Conquering the Void: Transforming Data into Action

by

Amber Paige Osborne

2094920

Dr Katarzyna Stawarz, Supervisor

Dr Argenis Ramirez Gomez, Co-supervisor

A dissertation submitted as per

the requirements for the

Degree of Masters of Science

in Computing.

Word Count: 18007

The School of Computer Science and Informatics

Cardiff University

October 7, 2021

Acknowledgments

I would like to thank my ever dedicated supervisory duo: Kathy Stawarz and Argenis Ramirez Gomez for their unwavering support and their reminders for me to "stop overthinking!". Thank you to all my participants, without whom this research would not be possible. Thank you also for the support of my cohort, friends and family, this project would have been impossible without your love and compassion. A special thanks to Stefan, for believing in me before I believed in myself and ensuring I got much needed fresh air throughout this process.

Abstract

Data collection and the technologies that facilitate it are incredibly popular within sport and fitness. However, there is a void between data collection and action. The problem lies in the fact that users currently collect data, but do not use it effectively to change their behaviours. Existing research on personal informatics, behavioural change and data visualisation theory focuses on data collection and analysis for reflective visualisations. However, the design of visualisations and the impact of predictive visualisations on behaviour change tends to be overlooked. Through an iterative design process using feedback from two user studies, this project has designed and developed a predictive visualisation to examine the usefulness of such a concept in supporting behaviour change, bridging the gap from data collection to action. This project resulted in the positive reception of a predictive bubble plot visualisation that recommends an exercise plan based on their historic data. The resulting design recommendations for data visualisations were produced as a result: gain user trust by producing a transparent, easily understandable visualisation, allow high levels of flexibility, focus on aesthetics, and utilise interactivity. These guidelines can be used to create meaningful data visualisations that users can engage with to make the most of their data and bridge the gap from data collection to action.

Contents

Acknowledgments **i**

Abstract **ii**

1 Introduction **1**

1.1 Aims & Objectives 3

1.2 Chapter Outline 5

2 Literature Review **7**

2.1 What is Personal Informatics? 7

2.2 Behavioural Change Principles 11

2.3 Data Visualisation Principles 14

2.4 Studies with Purposeful Visualisation Design 16

2.5 Commercial Platforms 17

2.6 The Problem to be Addressed 20

3 Methodology **22**

3.1 Project Timeline 22

3.2 Online Survey 24

3.3 Interviews 25

3.4 Ethics 26

4 Study One: Survey **27**

4.1 Methods 27

4.1.1 Participants 27

4.1.2 Procedures 28

4.1.3 Materials 29

4.1.4 Analysis 29

4.1.5 Survey Design 29

4.2 Findings 32

4.2.1 Questions on Visualisation One: Training Calendar 32

4.2.2 Questions on Visualisation Two: Training Log 34

4.2.3 Questions on Visualisation Three: Fitness & Freshness 35

4.2.4 Questions on Use of Activity Data 37

4.3 Summary 38

5 Designing & Developing Visualisations **39**

5.1 Requirements 39

5.2 Visualisation Design Process 40

5.3 Visualisation Development Process 48

6 Study Two: Interviews **55**

6.1 Methods 55

6.1.1 Participants 55

6.1.2 Procedures 56

6.1.3 Materials 57

6.1.4 Analysis 57

6.1.5 Pilot Study 58

6.2 Findings 59

6.2.1 Participants' profile 59
6.2.2 Theme: Visualisation for Reflection 60
6.2.3 Theme: Data Lead Development 60
6.2.4 Measuring the Visualisations Interpretability 61
6.2.5 Theme: Aesthetic Design 62
6.2.6 Theme: Motivation 63
6.2.7 Theme: Flexibility 64
6.2.8 Theme: Additional Requirements 64
6.3 Summary 65

7 Discussion **66**
7.1 The Studies 66
7.2 Contextualising the Results 67
7.2.1 Using Visualisations to Support Behavioural Change 67
7.2.2 Ensure Users Trust the Visualisation 68
7.2.3 The Impact of Aesthetic Design 69
7.2.4 Comparing the needs of Ex-Professional and Novice Users 70
7.3 Design Recommendations 71
7.3.1 Trust and Integrity 72
7.3.2 Flexibility 72
7.3.3 Aesthetics 73
7.3.4 Feedback Loop 73
7.3.5 Design Recommendations in Practice 74
7.4 Limitations & Future Work 75

8 Conclusions **77**

9 Reflection **79**

Bibliography **83**

Appendix A Survey Participation Information Sheet **93**

Appendix B Survey Participation Consent Form **97**

Appendix C Interview Participation Information Sheet **99**

Appendix D Interview Participation Consent Form **104**

List of Figures

2.1 The Stage-Based Model of Personal Informatics [30] 10

2.2 Garmin Connect Dashboard Example [4]. 18

2.3 Strava's Fitness & Freshness plot [16]. 19

3.1 A Section of the Gantt Chart 23

3.2 Timeline of Specific Tasks 24

4.1 Participant age-bands. 28

4.2 The Training Calendar Visualisation. 30

4.3 The Training Log Visualisation. 31

4.4 The Fitness & Freshness Visualisation. 31

4.5 Results from Thematic Analysis from Visualisation One. 33

4.6 Results from Thematic Analysis from Visualisation Two. 34

4.7 Results from Thematic Analysis from Visualisation Three. 36

4.8 Likert Scale on Participant Use of Activity Data. 37

5.1 Examples of Design Sketches. 41

5.2 Design Development: Table. 42

5.3 Design Development: Stacked Bar Chart. 44

5.4 Design Development: Scatter Plot. 45

5.5 Design Development: Bubble Plot Features. 46

5.6 Design Development: Pilot Study Visualisation. 47
5.7 Design Development: Overlapping Bubbles. 48
5.8 Design Development: Visualisation to be Implemented. 49
5.9 Development Flow Chart 51
5.10 Calculating Counts and Duration. 53
5.11 Implementing a random sampling technique 54

6.1 Sports of Participants. 59

7.1 Iteration of design incorporating attainment feedback and building to goals for novice users. 74
7.2 Iteration of design incorporating specific sessions, races and rest days for experienced users. 75

Chapter 1

Introduction

Data has become a powerful commodity, highly valued within society [33]. Simultaneously, use of self-tracking technologies is ubiquitous, particularly within a health and fitness context [19]. From general fitness watches, to sport specific devices and smartphones, we are constantly generating data. As data collection has increased, as has the demand for methods of collating and visualising the data for users [15]. The phenomenon of self-tracking and personal informatics examines the personal growth aspired to in the act of self-monitoring [42]. The quantified self is the pinnacle of how data collection can lead to insight. However, there is a gap between data collection and action [43].

Copious research has been undertaken into the best methods of data presentation to users. Behavioural change principles such as how the concepts of persuasion [22], personalisation [3] and reinforcement [34] can help engage a user and encourage action. Data visualisation theory on aesthetic design, simplicity versus complexity [8] and interaction [10] are key frameworks for creating the perfect design. There are also many commercial platforms available that aim to encapsulate these concepts, such as Strava [48] or Garmin Connect [24].

However, despite the availability of data and research into the behaviours of users and the developments of data analysis and visualisation there is a problem. The average user of a

smartwatch or tracking application stops using it after as little as six months and throughout that period may fail to action any of the feedback they are gathering from their data [38].

Existing research empathises the importance of visualisation design, indicating that current methods are ineffective in bridging the gap between data collection and action [29]. Users are collecting data but they are not effectively using it for personal growth and behaviour change. To counter the dominance of reflective visualisations, this project will investigate the impact of predictive visualisations to understand whether they are more effective in facilitating behavioural change. The aim is to bridge the gap from data collection to action.

The problem will be addressed by reviewing existing research in behaviour change, visualisation theory and existing commercial platforms; using this as a foundation for a series of user studies that will be conducted as part of an iterative design process. Participants will provide feedback on visualisation design as well as being used to gather information on requirements, how they use visualisations and their current behaviours. Due to the length of this project, it is impossible to measure behavioural change as a result of the predictive visualisations. Therefore, evaluation of success will be determined by indications of behavioural change taken from participants in a semi-structured interview.

The online service and mobile application Strava will be the focus for this project as a source of data as well as a baseline for visualisation design. Strava is a data aggregation system for active users where data is shared and visualised to increase user performance. It is the leading and fastest growing platform [15] for self-tracking exercise and offers compatibility with a broad range of sports targeting all abilities. Due to its popularity and range of visualisation types Strava will be used for this project as a source of data as well as a baseline for visualisation design.

1.1 Aims & Objectives

The aim of this project is to bridge the gap between data collection and user action through predictive data visualisations that support behaviour change.

Objective One -

Select visualisations from Strava that support future planning and organisation. Initial choices are informed by my own experiences of Strava visualisations.

Objective Two -

Establish effective methodology. Structuring studies for iterative design, recruitment, materials, data collection and ethics.

Objective Three -

Study one, survey current Strava user's attitude to the visualisations selected. Ascertain uses, benefits and criticisms across functionality as well as design. This will begin the iterative design process for developing the predictive visualisation, establishing functionality requirements, aesthetic elements and discussion points for further studies.

Objective Four -

Evaluate existing research across:

- Personal informatics.
- Behavioural change principles.
- Studies with purposeful visualisation design.
- Data visualisation principles.

- Existing commercial platforms.

These will be used to inspire development alongside the feedback gathered in study one.

Objective Five -

Design a visualisation informed by my investigations that will be effective at engaging my participants. Use my own data to carry out the design process using the Strava application programming interface (API). Use a pilot study to evaluate effectiveness before involving participants.

Objective Six -

Collect data from users to create the predictive visualisation designed using findings from earlier objectives.

Objective Seven -

Present visualisations to participants in a second study, individual interviews. Hold the interviews with participants and evaluate visualisation's usability, establish current their current behaviours, question them on its design and its usefulness. Establish indications of longer term behavioural change as a result of the visualisation. Use feedback for further mock-up designs.

Objective Eight -

Create a set of design recommendations to aid future research and to guide others in the creation of predictive data visualisations that support behaviour change.

1.2 Chapter Outline

Chapter Two -

The next chapter will contextualise the project, covering relevant existing research such as personal informatics, behaviour change, data visualisation theory and existing commercial platforms. This will provide the relevant context to the project and highlight why further improvements should be made. The chapter will conclude by addressing why there are gaps in this research and how they can be filled by this project's contribution.

Chapter Three -

Chapter three will explain the methodology behind the project, exploring the approach undertaken for each component; from research methods, to data collection and the organisation of the studies. It will also cover the ethical considerations.

Chapter Four -

Chapter four will explore study one, the survey that will be used to gather baseline opinion from participants regarding the current offering of Strava visualisations and establish requirements for the design process. This chapter will explain the methodology and procedure for the study, as well as the results.

Chapter Five -

Chapter five will cover the design and development process of the visualisation. Beginning with the requirements identified from study one and then explaining the iterative design process and eventual implementation of the final visualisation design.

Chapter Six -

Chapter six will explore study two, the semi-structured individual interviews where partic-

ipants will be shown the visualisation and asked to evaluate its usefulness and gain insight on participant behaviours. It will cover the methods and results from these interviews.

Chapter Seven -

Chapter seven will first section will summarise the studies and examine the results within the context of existing research. The design recommendations will be presented and further mock up iterations will be explored as a result of my research. It will conclude with the project's limitations and future work.

Chapter Eight -

The penultimate chapter will conclude the project, outlining the project as a whole before adding final remarks.

Chapter Nine -

The final chapter will cover a personal reflection on the research process, addressing the important learning and self-development that this project has allowed me to achieve and how I can apply what I have learnt to future work.

Chapter 2

Literature Review

This chapter will explore existing research within the fields related to this project and existing commercial platforms. This review will provide clarity on why this research is necessary. My research topic centres on the impact of predictive data visualisations on behaviour change in active users. The first sections will explore the concept of personal informatics as a vital concept within this project. I will go onto explore behaviour change theory, visualisation concepts and studies with purposeful visualisation design in order to use existing research as a foundation for development within this project. This chapter will also review current commercially available platforms, specifically the application Strava, explaining what it is and why it was selected for design inspiration.

2.1 What is Personal Informatics?

A key concept for this project is self-tracking, especially the act of collecting data and gathering insight from it. A key problem is discovering why users stop tracking themselves or do not use the data collected to gain a greater insight into themselves. This is significant as it represents a gap between data collection and action that users struggle to overcome.

Personal informatics is the act of collecting personally relevant data about oneself, for the

purpose of self-reflection and development [42]. Users self-track to learn more about themselves, for example identifying their behaviours, thought patterns or habits. The emphasis of this field is the concept of " *the self*" to draw insight, and this form of self-surveillance is used as the human memory is often flawed by a lack of capacity or being unable to collect data at certain times, for example during sleep [31]. This overlaps with the "Quantified Self", a phrase coined by Gary Wolf and Kevin Kelly [58], which refers to self-tracking behaviours for "self-development" or "optimisation" . Personal informatics and the quantified self both share the idea of "the self" and self-reflection, and are often utilised by users for the sake of self-improvement or development. For example, a user may collect data while running, tracking distance and time. They can then use this data in the future to run further or faster.

Self-tracking is a highly effective form of motivation for participants when it comes to maintaining a healthy level of physical activity and avoiding a sedentary lifestyle [23]. Reflection can assist in motivation, but there are other factors. Researching the correct methods of motivating users is key to helping them use their data for self-improvement. Other motivational factors include organisation and the scheduling of regular exercise as part of a user's daily routine [29]. Lazar et al writes, "routines serve as factors contributing to motivation" [29]. A key lesson from their study was that the driving force of self-tracking and evaluation comes from motivating and engaging users long enough for routines to form. The important factor is forming routines before users cease to engage. Rawassizadeh et al. observes that despite the market success of smart fitness devices, many users cease to use them after six months [43]. Without the motivation provided by such devices, many users cease regular exercise altogether. The study claims device and sport abandonment is due to an over-emphasis on data collection within tracking devices and a lack of compelling data visualisations for end users [43].

The behaviour of self-tracking for betterment is not exclusive to technology, with personal

diaries being an example of a handwritten equivalent. Analogue tracking is still popular, with trends in hand-crafting items, such as scrap books and mood diaries, on the rise [51]. Regarding physical activity, technology has made self-tracking more prevalent than ever. Data collected from wearable sports devices means that data aggregation services like Garmin Connect or Strava are incredibly popular, and technology based tracking is the exclusive choice for self-tracking physical activity. The role of commercial services will be explored later in this chapter.

A key framework in self-tracking and understanding the journey users take when tracking is the Stage-Based Model of Personal Informatics Systems formulated by Li et al. [30]. Their study involved a selection of participants that engaged in some form of self-tracking behaviour, from sports tracking to sleep patterns. Through user studies they analysed how the participants decided to collect data, investigated how the data is visualised and how users interacted with and used the visualisations. As a result they concluded that throughout a self-tracking journey, users undertake the following five stages: preparation, collection, integration, reflection, and action, see Figure 2.1. This introduces the importance of data visualisations within the self-tracking process.

- **Preparation** - Consciously deciding to begin self-tracking and planning how this will occur.
- **Collection** - Actively collecting data through a chosen method.
- **Integration** - Data processing and presenting the data in a way that it can be reasonably be interpreted and understood. This is where the data is visualised. The average user will not do this manually. They will instigate the process through synchronising their device to a web based platform that will visualise on their behalf. The way in which data is visualised is vital in ensuring an engaging reflective process.
- **Reflection** - A user reflects and explores the information collected. This can be a long-

term or short-term process. A user may reflect immediately or look back to something they did a year ago. The quality of the visualisation, in terms of engaging the user and displaying useful information, will impact how successful this process is.

- **Action** - The user acts on the information that they have explored and incorporates what they have learnt into future activity. This process will only occur if the user learns from the visualisation, establishing a positive feedback loop.

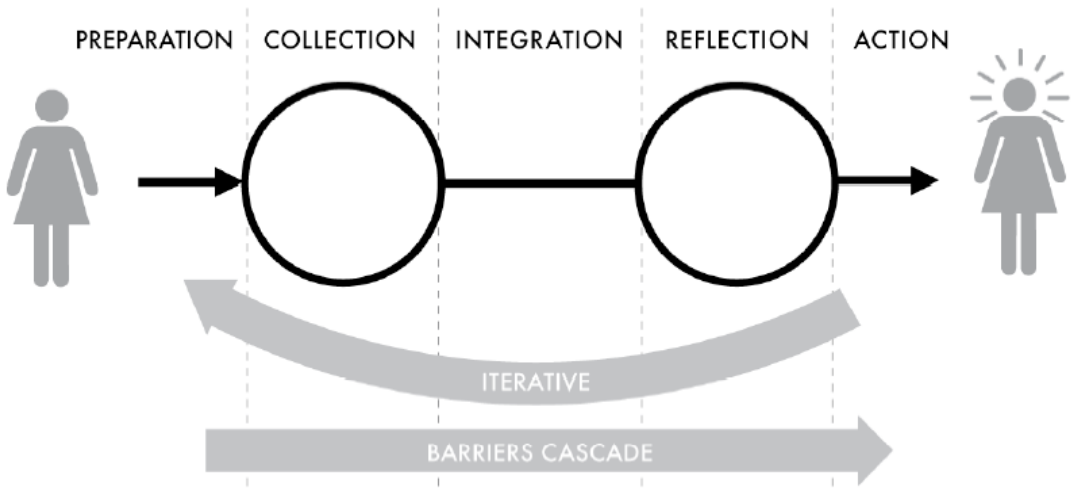


Figure 2.1: The Stage-Based Model of Personal Informatics [30]

The first two stages: preparation and collection, refer to the context leading up to the decision to collect data and the collection of data itself. Participants within my own study will have already undertaken these processes and the focus will be on the latter three stages: integration, reflection and action. Current research indicates users are not acting on their insights from the data visualisations which breaks a link in the feedback loop. [43]. As a result, this project questions why that is the case.

The reflection stage occurs as a direct result of the integration stage, linking a lack of engagement to ineffective methods of data visualisation. Li et al. recognised that within their model barriers to actioning self-tracked data cascade from one part of the model to the

next. Problems impacting earlier stages have consequences on latter stages, and therefore they undertook a holistic user-driven design process for the system as a whole to increase the likelihood of action [30]. One issue raised within the action stage was that "most systems do not have specific suggestions on what to do next, which is a barrier to understanding" [30]. Therefore, rather than looking at a system as a whole, this project will expand on this factor, examining the integration and action stages in isolation, experimenting with the impact of the visualisation design on promoting action.

2.2 Behavioural Change Principles

Investigations into related research have highlighted the importance of incorporating recognised behavioural change strategies into systems or visualisations to aid resonance with users. Pre-existing work indicates behaviour change is more likely if systems or visualisations have been developed with these concepts in mind [3].

A key distinction to highlight is the difference between reflective technologies and persuasive technologies. Many visualisations on platforms such as Strava, are entirely reflective. Fritz et al. explores the use of different motivating technologies on users when it comes to changing their behaviours [22]. The study provided groups with two types of pedometer, one that displays the number of steps and the other that displays the number of steps alongside how many kilometres those steps equate to. Pedometer one is a reflective technology, providing data which a user may find interesting but is uninformative as few users are able to identify how many steps equal one kilometre. The second pedometer is a persuasive technology as it measures distance as well as steps which, according to Fritz et al., is more likely to bridge the gap from reflection to action as the data presented is more concrete [22]. Adding universal metrics, such as kilometres rather than steps, also aids motivation across multiple sports by encouraging a user to more actively connect their activities together.

Research conducted by Fogg in the 1990s, explored the concept of persuasive technologies

describing them as "computing systems, devices or applications intentionally designed to change a person's attitude or behaviour in a certain way" [20]. His work links the act of persuasion directly with behaviour change and emphasises the importance of system design to enact change. A persuasive technique often utilised is the construct of a social community. For example, within Strava a user may be encouraged to exercise through the social media-like functionality of "kudos" or likes. Virtual support acts as a form of positive reinforcement [54]. However, social comparison can also be anxiety inducing and according to Aldenaini et al., identifying the personality type of users is vital in identifying an effective method of reinforcement [3].

Tailoring visualisations uniquely to users however, also has limitations. Several studies [50, 3] have found that overly specific visualisations can have an adverse impact on behaviour change due to fear and suspicion around the technology used to create them. Aldenaini et al. argues there is "only a tiny difference between encouragement and persuasions on one side and coercion and deception on the other" [3]. When accessing data to produce visualisastions for a user to provoke behaviour change, the method behind the process must be transparent and not in any way remove or impact the users' autonomy [37, 13].

The "Elaboration Likelihood Model (ELM) of persuasion" [40] is a key model from the 1980s. This theory works off the basis of two routes, the "central" and "peripheral" pathways. The central route is defined by a user's careful consideration of possible outcomes and making an informed decision, resulting in high cognitive consideration. The peripheral route has an absence of information, resulting in a user's reliance on feelings rather than facts, and as a result has a lower cognitive load [40]. This model relates to behavioural change design as studies have used these principles to examine the different interpretations of data. For example, Pandey et al. suggests that news articles that use statistical analysis are more believable than those without [37]. Readers are provided with concrete information to evaluate and can use it to form an interpretation, finding the information more believable

[37]. This is significant as visualisations with concrete and a believable representation of a user's data will be more impactful and more likely to influence their behaviour as a result.

Alongside persuasive technologies, an important factor is motivation and encouragement. Michie et al.'s COM-B model states that the success of a persuasive system for behaviour change relies on three factors: capability, opportunity and motivation [34]. Motivation is the desire to achieve a goal and in order to "open the gates" for motivation, the desired behaviour must be facilitated by the belief that the user is capable and there must be the relevant opportunity to carry out the behaviour. Behaviours can change after a period of repeated activity, becoming more automatic the longer they are performed [28]. Setting realistic and attainable goals can also be a challenge for users, which is where predictive visualisations may be beneficial.

Reinforcement through visualisations that aid planning may be a method of encouraging routine formation, as these visualisations act as a form of prompt. Humans are more likely to engage in new behaviours when the resources to do so are plentiful and the effort to perform the behaviour is low [28, 34]. Visualisations that are part of the environment remove barriers to participation in sporting activities, such as time management or inexperience.

Motivation also aids behavioural change with research suggesting two types of motivation: intrinsic and extrinsic. Extrinsic behaviour change relates to an external demand or social pressure. Intrinsic motivation is stronger as it relates to a personal values and needs [28]. Participants within this study are all pre-existing active users that have joined the study to gather insight, and therefore are more likely to possess the internal motivation that makes them susceptible to longer-term change.

