Narrative Construction in Information Visualization

Donia Yousef M. Badawood

A Thesis submitted for the degree of Doctor of Philosophy in Information Science

giCentre, Department of Computer Science
City University London

June 2015

Contents

List of Figures

List of Tables

Acknowledgements

I would never have been able to finish this thesis without the support from many people:

First and foremost, I am very grateful to my first supervisor, Professor Jo Wood, for his continuous support, understanding, encouragement, enthusiasm, and constructive feedback throughout the entirety of my research work, from when I began my MSc dissertation up until the completion of this PhD thesis. Honestly, I could not have wished for a better supervisor.

I am also thankful for my second supervisor, Professor Jason Dykes, for his support, constructive feedback, reading and reviewing papers, and especially for his support during various conferences and for introducing me to well-known academics in our field.

I would like to thank all the giCentre members whether still at City or elsewhere for the useful discussions during our meetings and my presentations to them, and for all their support throughout the past years. Special thanks to Dr. Aidan Slingsby and Sarah Goodwin for qualitatively coding a sample of my data to help me assess the reliability of coding.

Many thanks also to Dr. Simone Stumph for being the internal reviewer of my MPhil/PhD upgrade report, and for her constructive feedback at that stage.

Thanks to Mark Firman and Nathalie Chatelain, for all the administrative tasks, the processing of applications for fund to attend conferences, and the endless letters requests.

My thanks also go to the participants in the experiments I carried out.

I would like to acknowledge all the academics I met during various conferences for the constructive discussions about my work.

I would like to express my deepest gratitude to my parents for all their help, sacrifices, and encouragement throughout my whole life.

My thanks to my husband, and my lovely daughter, Hala, for understanding, supporting, and sometimes very needed distraction.

My thanks to my family and friends for all their best wishes and understanding.

Last but not least, my deepest gratitude to the souls of my grandfathers, who were so proud of me and who were waiting for the day I would come home after completing my PhD. Sadly, they passed away during the course of my PhD, but I am certain that they are proud of me.

Abstract

Storytelling has been used throughout the ages as means of communication, conveying and transmitting knowledge from one person to another and from one generation to the next. In various domains, formulating of messages, ideas, or findings into a story has proven its efficiency in making them comprehensible, memorable, and engaging. Information Visualization as an academic field also utilises the power of storytelling to make visualizations more understandable and interesting for a variety of audiences. Although storytelling has been a a topic of interest in information visualization for some time, little or no empirical evaluations exist to compare different approaches to storytelling through information visualization. There is also a need for work that addresses in depth some criteria and techniques of storytelling such as transition types in visual stories in general and data-driven stories in particular.

Two sets of experiments were conducted to explore how two different models of information visualization delivery influence narratives constructed by audiences. The first model involves direct narrative by a speaker using visualization software to tell a data-story, while the second involves constructing a story by interactively exploring the visualization software. The first set of experiments is a within-subject experiment with 13 participants, and the second set of experiments is a between-subject experiment with 32 participants. In both rounds, an open-ended questionnaire was used in controlled laboratory settings in which the primary goal was to collect a number of written data-stories derived from the two models. The data-stories and answers written by the participants were all analysed and coded using data-driven and pre-set themes. The themes include reported impressions about the story, insight types reported, narrative structures, curiosity about the data, and ease

of telling a story after experimenting with each model. The findings show that while the delivery model has no effect on how easy or difficult the participants found telling a data story to be, it does have an effect on the tendency to identify and use outliers' insights in the data story if they are not distracted by direct narration. It also affects the narrative structure and depth of the data story.

Examining some more mature domains of visual storytelling, such as films and comics, can be highly beneficial to this new sub-field of data visualization. In the research in hand, a taxonomy of panel-to-panel transitions in comics has been used. The definitions of the components of this taxonomy have been refined to reflect the nature of data-stories in information visualization, and the taxonomy has then been used in coding a number of VAST Challenge videos. The transitions used in each video have been represented graphically with a diagram that shows how the information was added incrementally in order to tell a story that answers a particular question. A number of issues have been taken into account when coding transitions in each video and when designing and creating the visual diagram, such as nested transitions, the use of sub-topics, and delayed transitions. The major contribution of this part of the research is the provision of a taxonomy and description of transition types in the context of narrative visualization, an explanation of how this taxonomy can be used to code transitions in narrative visualization, and a visual summary as a means of summarising that coding.

