UNIVERSITY OF LINCOLN

Visualising Player Data for Video Game Designers

by

Tom Feltwell

A thesis submitted in fulfillment for the
degree of Master by Research in Computer Science

in the
School of Computer Science
College of Science

July 2016

Declaration of Authorship

I, Tom Feltwell, declare that this thesis titled, 'Visualising Player Data for Video Game Designers' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: *11^{th} July 2016*

"Character itself is forged on the anvil of adversity"

Paul Sellers

UNIVERSITY OF LINCOLN

Abstract

School of Computer Science
College of Science

Masters by Research

by Tom Feltwell

The collection and analysis of videogame players' actions in the game world, known as game telemetry, is a common technique for understanding the behaviour of players. This process, known as Game Analytics, often uses data visualisation to allow designers to manually analyse features of the data. Heatmap visualisation, a grid-based visualisation showing how often an event occurs across the game world (e.g. firing of weapons), is used widely in the games industry for visualising aggregated data, but has limitations when used to classify player behaviour at an individual or group level. Existing works using clustering to identify player behaviour yield results that must be interpreted by an expert, a problem acknowledged by existing research. Motivated by these limitations, this work presents the novel application of dendrogram visualisation as a means to interpret large datasets of heatmaps, through the use of hierarchical clustering, to aid designers in exploring and analysing player behaviour. This allows an intuitive and well-understood visualisation technique (heatmaps) to be used for cluster analysis, presenting intelligible results to a game designer, in a format they are familiar with.

To evaluate dendrograms as a design tool, a system was designed and implemented to visualise player data, using heatmaps, with hierarchical clustering being performed on these heatmaps, the results displayed as a dendrogram. A feasibility study was conducted with a set of game designers, to understand the opportunities and limitations of dendrograms as a game analytics tool. The results affirmed the utility of heatmaps for visualising aggregate data, but visual complexity increases in large quantities. Dendrograms were found to be initially difficult to read, but showed promise for analysing large sets of data and guiding the designer to interesting areas of the data, provided they could "drill down" into the base data (heatmaps). In light of these findings, a usability study was designed and conducted with a set of 40 game development students, where they were presented with realistic game design scenarios, and asked to find answers to analytics questions using heatmaps and dendrograms. The results showed that whilst dendrograms were initially difficult to understand, they were used to successfully

explore and understand cluster relationships, with participants providing the correct answers grounded in the data. Furthermore participants reiterated the need to explore the base data (heatmaps) to understand the cluster relationships of the dendrogram.

This work concludes that dendrograms represent a viable and useful tool for identifying interesting behaviour patterns within a heatmap dataset. Whilst some familiarity is required with the tool, it is possible to use dendrograms to explore behaviour clusters within a large dataset, and this work presents a solution to the limitations of analysing player behaviour through the use of heatmaps in large datasets. This work highlights a number of avenues for future work, such as deploying and studying dendrograms in a game production setting, or evaluating the dendrogram visualisation in different game genres.

Acknowledgements

I am deeply thankful to my supervisors Grzegorz Cielniak and Patrick Dickinson, whose advice and encouragement have helped this work come to fruition. I would also like to thank Ben Kirman, Conor Linehan and Shaun Lawson for their continued pub-based support. I am deeply grateful to my parents, John & Carol for everything they have done for me. Finally to my partner Nichola, who has helped, supported and encouraged me through the years.

Contents

List of Figures

List of Tables

Abbreviations & Acronyms

AI	**A**rtificial **I**ntellience
BLOB	**B**inary **L**arge **OB**ject
FPS	**F**irst **P**erson **S**hooter
GA	**G**ame **A**nalytics
GUR	**G**ames **U**ser **R**esearch
GIS	**G**eographic **I**nformation **S**ystem
HCI	**H**uman **C**omputer **I**nteraction
IDE	**I**ntegrated **D**evelopment **E**nvironment
RO	**R**ed **O**rchestra: Ostfront 41-45
RTS	**R**eal **T**ime **S**trategy
SDK	**S**oftware **D**evelopment **K**it
UCD	**U**ser **C**entred **D**esign

Chapter 1

Introduction

In the videogame industry, the recording and analysis of player behaviour, through the collection of game telemetry, is widespread, with more and more off the shelf solutions being developed to help designers and developers understand what their players are doing, which in turn helps inform decision-making in the design and development process. As videogames rise in popularity, the importance of understanding player behaviours is key to achieving and maintaining fun, challenging and rewarding game play experiences. This thesis centres on this process of game design, and the tools with which game designers understand their players. Game designers commonly use telemetry systems programmed within their games to collect data about players, often in the form of spatial data. This data can be visualised at an aggregated level using heatmap visualisation, but this visualisation approach becomes too visually complex when used to analyse multiple heatmaps at an individual level. Work has been done to mitigate this shortcoming by using clustering methods, but as yet none preserve the base spatial data in heatmap form. This work proposes the use of an existing visualisation technique, dendrograms, in order to display the clustering results of heatmaps. This work sets out to design, implement and evaluate the utility of dendrograms for game designers as a means of exploring the results of clustering. In the ensuing chapter, the context of the work within the field of Game Analytics is provided, the problem addressed by this work is outlined, followed by the research aim. Following this the objectives, and contributions of this work are discussed.

1.1 Game Analytics

The rise in popularity of video games since their initial beginnings in the 1950s, through to the multi-million pound industry they occupy today has lead to increasingly large teams of developers and designers required to create them. Complexity of these games has also increased significantly, not only in gameplay terms but also in technical and design terms. As sales and player numbers have increased, the task of understanding the behaviour and desires of players has become more difficult. Modern high-budget commercial games often have overall player numbers in the millions, with potential for hundreds of thousands of concurrent players across the globe [3]. In response to this growth in popularity and complexity, much work has been done in the games industry and academia to study and understand games, players and the development process, and this is represented through the field of **Game Analytics**.

The primary purpose of game analytics is to use data to support the decision making process during game development [17]. To do this, the field is focused towards the collection of data about games, also known as **game telemetry**, and the computation of this game telemetry into measurable **game metrics**. El-Nasr et al [17] outline the three core areas of game metrics:

- **User metrics** - Metrics related to users, from **a)** the customer angle, and **b)** the user angle. For example, **a)** will cover metrics such as daily expenditure on game content, whilst **b)** will cover kills per level and trajectory through the game levels.
- **Performance metrics** - Metrics related to the hardware and software processes that support the game. For example the graphics frame rate or the number of game crashes.
- **Process metrics** - Metrics related to the development process itself. For example the daily task completion rate of the programming team.

The study of *user metrics* is known as **Games User Research** (GUR), and this is the focus for the research presented in this thesis. GUR encompasses a number of methods and approaches used to understand the gameplay experience and the experience of those playing the game. GUR is split into two important areas, **user-testing** and **game telemetry analysis**.

User-testing is an established technique for testing products and software in the wider HCI field [1]. This approach is often used to attain detailed, qualitative feedback, as well as allowing testers to provide suggestions and ask questions of the designers. Within games, this typically involves bringing players into a laboratory-type arrangement, where they will play the game whilst being observed. Various methods can be used to collect the data from these players, including audio/visual recording of players' input, player experience questionnaires, interviews and biometric measures (heart-rate, galvanic skin response, and so forth). These collection methods produce a wide range of rich qualitative and quantitative data, some of which can be computed into game metrics. All of this data can then be analysed by game designers to gain insights into the player experience [35].

GUR also focuses on **game telemetry analysis**, which concerns the collection and analysis of data from remote players "in the wild", who are playing games that have already been released to the public. This collection of data is supported by the architecture of the Internet, and allows game developers to remotely collect data about their game and what users are doing within it in near real time. Subsequently, this data can be processed and used to understand what players are doing within the game world. With the rise of online games and Internet-connected PCs and games consoles it is easier for developers to collect telemetry data from tens of thousands of players concurrently, leading to large amounts of data that require specialised processing and analysis to make insightful for the game designers. This thesis focuses on this latter theme within GUR concerning remote collection of game telemetry, and its analysis by game designers in order to develop, change and maintain a fun and enjoyable play experience.

1.1.1 Visualisation & Analysis

Some user metrics can be processed and used to answer business intelligence questions at a broad scale. For example, a developer is able to ask the question "How many players are using this feature every day?" and is able to compute a numerical answer from the game telemetry data. The game metrics provided can be used to inform a decision, for example about whether to keep the feature. In this way, using game analytics allows developers to make more informed decisions regarding financial, strategic and gameplay matters, as opposed to using design intuition or experience alone.

By design, videogames feature a virtual environment which the players move through and interact with. This produces spatial information over the time players are moving and interacting (spatio-temporal data), that can be collected via game telemetry, then processed and visualised. For instance, the first-person shooter (FPS) game *Half-Life 2: Episode 2* sees players navigate through a series of levels within the virtual game world. Spatial data can be collected and processed for multiple players, which can subsequently be visualised in aggregate, showing the patterns of behaviour for multiple players. For example, Figure 1.1 shows a heatmap visualisation of player deaths in the game level "Outlands" of *Half-Life 2: Episode 2*. Analysis of spatial information becomes highly complex when used in large quantities, and by presenting spatial information in more intelligible forms, such as visualisations, designers are able to understand this complex information, and thus the behaviour of players within the virtual environment [11].

The analysis and visualisation of spatial data is highly-scalable, and as game environments become more complex and the number of players increase, game telemetry recording and analysis has become widely adopted by large game developers such as EA [30], Microsoft Game Studio [27] and Bioware [49].

1.2 Problem Statement & Motivation

Heatmaps are an established visualisation technique for spatial game data, with many practical examples of their application within the game development process (see [39], [2], [11]). Heatmaps are 2D grid-based representations of the frequency of a data type over time, overlaid onto a meaningful spatial image/map. In the games context this is usually the topography of a game level. They allow the quick identification of the distribution of data observations over a spatial area. Heatmaps derive their name from their use of a colour gradient to represent the frequency of the events, most often red/white "hot" colours for high frequency, and blue/black "cold" colours for low frequency. For example, the games developer *Valve* used heatmaps to visualise the locations of player deaths within a game level, where "hot" areas of the heatmap denote a high number of deaths, and "cold" areas denote a low number of deaths [2]. An example of this can be seen in Figure 1.1.

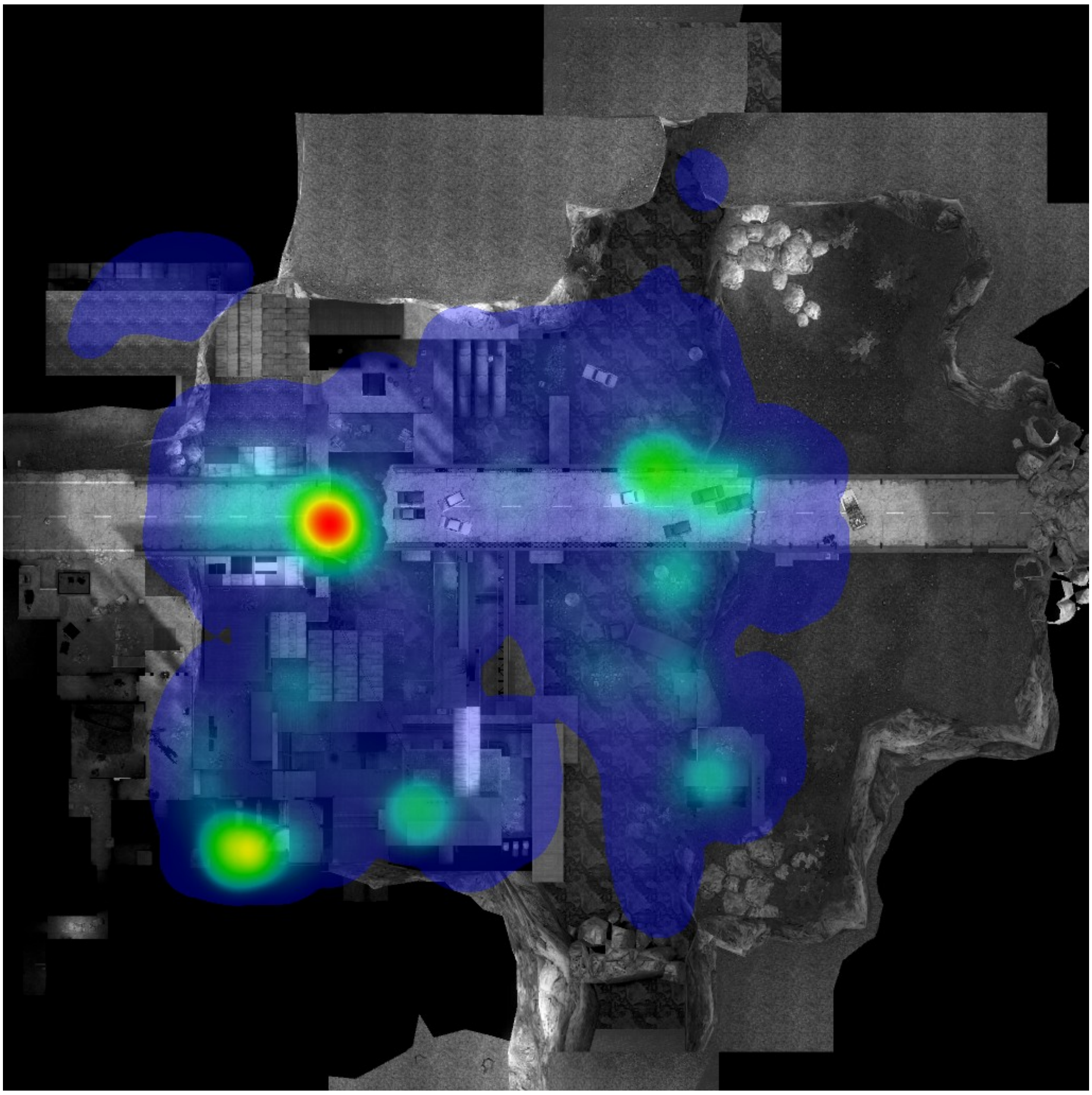

FIGURE 1.1: Heatmap of player deaths from *Half-Life 2: Episode 2*, on the map *Outlands*, courtesy of Valve Corporation. [26]

Although heatmaps are used often in the games industry for providing an overview of player behaviour using aggregated data, they have a number of shortcomings in application. Due to the reliance of the visualisation technique on colour gradients, it becomes difficult to visualise more than one data type, as this requires one colour gradient range per data type. The overlap between the colour gradient ranges becomes difficult to understand, and as such makes the heatmaps unintelligible. Solutions have been proposed to this shortcoming, notably the use of Geographic Information Systems (GIS) to display multiple data types of spatial game data as a series of layers which can be switched on and off at will [11]. GIS are used often for geographical applications (mapping, public health, crime mapping, and so forth), and one criticism of using GIS for game data visualisation is the prohibitive cost for small and medium sized game developers, as well as a lack of flexibility of commercial tools for games applications [21] [34].

In addition, with large quantities of telemetry data being generated by large numbers of players, it is often desirable to inspect individual behaviours in order to identify patterns, or to identify groups of behaviours within the overall player population. It is impractical for designers to compare multiple heatmaps of individual players, as visual complexity increases quickly based on the number of players being analysed. Game designers require game analytics tools to provide them with functionality to perform “first pass” filtration on large sets of telemetry data, allowing them to see the pertinent features within a data set [32]. Indeed, work has been done in this area to turn high-dimensionality “raw” game telemetry data into actionable and understandable insights for game developers, and this has been achieved through the use of classification of player metrics [45] and the clustering of player data [12] [13], though these latter approaches often require statistical expertise to process the final results into information that can inform game designers, such as behavioural profiles of the different players.

As a result, it is clear that there is a requirement for data summarisation techniques that present results that are understandable by a game designer, as opposed to a data visualisation expert. Furthermore, whilst heatmaps are ubiquitous for visualising aggregated data, they are seldom used for analysis of individual player data at anything above small scale, due to the high visual complexity of comparing large numbers of players.

1.3 Aim and Scope

Motivated by this problem space, the aim of this research is to design and implement a clustering visualisation technique, dendrograms, and evaluate the utility of them, to game designers when used to analyse large quantities of individual spatial player data.

It should be noted that this research is limited to spatial game telemetry related to player behaviour. A number of solutions have been proposed for clustering of non-spatial player telemetry features (see Chapter 2), and this limitation focuses the work towards the little researched area of spatial telemetry clustering. Additionally, the work concerns the visualisation of spatial data and does not consider the further analysis of clustered spatial player telemetry, concentrating the work on initial visualisation of hierarchical relationships between individual player heatmaps in large sets of spatial data.

1.4 Research Question

Based on an exploration of existing literature concerning user-centred design for game design tools, and game telemetry visualisation techniques, a full review of which can be found in Chapter 2, and building on the problem outlined above, it is clear that large amounts of player data are available through the use of game telemetry collection. The games industry are commonly using heatmaps to visualise this data, due to the ease of interpretation of heatmaps. Much work has been done presenting clustering of player data and/or classification of behaviour, yet little of this work relies on the commonly used heatmap visualisation. Motivated by the growing trend of game analytics research in a user-centred fashion, the research question is described as follows:

- Is an existing hierarchical clustering visualisation technique, the dendrogram, useful and valuable within the game design process to help summarise sets of individual heatmap visualisations in order to highlight patterns and pertinent features within the data?

1.5 Objectives

In order to explore this research question and address the aim of the project, a clear set of objectives was required. Each of these is detailed below:

- *Explore the usage of spatio-temporal player telemetry, and it's application within the game industry and academia.*

  Survey and review existing methods of analysing spatio-temporal player data, in both academia and in the games industry. This exploration will focus on visualisation approaches predominantly, such as heatmaps.

- *Understand the existing works and methodologies for designing game design tools.*

A thorough literature review of research in this area, as well as existing design approaches and practices of game designers. Taking into consideration data analysis methods in Game Analytics, as well as existing works within Games User Research that focus on how designers use and interpret gathered data, and tools and approaches developed for or in conjunction with practitioners themselves.

- *Implement telemetry collection framework and collection of preparatory dataset*

  Design and implement a telemetry collection framework within a game, and conduct data collection sessions with participants in order to collect a preparatory dataset for use in data visualisation.

- *Implement hierarchical clustering using dendrogram visualisation for use as a high level player telemetry summarisation tool.*

  To address shortcomings in the heatmap visualisation and existing behaviour classification and clustering approaches, a data summarisation tool will be created using hierarchical clustering of large sets of individual player heatmaps, generated from player telemetry. The results of this hierarchical clustering are then visualised through a dendrogram, a type of tree-like navigable graph, which presents the relationships between individual player heatmaps, allowing designers to identify interesting patterns and features within the data set.

- *Analyse the feasibility and usability of dendrograms as a tool for use by game designers in the exploration of spatial data sets.*

  Evaluating the dendrogram approach through a feasibility study (Chapter 5) with professional game designers, the potential limitations and opportunities posed by dendrograms as design tool can be understood. Informed by this, a usability study (Chapter 6) was conducted the evaluate their utility for actual game analytics tasks.

1.5.1 System Overview

To summarise the visualisation approach presented in the objectives, Figure 1.2 situates the proposed dendrogram visualisation technique within the game design process. In line with existing games industry practice, user testing is conducted which yields a corpus of spatial telemetry data. This corpus is then processed into heatmaps of individual player actions. Hierarchical clustering is then performed on this set of heatmaps, producing a visualisation in dendrogram form representing the relationships between each heatmap in the data set. This dendrogram can then be analysed by a game designer, who uses the insights gained from the dendrogram to implement changes into the design process.

1.6 Contributions

This thesis presents contributions to knowledge in four areas.

- At a practical level, the unsuitability of heatmap visualisation as a technique to identify and compare individual spatial behaviours is an understood problem of visual complexity. This work proposes a solution to this problem, by allowing the summarisation of individual spatial visualisations (heatmaps) through hierarchical clustering. These are displayed in an established visualisation form (dendrograms), which allow the relationships between individual spatial behaviours to be identified, whilst preserving the underlying base observations (individual heatmaps).

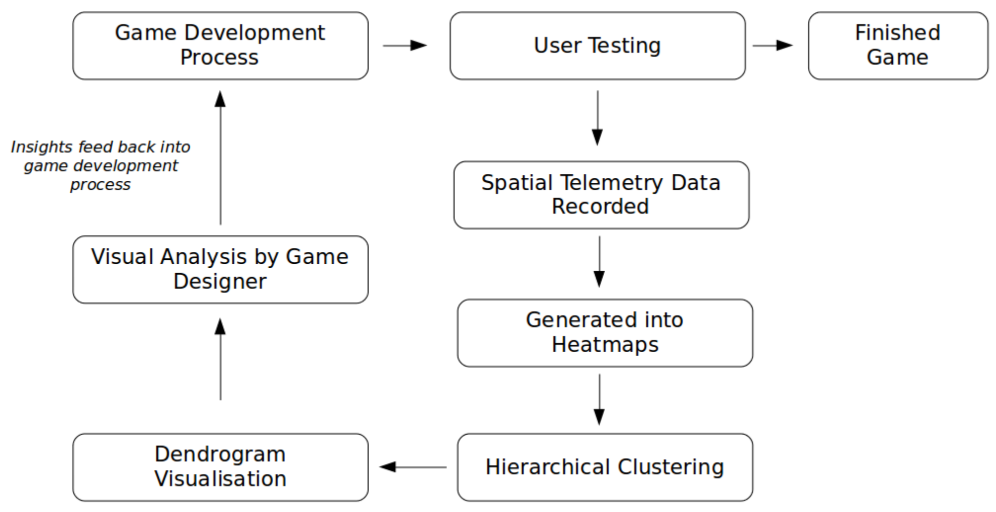

FIGURE 1.2: Overview of dendrogram visualisation within overall game-design process.

- This work proposes the novel application of dendrogram visualisation to the domain of game analytics, specifically in the visualisation and summarisation of relationships between spatial data observations, manifested in large sets of individual heatmaps.
- This work contributes to the growing body of work using user-centred design methodology to design and evaluate game design tools for and with their intended users, game designers. In Study 1 an understanding of designers' existing use of heatmaps and game analytics in the design process is contributed, as well as their considerations of data overload and a specific problem space for data summarisation in game analytics. Study 2 contributes an understanding of dendrogram visualisation in realistic game design scenarios, demonstrating the value of dendrograms as a means for isolating outliers, and analysing clusters of player behaviour.
- This work provides details of practical design and implementation of telemetry framework into a game engine, which can help inform future research into telemetry implementation within FPS and analogous games.

1.6.1 Publications

Details of the telemetry framework implemented into the FPS *Red Orchestra*, presented in this thesis, and an exploration of the preparatory dataset was published at the *GAMEON 2012* conference as a full main track paper. This paper presented the player telemetry framework, as detailed in Chapter 3. Full citation as follows:

> Tom Feltwell, Patrick Dickinson, Grzegorz Cielniak. A framework for quantitative analysis of user-generated spatial data, *GAMEON 2012 Conference*, 2012. [18].

The feasibility of dendrogram visualisation as a game design tool was published at the *CHI Play 2015* conference, as a work-in-progress with poster presentation, which can be seen in Appendix A. This paper explored Study 1 of this thesis (Chapter 5), where interviews were conducted with game designers in order to understand the feasibility and applicability of dendrograms and heatmaps within the design process. Full citation as follows:

Tom Feltwell, Grzegorz Cielniak, Patrick Dickinson, Ben J. Kirman and Shaun Lawson. Dendrogram Visualization as a Game Design Tool. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI Play 2015).* Pages 505-510, ACM, 2015. [19].

1.7 Overview of the Thesis

This thesis consists of six further chapters, which are broadly separated into three main parts. In PART I (Chapters 2, 3 and 4) a review of existing literature around game analytics, spatial clustering & visualisation and user-centred game design tools is detailed. This is followed by preparatory work implementing and collecting a corpus of game telemetry data, with the following chapter detailing the technical design and implementation of the proposed visualisation techniques. PART II (Chapters 5 and 6) comprise two studies: the first a feasibility study of dendrogram visualisation with game designers, the second a usability study of dendrogram visualisation in game analytics tasks. PART III (Chapter 7) explores the findings of these two studies, and presents the conclusions of this work.

Chapter 2

Literature Review

The context of this research is within the field of Game Analytics, the study of games, their players and the game development process. Specifically this work is focused towards the field of Games User Research (GUR), the study of a game's players, their behaviour and their playing experience. Within this literature there is work aimed towards the presentation of novel methods for collecting, analysing and visualising game telemetry data, using learning algorithms, clustering and other techniques. There is also a growing body of literature directed towards the game design process itself, and how to design tools in a user-centred way, where designers at put at the centre of the design process. As such, this work is grounded in two key areas of research: the technical processing and representation of game telemetry data, and the user-centred design of tools for game designers. The following literature review is divided accordingly into these two areas.

2.1 Game Telemetry Processing and Representation

The field of **Game Analytics**, a specialised branch of business intelligence, is concerned with understanding the way customers and users interact with games, in order to inform decisions about game development, maintenance and improvement. An extensive overview of the game analytics field is provided by El-Nasr et al [17], where they summarise the focus of game analytics:

> "The goal of game analytics is to support decision making at operational, tactical and strategic levels within all levels of an organisation - design, art, programming, marketing, user research, etc. Game analytics forms a key source of business intelligence in game development, and considers both games as products, and the business of developing and maintaining these products." [17] p.5.

A sub-field of game analytics, **GUR** studies player experience. This encompasses the behaviour of players within the virtual environments, as well as their experience and behaviour at the user level. Nacke et al. [35] describe the four key areas of research within the emerging GUR field in 2009: *Heuristics & expert review, biometric feedback, Game metrics* and *self-reported and observed behaviours.* These remain the four key strands of GUR, and a comprehensive overview of challenges in the GUR field can be seen in [14]. Drachen et al [14] call for more work to be done towards "low-cost testing" which does not require expert training, or laboratory conditions. Furthermore they call for this work to be focused towards small and medium sized game developers, who do not have large reserves of capital or work-hours with which to implement complex and time consuming game analytics approaches. The authors also point out the difficulty, for all developers, in cross-correlating and combining data of different types and from different sources, especially when working with large quantities of data.

GUR covers a wide variety of methods for extracting information about player behaviour, such as biometric data recording in a laboratory [33], the use of subjective user-surveys [5], as well as interviews and analysing video of players' gameplay experience [45]. The recording of player behaviour within the virtual environment, game telemetry or instrumentation data, can range from the button presses of each player to their spatial movements within the game world, and their interaction with objects and characters in the game world. As such, this raw data can be processed into game metrics, statistical summaries and representations of behaviours, which can then be used by game designers, in combination with existing user testing methods, in order to inform and understand game design decisions [45]. The importance of statistical analysis of user generated data has been recently noted by the game industry and resulted in a number of existing systems for data collection and analysis [27],[46].

Often telemetry processing and data analysis are developed into frameworks to be used by a large game developer. Zoeller [49], a developer at Bioware, presented a detailed overview of an in-house game telemetry analysis and exploration tool, used in the development of large AAA game titles, such as *Mass Effect, Knights of the Old Republic* and *Dragon Age: Origins.* The system is predominantly intended for bug-tracking and fixing, and in order to do this more effectively the tool uses an in-game telemetry system to combine a user-input bug report with the telemetry from their game play session, such that bug reports are contextualised and provide explorable data for the game developer. Additionally, they also provide a set of recommendations to other game developers about how to design an effective telemetry system that is usable by designers. This is an important insight to industry practice, as the confidentiality of industry data is often cited as a barrier for knowledge sharing between the games industry and academia [17].