Using an awareness of persuasive technology frameworks and the concepts of motivation and reinforcement are key to a visualisation that enacts change.

2.3 Data Visualisation Principles

Behavioural change principles can aid the content design, but visualisation principles define the aesthetics. Users require effective visualisations to understand their raw data and learn from it. Card et al. writes that "information visualisation", or data visualisation is "the use of computer supported interactive visual representations of abstract data to amplify cognition" [11]. Here, data visualisations are fundamentally linked to provoking thought.

There are many "rules" that different scholars will argue constitute to an effective visualisation, but visualisations are highly subjective. Edward Tufte's rules are foundational in the establishment of "effective" visualisations. One rule is the concept of "Graphical Excellence" [56]. This is achieved by:

- Complex ideas being presented efficiently and simply.
- Adhering to a good Data:Ink ratio. This relates to presenting the most ideas with the least amount of 'ink' within the smallest amount of space.
- The avoidance of "chart junk".
- Telling the truth and not distorting the data. The Lie Factor measures the size of effect in data next to the size of effect within the visualisation.

Tufte also uses "Graphical Integrity" as a measure of success [56]:

- The importance of labelling.
- The data is contextualised.
- The dimensions of the visualisation match those of the data.

Tufte believes that a visualisation should be as accurate, simple and efficient as possible in order to be effective. Tufte defines chart junk as "excess" aesthetic elements which can

distract from the visualisation. This factor can be tied to his position as a mathematician, where aesthetics are not required for him to be engaged in the visualisation. This is not necessarily the case for a lay audience.

Tufte's rules focus on creating a "successful" visualisation and a key feature of success is engagement and memorability. Memorability of a visualisation is a significant factor to its usefulness, and a study by Borkin et al. suggests that the aestheticism of a visualisation impacts of the usability and memorability [8]. They argue that visualisations are more memorable when they look certain ways. Using dummy data, they experimented on the ways in which the same data could be presented in different ways and how that impacted its memorability [8]. Borkin et al. provides the following structure to enhance a visualisations notability [8]:

- High graphical quality.
- Familiarity for understanding, but unique qualities for memorability.
- Shapes can be used to trigger emotions.

Remembering and engaging with visualisations can be challenging due to the quantity of visualisations available. Current research suggests that users are overloaded with visualisations and are overwhelmed with the amount of data they have available [32]. The important factor is finding a way in which we can transform raw data into a manner that provides "unobtrusive" feedback to users [32]. Visualisations that are high in data density can be cluttered with features. A way of reducing the features available within visualisations is through interactivity. Bruggemann et al. investigated the impact of interactivity on the engagement of visualisations and found that visualisations should be designed with interactivity in mind, rather than adding interaction retrospectively [10]. Interactivity can be an effective method of encouraging cognitive stimulation within a visualisation, making it clearer, simpler and more memorable. This is significant in data sets with a high numbers of

features. The concept of feature reduction is the flattening of features shown within a visualisation through interactivity, which increases the user's understanding as it also encourages "exploratory analysis" [44].

User engagement is vital to enact long-lasting change. Visualisations must force participants to think in order to increase memorability. By engaging cognitively, there is an increased likelihood of longer-term commitment to a visualisations teachings [12].

2.4 Studies with Purposeful Visualisation Design

Copious research is devoted to the impact of wearable devices on users' exercise behaviours and health, but less so on the impact of the visualisations produced by those devices. Hallnas et al. researched the impact of "slow technology", creating visualisations that were designed to initiate reflection. These slow technologies juxtapose "fast technology" that they argue is designed to save time but not provoke insight [25]. Their study emphasised the importance of a visualisation not being "time consuming but time productive, using time for reflective activities" [25]. While not linked to physical activity their research emphasises the importance of meaningful visualisation design that is designed to encourage reflection not merely report data. The latter may be faster, but will not necessarily result in the visualised data being actioned. Their work involved the use of ambient technologies such as sounds or lighting to prompt events, emphasising the importance of aesthetics in design. Bice et al. investigated the use of pedometers in motivating physical activity and expressed the importance of "immediate visual feedback increasing activity awareness" [7]. Their study found that visual feedback is key to forming a feedback loop which may reinforce behaviours through self-evaluation. In addition, providing individuals with the ability to "measure progress for goal attainment" prompted action from their participants. Medical research by Conroy et al. highlights the importance of supporting intentions in exercise. They write "forming intentions is rarely sufficient for changing behaviour and further motivational support is

often needed for people to implement their intentions" [14]. This study evaluated the top-rated fitness tracking apps to investigate persuasive techniques within them and discovered that there are significant gaps in intention supporting visualisations; "Surprisingly, the most well-established technique for bridging the intention-behaviour-gap, action planning, was relatively rare" [14]. This highlights that there are gaps within visualisations that have been identified but are still missing from popular applications like Strava. My own visualisation designs will consider these key aspects from research.

2.5 Commercial Platforms

There are many existing platforms that host and display data visualisations for end users. Devices have their own platforms, for example Garmin has "Garmin Connect" [24] or TomTom's "MyRun" [53]. Third-party systems are also available such as TrainingPeaks [39] or Strava [48] where a range of devices can be used. Strava will be used as a foundation for this project due to its popularity and the range of devices it supports, resulting in the largest user base. Strava is the leading sports and exercise tracking platform, as well as the fastest growing [15]. It also offers the broadest range of visualisations. I also have experience with Strava and can use it to inform early designs as part of the iterative process. Garmin Connect is another popular option and will also be investigated due to the broad visualisation options it offers.

Garmin Connect has a range of planning based visualisations that could be adapted to a predictive style, see Figure 2.2. They have a broad range of calendar-style training schedules, that users can populate to help them plan their activities, as well as a variety of training programmes to schedule activity based off user-defined goals.

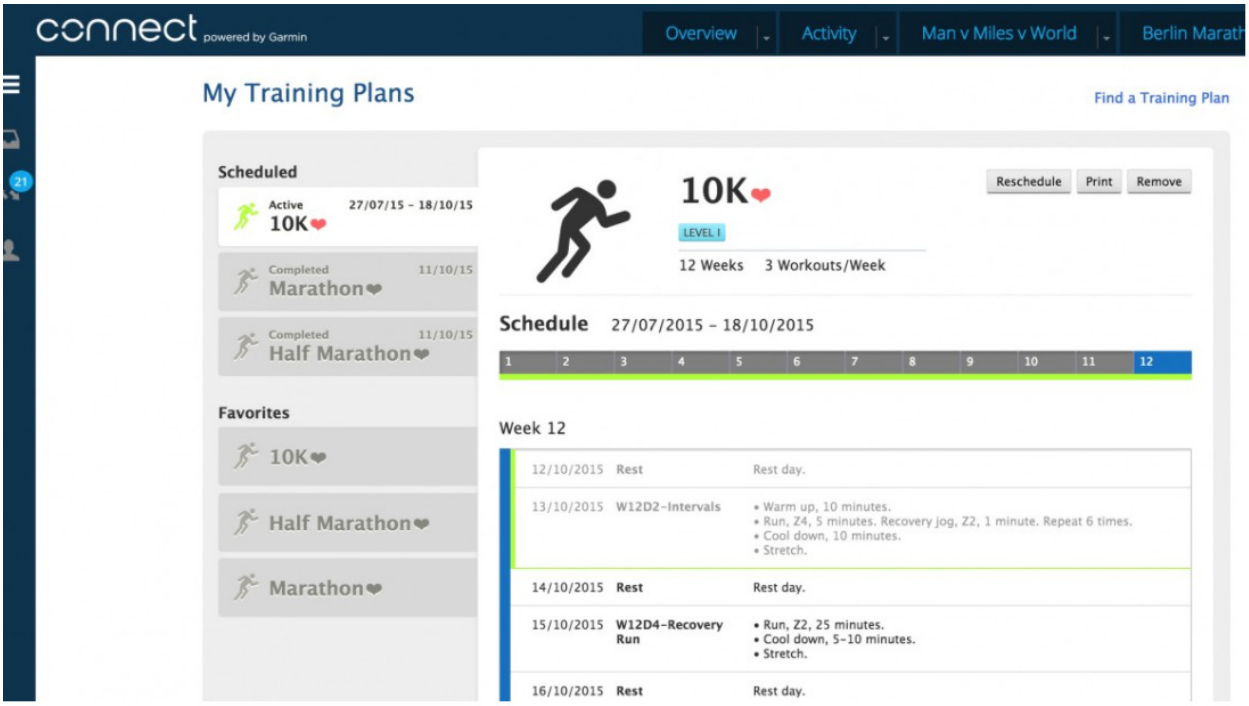


Figure 2.2: Garmin Connect Dashboard Example [4].

Strava offers a range of calendar style visualisations much like Garmin Connect. With a premium membership however, users can access the Fitness & Form visualisation which predicts when a user is at peak performance and other more professional looking visualisations, see Figure 2.3. Given the correct metrics (heart rate, power etc...) this visualisation is designed to use an athlete's relative effort in past activity and an "impact-response model to quantify its effect over time" [16].

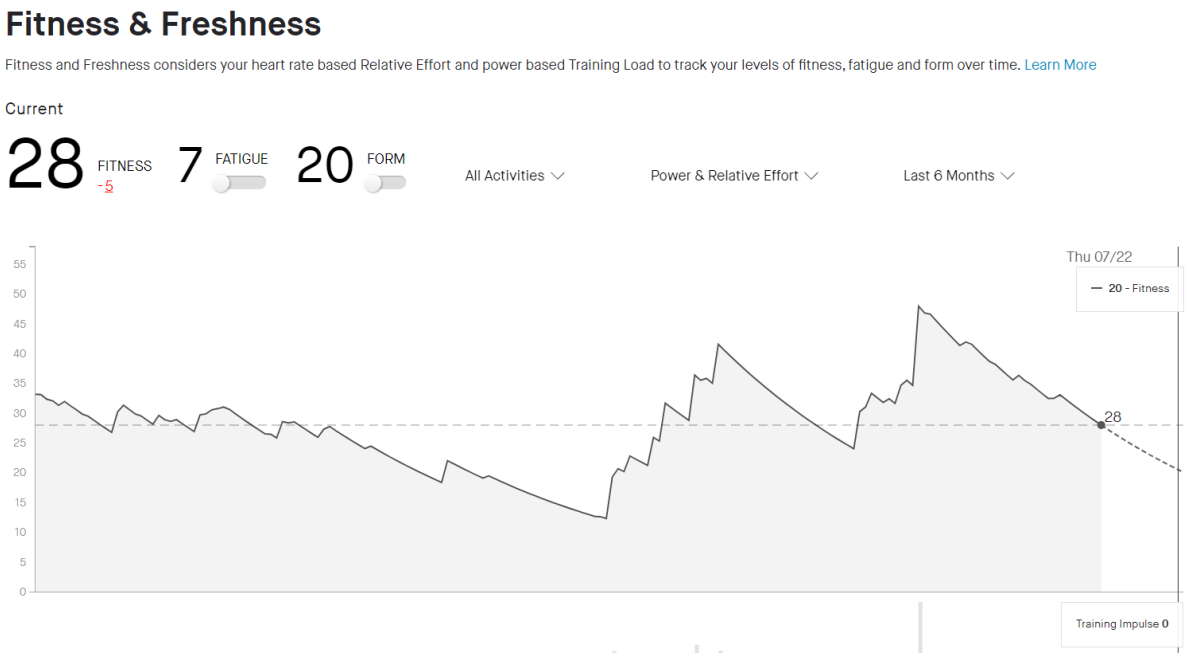


Figure 2.3: Strava's Fitness & Freshness plot [16].

These calculations are based off the 1975 works of Eric W. Banister which were later adapted for the professional cycling community by Dr. Andy Coggan [17]. Strava's use of an algorithm to calculate predictions of future performance are an example of prediction being used to alter behaviours and routines. Its incorporation within a popular application identifies an interest in using future performance to instigate behaviour change.

However, there are flaws in this visualisation, this visualisation relies upon indication of stress [46]. Stress can be measured by a range of measurements such as heart rate, power metrics or cadence. Not everyone has access to this data and without at least one of these things these calculations are not possible. Secondly, while Strava provides brief explanations of how these visualisations are calculated and how to interpret them, they are overly complex and likely to confuse the average user.

2.6 The Problem to be Addressed

A key gap in research is solving the issue of a lack of engagement and motivation to use collected data to action behavioural change. As addressed, currently many users collect data but they do not utilise the findings to help their future performance. Current research predominately focuses on the devices or platforms as a whole, rather than the production of insightful visualisations [43]. As a result this project will centre on using a predictive visualisation designed specifically to engage and motivate users into action.

Using the Stage-Based Model of Personal Informatics Systems [30] this project will focus on bridging the gap between the integration and action stages to encourage behaviour change and increase motivation. This will be facilitated by an in-depth exploration of visualisation design and by introducing a predictive element to it.

The primary benefit of this research will be allowing users to empower themselves through their data. By showing participants a new kind of visualisation and encouraging them to reflect on their motivations for exercise, they will be able to use the information for personal development. This project will also contribute to the broader movement of helping users take full advantage of technology to improve their health and well-being. It has been established that using wearable technology is beneficial to health and behaviour change, but there are gaps in user engagement which this project aims to assist. Existing engagement research centres on intervention, but prediction may be a new method of engagement without cues or interference, broadening the reach to users who do not find those methods beneficial.

A challenge to this research is creating a visualisation that is beneficial to all users and their personalities. Creating tailor-made visualisations for users is not an option, as already discussed [50] [3] and Strava's current approach of overwhelming users with visualisation options should also be avoided. To combat these issues, the visualisations designed will be informed by visualisation and behavioural change frameworks as well as feedback from current Strava users. The theory will result in producing a visualisation that engages and

motivates users by its design and paired with participant feedback over several iterations will ensure that the functionality and data shown is insightful. The next chapter will cover the methodology applied to carry out this work successfully.

Chapter 3

Methodology

This chapter will overview the methodology chosen to carry out this project and why it was selected. As a human-computer interaction (HCI) project, iterative design was at the heart of the process. This project was designed around a constant cycle of user feedback to ensure the visualisations met user requirements. From Nielsen's studies on designing usable interfaces he concludes "Because even the best usability experts cannot design perfect interfaces in a single attempt, interface designers should build a life-cycle around the concept of iteration" [36]. Iteration is the process of designing, demonstrating to participants, then using participant evaluation to improve the original design [36]. As a HCI project, it is essential design choices are underpinned by user feedback [3].

3.1 Project Timeline

The first priority was designing an approximate timeline of what was required throughout the project. A Gantt chart (see Figure 3.1) was utilised to manage the project effectively. This method is appropriate as it allows for convenient control of multiple tasks that occurred simultaneously and allowed the setting of key milestone goals, which was crucial for the success of the project [27]. In addition to the Gantt chart, I also created a timeline of specific

tasks, distilling the information within the Gantt chart, that were in need of completion sorted by date and type of activity (Figure 3.2). This combination of project management tools allowed for an overall schedule and more detailed notes on which each task involved.

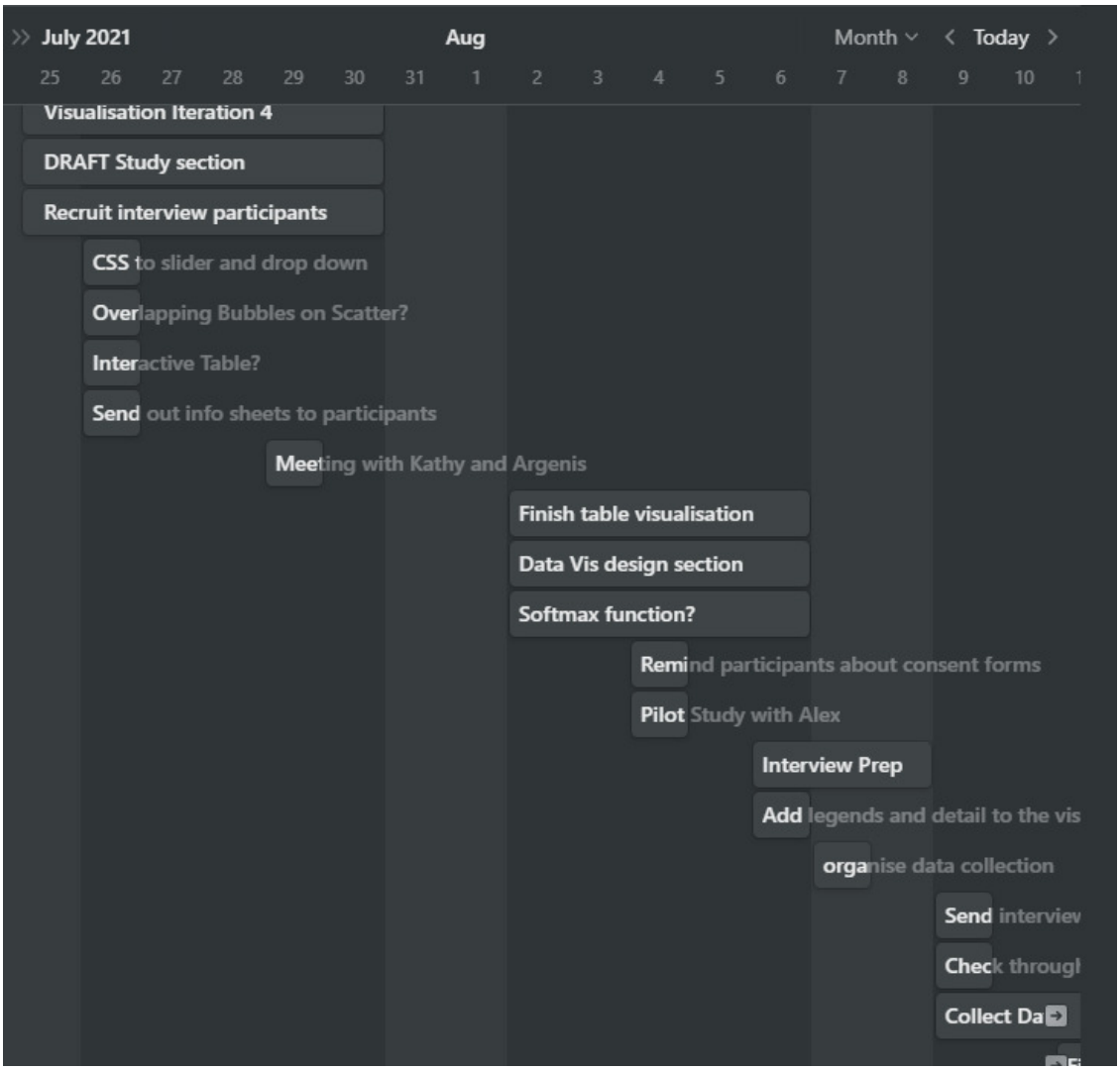
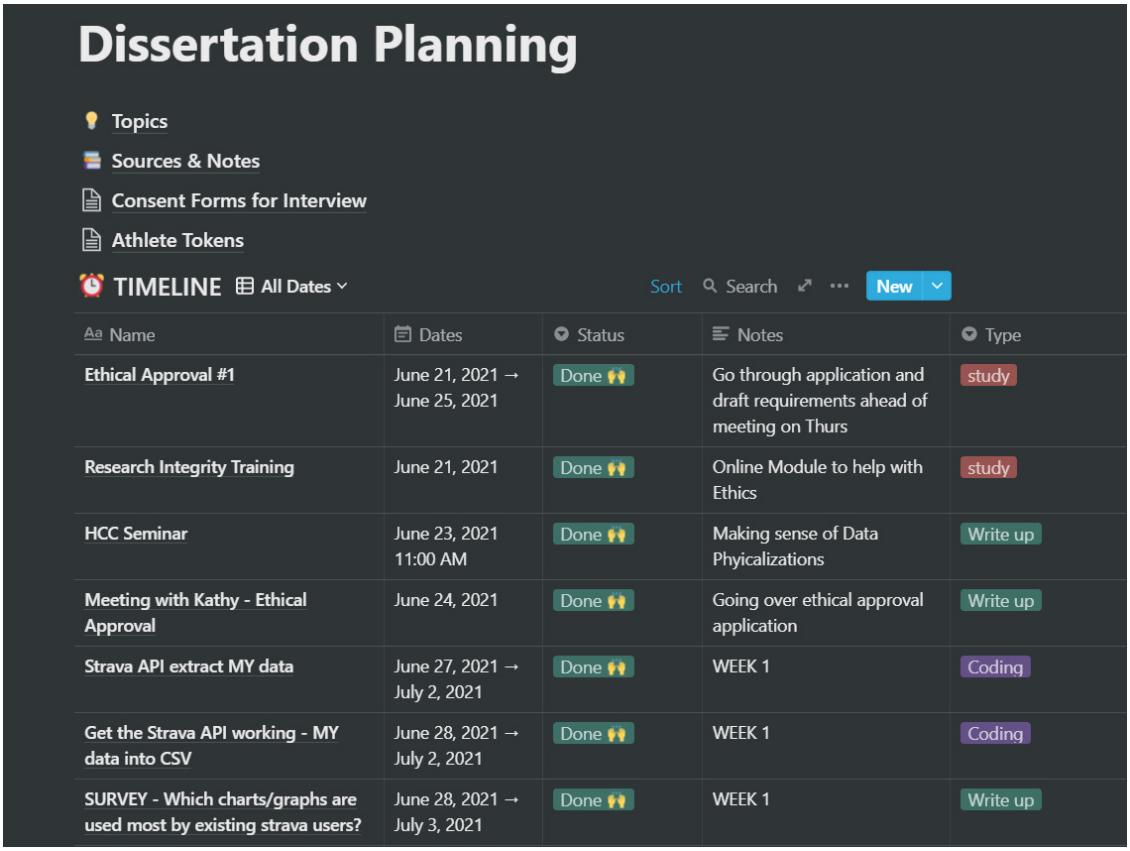


Figure 3.1: A Section of the Gantt Chart



Dissertation Planning

Topics
Sources & Notes
Consent Forms for Interview
Athlete Tokens

TIMELINE All Dates Sort Search New

Name	Dates	Status	Notes	Type
Ethical Approval #1	June 21, 2021 → June 25, 2021	Done 🙌	Go through application and draft requirements ahead of meeting on Thurs	study
Research Integrity Training	June 21, 2021	Done 🙌	Online Module to help with Ethics	study
HCC Seminar	June 23, 2021 11:00 AM	Done 🙌	Making sense of Data Phyicalizations	Write up
Meeting with Kathy - Ethical Approval	June 24, 2021	Done 🙌	Going over ethical approval application	Write up
Strava API extract MY data	June 27, 2021 → July 2, 2021	Done 🙌	WEEK 1	Coding
Get the Strava API working - MY data into CSV	June 28, 2021 → July 2, 2021	Done 🙌	WEEK 1	Coding
SURVEY - Which charts/graphs are used most by existing strava users?	June 28, 2021 → July 3, 2021	Done 🙌	WEEK 1	Write up

Figure 3.2: Timeline of Specific Tasks

3.2 Online Survey

Collecting feedback from participants was an essential part of the developmental process. The first interaction with participants was in the form of a short online survey regarding their current thoughts and opinions on the visualisations Strava offered them. An initial survey was selected as once it had been created it could be distributed to numerous participants [1]. This allows for a large quantity of feedback to be received quickly as a foundation for development. Within the survey a mix of quantitative and qualitative questions were asked. Quantitative questions allowed for easily interpretation of the responses, and the qualitative questions provided further insight that could be developed in interviews. Qualitative feedback is

important within the context of this project as it is impossible to measure behaviour change within a short time frame, therefore capturing thoughts and opinions through qualitative responses was a key measure of success [2].

