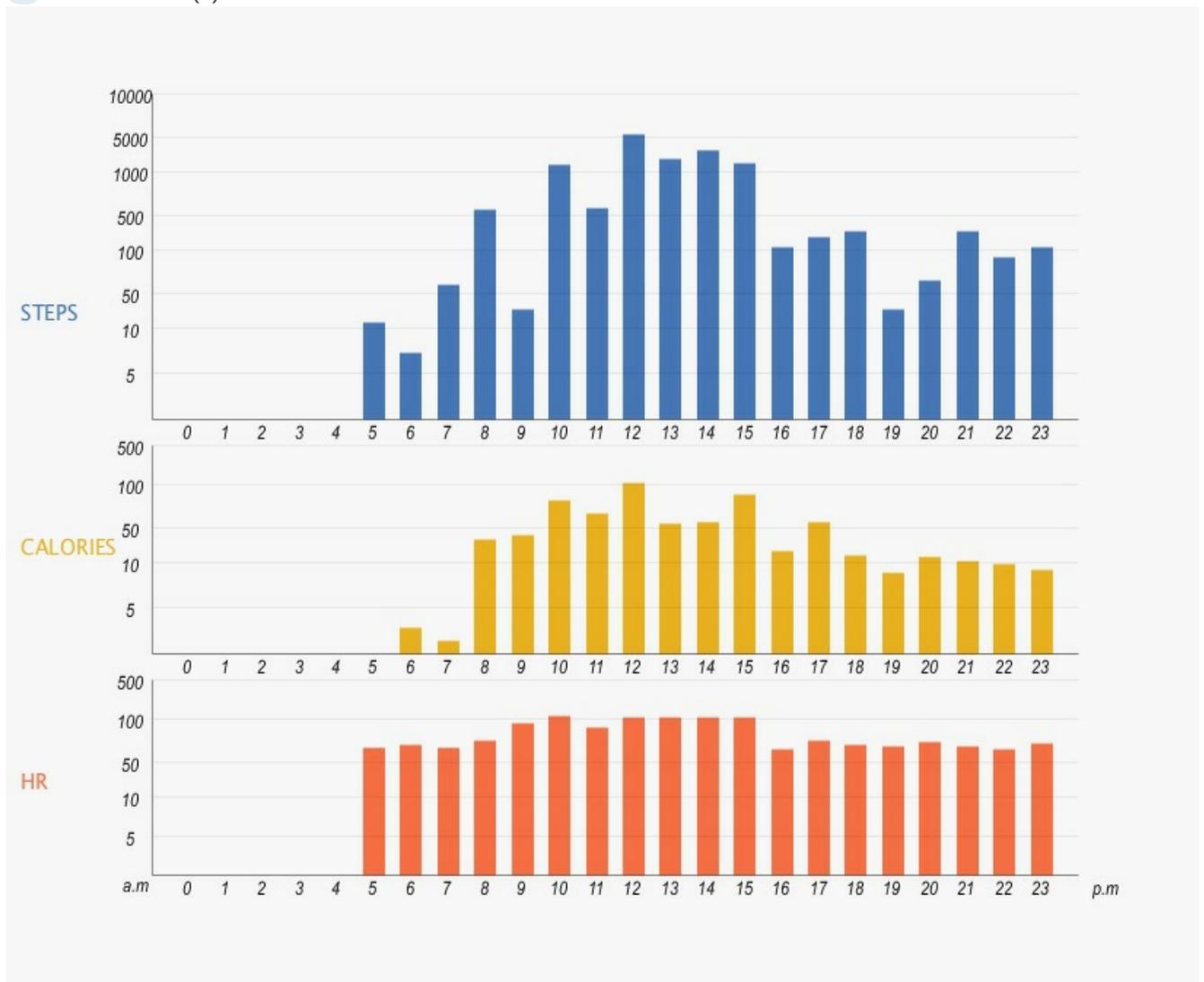
Page 3: Visualization Designs

Visualization Designs

This section displays eight different visualisation designs for three different variables of physical activity data: step count, active calories and heart rate (HR). These questions will ask you to rate each design in terms of **readability** (easy to read and understand) and **attractiveness**.

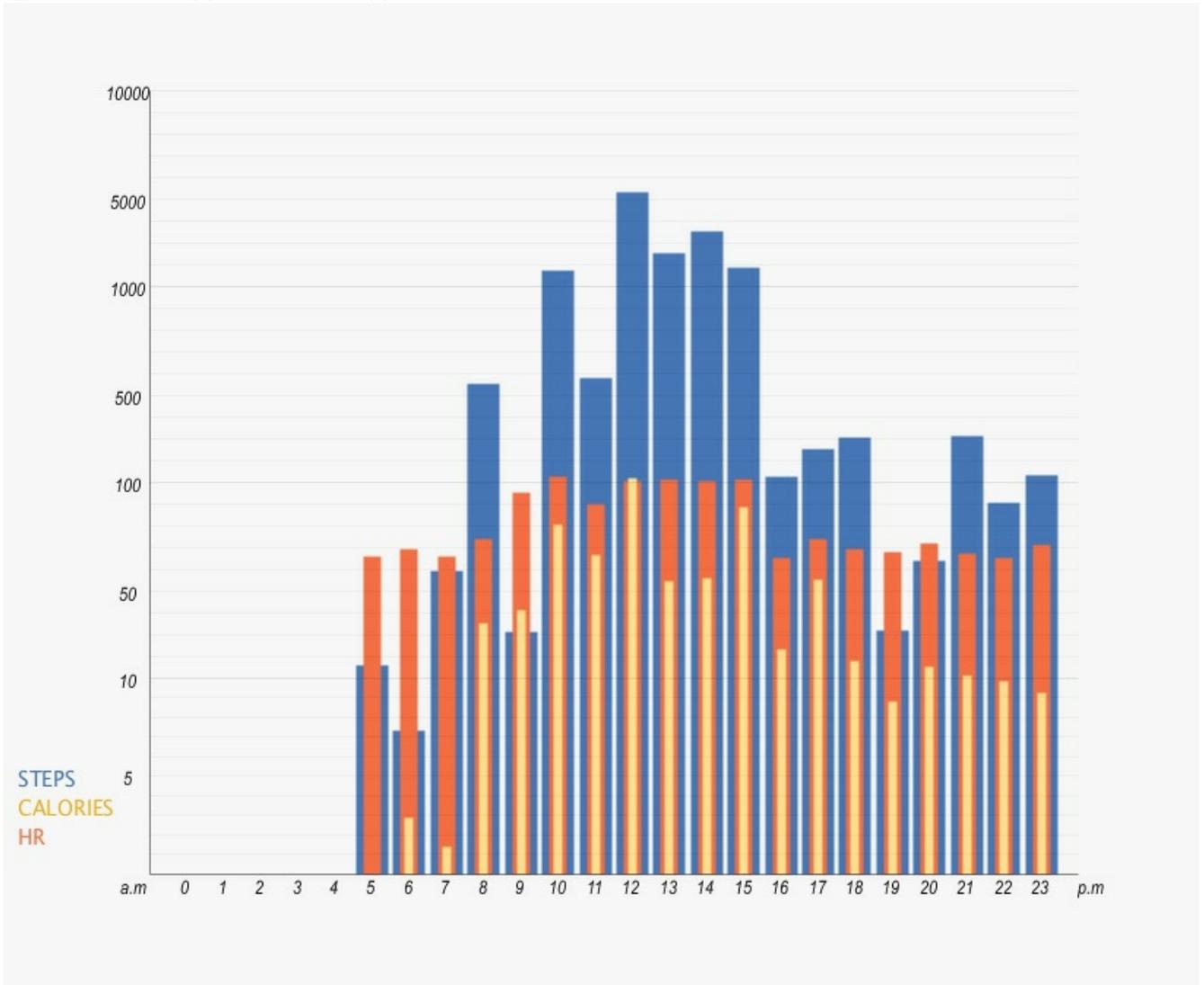
Rate the Following Visualization Designs in Term of its Readability and its Attractiveness

6. Visualization (1): Linear Stcked Bar Chart



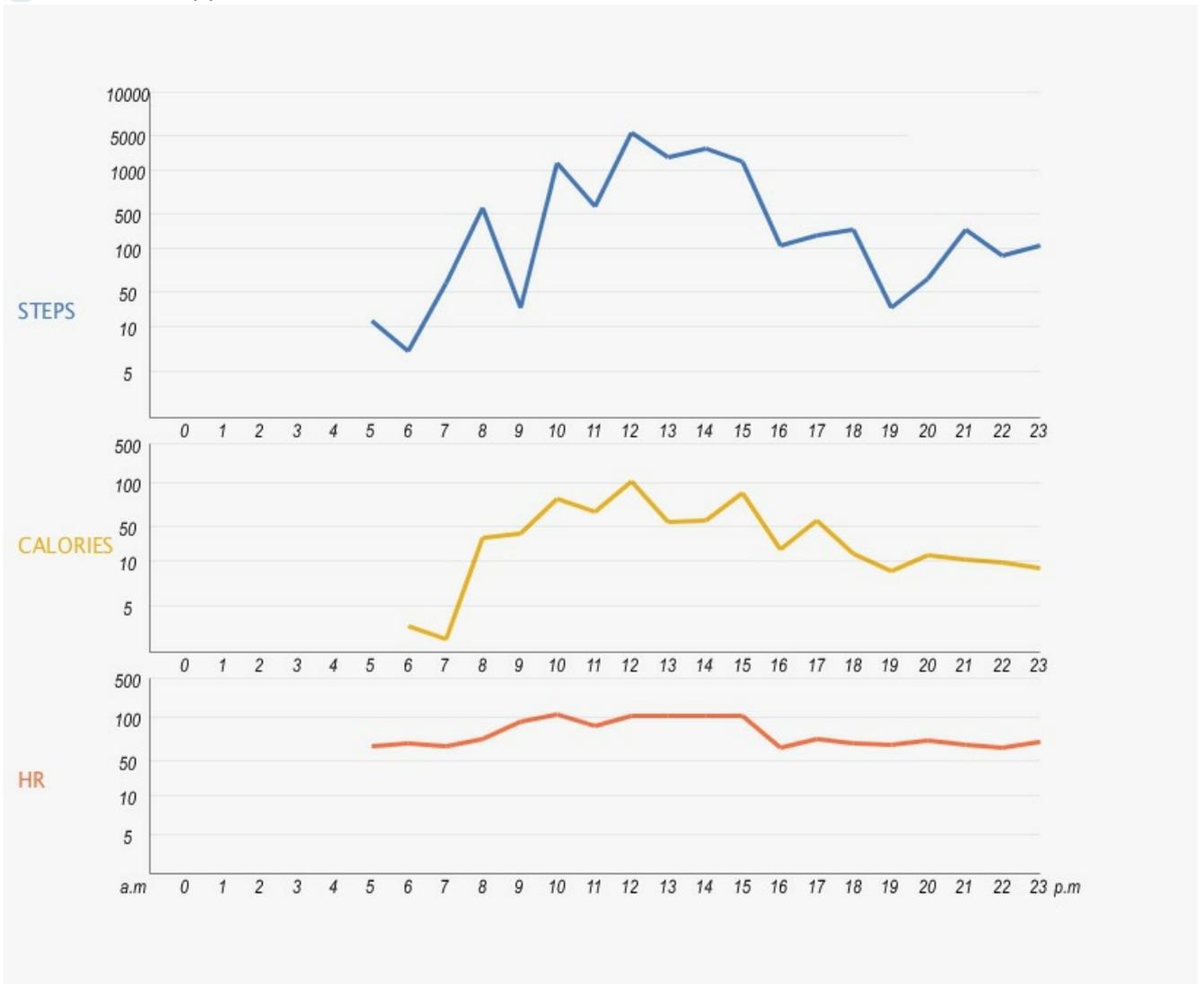
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

7. Visualization (2): Linear Overlapped Bar Chart



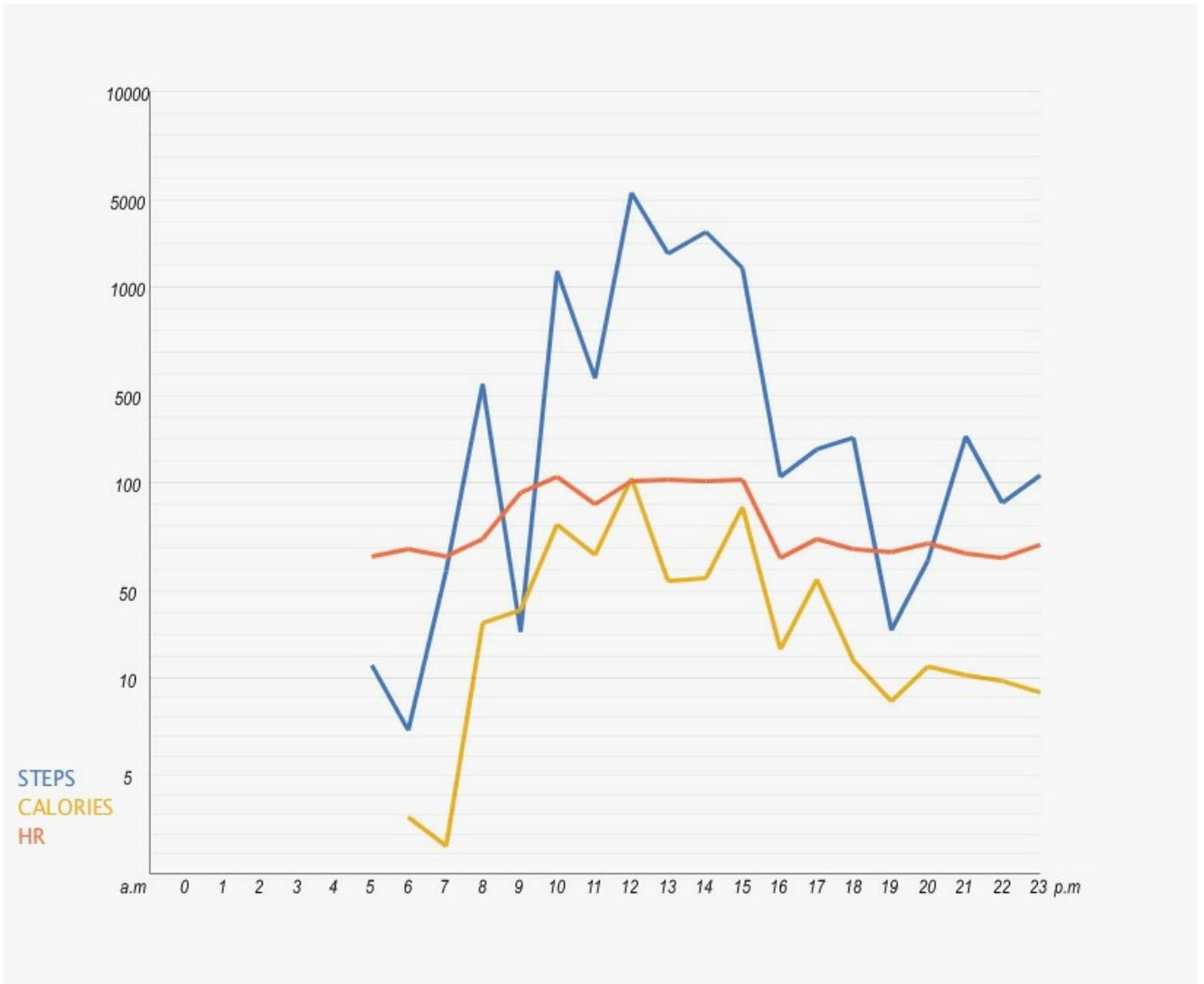
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

8. Visualization (3): Linear Stacked Line Chart



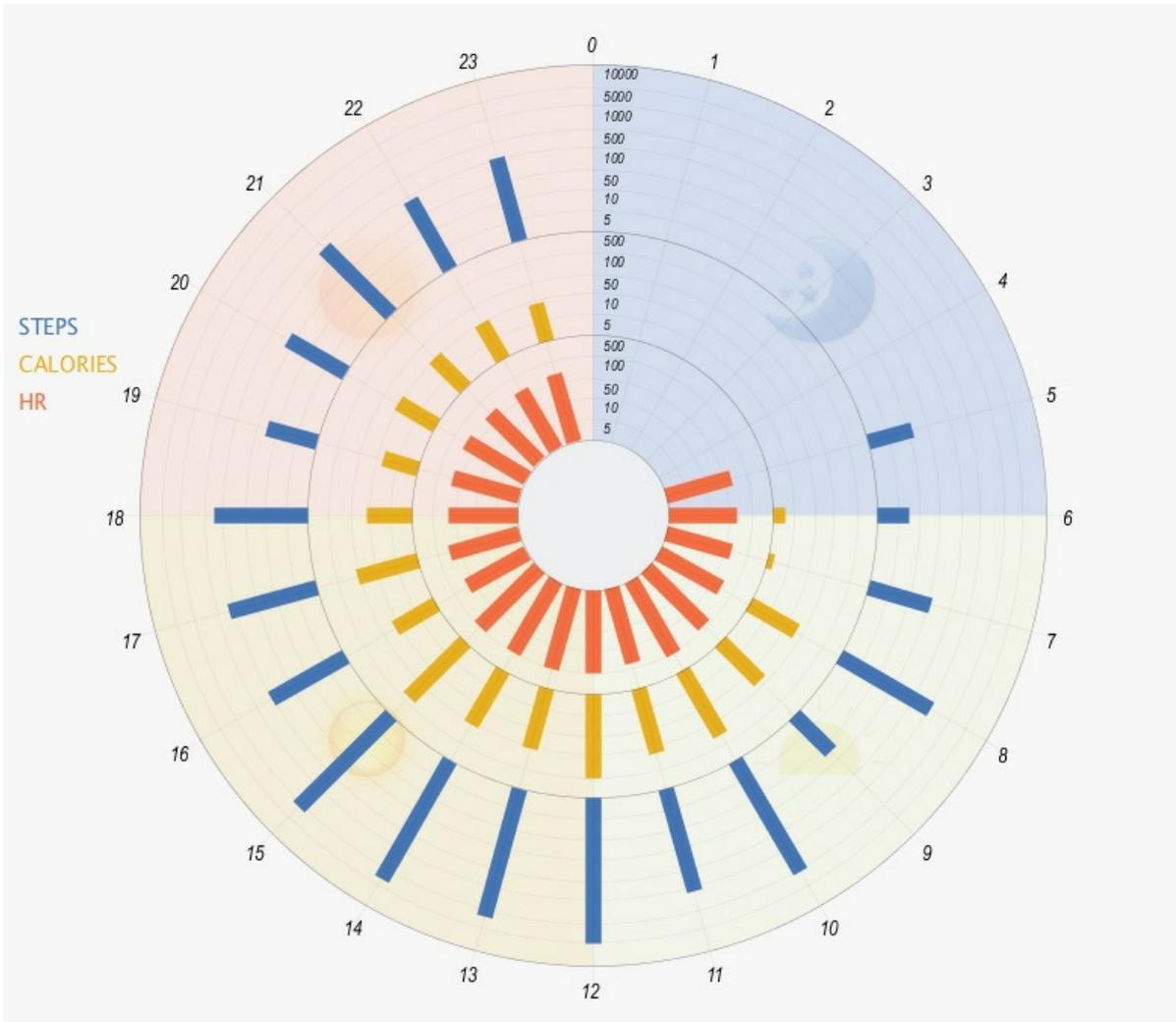
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

9. Visualization (4): Linear Overlapped Line Chart



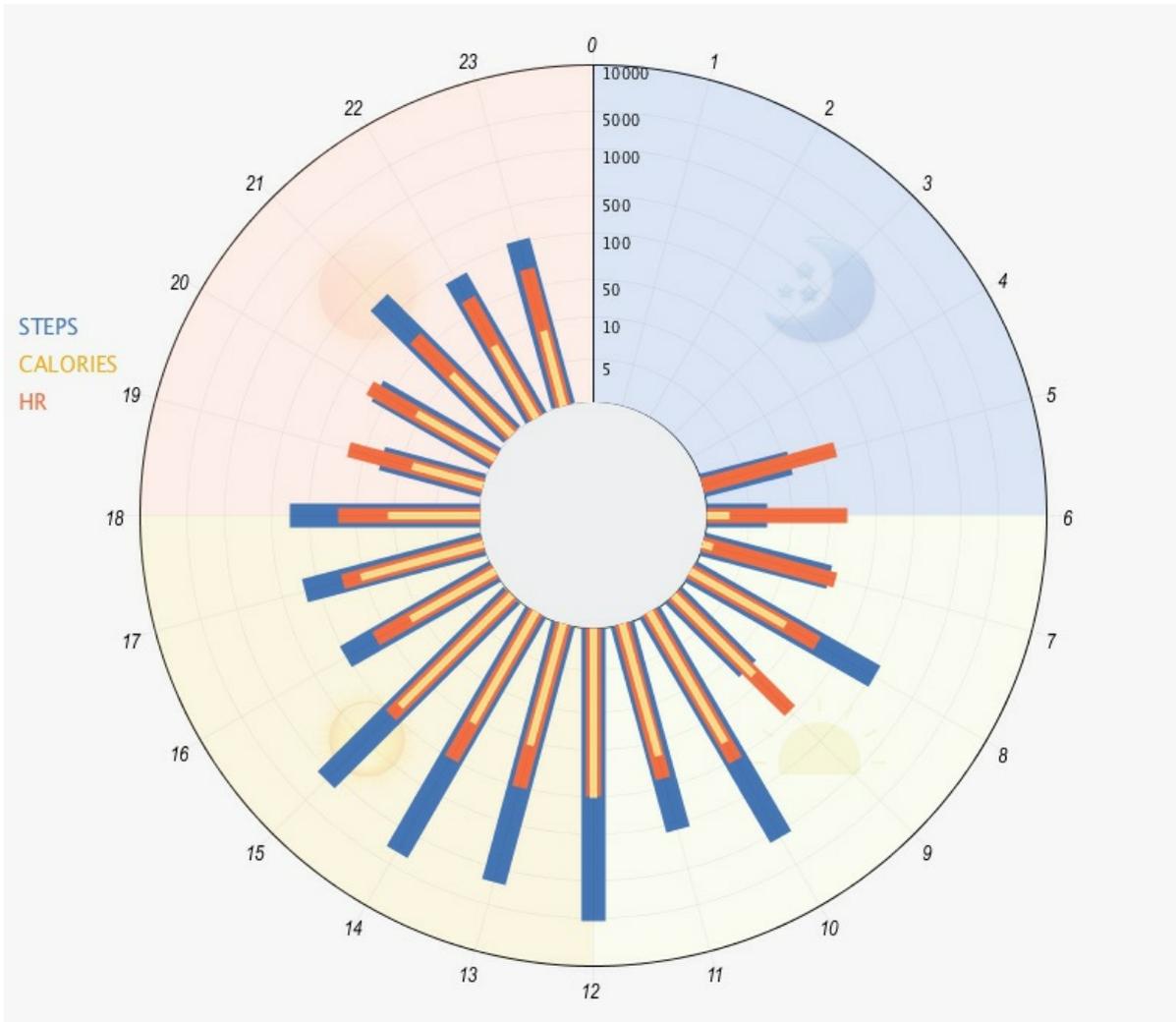
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractevness	<input type="checkbox"/>				

10. Visualization (5): Radial Stacked Bar Chart



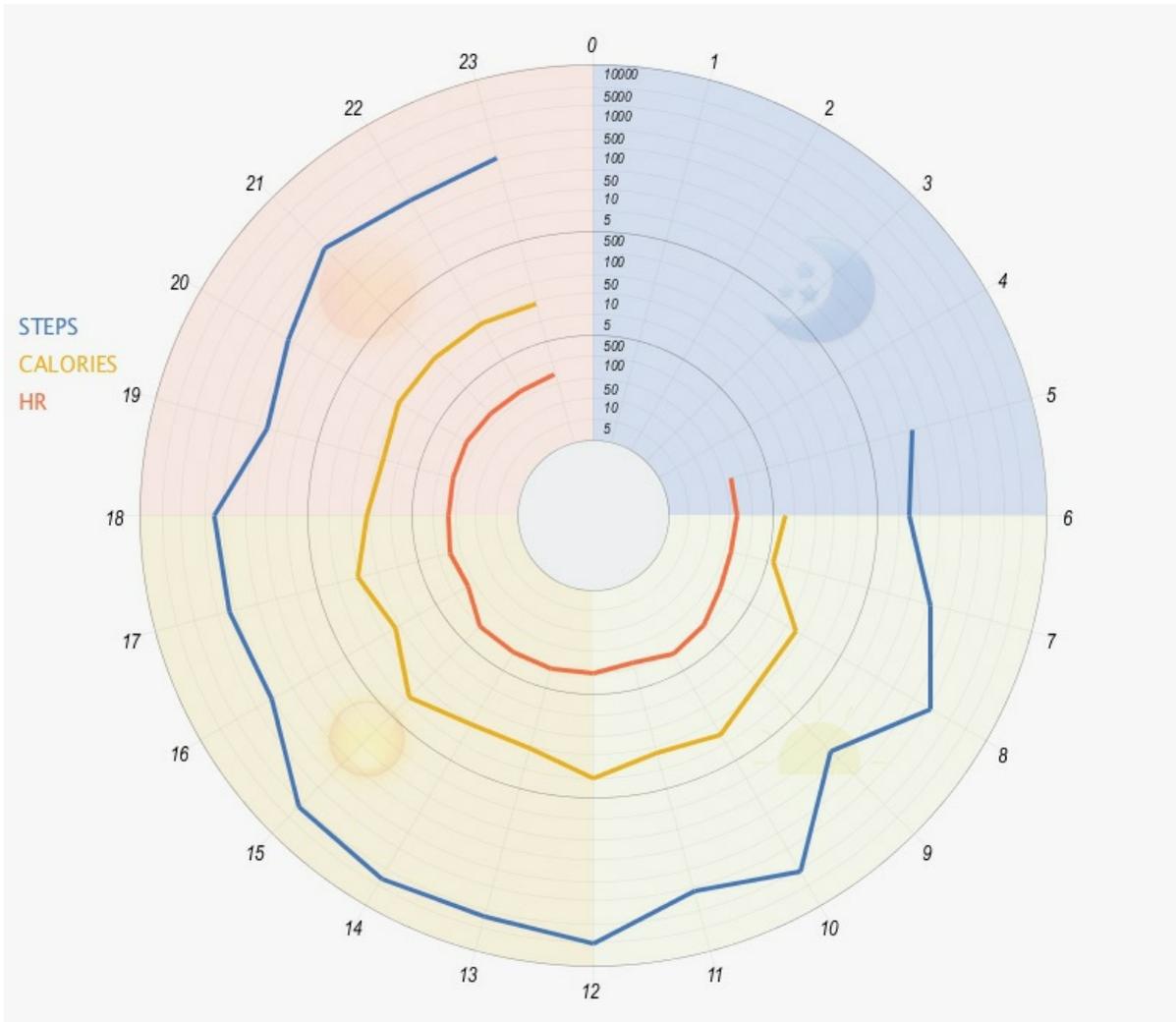
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