2.1.1 Spatial Data Visualisation

Narrowing the analysis of game telemetry down to spatial data, there is much research into different methods to visualise spatial game telemetry data in forms that are comprehensible and insightful for the game developer. A commonly used visualisation in the games industry are **heatmaps**, a form of spatial histogram, and heatmaps have been used to visualise a broad spectrum of spatial player data, such as location of player deaths, or movement of players within a game level [10].

Wallner et al [47] present a comprehensive literature review of game data analytics and visualisation, principally focused on spatial data. They note that the research into visualisations of game telemetry data spans both game developers and game players alike, with the latter being a recent emerging trend. They outline five types of common data visualisations/representations in industry and academia: charts & diagrams, heatmaps, movement visualisation, self-organising maps and node-link representation. Notably, they discuss how charts and diagrams are commonly used for answering specific questions about gameplay, but are less well suited to exploratory analysis, such as pattern detection. They also outline the strong industry usage of heatmap representation for explorative analysis of aggregated player data, such as movement of players over time. This comprehensive literature review is expanded upon by Drachen et al [11] who describe the academic and industry methods for analysis of player behaviour, i.e. the examination of player actions patterns over time. Furthermore, they detail a set of future directions for game analytics research, as well as a set of open problems in the field. Two of the presented flagship areas are of note, *Spatial Outlier Detection* and *Spatial Clustering*: The authors highlight the possibility to compare spatial observations between one another and using this to detect outliers, with the potential benefits of isolating exploitable problems within level design or noticing suspicious behaviour. The ability to cluster spatial observations into groups would allow for the identification of player behaviour patterns, and the authors call for more work in the area of spatial clustering and outlier detection.

Heatmap visualisation is commonly used for telemetry analysis in order to visualise the spatial distribution of data, and a full technical description of the visualisation technique can be found in Chapter 4. Heatmaps have a wider variety of applications, and are often used when spatial information needs to be visualised over an existing map, which gives the spatial information context. For example, Gourinchas et al [23] uses heatmaps to visualise the transfer of wealth around the globe after the 2008 global finanacial crisis, and visualising the wealth changes of nations around the world, is able to compare and contrast neighbouring countries, as well as quickly understand the terrain of wealth around the world after 2008 (i.e. hotspots of wealth reduce across the western world). Heatmaps have been used within HCI, for example Nielsen et al [36] use heatmaps to display the eye movements of users when viewing different website designs. The spatial information of each users eye movements were projected over the website itself, allowing the identification of hotspots where users would commonly look, and cold spots where users would not look often. This information can then be interpreted by the website's designer in order to understand whether the right areas of their website are attracting attention.

Heatmaps lend themselves well to viewing aggregated data, in order to identify hotspots and cold spots, areas of high and low activity respectively. Romero [39] describes the use of telemetry and subsequent visualisation as part of the game development process, outlining the utility of heatmaps when developing *Halo 3*, in order to identify where players were dying in their levels, and whether that was as expected by the game designers. The use of heatmaps by game developer Valve to explore the player behaviour in games such as *Team Fortress 2* were presented by one of their designers, Mike Ambinder, who details how they integrate heatmaps into the user-testing process, allowing

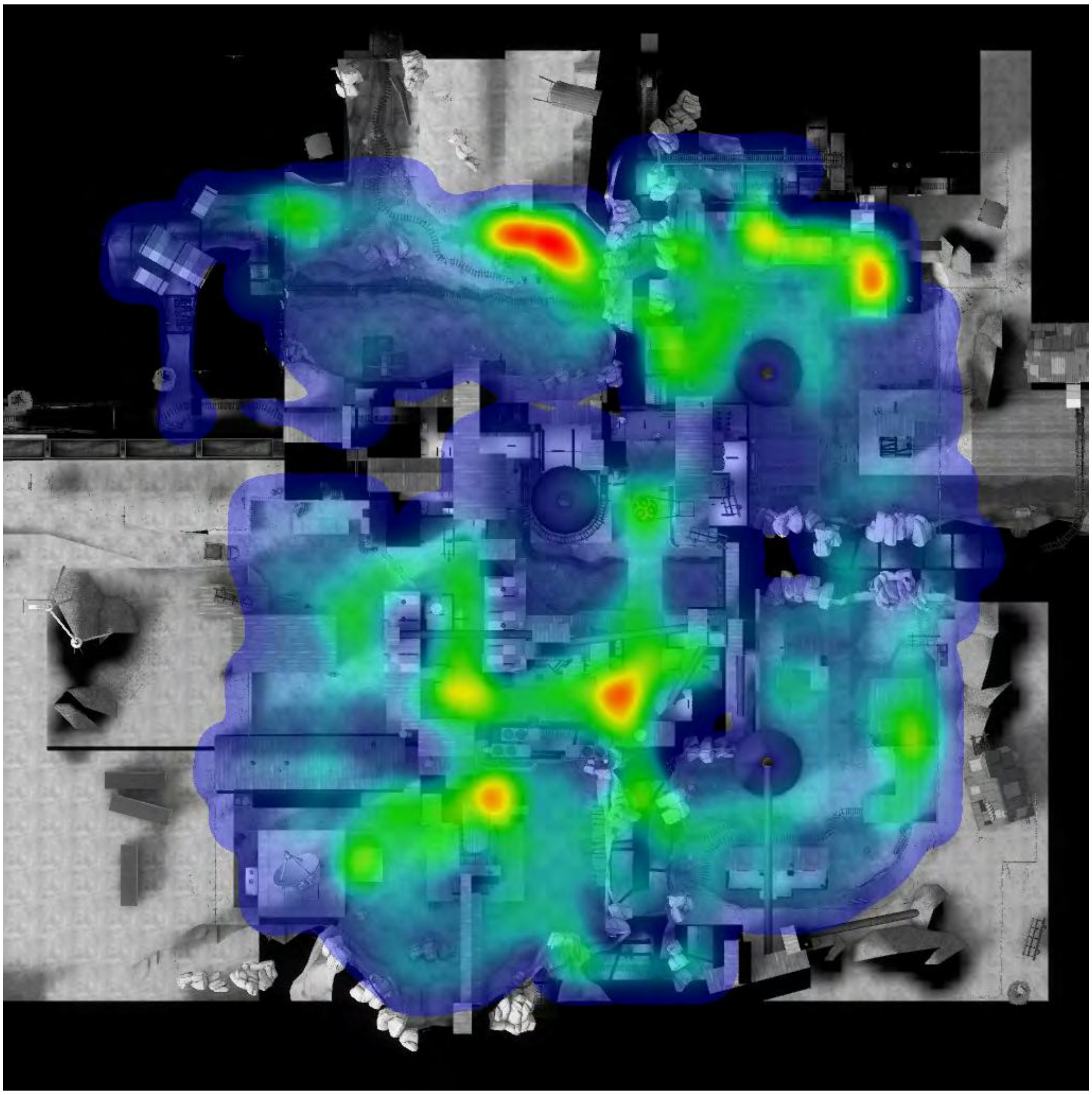

FIGURE 2.1: Heatmap of player deaths in the game *Team Fortress 2*, as presented by Ambinder [2]

them to formulate hypotheses about player behaviour, implement changes and measure the results using game telemetry data, including heatmaps [2] (see Figure 2.1). In their literature review of current spatial game data analytics, Drachen et al [11] provide a range examples of industrial use of heatmaps, in games such as *EVE Online, Just Cause 2* and *Transformers: War of Cybertron.*

The widespread adoption of heatmaps as a game analytics tool is evident by the number of tools, frameworks and research that utilise them as a fundamental visualisation technique: Kim et al [27] use heatmaps as part of an overall game telemetry analytics platform and present a case study of its use in the game development process, and

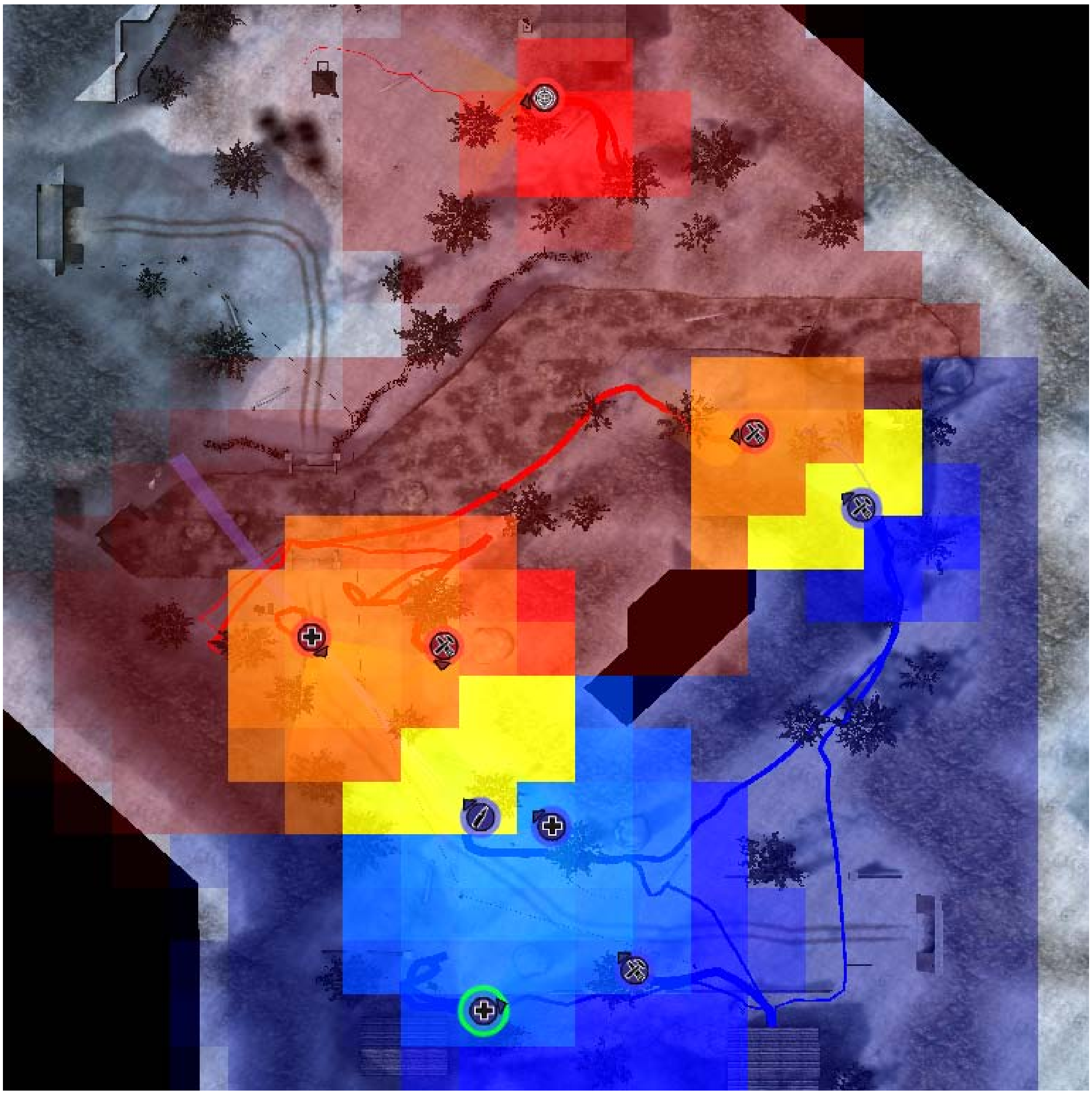

FIGURE 2.2: Real-time heatmap of team occupation in the game *Return to Castle Wolfenstein: Enemy Territory* presented by Hoobler et al. [25]

Hoobler et al [25] use heatmap-type visualisations to display live gameplay information to spectators in an online multiplayer FPS game, an example of which can be seen in Figure 2.2 Each of these examples are fully discussed in section 2.2 due to their focus on user-centred design. Drachen et al [10] describe a case study working with a corpus of data gathered from the action-adventure game *Tomb Raider: Underworld*, and how they used heatmaps to understand the locations where players were most likely to activate a "help-on-demand" mechanic within the game.

However there are limitations with heatmap visualisation, such as their use for visualising large data sets, and when visualising multiple variables. Using heatmaps to visualise aggregated data, such as the spatial distribution of player deaths for 300 players, is useful in identifying areas of game level that require attention, or locations of particular behaviour, they are not well suited to visualising large quantities for individual-level player data. One criticism of heatmaps used to represent aggregated data, is their ability to smooth over and hide individual behaviours. As such Nielsen et al [36] suggest the analysis of individual heatmaps allows a better understanding of individual user behaviour. In the games context, motivated by this desire to identify individual patterns in player data, using heatmaps to visualise the movement of each player in a 30 player data set, would create 30 individual heatmaps, and comparing each together would create n^2 comparisons. This creates a large amount of manual analysis work for a designer, and whilst manageable at small scale, this approach does not scale well to the large data sets available through game telemetry. This situation, often referred to as *data overload* [25], [27], reduces the efficacy of heatmaps as a data analysis method. Indeed, the creation of large volumes of data and requirement for automated support to understand the data is a well understood issue within HCI, especially in relation to usability testing, by example Hilbert et al present an extensive survey of different data analysis and summarisation approaches for processing the hugely voluminous telemetry output created through operating system interface usability testing [24].

It is also difficult to visualise multiple variables per heatmap, due to the use of colour gradients to visualise frequency. A separate colour gradient would be required for each variable being displayed, and this would quickly become confusing due to overlapping colour gradients, and whilst not impossible to create, the overlap between colour gradients would be confusing and reduce the heatmap's ready interpretability. A number of alternative methods and tools are presented to address this issue, in particular GIS tools proposed by Drachen et al [11].

For example, Drachen *et al* [9] present a showcase of approaches and methodologies, in the form of two case studies from commercial game development where various types of visualisation are used as ways to answer analytics question. The first case study covers the analysis of player behaviour in the game *Fragile Alliance*, specifically where in the game level players chose to betray their own team using an in-built *traitor* mechanic. The second case study covers the analysis of players' pathing through a level, to look for those who deviate from the "perfect" path intended by the game designers. In both of these cases, the spatial points of player data are plotted as layers in GIS software, with each layer displaying similarly to a heatmap. This approach allows the game level to be projected underneath the data point visualisation, allowing a developer to see the distribution and frequency of a particular game metric in the game level. Although the paper uses only experimental data collected from play testing within the game development company, it is an early piece of work extolling the virtues of game analytics, specifically with spatial data, as part of the game design and development process. In addition to this, Drachen et al [8] detail a case study of game analytics research with EIDOS Interactive, where they used a GIS tool to analyse telemetry data collected from the game *Tomb Raider: Underworld*, once it had been released. They propose that using GIS layers of different variables, such as a layer of spatial information for each cause of death, designers were able to switch the layers on and off independently, as well as manipulate opacity functions to view multiple layers simultaneously. The authors present GIS as a viable alternative to static single-variate heatmaps, allowing designers the ability to visualise multiple variables simultaneously. Whilst the benefits of GIS tools for multi-variate visualisation are clear, criticism has been levelled at commercial tools such as GIS, due to the prohibitively high cost (est. $10,000 per license) for small and medium sized game development companies, and thus create a barrier to adoption by those parts of the games industry [21].

Indeed, the unsuitability of commercial data analytics and mining tools for the specific application of game analytics is attributed as a driver for new and more specialised forms of game data visualisations research, noted by Moura et al [34], who details a node-based visualisation of spatial data, and Gagné et al [21] who develops an analytics approach to real-time strategy games.

2.1.2 Behaviour analysis & clustering

In order to further the analysis of spatial data, research has been done into the automated analysis of spatial data and the subsequent visualisation of these results. As outlined in the previous section, heatmap visualisation is useful for analysing behaviour in an aggregated manner, but becomes difficult when looking for individual behaviours. To address this, clustering and behaviour analysis research has been conducted with the aim of automating part of the analysis process of individual behaviour data, allowing designers to focus a set of results which classify or typify the behaviours present within a dataset. Often these methods rely on forms of clustering in order to group player data, in behavioural terms, into meaningful behaviour profiles and/or types. Furthermore these methods can be used to automatically detect and predict patterns and outliers within a data set, drawing attention to those features of the data.

Much clustering and behaviour analysis work has focused predominantly on non-spatial data, in order to describe and cluster player behaviour in a broad way. Tychsen et al [45] present a framework for classifying players into "personas" using game telemetry data. This framework outlines a number of characteristics, such as "Personality" and "Physical behaviour" with which to classify players, and provides a case study of *Hitman: Blood Money* which describes how these characteristics are mapped onto game telemetry data recorded from the game. The game telemetry used includes spatial information such as movement, but also considers a range of other data points, such as players' interaction with the story and interaction with the game world. The resultant classifications provide a multi-faceted persona for each character, for example a player having a stealthy navigation characteristic, with a silent assassin style. This framework provides a rich summary of players, but it requires a degree of work for each game to identify which game telemetry data item maps to each persona characteristic, as well as manual interpretation of these personas by a designer.

Work has been done by Drachen et al [12] to produce a clustering algorithm which can automatically provide succinct clustering of player behaviour via telemetry data. A study was conducted with telemetry data collected from 1,300 players of *Tomb Raider: Underworld*, and non-spatial data that described their overall player experience was used, e.g. total time to complete the game and number of deaths per level. Using k-means clustering and hierarchical clustering to pre-process the data, this clustered data was then passed through an unsupervised learning algorithm (a Self-Organizing Map), which produced four explicit clusters. The authors then outline that these clusters can be described, through a process of manual analysis of each cluster, into easy to understand player profiles. This work provides an important step towards scalably clustering and classifying players into a small number of broads groups. However, interpretation of the resultant learning algorithm data is complex and requires an expert, an example output can be seen in Figure 2.3 therefore limiting the potential of the method.

To address the problem of high-dimensionality game telemetry data of large scale,

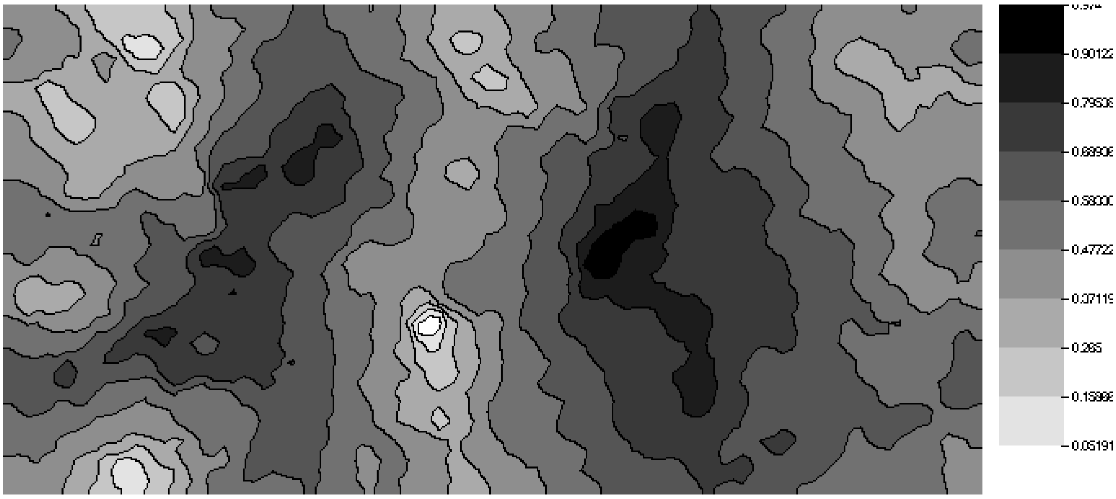

FIGURE 2.3: Output from Self-Organizing Map clustering on *Tomb Raider: Underworld* dataset, using causes of death by falling data. As presented by Drachen et al [12]

Drachen et al [13] illustrate the applicability of two clustering techniques in reducing dimensionality down to player behaviour classifications framed in language accessible to game designers. Using two case studies, they apply k-means clustering and Simple Volume Maximization clustering to datasets gathered from the online role-playing game (RPG) *Tera* and online multiplayer FPS *Battlefield 2: Bad Company*, comprising 25,000 and 10,000 player character each, respectively. Again, as in the work of [12], the results of each clustering algorithm are analysed by an expert, who then describes each cluster in language used by game designers, for example with cluster descriptions such as “Elite”, “Stragglers” and “Average Joes”. Of note, Simple Volume Maximization is described as providing an excellent identification of extreme, i.e. outlier, behaviours within the data sets analysed, and the authors note it’s potential utility for identifying problematic behaviour such as cheating and exploiting game mechanics. A degree of manual analysis is still required by an expert before a designer is able to action the insights, however the reduction in dimensionality in the large data sets presented is significant.

Focusing on clustering of spatial data, an extension to spatial heatmap visualisation is presented by Wallner et al [46], who include cluster analysis of spatial data within the heatmap form. This works by using geometric shapes superimposed over the game level, with the shapes, known as enclosure, representing the different clusters computed. In effect this allows a designer to see which clusters of behaviour (i.e. death by falling, death by NPC, and so forth) occur in particular areas of the game level. This approach is innovative and allows the visualisation of cluster information in "traditional" heatmap format, it removes the colour gradient feature of heatmaps and increases the amount of data being represented. The authors highlight the need to use their interactive tool in order to understand the clusters and their distribution on the heatmap.

One notable feature of the behaviour and clustering analysis work presented in this section is the lack of spatial data for clustering and behaviour analysis. Furthermore, whilst work is beginning to address this shortcoming (i.e. Wallner et al [46]), the original spatial data visualisations are not preserved. In the following section hierarchical clustering, dendrogram visualisation and their ability to preserve the original data observations is discussed.

2.1.3 Hierarchical Clustering & Dendrogram Visualisation

Clustering algorithms can be divided into hierarchical and partitional [43]. Hierarchical clustering methods iteratively build a hierarchy between each observation, whilst partitional clustering methods identify all clusters simultaneously, without any hierarchy within the data. Many examples of hierarchical and partitional clustering exist, and Steinbach et al [*ibid*] note that *single linkage* and *complete linkage* are the most common hierarchical methods used, with *k-means* being the most commonly used partitional method. K-means is fast, and produces distinct clusters within a dataset, but has limitations, most notably the need to define the number of clusters before cluster analysis is performed, and a lack of repeatability.

Hierarchical clustering does not require the number of clusters to be known, however because the method produces a hierarchy among the data, the clusters can be viewed at different levels of granularity, yielding different numbers of clusters. Hierarchical clustering is also dependent on the use of a similarity matrix, and as such is only suitable for data observations that can be compared for similarity in some way. In the games context, hierarchical clustering can be used to cluster any observations that can be compared, for example Drachen et al [12] performed hierarchical clustering on a similarity matrix of six non-spatial gameplay features of player gameplay. These features, such as cause of death, game completion time and the number of deaths, were compared across 1,365 players and represented as a similarity matrix. Hierarchical clustering was then performed on this data, producing a dendrogram showing the clustering of players (see Figure 2.4). The results of this hierarchical clustering were then used as base data for a more advanced learning algorithm (self-organizing models) to create player classifications.

Dendrograms are widely used as a data visualisation tool for representing hierarchical clustering, and therefore are found in most data-driven fields. Wilkinson and Friendly [48] present a history of the "cluster heat map", a visualisation that displays the cluster structure of a data matrix as a dendrogram, rendering the data matrix in the heatmap colour scheme. An example of this can be seen in Figure 2.5, which shows the hierarchical clustering of social statistics (e.g. life expectancy, literacy) across a range of countries. The "cluster heat map" preserves the base data observations (the data matrix of social statistics comparisons across countries) and augments it, displaying the cluster relationships between the data in the form of a dendrogram.

A further example of dendrograms for exploring hierarchical clustering can be found in the field of computational biology, dendrograms are used to display the results of hierarchical clustering of genomes [16], which visualize similarities between different genes and allow visual analysis of the relationships between different genes. In this context onus is placed on the researcher/user to visually inspect and analyse the relationships presented by the hierarchical clustering. By presenting the base data observations (the distance matrix/comparison between observations), an analyst is able to understand the relationships represented in the hierarchical clustering.

As Drachen et al [12] perform, but do not capitalise upon, the comparison of gameplay features, spatial or non-spatial, can be easily represented through hierarchical clustering, and therefore be visualised using a dendrogram or "cluster heat map". Adopting the dendrogram as a visualisation tool, rather than a pre-processing step for further computation, allows, as in other data analysis fields, the exploration and analysis of the cluster relationships in a set of data. The proposed use of dendrograms to visualise sets of spatial data (heatmaps) presents a union of existing cluster analysis practices in analogous fields outside games, and the desire for research and methods in the game analytics context that present clustering results that are actionable and interpretable by a game designer.

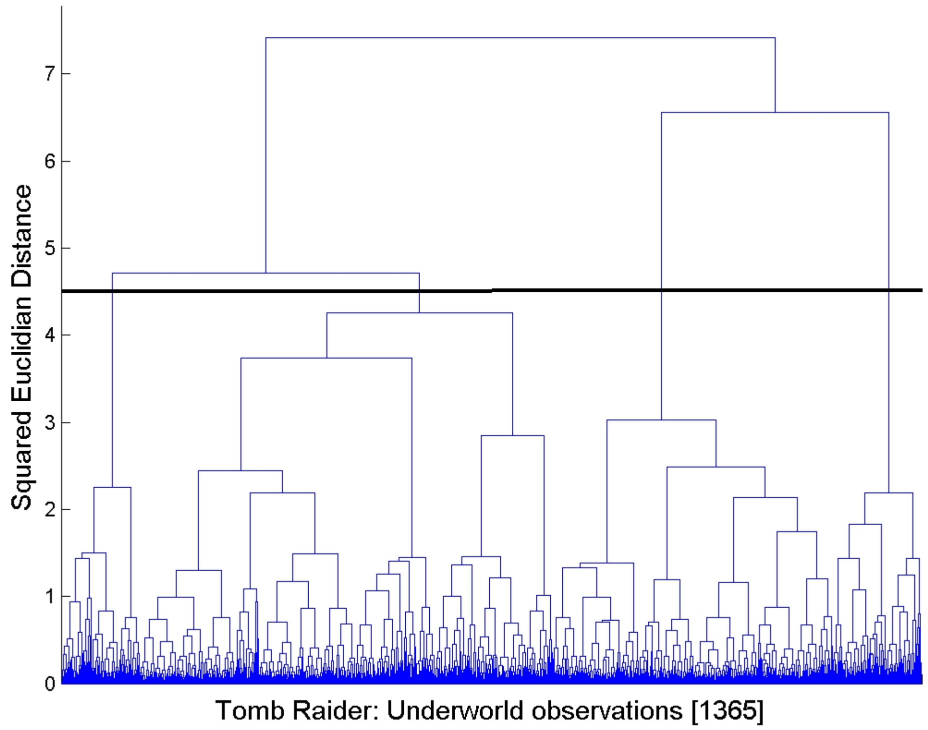

Figure 2.4: Dendrogram displaying results of hierarchical clustering performed on 1,365 players across 6 gameplay features. As presented by Drachen et al [12]

2.2 User-Centered Analytics Tools

Researchers are actively investigating how designers interact with and use instrumentation and telemetry data as part of their professional work flow within the games context, and this is informed by a long history of user-centred design (UCD) and usability testing established within the HCI field. This section is divided to discuss the existing work around these two methodologies.