Using Cox et al.'s guidance on conducting HCI experiments, the following factors were considered while planning the participant questioning process [1]:

- Why the questions are being asked?
- Who the results are for?
- What are the expected findings?
- How is the data from them going to be analysed?

The research question was used to design the short and simple survey, with a completion time of less than 10 minutes. This ensured that participants clearly knew how to respond and did not lose patience with the questions, resulting in more meaningful responses [1].

3.3 Interviews

The interview process was semi-structured, allowing for the maximum amount of information to be gathered. Qualitative interviews were used as they allowed the participants to reflect and evaluate information more effectively [21]. Set questions were asked with key topics and themes to keep the interview on track, however deviations that are relevant to the overarching themes of the project were welcome to ensure genuine responses and provide insight that was not anticipated during the planning process.

3.4 Ethics

As participants were recruited and their data collected for this project, research ethics was crucial in order to protect the rights, dignity and safety of those participating in research. To adhere to research ethics set out by Cardiff University and the School of Computer Science and Informatics, I undertook the research integrity training prior to the research and submitted an application to the ethics committee, which was discussed and approved (SREC reference: COMSC/Ethics/2021/084).

This project collects data that can be related directly to an individual. The project involved exclusively adult participants and no vulnerable adults or groups. None of the topics discussed with participants would be considered highly sensitive or invasive. Before participation information sheets were provided to all participants and signed consent forms were received before their inclusion within the studies. The only personal identifiable information collected were the email addresses of those who opted into the interviews, to allow contact with participants. The survey was answered anonymously and those interviewed were anonymised through participant identification numbers.

The only significant ethical risk identified was the potential for slight emotional distress for interviewed participants while reflecting on their data. To negate this, participants were reminded that they can remove themselves from the study at any point without reason to protect their well-being.

By using this methodology the project was systematic, organised and ethically sound. All interactions with participants were effective and directly related to the research question. The next chapter will explore study one, the survey, in further detail and overview its results.

Chapter 4

Study One: Survey

This chapter will detail the first of two studies, the survey. It was used to evaluate initial opinions on the current Strava offering as a foundation for development. This chapter will discuss the methods, results and conclusions from the survey.

4.1 Methods

This survey was used as a foundation for idea development. Participants shared their thoughts and opinions about the visualisations selected. They were asked to express thoughts on the visualisations design as well as their usefulness.

4.1.1 Participants

The average response time for the survey was four minutes and 26 seconds. The total number of responses was 41 exceeding the number of required participants (30) and provided enough data for reasonable insight. All participants passed the calibration question confirming that they had used Strava within the last four weeks to ensure they could provide enough detail in their responses. Gender and age-bands were collected to ensure a range of demographics were surveyed, with an option to opt out, which none of the participants chose. Nineteen

(46%) of the participants were male and 22 (54%) were female. Ten (24%) participants were aged 18-24, 11 (27%) were aged 25-34, six (15%) were aged 35-44, 12 (29%) were aged 45-54, two (5%) were aged 55-64 and none were 65+ (see Figure 4.1).

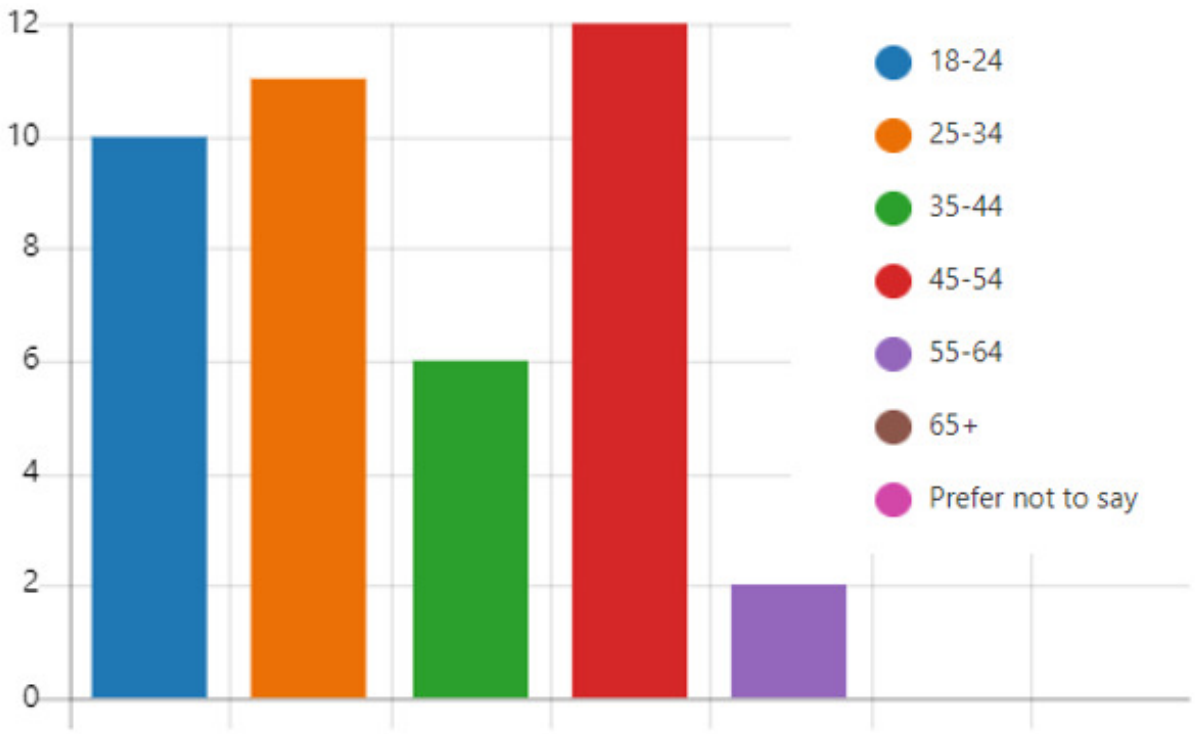


Figure 4.1: Participant age-bands.

4.1.2 Procedures

The procedure for this study began with logistical tasks such as providing information sheets, collecting consent forms and answering any questions from participants. Participants for the survey were recruited online using social media (Facebook, Instagram and LinkedIn). Online recruitment was chosen as it was the fastest and easiest way to reach the maximum number of participants. From sharing online many participants asked if they could share amongst their own clubs and peers which was an additional method of recruitment.

4.1.3 Materials

An online survey was used to collect the data. Within the survey was the information sheet and consent forms (see Appendices A and B) which they had to read and sign before answering any questions. The data collected from survey participants included their thoughts and opinions on three visualisations provided. They were also asked questions on their use of the visualisations and general opinions on their usability and design.

All work for this project was conducted on a university laptop to protect the data of participants and ensure all work was backed-up on the university protected OneDrive. The survey was conducted using Microsoft Forms as it is the university recommended platform and shareable online. This project gained favourable ethical opinion from the ethics committee, the SREC reference is COMSC/Ethics/2021/084.

4.1.4 Analysis

The results from the quantitative questions in the survey were analysed statistically. The survey also allowed for qualitative responses and these responses were analysed thematically and counts of each theme were used to report the results. Themes included praise or criticism and displayed the spread of opinion.

4.1.5 Survey Design

Three visualisations were selected for participants to evaluate during the survey: the Training Calendar (Figure 4.2), Training Log (Figure 4.3) and Fitness & Freshness (Figure 4.4). These were selected as they are all visualisations that can be used to plan and organise activities. Planning and routines are key to motivate behaviour change and were the focus for the studies. Both the Training Calendar and Training Log display the same data, which sporting activities have occurred and when, but in different ways. They have both been chosen to evaluate which design choices are preferred amongst participants to guide my own

design choices. Fitness & Freshness is Strava's only predictive visualisation, and it aims to inform users when they are at peak fitness. This has been chosen to evaluate how useful participants find predictive visualisations. It is also a highly complex visualisation that requires explanation, so these qualities will also be evaluated from a usability perspective. Training Log and Fitness & Freshness are Strava Premium visualisations, and therefore users must pay a premium to access. This could have limited the number of participants. However, the benefit of these visualisations outweighed the risk of a lack of participants as by using them I could evaluate the power of design, by comparing the Training Calendar to Training Log, and also the perception of a predictive visualisation.

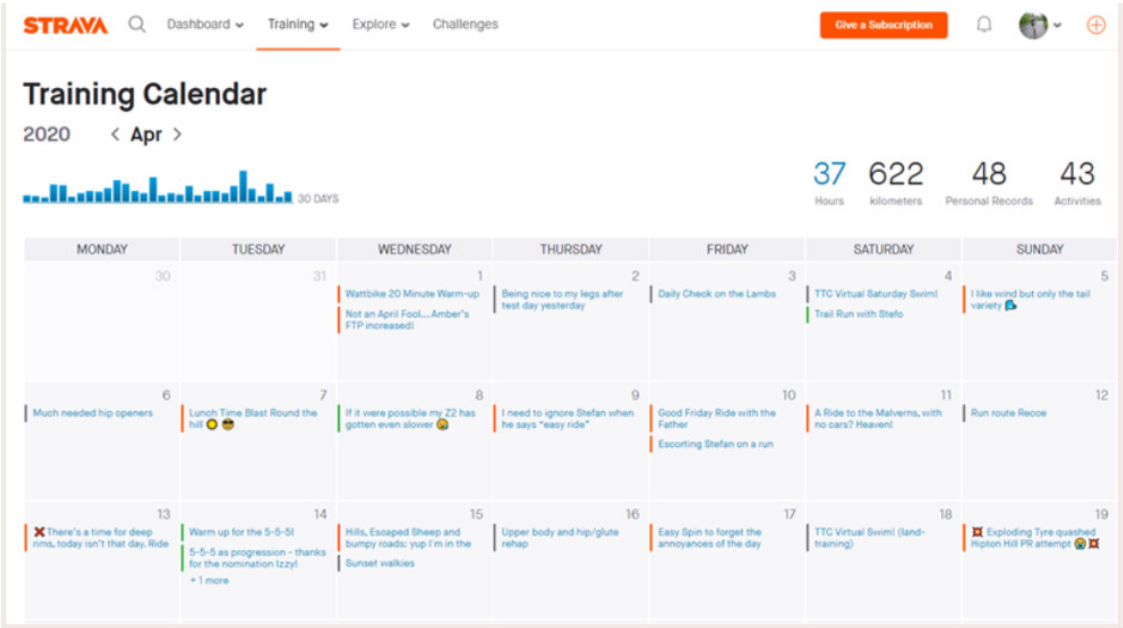


Figure 4.2: The Training Calendar Visualisation.

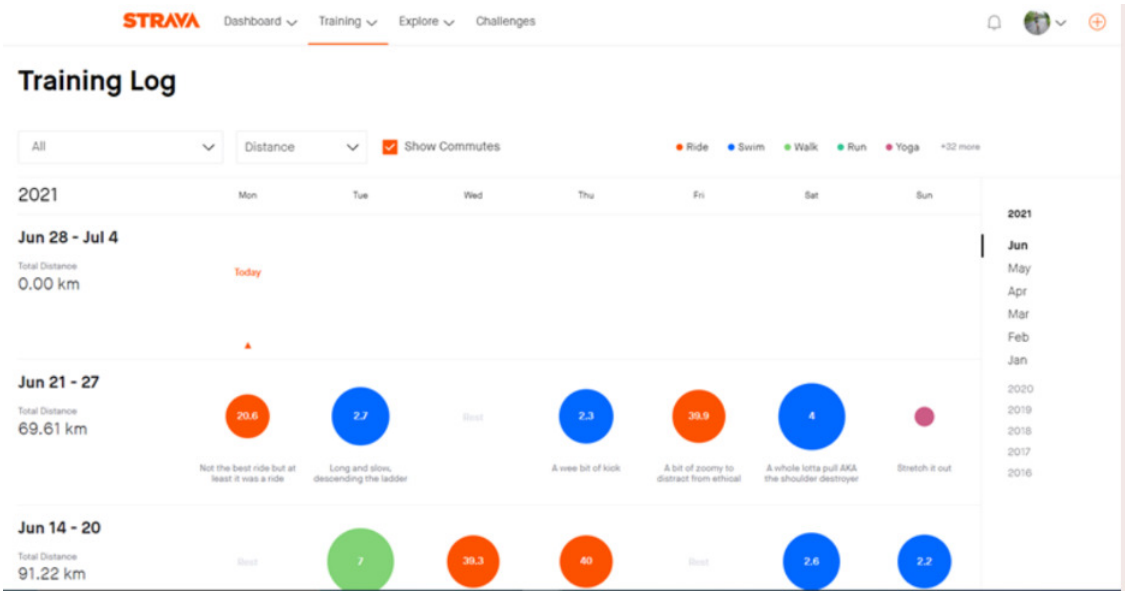


Figure 4.3: The Training Log Visualisation.

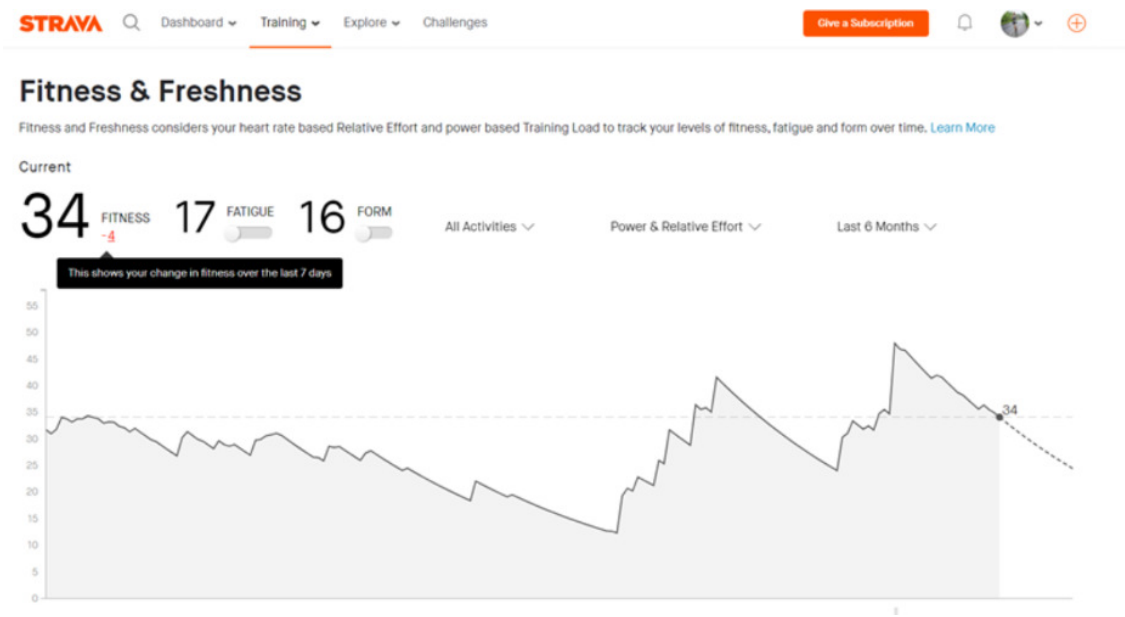


Figure 4.4: The Fitness & Freshness Visualisation.

The first part of the survey centred on the three selected visualisations. The same three questions were asked of each visualisation:

- Do you use this visualisation? (yes or no).

- Why do you use it? (to see what I did in the past, to plan future activity, all of the above).
- Any additional comments? (text box for free-form answers).

The second half of the survey used a Likert scale [5] as a method of gathering opinion toward themes to be asked in the subsequent interviews. It was chosen as it is useful for gathering feeling, such as motivation or reasoning, behind a behaviour [35]. Participants were asked to detail to what degree they agreed or disagreed with the following statements:

- Visualisations (like the examples shown within this survey) help motivate me to exercise and form exercise habits.
- I use Strava to reflect on past activity rather than plan future activity.
- Suggestions of when I will carry out sport in the future would be useful and help me plan my daily life.

These questions were asked to gather a broader understanding of how participants make use of their data and begin to establish potential requirements.

4.2 Findings

4.2.1 Questions on Visualisation One: Training Calendar

The first visualisation shown was the Training Calendar. Thirty participants (73%) did not use it and 11 participants (27%) said they did. Out of those that used the feature 10 out of 11, (91%) used the visualisation to see what they had done in the past and the other participant chose "all of the above" using the visualisation for both past reflection and future planning. The third question was analysed thematically and split into categories (see

Figure 4.5). The largest proportion of participants, 19 (47%), stated they knew the Training Calendar existed but they did not see its purpose.

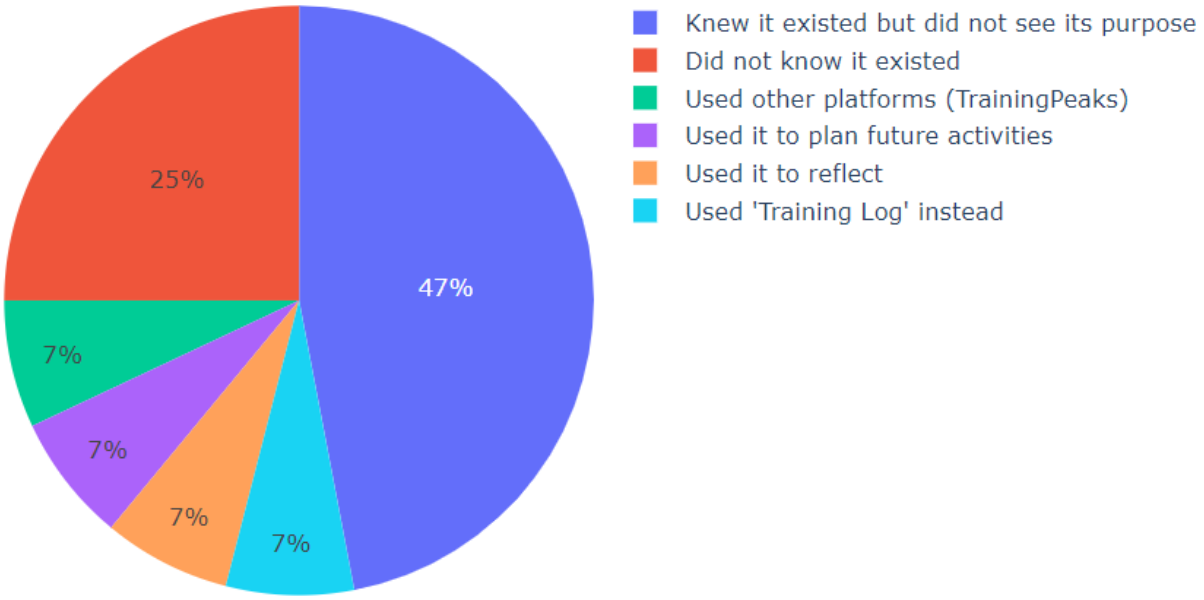


Figure 4.5: Results from Thematic Analysis from Visualisation One.

Comments from participants displayed their lack of awareness of this feature. Participant eight (P8) wrote "I very rarely use Strava on desktop and am not aware of it on the mobile app". Others gave their training as a reason for not using it, P10 wrote "don't use as training more adhoc". Some participants did find it useful, P25 stating "it can be useful for seeing how far I have come and what I did when" and P6 also suggests this "I have a half marathon training plan I am currently following and I like to tick off on the sheet which is quite satisfying". P2 uses their past activity to plan their future "I've used [it] to see previous training loads, when I hit a specific 5k time, I wanted to see what my training block was like in the months leading up to it". This suggests that users have many visualisations to choose from and use them for different reasons. Some participants are also not aware of all available options. It also indicates that users may need extra guidance to help them realise

the value of visualisations. For participants that did use the visualisation, it demonstrates the value of planning and structure to help continued motivation.

4.2.2 Questions on Visualisation Two: Training Log

The second visualisation was the Training Log. This feature is only available to Premium Strava users and as a result 15 participants (37%) could not evaluate its usefulness. Twenty (49%) individuals stated they had used it and six (15%) said they had not. Removing those who are not premium members leaves 20 out of 26 (77%) of participants having used the Training Log and six out of 26 (23%) not using it. All 20 of those using the visualisation (77%) stated it was used to reflect on what they had done in the past. Thematic analysis of the comments left in the final part of this question lead to the following statistics in Figure 4.5.

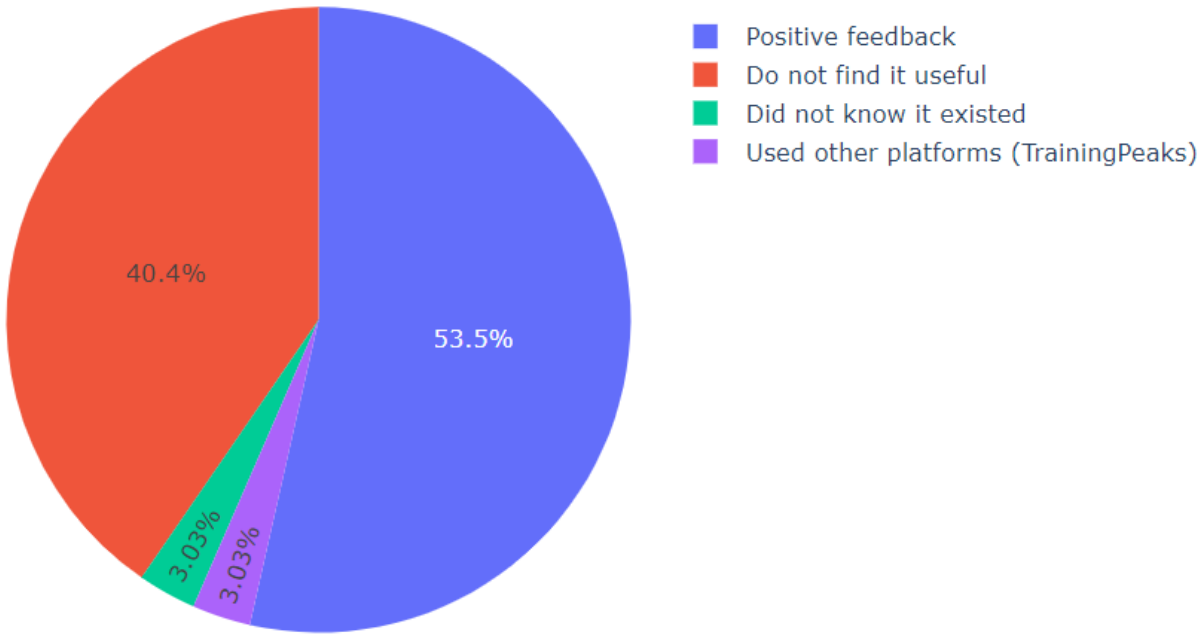


Figure 4.6: Results from Thematic Analysis from Visualisation Two.