11. Visualization (6): Radial Overlapped Bar Chart



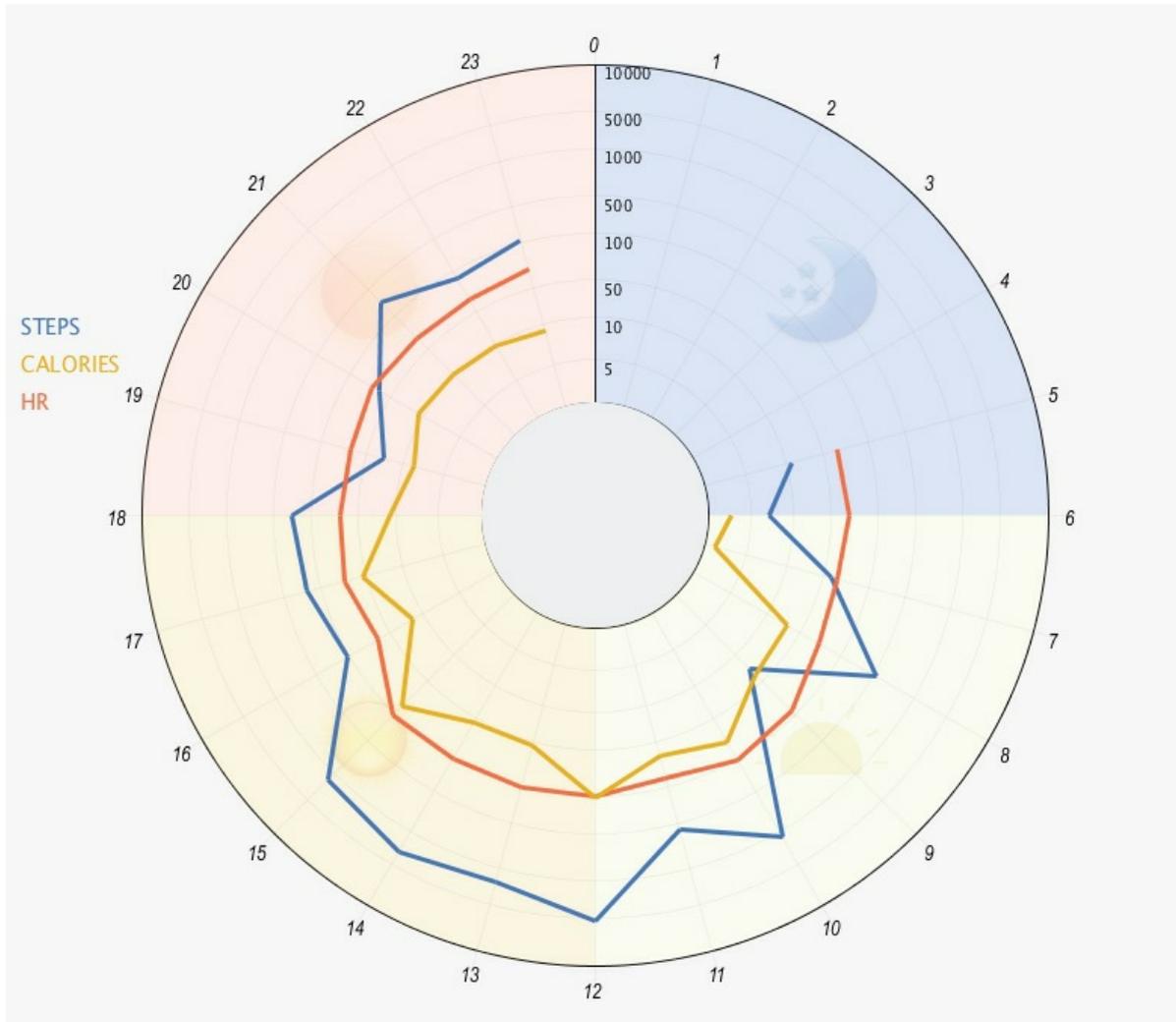
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

12. Visualization (7): Radial Stacked Line Chart



	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

13. Visualization (8): Radial Overlapped Line Chart



	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

14. Which of the previous visualization designs would you prefer to visualize your daily physical activity data?

- Visualization (1): Linear Stacked Bar Chart
- Visualization (2): Linear Overlapped Bar Chart
- Visualization (3): Linear Stacked Line Chart
- Visualization (4): Linear Overlapped Line Chart
- Visualization (5): Radial Stacked Bar Chart

- Visualization (6): Radial Overlapped Bar Chart
- Visualization (7): Radial Stacked Line Chart
- Visualization (8): Radial Overlapped Line Chart
- I don't Prefer any of the visualizations

15. Please justify your choices and clarify why did you not like other visualizations

.7 Main Evaluation Study: Post-Experiment Questionnaire



Main_Study_Post-Experiment_Questionnaire_1

Page 1: Visualization Designs

Participant Information

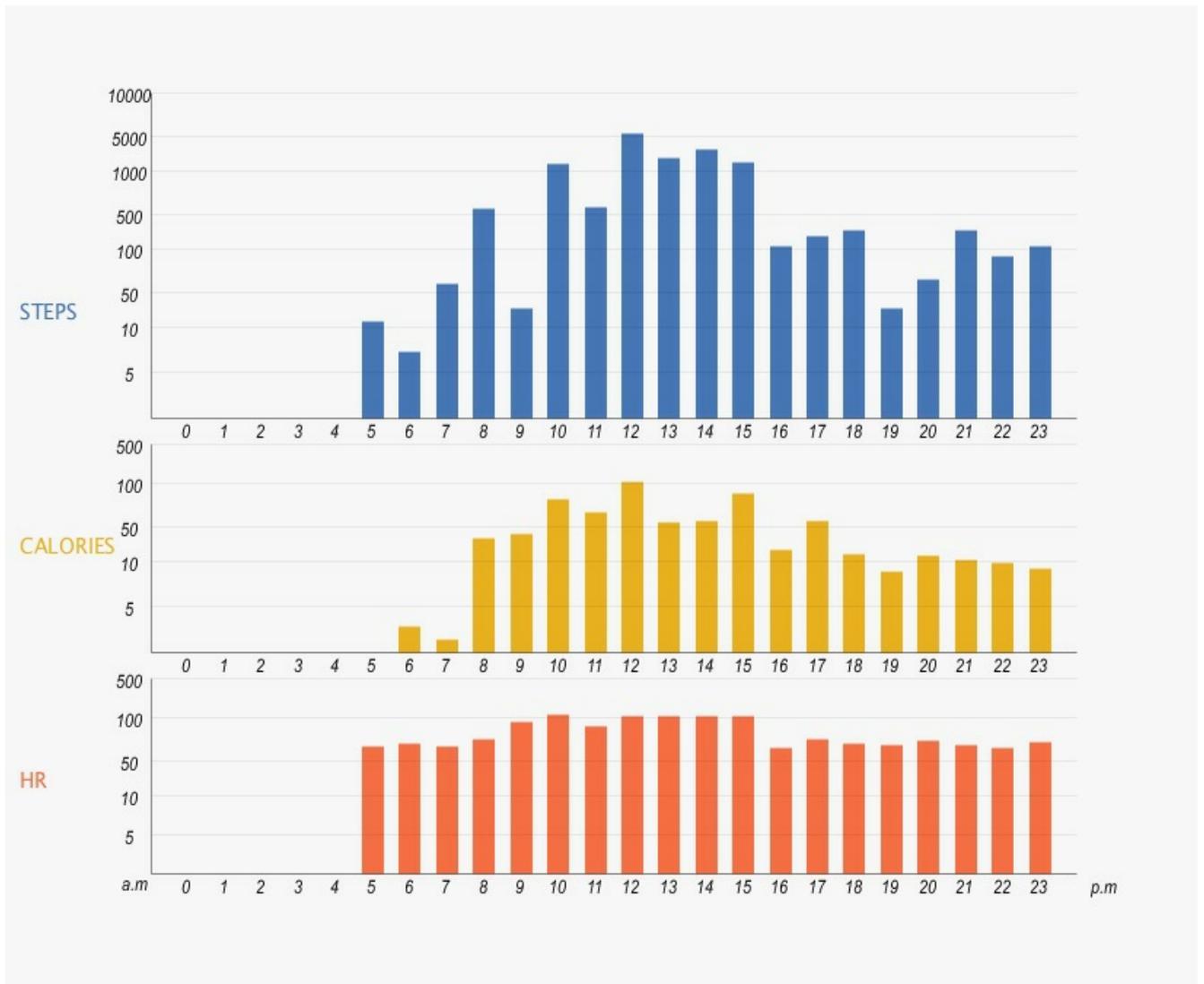
Please enter your participation ID * *Required*

Visualization Designs

This section displays eight different visualisation designs for three different variables of physical activity data: step count, active calories and heart rate (HR). These questions will ask you to rate each design in terms of readability (easy to read and understand) and attractiveness.

Rate the Following Visualization Designs in Term of its Readability and its Attractiveness

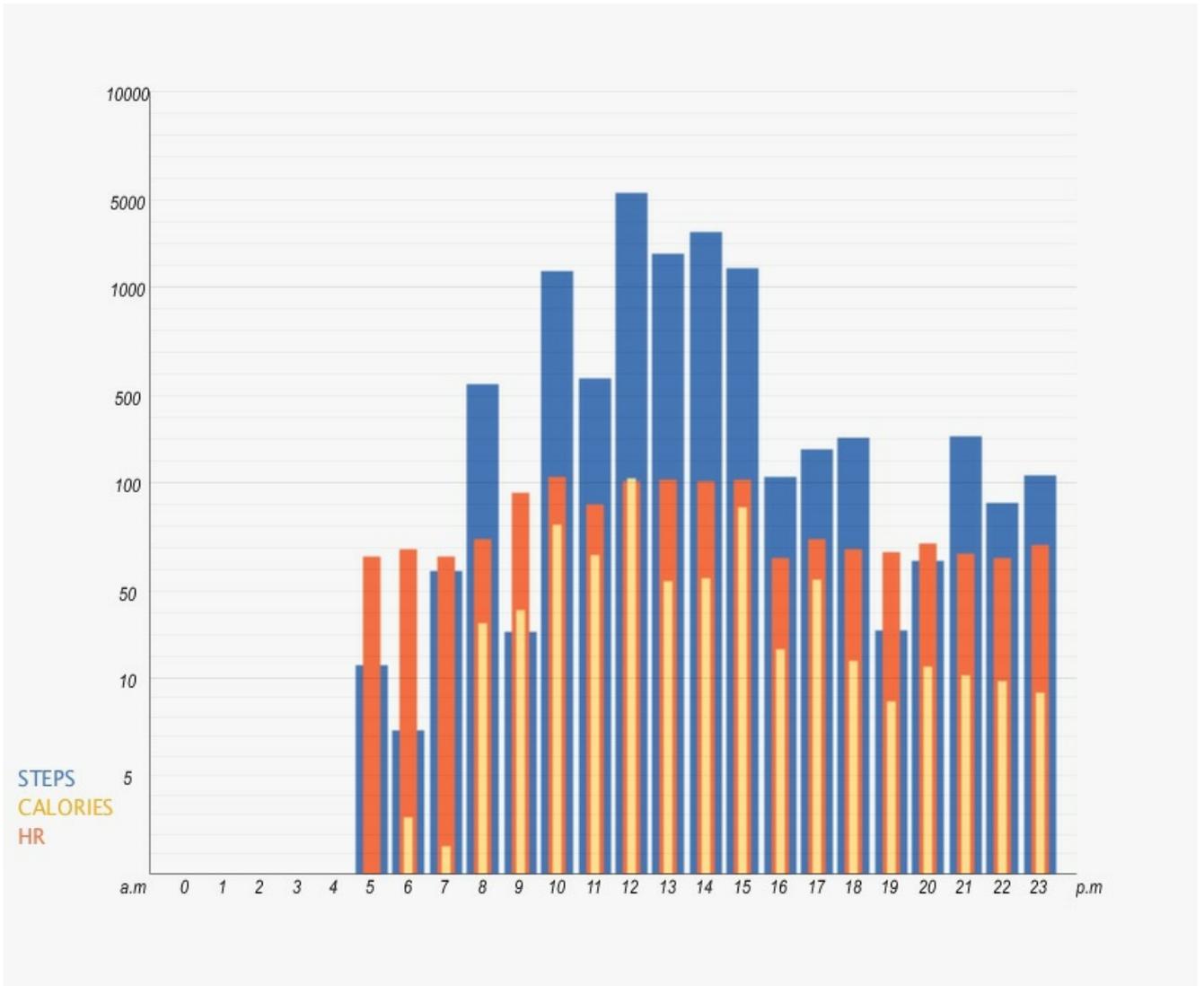
Visualization (1): Linear Stacked Bar Chart



Please don't select more than 1 answer(s) per row.

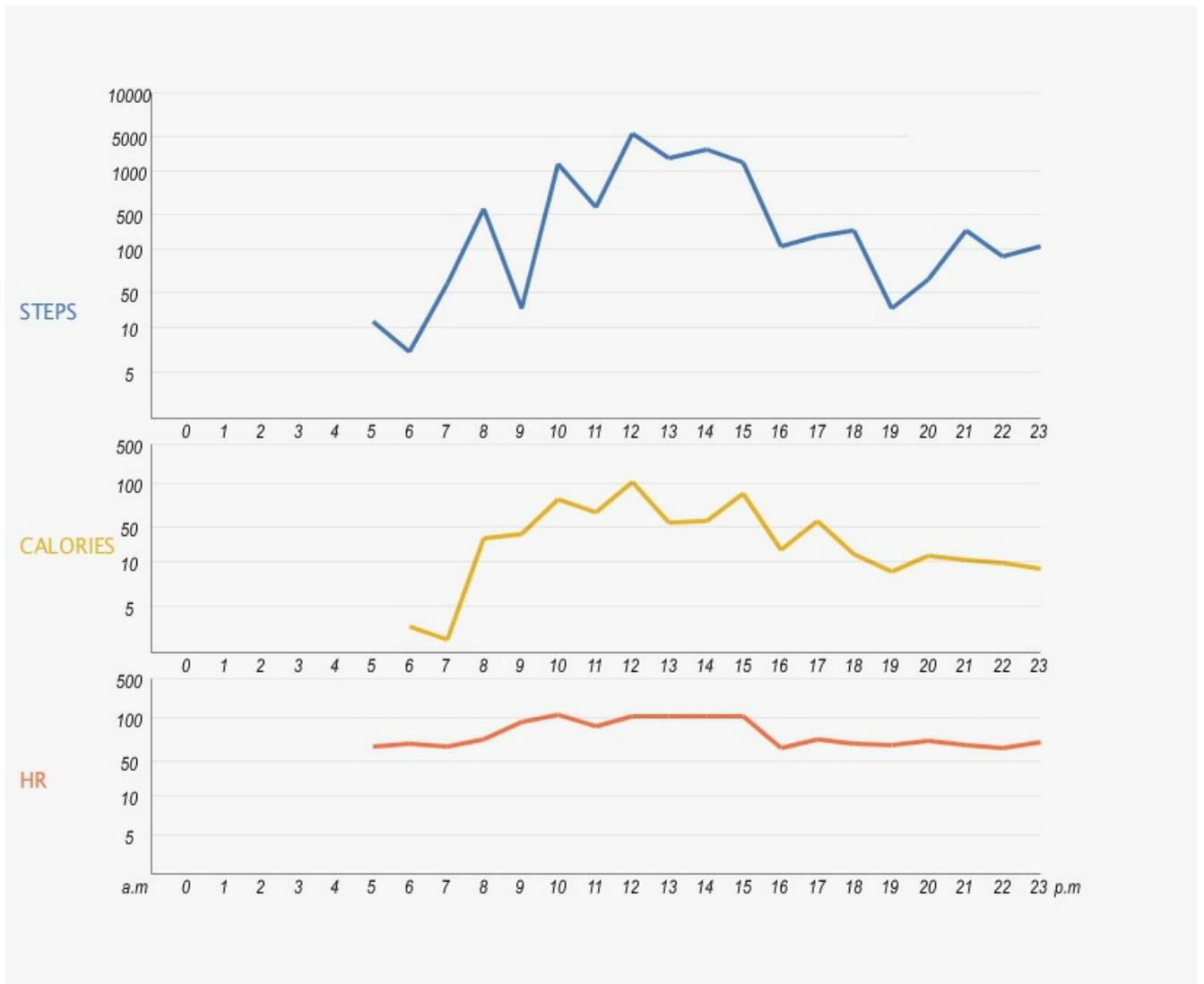
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (2): Linear Overlapped Bar Chart



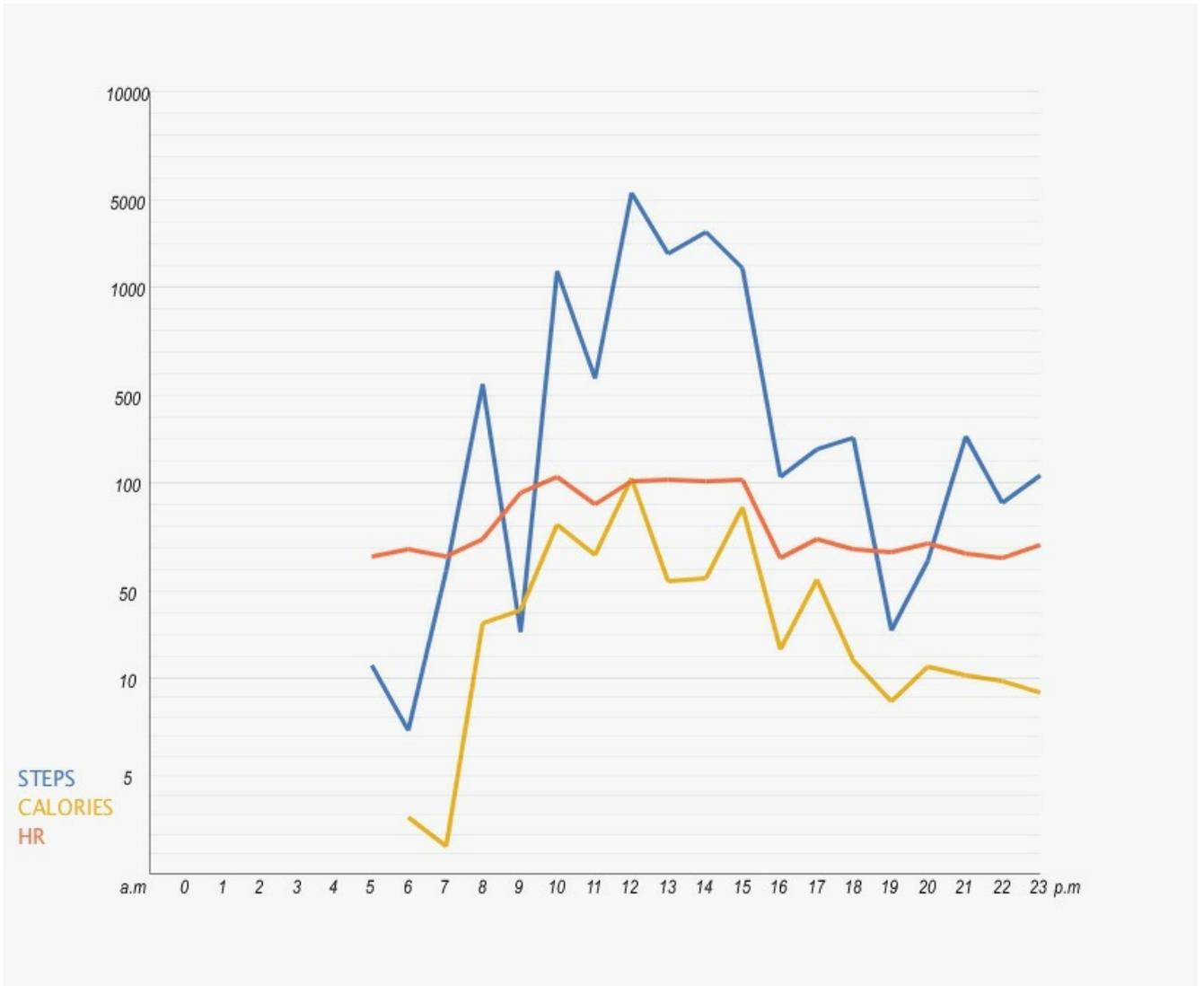
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (3): Linear Stacked Line Chart



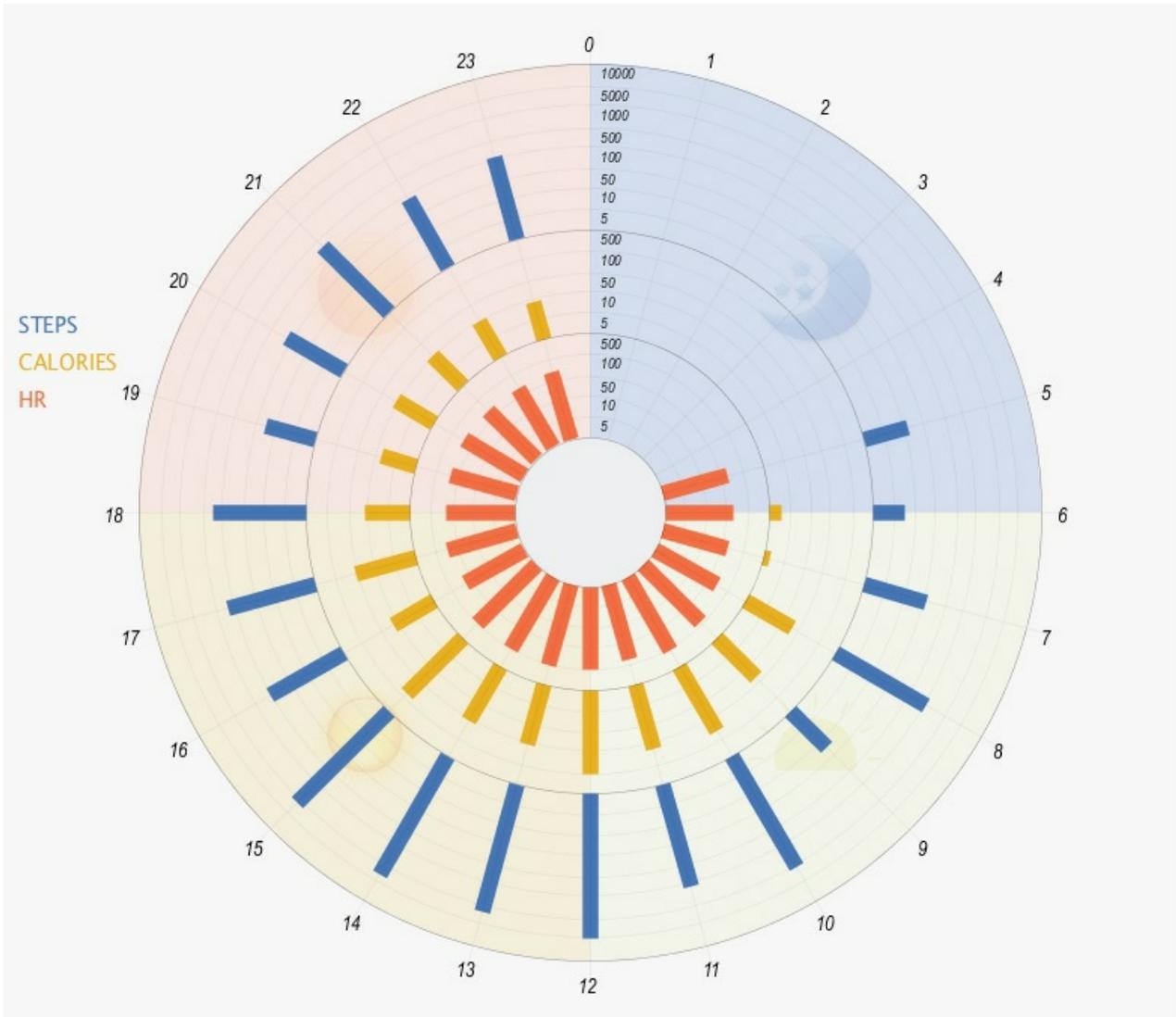
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (4): Linear Overlapped Line Chart