2.2.1 User-Centred Design

UCD is a methodology used in order to place the user of a product at the centre of the design and development process. The role of the designer becomes as a facilitator between the product design and the end user, ensuring they are able to use the product as intended, and that it is easy to do so [1]. The UCD approach draws on a long-established focus within HCI around the usability of objects and interfaces, and

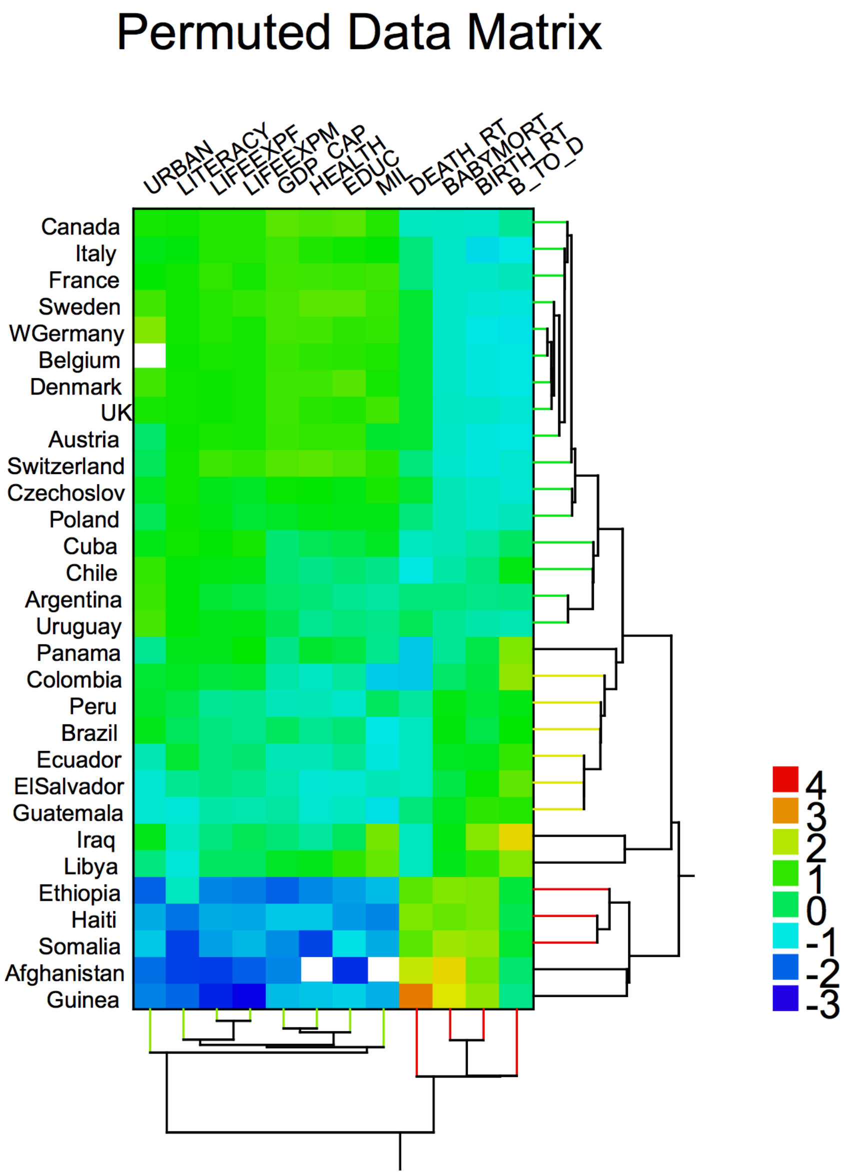

FIGURE 2.5: Cluster heat map of social statistics showing the results of hierarchical clustering, presented by Wilkinson and Friendly [48]

particularly how to make them intuitive, easy to use and efficient [37], [1]. Pagulayan et al [38] explore the challenges faced when using UCD methodology in the context of games, noting that games differ from the software and products usually designed with UCD, such as productivity tools. In a game, failing a difficult level may be desirable and intended by the game designer, in order for players to feel achievement when eventually beating it, whereas failing to format a paragraph in word processing tool would not be a desirable characteristic. As such, the authors note the careful consideration and adaptation of UCD methdology required when working with games, in order to consider both the tangible aspects (mislabelled interface buttons, difficult interaction) as well as the creative, emotional aspects (establishing difficulty in games, fostering cooperative play).

An early work by Southey et al [42] explored the possibility of using game data driven tools in order to test the artificial intelligence (AI) in a football game. They worked with the developers of *FIFA'99*, a football (soccer) game developed by EA. The game involves players taking control of various members within a football team throughout football matches, performing actions such as passing the ball, tackling opponents and shooting at the goal. The game's developers identified a problem space around the AI for a computer-controlled goal keeper. During development they would implement changes to the goalkeeper AI, and therefore wanted a way to quickly and thoroughly test its behaviour. Specifically, they were interested to find spatial points within the game world, also known as "sweet spots", where a player-controlled character would have a higher-than-normal chance of scoring a goal, potentially caused by errors or bugs in the computer-controlled goal keeper AI. The authors developed a tool that would iterate through all possible positions of player, ball and goal keeper. The success (or failure) of shots at these locations were then mapped out using a heatmap, which is then presented to a designer. The tool would then use learning algorithms to identify a set of rules for this information, such as "Striking from this point leads to a guaranteed goal every time". This is then presented to the game's developers for them to understand the impacts and or issues with the goalkeeper AI. Whilst the tool is not formally evaluated, the authors note the *FIFA'99* developers received the technology very well. The framing of this tool by the authors as a semi-automated tool is important because it allows the automation of time consuming parts of the game design process, but particularly they note the importance of the designer because, in the authors words *"such tools cannot replace the human designer, but only assist"* [42]. Moreover, the work focuses on presenting intelligible, metric summaries of the analysed data, and particularly in this case converts those metrics into rules.

Kim et al. [27], of Microsoft Game Studios, published a framework for combining game telemetry and instrumentation data together, and utilising this within the game development process. The authors present a system created that allows the linking and synchronisation of spatio-temporal player data (as heatmaps) with screen-capture video. This data, collected from lab-based user testing, allows designers to see the in-game actions of participants through the spatial data, and correlate that with the video of the participant playing the game. The data is presented in the form of an overall report, tailored to the specific aims of user-testing, with the ability to "drill down" to the underlying data of heatmaps, screen-capture video or other data. This affords the designer the ability to see a high-level summary of the data initially, as well as specific, low-level snapshots of the data when required. This framework, TRUE, was tested in two scenarios. The first used *Halo 2* play tests, combined with attitudinal surveys of participants, which the *Halo 2* designers used to tweak and adjust their level designs based on the feedback. Their adjustments were verified using a new set of participants, with the data showing the success of the changes. In the second scenario, the framework was used by the developers of a game whilst in pre-release beta stage of development, who used the insights from the data to revise and update the game during its development. Importantly, this work frames the development of the tool as part of user-centred design, which means the framework is implemented so as to help designers recognise the nuances of the virtual game environment they are working on, as opposed to replacing this knowledge discovery process. The ability to provide a high-level of data summarisation using "hierarchically organized reports" further allows designers to explore the subtleties of specific areas or behaviours within the data.

Medler et al. [30] worked with the development team of *Dead Space 2* to design a web-based system of analytics tools, Data Cracker, which visualises different types of game telemetry data, and is designed in such a way as to help answer typical analytics questions the development team may have. The system automatically processed telemetry data logged from players, and presented the data to the game development team through a web interface. A large variety of metrics were presented, such as the number of kills made by a player, the average number of players in a time period, and so forth. These are displayed through 2D graph visualisations, which the development team can explore particular areas through more detailed graphs, i.e. viewing the number of kills made by a player on a specific map. The authors use this example to inform a rich discussion about how to design a visual game analytics tool while working with a development team. This is presented as a set of guidelines for designing tools with a user-centred approach, which include physically meeting with the team often, designing the tool to be usable by a broad audience (programmers, artists, producers) and facilitating new and different types of telemetry data as development progresses. This work provides a rich, qualitative insight and discussion around the particular issues of designing analytics tools for an entire development team.

Mirza-Babaei et al [33] developed a new user-testing method which combines quantitative gameplay metrics and qualitative observations with physiological information about a player's state. Motivated by a lack of understanding around the usefulness of biometric data in user testing methods, they propose a new method, called Biometric Storyboards, which they test against traditional, long-established methods of user testing. The Biometric Storyboards system visualises physiological data, recorded from laboratory-based participants, and compares that to a graph composed by the game's designer, which visualises the intended player experience. This allows the comparison between how a designer intended a player to experience the game, and the way they actually felt the game. In order to evaluate this user-testing method, they created a prototype game, and collected a set of 6 participants play experiences, including their physiological experience data. The game development process was then simulated, with 6 game designers recruited, and asked to suggest improvements to the game as if they were developing the game themselves. The designers were split into three conditions, one where they were provided with the user-testing data in industry standard user-testing format, another where the data was provided in biometric storyboard format, and the final condition where no user-testing data was provided and they were asked to use their expert opinion alone. The subsequent improvements were then implemented, creating three different versions of the game. Finally, these three versions were tested blind by a group of participants, who rated their experiences of each game using four questionnaires. Their results show the game that utilised biometric storyboards for their gameplay improvements was rated better by all participants. The authors surmise that this user-testing method allows game designers to be more informed about the actual play experience when compared to the intended play experience, and can focus on and modify areas of the game that do not perform as expected. The actual usage and evaluation of the user-testing method within the design process produces a rich set of insights which can be drawn upon by future work around game design tools.

Gagné et al [21] worked with an independent game developer to create a vector-based 2D visualisation in order to show how players moved and made decisions within an RTS game. They note that existing approaches commonly used for FPS games, such as aggregate heatmaps and GIS, are not well tailored to the type of gameplay which occurs in RTS games, due to their different game mechanics, and the fact that a player normally controls more than a single agent. Their tool uses vectors to visualise where each player's agents move, as well as the locations that actions were occurring, i.e. deaths, capturing points, and so forth. Particularly, the tool was designed in consultation with the game developers, and therefore was designed to provide answers to their game analytics questions. These were split into macro-level questions, which deal with the players' actions in between and across matches, and micro-level questions, which concern the behaviour of players, in isolation, within single matches. They found their tool was useful for visually identifying strategies for win and loss, and demonstrate how it can be used to answer typical analytics questions, such as "Are players doing what the designer expected?" Furthermore, the authors acknowledge one limitation to their system is the inability to select and "drill down" into specific features of the data, and call for future work in this area, with respect to RTS data visualisations.

Analogously, Hoobler et al [25] studied gameplay telemetry visualisation from the context of the player, rather than game designer, presenting a system of data visualisations to augment the information provided to spectators of digital games. They identify two rather limited options that are available to spectators of multi-player online games (using FPS games as an example) with regards to the information they are able to view. The first option is to focus on a specific area of the game, and monitoring the events that take place there. The second option is to view a larger area, with the hope of achieving a more overall view. Each options have potential for spectators to either miss events, or be overwhelmed with the number of events taking place - motivated by this problem, they implement a set of data visualisations into the spectator interface of the game *Return to Castle Wolfenstein: Enemy Territory.* They present two types of visualisation functionality for spectators, local visualisations and global visualisations. Local visualisations include visualising the line of fire and line of sight for players, as well as their team and game role information as icons. Global visualisations use a heatmap approach to display the occupation of areas of the game level by the different teams, as well as the efficacy of the medic game role. By providing global information and select local information within the same interface, they are able to enrich the spectator experience, and as the authors highlight this may be useful to professional players. Whilst no user studies are conducted to study the effectiveness of the system, this work presents interesting research into the little explored area of game player (or spectator) data visualisations and enrichments.

2.2.2 Usability Studies

Dumas et al [15] provide a detailed discussion of usability testing and its benefits as a process for developing software and interfaces. They state that usability testing is focused towards improving software and products, and this is done by involving real users who are completing real tasks, whereupon you record their interactions and analyse the data to detect problems. By using real users you are able to see exactly how the product will be used when released to the wider public, and are able to evaluate whether it is being used as intended.

Within the games context, Medlock et al [31] describe their use of a usability testing method used in the production process to rapidly iterate the design of a tutorial level for *Age of Empires 2*. This process involved recruiting participants to a usability study, analysing their game telemetry data, implementing changes based on this, and re-testing to measure the efficacy of the changes. This work shows how usability studies can be used by game developers as part of their user-testing process to quickly develop and change sections of the game. It is of note that the approach used broad game telemetry metrics such as "number of failures" per player, but it did not use any spatial data.

Kriglstein et al [28] studied gameplay telemetry visualisation for players of an FPS game, employing a usability study to measure the usability and interprettability of the visualisations. The work was motivated by the growing use of player data visualisation being presented to players. For example, they cite the game *Call of Duty: Elite*, which allows players to view heatmaps of their own multiplayer activity on the various game levels, giving them the opportunity to compare and show off to others, as well as plan their future play actions. They conducted a laboratory-based usability study for players to use a set of visualisation techniques for gameplay analysis. These visualisation techniques were "traditional" single variable heatmaps, representing gameplay features such as deaths, movement, and so forth, along with two novel heatmap-based multi-variate clustering visualisations, one using polygons on the heatmap itself for cluster enclosure representation, and another using circular statistical visualisation (similar to a pie chart) of each cluster. Participants, who all had experience playing FPS games, were asked to perform 6 tasks of gameplay analysis on a corpus of data collected from the FPS game *Team Fortress 2*. Tasks involved finding the most contested areas, detecting areas where specific weapons excel, and other similar games analytics type questions. They found that the circular cluster visualisations allowed participants to gain a high-level overview of the data, and observed that they then investigated heatmaps of single variables for hotspots, as well as using the enclosure representation as a means to understand the spatial distribution of data. As noted in the paper, this work contributes greatly to the under explored area of non-expert analysis of game analytics data, specifically around game analytics in a non-developer context. Furthermore, the task-oriented nature of the usability study encourages participants to reflect on the data they are being presented with, along with their own data.

Usability studies represent a useful methodology for understanding how a user interacts with a software product or tool, and the work of Kriglstein et al [28] demonstrates the method's utility around visualisation, however it is of note that the method has not been used to evaluate the usability of game design tools from the game designer's perspective.

2.3 Summary

In this chapter the field of Game Analytics and the sub-field Games User Research (GUR) have been introduced. The literature focused on the collection and analysis of player data, through the process of game telemetry recording. The popular visualisation heatmaps were presented, with their popularity in the games industry evident in the literature. The limitations of heatmaps were discussed, along with the various alternatives and works done to address these limitations. Following this, an exploration of hierarchical clustering and dendrogram visualisation was detailed, as a solution to the limitations of heatmaps. In the latter half of the chapter, existing works that place the game designer at the centre of the design process were discussed, particularly noting the findings of each. This section was concluded with an outline of the usability study methodology as a method for evaluating the utility of game design tools with the designers who will use them. The following section details preparatory work to design and implement a telemetry framework, and the collection of a set of player telemetry data, to be used as a base dataset for the evaluation of data visualisations.

Chapter 3

Collection of Telemetry Data

In this chapter preparatory work, conducted in order to collect a set of player telemetry data for use as sample data when implementing and evaluating game data visualisations is detailed. The majority of games released do not provide an access to their source code, primarily due to the confidential and highly competitive nature of the games industry. Some games provide limited access to their source code, but this access is usually limited to code that controls game play and user interface functions, and rarely the low-level parts of the game code that deal with telemetry or computer hardware. Additionally, games companies do not often release sets of telemetry data, again due to the confidential nature of their business. Practically, this means that to obtain a data set of game telemetry, a framework needs to be implemented into a game which provides sufficient access to its source code and underlying functions. This chapter outlines the design and implementation of a telemetry recording framework in a game, in order to gather a set of spatial game telemetry for use in in spatial visualisations.

3.1 Design

To collect a suitable spatial data set, it was imperative to identify a game that provided access to the right parts of the game engine in order to implement telemetry recording functionality. It was decided to select a game that featured a spatial environment (i.e. 3D game world) and that had an accessible code base where telemetry recording could

FIGURE 3.1: Screenshot of FPS game *Red Orchestra: Ostfront 41-45*, showing German soldiers in RO-Lyes_Krovy

be implemented. Often, videogames do not provide access to telemetry recording infrastructure due to the proprietary nature of this process, and thus implementation of a telemetry framework is required in order to collect a dataset for experimentation. This data provides a corpus for computation in spatial data visualisations.

The team-based multiplayer first-person shooter *Red Orchestra: Ost Front 41-45* (RO) provides an open and accessible code base, and was selected to implement a telemetry collection framework. Developed by Tripwire Interactive, and released in 2006, RO is an FPS set during World War II, and is focused on the conflict on the eastern front between German and Russian forces. The games feature both infantry combat, and vehicular combat, with a number of game levels that focus solely on infantry combat. Screenshots of the game can be seen in Figure 3.1 and 3.2. The infantry combat modelled in focused towards tactical gameplay, which is similar to other games in the FPS genre, often referred to as "tactical shooters", such as the *ArmA* series, some titles of the *Call of Duty* series, and the *Rainbow Six* series. In these games, along with RO, gameplay is

FIGURE 3.2: In-game screenshot of FPS game *Red Orchestra: Ostfront 41-45* showing waterfront combat in RO-Danzig.

predominantly concentrated around multiplayer gameplay, with players split into different teams, with an emphasis towards realism and the use of tactics in gameplay. There are a number of in-game classes which players are asked to chose between, which sees them using different weaponry and equipment, such as sniper rifles, sub machine-guns and heavy machine guns. As such, the gameplay of RO is widely representative of other games in the FPS genre, and with an accessible code base, RO presented a suitable game in which to implement a telemetry recording framework.

Tripwire Interactive released the Software Development Kit (SDK) for RO in 2006 alongside the game's official release. This tool is distributed via their website, as well as through the *Steam* digital games platform. The SDK gives access to the editable source files for all game levels, models and a large portion of the source code. It also includes a level editing tool and a model viewer. RO is build on the Unreal 2.5 engine, a proprietary game engine developed by Epic Games, released in 2006. As such, the RO SDK does not give access to source code relating to the base engine, as this is proprietary.

However, it does provide access to all code used in RO, meaning that total modification of the game is possible.

The code base of the RO SDK contains over 100 object classes, and is written in a mixture of C++ and *UnrealScript*, a Java-like proprietary scripting language [44]. This architecture reflects the architecture of Unreal Engine 2, which uses C++ for low-level functions such as ray tracing and rendering, and UnrealScript for higher-level functions, usually pertaining to game logic. Due to the multiplayer architecture of the game, the code base included both client and server versions of game code.

To complement the SDK provided by Tripwire Interactive, the community has developed *WOTGreal* as an integrated developer environment (IDE) that can be used to explore UnrealScript, outlining and highlighting links between classes (inheritance) as well as providing tailored IDE functions such as searching for usage of an individual class, searching for definition, and so forth.

3.1.1 Telemetry Framework Design

RO uses a client-server architecture when running multiplayer games. This involves a central server "hosting" the game, with multiple clients connecting. The server controls the game environment, and thus processes and tracks every action in the game world. This client-server architecture can be leveraged in order to collect telemetry data from players, as explained by Duchenaut and Lee ([17], p.641-664) when discussing this architecture in Massively Multiplayer Online Games. In the RO context, this would entail recording the actions of players at the server side, removing the need for game clients to transmit extra information.

The Unreal 2.5 engine on which RO is built does not contain any capability for interfacing with a database, and as such an alternative method was found to output the telemetry data into a parsable format. The functionality of the game engine was explored and it was found to contain functionality to output information to the developer console, initially intended for debugging use during the game's development. The practice the developer console is filled with many warnings and debugging messages during normal gameplay. The entire contents of the developer console is saved at the end of each game as a .log file in the game's root folder. Due to the quantity of developer messages these log files, it was impracticable to do this for telemetry recording. By modifying the source code of RO, it was possible to create individual telemetry files for each player for each game match, thus maintaining clarity and separation of player data during the parsing process.

3.1.1.1 Gameplay Features

In line with existing Game Analytics practice ([17], [10]) a set of in-game features/events were selected that encapsulate the gameplay experience in RO. These features can be seen in Table 3.1, and broadly cover the key parts of gameplay, including movement, combat and deaths. A standard format was defined for each telemetry record, in order to capture key contextual information related to the player. This format can be seen in Figure 3.3. In 3.3 *eventID* refers to which type of event is being logged, as defined it Table 3.1. *timestamp* is the time stamp of when the event happened, in HH:MM:SS format. *xyz* is the X, Y and Z coordinates in the game world where the event happened. *running?, crouched?, prone? and wounded?* are all flags which denote the key states a player can be in.

```
<eventID>|<timestamp>|<xyz>|<running?>|<crouched?>|<prone?>|<wounded?>
```

FIGURE 3.3: Standard telemetry framework data format for RO.

EventID	Name	Description
0	Shoot	Firing of weapon
1	Move	Location of player
2	Reload	Reloading of weapon
3	Die	Death of player
4	TakeDamage	Receiving damage
5	Change Class	Switching player class

TABLE 3.1: RO telemetry framework event type IDs and their respective in game data types.

The logging of player movement, through the recording of the spatial location of the player, at a rate of 1 hertz, was experimentally determined as providing an acceptable balance between data granularity and data overload. Due to the movement speed of players in RO, the distance moved during the 1 second recording interval, even at the fastest player speed, was not enough to cause recording issues with player movement. When a player receives any damage, whether self-inflicted, this is recorded using the

Event Type	Specific Information
Shoot	Weapon Name, Ammunition Count, Weapon Rotation
Movement:	Stamina
Reload:	Weapon Name, Ammunition Count
Die:	Killer
TakeDamage:	Damager, Location on Body, Damage Amount
Change Class*:	Player Name, Class Name
*Only records this information.	

Table 3.2: Event specific information recorded for each event type recorded in RO.

TakeDamage event. In cases where a player takes more damage than they have health, the Die event is triggered. Shoot and Reload record when and where players fire and reload their weapon. Change Class records when players switch in-game class.

3.1.1.2 Event Specific Information

Some gameplay information in RO is event specific, and therefore it would be useful for subsequent analysis to capture this event-specific information in the telemetry. A breakdown of this information and their related event types can be seen in Table 3.2.

For the *Shoot* action, it is pertinent to know which weapon is being fired, how much ammunition it has and which direction it is facing. The latter is important if using an approach such as trajectory visualisation, as outlined in chapter 2. When recording *Movement*, Stamina is a decimal value between 20.0 and 0.0, representing how tired the player is, with 0.0 representing complete exhaustion. Similarly to the shoot action, *Reload* records the weapon and ammunition count when reloading. For *Die* the person who provided the killing hit is recorded. *TakeDamage* records who is doing the damage, along with where and how much damage to apply. Finally the *ChangeClass* action only records the player name and class name, as the event occurred infrequently and often when players had died, therefore providing no useful spatial information.

3.2 Implementation

The initial implementation work focused on modifying the RO game engine to record and output the telemetry data in the manner described in the previous section. With this accomplished an experiment was conducted, involving participants playing the modified version of RO, in order to generate data for analysis later in the study.

A data logger was implemented within the RO engine to log files for each player during each in-game match. To ensure the telemetry recording was only performed by the server, the data logging functions were placed in the server-side parts of the code that controlled the players connecting to the server. To maintain order when working and parsing the log files, a standard naming format was created for the log file names, which can be seen in Figure 3.4. Each time a player connected to the server, or a new match was started, a log file would be generated on the server using the specified format, where all of their telemetry data would be recorded for the duration of that match.

```
<year>_<month>_<day>_<hour>_<minute>_<mapName>_<playerName>.log
```

Figure 3.4: Standard form for telemetry log file names in RO.

The player class and associated weaponry classes of the game engine provided access to all of the core gameplay features detailed in 3.1, and logging calls were placed in the relevant methods in order to capture the telemetry data required.

With this work done, a brief testing period was initiated by the author using multiple PCs, to ensure that as players connected a log file would be created and would record their actions through the game correctly. These modifications were then compiled and packaged into a binary UnrealScript file (*.u). The modifications above are limited to a single module of the game engine, and as such only a single file, *ROEngine.u* was modified from the original. During deployment and testing the authors discovered another technology implemented by the developers to maintain game continuity (and freedom from bugs). The server requires that all clients are running the same version of the game without any modifications. With the server running a modified version of the game (ROEngine.u), it was required to distribute this file to client machines that were to connect to the server. ROEngine.u totals 2.5Mb in file size and therefore was not difficult to distribute and install using a batch script.

3.2.1 Data Collection Procedure

A convenience sample using staff and students at University of Lincoln provided the quickest method of collecting an initial data set, to be used in the generation of visualisations. The rationale for using this method was that a large, generic dataset was required for use as input data for the visualisation process, and this sample would provide a wide range of player behaviours due to differing playing abilities and playing styles. To this end, laboratory space at the university was used.

Informed consent was obtains from participants via a consent form, which explained that the data collected through this collection process would be fully anonymised, and that the anonymised data would be used in visualisations in later experiments. A laboratory was sourced at the School of Computer Science, within the University of Lincoln, UK. The laboratory consisted 60 desktop PCs, each one with a single 24-inch widescreen monitor. All PCs were identical Dell XPS gaming machines, with Intel Core2 Duo Quad processors, 8GB of RAM and two NVidia 240 GTS graphics cards running in SLi configuration. Access to Red Orchestra game was provided by the School of Computer Science's Steam CyberCafe account. This is a service provided to Internet cafés, as well as academic institutions, providing access to a wide range (1000+) of games on available on the Steam network, including RO.

Each machine was installed with RO, along with the modified binary *ROEngine.u*. Another machine was configured to run the game in dedicated server mode. This runs the game in server mode, from a command line prompt, advertising the game to other players. The server was stored in the Technician's office, being allowed to remain on at all times.

The laboratory used was also an active teaching space throughout the week, and as such all data gathering sessions were not able to be run consecutively. In light of this, it was arrange the data gathering would be conducted on Fridays at 4pm until 6pm. Due to the teaching usage of the laboratory throughout the week, it was the authors responsibility to check that each machine conformed to the experimental configuration (RO installed, modified binary installed) before commencement each week.

Participants were recruited from the staff and students at the University of Lincoln. The sessions were advertised amongst the Schools staff and students, via email as well

Figure 3.5: Computer laboratory used for data collection sessions.

as promoted at undergraduate Games Computing and Games Production lectures. It was envisaged to run the game server accessible through the Internet, but due to the restrictive University network policies, it was not possible to do this, and the server was restricted to the local area network.

Overall 50 sessions were run over the space of one year, with the majority of sessions being clustered in the academic year (September - May), due to the availability of participants. If too few participants were available to play a match (less than 6), it was decided to postpone the session, in order to maintain integrity in the data collected. Each session lasted approximately two hours, usually consisting of 2 matches, with players joining one of the two available teams. Players were allocated between the two teams to ensure the number of players per team was as close to even as possible. The sessions were held between February 2011 and August 2012.

The game server rotated through three game levels during the experiment, to simulate the behaviour of online multiplayer game servers, as well as provide variety to the player experience. The three levels selected were RO-Danzig, RO-Basovka and RO-Lyes_Krovy. The server randomised the order of play every time. These levels were selected due to their infantry-only focus, providing a telemetry dataset that is broadly generalisable to other FPS games. Furthermore, the three levels selected represented a wide range of combat environments: RO-Danzig (Figure 3.6 and 3.2) is an urban level with many routes through buildings, with combat taking place on multiple building stories, and with relatively short line of sight for players. RO-Basovka (Figure 3.7) is an exposed, rural setting, with combat taking place in a field with relatively little cover, and a set of trenches, thus encouraging long-range combat, as well as extreme close-range combat. RO-Lyes_Krovy (Figure 3.8 and 3.8) is a forest-based level, with combat taking place through fortified trenches set into a hillside.