Participants made more use of this visualisation with 22 participants (53.5%) providing

positive feedback, two additional participants praising it despite not having used it. P11 comments on the improved layout as a reason for their use "better layout, I like the circle diameters for the length of activity and best it lets you 'mouse hover' to see more detail". P34 writes "I like that you can break it up by activities and I use it to track weekly time/distance". P39 wrote "This is a good feature, but I don't use Strava enough to warrant it". However, 17 participants (40.4%) stated that they do not find it useful. P19 wrote its "not relevant" and P22 was "unsure on how it would benefit me". This suggests the power of the aesthetic design, with more positive feedback being received within this visualisation in contrast to the previous, despite the same information being displayed. However, there are still comments on the lack of use or relevancy, suggesting that some participants still find a record of what they have done useless and not engaging, suggesting that displaying the right content within the visualisation is still a vital component to its design.

4.2.3 Questions on Visualisation Three: Fitness & Freshness

The third visualisation was "Fitness & Freshness". This was also a Premium feature and therefore a percentage of the participants were unable to respond. Removing the non-premium users, 14 out of 27 participants did not use this visualisation (52%) and 13 out of 27 participants (48%) did. When asked how the visualisation was used, seven (54%) participants reported they used it to plan future activities. The categorical analysis of the commented responses resulted in the following Figure 4.7.

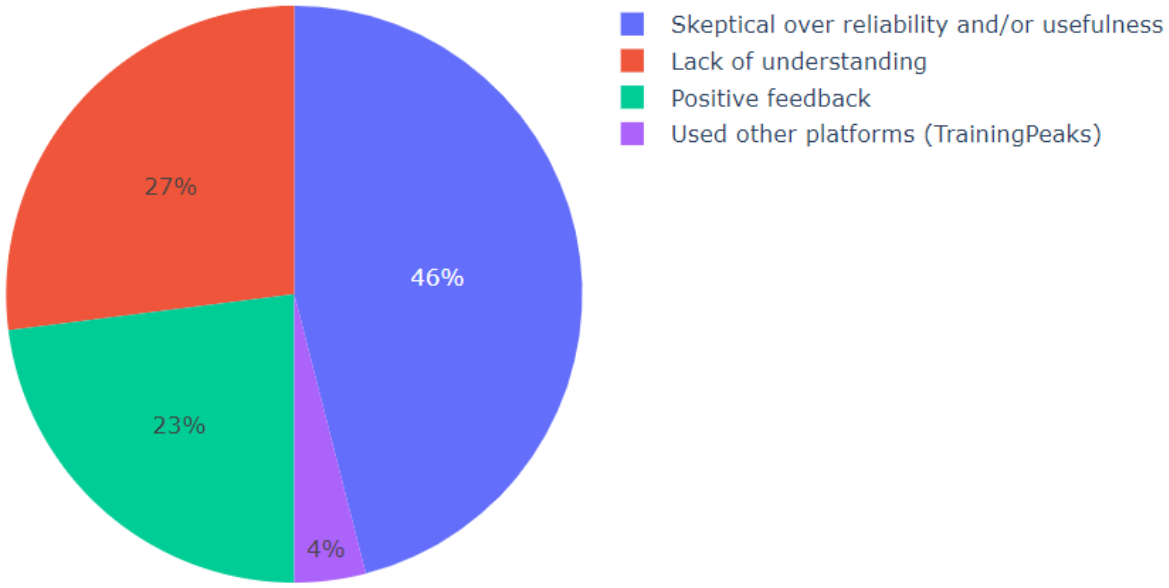


Figure 4.7: Results from Thematic Analysis from Visualisation Three.

Participants commented that they may consider using this type of visualisation but it was unclear or they were skeptical about how accurate the predictions were. P8 commented "used to use it when I used HR for every session, but no longer do so this is not an accurate reflection". P41 wrote "I like this feature but it is not easy to find and isn't clear what it's showing/how its getting that information". Some participants did not understand visualisation. P26 wrote "I have tried [...] as I like the prediction. But I find it confusing and do not really understand it". Only nine out of 41 participants (23%) had positive comments. P12 stated "I like this feature of Strava and use it to check I'm not losing fitness or over training". This suggests the potential opportunity to creating a predictive visualisation with many participants praising the concept. Skepticism was raised, but not toward the concept of a predictive visualisation but by the way in which it was produced. This feedback presents a key requirement, that if prediction is to be successful users must believe its information is valid and that it is trustworthy. The importance of design has also been raised, with a lack

of clarity frustrating participants. This emphasises the importance of usability.

4.2.4 Questions on Use of Activity Data

The final question within the survey was designed to ascertain how participants are using their data and they were asked to state to what degree they agreed or disagreed with the following statements, with the results shown in Figure 4.8:

- Visualisations (like the examples shown within this survey) help motivate me to exercise and form habits.
- I use Strava to reflect on past activity rather than plan future activity.
- Suggestions of when I will carry out sport in the future would be useful and help me plan my daily life.

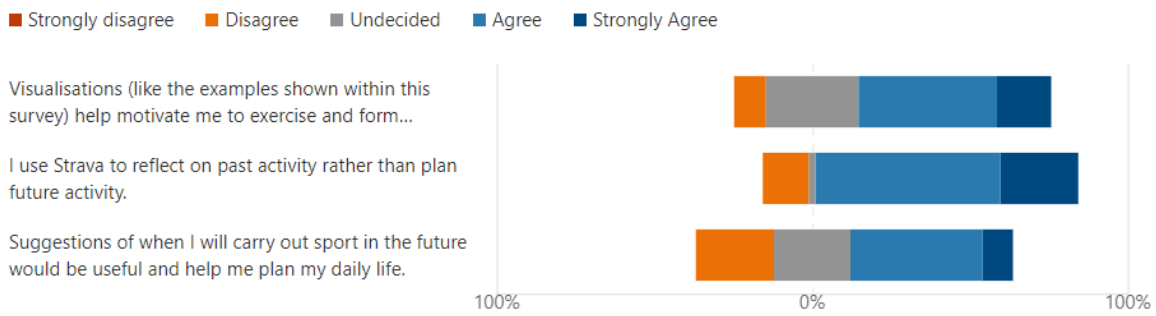


Figure 4.8: Likert Scale on Participant Use of Activity Data.

The results from this question allow an overview on how participants are using Strava. Twenty-five participants (61%) agree or strongly agree that visualisations help or motivate them. Thirty-four participants (83%) agree or strongly agree that they currently use Strava for reflection rather than future planning. Twenty-one participants (51%) agree or strongly agree that predictions may help them plan their daily lives. Overall their responses suggest

that data visualisations are a factor in motivating participants to exercise and currently they are predominantly being used for reflection. There is also an indication that incorporating predictive visualisations may be a way of helping plan and structure future activity.

4.3 Summary

The survey provided a broad insight into how participants are currently using their data through visualisations. This feedback was used to inform design specifications and the requirements (see Chapter Five) for the visualisation that this project will develop. Participant feedback from this survey highlights a lack of clarity and understanding in how visualisations can be used. Many do not use them but have expressed desires to do so, if they were produced in a certain way. Circular designs of the Training Log are more appealing than tabular forms and prediction is intriguing, but only when it is presented transparently in an accessible fashion. Participants use their visualisations primarily for reflection, presenting an opportunity to design predictive visualisations that are trustworthy and more user-friendly.

The next chapter will explore the design process and the technical development aspects of the project including how the data was retrieved, sorted and presented as well as the design process of creating the visualisation based on an interactive feedback-driven cycle.

Chapter 5

Designing & Developing Visualisations

This chapter will explore the creation of the data visualisation shown to participants. Covering the requirements identified from study one, the design process and the development of the visualisation.

5.1 Requirements

Using feedback from the survey (see Chapter Four) and existing literature (see Chapter Two), this section will outline the requirements for the visualisation and inform the development of a prototype.

The key requirements of the visualisation are:

- High user engagement, the visualisation must draw the attention of participants through high degrees of aesthetic quality, this will enhance memorability and cognitive involvement [8, 12, 11].
- The visualisation is simple and easy to use [56].

- Interactive elements [10].
- Telling the truth and not distorting the data [56]. Additionally, ensuring participants trust the validity of the visualisation.
- It is motivational and persuasive [23].
- Creates an "action plan" [14].

5.2 Visualisation Design Process

This project aims to increase the ability of visualisations to encourage action. The design of the visualisation is important for maximum engagement and this section will outline this process. The requirements, based upon existing literature and participant feedback, were used as the foundation for design.

Tufte's rules [56] and Borkin et al.'s [8] research into memorability in data visualisations were used as the foundation of the design as both closely relate to feedback from participants and provide a structured design framework.

With visualisation design being highly subjective, the context of the project was closely considered. Taking a flexible approach with using theory as part of a design framework is important. An example of this regards Tufte's rule: the Data:Ink ratio. Tufte states a good design will provide the most amount of data in the least amount of ink [55][56]. This is not applicable to this project as the aspect of colour and visual elements, such as use of shapes, have been highlighted as significant within participant feedback and across alternative literature [8]. It is an example of Tufte, a mathematician, possessing differing design requirements to participants within this project, who are non-technical sports-people. P27 writes on the Training Log "It's easy to use and I like the colours, I can see activities at a glance" and P30 agrees stating, "I like the circles of colour and the sizes". Tufte's concept

of "graphical excellence" however, is significant as it describes the concept of providing "optimum access to the relevant aspects of [user] data" [45].

Combining the requirements with my own initial experiences with Strava, I began to develop visualisation sketches, as seen in Figure 5.1. Sketching was a useful method of design as it allowed an unrestricted creative process. Sturdee et al. advocates the use of "offline" creative arts within HCI development, writing that sketching is perfect for ideation "because of its non-final state... [sketches] can be edited and discussed" [49].

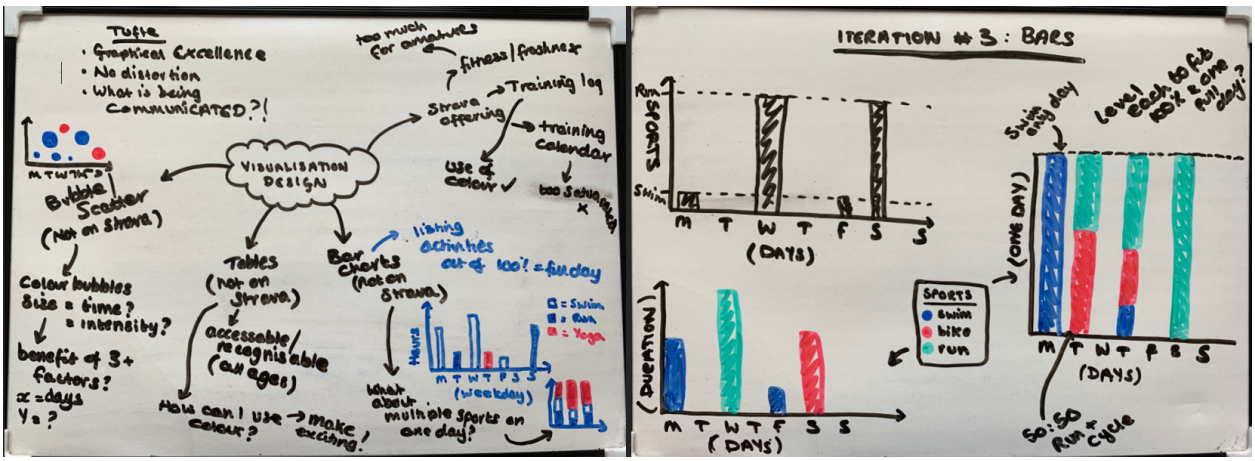


Figure 5.1: Examples of Design Sketches.

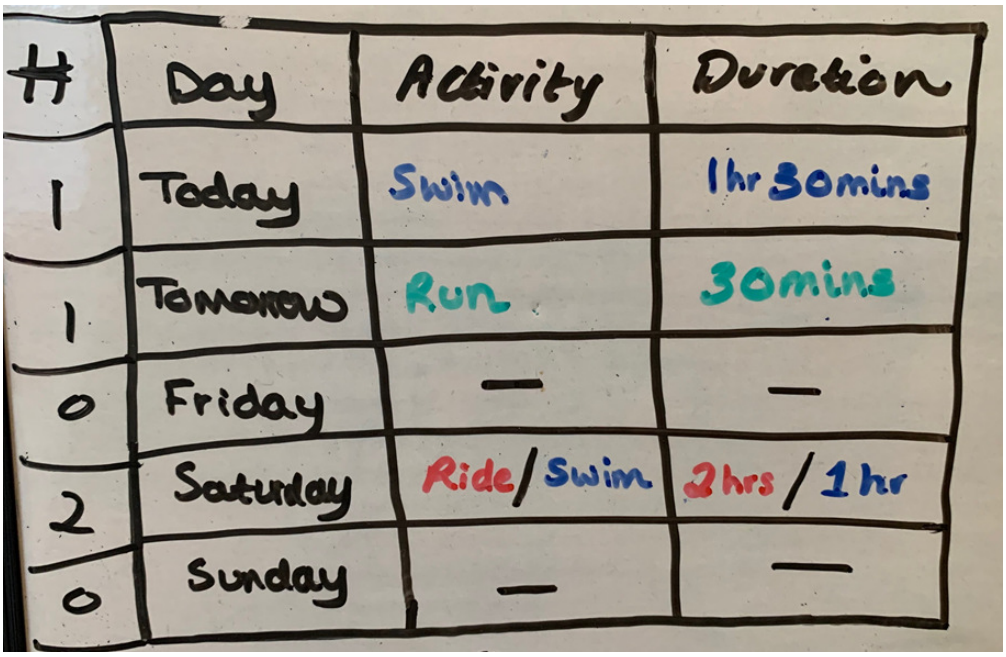
Based on Tufte's rules [55], an efficient and tabular visualisation was designed first to cement the required features for visualising. The absence of chart junk [56], allowed for clarity in the initial development phase. Using data already visualised in Strava's Training Calendar and Training Log, the following features were chosen for visualisation:

- The sports or activities that the user is going to participate in (for example: swimming or weight lifting etc...).
- Duration of the activity (in hours and minutes).
- What day it may occur (for example: Monday).
- Time of day (morning or afternoon).

A table was also considered as a possible visualisation due to its simplicity and ease of interpretation, factors identified as significant from Study One (see Chapter Four). A tables commonality is also beneficial with research suggesting that simple and "recognisable data visualisations" are key to communicating data to non-technical audiences [26].

The first iteration was based upon a weekly calendar displaying the most likely activity for each day. The visualisation was useful but inflexible and not engaging due to its lack of aesthetic qualities. Borkin et al. writes that additional aesthetic qualities forces the reader to engage cognitively and results in improved memory retention [8].

Next, I explored interactivity to encourage user engagement [10] [6]. A new table concept displayed set days and times dependent on the date it was viewed and allowed a number of sessions to be displayed, as can be seen in Figure 5.2.



#	Day	Activity	Duration
1	Today	Swim	1hr 30mins
1	Tomorrow	Run	30mins
0	Friday	—	—
2	Saturday	Ride / Swim	2hrs / 1hr
0	Sunday	—	—

Figure 5.2: Design Development: Table.

Interactivity was introduced in early iterations as Bruggemann et al. writes that a flaw of visualisations is that they are often retrospectively interactive rather than designed with interactivity in mind [10]. Interactivity provides the power of personalisation to users

without coercion or intrusiveness [50] [3]. Introducing the concept of interactivity was an improvement to the design, but tables as a visualisation lack the graphical complexity to fulfil the requirements.

Borkin et al. conducted a study on the impact of visualisation design on visualisation memorability [8]. They compared how participants remembered the same data in pictographic form versus a tabular form. The study suggested that a higher degree of graphical elements makes the visualisation more memorable. These findings informed my decision to increase graphical complexity. Memorability is a key factor for action as it makes participants more likely to engage and invest in the visualisation longer term [8].

The next iterations were bar charts. They increased the graphical complexity, but maintained the familiarity. A stacked bar chart was built as a highly visual method of representing the likelihood of participating in selected sports for every day of the week (See Figure 5.3.). When representing parts of a whole, a pie chart could have been used. However, linear lengths are easier to interpret than angles, therefore bars result in a clearer visualisation [47].

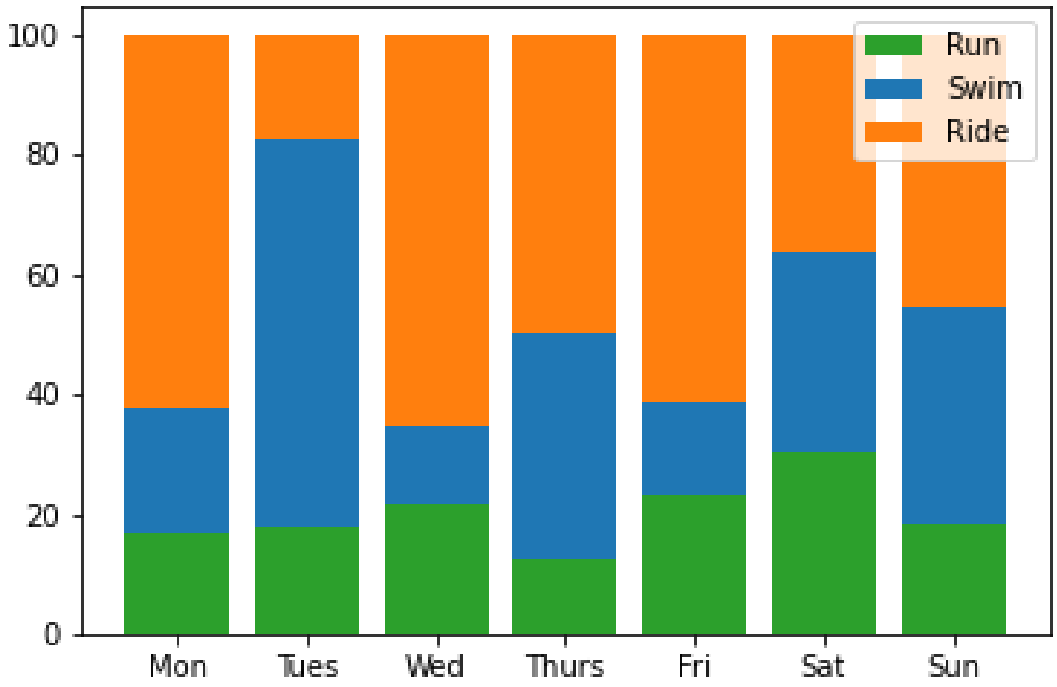


Figure 5.3: Design Development: Stacked Bar Chart.

This iteration was highly visual, but lacked interaction and provided descriptive statistics as opposed to desirable predictive data. Participants in study one liked the specific data they could read from existing visualisations on Strava, rather than purely times and dates (see Chapter Four).

Part of Borkin et al.'s findings on memorability counter assertions on the importance of familiarity regarding data visualisations. They found that overly familiar or common data visualisation types are less memorable as they require less cognition [8]. Familiar visualisations with novel aspects are more memorable, but an important caveat is that when designing unique visualisations, the design should not distract from the value of the data [8]. A balance was negotiated between developing a visualisation that was recognisable and easy to understand, but is different enough to be distinctive, memorable and increase the

likelihood of action.

Informed by Borkin et al., the next iterations balanced familiarity with distinction. Scatter (see Figure 5.4) and bubble plots were chosen in order to include more features and additional data in contrast to the bar charts. Bubble plots are particularly beneficial for adding many features without visual complexity [59, 57]. These iterations allowed experimentation with design of the circle markers, a design choice praised in the Training Log visualisation in study one (see Chapter Four). Circular designs are also more memorable than "sharp corners" [8].

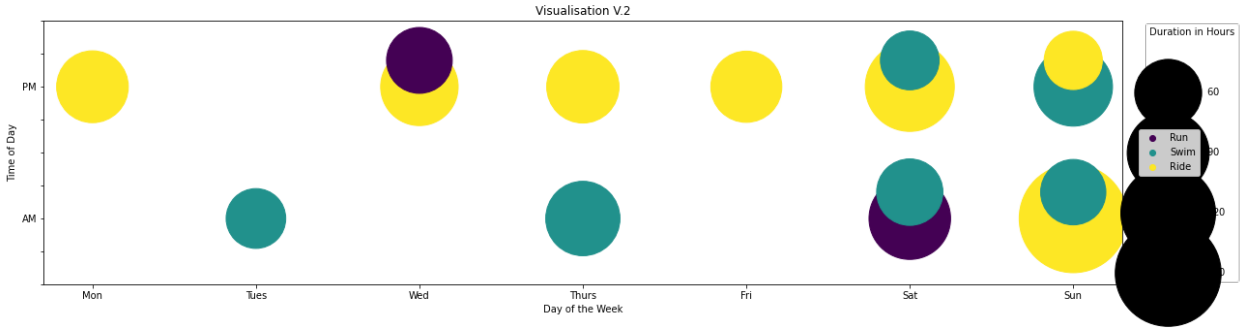


Figure 5.4: Design Development: Scatter Plot.

The bubble iterations included features such as hover and bubble size based on duration, quick visual methods of interpretation of precise data points. The bubble plot and interactivity resulted in a simple visualisation with high levels of user-driven personalisation, as shown in Figure 5.5.

