Chapter 1

Introduction

"The ability to collect, store, and manage data is increasing quickly, but our ability to understand it remains constant."

—Ben Fry, 2004

1.1. Temporal Information Visualisation

Information visualisation is a set of techniques that enable us to take information and represent it visually, so that complex data can be made easier to understand. For years, people have been using visual representations of abstract data to make discoveries, support decisions and explain phenomena. Lambert and Playfair used graphs to represent statistical data in the 1700s (Tufte, 83), and the techniques have been evolving and improving ever since. As a discipline, computer-supported information visualisation has its roots in the beginnings of the field of human-computer interaction (HCI) in the 1980s.

Just as there are many types of data we may want to analyse, there are also a range of techniques used to represent these data. There are three broad roles that the use of information visualisation (or "infovis") fall into:

EXPLORATION—These visualisations are generated to allow a user to explore a data set, generating hypotheses and then finding answers and explanations as they go. These visualisations will often incorporate interaction between the viewer and the visualisation tool.

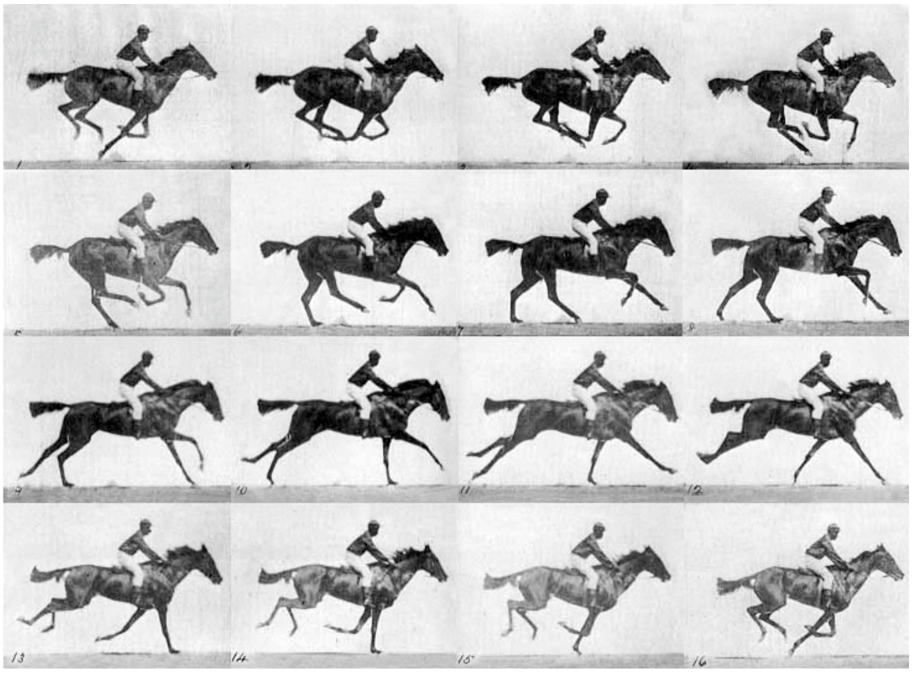
CONFIRMATION—These visualisations are used when the user has a specific query about the data that they want to answer, or a hunch about properties of the data that they would like confirmed. Having these questions as a starting point may change which visualisation technique is chosen.

PRESENTATION—When a salient property of a data set has been identified, a third type of visualisation may be used to best present this information to other observers, with a focus on communication and aesthetics.

The goals of visualisation have variously been described as to "facilitate comparison, pattern recognition, change detection" (Hearst, 2003), and to "amplify cognition" (Card, 1999). Infovis helps the viewer to turn a picture of abstract data into a *mental model*—an internalised understanding of the properties, features and scope of a particular collection of data. Challenges arise in the designer's choice of representation for data along with their use of colour, space, visual association and level of detail. The field has advanced considerably through the application of knowledge gained from research into human vision and perception (Ware, 2004).