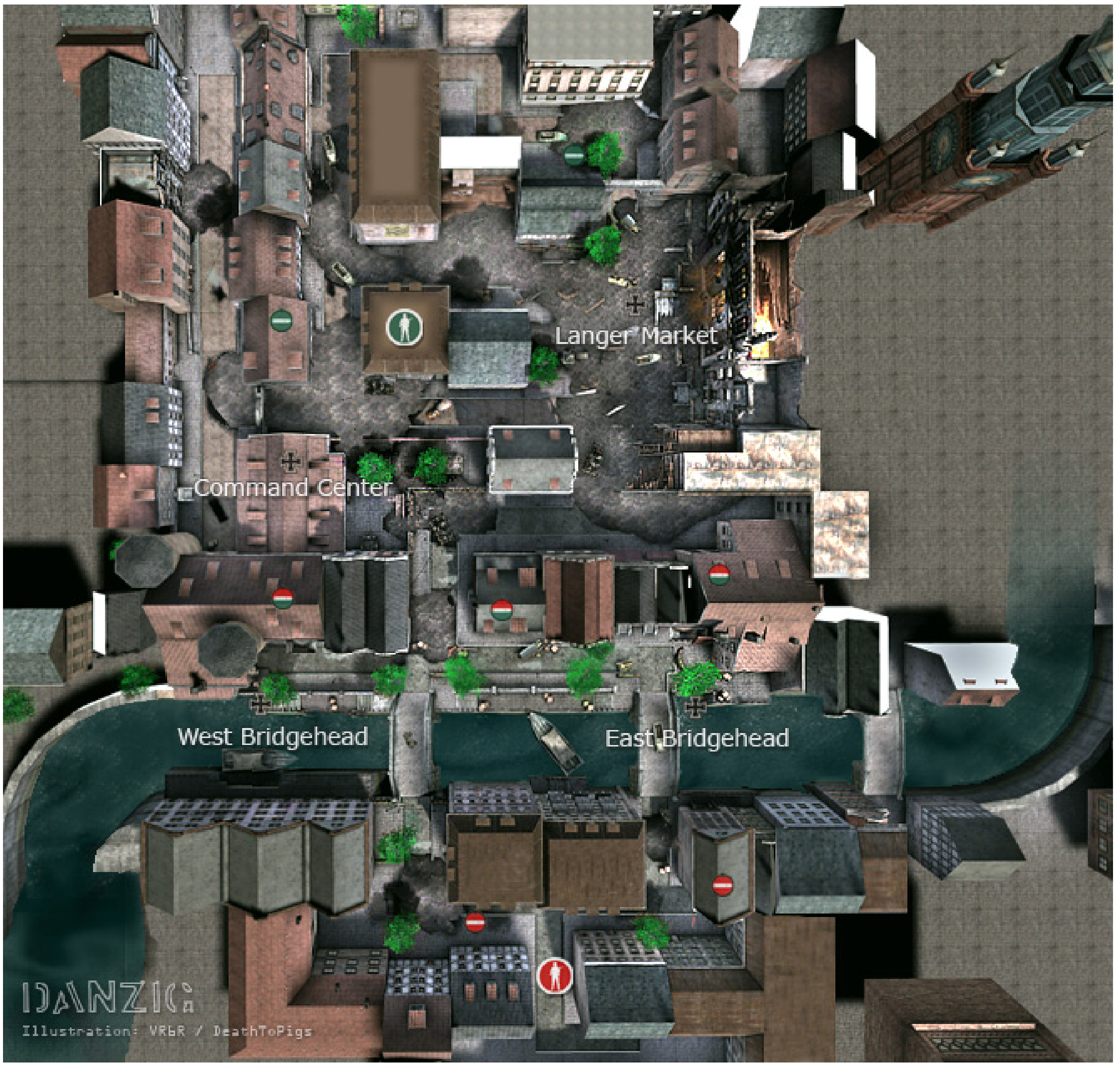

FIGURE 3.6: Overhead view of RO level RO-Danzig.

3.2.2 Telemetry processing

Following the collection of player telemetry, the final step to prepare the dataset was to parse the raw telemetry data into a database so it could be queried. In line with industry

FIGURE 3.7: Overhead view of RO level RO-Basovka

standards [17] it was decided to process the data into an SQL database, specifically SQLite[1] due to its portable file format. To parse the raw telemetry data, as well as interface with the SQLite database, the programming language Python[2] was chosen.

The initial technical hurdle when processing raw player telemetry, as in all data mining operations, involves *data scrubbing* and *data cleaning*, the process of checking, correcting or removing data that does not conform to the required standard [17]. In this instance, after inspection of the raw data, it was decided to use a set of criteria for data cleansing, defined as follows:

1. Log file must contain minimum a single record of player data.

[1] www.sqlite.org
[2] www.python.org

FIGURE 3.8: Overhead view of RO level RO-Lyes_Krovy

2. Log filename must be correctly formatted, as defined in Figure 3.4 on p38.

3. Log fields must conform to specification, as defined in Figure 3.3 on p36.

To ensure the data was stored in the most robust and accessible fashion, a comprehensive schema was defined for the database, which can be seen in Table 3.3. In order to reduce data redundancy, the structure and attributes of the input data was normalised, particularly around the replication of player details and game level details. This resulted in a database schema where player telemetry records were stored in the *Events* table, with relational links to player information, game level information and gameplay feature specific tables.

Name	Function
Event Types	Types of events logged
Events	Store each log for event
Game	Games (Red Orchestra, other games)
GameInstances	Multiplayer matches
Maps	Map details
Players	Player details
ProcessedLogs	Record of filenames parsed into database
RODeathData	Specific data for death records
ROShootData	Specific data for shoot records
ROWeaponID	Weapon information
TeamData	Players per team records
TeamTypes	Team types and their IDs
histograms	Raw histogram data

Table 3.3: Database schema used for storing parsed RO player telemetry data.

No. players:	No. matches	Total records
152	171	1,296,877

Table 3.4: Total number of players, number of matches and total number of records in preparatory dataset.

Python scripts were created to cleanse the data, and simultaneously import it into the SQLite database. With the data in a database format, it was possible to retrieve data quickly for further processing via queries. An example SQL query to gather the total number of shoot event records for a specified player:

```
SELECT COUNT(*) FROM Events WHERE playerID=? AND eventTypeID=0
```

In order to facilitate the querying process, a library object was created, named *dbtools*, which contained the database connection, and a large number (40+) of library functions created in order to reduce programming repetition. This library allowed the quick access of functions and complex queries, reducing development time.

3.3 Results

In total, 5681 individual log files were produced during the data collection sessions, totalling 85MB in file size. An overview of the collected dataset can be seen in Table 3.4, with a breakdown of all telemetry records by event type can be seen in Table 3.5.

Move	Shoot	Take Damage	Death	Reload
1,150,389	97,811	27,924	14,276	6,477

TABLE 3.5: Breakdown of total number of records by event types from preparatory dataset.

3.4 Summary

In this chapter, the design and implementation of a telemetry recording framework has been detailed, noting the considerations required to capture the desired data features that will be used in further analysis. Details of data collection sessions conducted over 18 months were noted, where a variety of participants were recruited in order to collect a large dataset of player telemetry. Having been parsed into a database, this large corpus of telemetry data collected from the game *Red Orchestra* provides a dataset representative of many games in the FPS genre, and therefore this data can be used with confidence during the evaluation of visualisation techniques in the following chapters.

Chapter 4

Data Visualisation Techniques

In this chapter the theoretical grounding of the game data visualisation techniques being drawn upon in this work is provided, and the design and implementation of these techniques, in order to work together as a usable system for game designers to analyse spatial data, is detailed.

4.1 Technical Introduction

The following sub-sections describe the technical process behind the data processing and each of the visualisation techniques presented in the thesis, along with examples explaining their function.

4.1.1 Heatmaps

A heatmap is comprised of two components: a two-dimensional histogram representing the frequency of data over an area, and a meaningful underlay image. The two-dimensional histogram is rendered using a hot-vs-cold colour scheme, allowing the viewer to distinguish between the values of each bin within the histogram. The underlay image is usually a two-dimensional image of the spatial area being studied, such as a topographic view of a game level (see Figure 4.3a). Heatmaps receive their name from the use of the hot-vs-cold colour scheme, which displays a high frequency as a hot colour (red/white) and a low frequency as a cold colour (blue/black). This intuitive visualisation allows a viewer to quickly understand the distribution of a data observation over, say, a game level.

A spatial histogram is a discrete representation of the environment, which consists of an n x m number of bins (see Figure 4.3b). Each bin relates to a rectangular region of the spatial environment, and each bin stores the total frequency of the occurrence of the data observation. Each data observation is processed, and the value of a bin is incremented when a data observation falls within that bin. After all data has been processed, the resultant histogram will show distribution of the frequency of the data observations across the spatial area.

By combining the resulting histogram, visualised with hot-vs-cold colour scheme, with the underlay image, a heatmap is created. Examples of this can be seen in Figure 4.3. There are two critical parameters which affect heatmap visualisation: the number of bins over the spatial area, and the amount of data observations being processed. Using a higher number of bins means each bin covers less spatial area, becoming higher resolution, with a lower number of bins decreasing the resolution, see Figure 4.3 d and c respectively. It should be noted that the resolution of a histogram should be tuned to the quantity of data being processed, as too high resolution can result in overly sparse histograms.

4.1.2 Hierarchical clustering & dendrograms

Clustering is the process of "partitioning a data set into subsets or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters to some defined distance measure" [7]. Specifically, hierarchical clustering is the process of clustering that involves measuring the distance between each data observation and every other data observation, in order to build a distance table. This distance table represents the distance (of a specific measure, such as Euclidean distance) between each observation, and can be used to build a hierarchy of relationships, detailing how (dis)similar each observation is to every other.

Hierarchical clustering results are often visualised as a dendrogram. Dodge [7] defines a dendrogram as "a graphical representation of different aggregations made during a cluster analysis... The structure of the dendrogram is determined by the order in which the aggregations are made." From the Greek *dendro* for tree, a dendrogram is a tree-like representation of the relationships between a set of observations. Extending the tree nomenclature, *leaves, branches* and *trunk* are all parts of a dendrogram. The *leaves* of a dendrogram are the original observations. Each of these leaves is connected to the rest of the structure using *branches*. The length and path of these branches denotes the relationship, or distance, between two leaves. Eventually, every leaf, via a set of branches, is connected to the *trunk* of the dendrogram. This is the base of the graph, and encompasses of the entire data set of observations. It is therefore possible to connect any leaf to any other leaf, using a number of branches. A visual illustration of this structure can be seen in Figure 4.1.

An example of a dendrogram based on the hierarchical clustering of a set of gameplay data histograms can be seen in Figure 4.4.

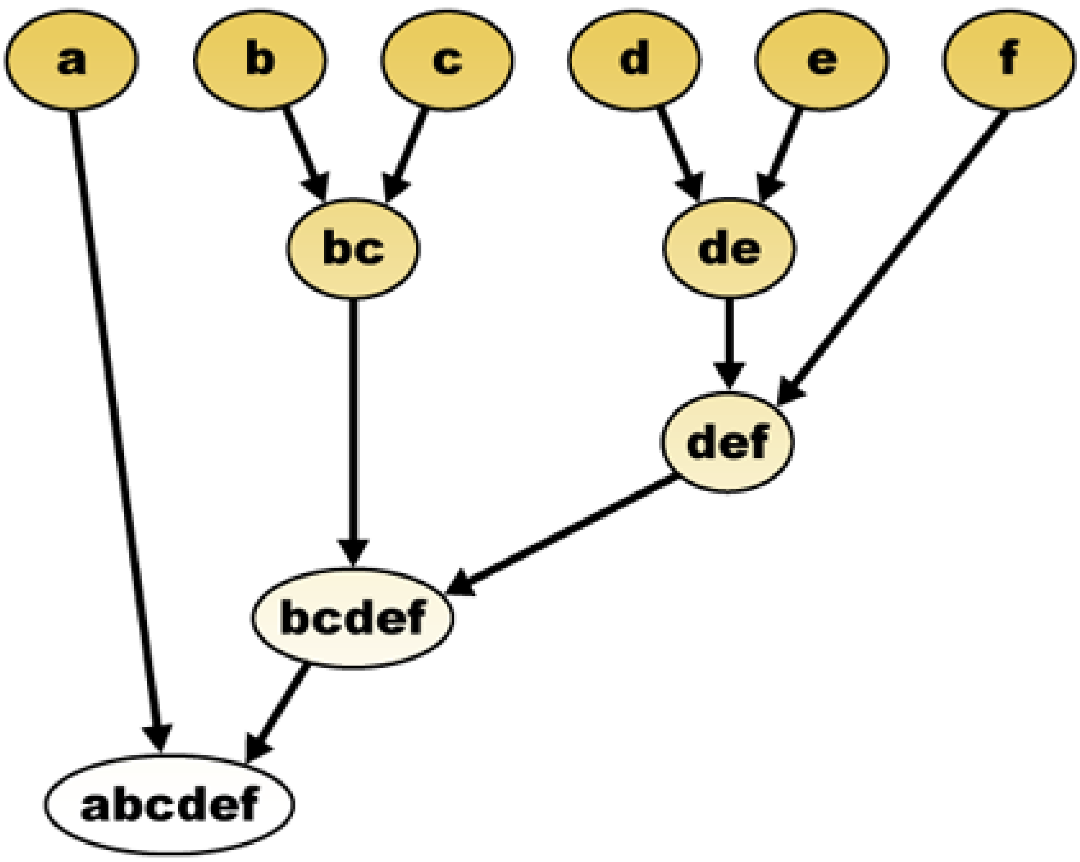

FIGURE 4.1: Example of hierarchical clustering via a dendrogram. Diagram by Stathis Sideris available under Creative Commons Attribution-Share Alike 3.0 Unported.

4.2 Implementation

The following section details the technical implementation of programming code to generate heatmaps and dendrograms from the pre-processed game telemetry data. To summarise the overall function of the proposed system of visualisation, Figure 4.2 describes the position of the system in the overall game design process. Following user-testing data being collected and processed, heatmaps are generated for each individual player, for each of the core data features. These heatmaps are then hierarchically clustered using Euclidean distance comparison, and then visualised on a dendrogram. These dendrograms are then displayed to a game designer for visual analysis. In order to implement this system, the Python programming language was chosen as the scripting language with which to manipulate data and produce visualisations. This is due to it's flexibility, expansive built-in libraries catering to data visualisation and easily-readable syntax.

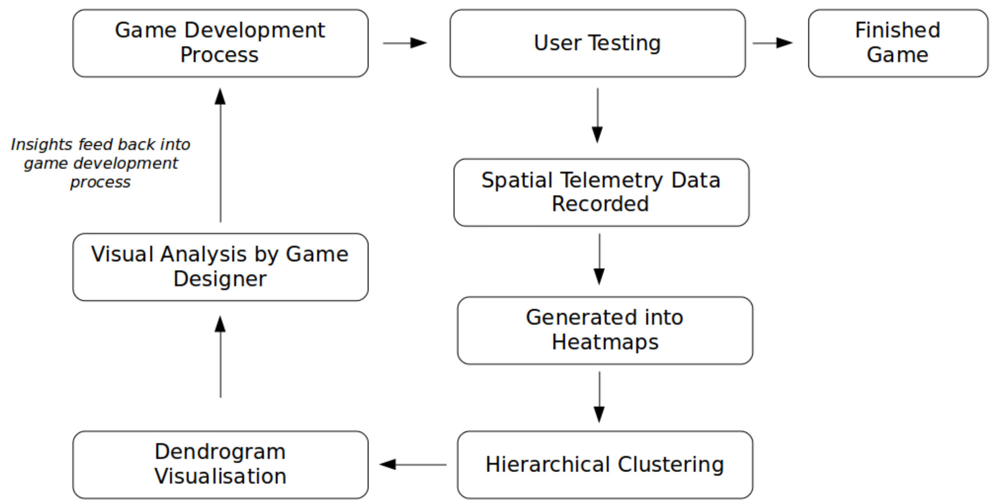

Figure 4.2: Overview of dendrogram visualisation within overall game-design process.

4.2.1 Heatmaps

To generate histograms, the *NumPy*[1] and *Matplotlib*[2] Python libraries were used, which have extensive high-dimensionality array functionality, and graph plotting functionality, respectively. As outlined above, Python was used to query the SQL database. A histogram was generated for each player on every match of RO-Danzig within the data set. This involved an iterative process, requesting the telemetry records for each player per match for each of the basic event type (*shooting, moving,* etc). From these records, only the x and y coordinates were used, which were stored in a *NumPy* array. The result was an array filled with all the spatial information related to that event.

The next process was to create a histogram from each of these. NumPy's *histogram2d* function was used to produce a 2D histogram from input data. The *dbtools* library was used to initialise some default variables associated the creation of histograms, such as the number of bins in the histogram. Whilst work has been done to calculate the optimal bin size [41], these works are often built on assumptions such as normal distribution of data observations, which is not always the case with game telemetry data.The number of bins was experimentally defined, and a usable range was found between 16x16 through to 128x128. Once a histogram was generated by NumPy, Matplotlib was used to visualise the histograms. An example can be seen in Figure 4.3c and d.

An unforeseen technical implementation was that of another SQLite3 database to store the generated histograms. Based on the number of telemetry records multiplied by the amount of histograms being generated at different bin sizes resulted in significant processing time (10 minutes+) for large queries. It was decided that storing these generated histograms as Binary Large Objects (BLOB), in a database. The BLOB data type allows the storage of binary objects, allowing Python to quickly load histograms without need for converting or parsing the information. As such processing time was reduced to sub-2 minute for most queries. In order to further reduce processing time, it was decided to generate each histogram at different bin numbers from 1x1, through to 64x64.

As a result, the database contained 91 players in total. For each of these players, histograms were generated for each match they played in. For each of these matches,

[1]www.numpy.org

[2]www.matplotlib.org

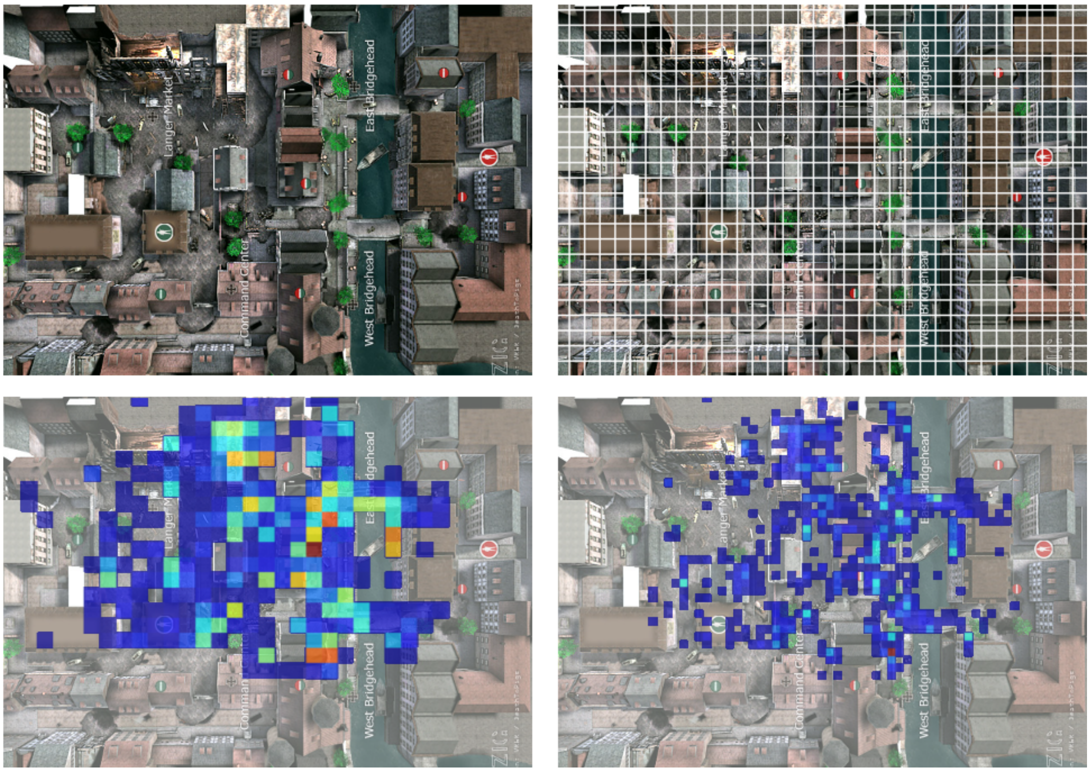

FIGURE 4.3: Illustrated example of the creation of a heatmap, and the effect of variation in bin number.
Clockwise from top left: a) RO-Danzig Map. b) RO-Danzig with white grid lines depicting histogram bins. c) RO-Danzig map with overlaid histogram (32 bins). d) RO-Danzig map with higher resolution (64 bins)

histograms were generated for each of the five event types. For each event type, histograms with bin numbers from 1x1 through to 64x64 were generated. This can be approximated into the following formula per player:

`no. of matches x (64 bin numbers x 5 event types) x teams`

It is important to note that when players switch team, they create a separate histogram, and thus there can be an Allied and an Axis team histogram for each player for each match. As a result, the total number of generated histograms for 91 players is 79,770. Whilst this process took a large amount of time to complete processing, the subsequent time-saving was significant, with a resultant database using 2.2GB of disk space.

4.2.2 Hierarchical Clustering & Dendrograms

With a corpus of individual heatmaps generated for each player for each of the event types, computation of hierarchical clustering and dendrogram visualisations could then be generated. In this implementation, the base data observations were the individual heatmaps of player events. Euclidean distance between each heatmap is used to compare all heatmaps. In effect this means heatmaps which are visually similar will be clustered together, whilst heatmaps that are visually dissimilar will not be clustered together. When analysing player movement, for example, clusters represent players whose spatial movement patterns are similar.

NumPy and Matplotlib Python libraries were used, as before, to manipulate arrays and plot/visualise data as previously, along with the *SciPy*[3] scientific Python library. Clustering was performed on a game match by game match basis, producing a set of hierarchical clustering data for all players within a specific game match. Each event type was processed separately, ending up with five sets of hierarchical clustering.

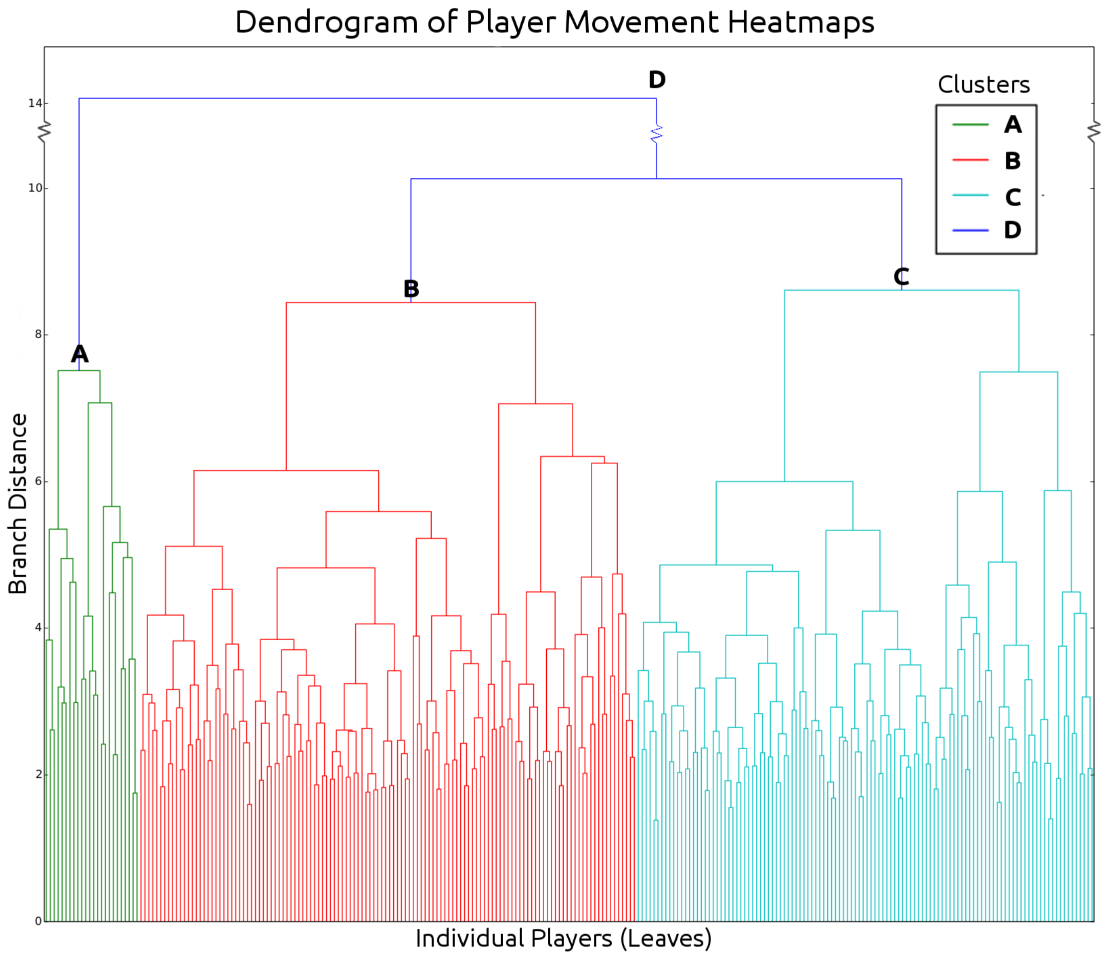

FIGURE 4.4: Dendrogram visualisation generated using 300 gameplay instances from RO-Danzig.

[3]www.scipy.org

In order to produce a set of hierarchical clustering data, the heatmaps for each player within a match were loaded from the SQLite database and stored in an array, of length m. An L1 norm (Euclidean distance) comparison was then performed between each observation, resulting in a distance matrix of m x m dimension. The SciPy function *cluster.hierarchy.linkage* was then used with the distance matrix to perform hierarchical clustering. There are two commonly used algorithms used in hierarchical clustering, single-linkage and complete-linkage. Dodge [7] notes the two algorithms are similar, with single-linkage measuring the minimum distance between two observations, and complete-linkage measuring the maximum distance between two observations. SciPy includes both of these methods, and through a process of experimentation complete-linkage was found to produce the most distinct clustering results.

Following this, the *cluster.hierarchy.dendrogram* function can then be used to plot the hierarchical clustering results. At this stage it is important to specify the threshold at which clusters are defined. This threshold is crucial to define in order to obtain a useful number of clusters within the visualisation, as too high will result in only 1 visible cluster, whilst too low could result in n number of clusters. The threshold does not change the actual clusters of data on the dendrogram, merely the colour coding. In Figure 4.4, it can be seen that the threshold is set to between 8 and 10 on the Y axis, resulting in four clusters. If the threshold was set to 2 on the Y axis, it would result in hundreds of clusters. For each event type, it was experimentally defined to suit each dendrogram.

To explore the stability of the clusters within the dendrogram, the dataset of 300 gameplay instances seen in Figure 4.4 was randomly split into two sub-datasets (1 and 2) of 150 gameplay instances. Each of these sub-datasets had hierarchical clustering performed on them, followed by visualisation in a dendrogram. The resultant dendrograms for the two sub-datasets can be seen in Figure 4.5 and 4.6.