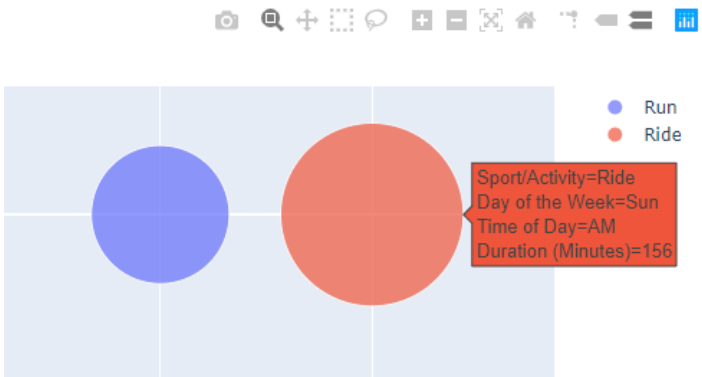


Figure 5.5: Design Development: Bubble Plot Features.

Further iterations included more interactivity exploring the use of tick boxes, search bars, sliders and drop-downs selectors. A searchable drop down was added to allow users to select the sports visualised depending on their needs. A slider was used to allow users to choose how many activities displayed on the visualisation, making it more flexible and meaningful. The default activities displayed are the participant's most typical activities based off historic data, providing additional insight to users. An input box was considered, but a slider removes the possibility of user error. The visualisation with improved interactivity is shown in Figure 5.6.

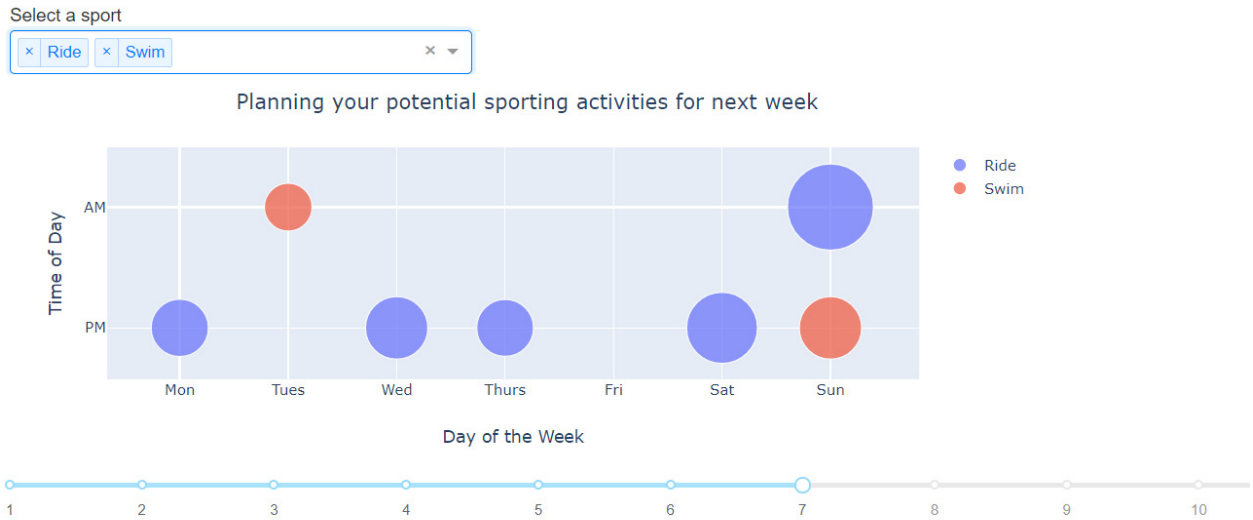


Figure 5.6: Design Development: Pilot Study Visualisation.

A pilot study was run, with a participant not involved in the studies, to test the usability of the visualisation. The visualisation was shared virtually with the participant controlling my screen to see how they interacted with the visualisation before asking comprehension questions. As this iteration used my data the pilot study evaluated usability and engagement, rather than the impact on behaviour change.

The pilot revealed that the legend and labelling were insufficient and unclear, the participant not understanding what the slider did and confusing the x-axis label with a slider label. They also suggested highlighting interactivity options to ensure the users were aware of all possible ways of interacting with the visualisation. As a result more labels were added alongside text for clarity. They also drew attention to an issue when two activities occurred at the same time, the bubbles overlapped each other, hiding one activity. The opacity of the bubbles were changed and they were plotted with an offset (see Figure 5.7).

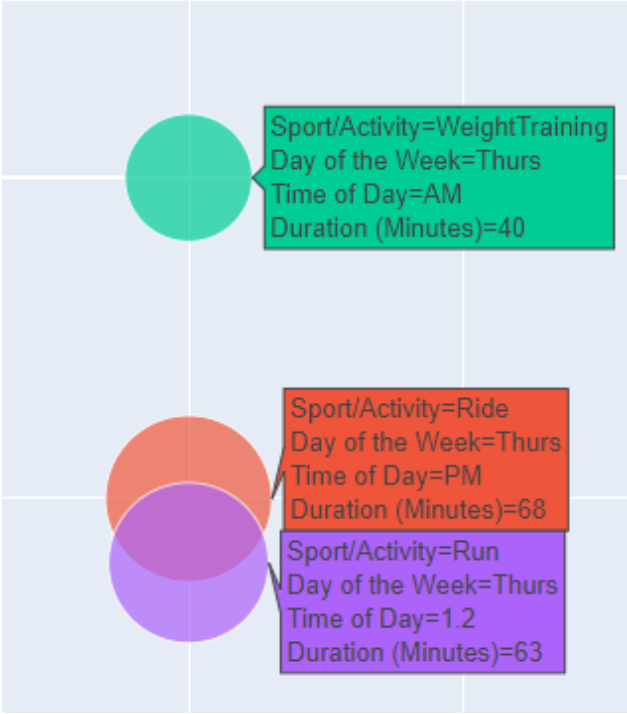


Figure 5.7: Design Development: Overlapping Bubbles.

This feedback formed part of the iterative design process and allowed further improvements resulting in Figure 5.8 for use in study two.

Through an iterative design process many versions of the visualisation were designed based off the original requirements before being shown to a participant to gather in-depth feedback for further development. The next section will cover the technical aspect of creating the visualisation.

5.3 Visualisation Development Process

This section will discuss the final visualisation as well as explaining its technical elements: how the data was acquired, cleaned, processed and visualised.

The final visualisation (Figure 5.8) design was a bubble plot that depicted the days of the week on the x-axis and morning and afternoon on the y-axis. The visualisation also

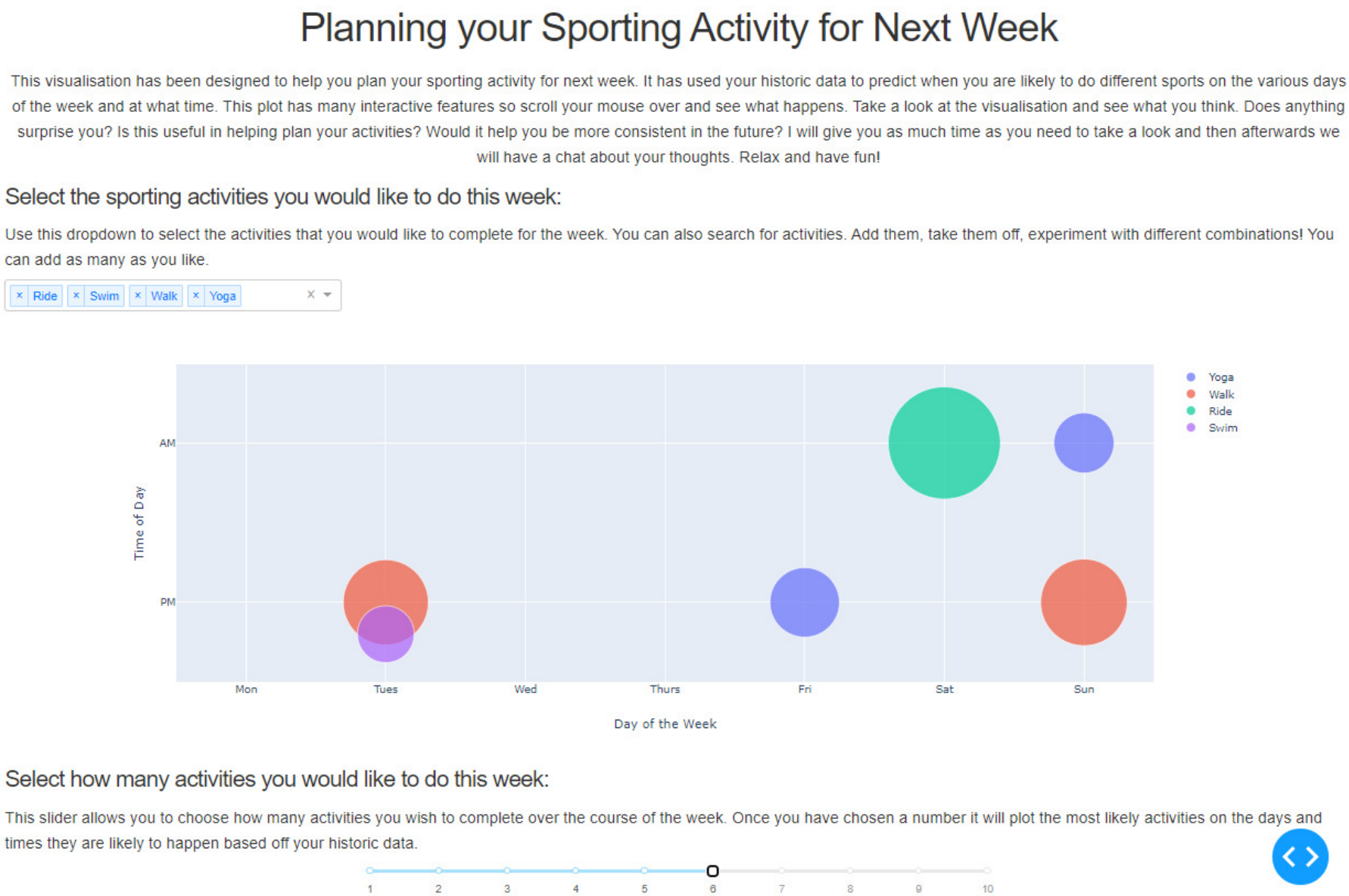


Figure 5.8: Design Development: Visualisation to be Implemented.

contained a slider for participants to alter the number of activities plotted and a drop-down menu to choose which sports were plotted. Once these features were selected, activities are displayed as coloured bubbles on the times and days most likely for participation, with the size of the bubble indicating the duration. Participants can scroll over the bubbles to gather further detail such as specific duration. The plot also contains a legend which displays which colour bubble represents which activity.

The flow diagram in Figure 5.9 shows the process followed to generate the visualisation from data collection to user interaction. More in depth information can be found within the README.md in the GitLab repository for this project [1].

[1] https://git.cardiff.ac.uk/c2094920/msc-dissertation

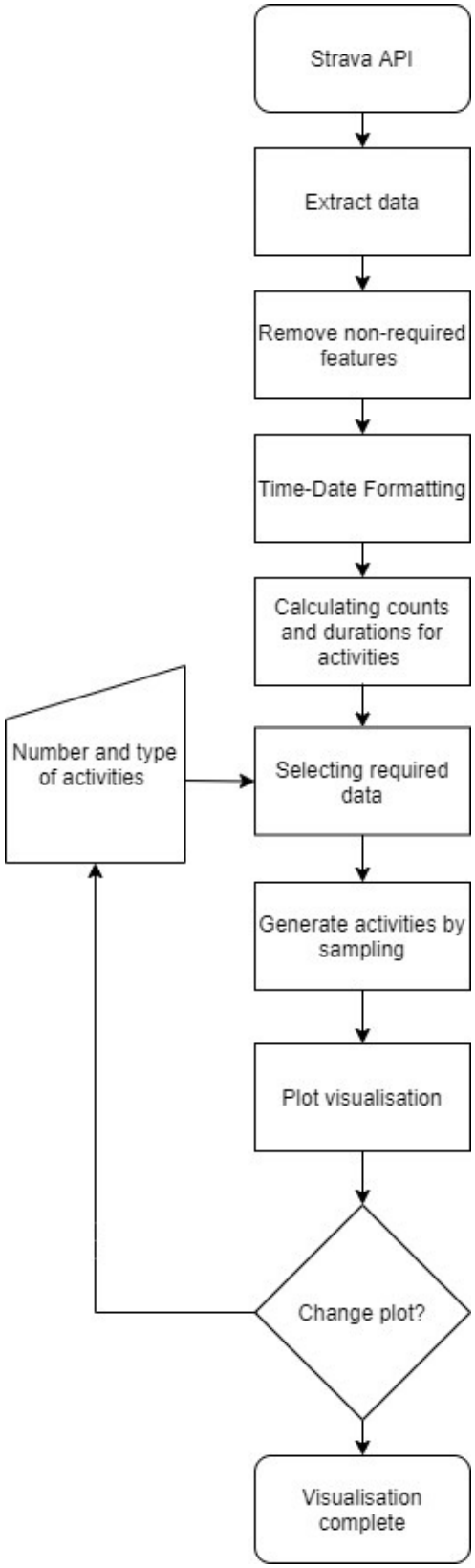


Figure 5.9: Development Flow Chart

This project was written in Python because of its range of libraries for exploratory analysis and data visualisation. Pandas was used for data analysis and exploration and Matplotlib, Plotly and Dash were used to create the visualisation. Coding was carried out using the PyCharm integrated development environment (IDE) for fast development and debugging. Initial data exploration was carried out using Jupyter notebooks within Pycharm which were useful for experimentation. All coding for this project was completed on a university laptop, stored on the OneDrive and backed-up using a private GitLab repository.

The data was collected from Strava as it is a widely used multi-device platform. The Strava application programming interface (API) was used to access the data required [18, 52]. The API access information (`client ID`, `client secret` and `access token`) for each participant were stored within a configuration (config) JSON file. The `get_data_from_strava` function accesses the Strava API using the provided `client ID`, `client secret` and `access token` and extracts the selected features from all activities into Pandas DataFrame and then saves to a CSV. The features selected correlate to the required data identified in the design section (see section 5.1 of this chapter) and include: date and time, activity type, distance, elapsed time, country, latitude and longitude.

The `clean_data` function prepares the data for analysis and plotting. It formats the date and time splitting it into two features and then separating the date into day, month, year and the time into hours and minutes, removing seconds. It then uses the time to add a morning (AM) and afternoon (PM) feature. Next, the data types are corrected and data older than 365 days is removed and then saved overwriting the CSV. Only one year worth of data was required for the project and therefore it was removed at source to prevent unnecessary data collection.

The programme is run using the `main.py` file which takes two command line arguments, `athlete_ID` and the options of `extract` for data extraction, or `plot` for plotting the extracted data. This procedure was used as the Strava API access tokens are time sensitive.

This method allowed for the data to be extracted then saved as soon as the access codes were received from participants, but the plotting could occur at a later stage. The access codes were sent once, immediately used and then deleted.

The `data_vis.py` file contains the `bubble_plot` function which creates the visualisation. The first step within the function is to read the CSV and then calculate the count (frequency of occurrence) and average time of every activity within the data set, for every day of the week. These occurrences are then further split by specifying whether they occurred in the morning or afternoon. The result is two 3-dimensional NumPy arrays, one that contains the counts and the other average duration, which are the data sources for the visualisation. The code for this step is shown in Figure 5.10.

```
for i, activity in enumerate(activities):
    for j, is_morning in enumerate([True, False]):
        data_loop = data[(data["type"] == activity) & (data["morning"] == is_morning)]
        count = data_loop["weekday"].value_counts(sort=False).reindex(index=list(range(7)), fill_value=0).to_numpy()
        time = [data_loop[data_loop["weekday"] == i]["elapsed_time"].mean() for i in range(7)]
```

Figure 5.10: Calculating Counts and Duration.

The end user of the visualisation can input the number of activities and the activity types. The selected activities were used to look up the corresponding data which were used to produce the visualisation. The predictions were originally calculated exclusively via greedy selection whereby activities with the highest number of occurrences were displayed. When trialling this design on my own data, it presented cycling as the most likely activity due to cycling being a key sport in the last year but this is not representative of my typical habits and is inaccurate. Greedy selection also resulted in deterministic predictions which lacked variety to keep the user engaged.

To improve on the selection criteria a random sampling technique was used to identify activities. Specifically, a multinomial distribution was created using the normalised activity counts as event probabilities (`pvals`). The required number of samples were drawn without replacement to avoid repetition (see Figure 5.11). PyTorch was used to implement this

due to its ability to sample without replacement. This introduced stochasticity as activities are selected in proportion to their likelihood rather than just choosing the most likely. This prevented the same activities being presented every time and provided different combinations of activities with high likelihoods, which is more useful as a form of suggestion to encourage user engagement.

```
pvals = selected_sport_counts.flatten() / selected_sport_counts.sum()
max_counts = torch.multinomial(torch.Tensor(pvals), num_of_sessions, replacement=False).numpy()
```

Figure 5.11: Implementing a random sampling technique

The bubble plot design chosen (see section 5.2) was plotted using Plotly and Dash. Plotly provided the initial visualisation. Dash allowed for the additional functionality and annotation. Plotly was chosen due to its extensive range of visualisations. Plotly.py is built upon a JavaScript equivalent and allows for a D3 style customisation without the complexity. Using the Dash library allowed the addition of explanatory HTML and CSS components for customisation. Dash uses Plotly.js to render additional interactivity which was used for the slider and drop-down functionality. These interactive elements acted as the input functionality for end users to dynamically alter the plot and specify which activities they wished to plan for and how many sessions per week.

The next chapter will discuss the second study, the interviews. It will establish the methods, results and further design iterations from participant feedback.

Chapter 6

Study Two: Interviews

This chapter will detail the second study, the interviews. The interviews were used to gather feedback on visualisation developed, investigate facilitators and barriers to behaviour change in sport and continue to ascertain the usefulness of predictive data visualisations. This chapter will discuss the methods, results and conclusions from the interviews.

6.1 Methods

6.1.1 Participants

Nine participants were involved in this study and they were split across a range of age groups. One participant was in the 18-24 range, three were in the 25-34 range, two were in the 35-44 range, two were in the 45-54 range and one in the 55-64 range. The gender split was five female and four male. Their sporting background and experience ranged broadly from ex-professional athletes to those newly exercising for personal well-being. Each participant remained anonymous and will be referred to using their participant number. Recruitment for the interviews came from an option to opt in during the survey. The target number of participants was between eight and 10. Participants for the interview must have used Strava

for at least one-year in order to possess enough data for a meaningful visualisation to be created.

6.1.2 Procedures

After reading information sheets, answering a questionnaire and signing consent forms (available in appendices C and D), participants were then asked to complete an additional survey to collect the required codes for data extraction using the Strava API (see Chapter Five). The visualisations were generated after extracting the data and then shown to the participants for the first time during the interview. The interview was semi-structured, making allowances for relevant comments away from my prompts.

Interview participants were asked to provide approximately 365 days of data from their Strava profile, extracted using the Strava API in order to provide them with a visualisation to evaluate. Other data collected included their current exercise habits, how they used Strava and their thoughts on the usefulness of the visualisation shown to them. Participants were asked to interpret different aspects of the visualisation, this data was collected to test usability.

The following framework for conducting interviews was used [1]:

- Background: understanding how the participant uses Strava, their current behaviours, experience levels and attitudes toward exercise.
- "Letting off Steam": allowing participants to state the opinions they wish to say at the beginning, even if not relevant, to prevent these thoughts from re-surfacing throughout the interview.
- The Issues: Asking further questions that have not already been introduced, being careful to avoid leading any responses. Depending on what had already been stated,

this includes questions on motivation or questions that indicate their understanding of the visualisation by asking them to interpret or explain aspects of it.

- Debriefing: Rounding up the interview by summarising thoughts and giving time for any other thoughts to be said.

6.1.3 Materials

Data in the form of participant opinion was collected through individual interviews online. The interviews were conducted on Zoom using Cardiff University login credentials. Each interview was recorded for analysis purposes and transcripts from the interviews were typed on Microsoft Word. To take part participants only required a laptop or computer that they could use to clearly view the visualisations that were shared on screen and remote control was used to allow participants to explore the visualisation independently. All transcripts were saved on the university OneDrive to secure participant data.

6.1.4 Analysis

The interviews were designed to investigate the usefulness of the predictive visualisation and identify areas for further development. There were a set of questions each participant was asked, but throughout the interview deviations on related topics were encouraged to gather authentic feedback. Thematic analysis was used to analyse the results as a useful method of "identifying, analysing and reporting patterns within data" [9]. Braun and Clarke's 6-stage process was used to conduct the analysis [9]. After each interview the recording of it was reviewed multiple times and detailed notes of the conversation were taken for analysis. Direct transcription was taken at significant points during the interview while reviewing the recordings. Thematic analysis was then carried out, involving [9]:

- Familiarisation with the data: transcription, re-watching and idea generation from the

interviews to ensure a comprehensive understanding.

- Code Generation: creating initial codes that group meaningful parts of the interview feedback together within its context[1].
- Themes: grouping together similar codes according to broader themes.
- Reviewing Themes: revisiting the categorisation and adjusting or adapting where necessary.
- Report: providing the themes in writing alongside examples of the theme to justify the analysis.

Six themes were produced: visualisation for reflection, data lead development, aesthetic design, motivation, flexibility and additional requirements.

6.1.5 Pilot Study

As part of the design process, my own Strava data was collected to create the designs for a pilot study before using participant data for the interviews. Using a student not involved in the study, the visualisation and interview questions were trialled on them. This allowed for continued development before the studies commenced. As a result of the pilot I added contextual questions to my interview prompts to establish the participant's experience and current sporting behaviours as a foundation for analysis.

[1]Fourteen codes were identified: recognises own data, data driven, difficulties with current visualisations, immediate refection, long-term reflection, acting upon their data, motivation, predictions are useful, preference to suggestions or guidelines, new features, automation, rewards, flexibility, aesthetics.

6.2 Findings

6.2.1 Participants' profile

The opening question established what sports the participant undertakes, how long they have been participating and how they currently use Strava, or other similar applications. Five participants were semi-experienced athletes, having participated in their sports relatively seriously for a year or more, with either their experience or support of a club to help them. Three were highly experienced or ex-professional athletes. Those in this group were highly independent and reliant on their extensive experiences to be able to plan and support their exercise. One participant was a novice, taking part in sport for personal enjoyment and not having the experience to plan their activities or any form of coaching or club structure. See figure 6.1 for the breakdown of sports carried out by participants.