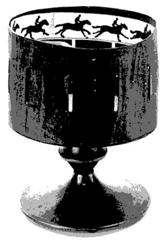
Many of the most traditional information visualisation techniques that have been developed over the years—like bar charts, scatterplots, maps, matrices and graphs—were designed to be used with *static* data that had been collected and could now be visualised all at once, such as survey results and collections of experimental readings.

Temporal data are those data that vary over time. For example, one may want to see the historical trends of a particular company on the stock market, or a researcher may want to analyse the structure of a complex network at different stages of its evolution. *This thesis is concerned with the visualisation of temporal data.* Visualisations of these complex dynamic data introduce new challenges beyond those encountered when working with only static data. The goal of this work is to introduce new techniques that are designed to maximise a user's understanding of a complex temporal visualisation.



1.1 Eadweard Muybridge's famous 1878 sequence of high-speed photographs which proved for the first time that all four of a horse's legs leave the ground when they gallop. (Hendricks, 75)

Time is often an essential aspect of a data set. Often thought of as a dimension, time is frequently visualised as one too (i.e., as an axis in a two-dimensional chart). Take Figure 1.1, which depicts the motion of a galloping horse in a dense format known as *small multiples* (Tufte, 90, p. 67). In this sense time lends *order* to data, allowing a visualisation designer to show multiple successive values of a variable at different times, and allowing the viewer of such a visualisation to perform local comparisons between these states, and identify paths, progressions, trends and causalities.



When these individual frames in Figure 1.1 are played together, each one replaced with the frame that follows it after a brief delay, or through a rotating zoetrope machine (inset), the result is *animation*: the illusion of motion which makes it easier for a viewer to follow the information that they are presented with.

In *The Illusion of Life: Disney Animation*, the authors say, “Man always has had a compelling urge to make representations of the things he sees in the world around him…” (Johnston, 95). In information visualisation, it is often that which *can't* be seen by vision alone that we are interested in representing.

1.2. Temporal Data

Many modern data sets are both large and dynamic. Though the processing power and storage capacity of modern computer systems has continued to advance significantly, the expansion of data available to process has increased hugely in that same period, generally outpacing our ability to analyse and make sense of it.

The sum total of human knowledge is now doubling every five years (Bennett, 99), not to mention the amount of data upon which this knowledge is based. Furthermore, our own finite capacity for decoding and absorbing such information presents a restriction. In an increasingly complex world, we will find ourselves turning towards visualisation more and more, as the high bandwidth of visual representation and perception allows large amounts of information to be presented at once (Ball, 02).

Sources of temporal, or dynamic data abound, and differ in a number of ways, from the frequency that they are updated, to how much they change over time. On one end of the scale, data sets like online social networks will generally change slowly, typically evolving by growing ever-larger of the order of one or two changes per week. Program source code in a multiplicity of programming languages can be similarly analysed as it evolves (Eick, 92).

1.2 A view of Apple Inc's (AAPL) stock price for the year starting October 23 2007–October 22 2008.

The top line graph is the stock price, which fluctuates between a high of $201 and a low of $91 over the 52 week period.

The bottom bar graph shows trading volume (in millions of shares).

Other sources of data change more rapidly. Examples of these type of data include environmental sensor networks which are constantly recording updated readings for factors like temperature, ambient light and sound. On the far end of the scale are realtime data like the stock market, which is constantly churning and making small and large leaps in value every day. Figure 1.2 shows the value of a stock over a one-year time segment. Data like this is almost impossible to divorce from its temporal context. Though a view of the historic trends in this data may be edifying and interesting, it can often be irrelevant to future movements in value, as the events of 2008's global economic slowdown made so clear.

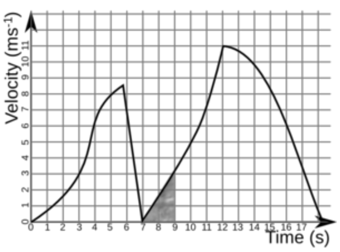
1.3 A chess game in which the Queen's Gambit move has been accepted by Black.