Figure 4.4 contains three main clusters (A, B, C). The base observations used were heatmaps of individual game instances for each player, with the heatmaps representing the player's movement in the game world for a given game instance. Broadly, cluster A contains players who were on the Axis team, whilst cluster B contains players who were on the Allies team. RO contains areas of the game level that are restricted to one team (e.g. Allied spawn point is restricted to Allied players only) and as such this asymmetry in player movement is expected. Cluster C contains players on both teams, however their movement is concentrated in specific areas, and may have either been camping or playing for a short amount of time.

Referring to the sub-datasets 1 and 2, Figure 4.5 contains three main clusters (A, B and C). The two largest clusters, B and C, contain predominantly Allied and Axis players (respectively) who have heatmaps well populated with movement data, showing movement across many parts of the game level. Cluster C, represents players from both teams whose movement heatmaps are sparsely populated, or with short linear movement trajectories through the game level. Figure 4.6 contains four main clusters, (A, B, C, D). The two largest clusters, A and B, contain Allied and Axis players (respectively) who have heatmaps well populated with movement data, with movement across the game level. Cluster C comprises predominantly Allied players with long, linear movement paths through the level, thus differing to those contained within cluster A. Cluster D contains players from both teams, with sparsely populated heatmaps or short linear paths through the game level.

Comparing the two dendrograms produced by clustering sub-dataset 1 and 2 (Figures 4.5 and 4.6), it can be seen that they have similar structure, containing a large cluster of Allied players and a large cluster of Axis players, who have lots of movement data in their heatmaps. They also contain at least one outlier cluster, representing players with less-well populated or differing movement patterns in their heatmaps. Comparing the structure of dendrograms for the two sub-datasets with that of the "main" 300 player dendrogram (Figure 4.4), the Allied/Axis/Outlier structure is identical.

It should be noted that in both Figure 4.5 and 4.6, different thresholds were used in order to obtain colour-coding of clusters at a high-level (with a small number of clusters containing a large number of observations). During the evaluation of the dendrogram tool this is a function that is not exposed to users and is experimentally defined (as detailed previously in this chapter). However, a discussion of exposing this function as a user-manipulatable control can be found in Chapter 6.

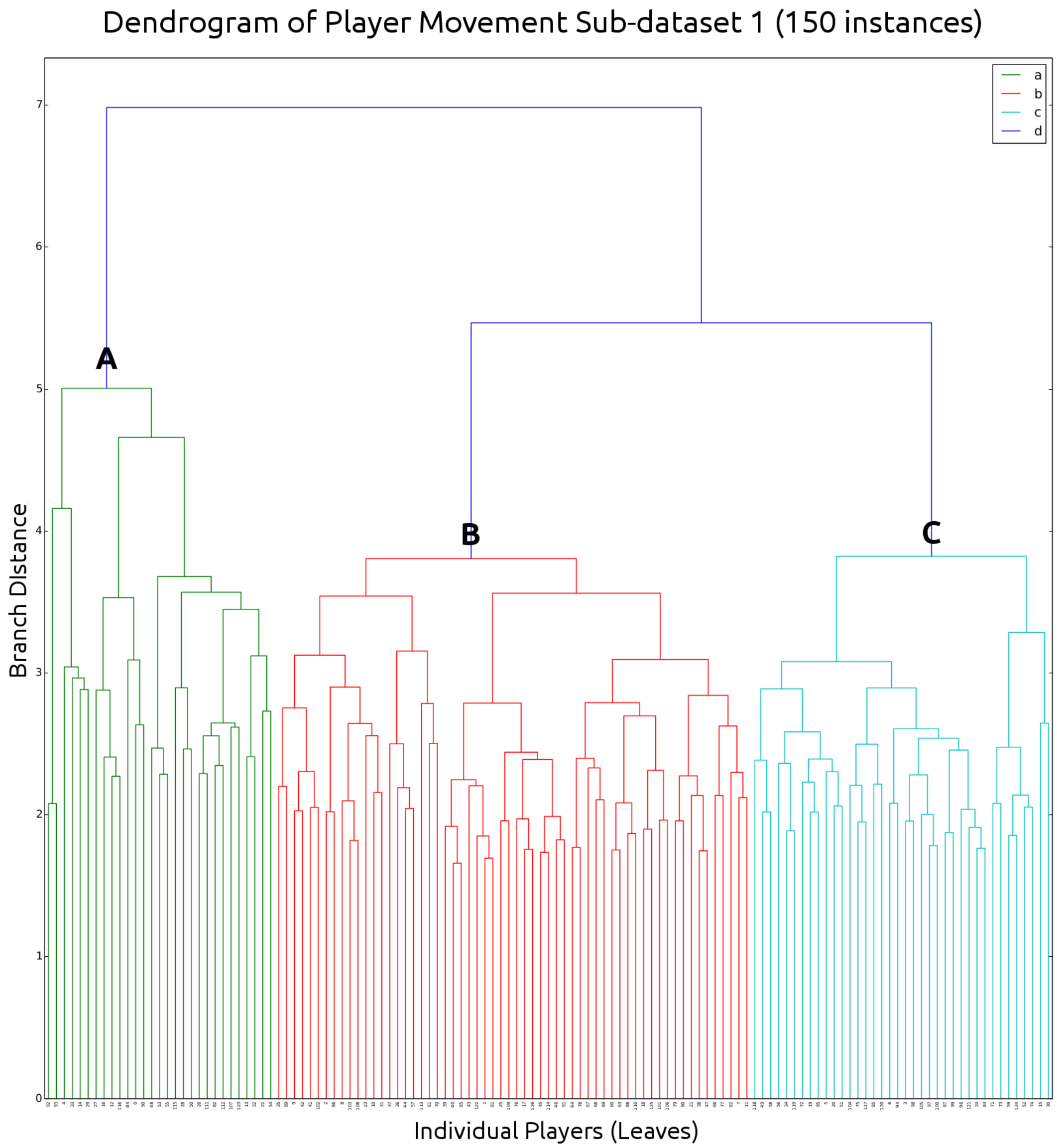

FIGURE 4.5: Dendrogram visualisation generated using sub-dataset 1, containing 150 gameplay instances from RO-Danzig.

4.3 Summary

The first half of this chapter provided the formal definitions and technical details of heatmaps, hierarchical clustering and dendrogram visualisation. In the second half of the chapter the technical implementation of the dendrogram visualisation system was detailed. Using the player telemetry data collected in Chapter 3, the Python programming language was used to generate heatmaps of the spatial data. These heatmaps, stored in a database, were then used as the base observations for hierarchical clustering. The results of this hierarchical clustering were then visualised using a dendrogram.

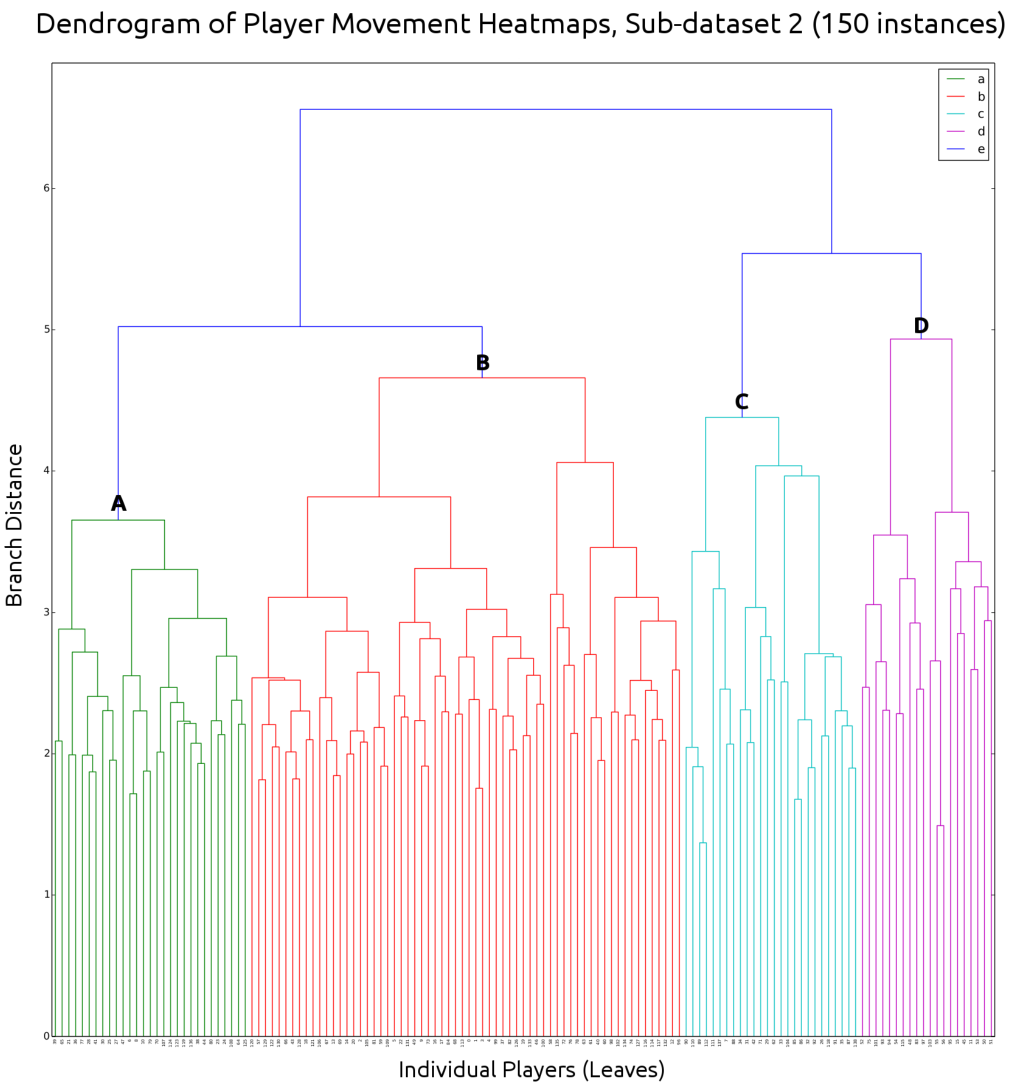

FIGURE 4.6: Dendrogram visualisation generated using sub-dataset 2, containing 150 gameplay instances from RO-Danzig.

The following chapter describes a feasibility study conducted using the heatmaps and dendrograms, in order to explore their utility as a game design tool.

Chapter 5

Study 1: Feasibility of Using Dendrograms in the Design Process

Having collected a set of game telemetry data (Chapter 3) and implemented a system for clustering and visualising spatial data (Chapter 4), this chapter details the design, implementation and results of a study to evaluate the feasibility of dendrograms as part of the game design process.

5.1 Design

To understand the way dendrogram representation might be used by game designers when exploring design queries, a pilot study with professional game designers was designed. This would allow a group of professionals to provide feedback on the potential for dendrogram representation to be used in the design process as a tool for feedback about player behaviour. As explored in the literature review (Chapter 2), working with and discussing game design tools with the professionals who will use them is important, as they are better able to judge how a tool might be used in practice. Using the guidelines presented by Medler et al [30], specifically "produce early visual prototypes", we designed a feasibility study using game designers in order to show some examples of dendrogram visualisation as a solution to clustering of spatial data sets. In order to receive rich feedback from the game designers, hypothetical design scenarios were created, with the designers asked to speculate how they might use different spatial tools (heatmaps and dendrograms) to solve the scenario. These scenarios focused on the game design process and tools place within it, by asking designers how the tools might be used, rather than requesting the tools be used to receive correct answers. As such, the designers were able to speculate on the suitability of the tools, as well as improvements or changes required to fit their desired usage.

To this end, the interview was divided into two distinct parts, with the first part covering the interviewee's design experience and usual methods for obtaining feedback on their game design decisions. The second part consisted a set of hypothetical design scenarios where the interviewees were asked to suggest ways of balancing a multiplayer first-person shooter level by using the visualisations provided to them. In total four design scenarios were devised, with two concentrating on the usage of heatmap visualisations, and two concentrating on the usage of dendrogram visualisations. Each scenario was designed to reflect the question-based utilisation of game analytics data by game development teams (e.g. [21], [17] p.138). Questions are structured as follows:

`<Set scene><Problem through data><Use data to understand>`

By creating scenarios in this way, it is possible to tailor the interview to the participants by couching them in language familiar to them (that of game design), thus allowing more natural responses. The full interview schedule can been in Appendix B, which includes a description of each scenario.

5.2 Study

In order to recruit participants, adverts were placed on public message boards, specifically the GameDev section of *Reddit*[1] and the Tripwire Interactive online forums[2]. Professional connections of the author, as well as the School of Computer Science were also used for recruitment. One designer was recruited by the online adverts, with the remaining three recruited through professional connections. One participant (P4) was

[1]http://www.reddit.com/r/GameDev
[2]http://forums.tripwireinteractive.com

a lead designer at a major game developer, two participants (P2, P3) were independent game developers and one participant (P1) was an amateur level designer for *Red Orchestra: Ostfront 41-45*, along with *Killing Floor*, another Tripwire Interactive title. All of the designers were familiar with the concept of heatmap visualisations, but had not used them in their own work. None of the designers were familiar with dendrogram visualisation, and the concepts were explained to them before the interviews commenced.

Due to geographical distance, two of the interviews were conducted via Skype, with the remaining two conducted in person. Each interview lasted approximately 20 minutes, and followed a semi-structured approach. Interviews were conducted between October and December 2014. The interview schedule can be seen in Appendix B. Each participant completed two scenarios, one heatmap and one dendrogram. The scenarios were allocated according to Table 5.1.

Participant	H1	H2	D1	D2
P1	X	-	X	-
P2	-	X	-	X
P3	X	-	-	X
P4	-	X	X	-

Table 5.1: Allocation of participants to different scenarios

In order to complete these scenarios, the interviewee was provided with a set of visualisations. For the heatmap scenarios, they were provided with a top-down image of the game level, and a single heatmap showing the movement of 40 players from a single game session. For the dendrogram scenarios, they were again provided with the same top- down image and heatmap of player movement, as well as a dendrogram visualisation of the same data. These materials were provided either on paper or as a PDF, dependent on the interview method.

Before beginning either scenarios, the concept of heatmaps and dendrograms, respectively, were explained to the interviewee. After completing the two allocated scenarios, the interview concluded with a short professional reflection section, discussing the ways they use player telemetry data in their own design process. Following this the participants were thanked and debriefed. The interviews were recorded using hand-held digital audio recorders, with the resultant audio data being transcribed by the author, totalling 12,800 words for all participants. These transcripts were then analysed using thematic analysis.

5.3 Results

An inductive thematic analysis was conducted on the transcribed interview data, with game design process and usage of visualization being the focal aspects. Audio recordings were transcribed, and then split into meaningful sentence fragments. This totalled 74 sentences, which were coded into 10 themes by a single coder using a process of inductive thematic analysis [4]. Another coding iteration was undertaken by the same coder, which consolidated the data into four themes, discussed below.

5.3.1 Value of Metric Feedback

P2, P3 and P4 expressed the value of metrics to understand player behaviour in the game industry. For example, P2 outlined how they gather data about players in their game: *"[When using our metrics] if there are players spending an abnormally long time in a room, maybe the enemy's health is too inflated and needs to be changed...to make sure [the level] has a nice [difficulty] curve."* P4 talked about the testing process: *"They put a lot of analytics into it [the game] and sent the game to people remotely so they can track how long they were playing for and what they were using."*

5.3.2 Issues for Design

The way analytics are used to answer questions on a case-by-case basis during the design process was discussed by participants. P4 illustrated how they combat data overload from quantitative metrics: *"We tracked everything... you've got reams of data, which is counter-intuitive. We tried to figure out what questions we wanted to ask and then put in the right amount [of analytics] to answer those."* P3 echoed this statement, outlining what kind of questions they ask when using metrics: *"I'd like to be able to take out...the guys doing really well, what's he doing different and how can I bring him closer?"* and *"the guy who's really struggling, why is he struggling?"*

5.3.3 Heatmap Utility

The utility of heatmaps for the design process was discussed. P2: *"You could see if two classes are working together they're gonna have a similar heatmap, that's a good* *dynamic"* and P1: *"It would be useful to find out how the map is flowing, how they're moving from one part to another."* and also *"my eye is drawn to the hotspots but the [small] points are still important, you could argue they are more important."* The issue of large data sets of heatmaps was raised by participants. P3: *"If it's 50 [heatmaps] for each level, there's too much data to actually analyse usefully. You want it down into something more succinct"* and P2: *"I would definitely look at it [all], but I probably wouldn't pay as much attention to the smaller details"*.

5.3.4 Dendrogram Utility

None of the designers had encountered a dendrogram before, however, when interpreting the dendrogram representation during the design scenario, points were made regarding its use. P2: *'You can instantly see where the problems are lying then you could look at the heatmaps to find it."* P4: *"this would help show you how similar all the players are to each other, or not. Maybe you would know if you needed to look further."* P3: *"With the more players you get, the more useful something like this is going to be."* P4: *"As long as I can do that [click on items], it's quite easy to click on it and say what is it that's made them similar. I think that could be really useful."* Problems with the representation were also highlighted. P3: *"I'm seeing a lot of noise. I think you'd need to use it a few times before you really understood. There's a lot of data there and it's knowing how it's gone together"*. P4: *"When you first look at it and are uneducated as to what it is, it looks more like it's only the lines that are important and that the white space doesn't represent anything."*

5.4 Discussion

Our results confirm the question-based analytics approach used in the games industry ([30]), as shown by P4: *"We tracked everything... you've got reams of data, which is counter-intuitive. We tried to figure out what questions we wanted to ask and then put in the right amount [of analytics] to answer those."* This suggests that they are using the analytics process to reduce and summarise telemetry data into the most pertinent features, and this is used to directly inform design decisions, in line with practices presented by Mellon et al [32] and DeRosa [6].

The utility of heatmaps as part of the game analytics process was evident in the data: *"It would be useful to find out how the map is flowing, how they're moving from one part to another."* [P1], but this utility is couched against their limitations when using large datasets as they present *"too much data to actually analyse usefully. You want it down into something more succinct"* [P3].

Mellon discusses the use of game analytics processes as a "first pass tool", giving the designer a starting point to launch deeper analysis and exploration. The discussion of dendrograms demonstrates their utility for this function, displaying whether an analyst would need to *"if you needed to look further."* [P3]. Furthermore, P2 notes how the dendrogram could be used to isolate outliers, pointing them to the area of heatmap data for further analysis: *"You can instantly see where the problems are lying then you could look at the heatmaps to find it."* This has the effect of guiding them towards the most pertinent areas of the data for their analysis.

Some participants found the dendrogram difficult to understand initially, however they noted that this could be mitigated once familiar with using the tool and understanding the data: *"I'm seeing a lot of noise. I think you'd need to use it a few times before you really understood. There's a lot of data there and it's knowing how it's gone together"* [P3]. In the design scenarios, dendrograms and heatmaps were only provided as a fixed representation (i.e. on paper for the physical sessions), and one participant detailed having access to the underlying heatmap data was important to understand the dendrogram: *"As long as I can do that [click on items], it's quite easy to click on it and say what is it that's made them similar. I think that could be really useful."* [P4]. The requirement for interactivity with dendrograms is also a concern within computational biology, where Gilbert [22] calls for the tools which help an analyse view the underlying source data observations.

Our results show the designers identifying a problem space around large quantities of spatial data, in the form of heatmaps, and they call for data summarisation in order to take away actionable insights, e.g. The designers interviewed called for meta-visualisation and data summarisation to address the difficulty of analysing large quantities of spatial analytics data: *"the guy who's really struggling, why is he struggling?"* [P3].

5.5 Summary

This chapter set out the design of a feasibility study in order to explore the potential utility of dendrograms as a design tool, conducted with four expert game designers. Using a set of hypothetical design scenarios, the study found that game analytics processes were important for designers, with the need to reduce visual complexity a concern. By considering heatmaps and dendrograms in the design scenarios, heatmaps were described as intuitive and useful at small scale, whereas dendrograms provided the possibility to explore large sets of heatmaps to identify outliers. However, the designers suggested that seeing the base heatmap data was important to understanding the cluster relationships of the dendrogram. The following chapter details a usability study, encapsulating the findings from this chapter, designed to understand how dendrograms can actually be used in the game analytics process.

Chapter 6

Study 2: Usability of Dendrograms in the Design Process

To evaluate the utility of dendrograms within the design process, a usability study was conducted. Informed by a feasibility study of heatmaps and dendrograms as tools for game data analytics, a set of game analytics tasks were created, requiring participants to use the visualisations to interpret a dataset. This chapter describes the design of the usability study, the running of the study, and a comprehensive presentation and discussion of the results.

6.1 Design

In Study 1 (Chapter 5), the results showed there is potential for dendrogram visualisation to be useful, however it was noted by those involved that the ability to interact with the dendrogram to explore the underlying data was required. It was also suggested that dendrogram visualisation might make it easier to analyse and understand group and individual behaviour when working with large sets of heatmaps. It was further found that dendrograms present a potentially useful tool for analysing spatial data in the form of heatmaps, but that participants noted they could not judge the efficacy without actually using the tool. Motivated by this, it was desired to evaluate the dendrogram visualisation technique in a production-like environment, and to do this a usability study using dendrograms was designed and conducted, in order to evaluate their effectiveness at summarising data during the game analytics process, facilitating game designers when answering analytics questions.

Due to time constraints, it was not possible to recruit an active game development team to a usability study where they would use dendrogram visualisation as an analytics tool during the design process. The difficulty of obtaining long-term involvement with the games industry for game analytics research is well documented as an open problem (see [11] and [47]), often due to the inherent time pressures game developers are under, as well as information and process secrecy. As such, it was decided to recruit a group of third year BSc Games Computing students from the University of Lincoln Computer Science department. Whilst the students are not active professional developers, their programme of study mandates they design and implement games throughout their three year study. As third year students they are considered advanced learners, as not only are they training in the discipline, they have completed a number of design tasks during the previous two years of study.

Usability studies provide a useful methodology for evaluating the efficacy of a tool in its natural environment, and a full literature review around usability studies can be found in Chapter 2. By providing participants with data and tools, and setting them a task where they are required to use that set of data and tools, this allows us to measure, both quantitatively and qualitatively, how they use the tools, what problems they have and whether they are able to perform the task correctly in a real situation. In this instance, a game analytics task, utilising the previously collected game data and histogram and dendrogram visualisations, was designed. A roleplay scenario was selected to form the main task for the usability study. This was chosen specifically to leverage the experience and perspective of the Games Computing students. Being in their final year of study, they will be forward-looking towards their future role within the games industry, and are thus receptive to experiences which are similar to that of the games industry. Furthermore, it creates a step change from their normal mode of teaching, as they will be presented with data that is new to them, and they will be required to perform tasks and provide answers that will provide actionable insights for other people.

6.1.1 Task Design

Utilising the heatmap and dendrogram visualisations generated using the corpus of player telemetry collected from *Red Orchestra* (RO) (Chapters 4 and 3 respectively), it was decided to create a roleplay task around Tripwire Interactive and RO. The roleplay for participants was defined as follows:

> You're working for Tripwire Interactive whilst developing *Red Orchestra*, doing game analytics. The Quality Assurance team have just run a batch of user tests on the most recent version of the game, on the level RO-Danzig. They tested nearly 300 people, and gathered spatial data for five in-game actions (shooting, moving, reloading, dying, taking damage). The senior producer has looked at this data briefly and wants you to look at the data and answer some questions they have about the gameplay, level design and player behaviour. You need to evidence your answers using the play test data, not your own experience.

The roleplay would involve each participant receiving a set of five questions about features of the dataset, and being expected to utilise the analytics information they were provided with in order to answer the questions and justify their answers. The data provided to them would be histograms and dendrograms based on gameplay of the RO-Danzig game level, which was collected as part of the first study. Specifically, the data provided consisted of 266 individual game sessions on the RO-Danzig game level, collected over the course of 32 game matches. Each individual game session had been generated into histograms generated from the five key gameplay features: *movement, deaths, damage, shooting* and *reloading*. This data was then provided on aggregate, split by teams and as an overall aggregate. The 266 individual game sessions were used as the basis for the dendrogram visualisation. Five dendrograms were provided, one for each of the five key gameplay features. In summary, the data provided was:

- Histograms of 5 gameplay features for each of the 266 players.
- Histograms of 5 gameplay features, using aggregated data of 266 players
  - Split into Allies, Axis and Overall.

- Dendrograms of 5 gameplay features.
  - Histograms used by dendrograms, divided into cluster group folders.

Based on feedback from the previous study, and in keeping with the roleplay scenario of the study, as set of 6 questions were designed. The full set of questions can be seen in Appendix C, with the answers in Appendix D. The questions were couched in language common to game design, such as "bottlenecks", "combat zones" and "spawn points", with each question being considered in terms of the tools and approaches needed to answer it. Question 1 is an introductory question about the game objectives, which can be answered using play experience or the analytics data provided. Question 2 and 3 ask about combat and death zones, and therefore can easily be answered with the histogram data, or by using dendrograms. Question 4 and 5 specify groups within certain dendrograms and therefore can only be answered by using the dendrograms. Question 6 provided a piece of fuzzy user feedback, and asks for that user behaviour to be identified in the data. It can be answered using histograms or dendrograms.

The answers to each of these questions is grounded within the data itself, therefore for a participant to provide a correct answer, they must have used the tools provided to them to analyse the data. Thus, it is possible to grade their responses based on how "correct" their response is. The complexity of the player behaviour represented in the data is such that participants may provide a partially correct answer to a question. It is also important to anticipate participants may use the tool in a different and/or unexpected way, and to recognise this usage, a category for this additional use was included. The classification scale developed for the written responses is detailed in Table 6.1. This classification scale provided a method for quantifying participants usage of game analytics visualisations to find the correct information required by the analytics questions.

Question No.	Classification	Details
1	Failed	Did not provide answer to question
2	Attempt, Wrong	Not correct or no evidence in data
3	Attempt, Partial	Partially correct
4	Attempt, Right	Correct
5	Attempt, Additional	Different insights than expected

Table 6.1: Answer classification for questions in the game analytics task.

6.1.2 Focus Group Schedule Design

Following the roleplay exercise, it was decided to run focus groups, conducted in a semi-structured manner, in order to capture some of thoughts and feelings of the participants towards the task and the tools and techniques used. A schedule was designed for the focus groups, seen in Appendix E. It was anticipated that two key areas of the roleplay task would provide useful insights, discussion around the task itself and their approach to the game analytics process, and the specific usage of tools and techniques to answer the analytics questions. As such the schedule is divided into an introduction, in order to get participants talks, a set of questions about the task itself, a set of questions about the tools used, and a final question to capture and changes or thoughts about the overall task that may have been omitted. In line with the semi-structured nature of the focus groups, the schedule was written as a set of prompts, with impetus on the researcher to foster discussion and debate amongst participants. It was decided to recruit a maximum of four participants per focus group, in order to preserve debate and discussion.

6.2 Study

Participants were recruited from the University of Lincoln, UK, from the undergraduate BSc Games Computing programme. All students recruited were in their third year of study. The usability study was conducted as part of their programme of study for one module, Game Engines. This module has a focus on physics and analytics within games, and as such the task was suitable for their programme of study. The usability of study was conducted twice, on the 30th September 2015 and the 14th October 2015. As the usability study was integrated within the undergraduate's programme of study, all were allowed to participate in the task, but only those who were happy to consent to their participation in the study had their data collected. The usability study was conducted twice to fit the schedule of the students.

In each iteration, a short introductory lecture was given, lasting approximately 30 minutes, discussing the task for the session, as well as some refresher information about game analytics. All students had been exposed to game analytics through previous lectures and workshops. After this introduction, students were given 30 minutes to play

Focus Group	No. Participants
FG 1	4
FG 2	4
FG 3	2
FG 4	2
FG 5	2
FG 6	4
FG 7	3

Table 6.2: Breakdown of focus group participants.