The approaches to data analysis and different storytelling axes, both in the experimental work and in proposing and applying the framework of transition types used, can be usefully applied to other studies and comparisons of storytelling approaches.

Chapter 1

Introduction

1.1 Research Context

For thousands of years, visual and oral storytelling has been one of the most widespread and commonly used forms of communication. In anthropology and philosophy, the word *myth*, which was derived from the Greek term *mythos*, was used to refer to a 'traditional plot or storyline which could be transmitted from one generation of tellers to the next' (Kearney 2002, p. 8). Mythic narratives can be divided into two main types: historical and fictional. As the terms suggest, historical or factual narratives describe real past events, such as biographies, while fictional narratives are about fantasy and using the imagination to fabricate events (Kearney 2002, p.9). From a critical theory point of view, what these two types of narratives have in common is *mimesis*, which is defined by Aristolte as the 'imitation of an action' (Kearney 2002). More recently, many scholars have also demonstrated the power of storytelling as a means of information transfer (Gershon & Page 2001, Tergan et al. 2006). Over time, storytelling has evolved from its origins as an oral tradition from generation to generation or as a simple visual stories found drawn on caves walls. This has expanded into new fields and new forms of communication such as novels, films, and computer games. Each of these storytelling forms has its own strategies, models, theories, and techniques. Recently, there has been great interest in storytelling through information visualization, resulting in two notable workshops on telling stories with data at VisWeek conferences in 2010 and 2011. There have also been a number of papers on the topic (Hullman et al. 2013, Kosara & Mackinlay 2013, Hullman & Diakopoulos 2011, Segel & Heer 2010). Although there are obvious differences in the details and ways stories are told through information visualization when compared with films or written text, and there are also differences in the nature and types of stories being told in each medium, there are certain concepts and characteristics of storytelling common to all fields.

To avoid confusion and uncertainty about the terminology that will be used throughout the thesis, it is important to define the terms used before proceeding. The words *narrative* and *story* are commonly used as synonyms, particularly in domains other than literature or philosophy. However, there is a difference between them. The Oxford English Dictionary defines *narrative* as ‘a spoken or written account of connected events’. It is the sequence of events that are organised into a story. On the other hand, a story is all events in the narratives, the kind of things that happen. Therefore, one could construct many narratives with the same underlying story.

To set the context and define the terms: *data story*, *constructed narratives*, and *data storytelling through information visualization*, we can start from these basic definitions. We could also look at what a story is made of, regardless of the domain of this story. Any story consists of some events and components; these can be characters (as in films or comics), facts, or pieces of information (as in news reports). These components are organised or ordered in a specific way or using steps to communicate a specific message, answer a question, or make a point. To achieve the intended goal of the story, this ordering or organisation of the components and events cannot obviously be random. It takes into account the relationships between different events or components, as well as the intended goal or message. A *data story* is a story that consists of data or facts. It can be structured and told using many *data narratives*. Combining this with the main essence of information visualization, we can explain *data storytelling through information visualization*. Card et al. (1999) defines information visualization as, 'The use of computer-supported, interactive, visual representations of abstract data to amplify cognition'. Hence, *data storytelling through information visualization* involves using visual representations of data, facts, or events, as well as the relationship between them to tell a data story that communicates a message or answers a question. This is very similar to the focus of data-driven journalism (Gray et al. 2012), where stories are told using some form of data; this particular form of storytelling has become more important and popular with the release and revolution of open data. Websites such as data.gov in the US and data.gov.uk in the UK provide rich sources for data stories that are of interest to many agents, ranging from local governments to business organisations to the general public.

1.2 Research Focus

Data storytelling is a broad topic that can be addressed from many different angles, such as the building of a system that supports specific elements of storytelling (Satyanarayan & Heer 2014, *Tablue* n.d., *Gapminder* n.d., Heer et al. 2008), the role of the language and body language of the storyteller, the flow of narratives and overall narrative structure, or the data type used in the story. To narrow the focus of the research, two specific areas of data storytelling through information visualization have been chosen. The first area is the delivery model, where this research explores what effects a direct narrative has compared with the interactive exploration of the story in information visualization, as well as how these impact the narratives constructed using each model. For this area, two delivery models were identified, the first model involves watching a video of a presenter/speaker telling a story using information visualization software, while the second model allows the users to explore an interactive visualization, as well as construct narratives by themselves. The research in hand aims to explore empirically how non-expert general users understand, construct, and tell stories using these two models of story delivery.