With this in mind, some data sets can be said to have a *historical context* which is information that is useful to have in understanding the current state of the data. Some applications of this data will not require this historical view of the patterns and trends that have occurred in the past, as it is more important to present the viewer with the most recent information up front.

For example, a view of the final cards you have been dealt in a game of five-card draw poker is less useful if you are not given any information about what cards you held originally before replacement. Conversely, a chess board in the early stages of a game, as in Figure 1.3, is perfectly playable without a move history, meaning that a chess player can sit down at a game that has already begun, and still have an excellent chance of choosing a good next move.

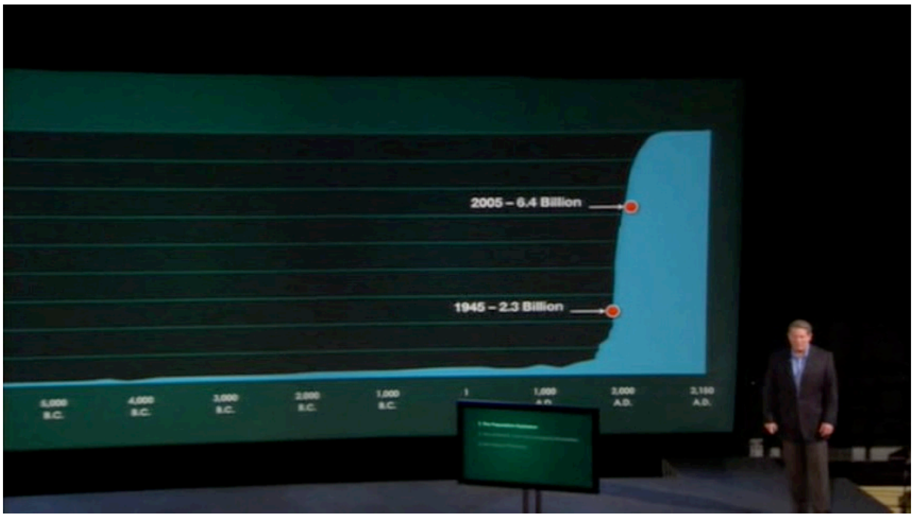
1.3. Visual Representations of Time

1.3.1. Encoding Time



Depending on the properties of the data set and the goals of the developer, there are many options available to represent time in visualisation. The straightforward *time series* have traditionally been used to show the changes in value of a single datum or multiple data over time. This is achieved by explicitly making time a spatial dimension in the visualisation. A vehicle's velocity over time is a good example of this, as the viewer can compare the value of the vehicle's velocity on the y-axis over a period of time on the x-axis (inset).

Time series have the capacity to show striking changes in value over time. They have long been used as a means to frame an argument for the benefits of various political and economic policies. As the line connecting the data points together has a certain curve, it can be extrapolated forwards in time while maintaining a similar trajectory, to become a predictive aid in understanding possible and probable outcomes. Recently time series like these have been used effectively to draw attention to crucial geo-political events, such as Al Gore's use of dramatic time series (Figure 1.4) in both the book and film version of his *An Inconvenient Truth* (Gore, 06a, 06b).