Red Orchestra, on the map RO-Danzig. They were allowed to play in multiplayer or singleplayer game mode, as long as they were playing RO-Danzig.

After 30 minutes, they were handed their question sheets, with the researcher making note of those who had consented to data collection and participation in the study. Participants were given 60 minutes to complete their task. Along with question sheets, a folder of data visualisations were provided to them, digitally, the breakdown of the contents of this folder are outlined in the previous section. Workshop demonstrators (hourly-paid support staff) were on hand to guide students who had questions about the task, as well as the researcher. However all were under strict instructions not to directly answer any of the questions for the students. At the end of 60 minutes, participants were randomly selected to take part in the 10 minute focus group.

In total, 40 students participated in the study, with all consenting to have their written answers collected on their question sheets. Within this group, there were 3 participants who had to leave early, and thus did not complete beyond question 3. For the selection of focus groups, three researchers were available, and each randomly selected up to four people to participate in their focus group. When a participant indicated they had finished the written task, a researcher asked whether they would be willing to participate in a short focus group to discuss the tools used in the task. Participants were only asked if they had completed the consent form and given consent. Once each researcher had four participants, or there were no more available participants in the workshop, they moved to small breakout areas adjacent to the computer laboratory to conduct their focus group. The breakdown of participant numbers can be seen in Table 6.2. In session one, three focus groups were conducted, and in session four focus groups were conducted, a total of 21 participants. This data was recorded on hand-held audio recording devices, which was then transcribed, the results of which are outlined below.

6.3 Results

The results for this study are separated into two sections, the first covering the results of the written responses provided during the games analytics exercise. The second section covers the results of the thematic analysis performed on the transcribed focus group audio.

6.3.1 Written Responses

In line with the nature of usability studies, the coding scheme designated in Table 6.1 was used to code the responses of the participants. This coding was conducted by two independent expert coders. Each coder used the classification scale to score each answer for each participant independently onto a spreadsheet, noting any information about their decision if applicable. One clear criteria for both coders was to ensure that answers should be evidenced in data. The classifications from each coder were then compared using SPSS to find the inter-rater reliability. The codes were measured for inter-rater reliability, and an initial Fleiss' Kappa was reported as 0.491, which represents moderate agreement using the guidelines presented by Landis and Koch [29]. On inspection, the majority of disagreement was around the use of classification 3 and 4. To address this moderate agreement level, these questions were isolated and both coders discussed their classifications, adjusting where necessary, as well as comparing their decision making notes. The roleplay and subsequent questions were explicitly clear to participants that answers should be evidenced in data, and as such it was mandated that to classify answers as *Partial, Right* or *Additional* they should be clearly evidenced in data. This distinction was a major source of discussion between coders as it was common for one to miss subtle evidence within a response. After this, inter-rater reliability was recalculated based on the adjusted classifications, with a resultant 0.929 Fleiss' Kappa. The breakdown of classifications made by each coder can be seen in Tables 6.3 and 6.4, and an overview of the data provided in Figure 6.1.

Classification	Q1	Q2	Q3	Q4	Q5	Q6
Failed	-	1	3	8	9	4
Wrong	1	2	1	5	10	5
Partial	13	11	7	9	6	12
Right	23	25	29	16	15	18
Additional	3	1	-	2	-	1

TABLE 6.3: Breakdown of written response classifications by Coder 1.

Classification	Q1	Q2	Q3	Q4	Q5	Q6
Failed	-	1	3	8	9	4
Wrong	1	3	1	5	10	6
Partial	14	11	8	9	7	10
Right	22	24	28	16	14	19
Additional	3	1	-	2	-	1

TABLE 6.4: Breakdown of written response classifications by Coder 2.

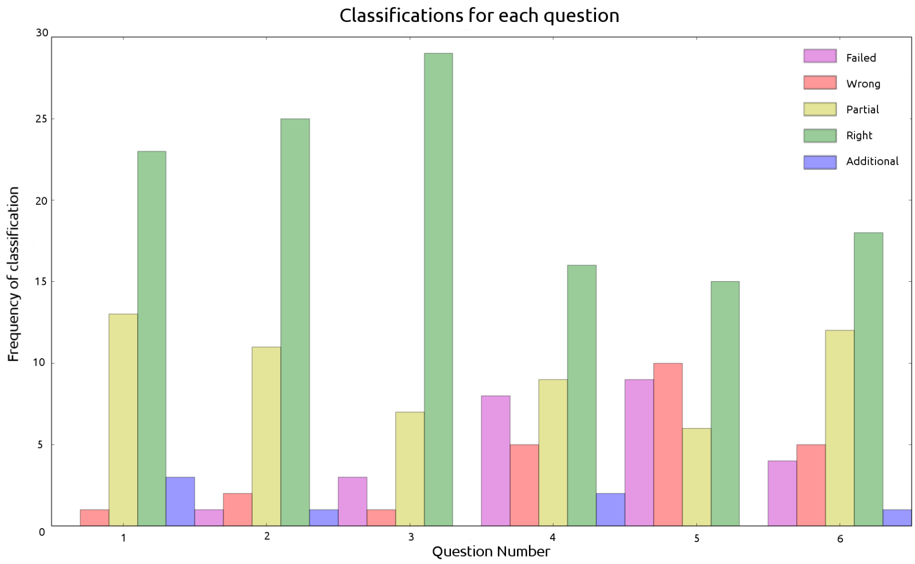

FIGURE 6.1: Bar chart of questions classifications

6.3.2 Focus Groups

An inductive thematic analysis was conducted on the transcribed interview data. The total dataset consisted of 11,684 words produced by seven focus groups and 21 participants. The thematic analysis was conducted by three independent expert coders. The first coder transcribed all the focus group data, in order to become familiar with the data. Once this was done, the first coder developed an initial set of inductively-generated category codes, based on terminology used by participants in the data. These category codes were applied to conversation fragments, and refined iteratively. The resultant codebook of category codes was discussed by all three coders, with duplicate codes being removed or combined into broader codes. To validate the codebook, each coder coded the interview data, independently to one another. With this completed, the coders compared their classifications, discussing and amending as necessary. Thematic analysis software, NVivo, was used to perform the coding, as well as for highlighting disagreement between coders. The initial overall agreement between the three coders was produces a Kappa of 0.39. As this represented fair agreement, as measured by Landis and Koch [29], it was decided to discuss each code that had a lower than average agreement. The three coders stepped through each of these codes, discussing the classifications. A number of these disagreements were due to a coder missing the presence of a code, which in most cases resulted in that coder amending their classifications. This was further compounded by the specific nature of some of the codes in the codebook, meaning there were a relatively small number of instances within the dataset (e.g. less than 5), therefore it was easy to miss their occurrence in the data - this was dealt with in the same way as other disagreements through discussion between the coders and correction of classification where needed. The resultant inter-rater reliability between the three coders produced a Fleiss' Kappa of 0.84.

It should be noted that as not all participants chose to participate in the focus group part of the study, the sample is self-selected. The focus groups were conducted straight after the usability tasks, thus it is possible that participants may have been intrinsically motivated to participate in the focus group to share their opinions (positive or negative) with the research team. Furthermore, participants keen to learn about game analytics or gain some extra insight to advance their career prospects may have been motivated to participate, as the participants are still in training for the games industry. As such, some individual opinions may be over-represented from the initial sample of 40 participants, and when interpreting the results it is important to consider the self-selection bias inherent in the focus group sample.

The next phase of the analysis involved the clustering of the individual codes into thematic categories. This was an iterative process, performed on a large white board, and involved the grouping of common codes through agreement of the coders. These themes were then grouped further into overarching themes. The resultant theme structure can be seen in Figure 6.2. A total of 51 codes were combined into 13 first-order themes, with these being further grouped into 4 second-order themes. A detailed account of

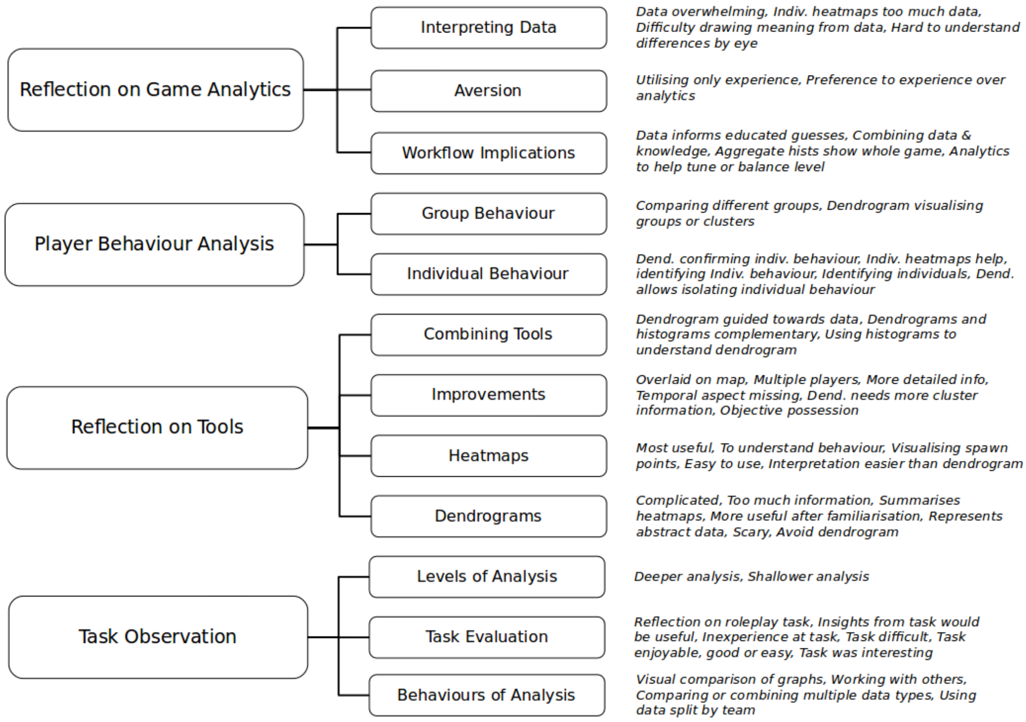

FIGURE 6.2: Structure of themes and sub-themes from focus group data.

the thematic analysis of this data can be found in Appendix F, and a summary of each theme is presented below.

6.3.2.1 Theme 1: Reflection on Usage of Game Analytics

This theme encapsulates participant's reflections on the usage of game analytics, particularly how they used the different tools provided, and how these tools and processes would be used in the games industry. Participants noted that analytics tools such as histograms and dendrograms could help a designer make educated guesses: *"When you get some feedback that is like "I had to run too far", you might have no concept of where the combat's actually taking place... so to have that ability to look at where the fighting is, enables you to make educated guesses about spawn points"* [P1, FG3] and *"I: Do you think you had to guess? P1: Yeah, if you can't ask the person who's playing a question, you're going to have to guess, an educated guess"* [P1, FG7]. One participant noted how they drew information from the game itself, the histogram and dendrogram visualisations to draw their answers: *"I looked at the maps and heatmaps, saw Axis had a more spread out heatmap for movement. I then went in for a practice round and realised I had* *to run really far. Then I spawned as an Allied and realised I didn't have to run far. So I made the assumption that it was probably the Axis team that were complaining"* [P1, FG3].

The ambiguity of the data, in terms of drawing a definitive answer, was cited by two participants as causing difficulty when trying to answer the analytics questions: *"It was really fun to do it, but it was hard... to come to a decision on what it's saying. I: How was it hard? P1: There's loads of different interpretations"* [P1, FG7] and *"I definitely changed my mind half way through... after looking at it for a while and getting to understand it better, I changed my mind"* [P1, FG3].

Participants discussed how the game analytics process would influence the gameplay of *Red Orchestra*: *"That sort of information would be perfect for level designers if they want to put in, bottlenecks or battle zones."* [P1, FG7] and *"It means a lot, if the data shows dramatic results, then you can re-make the map completely to change it to how you expect"* [P2, FG2]. However, a number of participants expressed a preference to using their own design expertise or play experience over using game analytics tools: *"I think I probably could have done if I'd looked for it. I thought the simplest way to do this is to get in and play and see what people might be saying. Obviously I could have looked at the map for spawn points and looked again for distance travelled."* [P1, FG3]

6.3.2.2 Theme 2: Player Behaviour Analysis

The ways that participants approached analysing the behaviour of players is the focus of this theme. The analysis of group behaviour using dendrograms was discussed: *"We were talking about group C splitting so soon, and realised it must be that Allies at their spawn point, because they've got two distinct directions to go in, so that made a distinct grouping* [on the dendrogram]*."* [P1, FG3] and *"I think another tier down* [the dendrogram] *one of the groups splits again... that's probably one group crossing the boat and another crossing the bridge."* [P1, FG3]. Refer to Figure 4.4 on page 52 for dendrogram in question.

The use of histograms and dendrograms for analysing the behaviour of individuals was another sub-theme emergent from the data: *"I found looking at different heatmaps from different players - which ones seemed to line up more with the feedback that was given - the movement and death heatmaps - trying to find a correlation between the two."* [P1, FG7]. Participants also used dendrograms to isolate individual players and inspect their histogram data: *"If you want to go into it in more depth, you can get the individual number of player, and then look at them as an individual"* [P2, FG1] and *"We were looking at a heatmap of a player who died a lot, and we were wondering, why are they dying so much? So we compared all their heatmaps for that player, who we identified on the dendrogram within this group of people we were looking at. I don't think we found anything really conclusive but it was still really useful to be able to look at all that data together"* [P2, FG7].

6.3.2.3 Theme 3: Reflection on Tools

This theme comprises participant's discussion of the tools used during the usability study, particularly in terms of their utility, drawbacks and potential improvements.

Using both histograms and dendrograms together to analyse the data was mentioned. One participant outlined that exploration of the base data (heatmaps) allowed them to comprehend and build an understanding of the relationships being represented in the dendrogram visualisation: *"When you put the two together though, you've got the dendrogram and the heatmaps, and you look at the two of them, you can figure out what the dendrogram is saying, but only when you've got the two together"* [P3, FG2] - see poster in Appendix A on page 90 for illustration of this process. Participants also noted that dendrograms presented a summarisation of the base data: *"I think having both produces better results. You can look at the heatmaps and it's hard to put it all together in your mind, but then the dendrograms kinda summarise it. So using both of them is good, but if you had just one, I think you'd get less of a result"* [P2, FG6] and *"Definitely for finding a group of people who do similar things. I think combining that* [the dendrogram] *with individual heatmaps makes it quite a lot more powerful a tool than it seems originally"* [P1, FG7].

Participants reflected on the histograms used in the usability study, that they were easy to interpret, and useful for exploring player behaviour: *"They were really easy to interpret, especially if you were displaying this data to somebody who maybe doesn't understand the game or map."* [P1, FG5], *"I found them really helpful just to be able to visually see it all and compare it straight to the map"* [P1, FG3] and *"Made life really* *easy to be able to see what kind of zones [of the map] were really intensified and where people were travelling to"* [P1, FG3].

Participants also reflected on dendrograms, with some participants describing the dendrogram as presenting too much information: *"It was useful to look at them, because they have the groupings all in one go, but because the actually data, individually, was quite difficult to read because there was so much of it"* [P1, FG5] and *"It was daunting at first. It was a lot of information, 300 players on each one - there's a lot of information to take in"* [P1, FG7]. To address this, participants cited the inspection of the underlying data helped to understand the dendrogram: *"I think if you had just the dendrogram you wouldn't stand much chance of understanding anything, because it doesn't really give anything away from itself, you need the heatmaps to back it up"* [P4, FG6] and *"When I was looking at two* [heatmaps] *that were very similar, according to the dendrogram, it didn't make any sense until I looked at the heatmaps. They gave it meaning."* [P1, FG7].

It is of note that participants expressed that dendrograms would require familiarisation before they could be used effectively: *"I think if we got better... if we were more well-versed in dendrograms... after you got the concept in your head of what it was actually representing, it was easier to find the data. Looking at the different groups, and the different branches. Once you got those and look at the data and figured it out, it was a bit easier to use."* [P2, FG7] and *"If I understood it a bit better I probably would have found it a bit more useful, but when I first started looking at it, it is a bit daunting if you've never seen one before to try and analyse it, but after awhile of looking at it, I kind of understood it."* [P1, FG3]. The poster shown in Appendix A on page 90 illustrates the process of examining the underlying heatmaps.

One participant stated they did not utilise the dendrogram visualisation itself, but used the clustering of histogram data created by hierarchical clustering to explore the data: *"I didn't really use the dendrogram that much anyway. I just looked through all the histograms in the folders, and compared them like that"* [P2, FG5].

A number of future improvements were suggested for the dendrogram visualisations, most notably more detailed information about player states: *"Sprint data would have been helpful. It was the task where we had to identify which group of players were having issues with running for ages. That would have been useful to see who's having to sprint a lot."* [P2, FG7] and *"The skill of the players perhaps. I couldn't really tell if they were* *running randomly around the map or actually heading to a point."* [P2, FG5]. Further to this, participants suggested multi-level dendrograms, or dendrograms with fewer players, in order to aid exploration of the base data: *"I think maybe some smaller dendrograms would have been [more useful]. It would have been nice to have the three separate groups* [as standalone dendrograms] *and the larger one. You could look at them a bit more clearly, because there was so much data in the large dendrogram "* [P1, FG7] and *"That might have been enough reason to split into another group, because then it would have been easier to find differences. It was hard to read, but if the second sub-group in cluster C were coloured differently, it would have been easier for me to read it quickly. You could also split the groups into more groups, because then you could colour it and it would be easier to read"* [P2, FG5].

6.3.2.4 Theme 4: Task Observation

This theme encapsulates the participants discussing the usability study tasks (game analytics scenarios) and how they went about finding information and answering the questions.

Participants noted there were some tools more suited for quick answers and others suited for more complex answers. Quick, shallow analysis could be achieved using histograms: *"If you're using the histogram, it's a lot easier to get a quick look"* [P4, FG1], *"What we would probably pick up was basic stuff, but there were probably underlying things that are way more advanced than we would have thought of."* [P4, FG6]. When looking to perform deeper analysis, dendrograms and histograms were used in conjunction: *"We were looking at the heatmap of a player who died a lot, and we were wondering why they were dying so much. So we compared all of their heatmaps by identifying them through the dendrogram, as they were in a group of players we wanted to look at."* [P2, FG7] and *"The dendrogram was useful because you can see what group was doing what, but also if you want to go into more depth, you can get the individual number of the player and then look at them as an individual, and how they moved on the map."* [P2, FG1].

Whilst participants enjoyed the game analytics tasks, they noted their lack of experience with game analytics: *"I think the exercise itself made us get into the role. Even though it was "role-play", it made you want to do it and to think about it. All of a sudden you're thinking 'If I was building this map, I would move this bit or do this here'"* [P4, FG6] and *"We probably missed things because we haven't got experience of it, but I feel like we understood what was going on"* [P1, FG4].

Participants outlined different working techniques and approaches when analysing the data, such as working with others to understand the data: *"Talking to people around me and having another look at it. We were talking about group C splitting so soon, and that was because they go in two different directions. You can kind of understand it having sat and spoken to people for awhile."* [P1, FG3] and *"I worked with other to deduce how to use it, but it's kind of self explanatory, in the ways that things are matching together"* [P1, FG5]. The combination and comparison of different data types and visualisations was noted by some participants: *"The shooting was quite useful, because you could see where the combat zones were. You could see they'd be moving through the map, come across an enemy, and see shooting but no deaths at that location"* [P1, FG5] and *"I looked at the maps - heatmaps - to see the Axis team had a more spread out heatmap for movement, whereas the Allied team was more focused."* [P1, FG3].

6.4 Discussion

The aim of this study was to test the usability of heatmaps and dendrograms as tools for use in the game design process, specifically when answering analytics questions. The usability study was conducted with a group of game development students, who were tasked with answering a set of game analytics questions using data visualisations provided to them. Their responses were recorded, and a sub-set of the participants were recruited to focus groups to specifically discuss their use of tools and the usability study. The results presented above provide some interesting insights into the efficacy of heatmaps and dendrograms in the design process.

A quick overview of the performance of participants in the usability study is provided by the quantitative classification of the written responses. Using Table 6.3 and 6.4 and Figure 6.1 there is a clear difference between questions 1 through 3, and questions 4 through 6. In questions 1 through 3, the majority of participants were able to answer the questions successfully. Whilst question 1 was a "warm up" question, 2 and 3 were heatmap-oriented questions about combat zones and game level architecture. These questions relied on the correct usage of heatmaps in order to find this information, and the fact that the majority of participants were able to answer them correctly reinforces previous work around histograms and heatmaps being a quick and easy tool for visualising aggregate-level player behaviour data, as discussed in Chapter 2, section 2.1.1.

On the other hand, it is of note that for questions 4, 5 and 6 participants did not fair so well, with a wider range of answers. There is still a small majority of participants who answer the questions correctly, for questions 4 and 5 there are almost double the number of *failed* and *wrong* classifications than for questions 1-3. This can be attributed to the required use of dendrograms to answer the questions. The classification code *failed*, denoted no attempt being made to answer the question. In order to unpick the subtleties and reasons behind, the focus group data provides details of the participants thoughts and reactions which can be drawn upon.

6.4.1 Dendrogram Familiarity

Throughout the data, one strongly emergent theme was the expression that dendrograms were daunting or scary to look at initially: *"It was daunting at first. It was a lot of information, 300 players on each one - there's a lot of information to take in"* [P1, FG7]. This initial aversion gave way to a reflection on the complexity of the tool and its potential after a period of familiarisation. Participants noted that they became used to reading and understanding what the dendrogram was represented as they progressed through the task: *"If I understood it a bit better I probably would have found it a bit more useful, but when I first started looking at it, it is a bit daunting if you've never seen one before to try and analyse it, but after awhile of looking at it, I kind of understood it."* [P1, FG3]. As noted earlier in this chapter, none of the participants had ever used a dendrogram. The discussion around familiarisation and usability of tools is a topic covered extensively, for example Sennett [40] describes the process of familiarisation with tools: *"Getting better at using tools comes to us, in part, when the tools challenge us, and this challenge often occurs just because the tools are not fit-for-purpose."*

However this is opposed by existing works within HCI that have strived to reduce the creation of digital tools that are "not fit-for-purpose" by adopting methods, such as usability testing and user-centred design, with the intention of understanding the processes involved when using tools, and how a proposed design might, or does, affect that process. The value-sensitive design agenda, as presented by [20] is embodied through the designer-centric work of Kim et al [27], Medler et al [30] and Mirza-Babaei et al [33]. Motivated by this, the feasibility study (Chapter 5) presented the dendrogram visualisation approach to professional game designers in order to gather feedback and thoughts about the efficacy as a game analytics tools. The usability study (Chapter 6) acted on the feasibility study feedback by providing designers with the opportunity to use the dendrogram visualisation along with the underlying heatmap data, to complete tasks. This showed us that, although the dendrogram visualisation seemed complex and daunting initially, as with other expert tools, designers were able to use them to provide player behaviour insights that would not have been possible with heatmaps alone.

Medler et al [30] note in their guidelines for tool design with game developers, "analytic prejudice" can be encountered, and this was encapsulated in the first order theme *"Aversion to Game Analytics"*, such as: *"I think I probably could have done if I'd looked for it. I thought the simplest way to do this is to get in and play and see what people might be saying. Obviously I could have looked at the map for spawn points and looked again for distance travelled."* [P1, FG3]. This aversion to game analytics may be explained by the participants lack of game analytics experience, a topic openly discussed by one participant: *"We probably missed things because we haven't got experience of it, but I feel like we understood what was going on"* [P1, FG4]. Whilst the task was primarily designed as an individual exercise, pair working was discussed as a means to comprehend the tools and the data being represented, notably in the *Behaviours of analysis* sub-theme: *"Talking to people around me and having another look at it. We were talking about group C splitting so soon, and that was because they go in two different directions. You can kind of understand it having sat and spoken to people for awhile."* [P1, FG3].

6.4.2 Levels of Analysis

Another emergent theme within the focus group data was the positioning of heatmaps and dendrograms as suitable for different "levels" of analysis. Participants stated that heatmaps could, and were, used to quickly answer simple analytics questions: *"If you're using the histogram, it's a lot easier to get a quick look"* [P4, FG1]. This confirms the well established understanding of heatmap usage in the games industry, notably for the ease with which fast answers can be obtained about aggregate behaviour (see discussion in Chapter 2). Participants stated that dendrograms were better suited to deeper, more time-intensive analysis: *"The dendrogram was useful because you can see what group was doing what, but also if you want to go into more depth, you can get the individual number of the player and then look at them as an individual, and how they moved on the map."* [P2, FG1]. These usability results have shown that dendrograms are used predominantly for deeper analysis of the patterns and features of the data, which confirms the sentiments of the game designers interviewed in Study 1 (Chapter 5), who noted that dendrograms would be suitable for summarising large datasets, and allowing the isolation of outliers and interesting patterns.

6.4.3 Cluster Analysis

It is clear that participants used both heatmaps and dendrograms to analyse and identify group behaviour. A number of participants discussed the behaviour of players in *group C* cluster, which was the focus of question 5 (see Appendix C), with participants asked to specifically typify what the characteristics of this group were. *"We were talking about group C splitting so soon, and realised it must be that Allies at their spawn point, because they've got two distinct directions to go in, so that made a distinct grouping* [on the dendrogram]*."* [P1, FG3]. In referring to *"splitting so soon"* they are referring to the large branch distance on the dendrogram between the first two sub-groups within group C, which can be seen in Figure 4.4 on page 52. This also demonstrates a correct answer to question 5 (see Appendix D), as they have identified the team correctly, and that the clusters are predominantly split due to different routing through the game level.

This sub-analysis of clusters could be facilitated in future tools designed around the heatmap and dendrogram concept. In the tools used in the usability study, the threshold for clusters in the dendrogram (described in Chapter 4) was set at a level to provide a small number of clusters, each with a large number of observations, for the purposes of testing the usability of dendrograms in the games context. However, it is possible to change this threshold, with a higher threshold reducing the number of colour-coded clusters on the dendrogram, and a lower threshold increasing the number of colour-coded clusters. When generating the underlying heatmaps for the visualisation process, the number of bins per heatmap needs to be defined. In this instance the bin number has been experimentally defined for the purpose of the usability study, however it is possible that a designer may wish to change this bin number in order to change the way the data clusters. As such, both the ability to tune the dendrogram threshold, and the underlying heatmap bin number could be exposed as functions within future tools. This would provide the design with the ability to explore different levels of clusters on the dendrogram, as well as the way the clusters change when heatmap bin size is manipulated.