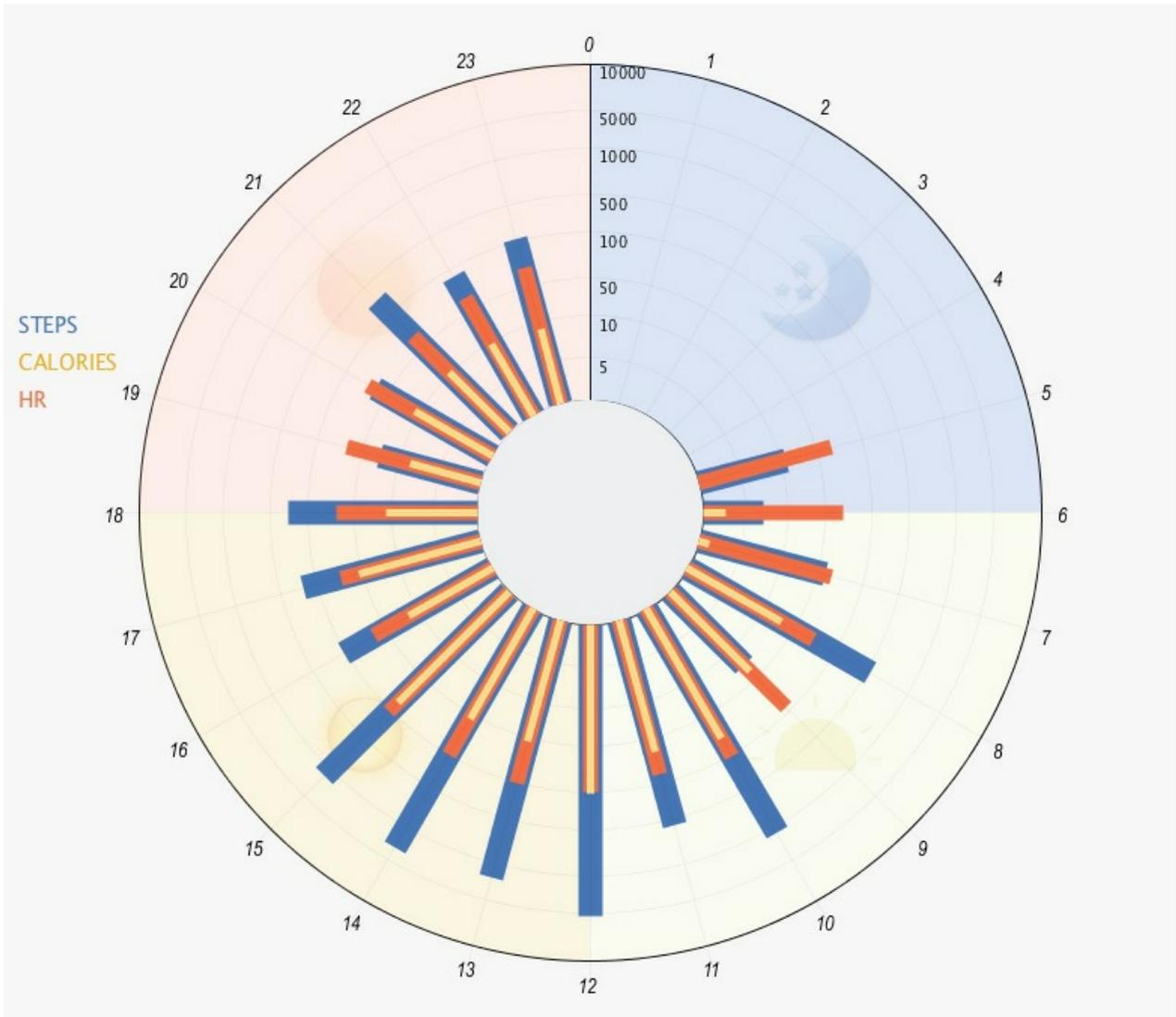
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractevness	<input type="checkbox"/>				

Visualization (5): Radial Stacked Bar Chart



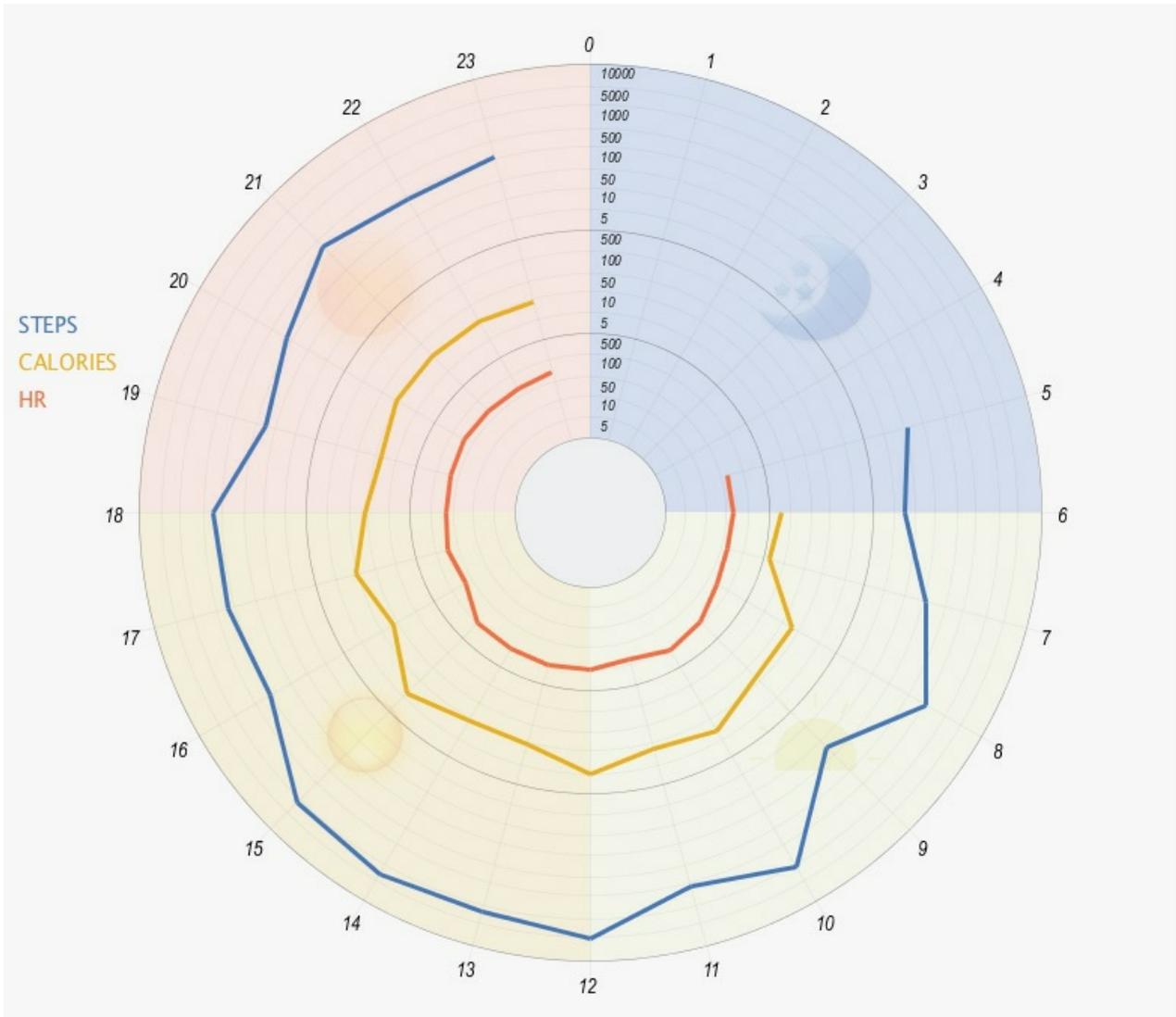
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (6): Radial Overlapped Bar Chart



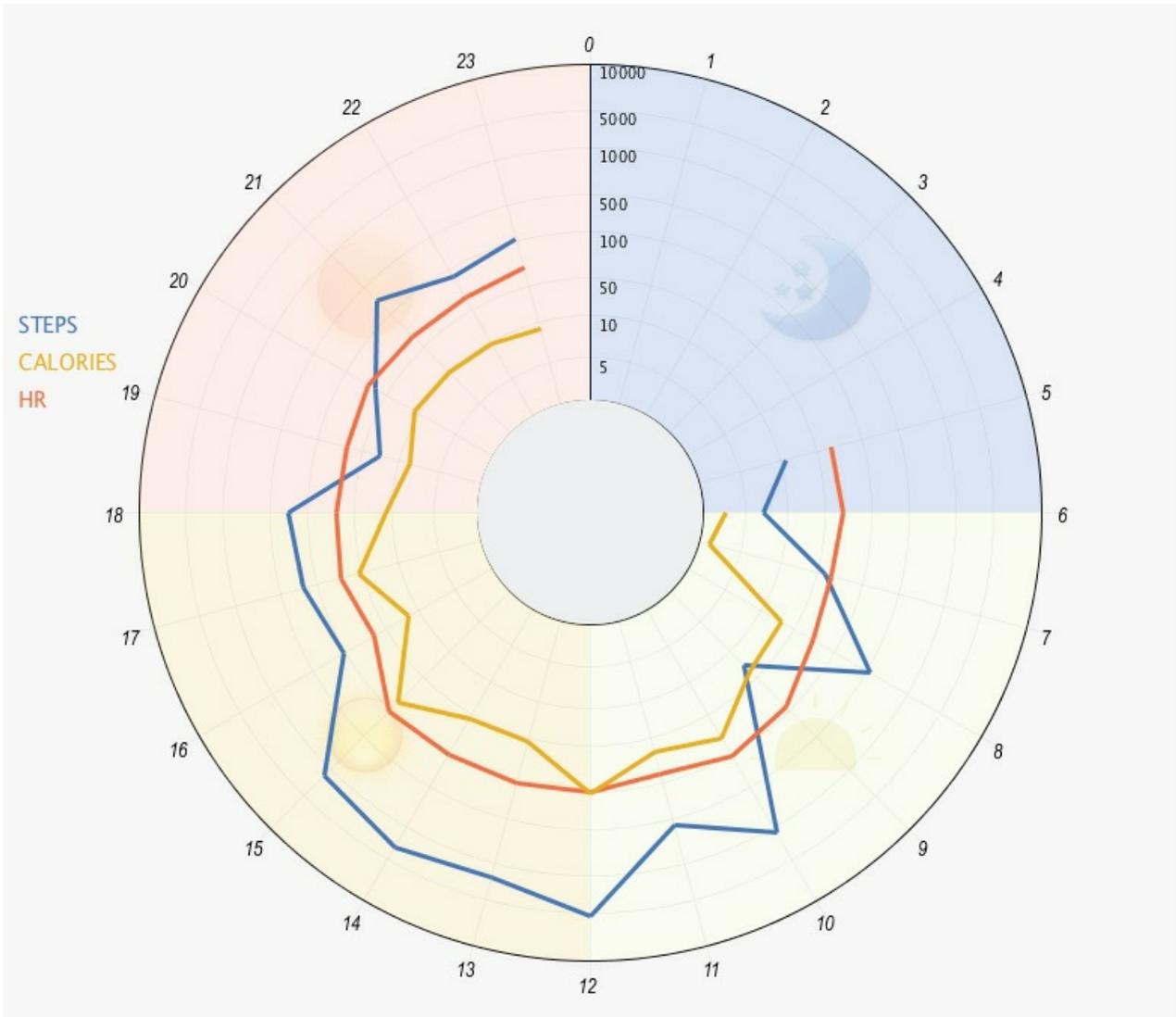
	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (7): Radial Stacked Line Chart



	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Visualization (8): Radial Overlapped Line Chart



	Very Poor	Poor	Acceptable	Good	Very Good
Readability	<input type="checkbox"/>				
Attractiveness	<input type="checkbox"/>				

Which of the Previous Visualization Designs Would you Prefer to Visualize your Daily Physical Activity Data?

- Visualization (1): Linear Stacked Bar Chart
- Visualization (2): Linear Overlapped Bar Chart
- Visualization (3): Linear Stacked Line Chart
- Visualization (4): Linear Overlapped Line Chart
- Visualization (5): Radial Stacked Bar Chart

- Visualization (6): Radial Overlapped Bar Chart
- Visualization (7): Radial Stacked Line Chart
- Visualization (8): Radial Overlapped Line Chart
- I don't Prefer any of the visualizations

Please justify your choices and clarify why did you not like other visualizations

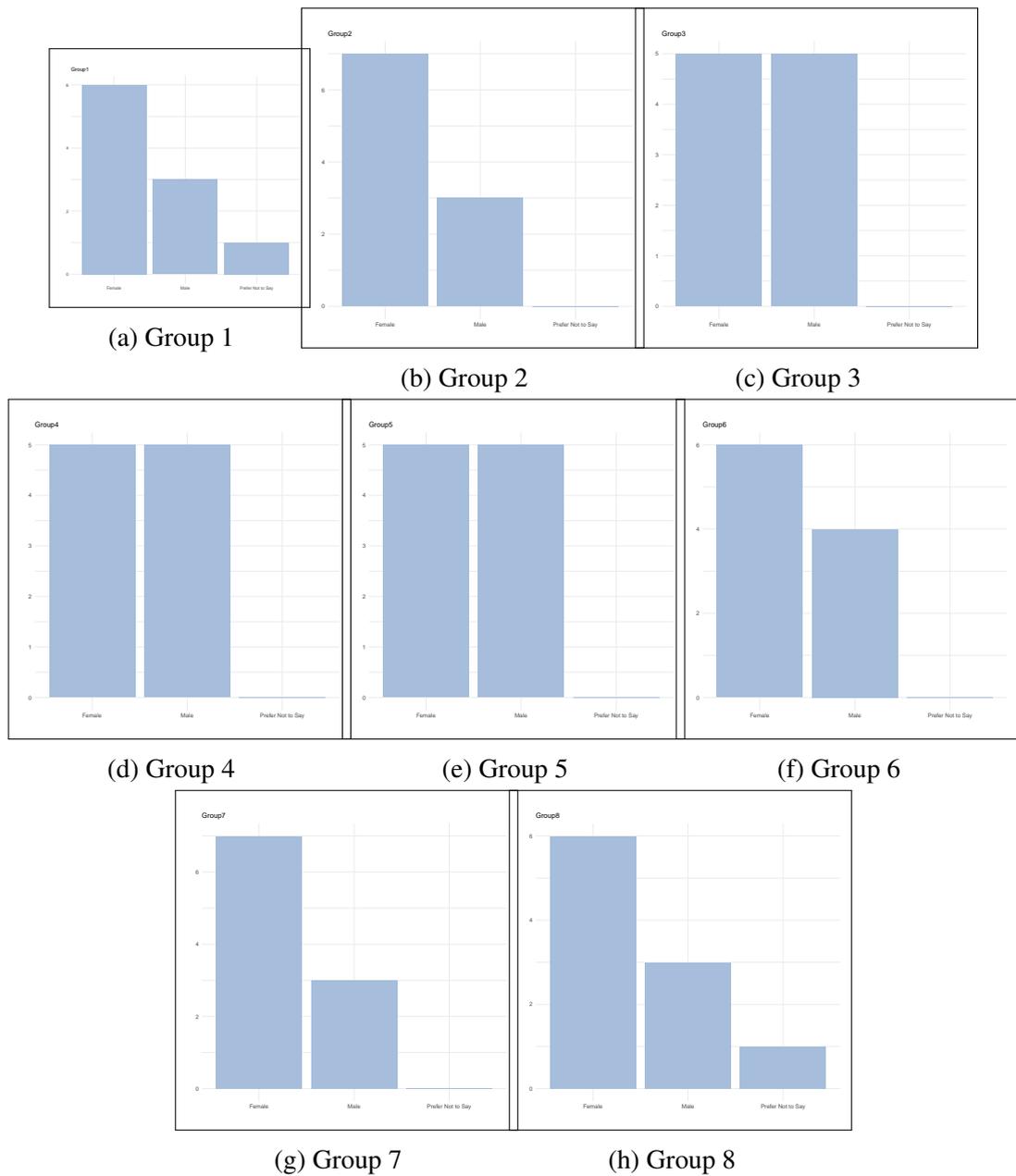


Figure 5: Gender Distribution in Each Group

.8 Groups Demographic Information

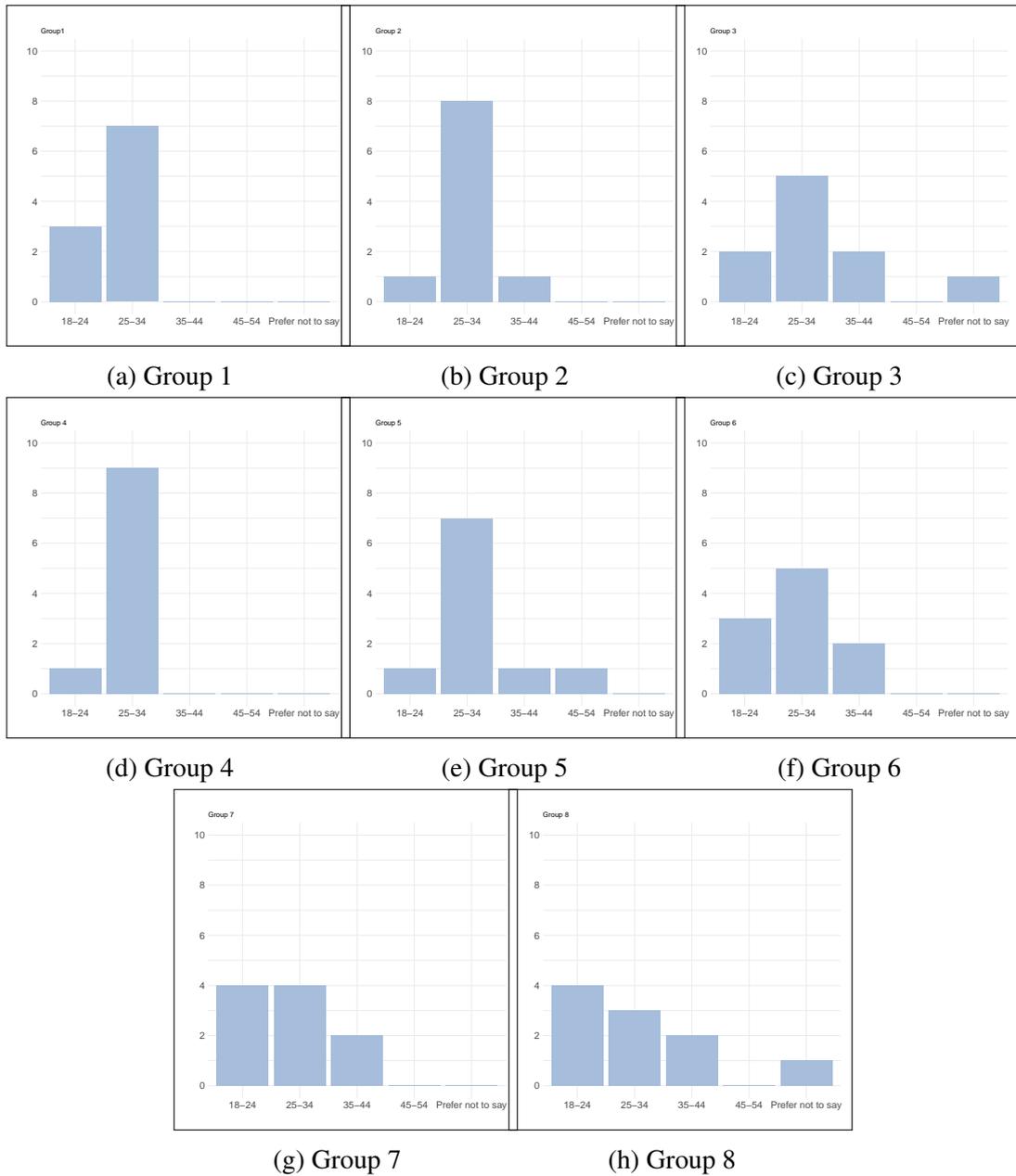
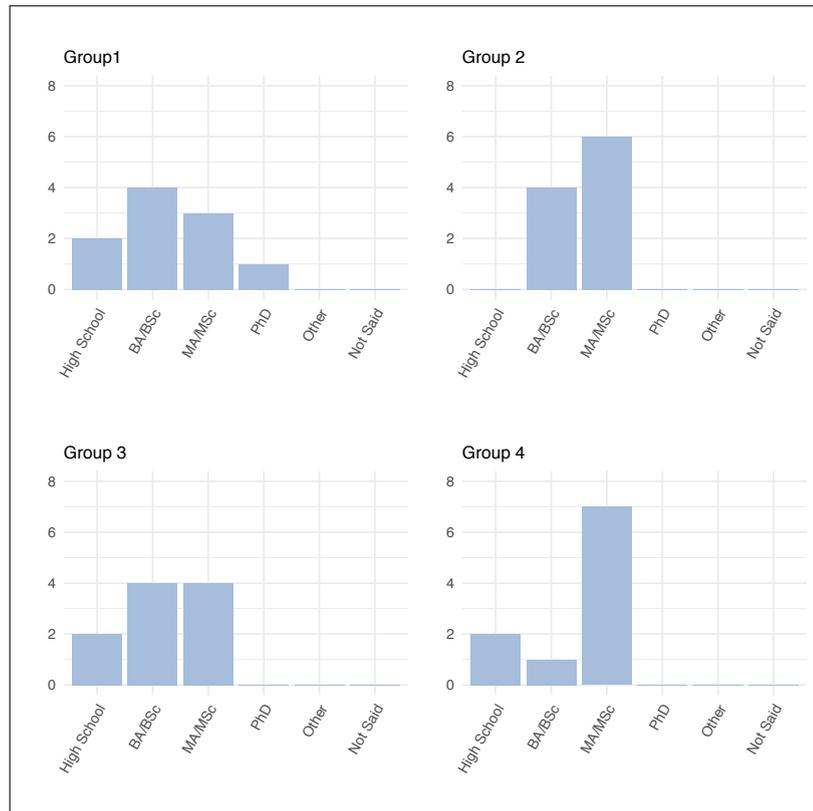


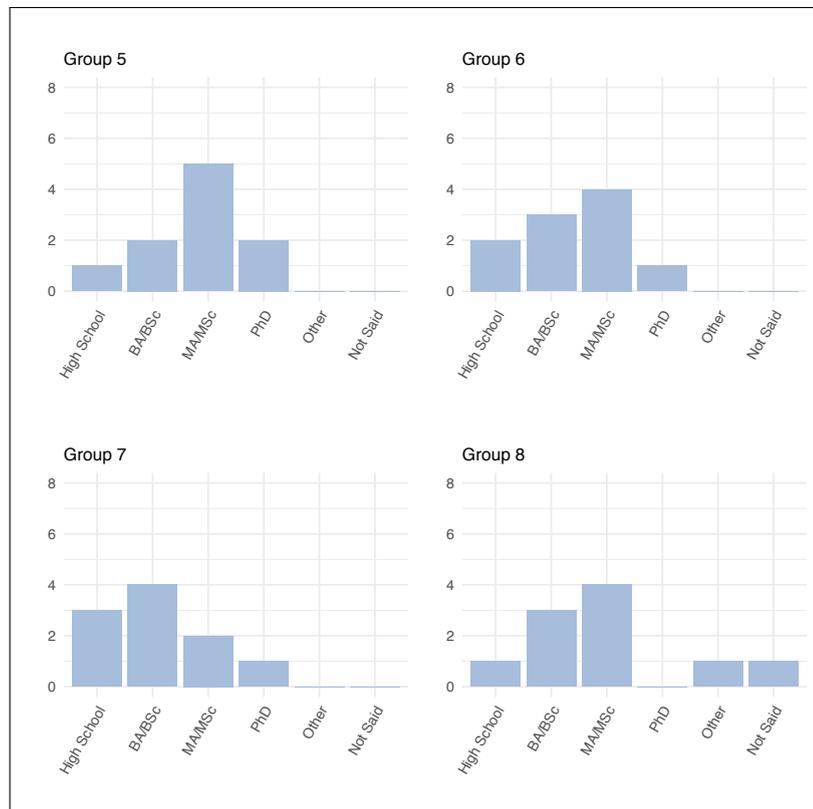
Figure 6: Age Distribution in Each Group. In the eight figures, the x-axis presents the age-ranges, and the Y-axis shows the number of participants in each age-range group. .

.9 Groups and Participants' Behaviour Analysis

In this section, we are describing and analyzing the participants' behaviour during the experiment with a focus on groups' performance and the participants' performance within each group. The aim is from this analysis is to identify how the participants in each group respond to the first viewed visualizations and to see if there is a learning effect in the experiment.