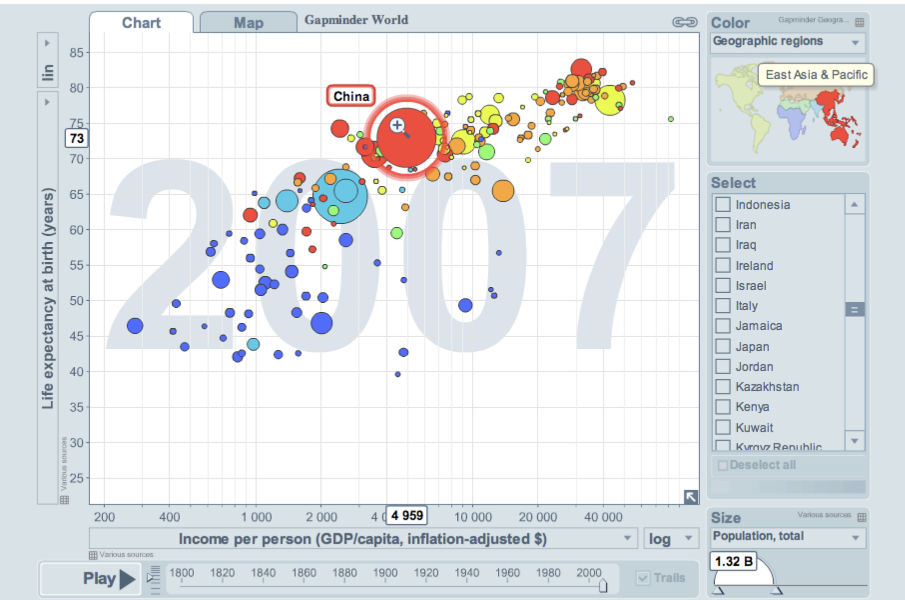


1.4 Al Gore presides over a time series visualisation depicting the dramatic and dangerous rise in carbon emissions that has occurred since the Industrial Revolution.

This presentational visualisation shows how many tonnes of CO_2 are pumped into the atmosphere through human activities. Even on the linear vertical scale used, the acceleration is remarkable.

The horizontal scale extends beyond the present day to predict the trend line for future years.

These kind of presentations can be made even more interesting through the use of animation, which can show dramatic changes among a group of variables as well as encode historical trends over time. Hans Rosling's talk at the TED conference in February 2006 demonstrated the power of these techniques by plotting geopolitical statistics using an animated bubble chart (Figure 1.5), made even more compelling by his accompanying spoken explanations (Rosling, 06).



1.5 The *Trendalyzer* visualisation tool from the Gapminder Foundation, subsequently acquired by Google. This advanced interface is used to compare the performance of the countries of the world. In this view life expectancy is compared to income per person, with the size of the bubbles indicating population per country.

Here time is presented as a separate timeline control at the bottom of the interface, and the current position in that index is rendered behind the chart. The data can be animated by pressing "Play", cycling the data points through their historical positions.

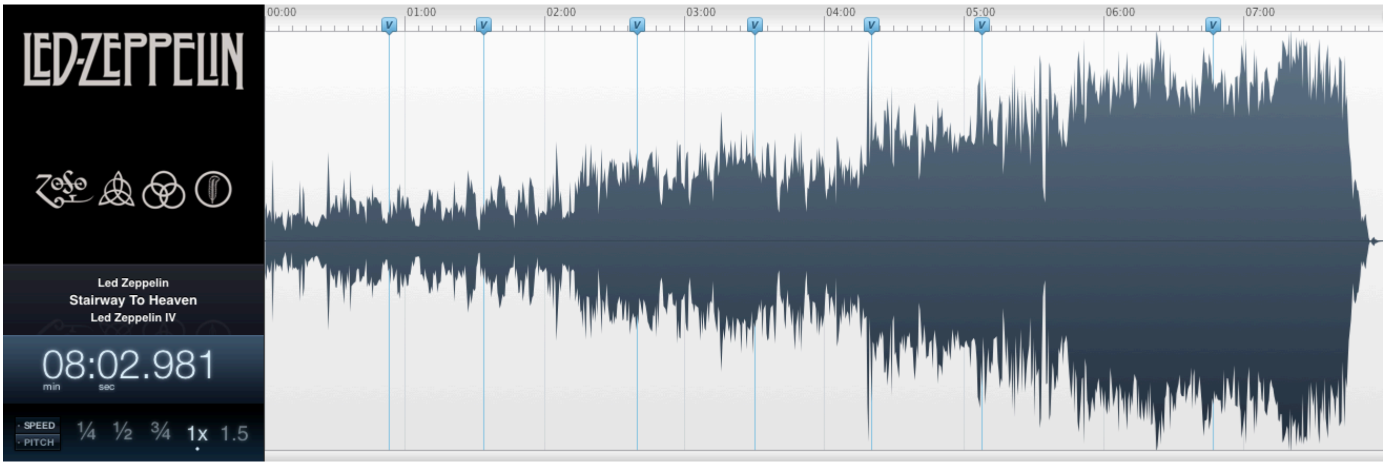
Various filtering controls on the right hand side allow users to interactively filter the data, winnowing them by population and toggling their individual visibility.