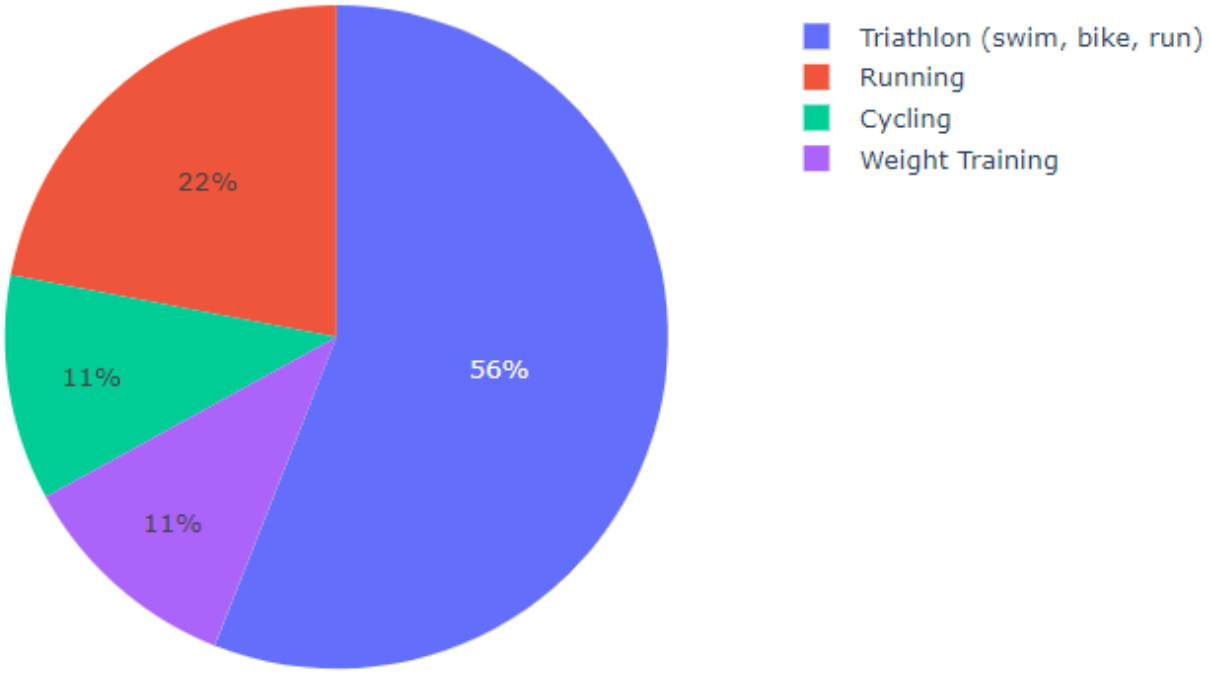


Figure 6.1: Sports of Participants.

6.2.2 Theme: Visualisation for Reflection

One of the themes identified during the analysis was how participants are using their data and how that varies. The most common approach was for reflection, often immediately after completing the activity. Some would then further reflect on significant activities such as races or personal bests. Participant two (P2) stated "I would look back to how I have been doing this week or if I have done a big event the year before, like reflecting on a Triathlon that I did this year and last". For participants training for specific events, the type of reflection depended on their goals such as P7, "If I am training for something I will look and see how I compare to previous runs". Another use within this theme was reflection for practical concerns regarding memorability. P8 used reflection for a practical purpose as they struggle to remember what weights they previously lifted and they use Strava to plan their next session. Reflection has a dual role, it can be used both before and after exercise to establish how future activities may be undertaken.

6.2.3 Theme: Data Lead Development

Eight out of nine participants identified specific metrics they look at while exploring visualisations, such as heart rate (P1, P2, P3, P4, P5), pace (P9, P4, P3) or distance (P9, P7, P6, P4). Therefore, visualised data metrics are a powerful tool in structuring how participants perform and plan their exercise. Participants enjoyed looking at these metrics and using them to help guide their performance.

The usefulness of these metrics seemingly depended on the participant experience level. The more competitive and professional participants used the data to aid their training and race preparation, such as P1, "I am interested in [data] and it supports my racing". This is further emphasised by the participant with the least experience not identifying any specific metrics that they look for in their visualisations.

The semi-experienced participants preferred to use the data as a guide and indicator of

their success. P2 uses it to ensure they are on track "I use Strava to see how far off my times I am [and] I look to see if my runs are worse or better". Indicators of success or reassurance seem to be a large factor within semi-experienced athletes, with P5 stating "Sometimes I look at my times and I am faster than I thought and that is great". This is perhaps due to their lack of experience where feeling alone is not enough to validate their efforts and the data helps to reassure them, keeping them engaged.

However, the dangers of data as a detractor as well as a reassurance were raised within this theme, with some participants expressing a lack of trust in the predictive visualisations or the accuracy of data visualisation more broadly. P8 stating "I do not believe its accurate [...] If I was convinced on how it works or there was a demonstrable history that it worked for others then I would [use it]". P7 uses Garmin Connect rather than Strava for certain activities claiming it "is more accurate than Strava". These concerns raise the issue of trust in a visualisation and suggest that end users will not use something they do not trust or understand.

On the contrary, some participants did recognise the data visualised as their own which lead to an increase of trust. P2 stating "I am very predictable aren't I?", referring to the fact their club swims every Tuesday, Thursday and Saturday which they were able to see in the visualisation. P4 recognised the different times they do certain activities, and P7 also recognised regular club activities in their plots. When questioned on the significance of this recognition both participants stated that it allowed further trust as they understood that the predictions presented were not random and had a purpose by using their personal data.

6.2.4 Measuring the Visualisations Interpretability

Participants were shown the visualisation halfway through the interview and then asked comprehension based questions to ascertain their understanding of the visualisation. These included simple show and tell exercises such as "show me how you would change the sports

displayed" or "tell me what time of day you will swim this week". This was to test usability. All participants were able to answer these questions instantaneously demonstrating an excellent understanding of what parts of the visualisation did or how they would control a requested aspect of it.

The importance of ease of understanding has been highlighted when discussing with participants their challenges with interpreting pre-existing visualisations on Strava, particularly surrounding the Fitness & Freshness visualisation. P4 stated "I do not understand it really" or "I do not know what it means to be honest so I do not look at it" (P6). P7 claims "I am not mathematical or techy so maybe I just don't understand [the data] as well as I should?". This suggests if a visualisation is not instantly clear, it will not be embraced by participants.

6.2.5 Theme: Aesthetic Design

Aesthetics is a theme as many participants commented on the aesthetics of the visualisation before being prompted. The most common aspects of the plot that participants commented on without prompt were the use of colour and it's simplicity. P1 stated "the biggest thing for me is the colours, its obviously easy to distinguish which is which [...] you can be overwhelmed with words on screen". P4 stated that anything more complex is confusing and there is a good balance of detail without being overwhelming.

The bubble plot design was also a feature praised, the ability to gather large amounts of information from a glance but with the addition of interactivity to gain more in-depth data was praised. Participants enjoyed the dual nature and differing levels of interpretation. P9 summarises this: "It gives you so much information on one visual very quickly, any other chart is too much reading".

P9 and P3 also raise a point regarding the importance of simplicity both commenting in some way the benefit of a lack of text, reading or manual inputting. Participants enjoyed the drop-down, slider and scroll options rather than typing which prevented excessive effort

from an end user perspective. P1 stated "I like the scale at the bottom, I think people will find that easier to use then putting in a number".

Aesthetics seemed to link with motivation with P6 and P7 commenting that if the visualisation "looks cool" then they are more likely to engage with it. Participants also criticised the number of visualisations and data available in current platforms. P1 commenting "if you click on an activity you've got at least 10 different things to look at [...] it can be quite daunting [...] for one activity you get a hell of a lot of data and I don't think I look at everything". This suggests that one simple, interactive plot is sufficient and participants do not want to be overwhelmed with multiple offerings.

6.2.6 Theme: Motivation

A significant theme identified surrounds the topic of motivation. What motivated participants largely centred on their attitude toward support. P5, a highly experienced athlete did not find the predictions as useful as P6, a novice. The more experienced the athlete, the less they struggled to form routines or felt they needed the support of the visualisation.

Many participants were motivated by intrinsic goals. P2, P8 and P9 were motivated by personal development goals, whereas P5 and P4 were performance driven. Many participants commented on a desire for structure and routine as a form of motivation but struggled to self-impose it. P2 expressed a desire for accountability, stating that their plan is largely in their head which makes it is easy to fall out of routine. More experienced participants were set in behaviours but agreed that the visualisation would be helpful in introducing new behaviours to pre-existing routines and helping with time management.

Participants liked the idea of prediction to help set behaviours, P1 stated "I am not good at setting myself things. I think a lot of people prefer having a structure set for them rather than self-imposing it". P2 also emphasised this saying they would "work harder" if they were being pushed to do so by a prediction. P7 described predictions as a way of "breaking

the barrier or breaking the habit of not doing anything is the hard bit. For something like running when I am doing it on my own [predictions] would help". These comments highlight how predictive visualisations could be useful in breaking down mental barriers to participation by holding users accountable to their exercise and by introducing a cycle of positive reinforcement.

6.2.7 Theme: Flexibility

The importance of flexibility when using planning based visualisations was strongly apparent. Participants gravitated toward the concept of suggestion if they disliked the rigidity of planning day by day. P1 stated they do not like purely adhering to the data and include how they feel as part of their planning process.

All participants have full-time commitments due to work, study or childcare. There was a divide in some participants wishing to put their faith entirely in a predictive visualisation, allowing it to plan their day. Others desired prediction that was fluid, adapting depending on how they are feeling, what goals they have and personal factors such as family time or work commitments. P9 suggested that "weekly buckets" may be more effective for them as it does not matter what specific day they exercise, but a weekly guide would be useful. Many participants described work, or similar full-time commitments as a reason for struggling to form consistent behaviours (P1, P2, P4, P6, P7, P9). P3 also cited consistently keeping to a schedule reinforces behaviours and is positive feedback, empathising the importance of generating a routine.

6.2.8 Theme: Additional Requirements

Participants were asked for any additional functionality requirements, common themes were:

- Being able to add specific training activities such as training types (intervals, easy,

endurance), the inclusion of rest days and building intensity based off goals or competitions (P2, P4, P6, P9).

- Being able to manually override activities for additional flexibility (P3, P8).
- Two-way interaction, such as activities turning green as feedback once complete (P4, P6, P7).
- The option to choose weekly or monthly recommendations instead of daily (P3, P9).

6.3 Summary

The interviews provided in-depth insight into how participants use the existing visualisations for reflection building on the information collected from the surveys.

Participants revealed the different ways that they are motivated to carry out sport which varied depending on their goals and experience. All participants cited structure as a factor for motivation. Participants with less experience enjoyed the rigidity of a strict predicted structure, while more experienced athletes relied more on their own planning but appreciated the accountability or the use of predictions to help them form new behaviours outside their existing routine. The vast majority of participants agreed that a predictive visualisation would support the building of new behaviours by helping introduce or maintain a structure to their training. The more experienced athletes did not always agree, but could understand why less driven athletes would find that support useful.

The interviews also provided insight into additional requirements which can be built into future iterations of the visualisation. The next chapter will discuss the results of both studies and review them in the context of existing literature. I will go on to explore a set of design recommendations for visualisations that support behaviour change as a result of this research.

Chapter 7

Discussion

This chapter will discuss the results across both of the studies, contextualising them with existing research. It will also establish a set of design recommendations as a result of the findings from this project, evaluate limitations and describe future work.

7.1 The Studies

The first study was a survey that established a baseline for iterative development. The project began using my own experiences that were enhanced from the feedback provided within the first study by 41 participants. Participants were asked to evaluate three provided visualisations from Strava and provide an insight into how they currently use sports based data visualisations. This provided a number of requirements that were used in the visualisation design process such as simple design, interactivity and creating an action plan (see Chapter Five for further detail).

The second study was a set of semi-structured interviews with nine participants. They were asked questions to establish their current behaviours, motivations for exercise and understanding of data visualisations, then shown the visualisation I developed and asked to comment on it. The key findings from the interviews emphasised the importance of a

simple aesthetic design and how visualisations are key to supporting exercise at all levels of experience (see Chapter Six for further detail).

7.2 Contextualising the Results

7.2.1 Using Visualisations to Support Behavioural Change

From a behavioural change perspective the studies confirm the effectiveness of persuasive visualisations over reflective ones. Reflective visualisations still hold an important role with multiple participants using simple statistics to help them plan their exercise. Participants praised the use of in-depth and specific, meaningful metrics to engage with the visualisations. Fritz et al.'s study into persuasive technologies concluded that providing concrete data within the visualisation increased the connection made with the visualisation [22], something also witnessed within the second study. This suggests that the balance of providing an easily interpretable visualisation with the option to gather in-depth data by voluntary interaction is important for user engagement. Users can gather information at a glance but also be informed of specific metrics such as duration or activity type by hovering over the bubble. This provides a detailed visualisation for those that require it, or a simple minimalist design for users that do not.

Participants also expressed a barrier to their exercise being motivation or introducing new behaviours to pre-existing routines, something that Conroy et al. investigated [14]. Their study found that setting intentions was a challenge that was not overcome in popular fitness applications. My visualisation facilitated the setting of intention by allowing users to select sports and their frequency of participation, which was praised particularly among the multi-sport participant's within my second study. This suggests that building intentions is a key requirement and that end users need the additional accountability to establish new behaviours, as passive planning is not always enough. Setting the plan and intention within

the visualisation can be an excellent method of gently guiding a user into a new behaviour.

Motivation and routine setting were also found to be fundamentally linked, which is also described by Michie et al. within their COM-B model [34] and reported by Kwasnicka et al. [28]. Participants within my studies frequently stated that establishing the routine was the challenge and that once a routine had been formed, motivation would increase and provide positive reinforcement. This means that creating a visualisation that not only engages a user but is trustworthy enough to guide a user every day is vital to successful behaviour change. Routines can be formed through repeating behaviours and the visualisation must engage and be accurate enough to allow the user to follow it. By using the visualisation regularly and abiding by its recommendation, a positive feedback loop is created and users feel better for completing their recommendation and will continue with the behaviour.

7.2.2 Ensure Users Trust the Visualisation

Another discovery from the studies seen in existing research is trusting visualisations[3]. P4 and P8 commented on their lack of trust in pre-existing Strava visualisations and scepticism over their creation. Participants in my studies highlighted that trust would increase visualisation engagement and use. This is an issue raised by Aldenaini et al. on the topic of over-personalising visualisations, particularly those designed to persuade [3]. This suggests the importance of building trust and confidence in users, as even the most sophisticated or impressive visualisations will be avoided by users if they do not trust the validity of what is presented or if users believe their data is being used for coercion. When it comes to creating visualisations it is important that users understand how they work, this can be done through supporting text. It is important to keep visualisations simple so that any additional information is accessible and not too dense or long as this will also dissuade participants.

7.2.3 The Impact of Aesthetic Design

Aesthetics and visual qualities are shown in studies and existing research to be beneficial to motivating exercise and encouraging habit formation. Bice et al. recognised the importance of highly visual feedback to create a positive feedback loop and the value of registering goal attainment [7]. Participants within my studies liked the use of bright colour and the bold design of bubbles. Participants from the interview confirmed this further by suggesting that further visual feedback be incorporated, such as bubbles turning green once complete, emphasising that users appreciate this form of feedback to engage them. The importance of design is highlighted in the survey participants responses to the Training Calendar versus the Training Log. Both displayed the same information but differ in design (see Figures 4.2 and 4.3 in Chapter Four) with the Training Log's colourful circular designs resulting in positive feedback. This suggests the impact of colour and shape within a visualisation. Different end-users may have differing opinions to design but general rules such as bold colour and visual, rather than tabular presentation can be highly important in user engagement. Universally recognised colours, such as green for positive reward and avoiding red (which can be overtly negative), can also aid the building of a positive feedback loop.

Being overwhelmed by the data available was a re-occurring factor from participants and existing research. Many survey participants replied that they did not use some visualisations because of the sheer volume of options and complexity, resulting in many being overlooked. Lazar et al. cites this as a reason for user abandonment of fitness wearables [29]. This suggests the significance of simple and user-tested visualisations where features are not hidden and it is clear how to get the most from the visualisation. This is especially important to consider within interactive designs where a simple interface could hide the available features. Simple and concise legends can help alleviate this problem. The use of interactivity can prevent users from being overwhelmed, as many features can be included into one design without having multiple visualisations.

A contradiction with existing research within the studies was levels of complexity. Frameworks such as the Elaboration Likelihood model (ELM) of persuasion [40] or research conducted by Borkin et al. [8] suggest that encouraging a high cognitive load increases the engagement and overly simplified visualisations may be more forgettable. However, during interviews participants frequently referred to overly complex visualisations being less useful, P2 citing that "less is more". Many participants stated the same assertions and if a visualisation was not immediately interpretable they would not use it. Participants criticised Strava for offering too much complexity and preferred a simpler design. Borkin et al. research centred on memorability, which was not a factor accurately tested within this project. However, as participants suggested that they would not engage with a more complex visualisation in the first place, it can still be considered an interesting contradiction. Participant preference fell more within the boundaries of Tufte's rules [55], where simplicity and a good Data:Ink ratio is maintained. Participants did however comment on the benefits of colour and the aesthetic elements. These contradictions suggest that a balance must be negotiated between providing a visualisation that is unique and engaging with colour and aesthetic design, but is not overly cluttered or distracting. The aesthetic elements and overall design must be interpreted by the end user as useful and significant rather than colourful without reason, which leads to the perception of over-complexity. Making a visualisation memorable is not necessary based on design alone and that the data presented within the visualisation is a significant contributor to memorability and usefulness, something that Borkin et. al's study did not consider. If the data is perceived useful to users, the design must facilitate easy access and interpretation of that data in order to be engaging and useful.

7.2.4 Comparing the needs of Ex-Professional and Novice Users

Participant feedback revealed there was an issue in certain participant groups with motivation and actioning their data. Participants with lower levels of experience in exercise stated

they were overwhelmed by large amounts of data. Existing research in personal informatics also suggests this consequence, with Rawassizadeh et al. highlighting that a lack of "compelling" data visualisations is the cause [43]. Some participants used Strava for the social motivation, particularly the female participants. Others cited a lack of data literacy which hindered taking fully advantage of their data.

A key factor in affecting motivation and behaviour change is experience, which is especially prominent in the area of sport and fitness, but is rarely considered in existing research. Levels of expertise and complexity have been considered in application interface design [41] but not explicitly with data visualisation design for behavioural change in physical activity. The design guidelines created from this project help to fill an area of research that is not highly saturated. A key output of this project are design recommendations for data visualisations that support behavioural change in physically active users, regardless of their experience levels.

The participants within the studies ranged broadly in ability and background. Originally, I believed that personality or data literacy would be a significant factor in engagement with the visualisation. However, the greatest factor was experience. An ex-elite athlete did not find the predictions useful, as they had the experience to make suitable plans. Whereas someone new to regular participation may not know how to maintain a consistent action driven training plan. As a result, different types of visualisations could be created for different target audiences. For example an elite athlete may appreciate being told what types of training they should do rather than merely the sport.

7.3 Design Recommendations

As a result of this research, the following guidelines have been proposed for producing predictive data visualisations that aim to influence behaviour change for active participants.

- Instil trust and integrity.

- Allow flexibility.
- Design aesthetics.
- Create a feedback loop.

7.3.1 Trust and Integrity

Trust and integrity in the visualisation is essential. Participants within the survey and interview frequently expressed doubts in pre-existing Strava visualisations, questioning the accuracy of the predictions or expressing confusion over the process of generating the plots. Future researchers should aim for transparency in both the data and the prediction process to ensure participants are satisfied with the quality and validity of the visualisations. This is particularly important with elite or semi-experienced athletes, who are more performance driven and small margins in training can have a big impact on their race performance. When participants are using data visualisations they have a strong desire to ensure that the visualisations are correctly processing their data and will benefit them in the long term. The best way of ensuring this is by using simple methods, allowing for easy explanations to lay audiences. However, this may impact the accuracy so more complex methods may be required. In this case supporting material should be included, but this must be done sparingly as participants dislike having to read large volumes of text.

7.3.2 Flexibility

Another key factor is the element of flexibility. Many participants were multi-sport athletes and balance many sports with other full-time commitments. The predictions provided to users need to have a broad level of flexibility to allow the user to use the predictions as a guide to impose structure and promote behaviour change. This can be incorporated by providing manual overrides or by providing weekly suggestions for activities and allow users

to plan which days they do the recommendations. Alternatively, the visualisation could use artificial intelligence (AI) to factor in additional features such as a work diary in order to more accurately predict the visualisations. However, this could potentially hinder trust and integrity due to the complex decision-making process associated with AI.

7.3.3 Aesthetics

Aesthetic quality and purposeful design are important guidelines. The balance of creating a visualisation that has high visual elements, but not overly complex is a challenge that must be overcome. Participants enjoyed the simplicity of the design created using the bubble plot as well as use of bold colours. A visualisation must be easily interpretable and information should be available at a glance. There should be little need for explanation on its purpose or functionality. However, different levels of use should be incorporated. The visualisation designed was simple to interpret with the option of further detail, such as specific timings or activities accessed through interactivity. Interactivity is essential for user engagement and allows for additional information to be displayed without excess clutter. Interactivity and simple but bold aesthetic qualities result in an engaging visualisation that participants can use without aggravation.

7.3.4 Feedback Loop

Creating a feedback loop is an important factor in two parts of development. Firstly, interactive design ensures the product is designed for users by using feedback from multiple evaluations during user studies. Secondary, an end user feedback loop promotes formation of behaviour change by use of positive reinforcement. Prediction can form positive reinforcement when users complete what is suggested, but further reinforcement can be implemented from colour changes or simple non-intrusive notifications that allow the user to gain a sense of achievement. When users do not achieve their predictions, encouragement and re-calculation

of the prediction according to their goals is essential to maintain motivation and avoid abandonment. Being highly conscious of the user-base experience is vital, as different levels of athletes will require different motivational factors. Those of an elite level will require less support than a novice.

7.3.5 Design Recommendations in Practice

Using the feedback from the user study further mock-up designs were created to further improve the visualisation as part of an iterative design process. These figures (7.1 and 7.2) demonstrate how the design recommendations could be applied in practice.