It was noted that participants used the dendrogram to isolate outlier patterns of behaviour that did not fit with the rest of the behaviour, when exploring the data *"We were looking at a heatmap of a player who died a lot, and we were wondering, why are they dying so much? So we compared all their heatmaps for that player, who we identified on the dendrogram within this group of people we were looking at. I don't think we found anything really conclusive but it was still really useful to be able to look at all that data together"* [P2, FG7]. Whilst the participants do not conclude anything meaningful from their analysis, they cite the ability to look at the behaviour within the context of the whole data set as an important part of their analysis. This theme is further discussed by participants, who note the abstract nature of the dendrogram visualisation: *"I think if you had just the dendrogram you wouldn't stand much chance of understanding anything, because it doesn't really give anything away from itself, you need the heatmaps to back it up"* [P4, FG6]

As a result, being able to explore and understand the heatmap data that is represented in the dendrograms, they noted it gave the hierarchical clustering relationships of the dendrogram some concrete meaning: *"When you put the two together though, you've got the dendrogram and the heatmaps, and you look at the two of them, you can figure out what the dendrogram is saying, but only when you've got the two together"* [P3, FG2]. This corroborates the feedback from the professional game designers interviewed in the feasibility study, who noted that although the dendrogram visualisation approach had potential, without being able to explore the underlying heatmap data it was difficult to judge, for example: *"As long as I can do that [click on items], it's quite easy to click on it and say what is it that's made them similar. I think that could be really useful."* [Study 1, P4]. This aligns with the findings of Kim et al [27], who found that the ability to "drill down" to the underlying data allowed the analyst to understand what a higher-level report was displaying. This can be applied to dendrograms, which are representing abstract hierarchical clustering relationships, and by exploring the underlying, concrete, and easily interpretable data, these abstract cluster relationships can become concretely founded in the data. This is an advantage over some of the behaviour analysis work discussed in Chapter 2, which has complex or abstract base information.

6.5 Summary

This chapter presents the design of a usability study with dendrograms, in order to evaluate their utility as a game design tool. The usability study used design tasks which involved answering questions grounded within the data, and as such participants were classified on the correctness of their answer. A set of 40 final year game development students were recruited. The results show that dendrograms were found to be initially hard to understand, but half of all participants were able to successfully use the dendrogram to correctly identify data features. Through the subsequent focus groups, participants reinforced the use of heatmaps and dendrograms together, the former helping them to understand the cluster relationships represented in the dendrogram. The following chapter draws together the conclusions and contributions, and proposed future directions for research.

Chapter 7

Conclusions & Future Work

This chapter brings together the findings and implications of the work presented, tying these back to the aim and objectives outlined in Chapter 1. Following this, the limitations of the work are presented, with speculation towards some directions of future research using dendrogram visualisation in the games context.

7.1 Addressing the Aims & Objectives

The aim of this work was to present an analysis technique for clustering large sets of individual player heatmaps into a form that is intelligible by game designers. Based on a thorough review of existing literature, this focused the research question towards the ability of dendrogram visualisation being used to display and summarise the cluster relationships between sets of individual heatmaps, and whether they could highlight the interesting patterns and features of the dataset to a game designer in a way a designer can understand.

To answer this question, a set of dendrograms and heatmaps was generated from a corpus of telemetry data, and used these in a feasibility study with a small group of expert game designers. By asking them how each of the visualisations might be used to answer game analytics questions, it was learned that dendrograms show promise as a visualisation technique for large sets of data, and that the ability to "drill down" into the original heatmap observations was imperative to understanding the cluster relationships being visualised. Based on this feedback, a usability study was designed and conducted with a large group of game development students, where they were required to use both heatmaps and dendrograms to answer a set of pre-designed analytics questions. By gathering both quantitative and qualitative data about the usability of both visualisations, the results showed that dendrograms initially presented a complex visualisation to the game designers, but with time they were able to understand the cluster relationships of the heatmap dataset, and were thus able to use the dendrogram to identify and describe player behaviour in different clusters, as well as isolating outlier individuals and clusters of individuals.

In Chapter 1, five objectives were set out in order to answer the research question. The following section discusses how this work addressed each of these aims in the below paragraphs:

The first objective was to explore spatio-temporal data analysis, specifically around visualisation. In the Literature Review (Chapter 2) different methods and systems employed by industry were presented, noting the popularity of heatmap visualisation in the games industry. Academic applications of clustering and behaviour analysis were also presented, with the application of the knowledge presented in this chapter utilised in the design of the telemetry framework in Chapter 3.

To understand the existing works and methods for UCD of game design tools, a thorough literature review of existing research in this area was conducted, presented in Chapter 2, documenting the growing body of research documenting the design of game design tools for practitioners, and their applicability and usability within the game development process.

In order to collect a preparatory data set of player telemetry for use in heatmap and dendrogram visualisation, an FPS game, *Red Orchestra*, was selected, which allowed modification to the source code, whereupon a telemetry recording framework was designed and implemented into the game source code. This allowed a group of players to be recruited, who generated a corpus of telemetry data with more than 1 million data points, using the telemetry framework, which was then processed ready for visualisation.

Following this, a hierarchical clustering visualisation system was implemented in the *Python* programming language, which initially produced individual heatmaps of player's actions from the corpus of telemetry data collected in the preparatory work. These sets of individual heatmaps were then hierarchically clustered, and the subsequent clustering information visualised using a dendrogram.

To collect qualitative and quantitative data from game designers, a feasibility study with game design experts was conducted. The results from this feasibility study were used to design a follow up study, a usability study, which used a set of game analytics tasks involving the use of heatmaps and dendrograms. Quantitative data was collected about the accuracy of participants using the visualisation techniques to answer the tasks, as well as a large set of qualitative data gathered from focus groups with participants after the tasks. This qualitative data provided insights into the participants views towards the different visualisation techniques.

In meeting these objectives and answering the research question, this work presents a number of contributions. It proposes a solution to the unsuitability of heatmap visualisation as a technique for comparing and identifying individual spatial behaviours. It does this by presenting the dendrograms as a means to explore and understand the relationships between clusters of individual player heatmaps, preserving the ability for a designer to “drill down” to the base observations that are in an easy to understand format (heatmaps). In effect this allows game designers to reduce the visual complexity presented by hundreds of heatmaps, and focus on heatmaps which display interesting or different behaviour, a use that was identified during usability testing. This is a novel use for heatmaps, a popular and well understood visualisation technique in games, extending the applicability of heatmaps to interpreting individual behaviour at large scale.

This work contributes a novel application of the dendrogram visualisation technique to the field of games, showing the utility of dendrograms as a visualisation technique for comprehension of hierarchical clustering results by game designers. Whilst research has been conducted towards clustering of player behaviour, this is predominantly based on non-spatial telemetry features, and in some instances requires a data analysis expert to interpret the results into a form intelligible by a game designer. This work uses dendrograms, a visualisation well suited for data analysis, for displaying cluster relationships spatial heatmap data, which allows designers to interpret the pertinent and interesting features and behaviours within the dataset.

The two studies presented in this thesis contribute to research around game design tools using a user-centred design methodology. The feasibility study (Chapter 5) provides insights into the way designers use a question-based approach to exploring data, and notably the value of heatmaps as a tool for small-scale, aggregate level data analysis, and the opportunities for dendrograms as a tool for clustering heatmaps. The contribution of this aspect of the work is further reinforced by the publication of the feasibility study at the *CHI Play 2015* conference [19].

Informed by this, the usability study (Chapter 6), presented a rich dataset of heatmap and dendrogram usage in a realistic game design scenario. This confirmed the value of dendrograms for isolating individual or group behaviours of interest by game designers, guiding them towards the underlying data. Whilst they presented an initially difficult tool to use, the designers saw the value of them for exploring the behaviours present in a large set of data.

The design and implementation of a telemetry recording framework into the FPS game *Red Orchestra* is a contribution to game analytics research, as it outlines the considerations and technical work required to implement telemetry data recording into a game engine. This contribution is published at the *GAMEON 2012* conference [18], where the insights presented can be used to inform future work that intends to design and implement telemetry frameworks in games.

7.2 Limitations

It is important to acknowledge the limitations of this work. As outlined in Chapter 1, this work is limited to spatial telemetry data, and within that utilises only six basic features of the gameplay of *Red Orchestra*. It is possible to draw more detailed and complex features from the gameplay features, as described in Chapter 3, for example the locations of deaths by specific weapons. Based on this, the behavioural profiles generated by the dendrogram visualisation approach are not as detailed as those generated using high dimensionality player telemetry data, such as those proposed by [45]. However, this limitation is intentional, in order to reduce the complexity of the data presented to participants in the usability study, as the focus of Study 1 and 2 were towards the reduction of large, single-variate sets of individual heatmaps, and whether users could make use of such visualisations. Presenting a wider variety of data features than the 6 provided would increase the complexity of analysis dramatically.

Moreover, the work is limited because it does not evaluate dendrograms in the design process in an “in the wild” setting, where real game developers are using it to inform design decisions being made during the development of a real game. Obtaining access to game designers during the development process is difficult, and often involves industry cooperation with academic research, and this problem is an acknowledged issue in game analytics research [46], often due to confidentiality, and differing time-scales. As the work of Medler et al [30] shows, when implementing and evaluating tools in the live game development process, it is difficult to obtain quantitative data, as feedback and utility is often more subjective and varied between team members, yielding large sets of qualitative data.

As regards the usability study of dendrograms, there is a limitation to the work due to the type of participants recruited. Based on the previously mentioned difficulty cooperating with game developers, the sample of game development students presents a limitation on the usability study. Primarily this is because students, whilst having received training and learned skills, are not fully qualified and/or skilled in their profession (game development), and thus their experiences and feedback during the usability study can be considered more error prone and the participants less skilled than that of expert game designers, meaning the findings will be different. Whilst this limits the generalisability of the findings of the usability study, the findings are encouraging because game development students were able to use both heatmaps and dendrograms to answer a set of expert-level game analytics questions, and were able to interpret and understand the cluster relationships, and subsequently profile the behaviour of players represented in the dendrograms.

7.3 Future Work

Based on the limitations of this work, a number of routes for future work become available. The utilisation of a more complex usability study could be employed, such as the iterative usability testing method used by Mirza-Babaei et al [33]. This method sees the decisions made from inspecting and analysing analytics data actually implemented into the game in question, where the player experience of the modified game is then evaluated by new participants. This method, if used with dendrograms as one of the testing conditions, would be able to extend the work presented in this thesis by measuring the efficacy of dendrogram visualisation on the player experience.

The proposed system of dendrogram visualisation of individual heatmaps could be evaluated in a development, or "in the wild" setting, where this system is used iteratively to analyse the behaviour of players during the game development cycle. Due to the low technical implementation barrier of creating heatmaps and dendrograms, this could be achieved by attaining cooperation from a small or medium sized game developer, if cooperation with a large game developer is not possible.

There is scope to apply the dendrogram visualisation approach to games of different genres than the FPS genre game presented here. Every digital game features a virtual environment through which players navigate, whether that is in the form of navigation through a 2D/3D environment with a player character, or navigating around a game interface. This spatial data can be, and is currently in the games industry, processed into heatmaps, and implementing the dendrogram visualisation approach would be a simple step. Furthermore, any spatial information containing large numbers of individual users could be visualised using dendrograms to understand spatial behaviour. For example, dendrograms could be used to visualise how individuals cluster into different routes or patterns of behaviour when walking around a country park.

As outlined in the limitations section previously, only a small set of basic gameplay features were used in study 1 and 2. A future direction of the work could be towards exploring how deeper and more complex game data features could be clustered and visualised using dendrograms, and whether this would be useful for designers.

Appendix A

CHI Play Poster

Poster presented at CHI Play 2015 for work-in-progress publication, full citation as follows:

> Tom Feltwell, Grzegorz Cielniak, Patrick Dickinson, Ben J. Kirman and Shaun Lawson. Dendrogram Visualization as a Game Design Tool. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI Play 2015).* Pages 505-510, ACM, 2015. [19].

See poster overleaf

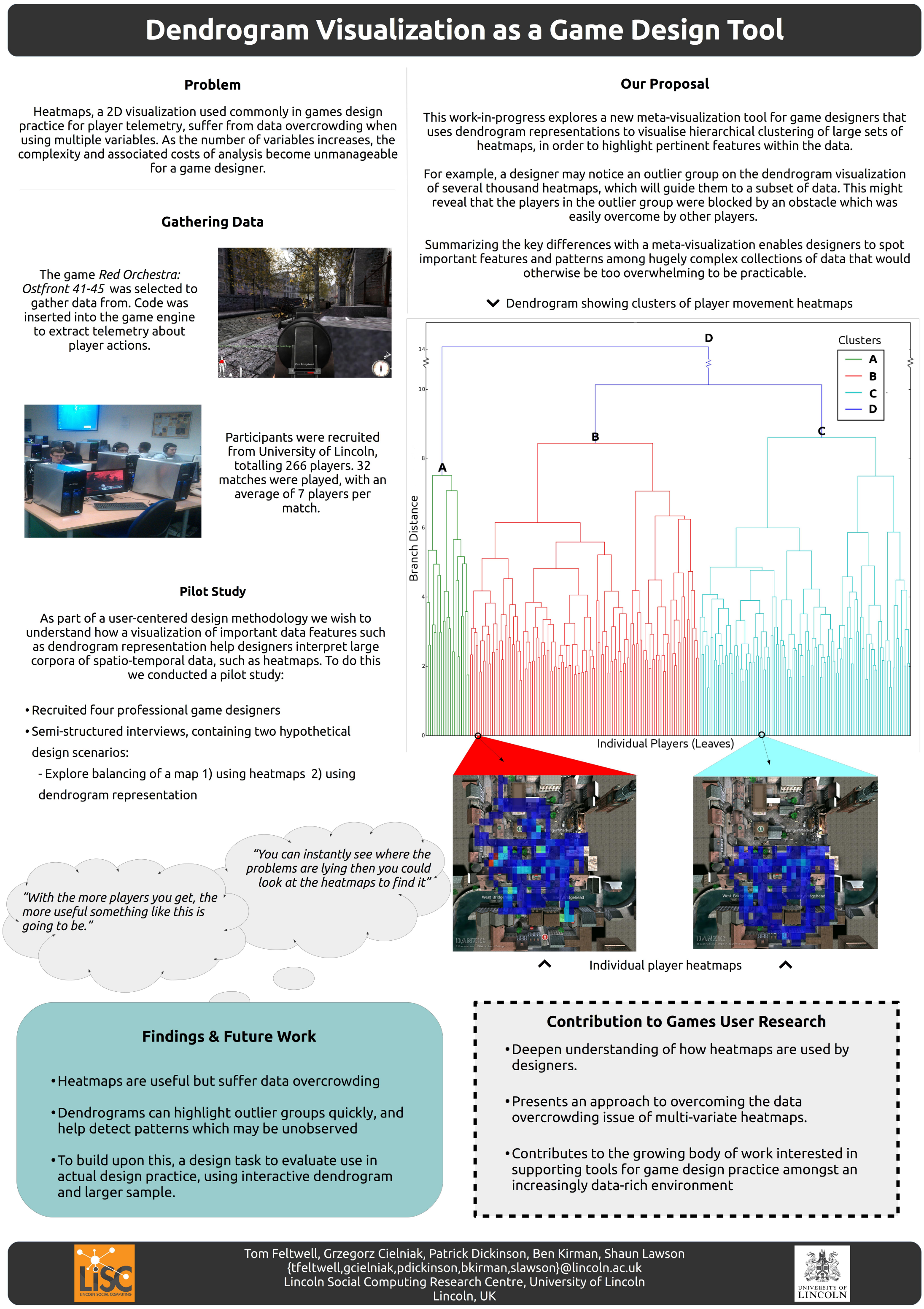

FIGURE A.1: Poster presented at CHI Play 2015 for work-in-progress publication titled *"Dendrogram Visualization as a Game Design Tool."*

Appendix B

Interview Schedule

B.1 Intro

[Discuss consent form, any questions?]

[Introduce researcher] I'm a research assistant at the University of Lincoln, and today would like to talk about semi-automated game design tools.

To clarify, semi-automated game design tools are tools which complement and aide a game designer. Their aim is not to replace a designer, but merely to provide useful information, and do some of the leg work to make a designer's job easier. This work forms part of my Masters by Research degree, and will go towards a research publication.

With this in mind, the first half of the interview will talk about your design experience, and the second part I will pose some hypothetical design scenarios along with some visualisations to see how you might solve them.

Importantly, don't tell me what you think is the right answer. I'd like to hear how you genuinely feel and what your real opinions are. Any questions?

B.1.1 Questions

- Briefly, tell me about your game/level design experience? [Prompts] How long have you been developing? / How did you get into it? / Which platforms have you worked on? / Have you collaborated with others?

- What is your feedback mechanism usually? (Tools, player feedback) How would the player feedback work? (Interview, view play test)

B.2 Tool Usage Scenarios

For this section, I'd like you to use your design experience to think how you might solve a hypothetical design scenario.

Heatmap is a graph/visualisation which shows the frequency of events in cells across the level. In this case, the brighter, or warmer, the colour, the more events happened there. Coloured like heat, blue is cold with red/white as the hottest.

B.2.1 Heatmap

Scenario 1: You are designing a FPS level for a multiplayer game in an urban environment, as part of a small team. When observing play one your colleagues notice there is some combination of player class and weaponry which is far too powerful and is making one team win more often. You are provided with a couple of heatmaps of movement data. [Refer example 1] How would you go about trying to understand and solve this problem using the heatmap?

Scenario 2: You are designing a FPS level for a multiplayer game in an urban environment, as part of a small team. Play testers report that on this level one team wins more often. You are provided with a couple of heatmaps of movement data. [Refer example 1]How would you go about trying to understand and solve this problem using the heatmap?

Question prompts: Is there data that is missing? Does this data allow you to solve the problem? If not, what would you need to solve the problem? What scenario would this be useful for? Would this approach fit into your design process?

B.2.2 Dendrogram

Let's say we have 80 heatmaps, each one shows the movement of single player for a single match. (Two games with 20 players per side). That's too much data to look at.

This visualisation is a summary of how similar the patterns of each players movement is to the others. Each of the nodes (at the bottom of the graph), represent one heatmap, and the lines note how similar they are. The further up the tree (closer to the root) they are, the less similar they are. The tree is nicely broken up into clusters, showing distinct groups that are similar. Currently on paper, but in practice you would be able to look each node or cluster to see the heatmap.

Scenario 1: You are designing a FPS level for a multiplayer game in an urban environment, as part of a small team. Play testers report that on this level one team wins more often. You are provided with a dendrogram of movement data. [Refer dendrogram example] How would you go about trying to understand and solve this problem using the dendrogram?

Scenario 2: You are designing a FPS level for a multiplayer game in an urban environment, as part of a small team. When observing play one your colleagues notice there is some combination of player class and weaponry which is far too powerful and is making one team win more often. You are provided with a dendrogram of movement data. [Refer dendrogram example] How would you go about trying to understand and solve this problem using the dendrogram?

Question prompts: Is there data that is missing? Does this data allow you to solve the problem? If not, what would you need to solve the problem? What scenario would this be useful for? Would this approach fit into your design process?

B.3 Professional Reflection

(3 to 4 questions max)

- How much do you balance feedback/data against gut instinct?
- If a player was telling you something,
- Would you always trust your design instinct over player feedback?
- Are there times when you make decisions without enough data?
- Are there times when you make decisions that are divergent from what the data is telling you? (gut instinct, past experience)

- How do you regard metrics? As a way of making money, improving playability?
- (if game designer) Is data used in all aspects of design or only for level design?
- What kind of questions do you ask when trying to solve a design problem? (i.e. Why would players do this? How did they get there?)

Appendix C

Usability Study Analytics Questions

Download the data: [hyperlink]

You're working for Tripwire Interactive whilst developing Red Orchestra, doing game analytics. The QA team have just run a batch of user tests on the most recent version of the game, on the level RO-Danzig. They tested nearly 300 people, and gathered spatial data for five in-game actions (shooting, moving, reloading, dying, taking damage). The senior producer has looked at this data briefly and wants you to look at the data and answer some questions they have about the gameplay, level design and player behaviour. You need to evidence your answers using the play test data, not your own experience.

1) Can you outline which are the most used objectives? Why? Are there any that are interesting to you?

2) What are the death zones and bottlenecks like on the level? What is the balance of teams like in those spots?

3) How many main combat zones are there? What route are players taking to get into the main combat zones?

4) In the movement-32 dendrogram, can you give me some detail as to what players are in the three groups? Also, what are the characteristics of the cluster A and why are they different?

5) Again on movement-32, why does cluster C split into two groups? What are the differences in behaviour of both sub groups?

6) We received a lot of feedback as follows: The spawn points are too far away from the combat zones, and I kept getting killed so spent ages running. Which team were these players on and which spawn points were they complaining about?

FINISH

Appendix D

Usability Study Question Answers

1. This is visible from the heatmaps of all players movement and deaths. Most used are Command Centre and West Bridgehead, followed by East Bridgehead. East has a large spike of deaths but not so much movement (= more contestation perhaps) Langer Market is least used.

2. This is visible from the heatmaps of all players movement and deaths. Death zones: a) Small concentrations of high-death areas in Command centre, waterfront around East Bridgehead. Building on the RHS of East Bridgehead, north side. b) Very large spread of deaths around West Bridgehead, although intensity not as high Bottlenecks: East & West Bridges, especially West. Waterfront north of bridges. Alleyway/crossroads directly north of West Bridgehead on way to command centre. Balance: Axis favour due to map forcing Allies to cross bridges.

3. Visible from heatmaps, looking at movement, death, shooting and damage locations. Both bridgeheads combined are most likely the biggest combat zone, followed by Command Centre. The alleyway north of West Bridgehead on the way to Command Centre could also be considered such, as well as the river line.

4. Must use dendrogram for this. Cluster B is the Axis players. Cluster C is the Allied players. B and C move around the map in a varied fashion, as each cluster almost completely encompasses the Allied/Axis teams. Cluster A can be on both teams (predominantly Allied) but tend to have less movement overall, and more focused (i.e. camping, idling, moving straight to objective etc). Clusters B and C are NOT only new players.

5. Again must use dendrogram for this question. The two groups within cluster C are both Allies. The group on the left (PID 184 236) have less spread out movement, and more camping, or hanging around in particular areas, e.g. running around inside command centre, bridges, etc. Left group and focused to LHS of map, right group are focused to centre and RHS of map. The group on the right (PID 127 124) move generally more spread out, with multiple paths being followed, e.g. flanking, running about between points.

6. It's possible to answer this question in multiple ways. The answer broadly is the north spawn for the Axis, which is only used after the bridgeheads have been secured. It's a long distance from this spawn to the Command Centre. Using the dendrogram of movement, they might find a specific player, such as 190 (although I did not pick one in particular) who is on the Axis team. Their heatmap shows they run often from spawn to command centre, and looking at their individual death heatmap it shows they often die there too.

This can also be seen using a couple of the aggregated heatmaps. Looking at deaths and shooting, it is clear there is very little action north-west of the map, near the Axis secondary spawn. This would indicate they are only running there, not fighting.

It is possible to reach this conclusion from just playing the game, it becomes abundantly clear the Axis have to run a long way however solely using experience to answer the question is wrong, as it is not clear they've used the data (combat zones etc)

Appendix E

Usability Study Focus Group Schedule

Researcher's note: We are interested in the way that these guys used the tools (histograms, dendrograms) provided to them, so try to keep the conversation focused on them. How did they use them to solve the problems, that sort of thing.

Discussion needs to be quite loose, we want the participants to talk, so discussion / debate between them is good. Do they all agree to certain points etc.

The below document is a rough set of prompts to talk through in the focus group:

OPENING:¡Read this first¿ Thanks for agreeing to take part in this study. In this focus group we'll talk about the way in which you approached the game analytics tasks set for you in the workshop, as well as how you used the tools provided.

TASK: First, we'll start off talking about the tasks themselves:

1. What did you think of the game analytics exercise? How difficult was it to find the information you needed?

4. Prompt: How well do you think you performed as a game data analyst? As a developer how do you think the insights you provided would be useful?

TOOLS:

2. What tool(s) did you find most useful? - Why?

3. What made the dendrogram useful in this task? - Why?

4. How did you approach understanding the dendrogram?

5. What data was missing from the data set that you needed to answer the exercise?

6. Think back to question 6, where you had to find spawn points based on fuzzy user-feedback. Where did you find the information to answer that question?

ENDING:

7. If you had opportunity to change or develop any of the analytics tools used today, what would you do?

Appendix F

Usability Study Thematic Analysis Results

Theme 1: Reflection on usage of Game Analytics

This second-order theme encapsulates the participant's discussions around the various processes of game analytics, and the impact that histograms and dendrograms have on that process.

Sub-theme: Interpreting data: This theme encapsulates discussions by the participants about different methods of understanding the data presented to them in the usability study, and specifically the difficulties they had when trying to complete the analytics exercise. Participants described how the dendrogram provided a lot of information on initial viewing: *"Because it* [the histogram] *was two colours, if it was closer to blue it was less happened, closer to red more happened, but with the dendrogram, with that much information...especially the first time seeing one, it was overwhelming."* [P1, FG7]. Another participant reflected on being provided with a histogram of every player: *"There was a lot of data we didn't really need. The individual player data, we didn't really need to look at..."* [P3, FG1]

Some participants said they found some ambiguity within the data and that they had difficulty drawing meaning from the data: *"It was really fun to do it, but it was hard... to come to a decision on what it's saying. I: How was it hard? P1: There's loads of different interpretations"* [P1, FG7], *"I definitely changed my mind half way through...*

after looking at it for a while and getting to understand it better, I changed my mind" [P1, FG3] and *"I couldn't really tell if group B or C were just randomly running around the map, or actually heading to a point, or thinking, I'm getting killed here so maybe I should go around a different way"* [P2, FG5]

It was also reported that it was difficult to carry out some of the analysis by eye, without software assistance: *"it was really clustered, so you couldn't see where the main bridge point was. So you didn't know what was someone actually killing someone, or what was someone jumping into the river"* [P2, FG2] and *"I couldn't get one of the answers... it's difficult to tell the difference between group B and C because the histogram doesn't give you much information...they looked pretty similar, they were relatively spread around the map but concentrated in a few areas, I couldn't really see much difference beyond that. Well they weren't different really"* [P2, FG5]

Sub-theme: Aversion to game analytics: This sub-theme represents participants describing an aversion to the games analytics process, in terms of using only their pre-existing design experience, or rejecting game analytics insights in favour of their pre-existing design experience. It was evident that some participants acknowledge the existence, and potential utility of the game analytics data presented, but chose to ignore it: *"I think I probably could have done if I'd looked for it. I thought the simplest way to do this is to get in and play and see what people might be saying. Obviously I could have looked at the map for spawn points and looked again for distance travelled."* [P1, FG3] and *"I just looked at the actual map itself, and the spawn points, where they are. It's not like, ok movement data would help, but if you just look at the layout, it tells you."* [P1, FG6]. Furthermore, some used only their game playing experience, or their pre-existing design knowledge to come to conclusions to the questions asked: *"It was like 'one team has found it difficult to get into battle' and obviously that's because battle was taking place on one side of the map. You can take that* [fuzzy] *feedback and switch the spawn points or spawn one team closer"* [P2, FG1] and *"All the Axis had to do was navigate a town and as soon as you find the quickest path, you can be there ready to kill them. Then you can run away and do something else."* [P1 FG6].

Sub-theme: Workflow implication of game analytics: When reflecting on the implications of dendrogram and heatmap tools on the overall design process, it was suggested that analytics data helps you, as a designer, make educated guesses: *"When you get* *some feedback that is like "I had to run too far", you might have no concept of where the combat's actually taking place... so to have that ability to look at where the fighting is, enables you to make educated guesses about spawn points"* [P1, FG3], *"I looked at the individual teams and looked at where most of the players were actually dying, and assuming that if they were normally dying there, they were running down from this spawn point"* [P3, FG7] and *"I: Do you think you had to guess? P1: Yeah, if you can't ask the person who's playing a question, you're going to have to guess, an educated guess"* [P1, FG7].