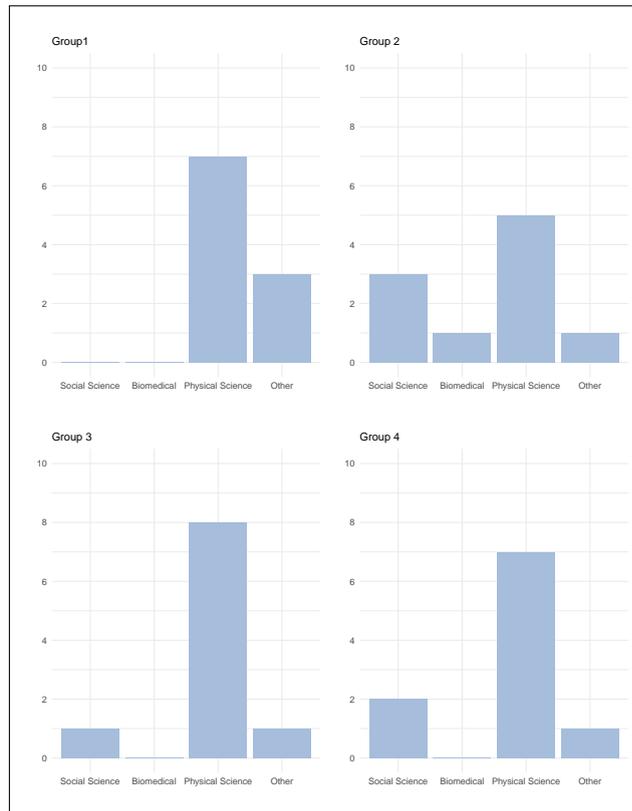


(a)

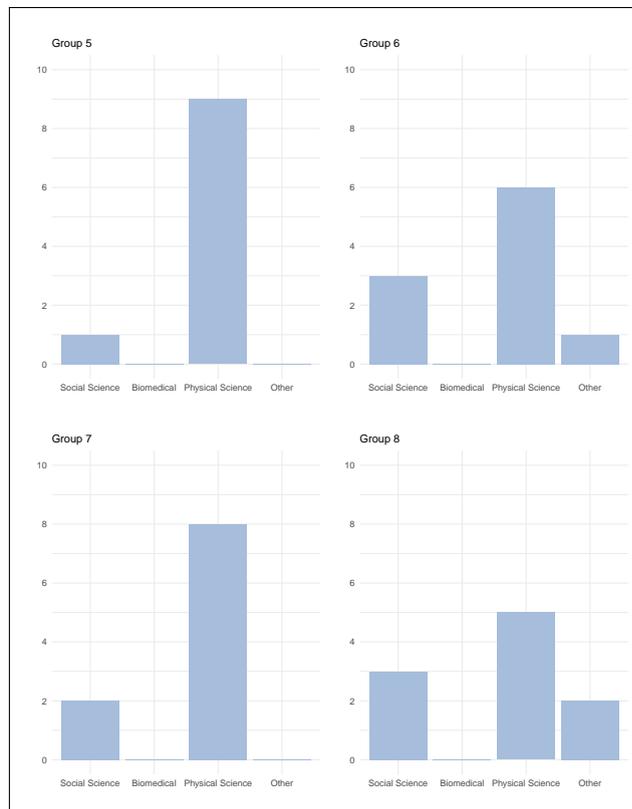


(b)

Figure 7: The Participants' Responses of their Education Degree (Distribution Based on Groups).

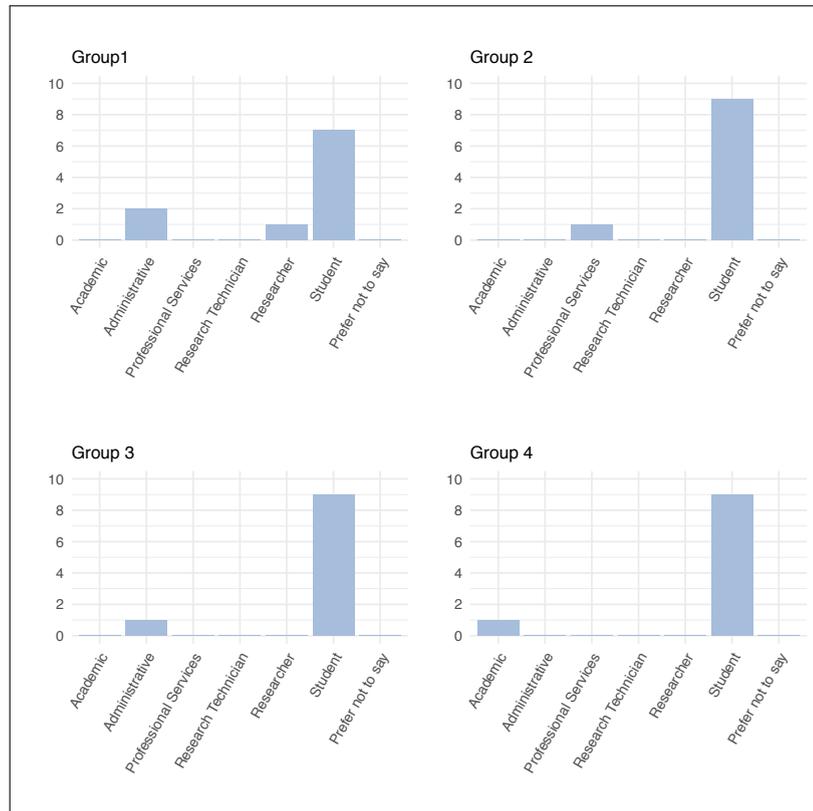


(a)

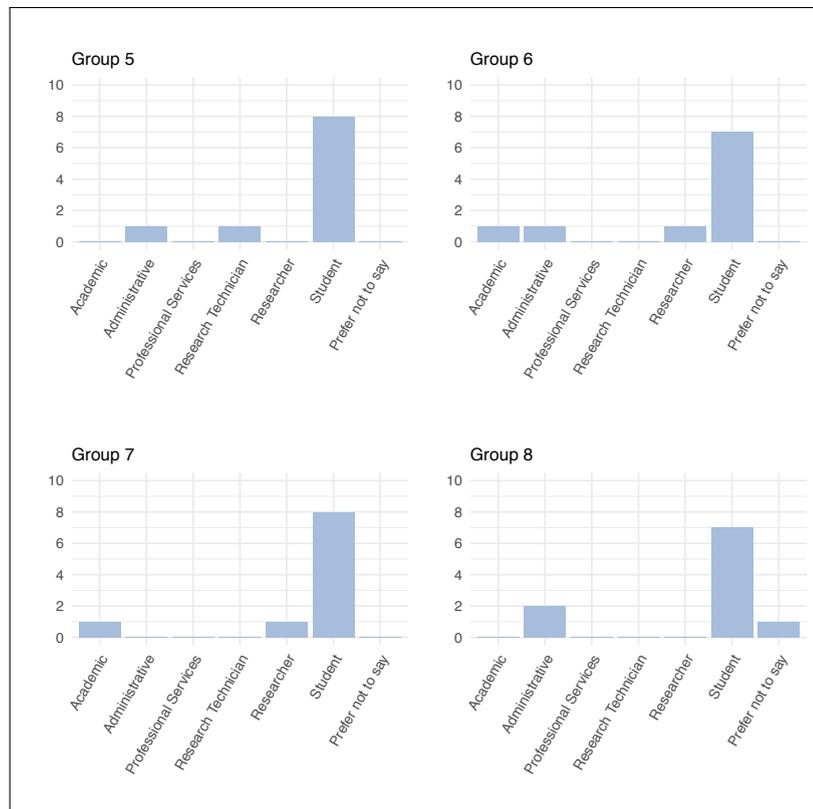


(b)

Figure 8: The Participants' Responses of their Education Field (Distribution Based on Groups).



(a)



(b)

Figure 9: The Participants' Responses of their Employment Status (Distribution Based on Groups).

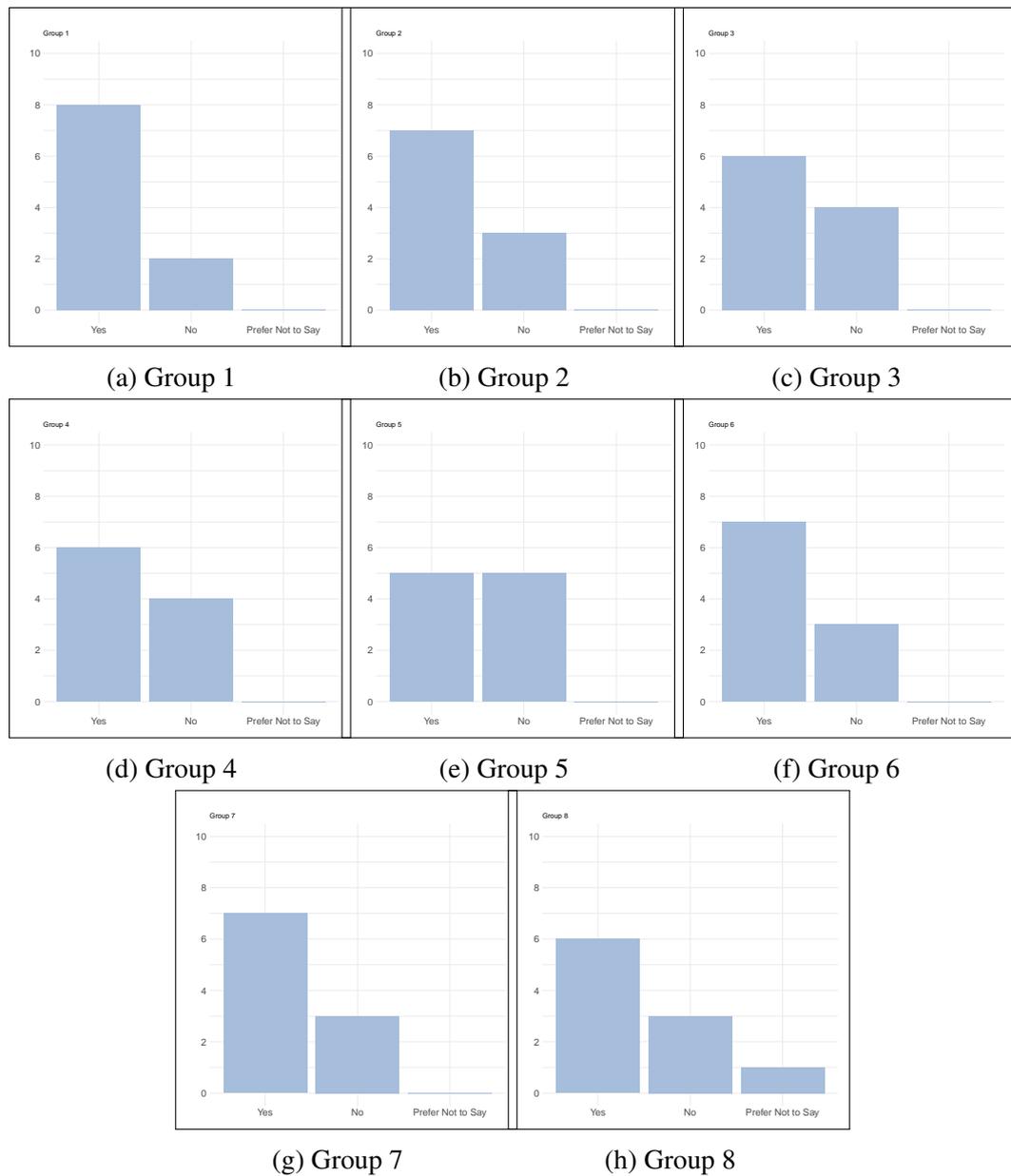


Figure 10: The Distribution of the Participants who Track and Do Not Track in Each Group. In the eight figures, the x-axis presents the participants' response to whether they track their health or physical activity data, and the Y-axis shows the number of participants.

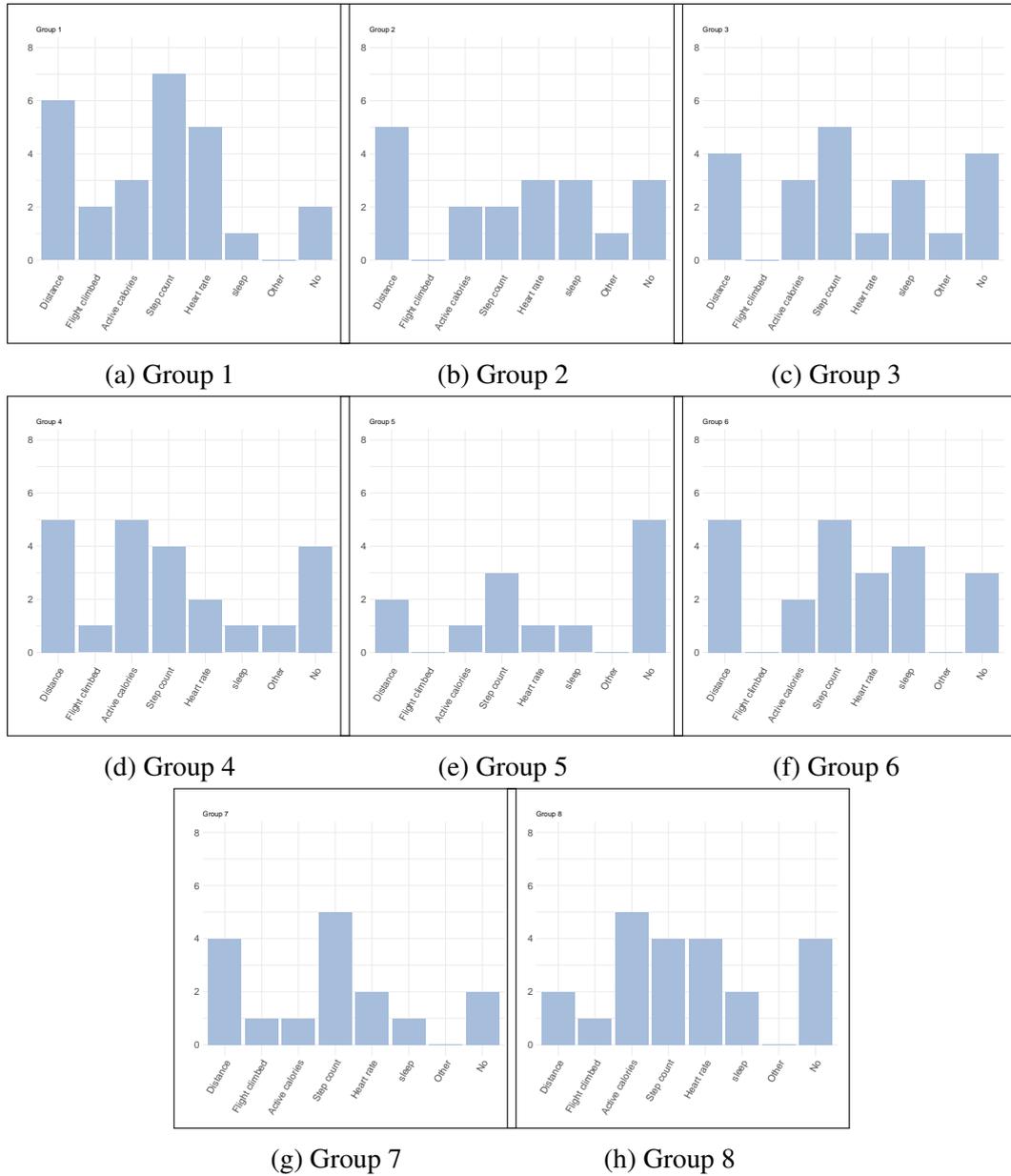


Figure 11: The Type of the Health/Physical Activity tracked by the Participants in Each Group. In the eight figures, the x-axis presents the types of the tracked data, and the Y-axis shows the number of participants who track each type of the data.

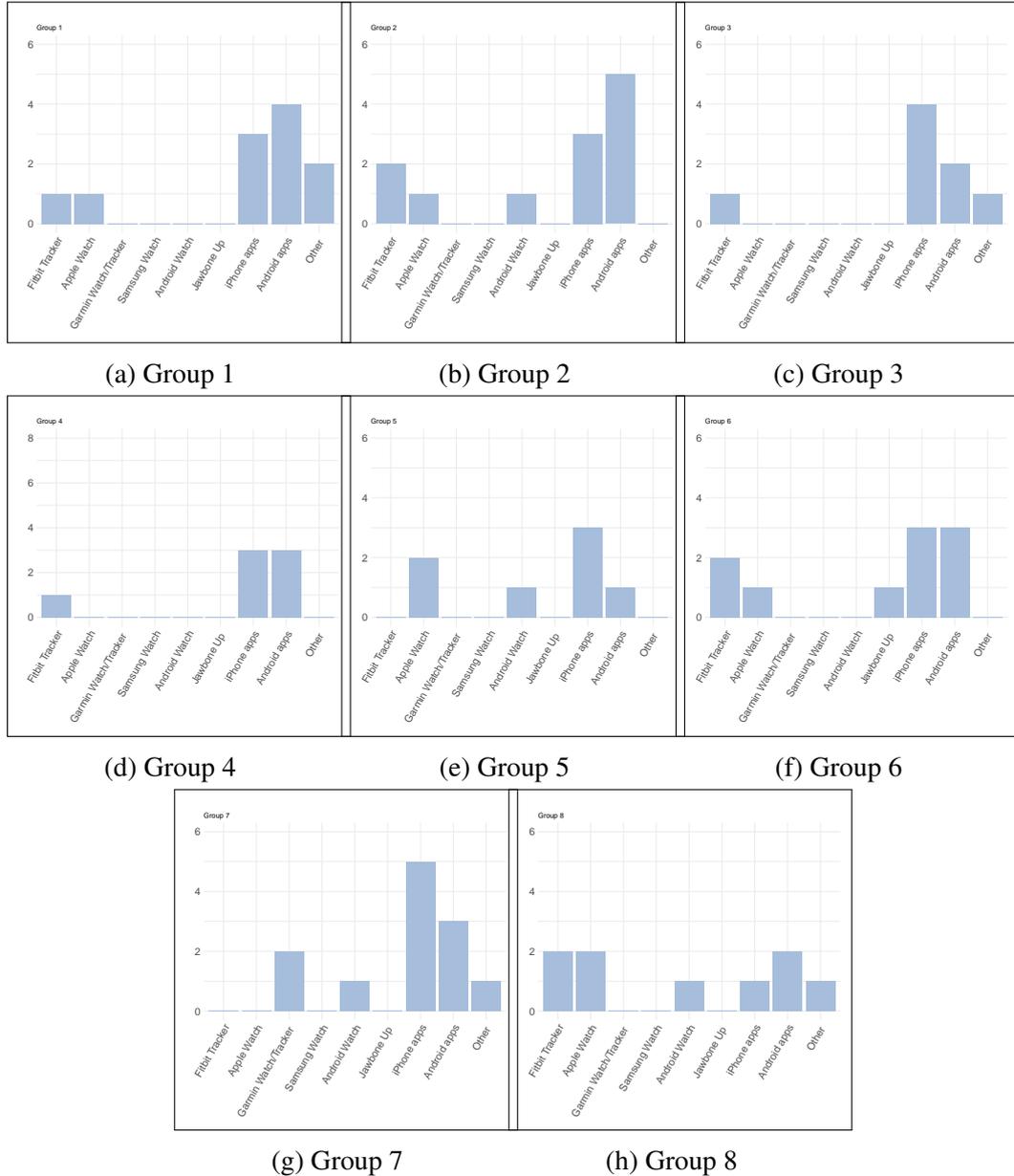
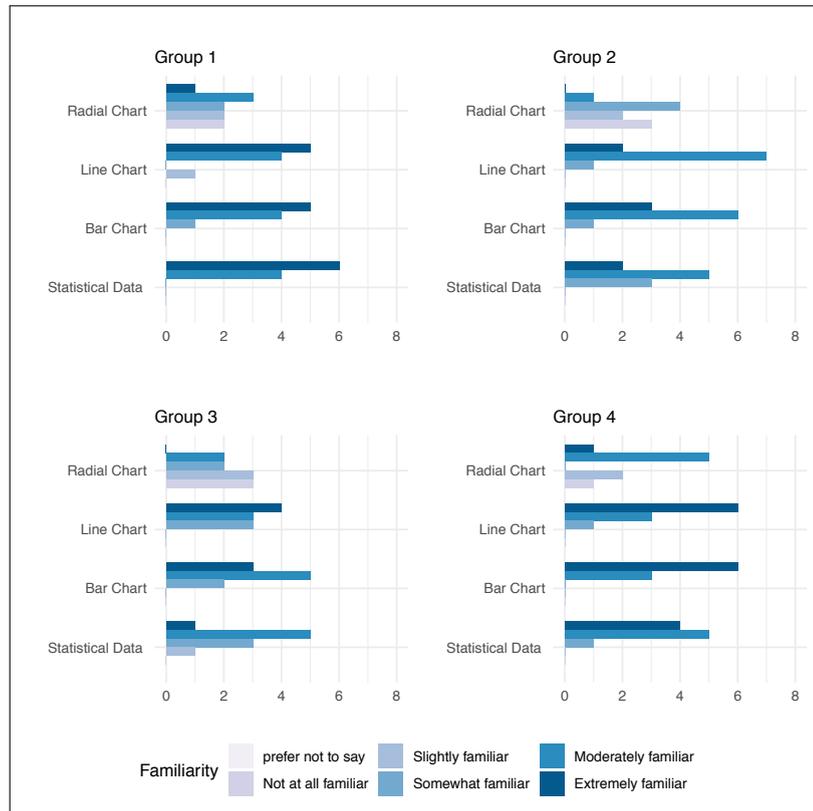
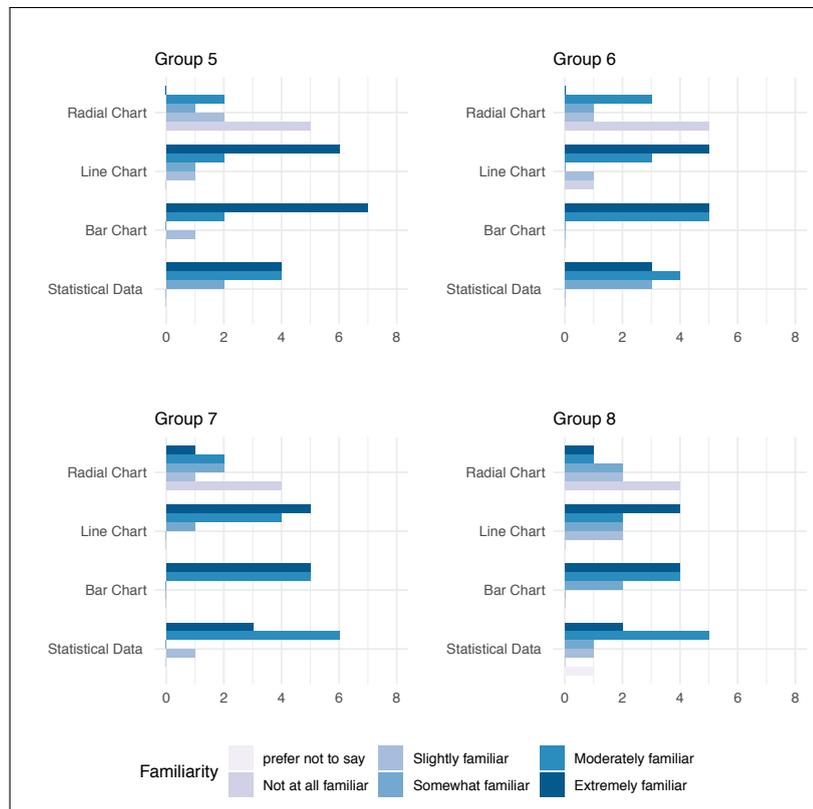


Figure 12: The Types of the Physical Activity Tracking Devices Used by the Participants in Each Group. In the eight figures, the x-axis presents the tracking devices, and the Y-axis shows the number of participants.

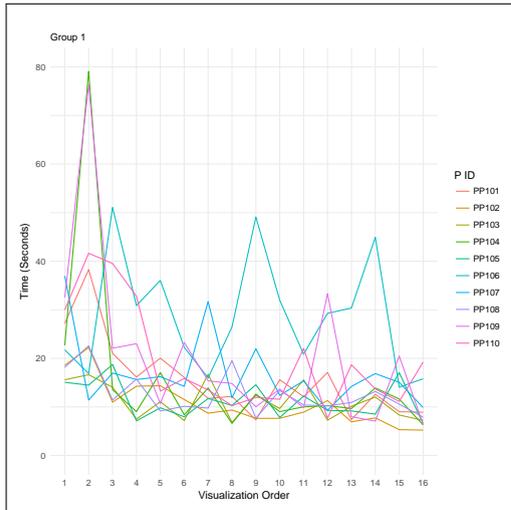


(a)

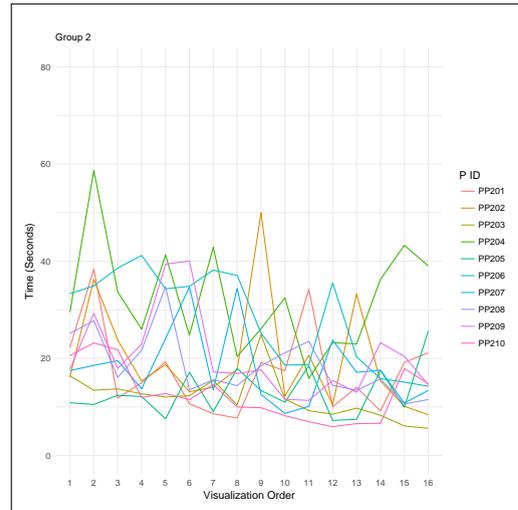


(b)

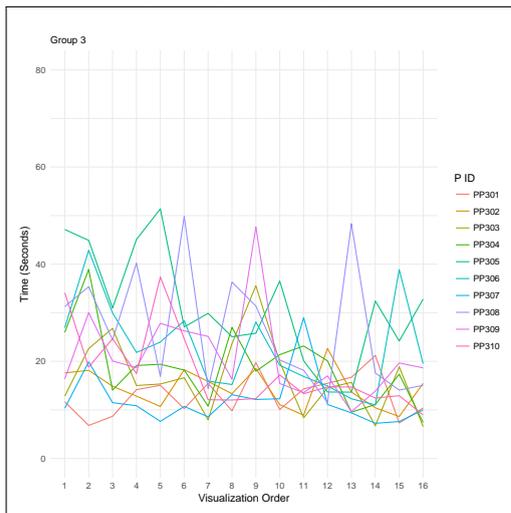
Figure 13: The Participants' Responses of their Familiarity with Different Chart and Statistical Data (Distribution Base on Groups).