The second area of the research in hand is concerned with the transition types usually used to tell stories with information visualization. It addresses the patterns in which these transition types are used, and how they impact on the flow of narratives, and provides a visual language to characterise the use of these transitions in narrative visualization. Within this second area, the chosen case study data to explore transition types is the set of VAST Challenge videos from 2009 to 2011 (*Visual Analytics Benchmark Repository* n.d.). VAST challenge is a series of tasks released annually as part of the world premier conference in information visualization, IEEE VIS, where a large dataset is provided and a number of tasks specified. There are typically two or three mini-challenges (or tasks) and a grand challenge. Every year a different dataset is released with different task types, and a compulsory video submission is required to address these tasks. These videos provide a valuable and suitable dataset because the videos of any specific year are based on the same dataset and address the same tasks, and they each have the same length and format; they thereby provide suitable data for valid analysis and characterisation of the storytelling styles and transition types used.

1.3 Overall Research Aim and Objectives

The overall aim of this research is to use existing theory and techniques from storytelling more generally, and to investigate their relevance to storytelling in information visualization, and to examine empirically the impact of different forms of storytelling visualization on narrative construction. Specifically, it aims:

1. To identify the characteristics of storytelling in general and as an emerging field within information visualization.

2. To investigate the effect of the information visualization delivery on narrative construction and development.

3. To propose and develop a framework of transition types used to tell data stories through information visualization based on theories in other storytelling mediums.

4. To propse a visual language to characterise transitions in narrative visualization based on the proposed framework of transition types.

5. To use the visual representation of transitions in narrative visualization to summarise a number of storytelling styles in narrative visualization.

1.4 Value and Contribution of the Research

The main contributions of this thesis can be summarised as follows:

1. The main contribution of the experimental work is to add to the work done in narrative visualization a step further by providing empirical results on the effect of the manner in which stories are delivered. Its remit is the effect of stories delivered using information visualization by a speaker and those constructed by users as a result of interactive exploration of information visualization. In addition to summarising some

differences between the direct narration and interactive exploration of data stories narratives, the results of the conducted experiments provide some advice and points for data storytellers to consider when delivering their stories with each delivery model (Chapter 3 and 5).

2. The major contribution of the work on transition types is to provide a taxonomy and description of transition types in the context of narrative visualization, and an analysis of how this taxonomy can be used by others to code transitions in narrative visualization. In addition, it provides a visual summary as a means of summarising that coding. This visual summary enables the comparison of different stories and their progression, and identification of a number of storytelling styles as a result of this comparison. The taxonomy of transition types should help data storytellers to initiate and control progressions in their stories. The themes found after analysis and visual representation of stories in VAST Challenge videos may also be considered as models that can be adopted as desired by storytellers (Chapter 4 and 6).

1.5 Thesis Structure Outline

The overall structure of this thesis takes the form of seven chapters, including this introductory chapter, **Chapter 1**, which provides information on the context and main area of the research in hand, together with the scope, aims and objectives of the research, and a summary of the research contributions.

Chapter 2 provides brief background information on the main studies on storytelling in information visualization, and a literature review on some aspects of storytelling in general and in the context of information visualization, such as narrative structure, transitions and sequences in storytelling, communication modes and story delivery. It also identifies the gaps and extensions needed to address this topic and the emergence of the research in hand from previous work.

The next two chapters are the methodology chapters. First, **Chapter 3** is related to the experimental work done on the effect of direct narrative compared to interactive exploration of data stories. This chapter provides information on the research strategy that has been chosen for this part, along with summary and justification of the data collection methods and the procedure followed when conducting the experiments. It also highlights a number of research questions to be explored. Next, **Chapter 4** provides information on the taxonomy of transition types, the process of using of this taxonomy to code transitions in VAST Challenge videos, and the visual encoding of the transition used in each video (the transition diagram).