The addition of the spoken narrative in this presentation suggests temporal visualisation's capacity for telling a story about data. When appropriate visual encodings are used for the elements of the visualisation, a viewer can be comfortably guided through a display of complex data.

1.3.2. Time Series Variations

Time series are generally simple to apply to even multivariate data, but can come in many forms depending on the intended application of the visualisation. For example, the following examples are three distinct visualisations of the same piece of music:

Figure 1.6 shows a sound-wave view of a piece of music (Led Zeppelin's "Stairway to Heaven"). This is a typical static visualisation of sound, and from it we can see the high-level patterns of the song. Intuitively, we can deduce sections of instrumentation and quieter passages, intros and outros and repeated segments of melody.



1.6 A visual representation of the music track, as seen in the music editor *Capo*. The start of each verse in the song have been marked along the top timeline.

The song's compositional progression from the gentle intro's single acoustic guitar to its much louder hard rock final third can be seen clearly in this view.

This view treats the entire song as one static data set, and so it is comprehensive, but it cannot be fully deconstructed into the individual notes that make up the song. One benefit of this representation is that it is graphically compressible. By resizing the viewing window, we can trade the view's representational fidelity for screen space.

The second image in this sequence shows the musical notation for a song, which is a visual instruction used by a musician to recreate a song themselves (Figure 1.7). This similarly static view breaks the song down into individual notes, and also gives the reader a sense of the rhythm and tempo of a song by introducing literal breaks in the notation.

1.7 Music notation for the guitar introduction to Led Zeppelin's "Stairway to Heaven".

This time series shows the notes to be played along a musical staff, with breaks and rhythms represented graphically with glyphs.

This representation requires much more visual resolution to depict.

This visualisation is considerably more complex, and requires knowledge of a sophisticated and expressive encoding system to understand fully, but a trained eye (and ear) will be able to deduce how it should sound. In this type of time series, the staff, which acts as both an x- and a y-axis, does not represent duration as you go from left to right; merely showing the *order* of notes to be played. The glyphs representing the notes encode their duration directly, with ♩ representing a quarter note, a ♫ representing a pair of eighth notes and so on. Together with the markings for rhythm (4/4 time in this case), the musician can determine how quickly to progress through the piece while playing.

The final image in this series, Figure 1.8 shows a screenshot from a modern video game, which presents the player with a dynamic, animated view of the notes that make up a song. As the song progresses, the constituent notes scroll down the screen towards the player at speed until they reach a group of five circles on the screen which represent the five buttons that the player can press to play the corresponding note. The player is graded based on how accurately they press the correct buttons as the notes reach this threshold at the bottom of the screen.

1.8 "Stairway to Heaven" being played in *Frets on Fire*, a music rhythm video game.

The player is challenged to keep up with the song as it increases in complexity over time, continually presenting additional information to the user, who must keep track of their movement.