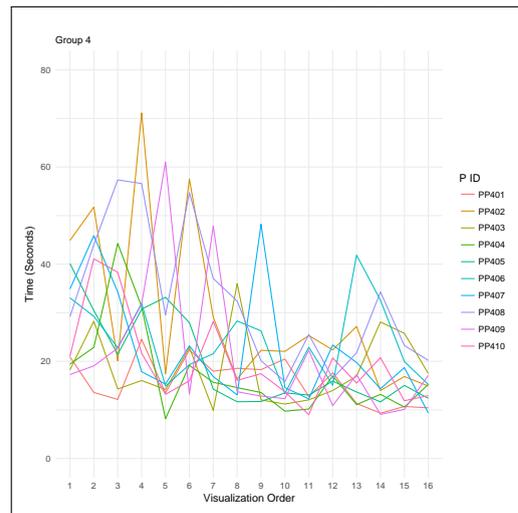
(a) Task completion time for 'Identify' task for the participants in group 1



(b) Task completion time for 'Identify' task for the participants in group 2

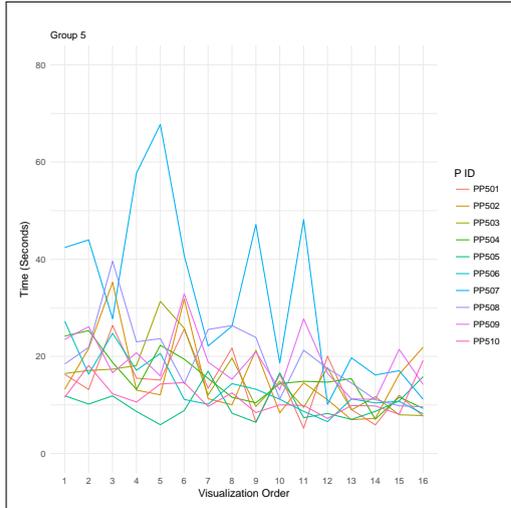


(c) Task completion time for 'Identify' task for the participants in group 3

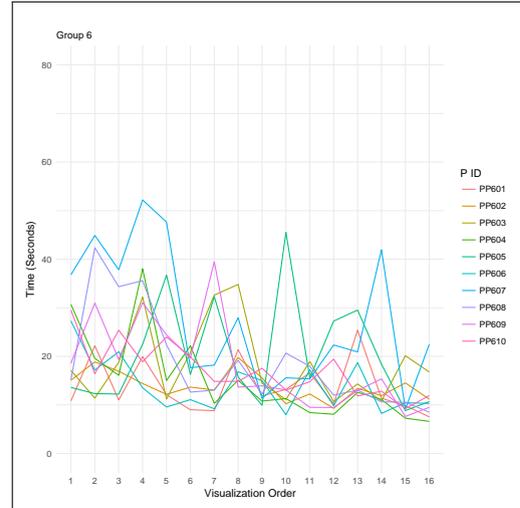


(d) Task completion time for 'Identify' task for the participants in group 4

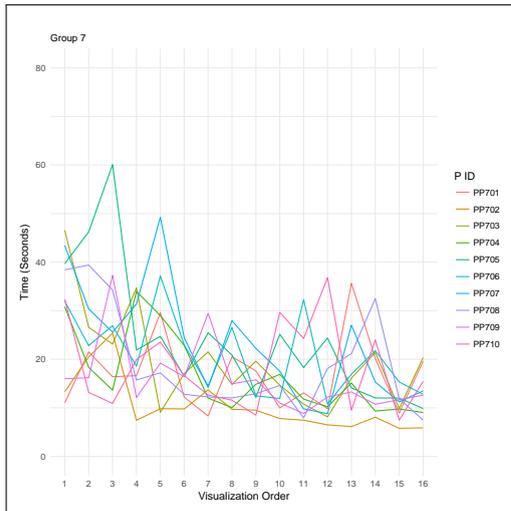
Figure 14: Task completion time for 'Identify' task for the participants in the groups (1, 2, 3 & 4). The Y-axis shows the time in seconds, and the X-axis shows the order of the presented visualizations (as each participant performs each task on 16 visualizations (8 designs with two datasets)).



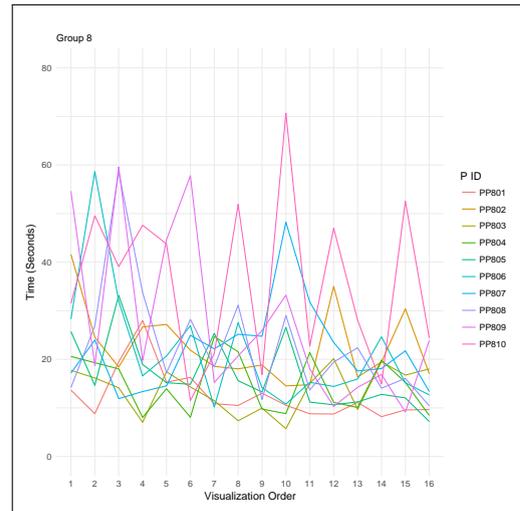
(a) Task completion time for 'Identify' task for the participants in group 5



(b) Task completion time for 'Identify' task for the participants in group 6

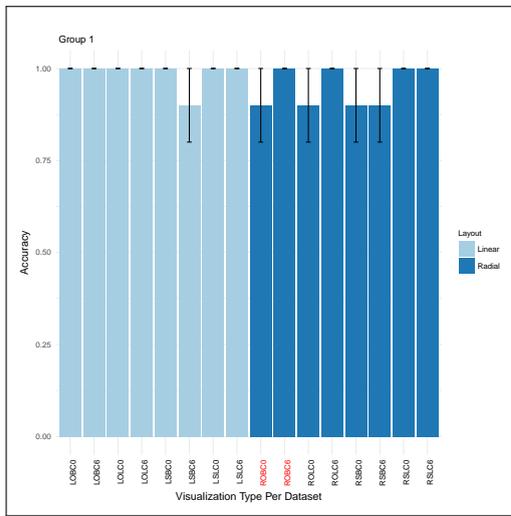


(c) Task completion time for 'Identify' task for the participants in group 7

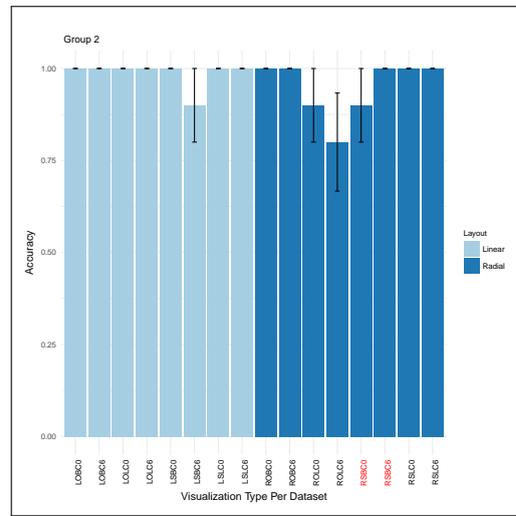


(d) Task completion time for 'Identify' task for the participants in group 8

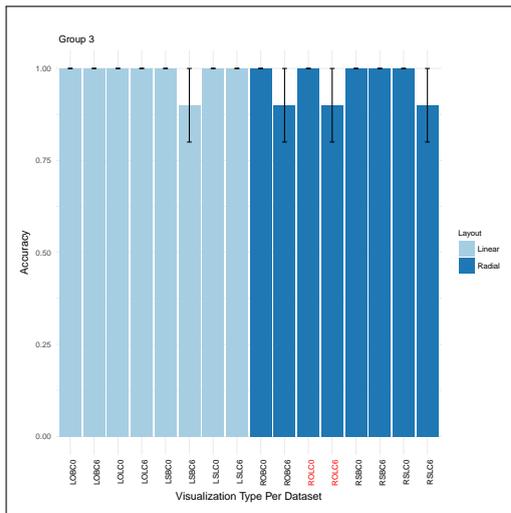
Figure 15: Task completion time for 'Identify' task for the participants in the groups (5, 6, 7 & 8). The Y-axis shows the time in seconds, and the X-axis shows the order of the presented visualizations (as each participant performs each task on 16 visualizations (8 designs with two datasets)).



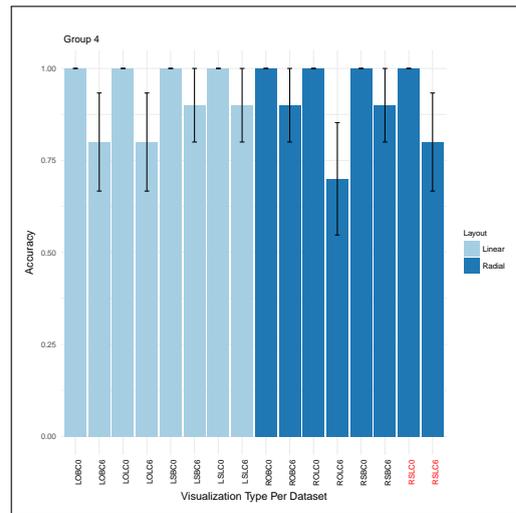
(a) Task Accuracy for the 'Identify' Task (Group 1)



(b) Task Accuracy for the 'Identify' Task (Group 2)

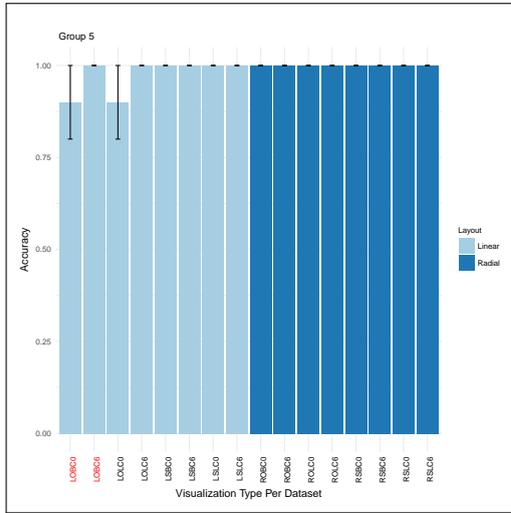


(c) Task Accuracy for the 'Identify' Task (Group 3)

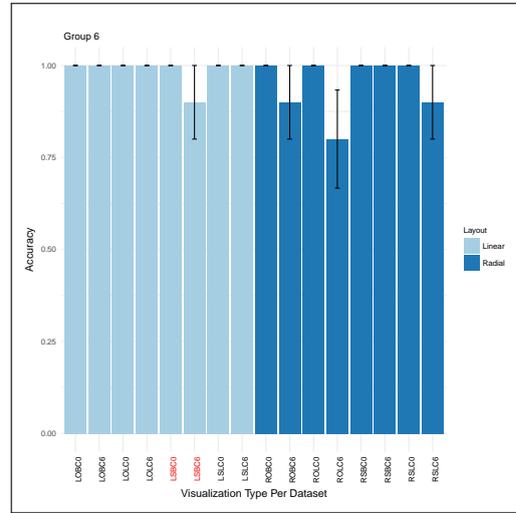


(d) Task Accuracy for the 'Identify' Task (Group 4)

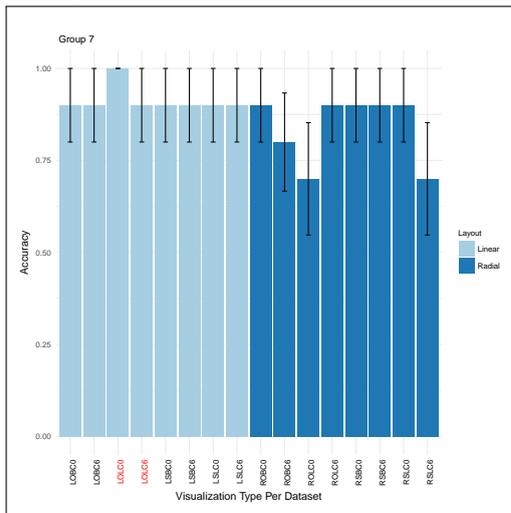
Figure 16: Task Accuracy for the 'Identify' Task on the Eight Visualization Designs Per Dataset (for the Groups 1, 2, 3 & 4). The graph shows the average accuracy for the 'Identify' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



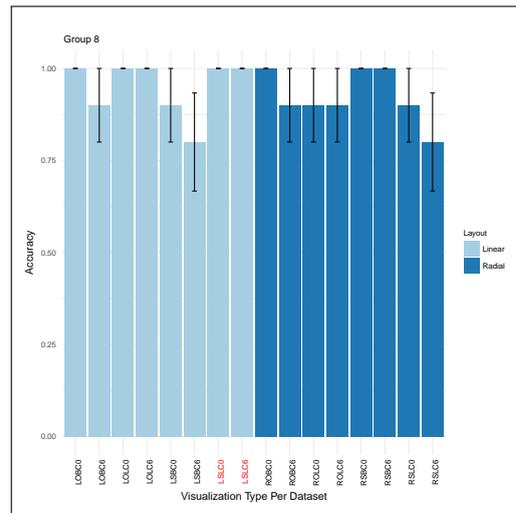
(a) Task Accuracy for the 'Identify' Task (Group 5)



(b) Task Accuracy for the 'Identify' Task (Group 6)

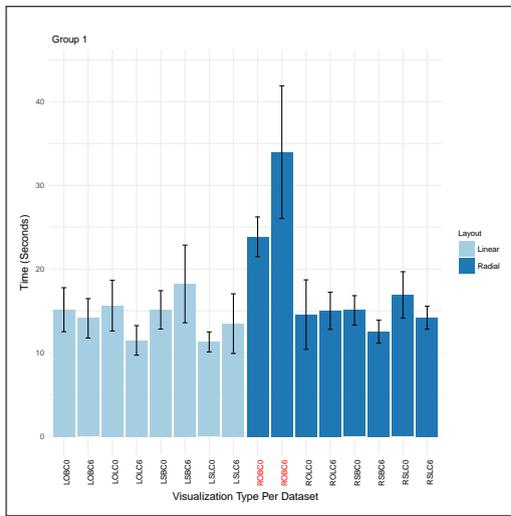


(c) Task Accuracy for the 'Identify' Task (Group 7)

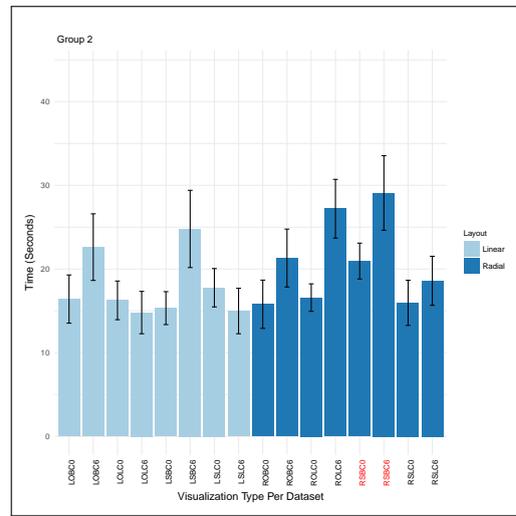


(d) Task Accuracy for the 'Identify' Task (Group 8)

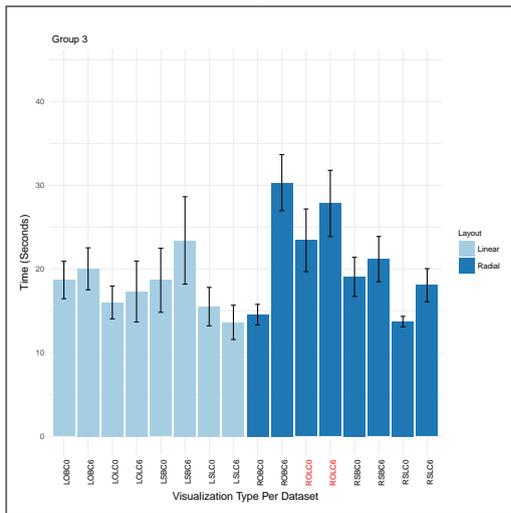
Figure 17: Task Accuracy for the 'Identify' Task on the Eight Visualization Designs Per Dataset (for the Groups 5, 6, 7 & 8). The graph shows the average accuracy for the 'Identify' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



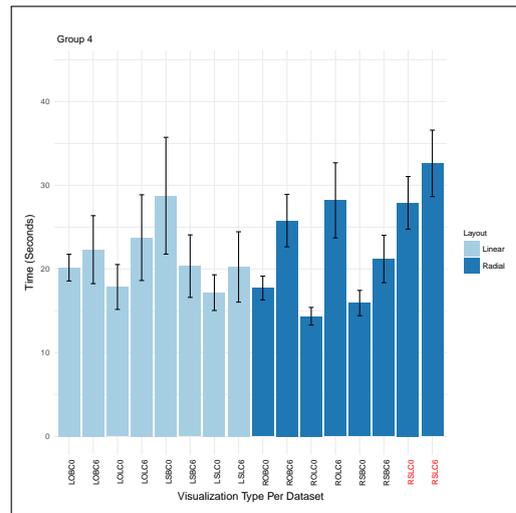
(a) Task Completion Time for the 'Identify' Task (Group 1)



(b) Task Completion Time for the 'Identify' Task (Group 2)

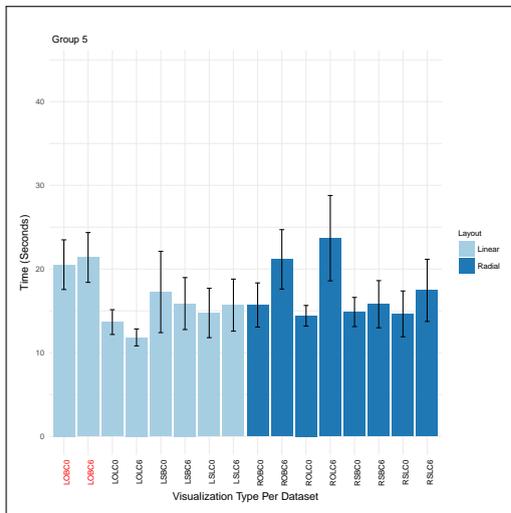


(c) Task Completion Time for the 'Identify' Task (Group 3)

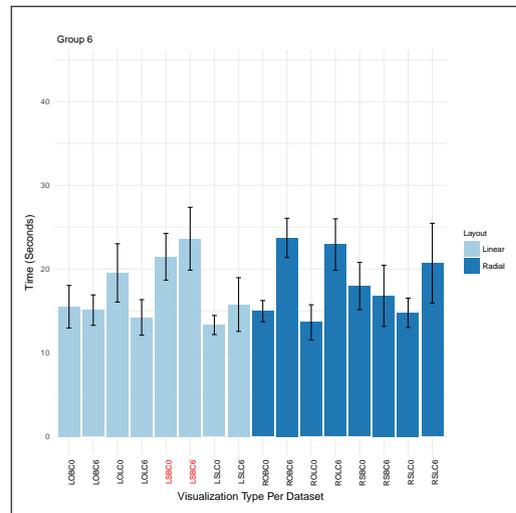


(d) Task Completion Time for the 'Identify' Task (Group 4)

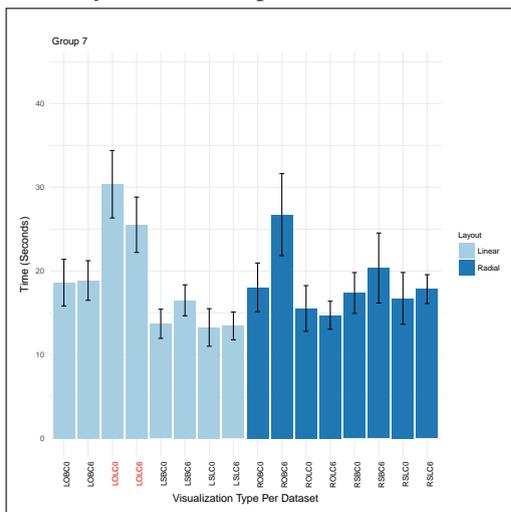
Figure 18: Task Completion Time for the 'Identify' Task on the Eight Visualization Designs Per Dataset (for the Groups 1, 2, 3 & 4). The graph shows the average time for completing the 'Identify' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



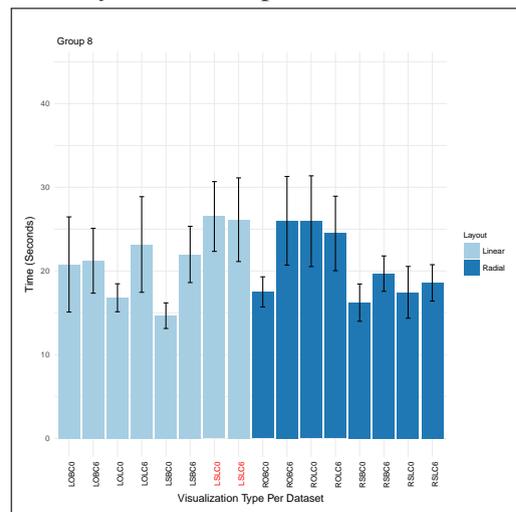
(a) Task Completion Time for the 'Identify' Task (Group 5)



(b) Task Completion Time for the 'Identify' Task (Group 6)

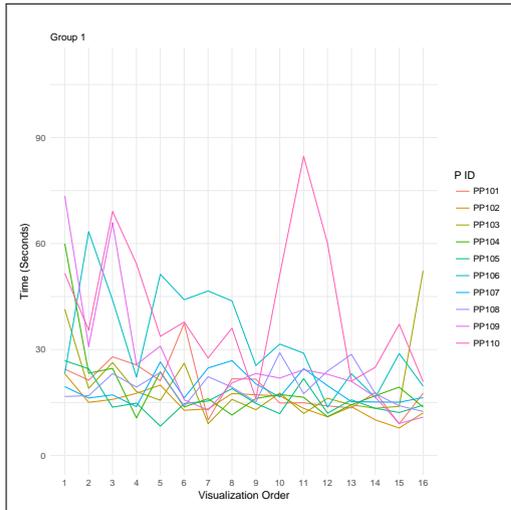


(c) Task Completion Time for the 'Identify' Task (Group 7)

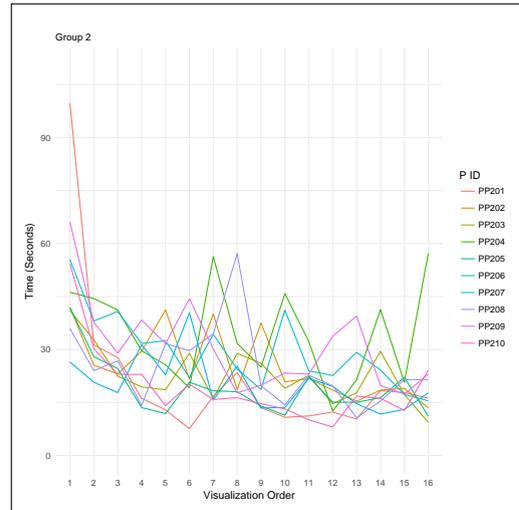


(d) Task Completion Time for the 'Identify' Task (Group 8)

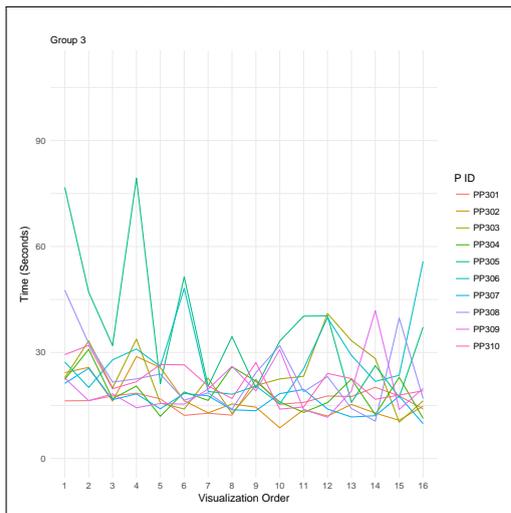
Figure 19: Task Completion Time for the 'Identify' Task on the Eight Visualization Designs Per Dataset (for the Groups 5, 6, 7 & 8). The graph shows the average time for completing the 'Identify' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



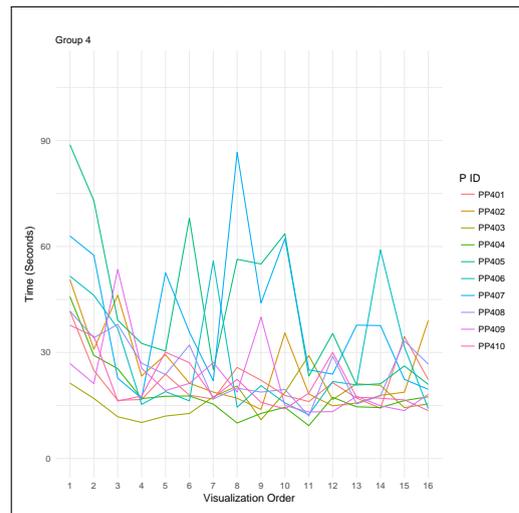
(a) Task completion time for 'Determine' task for the participants in group 1



(b) Task completion time for 'Determine' task for the participants in group 2

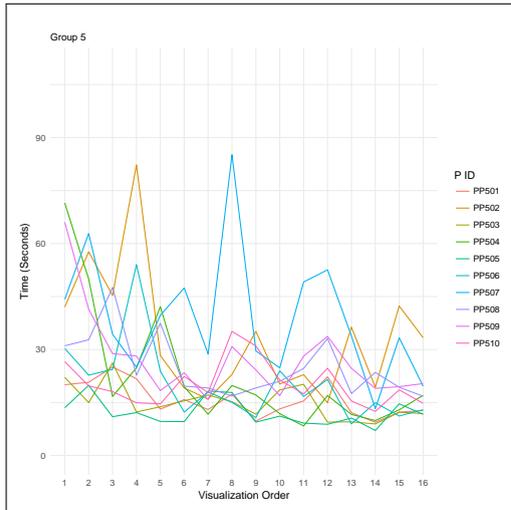


(c) Task completion time for 'Determine' task for the participants in group 3

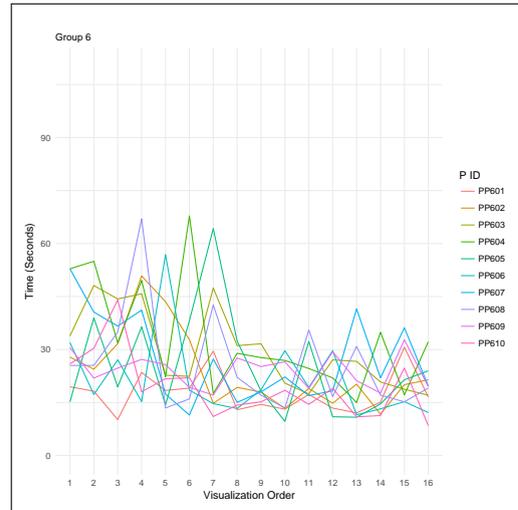


(d) Task completion time for 'Determine' task for the participants in group 4

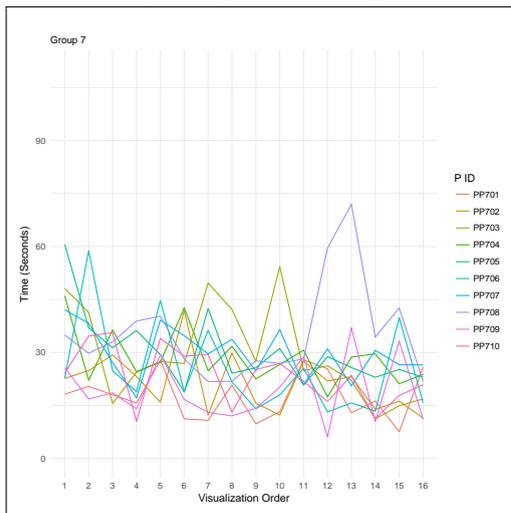
Figure 20: Task completion time for 'Determine' task for the participants in the groups (1, 2, 3 & 4). The Y-axis shows the time in seconds, and the X-axis shows the order of the presented visualizations (as each participant performs each task on 16 visualizations (8 designs with two datasets)).