After that come the analysis and findings chapters. First, **Chapter 5** provides information on the qualitative and quantitative analyses conducted on the experiments data, and the emergent findings. Next, **Chapter 6** summarises a number of styles of storytelling in VAST Challenge videos by examining the transition diagrams created along with the actual videos.

Finally, **Chapter 7** is the Discussion and Conclusion chapter, where the main research aims and objectives are revisited and findings, limitations, and possible future work are discussed.

Chapter 2

Literature Review

While most of the research objectives identified in section 1.3 will be fully achieved by conducting the research methodology, this literature review contributes substantially by providing the background information required to fully understand the research area. It mainly focuses on Objective 1, *to identify the characteristics of storytelling in general and as an emerging field within information visualization.* It starts with a summary and review of the main previous studies on narrative visualization and shows how this field emerged within that of information visualization (section 2.1). The scope of the literature review is then expanded throughout the subsequent sections (section 2.2 to 2.4) to include some theories and techniques of storytelling from other domains such as filmmaking, journalism, and comics and discuss their relevance to storytelling in the context of information visualization. Section 2.2 discusses the role of narrative structure and provides some insights into a number of different methods used to tell a story in many domains including information visualization. Section 2.3 provides more detailed insight into the meaning of transitions in storytelling and summarises a possible taxonomy of panel-to-panel transitions in comic stories and how this relates to storytelling through information visualization. Section 2.4 discusses the role of story delivery and communication mode and how these impact on the way audiences perceive and comprehend stories. Finally, section 2.5 summarises the main points in the literature review and the need for further work in the area of narrative visualization.

2.1 Storytelling in Information Visualization

Information visualization has two main purposes: exploration and analysis on the one hand, and presentation and communication of information to some target audiences on the other (Fulkerson et al. 2009, Keim 2002). Storytelling or narrative visualization is usually listed under communicative or presentational visualization as one of the main characteristics of any story is the message or the point the storyteller makes or communicates. However, it can be argued that storytelling visualization, particularly with interactivity functions in dynamic visualization designs where users construct their own stories, combines exploratory and presentational visualizations simply because it is not possible to construct and tell a story using data visualization without first exploring the data using available interactive features or examining the static visualization.

Interest in and publications on storytelling in the information visualization community have seen significant growth over the last few years, particularly since 2010. Tufte (1986, 2001) discusses the communication of narratives over space and time through the use of data visualization. A well known early example of this type of data stories is Minard's (1869) map showing the losses suffered in Napoleon's Russian campaign of 1812 (Figure 2.1). In 2001, Gershon and Page stated that using storytelling is an effective method of developing visualizations and that to do so, it is necessary to adopt a number of storytelling techniques such as building the picture, animating the events, creating continuity, filling gaps, and resolving conflict and ambiguity. Given

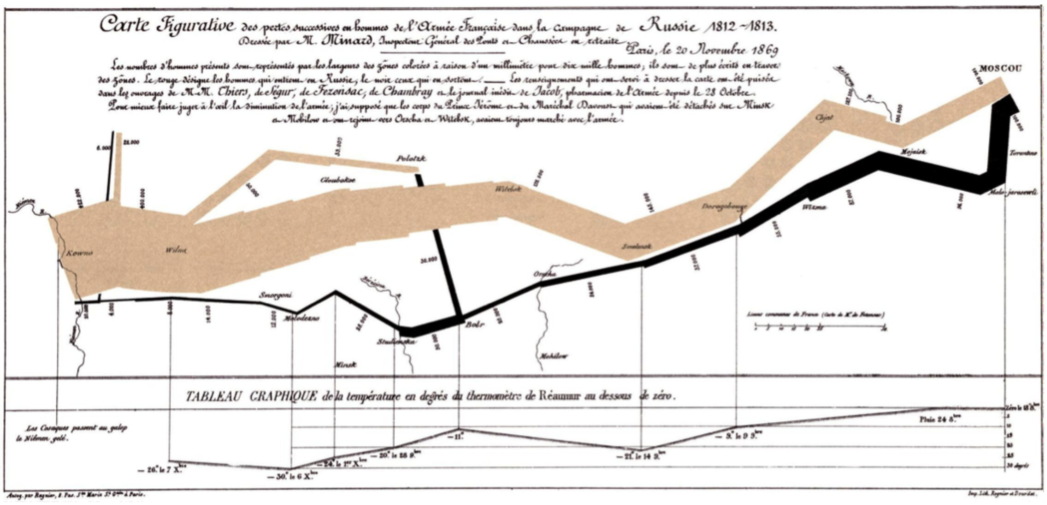


Figure 2.1: Charles Joseph Minard (1869) showing the losses suffered in Napoleon's Russian campaign of 1812.