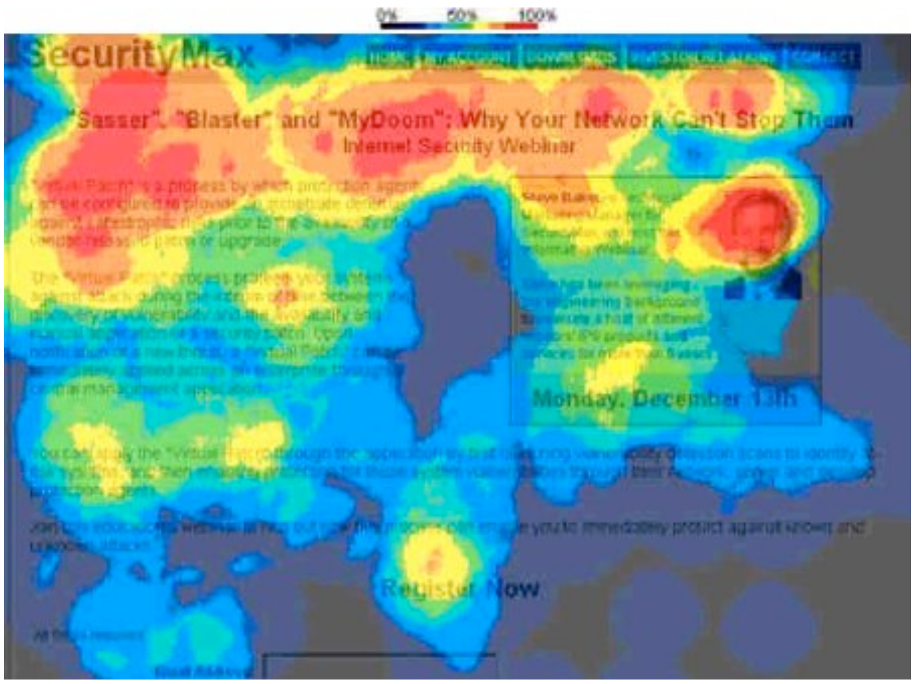
Visual techniques are used to draw the player's attention to changes and salient visual details in the centre of the visual field (while keeping the rest of the screen interesting to watch for onlookers). Through the use of animation, the viewer can build up a much better understanding of their rhythm and velocity through the data, and gain an intuitive sense of pauses and gaps in the data stream. This kind of view is particularly useful for situations where the previous state of the information is no longer useful; that is, it is immediately obviated by what is coming next.

These dynamic visualisations can be highly expressive and engaging, and show facets of the data set that are not as clear in a static visualisation. Again, a narrative can be built up of the data being visualised. However, dynamic animated visualisations do come with additional challenges, both for the visualisation developer as well as the eventual end user. The developer must take additional care to guide the viewer's attention to important visual details using appropriate visual encodings and avoid presenting data at an unreasonable rate. Similarly, the user must pay attention to the data as it is presented to them, so that they can perceive and understand all of the changes that are occurring, and remember events that have already occurred in the flow.

1.3.3. Encoding History

Time series explicitly represent the passage of time along a scale—presenting a spatial metaphor for time—but another possible representation of time is to render a view of activity over a period of time, and present this accretion of activity as another kind of visualisation.

Accretion-based views remove time as an axis in the view and instead show the *effect* of time, by allowing trends of activity to build up within a slice of time. For example, Figure 1.9 shows a *Heat Map* view of an eye-tracking survey performed on a website, which monitors the movement of a test subject's eyes as they interact with a webpage. The heat map is essentially a two-dimensional matrix, encoding the intensity of a subject's fixation on a point on the screen. The longer their eyes are fixed on an area of the page, the warmer the colour that builds up in the generated visualisation. As can be seen here, the user fixated most on the headline of the article, with secondary fixations on the author's photograph and first paragraph of text.



1.9 Visual heat map showing eye-tracking results. "Warmer" red and orange areas received more user eye fixations, for longer durations, while blue "cooler" areas were not fixated on.

These results suggest that users scan website headlines and the first few words of the opening paragraph. Logos and author photographs also draw fixations.

Aggregative visualisations like these forsake the ability to see how the data looked at a certain point in time, but allows the dominant patterns throughout time to emerge. Snapshots can be used to travel back to points earlier in time to allow the user to explore the history of the data.

1.4. Contributions of this thesis

As can be seen from the examples surveyed in this opening chapter, there are a variety of ways to visualise temporal data. Some techniques, such as animated data displays, use the passage of time to show the data progressing through many stages. Other techniques abstract the flow of time, and encode it as a visual axis in a data plot.