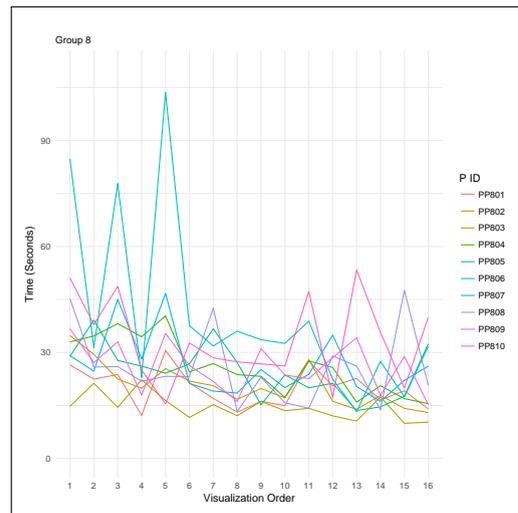
(a) Task completion time for 'Determine' task for the participants in group 5



(b) Task completion time for 'Determine' task for the participants in group 6

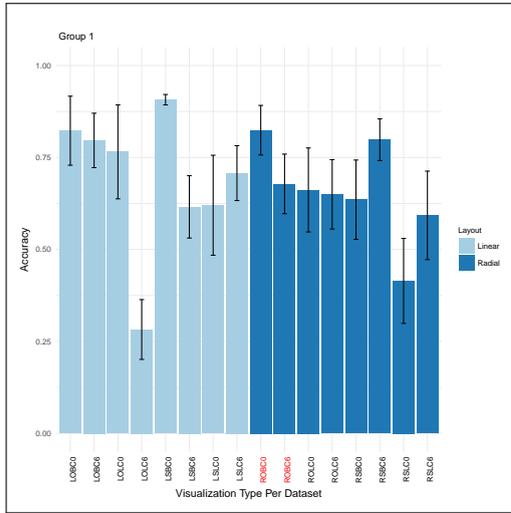


(c) Task completion time for 'Determine' task for the participants in group 7

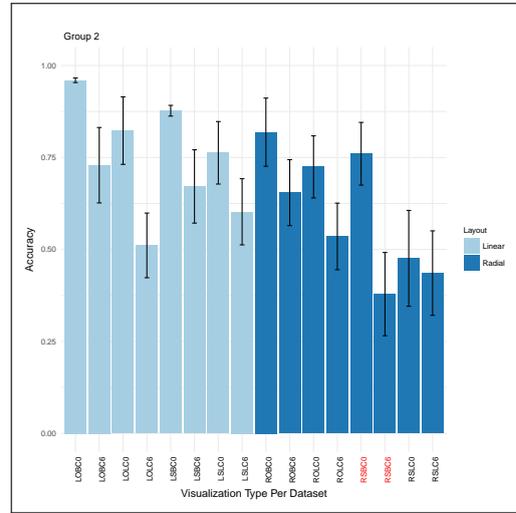


(d) Task completion time for 'Determine' task for the participants in group 8

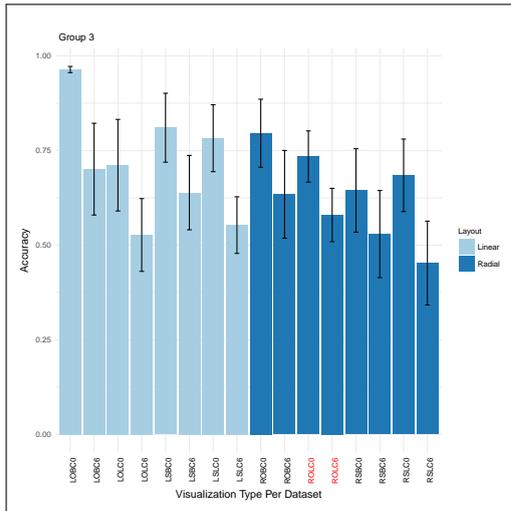
Figure 21: Task completion time for 'Determine' task for the participants in the groups (5, 6, 7 & 8). The Y-axis shows the time in seconds, and the X-axis shows the order of the presented visualizations (as each participant performs each task on 16 visualizations (8 designs with two datasets)).



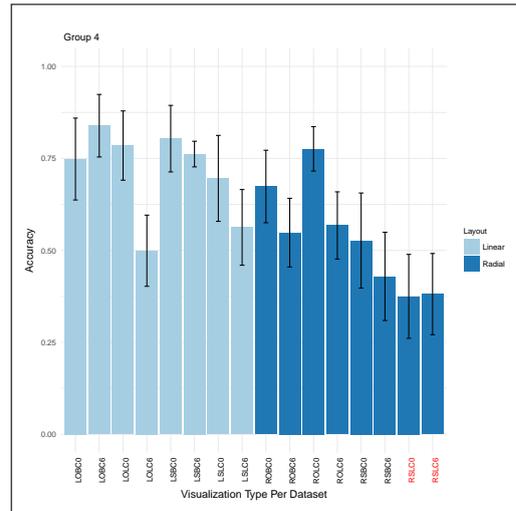
(a) Task Accuracy for the 'Determine' Task (Group 1)



(b) Task Accuracy for the 'Determine' Task (Group 2)

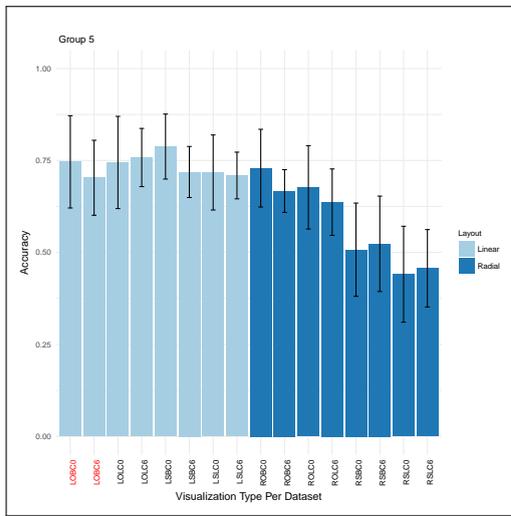


(c) Task Accuracy for the 'Determine' Task (Group 3)

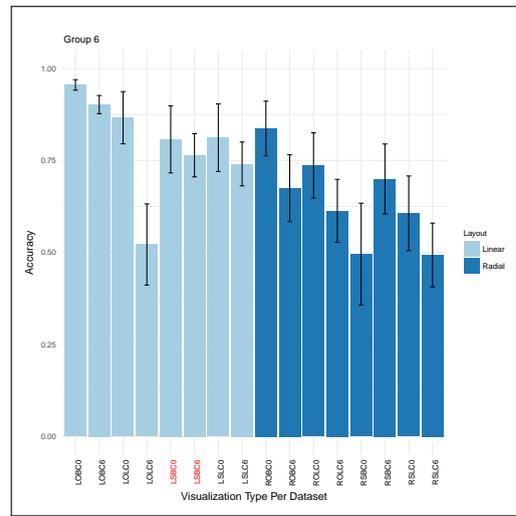


(d) Task Accuracy for the 'Determine' Task (Group 4)

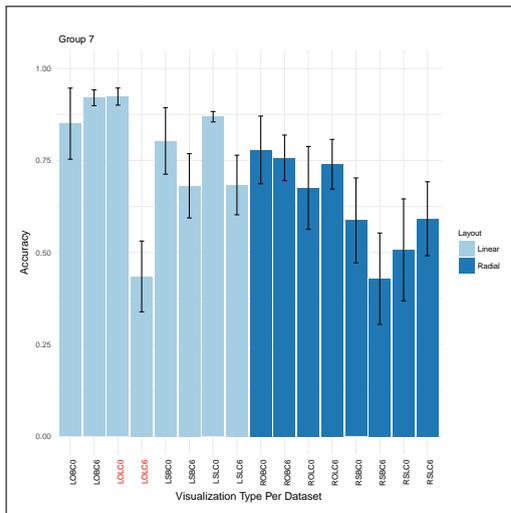
Figure 22: Task Accuracy for the 'Determine' Task on the Eight Visualization Designs Per Dataset (for the Groups 1, 2, 3 & 4). The graph shows the average accuracy for the 'Determine' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



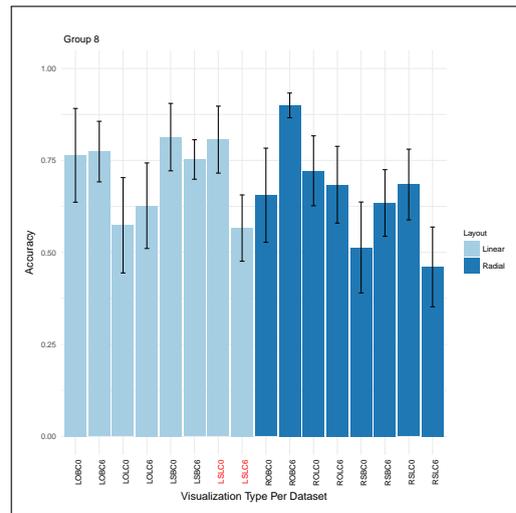
(a) Task Accuracy for the 'Determine' Task (Group 5)



(b) Task Accuracy for the 'Determine' Task (Group 6)

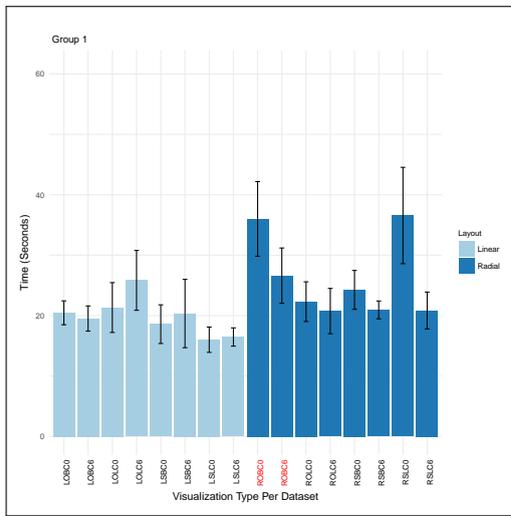


(c) Task Accuracy for the 'Determine' Task (Group 7)

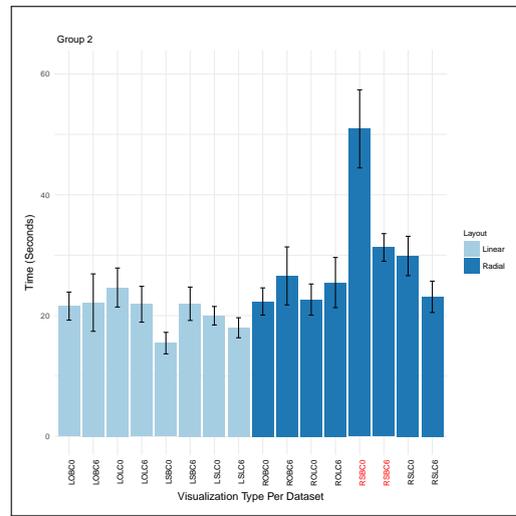


(d) Task Accuracy for the 'Determine' Task (Group 8)

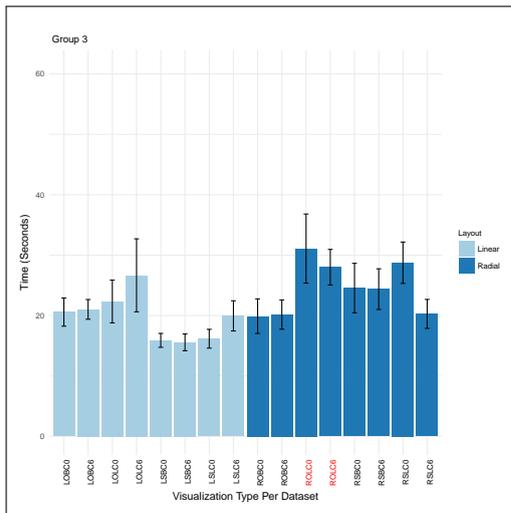
Figure 23: Task Accuracy for the 'Determine' Task on the Eight Visualization Designs Per Dataset (for the Groups 5, 6, 7 & 8). The graph shows the average accuracy for the 'Determine' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



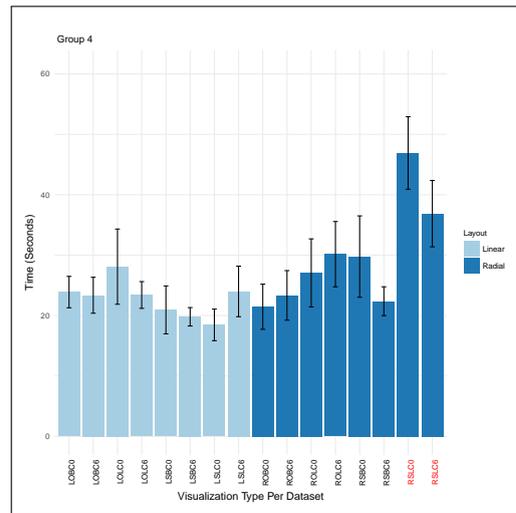
(a) Task Completion Time for the 'Determine' Task (Group 1)



(b) Task Completion Time for the 'Determine' Task (Group 2)

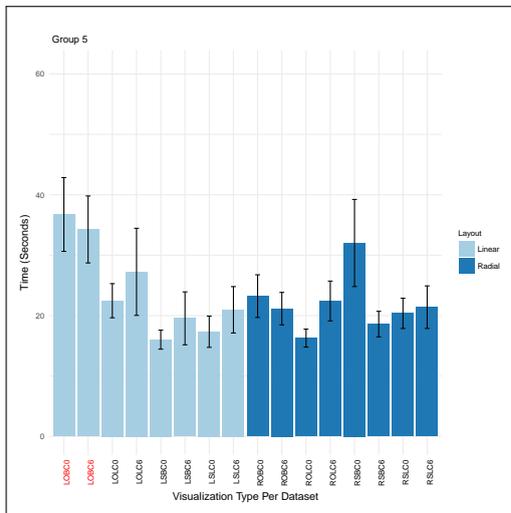


(c) Task Completion Time for the 'Determine' Task (Group 3)

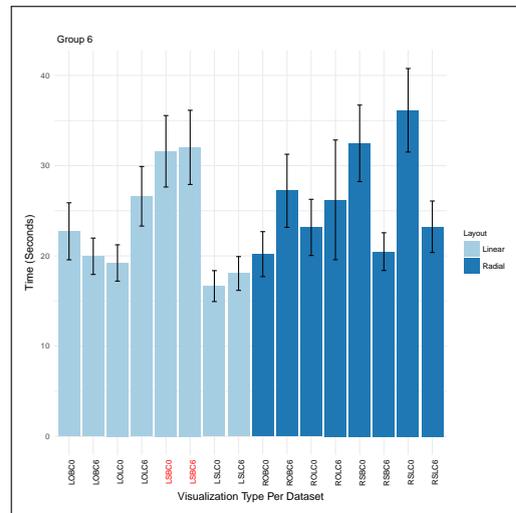


(d) Task Completion Time for the 'Determine' Task (Group 4)

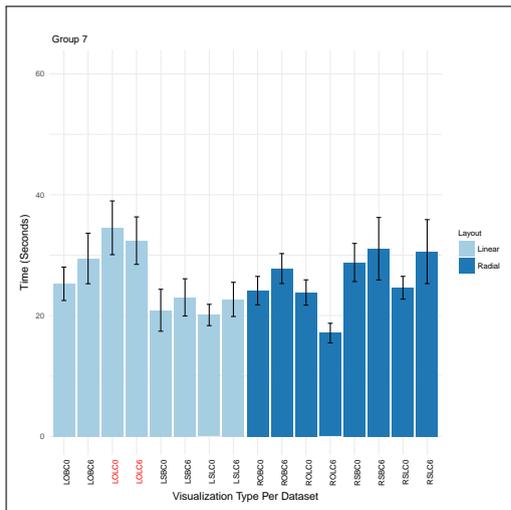
Figure 24: Task Completion Time for the 'Determine' Task on the Eight Visualization Designs Per Dataset (for the Groups 1, 2, 3 & 4). The graph shows the average time for completing the 'Determine' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



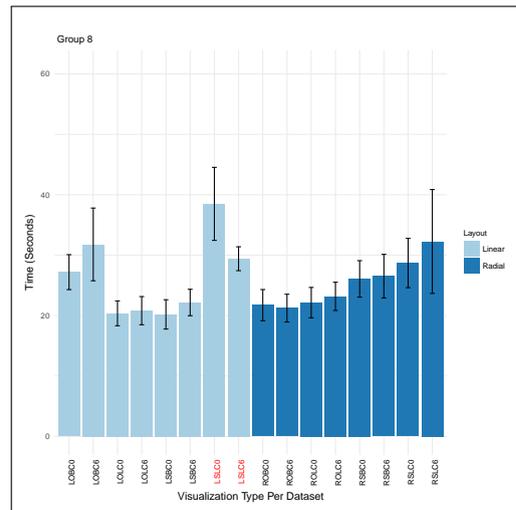
(a) Task Completion Time for the ‘Determine’ Task (Group 5)



(b) Task Completion Time for the ‘Determine’ Task (Group 6)

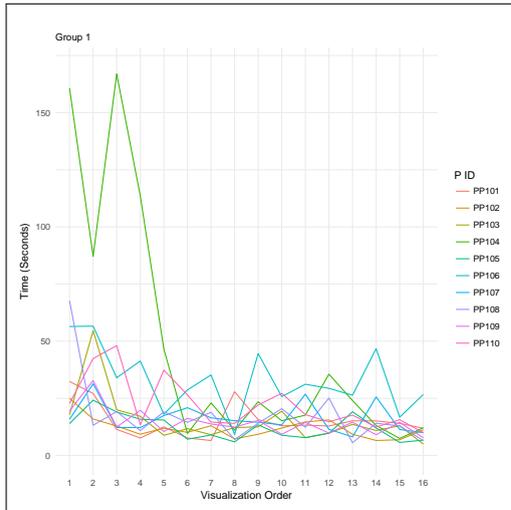


(c) Task Completion Time for the ‘Determine’ Task (Group 7)

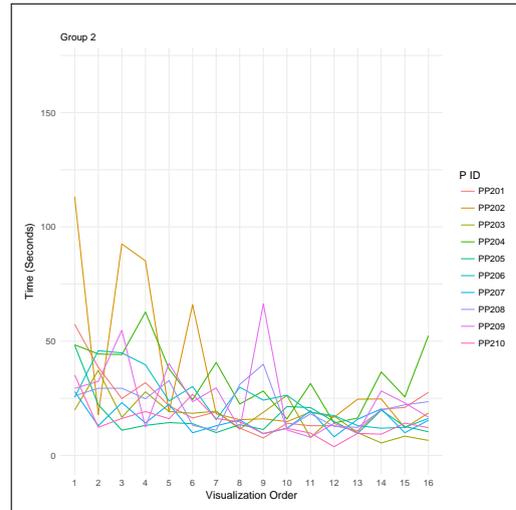


(d) Task Completion Time for the ‘Determine’ Task (Group 8)

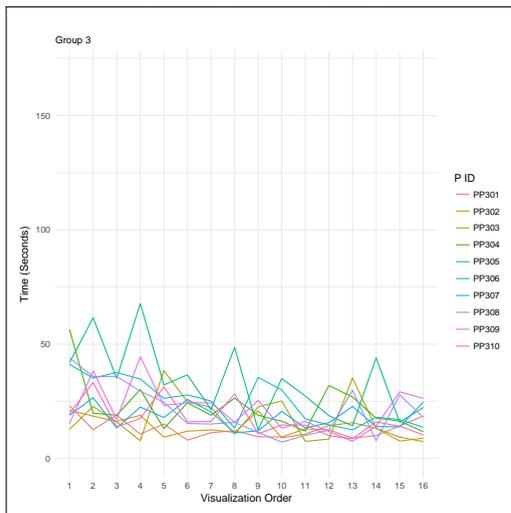
Figure 25: Task Completion Time for the ‘Determine’ Task on the Eight Visualization Designs Per Dataset (for the Groups 5, 6, 7 & 8). The graph shows the average time for completing the ‘Determine’ task and the standard errors. The numbers ‘0’ and ‘6’ after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



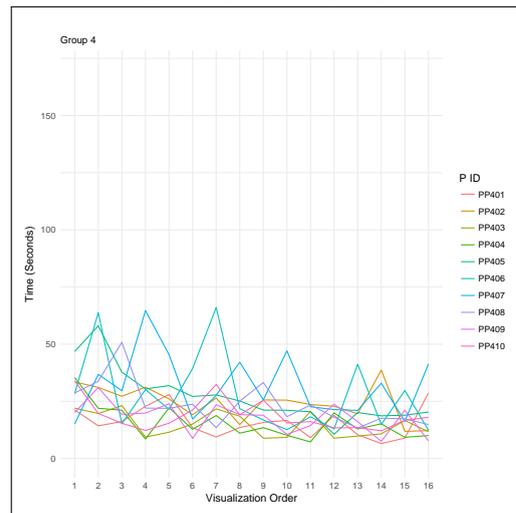
(a) Task completion time for 'Compare' task for the participants in group 1



(b) Task completion time for 'Compare' task for the participants in group 2

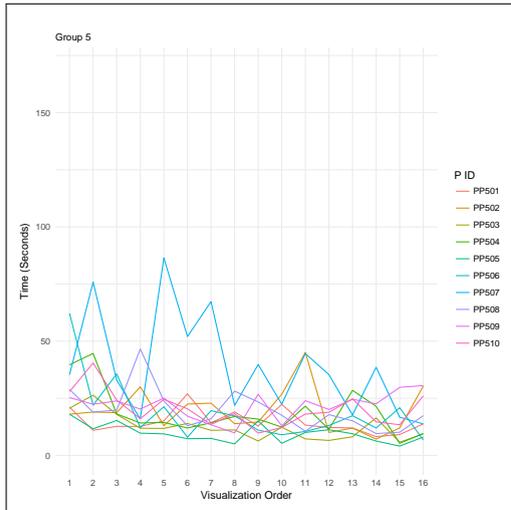


(c) Task completion time for 'Compare' task for the participants in group 3

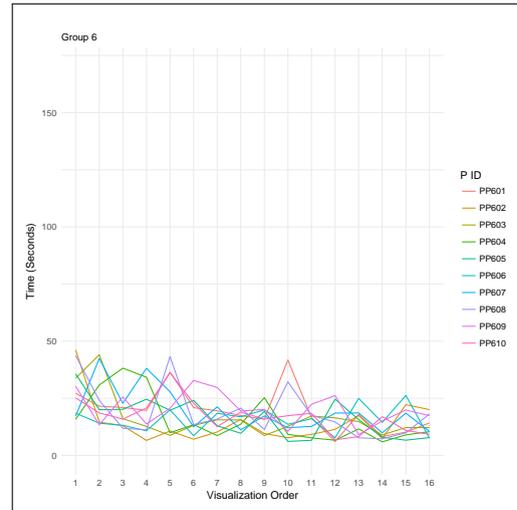


(d) Task completion time for 'Compare' task for the participants in group 4

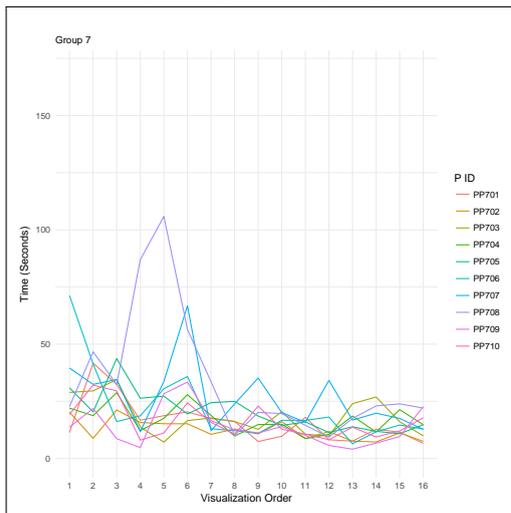
Figure 26: Task completion time for 'Compare' task for the participants in the groups (1, 2, 3 & 4). The Y-axis shows the time in seconds, and the X-axis shows the order of the presented visualizations (as each participant performs each task on 16 visualizations (8 designs with two datasets)).