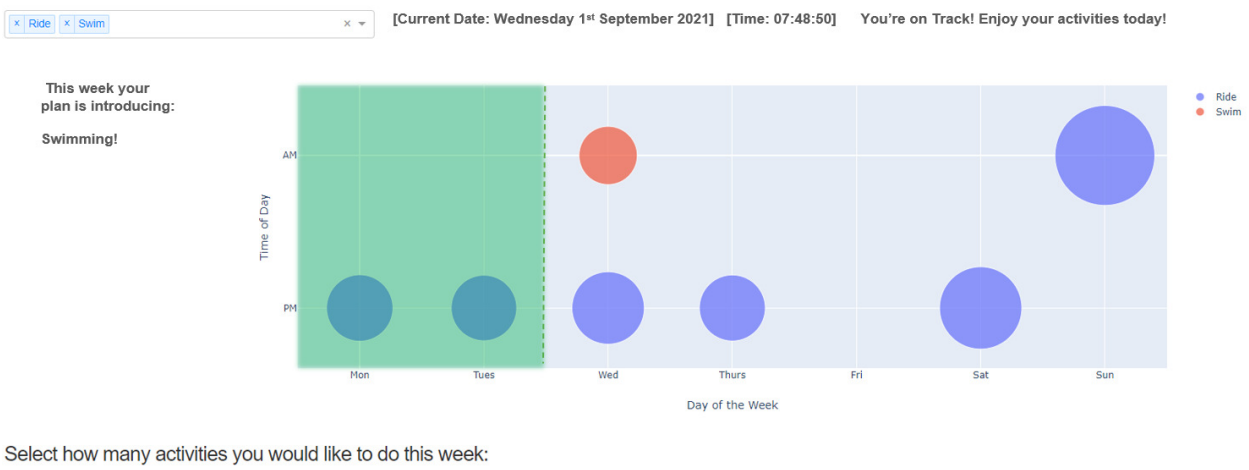


Figure 7.1: Iteration of design incorporating attainment feedback and building to goals for novice users.

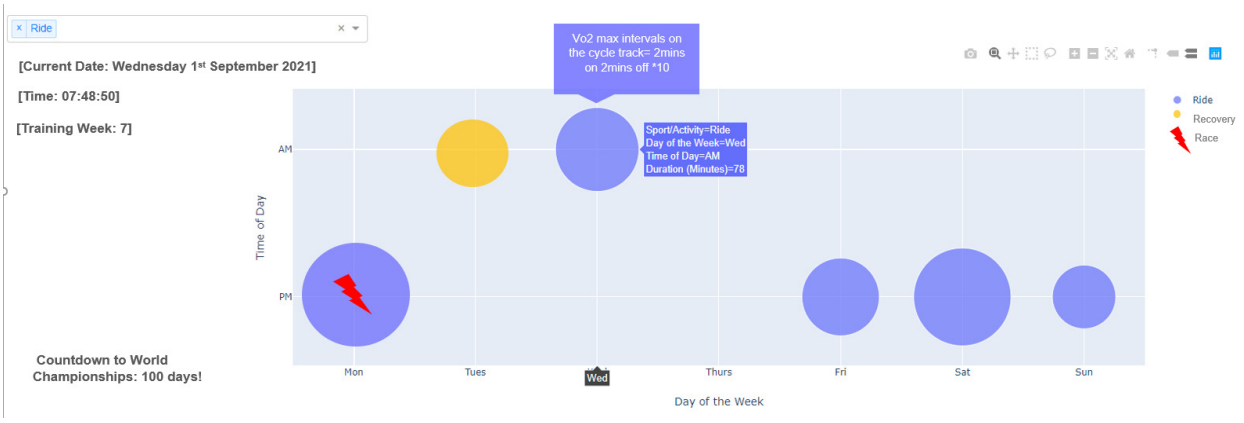


Figure 7.2: Iteration of design incorporating specific sessions, races and rest days for experienced users.

7.4 Limitations & Future Work

The main limitations to this research were the sample size and timescale. Although the recommended number of participants were reached for both studies, stronger conclusions could be made from a greater number of participants and their feedback. However, the data collected within the project was sufficient to draw conclusions. The project spanned a limited time and therefore planning, recruitment, design and evaluation had to be completed very quickly. With an extended period of time, more participants could be included in the study and more iterations could have been performed to provide further insight into the visualisations and their design. With further time I would produce more visualisations and present contrasting designs to measure the most effective concepts and iterate further for refinement. The evaluation of behaviour change could then have been made quantitively by collected and analysing more data as, opposed to qualitatively using interviews. Another opportunity could be to investigate how athlete experience levels, personality types or sports influence the preferred visualisation design or prediction type.

The studies were somewhat limited by my recruitment process, opting to use social media and word of mouth. This resulted in most participants competing in the same sports

(swimming, cycling, running or triathlon) and therefore this research is limited to aiding athletes from a few sports. Future research could include a broader variety to evaluate the impact of the sport on prediction requirements. However, as triathlon is a time-consuming sport and all participants in my studies did at least one of the three sports within a triathlon, the impact of predictions and the benefits of the visualisation related are still highly valuable and including further wildly different sports may result in confusing and irrelevant findings as the training needs would vary. In addition, the use of social media could have biased the demographics of the recruited participants, but the statistics did not indicate this.

Using Python based visualisation libraries was somewhat limiting in contrast to the creative freedom of the JavaScript library D3. Python was chosen for visualisation for ease of use thanks to its simplicity leading to faster development. If the project was longer and further detail could be added to the visualisation, D3 would be better suited for creating more detailed visualisations. However, for the short time frame Python was more than suitable.

Future work on this project would be best centred on improving the quality of the predictions. For participants new to fitness highly accurate visualisations are incredibly valuable. I lacked the time to successfully implement highly sophisticated predictions and centred more on the design process. I limited data collection to the last 365 days which was sufficient to produce predictions that participants recognised as their own data, however, on reflection being provided with all available data from each participant would allow for an increased accuracy in these predictions. Artificial intelligence would be an excellent future work opportunity to explore how machine learning could be used to produce a predictive model rather than use of descriptive statistics. This would produce more accurate recommendations. Features for the model could include weather, location, social or work calendars.

Chapter 8

Conclusions

This dissertation set out to gain an understanding into the factors that impact behavioural change in exercise. By reviewing existing research a re-occurring pattern of challenges emerged. In an increasingly technological world, users collect their own data but they do not always make use of it. This is especially true with exercise data. A void opens between data collection and actioning it to enact behavioural change. To overcome the problem, this project researched the impact of predictive visualisations as a way of encouraging users to engage and action their data to a higher degree and facilitate behavioural change.

As a human centred project, an iterative design process was established to ensure that the visualisation was designed with the users at the forefront. I began with my own feedback on the pre-existing platform Strava and selected three visualisations to use as a foundation in my first study, an online survey. Using feedback from participants in the study, partnered with existing research into visualisation and behaviour change theory, I designed and developed a series of visualisation concepts. Using a preliminary study, a visualisation was refined before being shown to another set of participants. The second study involved semi-structured interviews and explored participants current behaviours, motivations and opinions on the visualisation I created. Using interview feedback further refinements were made to the visualisation and a four-point design recommendations were created.

The studies confirmed many aspects explored in existing literature such as being overwhelmed with the quantity of visualisations available, the importance of interactivity for engagement and ensuring user trust. There were areas of contradiction particularly around the aspect of complexity with research suggesting more complex visualisations can be more engaging, when my studies suggested that simple visualisations had more impact. New areas of research arose through investigating the impact of user experience impacting their requirements for data visualisations. The ex-professional athletes within the studies had a differing level of requirements to novice participants.

Based off the research highlighted above, I devised a set of design recommendations to guide future researchers. They include: the importance of building trust and integrity with end users, allowing for flexibility, carefully considered design aesthetics and the creation of a feedback loop. The iterative design process was highly effective and ensured that user needs were catered for. Future work could build on this project by incorporating further user studies for continued refinement or improving the accuracy of the predictions through artificial intelligence.

This project has emphasised the substantial impact that data visualisations have on engagement and action. It has highlighted that copious data can be collected, but without meaningful visualisation data is meaningless. This project has shown that introducing structure to users can be highly motivating and if developers produce simple, engaging, predictive visualisations that cater to the fitness goals of the user, it significantly improves their chances of taking positive steps towards behavioural change. It is vital that users believe and trust what they are being shown, with the right levels of engagement and a solid action plan embedded within the visualisation. Whether the user is an ex-elite athlete or someone running their first five kilometres, we can empower users and help them transform their data into action.

Chapter 9

Reflection

The process of writing this project has required a considerable amount of personal and academic development. This chapter will reflect on the things learnt through the act of research, areas for improvement and knowledge acquired that will be taken forward to future work.

The first skill tested throughout this project was the concept of time and project management. This research was heavily restricted by time, especially in the final weeks before completion where I began full-time employment. Twelve weeks initially seemed like a long time but the nature of a HCI project, heavily dependent on user studies, meant that time had to be carefully planned not only for my own research and the schedule of my supervisor, but also considering my participants. I found this process incredibly challenging but was able to overcome the problem by using recognised project management tools such as Gantt charts and an adaptable timeline. This allowed me to plan a list of required activities and dynamically move them around depending on how long they were taking. The Gantt chart allowed for the overlapping of commitments and was vital to the success of the project. I would rely heavily on this method in future work. The key to my successful project management lay in a flexible approach, being honest with myself at the status and time each process within the project took. Knowing I was to start work before the project was due,

I worked to a soft deadline to ensure that I would have enough of the project completed to balance work and dissertation commitments. Ultimately I was able to create a pattern of proofreading before work and able to get the project completed on time.

A strength going into the project was my organisational skills, balancing when interviews were with creating the correct visualisation and beginning my writing. I was able to balance many tasks simultaneously which was thanks to my project management experience. I also have a significant personal drive and motivation. I spent extra time formulating a project I was deeply interested in to aid my motivation. I am also a highly driven person and was able to use this skill to complete the project independently. This work can be isolating and I made an effort to keep communication open with my supervisor and keep up to date with all work which lent to the success of the project.

A key lesson learnt in partnership with my supervisor is being wary of the validity of researchers claims. This project involved research into multiple academic fields and therefore the required literature review and reading was vast. As a student researcher I learnt that it is vital to not immediately trust academic work, just because it has been published. When citing research I have learnt to value the importance of identifying concrete research, often in the form of non-biased repeatable studies that back up their claims. I initially used old research that has since been proven false within the behavioural change section of my literature review. Working with my supervisor I was able to identify the problem and find more reliable research to compliment my arguments. It taught me to cast a critical eye on all research, even if it seems legitimate on the surface or has been cited by others. It is important that I do my own research to evaluate the claims and authority of other academics.

The two-study iterative process that my project followed and the effective design methodology were key success to the project. By carrying out an initial online survey I was able to get a baseline understanding of the kind of responses from participants, allowing for a more in depth discussion when interviewing participants. Carrying out a pilot study be-

fore the interviews was also highly successful as it allowed for further refinement with the visualisation I created and highlighted some improvements for my questioning style within the interviews. I was incredibly nervous carrying out the interviews as I feared not asking the correct questions or participants not answering successfully, so having the trial was a great way of building confidence. This resulted in interviews that were highly thought-out and purposely linked to my research question. Carrying out the user studies was incredibly time-consuming so it was vital that when they were carried out they were purposeful and well-designed, which the pilot studies allowed. In future work I will remember the importance of trials and pilots, including more of them to ensure that the big studies and major components of my research are perfected to avoid wasted opportunities with participants.

Carrying out this project assisted with continued personal development as it pushed me beyond my comfort zone, particularly when it came to interacting with my participants. Interviewing participants individually was a daunting prospect and I was able to practice my interview skills and develop interpersonal skills such as communication and mediating conversations with participants. The interviews involved forming a comfortable relationship with participants and allowing them to speak their views but also ensure that they remained on topic so time with them was used efficiently. I was able to cultivate a positive mental attitude throughout the interview process forming great relationships to get the best out of my interview participants despite my initial anxiety at the prospect of individual interviews. In future projects I will not fear this process thanks to personal development within this project.

My coding process faced the most problems throughout the project, partially due to my lack of experience but also due to my fear of errors or facing unsolvable problems. Throughout I created mock-up visualisations and concept designs which were very useful for the development process but I faced problems when it came to enhancing the code for more advanced iterations as I feared breaking pre-existing code. I was able to overcome

this through version controlling as it allowed to create a version that functioned and then continue to develop it without fear of losing the original. Breaking through the mental barrier of achieving "perfection" also helped with the progress as I realised that the act of coding is iterative. Overcoming the fear of development and mistakes was a massive breakthrough and is something I will carry forward in future research but also in my employment as a full-time engineer.

Throughout the preparatory reading and literature review for this project I discovered many new aspects that will be useful for future work. Many of these surround behavioural change theory and what end users consider when using a device or service. The most significant is the degree to which users are concerned with the accuracy and perceived validity of the data shown to them. Rather than trusting technology I have read studies and seen within my own participants how important it is to establish trust in applications, particularly if they are aiming to influence users' behaviour. These findings have made me more aware of my own trust in technology and also spending extra consideration on openness and transparency with future applications or visualisations that I design.

This project has been a highly developmental process, from discovering new academic fields, to the importance of evaluating research, to breaking down personal boundaries. There are aspects I will repeat in future work, such as my organisation and project management and things I will change such as adding more pilot studies to aid confidence and improve my processes. The most valuable lesson I will take forward is to have confidence and not aim for perfection all the time in the first instance. Computer science is a developmental process and I will take forward the importance of trusting my ability and to keep striving for improvements one step at a time.

Bibliography

[1] Anne Adams and Anna L. Cox. "Questionnaires, in-depth interviews and focus groups". In: *Research Methods for Human-Computer Interaction*. Ed. by Anna L. Cox and Paul Cairns. Cambridge: Cambridge University Press, 2008, pp. 17–34. ISBN: 978-0-521-69031-7. DOI: 10.1017/CBO9780511814570.003. URL: https://www.cambridge.org/core/books/research-methods-for-humancomputer-interaction/questionnaires-indepth-interviews-and-focus-groups/FA5B7278B9A5E489AFA83A030C816EB9 (visited on 08/25/2021).

[2] Anne Adams, Peter Lunt, and Paul Cairns. "A qualititative approach to HCI research". en. In: *A qualititative approach to HCI research*. Ed. by Paul Cairns and Anna Cox. Cambridge, UK: Cambridge University Press, 2008, pp. 138–157. ISBN: 978-0-521-87012-2. URL: http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521870122&ss=toc (visited on 08/26/2021).

[3] Noora Aldenaini et al. "Trends in Persuasive Technologies for Physical Activity and Sedentary Behavior: A Systematic Review". English. In: *Frontiers in Artificial Intelligence* 3 (2020). Publisher: Frontiers. ISSN: 2624-8212. DOI: 10.3389/frai.2020.00007. URL: https://www.frontiersin.org/articles/10.3389/frai.2020.00007/full (visited on 07/13/2021).

[4] Kieran Alger. *Garmin Connect review*. Oct. 2015. URL: https://www.wareable.com/garmin/garmin-connect-review (visited on 07/09/2021).

[5] Ankur Barua. "METHODS FOR DECISION-MAKING IN SURVEY QUESTIONNAIRES BASED ON LIKERT SCALE". en. In: *Journal of Asian Scientific Research* (2013), p. 5.

[6] Abha Belorkar. *Interactive data visualization with python: present your data as an effective and compelling story*. eng. Second edition. Birmingham, England: Packt, 2020. ISBN: 978-1-5231-3642-1.

[7] Matthew R. Bice, James W. Ball, and Steve McClaran. "Technology and physical activity motivation". In: *International Journal of Sport and Exercise Psychology* 14.4 (Oct. 2016). Publisher: Routledge _eprint: https://doi.org/10.1080/1612197X.2015.1025811, pp. 295–304. ISSN: 1612-197X. DOI: 10.1080/1612197X.2015.1025811. URL: https://doi.org/10.1080/1612197X.2015.1025811 (visited on 08/24/2021).

[8] Michelle A. Borkin et al. "What Makes a Visualization Memorable?" In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (Dec. 2013). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 2306–2315. ISSN: 1941-0506. DOI: 10.1109/TVCG.2013.234.

[9] Virginia Braun and Victoria Clarke. "Using thematic analysis in psychology". In: *Qualitative Research in Psychology* 3.2 (Jan. 2006). Publisher: Routledge eprint, pp. 77–101. ISSN: 1478-0887. DOI: 10.1191/1478088706qp063oa. URL: https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa (visited on 08/27/2021).

[10] Viktoria Brüggemann, Mark-Jan Bludau, and Marian Dörk. "The Fold: Rethinking Interactivity in Data Visualization". en. In: *DHQ: Digital Humanities Quarterly* 14.3 (2020), p. 16. (Visited on 08/05/2021).

[11] Mackinlay Card. *Readings in Information Visualization: Using Vision to Think*. en. Google-Books-ID: wdh2gqWfQmgC. Morgan Kaufmann, Feb. 1999. ISBN: 978-1-55860-533-6.

[12] Luis V. Casalo. "Be creative, my friend! Engaging users on Instagram by promoting positive emotions". en. In: *Journal of Business Research* 130 (June 2021). Publisher: Elsevier, pp. 416–425. ISSN: 0148-2963. DOI: 10.1016/j.jbusres.2020.02.014. URL: [http://www.sciencedirect.com/science/article/pii/S0148296320301089] (http://www.sciencedirect.com/science/article/pii/S0148296320301089) (visited on 08/02/2021).

[13] Philip J. Cash, Charlotte Gram Hartlev, and Christine Boysen Durazo. "Behavioural design: A process for integrating behaviour change and design". en. In: *Design Studies* 48 (Jan. 2017), pp. 96–128. ISSN: 0142-694X. DOI: 10.1016/j.destud.2016.10.001. URL: [https://www.sciencedirect.com/science/article/pii/S0142694X16300667] (https://www.sciencedirect.com/science/article/pii/S0142694X16300667) (visited on 07/13/2021).

[14] David E. Conroy, Chih-Hsiang Yang, and Jaclyn P. Maher. "Behavior Change Techniques in Top-Ranked Mobile Apps for Physical Activity". en. In: *American Journal of Preventive Medicine* 46.6 (June 2014), pp. 649–652. ISSN: 0749-3797. DOI: 10.1016/j.amepre.2014.01.010. URL: https://www.sciencedirect.com/science/article/pii/S0749379714000403 (visited on 08/24/2021).

[15] Jesse Couture. "Reflections from the 'Strava-sphere': Kudos, community, and (self-)surveillance on a social network for athletes". In: *Qualitative Research in Sport, Exercise and Health* 13.1 (Jan. 2021). Publisher: Routledge eprint, pp. 184–200. ISSN: 2159-676X. DOI: 10.1080/2159676X.2020.1836514. URL: https://doi.org/10.1080/2159676X.2020.1836514 (visited on 08/10/2021).

[16] Strava Developers. *Fitness & Freshness*. 2021. URL: https://www.strava.com/dashboard (visited on 07/09/2021).

[17] Strava Developers. *Fitness & Freshness: How it Works*. English. Apr. 2021. URL: https://support.strava.com/hc/en-us/articles/216918477-Fitness-Freshness (visited on 08/12/1935).

[18] Strava Developers. *Getting Started with the Strava API*. (Visited on 08/19/2021).

[19] Zach Emmanuel. *UK Wearable Technology Market Report 2020*. en. 2020. URL: https://store.mintel.com/uk-wearable-technology-market-report (visited on 09/29/2021).

[20] B. Fogg. "Persuasive computers: perspectives and research directions". In: *CHI* (1998). DOI: 10.1145/274644.274677.

[21] Bjarte Folkestad. *Analysing Interview Data Possibilities and challenges*. en. Tech. rep. 13. Publication Title: EUROSPHERE Working Paper Series (EWP). Eurospheres project, Dec. 2008. URL: https://ideas.repec.org/p/erp/ewpxxx/p0037.html (visited on 08/27/2021).

[22] Thomas Fritz et al. "Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness". en. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Toronto Ontario Canada: ACM, Apr. 2014, pp. 487–496. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557383. URL: https://dl.acm.org/doi/10.1145/2556288.2557383 (visited on 06/02/2021).

[23] Benjamin Gardner et al. "How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults". In: *Health Psychology Review* 10.1 (Jan. 2016), pp. 89–112. ISSN: 1743-7199. DOI: 10.1080/17437199.2015.1082146. URL: [https://www.ncbi.nlm.nih.gov/pmc/articles/

PMC4743603/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4743603/) (visited on 07/13/2021).

[24] Garmin. *Garmin Connect*. en. URL: https://connect.garmin.com/ (visited on 09/09/2021).

[25] Lars Hallnäs and Johan Redström. "Slow Technology – Designing for Reflection". en. In: *Personal and Ubiquitous Computing* 5.3 (Aug. 2001), pp. 201–212. ISSN: 1617-4909. DOI: 10.1007/PL00000019. URL: https://doi.org/10.1007/PL00000019 (visited on 08/24/2021).

[26] Helen Kennedy. *Impact of Social Sciences Blog: Simple data visualisations have become key to communicating about the COVID-19 pandemic, but we know little about their impact*. en. LSE. May 2020. URL: https://blogs.lse.ac.uk/impactofsocialsciences/2020/05/04/simple-data-visualisations-have-become-key-to-communicating-about-the-covid-19-pandemic-but-we-know-little-about-their-impact/ (visited on 08/05/2021).

[27] Pankaja Pradeep Kumar. "Effective Use of Gantt Chart for Managing Large Scale Projects". English. In: *Cost Engineering* 47.7 (July 2005). Num Pages: 8 Place: Morgantown, United States Publisher: American Association of Cost Engineers Section: CERTIFICATION PAPER, pp. 14–21. ISSN: 02749696. URL: http://www.proquest.com/docview/220442584/abstract/588BC70DFF43413DPQ/1 (visited on 08/19/2021).

[28] Dominika Kwasnicka et al. "Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories". In: *Health Psychology Review* 10.3 (July 2016). Publisher: Routledge eprint, pp. 277–296. ISSN: 1743-7199. DOI: 10.1080/17437199.2016.1151372. URL: [https://doi.org/10.1080/17437199.

2016.1151372](https://doi.org/10.1080/17437199.2016.1151372) (visited on 07/13/2021).

[29] Amanda Lazar et al. "Why we use and abandon smart devices". In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '15. New York, NY, USA: Association for Computing Machinery, Sept. 2015, pp. 635–646. ISBN: 978-1-4503-3574-4. DOI: 10.1145/2750858.2804288. URL: http://doi.org/10.1145/2750858.2804288 (visited on 08/10/2021).

[30] Ian Li, Anind Dey, and Jodi Forlizzi. "A stage-based model of personal informatics systems". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, Apr. 2010, pp. 557–566. ISBN: 978-1-60558-929-9. URL: http://doi.org/10.1145/1753326.1753409 (visited on 08/12/2021).