In this thesis we are focused on the design of information visualisation tools to present data in which time is an important property. To this end, we have identified three major classifications of temporal information visualisations: *time series*, *animated data structures*, and *versions*. These classes are described in detail in Chapter 2. Visualisations from each of these classes are in constant use, and opportunities abound for improving their efficacy. In particular we are looking for new methods that help *expose* and *explore* temporal information that is often left unrepresented in visualisations. This thesis presents three original visualisation tools we developed—one for each of the three temporal visualisation classes. Each tool was developed with a focus on their particular visualisation class; and taken together, we extract a set of design patterns for temporal information visualisation.

Our research question in each instance is "how can we visually express time in the most effective way in this interface to support the tasks that a user wants to perform on this data?" The visualisation of time should be clear, detailed, and sensitive to the perceptual capabilities of the human viewer in each case.

In developing these tools we have been exploring the visualisation space: looking for opportunities to harness useful properties of one type of visualisation in another. For example, Chapter 3 shows how a familiar visualisation technique for multivariate data can be applied to temporal data, while Chapters 4 and 5 show our technique for applying some of the benefits of time series (their easy readability and ability to show patterns) to animated data structures. Chapter 6 shows how we can introduce information about historical context into a familiar text editing interface that does not contain this information in existing tools.

Specifically, the goals of our three systems were to improve:

FILTERING capabilities for highly-complex, multidimensional data in time series, through novel interaction and clustering methods. (Chapter 3)

INFORMATION OVERLOAD issues in animated data structures through controlling and explicating the flow of incoming data. (Chapters 4 & 5)

UNDERREPRESENTATION of available temporal data in familiar user interfaces, by exposing this information in the interface. (Chapter 6)

We now detail the major contributions of this thesis:

1. The *Situvis* visualisation tool, a novel application and extension of the Parallel Coordinates technique for the investigation of time series data. Situvis allows for the interactive exploration of highly-multivariate time series data, showing thousands of records in a single view and making it easy for viewers to pick out high-level patterns, deviations and outlying data cases. Situvis was developed to support activities performed by researchers in the field of pervasive computing, and scenarios demonstrating its use are discussed, highlighting the features most useful to diagnosis and system design. (Chapter 3)

2. A set of experiments tracking subjects' understanding of the passage of time using visual change signatures representing the evolution of animated data structures such as node-link diagrams. Subjects were presented with a sequence of lightweight time-series visualisations of the activity within an animation of the development of a social network and asked about their understanding of the data. (Chapter 4)

3. The *Time Sequences* model, a new way of looking at animated data structure information visualisations. Our *Showtime* tool visually shows the passage of time in an animation. Time is manipulated to mitigate fast-moving events and remove long gaps of inactivity. (Chapter 5)

4. The *Deep Diffs* visualisation technique for comparing text documents between an arbitrary number of versions. Most file comparison tools available today only allow the comparison of two versions of a file, highlighting the differences between them. Our new technique allows many more versions of a text to be compared simultaneously, giving the reader or writer much more information about the recent history of the document. This tool was designed to support writing and editing tasks, and is validated with user studies in Section 6.6. (Chapter 6)

Some other supporting contributions from this work include:

1. A taxonomy of techniques for temporal information visualisation, breaking down previous developments into one of three classes. (Chapter 2)

2. A description of our approach to performing experiments using web-based tools and crowdsourced experimental subjects. (Chapter 4)

3. A collection of novel techniques for directing visual attention and encoding historical information in views. (Chapter 7)

4. A discussion of common design patterns which occur frequently in the design of visualisation tools for temporal data. (Chapter 7)

1.4.1. Relevant Publications

Elements of this thesis have been published previously in the following research papers:

A. *"Graphemes: Self-Organizing Shape-based Clustered Structures for Network Visualisations."* (2010)
Ross Shannon, Aaron Quigley & Paddy Nixon. In CHI '10: CHI 2010 extended abstracts on human factors in computing systems, New York, NY, USA, 2009. ACM. (Shannon, 10)