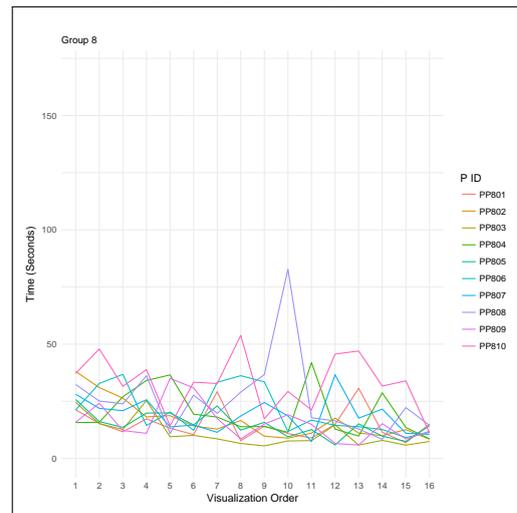
(a) Task completion time for 'Compare' task for the participants in group 5



(b) Task completion time for 'Compare' task for the participants in group 6

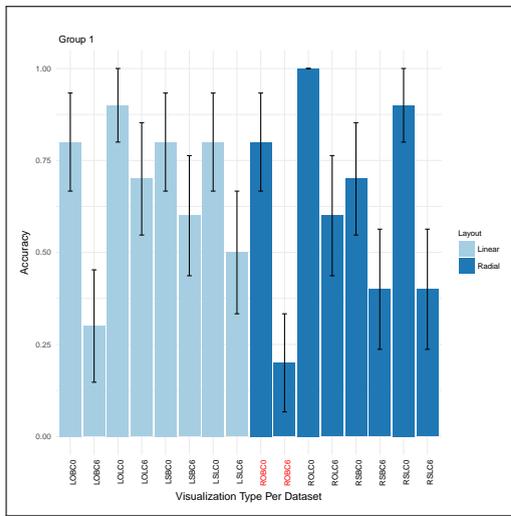


(c) Task completion time for 'Compare' task for the participants in group 7

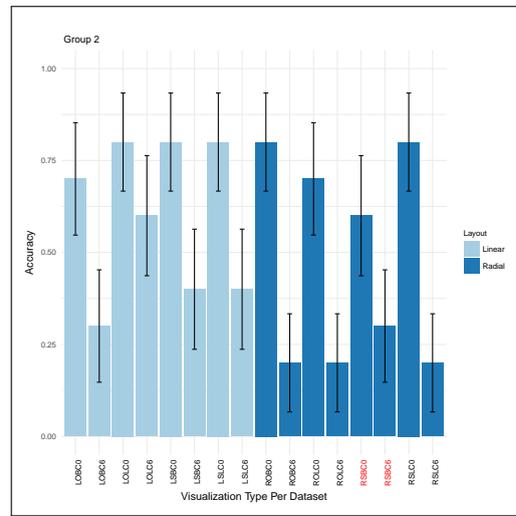


(d) Task completion time for 'Compare' task for the participants in group 8

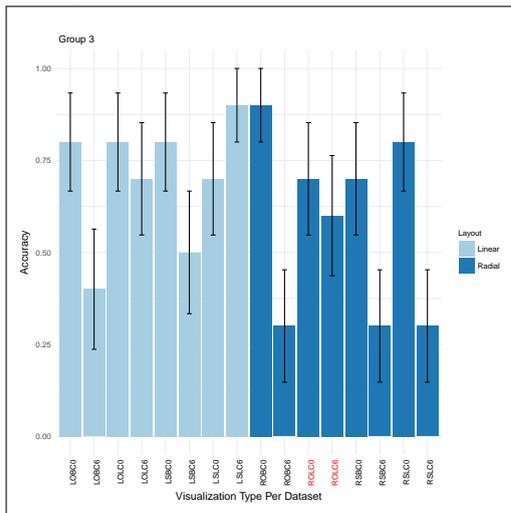
Figure 27: Task completion time for 'Compare' task for the participants in the groups (5, 6, 7 & 8). The Y-axis shows the time in seconds, and the X-axis shows the order of the presented visualizations (as each participant performs each task on 16 visualizations (8 designs with two datasets)).



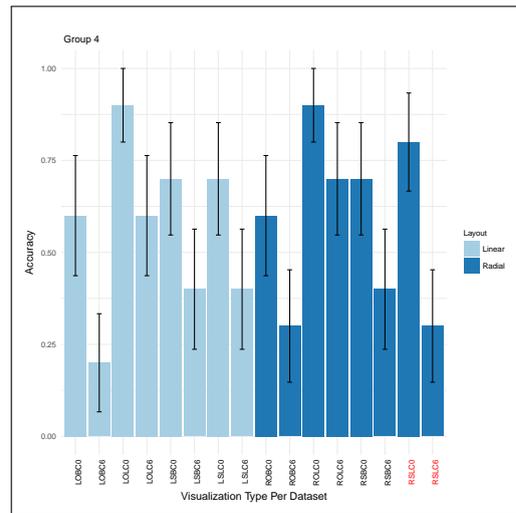
(a) Task Accuracy for the 'Compare' Task (Group 1)



(b) Task Accuracy for the 'Compare' Task (Group 2)

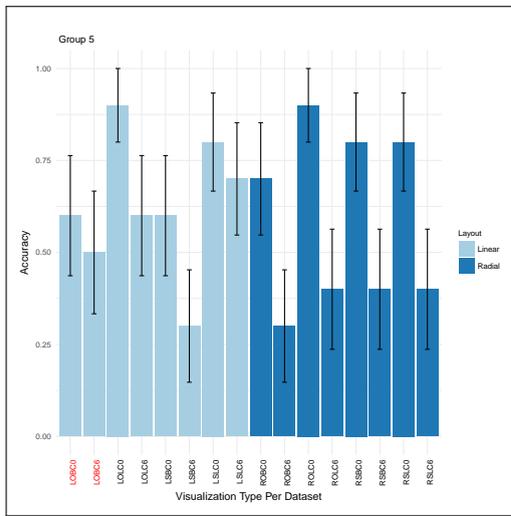


(c) Task Accuracy for the 'Compare' Task (Group 3)

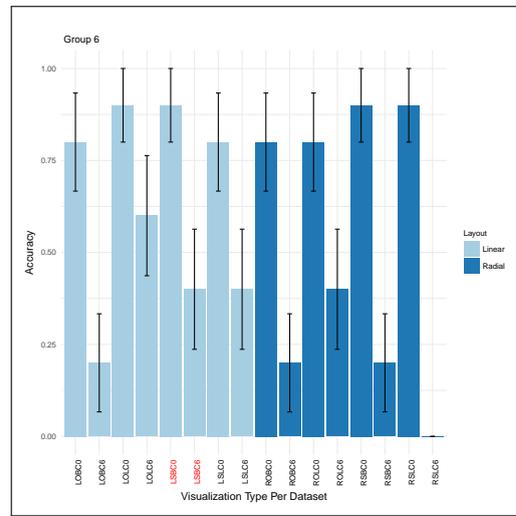


(d) Task Accuracy for the 'Compare' Task (Group 4)

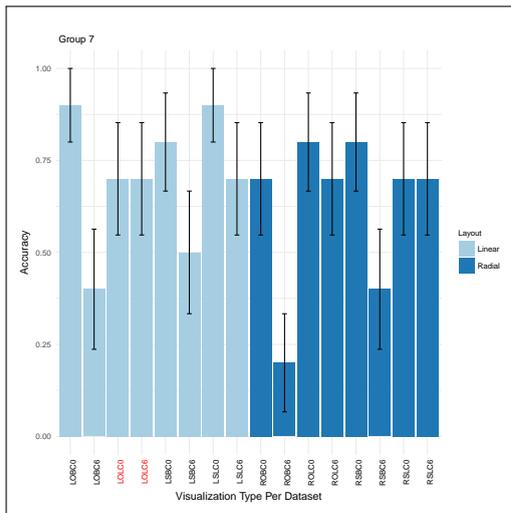
Figure 28: Task Accuracy for the 'Compare' Task on the Eight Visualization Designs Per Dataset (for the Groups 1, 2, 3 & 4). The graph shows the average accuracy for the 'Compare' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



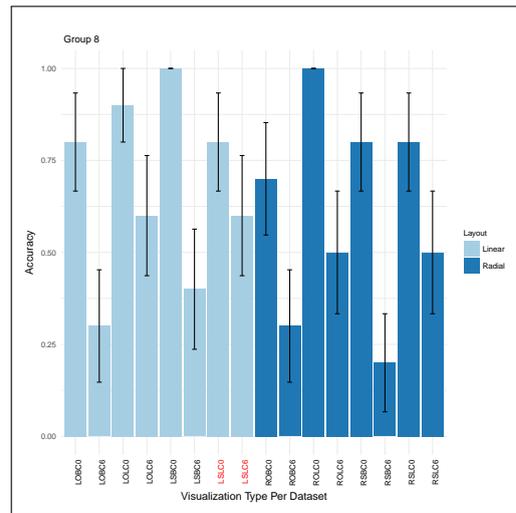
(a) Task Accuracy for the 'Compare' Task (Group 5)



(b) Task Accuracy for the 'Compare' Task (Group 6)

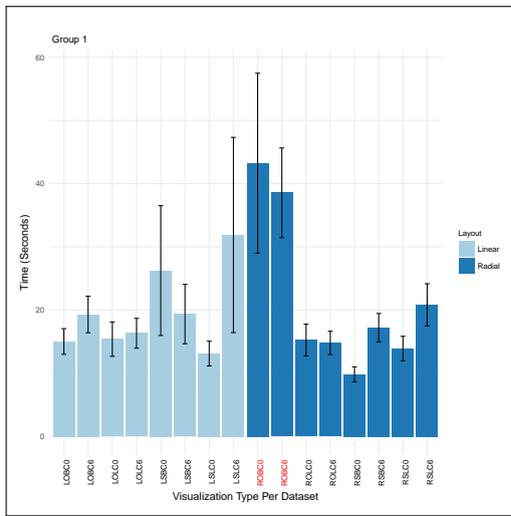


(c) Task Accuracy for the 'Compare' Task (Group 7)

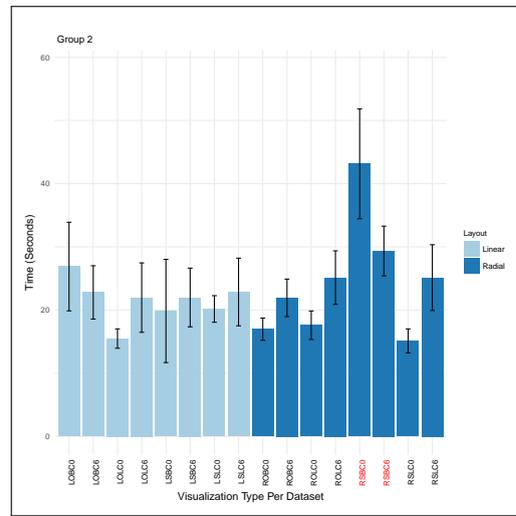


(d) Task Accuracy for the 'Compare' Task (Group 8)

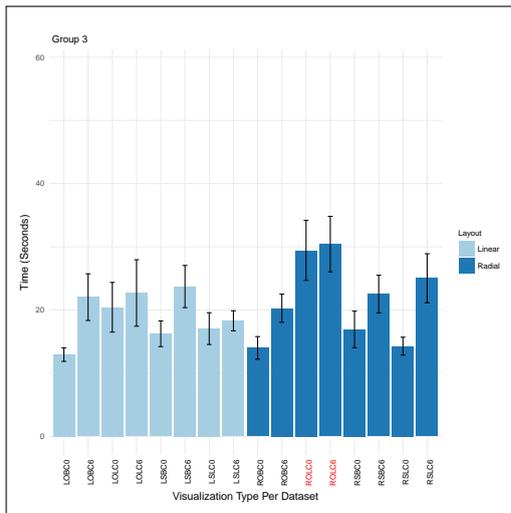
Figure 29: Task Accuracy for the 'Compare' Task on the Eight Visualization Designs Per Dataset (for the Groups 5, 6, 7 & 8). The graph shows the average accuracy for the 'Compare' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



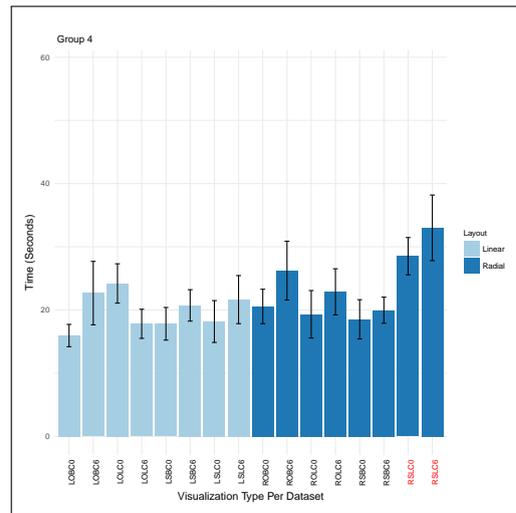
(a) Task Completion Time for the 'Compare' Task (Group 1)



(b) Task Completion Time for the 'Compare' Task (Group 2)

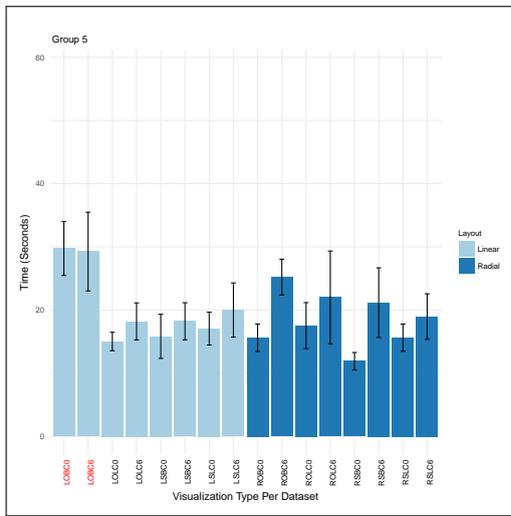


(c) Task Completion Time for the 'Compare' Task (Group 3)

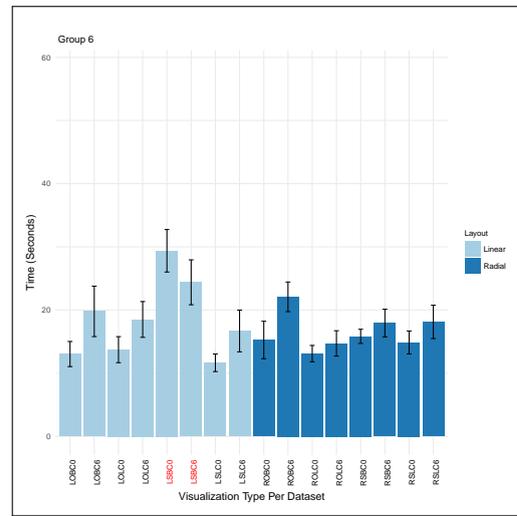


(d) Task Completion Time for the 'Compare' Task (Group 4)

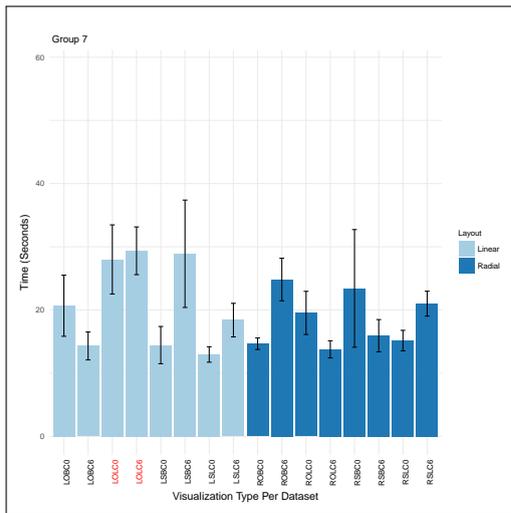
Figure 30: Task Completion Time for the 'Compare' Task on the Eight Visualization Designs Per Dataset (for the Groups 1, 2, 3 & 4). The graph shows the average time for completing the 'Compare' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



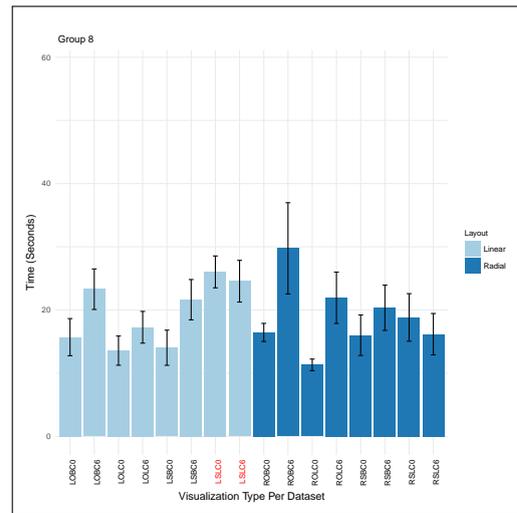
(a) Task Completion Time for the 'Compare' Task (Group 5)



(b) Task Completion Time for the 'Compare' Task (Group 6)

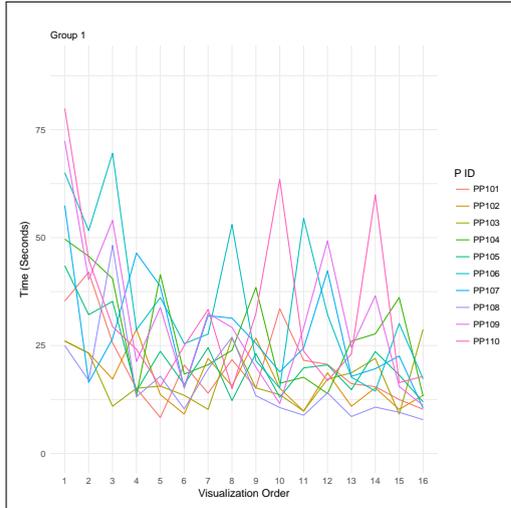


(c) Task Completion Time for the 'Compare' Task (Group 7)

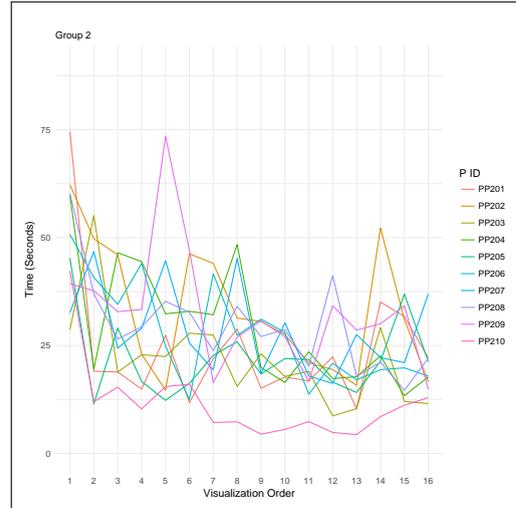


(d) Task Completion Time for the 'Compare' Task (Group 8)

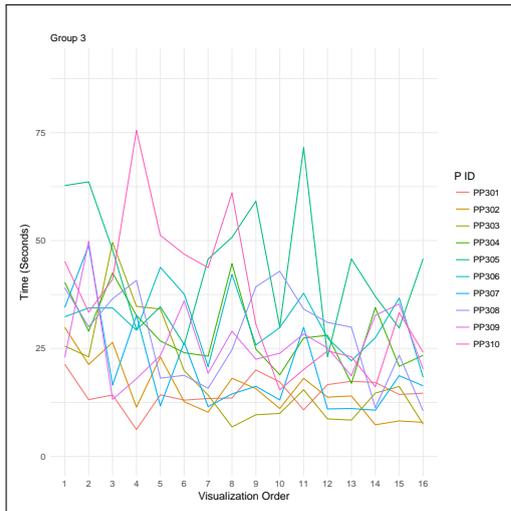
Figure 31: Task Completion Time for the 'Compare' Task on the Eight Visualization Designs Per Dataset (for the Groups 5, 6, 7 & 8). The graph shows the average time for completing the 'Compare' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



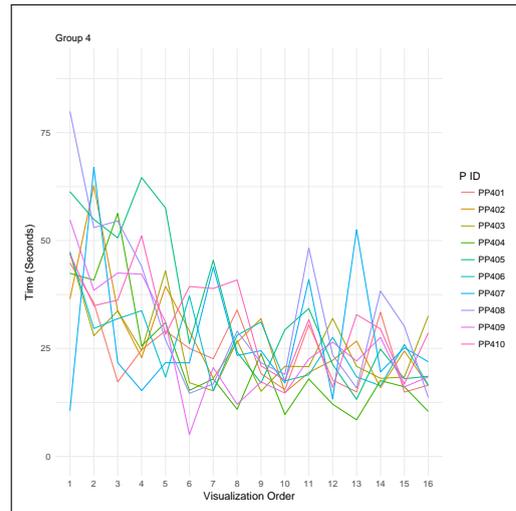
(a) Task completion time for 'Infer' task for the participants in group 1



(b) Task completion time for 'Infer' task for the participants in group 2

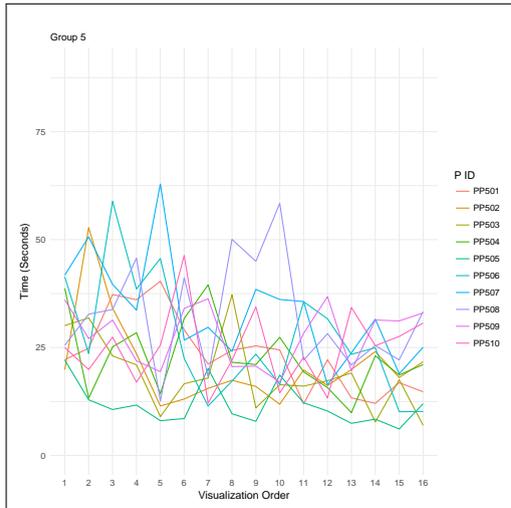


(c) Task completion time for 'Infer' task for the participants in group 3

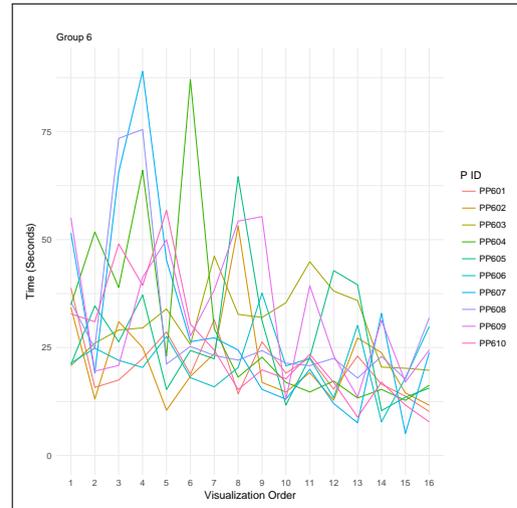


(d) Task completion time for 'Infer' task for the participants in group 4

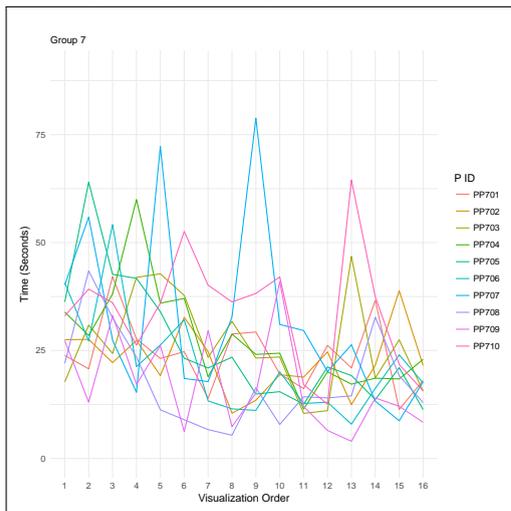
Figure 32: Task completion time for 'Infer' task for the participants in the groups (1, 2, 3 & 4). The Y-axis shows the time in seconds, and the X-axis shows the order of the presented visualizations (as each participant performs each task on 16 visualizations (8 designs with two datasets)).