It was also observed that participants detailed ways to combine game analytics insights, with knowledge either from play experience or existing design knowledge: *"I looked at the maps and heatmaps, saw Axis had a more spread out heatmap for movement. I then went in for a practice round and realised I had to run really far. Then I spawned as an Allied and realised I didn't have to run far. So I made the assumption that it was probably the Axis team that were complaining"* [P1, FG3], *''We identified the main combat zones using the heatmaps, and combined the knowledge we had from that with the overlay map. I think that made it fairly obvious who was having an issue"* [P2, FG7] and *"Allied deaths were concentrated below or on the bridge. Whereas the Axis team can go through the map and get to a bunch of sniper spots and cover quicker than the Allies."* [P2, FG5]

<u>Sub-theme: Gameplay implications of game analytics process:</u> This theme contains sentiment from the participants that the game analytics process performed in this usability study would have implications to gameplay, of *Red Orchestra* or in general. When talking about histograms, some participants said aggregate histograms are useful to visualise the whole game state: *"I used a lot of the ones that had the Axis and Allies on, just because I felt like that showed the game as a whole"* [P1, FG5] and *"Heatmaps are more overall - this is what's been happening in these areas, and then you can go and look more specifically at certain people"* [P3, FG7]. It could also be seen that the game analytics process would help tune or balance a level or gameplay: *"That sort of information would be perfect for level designers if they want to put in, bottlenecks or battle zones."* [P1, FG7], *"It means a lot, if the data shows dramatic results, then you can re-make the map completely to change it to how you expect"* [P2, FG2] and *"I: Do you think these analytics are useful for addressing fuzzy feedback? P2: I think so. You know exactly what's up with it, and why people don't like it. Then maybe where you should put*

something" [P2, FG3].

Theme 2: Player Behaviour Analysis

Broadly, this theme covers participants discussing the way tools were used to specifically analyse player behaviour, either as groups, or individually.

<u>Sub-theme: Group Behaviour:</u> An emergent theme was the discussion of identifying groups and performing behaviour analysis on them, particularly using the dendrogram: *"I think another tier down* [the dendrogram] *one of the groups splits again... that's probably one group crossing the boat and another crossing the bridge."* [P1, FG3], *"In the dendrogram you could identify a couple of different groups within the Allies, or a group of players who had a different strategy and preferred to camp or stay in one particular area of the map"* [P2, FG7] and *"We were talking about group C splitting so soon, and realised it must be that Allies at their spawn point, because they've got two distinct directions to go in, so that made a distinct grouping* [on the dendrogram]*."* [P1, FG3]. However this group analysis was not restricted to dendrograms: *"I used death* [histograms] *, from both sides, and compared the distance between one spawn point and the main cluster, and the main cluster to the other spawn point"* [P1, FG2].

<u>Sub-theme: Individual Behaviour:</u> The use of the game analytics processes to identify and analyse individual behaviours was a strong emergent theme among the data. The utility of heatmaps for identifying individuals was explained: *"I thought the heatmaps and histograms were useful, because they're split up and you can view each player separately if you need to."* [P2, FG5] and *"I found looking at different heatmaps from different players - which ones seemed to line up more with the feedback that was given - the movement and death heatmaps - trying to find a correlation between the two."* [P1, FG7]. Moreover, one participant noted individual histograms helped to identify the in-game roles of players through mapping their behaviour: *"I think having separate* [individual heatmaps] *helped, it allowed you to see who was actually being a sniper and who wasn't. Who - individual players, which ones were doing that"* [P1, FG6].

Participants discussed the ways that dendrograms could be used to confirm their assumptions about player behaviour: *"If you want to go into it in more depth, you can get the individual number of player, and then look at them as an individual"* [P2, FG1], *"I* *picked out this branch and this branch, and I could see that they were attacking different bridges... which is how I looked at it to understand it. I focused on individual players."* [P3, FG1] and *"We were looking at a heatmap of a player who died a lot, and we were wondering, why are they dying so much? So we compared all their heatmaps for that player, who we identified on the dendrogram within this group of people we were looking at. I don't think we found anything really conclusive but it was still really useful to be able to look at all that data together"* [P2, FG7]. This discussion was further extended to include using the dendrogram spot and isolate individual behaviours from clues on the dendrogram: *"A dendrogram would be following the movement of everybody, and if you saw something very erratic happen, you'd be like 'Oh that doesn't look right' and you could then find out what a certain person was doing"* [P3, FG7].

Theme 3: Reflection on tools

Reflection on the tools used in the usability study comprised the contents of this theme, with participants talking about improvements, drawbacks and specific ways they used the tools.

<u>Sub-theme: Combining tools:</u> When reflecting on how the tools and techniques could be combined, participants expressed that dendrograms guided them towards the underlying data: *"It helped me at least, find the information that made sense. It didn't feel like it quantified much for me, it doesn't explain much for me, but it seems like a very good tool for aiding me to find the information I want"* [P1, FG7], *"The dendrogram was useful because you can see what group was doing what, but also if you want more depth, you get the individual number of the player and then look at them as an individual and how they moved"* [P2, FG1] and ' *'If you looked at the dendrogram and looked at the playerID, then got their histograms up, you could look at what the players in that group got up to, made it easier to compare the different play styles, and identify the little things about them"* [P2, FG4]. The complementary nature of histograms and dendrograms was described: *"When you put the two together though, you've got the dendrogram and the heatmaps, and you look at the two of them, you can figure out what the dendrogram is saying, but only when you've got the two together"* [P3, FG2], *"I think having both produces better results. You can look at the heatmaps and it's hard to put it all together in your mind, but then the dendrograms kinda summarise it. So using both of them is good, but if you*

had just one, I think you'd get less of a result" [P2, FG6] and *"Definitely for finding a group of people who do similar things. I think combining that* [the dendrogram] *with individual heatmaps makes it quite a lot more powerful a tool than it seems originally"* [P1, FG7].

One participant also outlined how they used the histograms as a way of understanding wat the dendrogram was representing: *"Looking at the heatmaps helped. You look at the heatmaps and you're like OK.. especially with group A, you look at the heatmaps and you figure out they're going to this place, and staying there, so maybe they're snipers or something. Then you look at the dendrogram and you can see that there isn't much difference, and because you've made the connection you can see, you can make the rest of the connections in your mind, it's a bit easier"* [P2, FG6].

<u>Sub-theme: Improvements to tools:</u> Discussion of improvements that could be made to the tools used in the task comprises this theme. As the participants were presented with histograms (i.e. with no underlaid game map), a strong emergent theme was the desire for histograms to be overlaid onto a map, to form a heatmap: *"The histograms would be better if they were put on top of the map image, overlaid on top"* [P4, FG1] and *"That would be helpful because I was flicking in between, and that's quite easy, would be much nicer if you could just have a see through heatmap that you could flick and change"* [P1, FG3].

One participant suggested histograms could feature multiple players: *"You could probably have 5 players on one heatmap, just in different colours. There's so many heatmaps to look through, but if you just click on one then there's 5 players with different colours, that would help me"* [P2, FG6], whilst others suggested incorporating the temporal aspect of the data into the histogram *"In general, a time slider would be useful"* [P3, FG6] and *"I would turn all the heatmaps in GIFs, if possible, so you could see it over time, or how all of them move over time"* [P4, FG6]. Further to this, it was suggested that adding information about which team possessed the objectives over time would be useful: *"It would have helped to have, over time, so you could see when this objective was belonging to the Allies"* [P4, FG6] and *"You could see which objectives were mostly gone for, but you didn't actually see who got it"* [P1, FG6]

There were calls by some participants to provide more detailed information in order to conduct the game analytics tasks: *"Sprint data would have been helpful. It was the task* *where we had to identify which group of players were having issues with running for ages. That would have been useful to see who's having to sprint a lot."* [P2, FG7] and *"The skill of the players perhaps. I couldn't really tell if they were running randomly around the map or actually heading to a point."* [P2, FG5]. This was further discussed around dendrograms, specifically calling for more detailed, and different ways of representing the cluster information: *"I think maybe some smaller dendrograms would have been [more useful]. It would have been nice to have the three separate groups* [as standalone dendrograms] *and the larger one. You could look at them a bit more clearly, because there was so much data in the large dendrogram "* [P1, FG7] and *"That might have been enough reason to split into another group, because then it would have been easier to find differences. It was hard to read, but if the second sub-group in cluster C were coloured differently, it would have been easier for me to read it quickly. You could also split the groups into more groups, because then you could colour it and it would be easier to read"* [P2, FG5]

Sub-theme: Reflection on histograms: This sub-theme comprises participants reflecting on histograms and how they used them during the game analytics exercise. Many participants found they were easy to use and helped in a variety of ways: *"I found them really helpful just to be able to visually see it all and compare it straight to the map"* [P1, FG3], *"They were really easy to interpret, especially if you were displaying this data to somebody who maybe doesn't understand the game or map."* [P1, FG5] and *"I used a lot of the histograms that had Axis and Allies on, because I felt like that showed the game as a whole"* [P1, FG5]. Specifically, participants spoke about using histograms to understand the behaviour of players in the data: *"I think the spawn points change once you gain the objectives, so I was using the heatmaps for movement to see what distance they actually had to move from their spawn point to where the combat was happening."* [P3, FG7], *"Definitely like the movement, because you can see where they're moving from and they're obviously going to be moving towards the objectives"* [P4, FG1] and *"Made life really easy to be able to see what kind of zones [of the map] were really intensified and where people were travelling to"* [P1, FG3]. Also, using them to visualise and coordinate where spawn points on the game level were located was discussed: *"At the beginning of the match the defenders would be running a really long way from right at the back to the bridges, whereas by the time the end of the match had come about the attackers were running all the way to the centre"* [P2, FG3] and *"Depending on where* *the intensity of the play is [on the heatmap], you could try and keep the spawn points a bit away from that, so someone doesn't just spawn and die straight away"* [P2, FG3].

Some participants also compared histograms and dendrogram together, in terms of their interpretability: *"You can look at the intensity around each area. You can look at it and compare it to the map"* [P2, FG3], *"The histograms were really easy to interpret and look at side-by-side, but the dendrograms were kind of confusing"* [P2, FG4] and *"The heatmaps... because the dendrograms were just awkward to read, they're very intricate and it took us until the end to* [understand the clustering]*"* [P1, FG6].

Sub-theme: Reflection on dendrograms: A strongly emergent sub-theme within the data was participants reflecting on the use of dendrograms, and on the form of the tool overall. Initially, many found the dendrograms complicated or difficult to understand: *"When I first started looking at it, it is a bit daunting if you've never seen one before to try and analyse it, but after awhile of looking at it, I kind of understood it."* [P1, FG3], *"I mean, the histogram was pretty easy to look at, but the dendrogram, I'll be seeing them in my dreams tonight"* [P2, FG1] and *"It's a bunch of lines and a player ID, what am I meant to draw from it? OK, they're grouped, why are they grouped? It doesn't say."* [P2, FG5]. This sentiment was drawn on further, as participants expressed that dendrograms represent too much information: *"I don't mind data analysis personally, so I don't mind doing stuff like that. But the dendrograms, just no, too much information"* [P2, FG1] and *"It was useful to look at them, because they have the groupings all in one go, but because the actually data, individually, was quite difficult to read because there was so much of it"* [P1, FG5].

However, it was expressed by some participants that dendrograms would require a degree of familiarisation: *"If I understood it a bit better I probably would have found it a bit more useful, but when I first started looking at it, it is a bit daunting if you've never seen one before to try and analyse it, but after awhile of looking at it, I kind of understood it."* [P1, FG3] and *"I think if we got better... if we were more well-versed in dendrograms... after you got the concept in your head of what it was actually representing, it was easier to find the data. Looking at the different groups, and the different branches. Once you got those and look at the data and figured it out, it was a bit easier to use."* [P2, FG7]. Some participants reflected on the causes for the difficultly understanding dendrograms, isolating the abstract nature of the data being represented: *"When I was looking at two* [heatmaps] *that were very similar, according to the dendrogram, it didn't make any sense until I looked at the heatmaps. They gave it meaning."* [P1, FG7], *"It didn't feel like it quantified much for me, it doesn't explain much for me, but it seems like a very good tool for aiding me to find the information I wanted."* [P1, FG7] and *"I think if you had just the dendrogram you wouldn't stand much chance of understanding anything, because it doesn't really give anything away from itself, you need the heatmaps to back it up"* [P4, FG6].

Furthermore, participants discussed the way that dendrograms summarise histograms: *"If you just had the dendrogram you wouldn't stand much of a chance understanding anything, because it doesn't really give anything away in itself, you need the heatmaps to back it up"* [P4, FG6] and *"I think having both produces better results. You can look at the heatmaps and it's hard to put it all together in your mind, but then the dendrograms kinda summarise it. So using both of them is good, but if you had just one, I think you'd get less of a result"* [P2, FG6].

Within the data, it was emergent that some participants were confused or daunted by using dendrograms, with some actively avoiding using dendrograms. Particularly, the number of lines and information present on a dendrogram was isolated as a main concern: *"I think the graph itself was very confusing, with branches, because it's from very minute differences, I think it can sometimes overcomplicate the graph"* [P3, FG1], *"The dendrograms were kind of scary to begin with, like 'Whoa loads of lines!' It's harder to see"* [P1, FG6] and *"It was daunting at first. It was a lot of information, 300 players on each one - there's a lot of information to take in"* [P1, FG7].

It was also clear that some participants actively avoided using the dendrogram: *"I didn't really use the dendrogram that much anyway. I just looked through all the histograms in the folders, and compared them like that"* [P2, FG5] and *"After awhile of looking at it, I kind of understood it. I think at first I avoided it because heatmaps are so easy to look at."* [P1, FG3]

Theme 4: Task observation

This theme encapsulates the participants' reflection on the usability study task itself, such as how difficult they thought it was, and which tools they used for which parts of the task.

Sub-theme: Levels of analysis: A clear theme within the data was the reference to different levels of analysis, particularly identified were deep, long term analysis, and short term quick analysis. Reflecting on deep analysis, participants noted: *"I don't think it was difficult, it was common sense to understand what was going on on the maps. But there were some things that had hidden meanings that you had to look at see"* [P1, FG4], *"We were looking at the heatmap of a player who died a lot, and we were wondering why they were dying so much. So we compared all of their heatmaps by identifying them through the dendrogram, as they were in a group of players we wanted to look at."* [P2, FG7] and *"The dendrogram was useful because you can see what group was doing what, but also if you want to go into more depth, you can get the individual number of the player and then look at them as an individual, and how they moved on the map."* [P2, FG1]. This was counterposed by a number of participants who mentioned tools and approaches that could be used for quick, shallow analysis: *"If you're using the histogram, it's a lot easier to get a quick look"* [P4, FG1], *"What we would probably pick up was basic stuff, but there were probably underlying things that are way more advanced than we would have thought of."* [P4, FG6] and *"By finding the basic stuff you are addressing the initial problem, then you can go into the nitty gritty detail."* [P1, FG6].

Sub-theme: Task Evaluation: As part of the focus group schedule, participants were asked to reflect explicitly on the game analytics task they were given. This theme encapsulates their reflections on the task, as well as the tools used. In regards to the roleplay aspect of the task, participants were positive, but noted their lack of experience: *"I think the exercise itself made us get into the role. Even though it was "role-play", it made you want to do it and to think about it. All of a sudden you're thinking 'If I was building this map, I would move this bit or do this here'"* [P4, FG6], *"To begin with really difficult, because I've never looked at histograms or dendrograms or anything like that, so actually quite difficult"* [P1, FG6] and *"We probably missed things because we haven't got experience of it, but I feel like we understood what was going on"* [P1, FG4]. One participant noted that the written aspect of the task was difficult: *"It's much easier to point and explain. Using the data visually is a lot nicer than trying to transcribe data, because it's visual"* [P1, FG7].

Reflecting on how the insights they generated in the analytics exercise would be used in a professional game development setting, showed participants felt they found useful information, but lacked some experience: *"The insights would be useful, because my biggest point was about how little the LangerMarkt was used. I suppose the data that I produced could maybe be used to redesign the map"* [P1, FG5], *"We only had an hour to do it, I think eventually you could start picking it apart. With the information we gave, there's enough for someone to say 'They noticed this, we need to look at that'"* [P1, FG6] and *"We understood how things were, and what was going on. We're not trained to do these things so, just kind of attempting it"* [P2, FG4]

The participants also expressed that they enjoyed the task and found it interesting: *"Apart from questions 4 and 5, I think for the rest I got on straight anyway with it. They were quite enjoyable, I don't mind data analysis personally so I don't mind doing stuff like that"* [P2, FG1], *"It was really fun to do it, but it was hard to come to a decision"* [P1, FG7] and *"It was quite fun and an interesting thing looking at the player movement and stuff like that"* [P1, FG3]

<u>Sub-theme: Behaviours of analysis:</u> This sub-theme constitutes different behaviours used by participants during the game analytics task. Comparison between multiple on-screen graphs in order to conduct analysis was common, particularly when looking at histograms: *"I just do it with my eyes, because in Windows, in Explorer you can actually view all the histograms as tiles, so you can look and see there's differences between groups"* [P4, FG2], *"I found the heatmaps really helpful to be able to visually see it all and compare it straight to the map. Made life really easy to be able to see what kind of zones were intensified, and where people were travelling to"* [P1, FG3] and also with dendrograms: *"I used the histograms before it was made into a dendrogram, and I looked through all that data, because it was separated by folders"* [P2, FG5]. It was also noted that participants worked with others in order to understand and make sense of the game analytics task: *"Talking to people around me and having another look at it. We were talking about group C splitting so soon, and that was because they go in two different directions. You can kind of understand it having sat and spoken to people for awhile."* [P1, FG3] and *"I worked with other to deduce how to use it, but it's kind of self explanatory, in the ways that things are matching together"* [P1, FG5]

Combining and/or comparing multiple types of data and visualisations was also another approach used by participants: *"I looked at the histograms, because I looked specifically* *between the two teams - movement and shooter. You could see that at the Axis spawn point, there was a lot of moving, but no shooting, until they go further into the map, which I found useful."* [P1, FG1], *"The shooting was quite useful, because you could see where the combat zones were. You could see they'd be moving through the map, come across an enemy, and see shooting but no deaths at that location"* [P1, FG5] and *"If you looked at the dendrogram and looked at the playerID, then got their histograms up, you could look at what the players in that group got up to, made it easier to compare the different play styles, and identify the little things about them"* [P2, FG4]. Furthermore, splitting the data by team to aid analysis was another method used by participants: *"I used the histogram of movement, you could see from the Axis one they had every different route, whereas the Allies had to go straight across the bridge. You could see all the fighting was happening just over the bridge, which was quite far from the spawn points"* [P3, FG1], *"I looked at the individual teams and where most of the players were actually dying. Assuming that if they were normally dying there they were running down from this spawn point"* [P3, FG7] and *"I looked at the maps - heatmaps - to see the Axis team had a more spread out heatmap for movement, whereas the Allied team was more focused."* [P1, FG3].

References

[1] C. Abras, D. Maloney-Krichmar, and J. Preece. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4):445–456, 2004.

[2] M. Ambinder. Valves approach to playtesting: The application of empiricism. In *Game Developers Conference*, 2009.

[3] T. E. S. Association. The transformation of the video game industry. *The Entertainment Software Association*, 2010. URL `http://www.theesa.com/gamesindailylife/transformation.asp`.

[4] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[5] J. P. Davis, K. Steury, and R. Pagulayan. A survey method for assessing perceptions of a game: The consumer playtest in game design. *Game Studies*, 5(1), 2005.

[6] P. DeRosa. Tracking player feedback to improve game design. *Gamasutra, Aug*, 7, 2007.

[7] Y. Dodge. *The concise encyclopedia of statistics*. Springer Science & Business Media, 2008.

[8] A. Drachen and A. Canossa. Analyzing spatial user behavior in computer games using geographic information systems. *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era on - MindTrek '09*, page 182, 2009. doi: 10.1145/1621841.1621875. URL `http://dl.acm.org/citation.cfm?id=1621841.1621875`.

[9] A. Drachen and A. Canossa. Towards gameplay analysis via gameplay metrics. *Proceedings of the 13th International MindTrek Conference: Everyday Life in the*

Ubiquitous Era on - MindTrek '09, page 202, 2009. doi: 10.1145/1621841.1621878. URL http://portal.acm.org/citation.cfm?doid=1621841.1621878.

[10] A. Drachen and A. Canossa. Evaluating Motion: Spatial User Behavior in Virtual Environments. *International Journal of Arts and Technology*, 4(3): 294–314, 2011. URL http://www.scopus.com/inward/record.url?eid=2-s2.0-84858974835&partnerID=40&md5=ec34de0f16a3d271ecc26429c2f7813f.

[11] A. Drachen and M. Schubert. Spatial game analytics and visualization. *Computation Intelligence in Games (CIG)*, pages 1–8, 2013.

[12] A. Drachen, A. Canossa, and G. N. Yannakakis. Player modeling using self-organization in Tomb Raider: Underworld. *2009 IEEE Symposium on Computational Intelligence and Games*, pages 1–8, Sept. 2009. doi: 10.1109/CIG.2009.5286500. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5286500.

[13] A. Drachen, R. Sifa, C. Bauckhage, and C. Thurau. Guns, swords and data: Clustering of player behavior in computer games in the wild. *2012 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 163–170, Sept. 2012. doi: 10.1109/CIG.2012.6374152. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6374152.

[14] A. Drachen, P. Mirza-babaei, and L. E. Nacke. Flagship Areas in Games User Research. *Games User Research Workshop - CHI 2014*, 15, 2014.

[15] J. S. Dumas and J. Redish. *A practical guide to usability testing*. Intellect Books, 1999.

[16] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, 95(25):14863–14868, 1998.

[17] M. S. El-Nasr, A. Drachen, and A. Canossa. *Game analytics: Maximizing the value of player data*. Springer Science & Business Media, 2013.

[18] T. Feltwell, P. Dickinson, G. Cielniak, et al. A framework for quantitative analysis of user-generated spatial data. *GAMEON 2012 Conference*, 2012.

[19] T. Feltwell, G. Cielniak, P. Dickinson, B. J. Kirman, and S. Lawson. Dendrogram visualization as a game design tool. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, pages 505–510. ACM, 2015.

[20] B. Friedman. Value-sensitive design. *Interactions*, 3:16–23, 1996. ISSN 10725520. doi: 10.1145/242485.242493.

[21] A. R. Gagné, M. S. El-Nasr, and C. D. Shaw. A deeper look at the use of telemetry for analysis of player behavior in rts games. In *Entertainment Computing–ICEC 2011*, pages 247–257. Springer, 2011.

[22] D. R. Gilbert, M. Schroeder, and J. van Helden. Interactive visualization and exploration of relationships between biological objects. *Trends in biotechnology*, 18 (12):487–494, 2000.

[23] P.-O. Gourinchas, H. Rey, and K. Truempler. The financial crisis and the geography of wealth transfers. *Journal of International Economics*, 88(2):266–283, 2012.

[24] D. M. Hilbert and D. F. Redmiles. Extracting usability information from user interface events. *ACM Computing Surveys (CSUR)*, 32(4):384–421, 2000.

[25] N. Hoobler, G. Humphreys, and M. Agrawala. Visualizing competitive behaviors in multi-user virtual environments. In *Proceedings of the conference on Visualization'04*, pages 163–170. IEEE Computer Society, 2004.

[26] V. Inc. Half life 2: Episode 2 stats. `http://www.steampowered.com/status/ep2/ep2_stats.php`, 2008. Accessed: 2016-04-14.

[27] J. H. Kim, D. V. Gunn, E. Schuh, B. Phillips, R. J. Pagulayan, and D. Wixon. Tracking real-time user experience (true): a comprehensive instrumentation solution for complex systems. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 443–452. ACM, 2008.

[28] S. Kriglstein, G. Wallner, and M. Pohl. A user study of different gameplay visualizations. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 361–370, New York, New York, USA, 2014. ACM Press. ISBN 9781450324731. doi: 10.1145/2556288.2557317. URL `http://dl.acm.org/citation.cfm?doid=2556288.2557317`.

[29] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[30] B. Medler, M. John, and J. Lane. Data cracker: developing a visual game analytic tool for analyzing online gameplay. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2365–2374. ACM, 2011.

[31] M. C. Medlock, D. Wixon, M. Terrano, R. Romero, and B. Fulton. Using the rite method to improve products: A definition and a case study. *Usability Professionals Association*, 51, 2002.

[32] L. Mellon and D. Kazemi. Wake up and smell the metrics!: Metrics-driven development in online games, 2008.

[33] P. Mirza-Babaei, S. Long, E. Foley, and G. McAllister. Understanding the contribution of biometrics to games user research. In *Proc. DiGRA*, pages 329–347. Citeseer, 2011.

[34] D. Moura, M. S. El-nasr, and C. D. Shaw. Visualizing and Understanding Players Behavior in Video Games : Discovering Patterns and Supporting Aggregation and Comparison. In *Proceedings of ACM SIGGRAPH 2011 Game Papers*, pages 11–16. ACM, 2011. ISBN 9781450307758.

[35] L. E. Nacke, A. Drachen, K. Kuikkaniemi, J. Niesenhaus, H. Korhonen, W. Hoogen, K. Poels, W. IJsselsteijn, and Y. Kort. Playability and Player Experience Research. *DiGRA 2009*, 2009.

[36] J. Nielsen and K. Pernice. *Eyetracking web usability.* New Riders, 2010.

[37] D. A. Norman. *The design of everyday things: Revised and expanded edition.* Basic books, 2013.

[38] R. J. Pagulayan, K. Keeker, D. Wixon, R. L. Romero, and T. Fuller. User-centered design in games. In J. A. Jacko and A. Sears, editors, *The Human-computer Interaction Handbook*, pages 883–906. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2003. ISBN 0-8058-3838-4.

[39] R. Romero. Successful instrumentation: Tracking attitudes and behaviors to improve games. In *Game Developers Conference*, 2008.

[40] R. Sennett. *The Craftsman*. Yale University Press, 2008.

[41] H. Shimazaki and S. Shinomoto. A method for selecting the bin size of a time histogram. *Neural computation*, 19(6):1503–1527, 2007.

[42] F. Southey, G. Xiao, R. C. Holte, M. Trommelen, and J. Buchanan. Semi-Automated Gameplay Analysis by Machine Learning. *Artificial Intelligence and Interactive Digital Entertainment*, pages 123–128, 2005. URL `http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Semi-Automated+Gameplay+Analysis+by+Machine+Learning#0`.

[43] M. Steinbach, G. Karypis, V. Kumar, et al. A comparison of document clustering techniques. In *KDD workshop on text mining*, volume 400, pages 525–526. Boston, 2000.

[44] T. Sweeney and M. Hendriks. Unrealscript language reference. *Epic MegaGames, Inc*, 1998.

[45] A. Tychsen. Crafting User Experience via Game Metrics Analysis. *Computer*, pages 1–5, 2008. URL `http://www.cs.uta.fi/~ux-emotion/submissions/Tychsen.pdf`.

[46] G. Wallner and S. Kriglstein. A spatiotemporal visualization approach for the analysis of gameplay data. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems CHI 12*, pages 1115–1124, 2012. doi: 10.1145/2207676.2208558. URL `http://dl.acm.org/citation.cfm?doid=2207676.2208558`.

[47] G. Wallner and S. Kriglstein. Visualization-based analysis of gameplay data A review of literature. *Entertainment Computing*, 4(3):143–155, Aug. 2013. ISSN 18759521. doi: 10.1016/j.entcom.2013.02.002. URL `http://linkinghub.elsevier.com/retrieve/pii/S1875952113000049`.

[48] L. Wilkinson and M. Friendly. The history of the cluster heat map. *The American Statistician*, 2008.

[49] G. Zoeller. Development telemetry in video games projects. In *Game developer conference*, volume 2010, 2010.