B. *"Time Sequences."* (2009)
Ross Shannon, Aaron Quigley & Paddy Nixon. In CHI '09: CHI 2009 extended abstracts on human factors in computing systems, New York, NY, USA, 2009. ACM. (Shannon, 09)

C. *"Situvis: A Visual Tool for Modeling a User's Behaviour Patterns in a Pervasive Environment."* (2009)
Adrian K. Clear, Ross Shannon, Thomas Holland, Aaron Quigley, Simon Dobson & Paddy Nixon. In Pervasive 2009 (pp. 327–341). (Clear, 2009)

D. *"Situvis: Visualising Multivariate Context Information to Evaluate Situation Specifications."* (2008)
Adrian K. Clear, Ross Shannon, Thomas Holland, Simon Dobson, Aaron Quigley & Paddy Nixon. In Proc. of 2nd Workshop on Ubiquitous Systems Evaluation. Co-located with Ubicomp 2008, Seoul, South Korea. (Clear, 2008)

E. *"Multivariate Graph Drawing using Parallel Coordinate Visualisations."* (2008)
Ross Shannon, Thomas Holland & Aaron Quigley. Technical report. UCD Dublin. (Shannon, 08)

F. *"Visualising Network Communications to Evaluate a Data Dissemination Method for Ubiquitous Systems."* (2007)
Ross Shannon, Graham Williamson, Aaron Quigley & Paddy Nixon. In UbiComp 2007 Workshop Proceedings, Innsbruck, Austria, 2007 (pp. 288–291). (Shannon, 07)

G. *"Collaborating in Context: Immersive Visualisation Environments."* (2006)
Ross Shannon, Aaron Quigley & Paddy Nixon. In AVI '06: Proceedings of the international workshop in conjunction with AVI 2006 on Context in advanced interfaces, New York, NY, USA, 2006 (pp. 13–16). ACM Press. (Shannon, 06)

1.5. Summary of following chapters

The rest of this thesis continues like as follows:

Chapter 2 (“Temporal Visualisation: Background and Definitions”) continues in presenting models and techniques for presenting time in information visualisation. We review the common data types, processes and visualisation techniques. These techniques involve not just presenting time as an axis, but also showing the *effect* of time by showing the accretion of activity after a period, and by showing the state of a complex data structure at different points in time. We describe our tripartite taxonomy for organising the prior work in this area.

Chapter 3 (“Temporal Patterns: Finding Structure in Data”) describes how time scales are represented using time series visualisations. Interaction with these tools is described, before we introduce *Situvis*, our time-series-based visualisation tool for the exploration of data collected from sensors in a pervasive computer system. We show how the properties of the Parallel Coordinates technique used make it appropriate for data recorded about human activity over time.

Chapter 4 (“Temporal Flows: Models of Time”) begins in discussing our second visualisation classification, animated data structures, such as node-link diagrams and trees. The introduction of animation leads to many complications for the observer, due to the changing display. We describe our method for performing simple user studies online using a web service, and use this method in a study to investigate the abilities of observers to understand the flow of time in dynamic information visualisations.

Chapter 5 (“Temporal Manipulations: Event Models”) introduces our model of *event sequences*, which compartmentalises a time series into a set of ranged events and their knock-on effects. The *Time Sequences* technique is explained, which applies the event sequence model to complex visual structures to increase a user’s understanding of an evolving data visualisation. Experiments we have undertaken to prove the efficacy of the time sequences technique for improving a visualisation’s understandability, and capacity for knowledge transfer are detailed.

Chapter 6 (“Temporal Slices: Encoding History as Visual Context”) discusses the encoding of history in visualisations, and demonstrates some new techniques for adding historical information to textual data that typically isn’t presented with any. The design of *Deep Diffs*, our novel tool for collaborative writing is explained, and a user study that shows the tasks that it is useful for is presented.

Finally, Chapter 7 (“Discussion and Conclusions”) reflects on these findings and describes further implications for the design of temporal visualisation systems. Here we conclude the thesis and offer suggestions for future work.