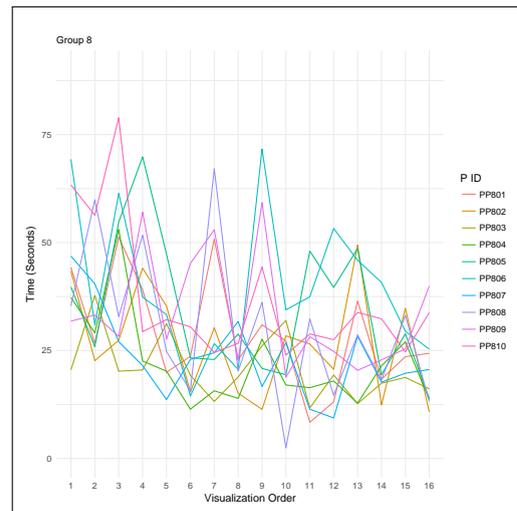
(a) Task completion time for 'Infer' task for the participants in group 5



(b) Task completion time for 'Infer' task for the participants in group 6

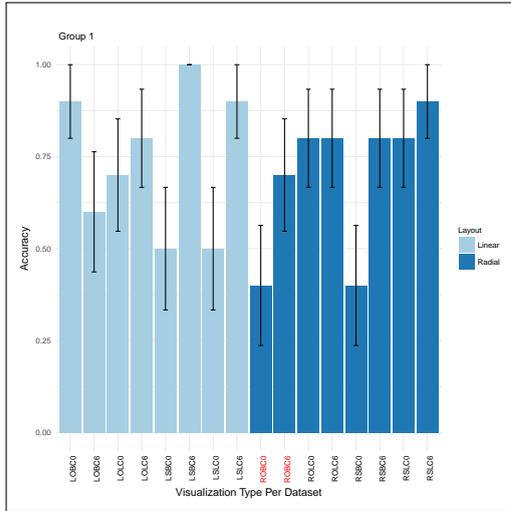


(c) Task completion time for 'Infer' task for the participants in group 7

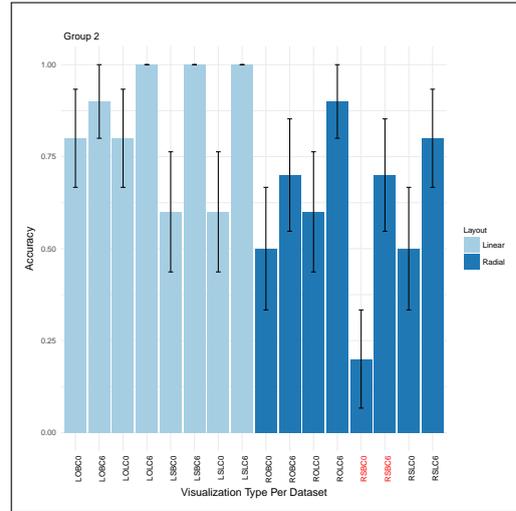


(d) Task completion time for 'Infer' task for the participants in group 8

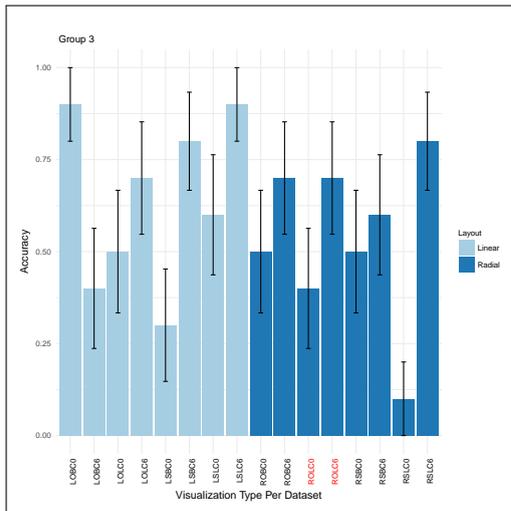
Figure 33: Task completion time for 'Infer' task for the participants in the groups (5, 6, 7 & 8). The Y-axis shows the time in seconds, and the X-axis shows the order of the presented visualizations (as each participant performs each task on 16 visualizations (8 designs with two datasets)).



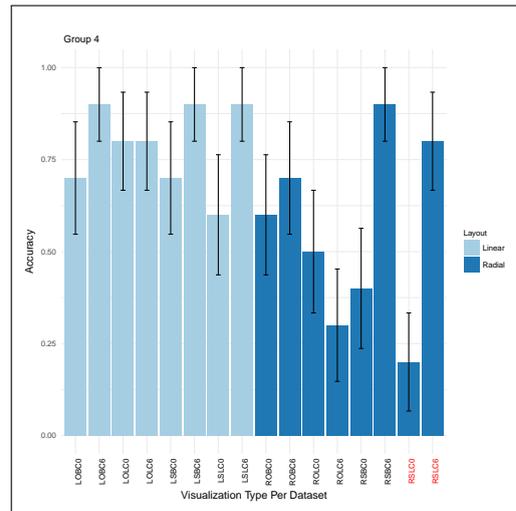
(a) Task Accuracy for 'Infer' Task (Group 1)



(b) Task Accuracy for 'Infer' Task (Group 2)

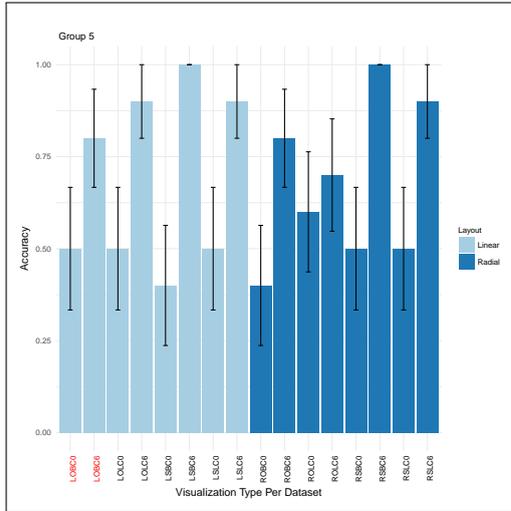


(c) Task Accuracy for 'Infer' Task (Group 3)

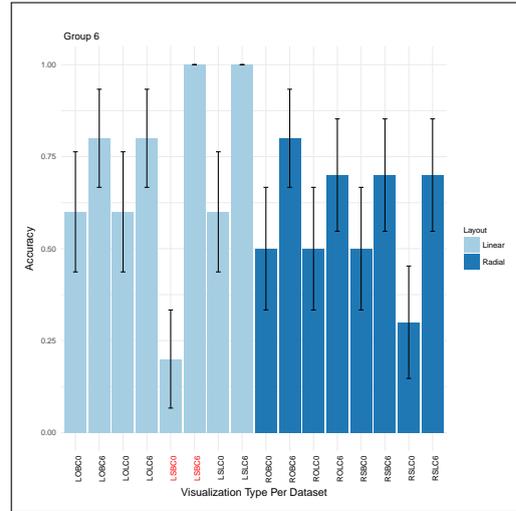


(d) Task Accuracy for 'Infer' Task (Group 4)

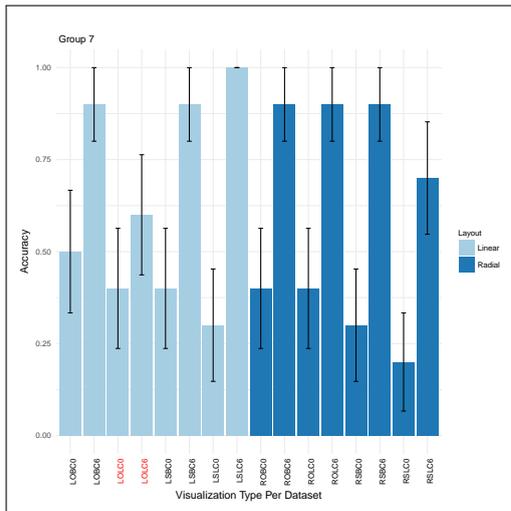
Figure 34: Task Accuracy for 'Infer' Task on the Eight Visualization Designs Per Dataset (for the Groups 1, 2, 3 & 4). The graph shows the average accuracy for 'Infer' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



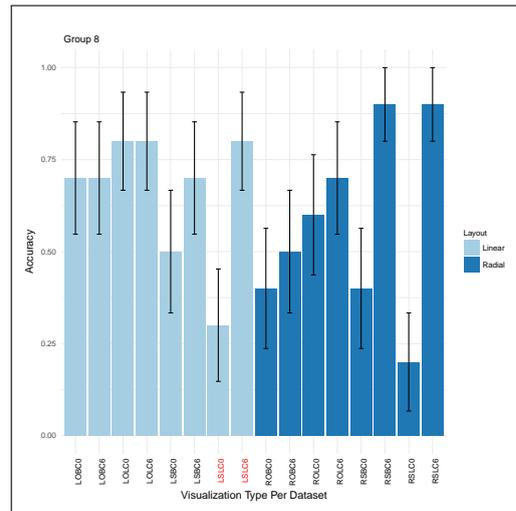
(a) Task Accuracy for 'Infer' Task (Group 5)



(b) Task Accuracy for 'Infer' Task (Group 6)

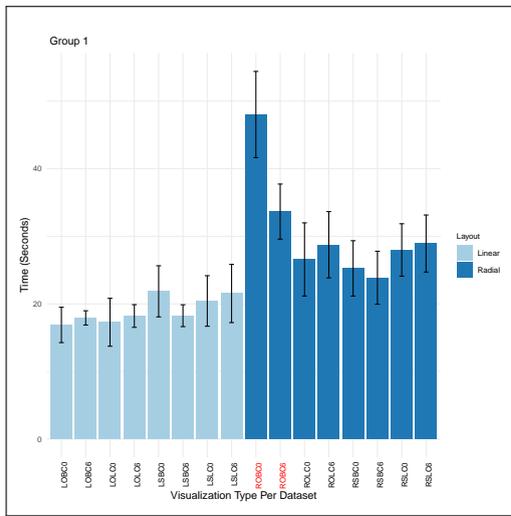


(c) Task Accuracy for 'Infer' Task (Group 7)

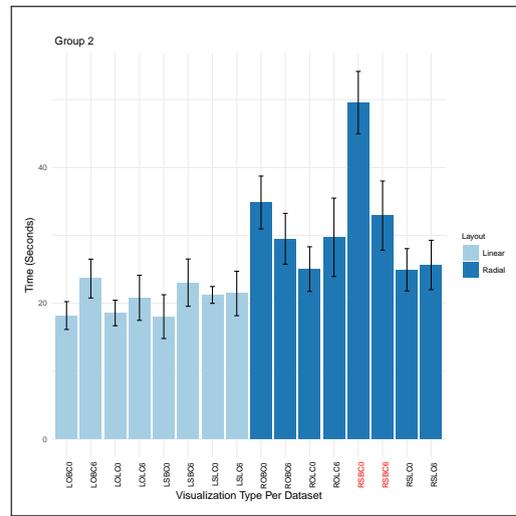


(d) Task Accuracy for 'Infer' Task (Group 8)

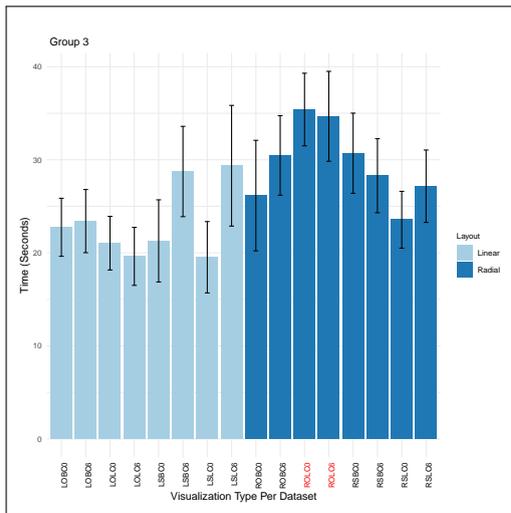
Figure 35: Task Accuracy for 'Infer' Task on the Eight Visualization Designs Per Dataset (for the Groups 5, 6, 7 & 8). The graph shows the average accuracy for 'Infer' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



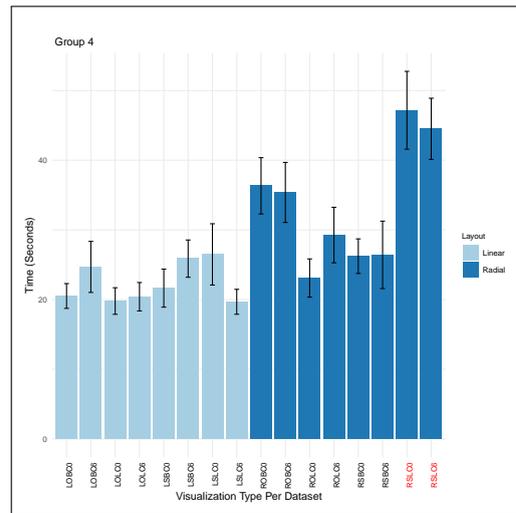
(a) Task Completion Time for 'Infer' Task (Group 1)



(b) Task Completion Time for 'Infer' Task (Group 2)

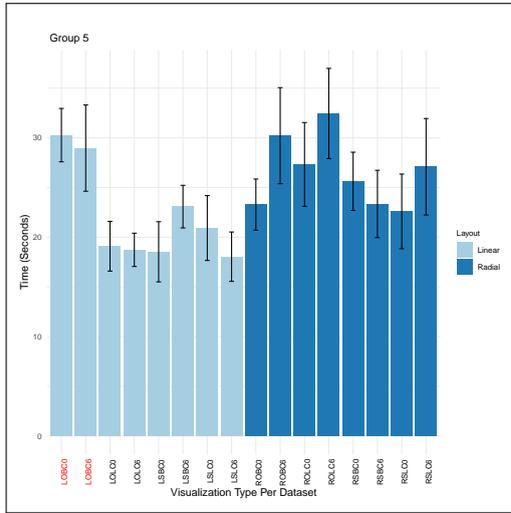


(c) Task Completion Time for 'Infer' Task (Group 3)

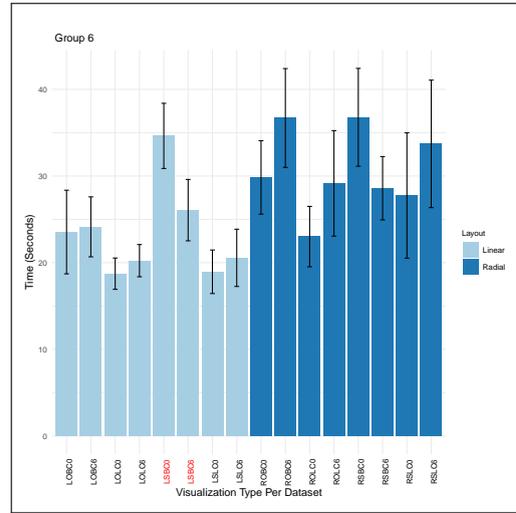


(d) Task Completion Time for 'Infer' Task (Group 4)

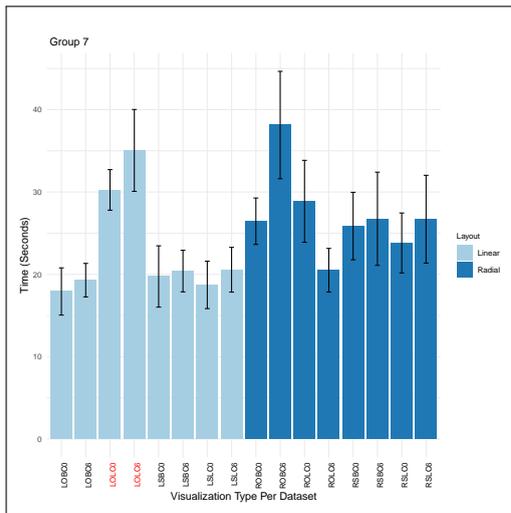
Figure 36: Task Completion Time for 'Infer' Task on the Eight Visualization Designs Per Dataset (for the Groups 1, 2, 3 & 4). The graph shows the average time for completing the 'Infer' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



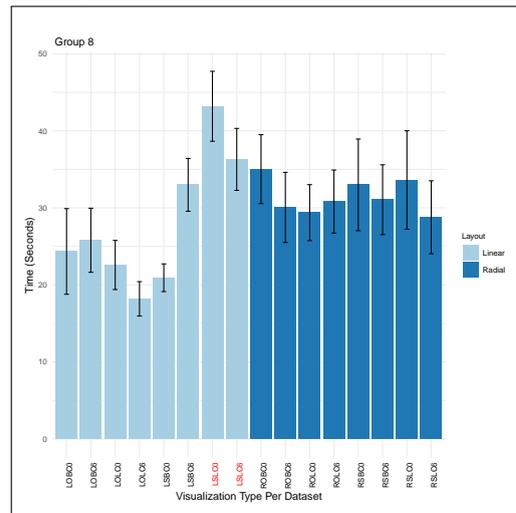
(a) Task Completion Time for 'Infer' Task (Group 5)



(b) Task Completion Time for 'Infer' Task (Group 6)

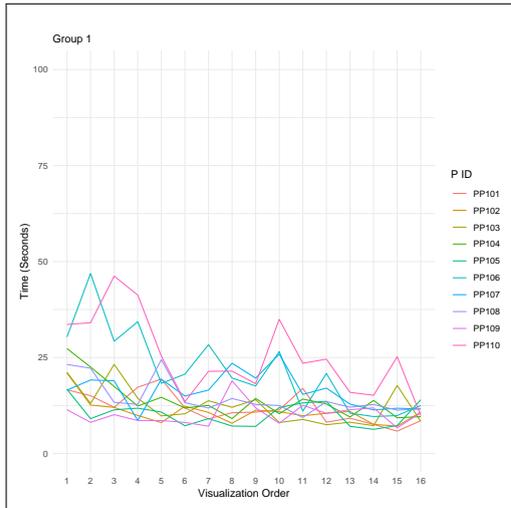


(c) Task Completion Time for 'Infer' Task (Group 7)

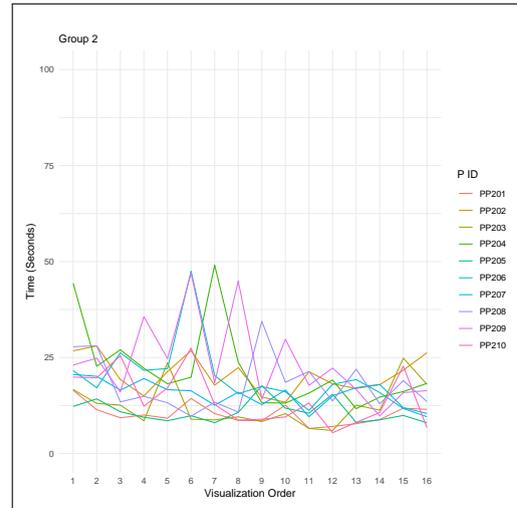


(d) Task Completion Time for 'Infer' Task (Group 8)

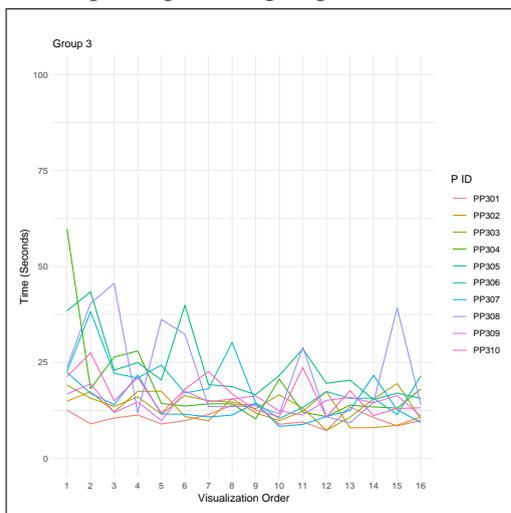
Figure 37: Task Completion Time for 'Infer' Task on the Eight Visualization Designs Per Dataset (for the Groups 5, 6, 7 & 8). The graph shows the average time for completing the 'Infer' task and the standard errors. The numbers '0' and '6' after each visualization name refer to the two different datasets. Visualization names with a red font (in x labels) refer to the first viewed visualization for each group.



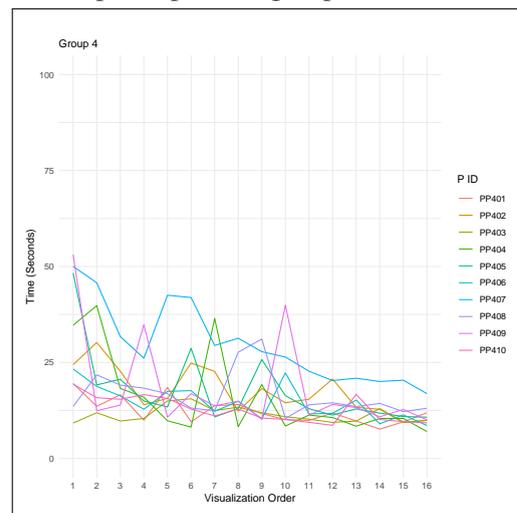
(a) Task completion time for 'Locate' task for the participants in group 1



(b) Task completion time for 'Locate' task for the participants in group 2

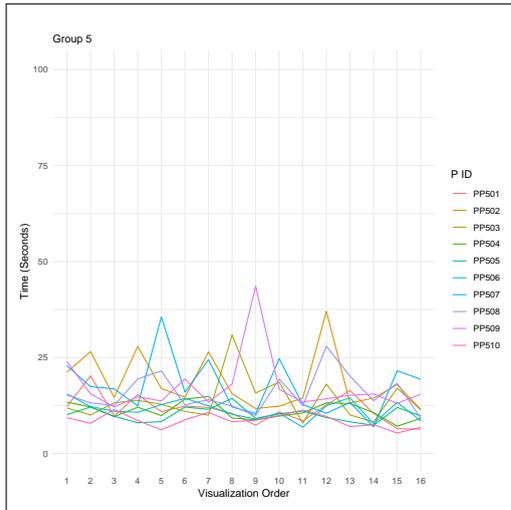


(c) Task completion time for 'Locate' task for the participants in group 3

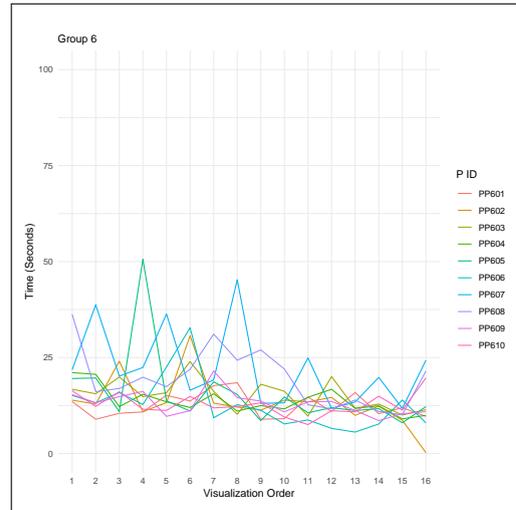


(d) Task completion time for 'Locate' task for the participants in group 4

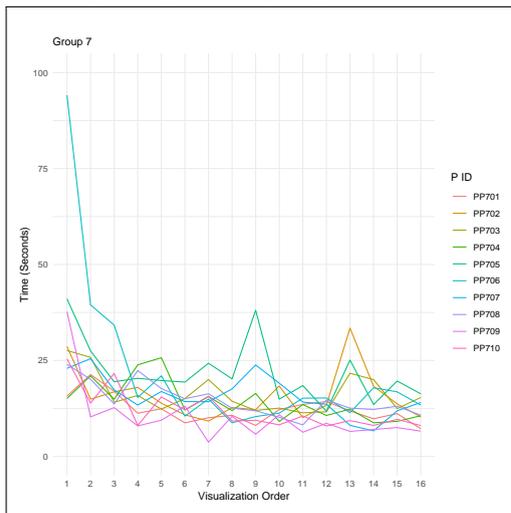
Figure 38: Task completion time for 'Locate' task for the participants in the groups (1, 2, 3 & 4). The Y-axis shows the time in seconds, and the X-axis shows the order of the presented visualizations (as each participant performs each task on 16 visualizations (8 designs with two datasets)).



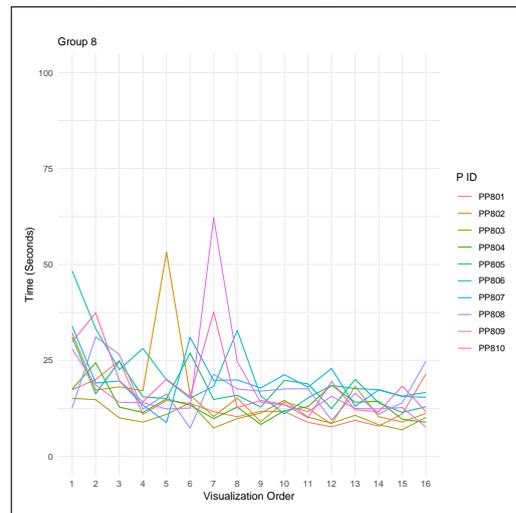
(a) Task completion time for 'Locate' task for the participants in group 5



(b) Task completion time for 'Locate' task for the participants in group 6



(c) Task completion time for 'Locate' task for the participants in group 7



(d) Task completion time for 'Locate' task for the participants in group 8

Figure 39: Task completion time for 'Locate' task for the participants in the groups (5, 6, 7 & 8). The Y-axis shows the time in seconds, and the X-axis shows the order of the presented visualizations (as each participant performs each task on 16 visualizations (8 designs with two datasets)).