[31] Ian Li, Jodi Forlizzi, and Anind Dey. "Know thyself: monitoring and reflecting on facets of one's life". In: *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, Apr. 2010, pp. 4489–4492. ISBN: 978-1-60558-930-5. URL: http://doi.org/10.1145/1753846.1754181 (visited on 07/13/2021).

[32] S. J. Maan et al. "Making it not too obvious : the effect of ambient light feedback on space heating energy consumption". English. In: *Energy Efficiency* 4.2 (2011). Publisher: Springer, pp. 175–183. DOI: 10.1007/s12053-010-9102-6. URL: https://research.tue.nl/en/publications/making-it-not-too-obvious-the-effect-of-ambient-light-feedback-on (visited on 07/14/2021).

[33] Amol Mavuduru. *Is Data Really the New Oil in the 21st Century?* en. Dec. 2020. URL: https://towardsdatascience.com/is-data-really-the-new-oil-in-the-21st-century-17d014811b88 (visited on 09/29/2021).

[34] Susan Michie, Maartje M. van Stralen, and Robert West. "The behaviour change wheel: A new method for characterising and designing behaviour change interventions". en. In: *Implementation Science* 6.1 (Dec. 2011). Number: 1 Publisher: BioMed Central, pp. 1–12. ISSN: 1748-5908. DOI: 10.1186/1748-5908-6-42. URL: http://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-6-42 (visited on 09/09/2021).

[35] Tomoko Nemoto and David Beglar. "Developing Likert-Scale Questionnaires". en. In: *Developing Likert Scale Questionnaires*. Kobe, Japan, 2013, p. 8.

[36] J. Nielsen. "Iterative user-interface design". In: *Computer* 26.11 (Nov. 1993). Conference Name: Computer, pp. 32–41. ISSN: 1558-0814. DOI: 10.1109/2.241424.

[37] Anshul Vikram Pandey et al. "The Persuasive Power of Data Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 2014). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 2211–2220. ISSN: 1941-0506. DOI: 10.1109/TVCG.2014.2346419.

[38] Endeavour Partners. *Inside Wearables Part 1: How behavior change unlocks long-term engagement*. English. Apr. 2017. URL: https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3 (visited on 08/09/2021).

[39] Training Peaks. *Training Peaks*. 2021. URL: https://www.trainingpeaks.com/ (visited on 09/09/2021).

[40] Richard Petty. *Communication and Persuasion - Central and Peripheral Routes to Attitude Change*. 1st ed. New York, NY, USA: Springer-Verlag, 1986. ISBN: 978-1-

4612-9378-1. URL: https://www-springer-com.abc.cardiff.ac.uk/gp/book/9781461293781 (visited on 07/13/2021).

[41] Vesna Popovic. "Expert and Novice User Differences and Implications for Product Design and Useability". en. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 44.38 (July 2000). Publisher: SAGE Publications Inc, pp. 933–936. ISSN: 2169-5067. DOI: 10.1177/154193120004403869. URL: https://doi.org/10.1177/154193120004403869 (visited on 09/26/2021).

[42] Amon Rapp and Maurizio Tirassa. "Know Thyself: A Theory of the Self for Personal Informatics". In: *Human–Computer Interaction* 32.5-6 (Nov. 2017). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07370024.2017.1285704, pp. 335–380. ISSN: 0737-0024. DOI: 10.1080/07370024.2017.1285704. URL: https://doi.org/10.1080/07370024.2017.1285704 (visited on 08/12/2021).

[43] Reza Rawassizadeh et al. "Lesson Learned from Collecting Quantified Self Information via Mobile and Wearable Devices". en. In: *Journal of Sensor and Actuator Networks* 4.4 (Nov. 2015), pp. 315–335. ISSN: 2224-2708. DOI: 10.3390/jsan4040315. URL: http://www.mdpi.com/2224-2708/4/4/315 (visited on 06/02/2021).

[44] Dominik Sacha et al. "Visual Interaction with Dimensionality Reduction: A Structured Literature Analysis". In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (Jan. 2017). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 241–250. ISSN: 1941-0506. DOI: 10.1109/TVCG.2016.2598495.

[45] David L. Schriger and Richelle J. Cooper. "Achieving graphical excellence: Suggestions and methods for creating high-quality visual displays of experimental data". en. In: *Annals of Emergency Medicine* 37.1 (Jan. 2001), pp. 75–87. ISSN: 0196-0644. DOI: 10.1067/mem.2001.111570. URL: [https://www.sciencedirect.com/science/

article/pii/S0196064401750639](https://www.sciencedirect.com/science/article/pii/S0196064401750639) (visited on 08/02/2021).

[46] Science4Performance. *Fitness and Freshness*. May 2017. URL: https://science4performance.com/2017/05/05/strava-fitness-and-freshness/ (visited on 08/13/2021).

[47] Harri Siirtola. "Bars, Pies, Doughnuts & Tables – Visualization of Proportions". In: *BCS Learning & Development* (Sept. 2014). Publisher: BCS Learning & Development. DOI: 10.14236/ewic/HCI2014.38. URL: https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/HCI2014.38 (visited on 08/05/2021).

[48] Strava. *Strava*. 2021. URL: https://www.strava.com/dashboard (visited on 09/09/2021).

[49] Miriam Sturdee and Joseph Lindley. "Sketching & Drawing as Future Inquiry in HCI". In: *Proceedings of the Halfway to the Future Symposium 2019*. HTTF 2019. New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 1–10. ISBN: 978-1-4503-7203-9. DOI: 10.1145/3363384.3363402. URL: http://doi.org/10.1145/3363384.3363402 (visited on 08/27/2021).

[50] Zhaoyuan Su et al. "Analyzing Description, User Understanding and Expectations of AI in Mobile Health Applications". In: *AMIA Annual Symposium Proceedings* 2020 (Jan. 2021), pp. 1170–1179. ISSN: 1942-597X. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8075490/ (visited on 07/13/2021).

[51] Jakob Tholander and Maria Normark. "Crafting Personal Information - Resistance, Imperfection, and Self-Creation in Bullet Journaling". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–13. ISBN: 978-1-4503-6708-0. URL: http://doi.org/10.1145/3313831.3376410 (visited on 08/12/2021).

[52] Tilde Ann Thurium. *Getting Started With The Strava API: A Tutorial*. en. Sept. 2017. URL: https://medium.com/@annthurium/getting-started-with-the-strava-api-a-tutorial-f3909496cd2d (visited on 08/19/2021).

[53] TomTom. *MyRun*. URL: https://mysports.tomtom.com/app/ (visited on 09/09/2021).

[54] Huong Ly Tong and Liliana Laranjo. "The use of social features in mobile health interventions to promote physical activity: a systematic review". In: *NPJ Digital Medicine* 1 (Sept. 2018), p. 43. ISSN: 2398-6352. DOI: 10.1038/s41746-018-0051-3. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6550193/ (visited on 07/13/2021).

[55] Edward Tufte. *The Visual Display of Quantitative Information (1983)*. de. Pages: 219-230 Publication Title: Diagrammatik-Reader Section: Diagrammatik-Reader. De Gruyter (A), Oct. 2016. ISBN: 978-3-05-009383-3. URL: http://www.degruyter.com/document/doi/10.1515/9783050093833-035/html (visited on 08/02/2021).

[56] Edward R. Tufte. *The visual display of quantitative information*. eng. 2nd ed. Cheshire, CT: Graphics Press, 2001. ISBN: 9780961392147.

[57] Matthew O. Ward, Georges Grinstein, and Daniel Keim. *Interactive Data Visualization: Foundations, Techniques, and Applications*. en. Google-Books-ID: Kk7NBQAAQBAJ. CRC Press, June 2010. ISBN: 978-1-4398-6554-5.

[58] Gary Wolf. *The quantified self*. en. 2010. URL: https://www.ted.com/talks/gary_wolf_the_quantified_self (visited on 08/12/2021).

[59] Mike Yi. *A Complete Guide to Bubble Charts*. en. Oct. 2019. URL: https://chartio.com/learn/charts/bubble-chart-complete-guide/ (visited on 08/11/2021).

Appendix A

Survey Participation Information Sheet

PARTICIPANT INFORMATION SHEET

CONQUERING THE VOID: TRANSFORMING DATA INTO ACTION

You are being invited to take part in a research project. Before you decide whether or not to take part, it is important for you to understand why the research is being undertaken and what it will involve. Please take time to read the following information carefully and discuss it with others, if you wish.

Thank you for reading this.

1. What is the purpose of this research project?

I am undertaking this research as part of my master's degree as it forms my final project: the dissertation. The aims of this study are to narrow the gap in wearable technologies between collecting data on smart devices such as watches or bike computers and taking tangible steps in using that data to form better habits when it comes to your physical activity. Current research suggests that data is collected, uploaded, and then forgotten about. There is a wealth of insight you can gather from your data, and I would like to investigate how different types of data, and how it is presented to you, can impact your habit formation.

2. Why have I been invited to take part?

You have been invited because you are user of the application Strava, you partake in exercise and have done for at least 4 weeks. You also have no academic education on the topic of data visualisation concepts or theories.

3. Do I have to take part?

No, your participation in this research project is entirely voluntary and it is up to you to decide whether or not to take part. If you decide to take part, we will discuss the research project with you and ask you to sign a consent form. If you decide not to take part, you do not have to explain your reasons and it will not affect your legal rights. Please note that if you are a student at Cardiff University you are welcome to participate but your involvement in this project will have no effect on your education or progression through your degree course.

You are free to withdraw your consent to participate in the research project at any time, without giving a reason, even after signing the consent form.

4. What will taking part involve?

Your involvement in the study will be answering a short survey on your current opinions, experiences, and any changes you make to your lifestyle because of the application Strava. This will be only carried out once and it should take you no more than 15 minutes maximum. You will be entirely anonymous as part of the survey; no identifiable information will be taken. You will be asked to provide a key word, that should not identify you, that you can use to withdraw your data from the study at any point should you wish. At the end of the survey, you will be asked if you wish to participate in the second half of the study for further research, there

is no obligation to do so. If you wish to find out more about this, see the additional information sheet.

5. Will I be paid for taking part?
No. You should understand that any data you give will be as a gift and you will not benefit financially in the future should this research project lead to the development of new methods of data presentation.

6. What are the possible benefits of taking part?
There will be no direct advantages or benefits to you from taking part, but your contribution will help me understand how data presentation can impact habit formation.

7. What are the possible risks of taking part?
There are no foreseeable discomforts, risks, or disadvantages for taking part in this survey.

8. Will my taking part in this research project be kept confidential?
All information collected from (or about) you during the research project will be kept confidential and any personal information you provide will be managed in accordance with data protection legislation. Please see 'What will happen to my Personal Data?' (below) for further information.

9. What will happen to my Personal Data?
This survey will not take any identifying information, when you submit your answers, you will be anonymised and assigned a name that is not your own, "Athlete 1", "Athlete 2" etc. I will also collect your gender and age-band, purely to ensure I have a range of demographics, you may choose to opt out of these questions should you wish. The only personal information that may be collected is your email address should you opt into the second half of the study so I may contact you. There is no obligation to opt into the second half of the study – see additional information sheet. For this survey, I am only interested in recording your opinion, not personal information.

Cardiff University is the Data Controller and is committed to respecting and protecting your personal data in accordance with your expectations and Data Protection legislation. Further information about Data Protection, including:

- your rights
- the legal basis under which Cardiff University processes your personal data for research
- Cardiff University's Data Protection Policy
- how to contact the Cardiff University Data Protection Officer
- how to contact the Information Commissioner's Office

may be found at https://www.cardiff.ac.uk/public-information/policies-and-procedures/data-protection
Printed copies of the above documentation can be provided upon request.

After data collection I will have your data in my possession until the end of the project which will be 8th October 2021.

After 8th October 2021, I will anonymise all the personal data I have collected from, or about, you in connection with this research project, with the exception of your consent form. Your consent form will be retained for the duration of the project and may be accessed by members of the research team and, where necessary, by members of the University's governance and audit teams or by regulatory authorities. Anonymised information will be kept for the duration of the project, estimated completion: 08.10.2021.

10. What happens to the data at the end of the research project?

After the research has been complete, the anonymised data may be shared within Cardiff University in the context of my dissertation being shared to aid future students form their own dissertations. I do not anticipate any future research using the data. The data will not be shared publicly with anyone outside of the University.

11. What will happen to the results of the research project?

This project is not likely to be formally published to any academic papers but once marks have been received around December time participants may request to obtain a copy of the project should they wish.

12. What if there is a problem?

If you wish to complain or have grounds for concerns about any aspect of the manner in which you have been approached or treated during the course of this research, please contact Katarzyna Stawarz, stawarzk@cardiff.ac.uk. If your complaint is not managed to your satisfaction, please contact Liam Turner, comsc-ethics@cardiff.ac.uk. If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, you may have grounds for legal action, but you may have to pay for it.

13. Who is organising and funding this research project?

The research is organised by myself, Amber Osborne under the guidance of my supervisor Dr Katarzyna Stawarz and co-supervisor Dr Argenis Ramirez Gomez from the School of Computer Science and Informatics at Cardiff University.

14. Who has reviewed this research project?

This research project has been reviewed and given a favourable opinion by the School of Computer Science and Informatics Research Ethics committee.

15. Further information and contact details

Should you have any questions relating to this research project, you may contact us during normal working hours:

Name: Amber Osborne
Email: OsborneAP@cardiff.ac.uk

Thank you for considering to take part in this research project. If you decide to participate, you will be given a copy of the Participant Information Sheet and a signed consent form to keep for your records.

Appendix B

Survey Participation Consent Form

Appendix 4

Participant ID no:
Do not include box for anonymised *samples*

CONSENT FORM

Title of research project: Conquering The Void: Transforming Data into Action

SREC reference and committee: School of Computer Science and Informatics Ethics Committee

Name of Chief/Principal Investigator: Amber Osborne

	Please initial box
I confirm that I have read the information sheet dated 01.07.21 version 2 for the above research project.	
I confirm that I have understood the information sheet dated 01.07.2021 version 2 for the above research project and that I have had the opportunity to ask questions and that these have been answered satisfactorily.	
I understand that my participation is voluntary, and I am free to withdraw at any time without giving a reason and without any adverse consequences (e.g. to medical care or legal rights, if relevant). I understand that if I withdraw, information about me that has already been obtained may be kept by Cardiff University.	
I consent to the processing of my personal information, thoughts and opinions on the application Strava, for the purposes explained to me. I understand that such information will be held in accordance with all applicable data protection legislation and in strict confidence, unless disclosure is required by law or professional obligation.	
I understand who will have access to personal information provided, how the data will be stored and what will happen to the data at the end of the research project.	
I understand that anonymised excerpts and/or verbatim quotes from my survey responses may be used as part of the research publication.	
I understand how the findings and results of the research project will be written up and published.	
I agree to take part in this research project.	

Name of participant (print) Date Signature

Appendix C

Interview Participation Information Sheet

PARTICIPANT INFORMATION SHEET

CONQUERING THE VOID: TRANSFORMING DATA INTO ACTION

You are being invited to take part in a research project. Before you decide whether or not to take part, it is important for you to understand why the research is being undertaken and what it will involve. Please take time to read the following information carefully and discuss it with others, if you wish.

Thank you for reading this.

1. What is the purpose of this research project?

I am undertaking this research as part of my master's degree as it forms my final project: the dissertation. The aims of this study are to narrow the gap in wearable technologies between collecting data on smart devices such as watches or bike computers and taking tangible steps in using that data to form better habits when it comes to your physical activity. Current research suggests that data is collected, uploaded, and then forgotten about. There is a wealth of insight you can gather from your data, and I would like to investigate how different types of data, and how it is presented to you, can impact your habit formation.

2. Why have I been invited to take part?

You have been invited because you are user of the application Strava, you partake in exercise and have done for around 1 year or more. You also have no academic education on the topic of data visualisation concepts or theories.

3. Do I have to take part?

No, your participation in this research project is entirely voluntary and it is up to you to decide whether or not to take part. If you decide to take part, we will discuss the research project with you and ask you to sign a consent form. If you decide not to take part, you do not have to explain your reasons and it will not affect your legal rights. Please note that if you are a student at Cardiff University you are welcome to participate but your involvement in this project will have no effect on your education or progression through your degree course.

You are free to withdraw your consent to participate in the research project at any time, without giving a reason, even after signing the consent form.

4. What will taking part involve?

Your involvement in the study will be allowing me to access all available Strava data on your account and then partaking in an interview answering questions based off how you interpret the data and what you think of it. The data collected will include: the type of activity (swim, cycle, run etc), time, duration, distance, elapsed time, country, start location (if available on your device). The location data will not be stored, I will be using it to cross reference your activity to weather data for my research. The weather data will be stored not your location data. The interview will be carried out once and it should take up to 45 minutes. You will be entirely anonymous as part of the interview; and the only identifiable information taken from you will

be your email address so we can arrange the interview. Your email address will not be used in the study and the data will be kept separate from that of the data collected during the interview and your Strava data. You will be asked to provide a key word, that should not identify you, that you can use to withdraw your data from the study at any point should you wish. I will record our voices as part of the interview process and store these recordings securely, this will be done so that I can refer back to our conversation at a later stage and accurately record your responses for the purpose of my research, I will not be taking video, it will be audio only.

5. Will I be paid for taking part?

No. You should understand that any data you give will be as a gift and you will not benefit financially in the future should this research project lead to the development of new methods of data presentation.

6. What are the possible benefits of taking part?

Users who take part will have access to predictions of their future sporting activities, created from their historic data. These predictions will be presented in a way that has been designed based off user feedback and visualisation principals. Those interested in their sports data and those curious about new epiphanies from their data may find the study beneficial to them personally.

7. What are the possible risks of taking part?

Participants should note that during interview I will ask questions about their current fitness habits and show them predictions for future habits. This will involve reflection on your life and sport during the interview. Therefore, there may be discomfort experienced for participants who may have experienced distress in the time-period of the data collected as you reflect on it. To mitigate this, I remind participants that should you experience discomfort you are welcome to leave the study at any point and do not have to give a reason for doing so.

8. Will my taking part in this research project be kept confidential?

All information collected from (or about) you during the research project will be kept confidential and any personal information you provide will be managed in accordance with data protection legislation. Please see 'What will happen to my Personal Data?' (below) for further information.

9. What will happen to my Personal Data?

The interview will not take any identifying information, when you state your answers, you will be anonymised and assigned a name that is not your own, "Athlete A", "Athlete B" etc. The only personal information that may be collected is your email address so I may contact you but this information will be stored separately and have no impact on the study.

Cardiff University is the Data Controller and is committed to respecting and protecting your personal data in accordance with your expectations and Data Protection legislation. Further information about Data Protection, including:

- your rights
- the legal basis under which Cardiff University processes your personal data for research
- Cardiff University's Data Protection Policy
- how to contact the Cardiff University Data Protection Officer

- how to contact the Information Commissioner's Office

may be found at https://www.cardiff.ac.uk/public-information/policies-and-procedures/data-protection
Printed copies of the above documentation can be provided upon request.

After data collection I will have your data in my possession until the end of the project which will be 8th October 2021.

After 8th October 2021, I will delete all the personal data I have collected from, or about, you in connection with this research project, with the exception of your consent form. Your consent form will be retained for the duration of the project and may be accessed by members of the research team and, where necessary, by members of the University's governance and audit teams or by regulatory authorities. Anonymised information will be kept for the duration of the project, estimated completion: 08.10.2021.

10. What happens to the data at the end of the research project?
After the research has been complete, the data visualisations created from your data may be shared within Cardiff University in the context of my dissertation being shared to aid future students form their own dissertations. I do not anticipate any future research using the data. The data will not be shared publicly with anyone outside of the University.

11. What will happen to the results of the research project?
This project is not likely to be formally published to any academic papers but once marks have been received around December time participants may request to obtain a copy of the project should they wish.

12. What if there is a problem?
If you wish to complain or have grounds for concerns about any aspect of the manner in which you have been approached or treated during the course of this research, please contact Katarzyna Stawarz, stawarzk@cardiff.ac.uk. If your complaint is not managed to your satisfaction, please contact Liam Turner, comsc-ethics@cardiff.ac.uk. If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, you may have grounds for legal action, but you may have to pay for it.

13. Who is organising and funding this research project?
The research is organised by myself, Amber Osborne, under the guidance of my supervisor Dr Katarzyna Stawarz and co-supervisor Dr Argenis Ramirez Gomez from the School of Computer Science and Informatics at Cardiff University.

14. Who has reviewed this research project?
This research project has been reviewed and given a favourable opinion by the School of Computer Science and Informatics Research Ethics committee.

15. Further information and contact details
Should you have any questions relating to this research project, you may contact us during normal working hours:

Name: Amber Osborne
Email: OsborneAP@cardiff.ac.uk

Thank you for considering to take part in this research project. If you decide to participate, you will be given a copy of the Participant Information Sheet and a signed consent form to keep for your records.

Appendix D

Interview Participation Consent Form

CONSENT FORM

Title of research project: Conquering The Void: Transforming Data into Action

SREC reference and committee: School of Computer Science and Informatics Ethics Committee

Name of Chief/Principal Investigator: Amber Osborne

	Please initial box
I confirm that I have read the information sheet dated 09/07/2021 version 3 for the above research project.	
I confirm that I have understood the information sheet dated 09/07/2021 version 3 for the above research project and that I have had the opportunity to ask questions and that these have been answered satisfactorily.	
I understand that my participation is voluntary, and I am free to withdraw at any time without giving a reason and without any adverse consequences (e.g. to medical care or legal rights, if relevant). I understand that if I withdraw, information about me that has already been obtained may be kept by Cardiff University.	
I consent to the processing of my personal information, my Strava exercise data from up to the last 365 days and responses during a reflective interview, for the purposes explained to me. I understand that such information will be held in accordance with all applicable data protection legislation and in strict confidence unless disclosure is required by law or professional obligation.	
I understand who will have access to personal information provided, how the data will be stored and what will happen to the data at the end of the research project.	
I consent to being audio recorded for the purposes of the research project and I understand how it will be used in the research.	
I understand that anonymised excerpts and/or verbatim quotes from my interview may be used as part of the research publication.	
I understand how the findings and results of the research project will be written up and published.	
I agree to take part in this research project.	

Name of participant (print)	Date	Signature
Amber Osborne	02.08.21	AOsborne
Name of person taking consent (print)	Date	Signature
Investigator		
Role of person taking consent (print)		

THANK YOU FOR PARTICIPATING IN OUR RESEARCH

YOU WILL BE GIVEN A COPY OF THIS CONSENT FORM TO KEEP