that various types of information may be visualized, and that the information visualization methods and techniques used do not lend themselves to a one-size-fits-all approach (Heer et al. 2005), storytelling through information visualization becomes a challenging task. Wojtkowski (2001) makes a similar point: data visualization's defining feature is the complexity of the content to be communicated.

In addition to choosing the best-fit solution or visualization technique to the data in hand, which is the norm with any visual design, there are multiple formatting decisions in visual storytelling to be made, based on audience, topic, and objectives. These are vital to ensuring that the message or the point of the data-driven story has been effectively delivered and perceived by the target audiences. This requires deciding on the right data, amount of detail, distance between the target audience and story context, and balance between the interactive nature of information visualization and delivering an intended story or message. An interesting example that can be used as a framework to classify the relationship between the audience and the story context has been provided by Quesenbery & Brooks (2010). Although it focuses on storytelling in the context of engineering design, the model is applicable to any domain that has different types of audiences and needs, as with information visualization. The example is the design of an electronic medical records system, to create a story that helps explain the daily routine of hospital nurses as they check in patients and keep records of their medical conditions. Within this example, the framework comprises a three column table: Audience, Relationship, and Need. There are multiple types of relationships: for instance, nurses constitute an audience who is part of the story context; a patient is an audience related to the story context but possessing a different perspective from that of the central characters; while database developers assigned to work on a healthcare system for the first time constitute an audience who knows the context from a technical perspective (Quesenbery & Brooks 2010). It can be argued that such a classification and understanding of the intended audience is highly relevant to any visualization design. It is also essential in deciding on the amount of context and detail necessary to deliver a coherent story, and on predicting the path that a particular type of user may take in a highly interactive exploratory narrative visualization. A number of studies have been conducted in order to understand the manner in which different users approach information visualization. The importance of knowing the target audience is also highlighted in a number of sources (Grammel et al. 2010, Yau 2011).

To better understand the design space of narrative visualization, Segel & Heer (2010) analysed approximately 50 information visualization case studies that have elements of narrative visualization, mostly from journalistic sources, and looked for recurring design techniques and strategies. They considered how users were led through visualizations and identified two approaches, author-driven and reader-driven. The author-driven approach to information visualization is based on purposeful linear ordering of scenes in which the author's message and strategic audience guidance prescribes cognitive acquisition of key agenda items and learning objectives (Segel & Heer

2010). Conversely, a reader-driven approach circumvents the rigid formatting of author-direction, allowing for free interactivity with the presentation (Segel & Heer 2010). This implies changes in the story path that each user or audience may follow and that may also change the overall outcome of the story. There are inherent benefits to each approach based on the audience, the primary objectives of the presentation, and the disseminated information, challenging the author to consider the value of each technique before subscribing to one approach or the other if he or she has to choose between them. Furthermore, Segel & Heer (2010) identify seven genres used in storytelling visualization: magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and film/video/animation. They also suggest three main narrative structures based on the author-driven and reader-driven approaches. These structures will be discussed in section 2.2.

Following the workshop on Telling Stories with Data held during the VisWeek 2010 conference, and the Segel & Heer (2010) paper on the topic, many information visualization scholars and practitioners expressed an interest in this area with many blog posts and papers on the topic. A subsequent study by Hullman & Diakopoulos (2011) looked at the area of narrative visualization from a different angle, focusing on some rhetoric elements when using information visualization to tell news stories. The principle 'rhetoric' is usually about persuation, which is common in storytelling in different contexts, not just in information visualization. For example, the aim of storytelling could be solving a problem or making a decision. This concept is also embodied in storytelling with data in the information visualization domain. To illustrate, the designer presents or uses data visualization to tell a story, such as in Hans Rosling's many talks on various topics (Rosling 2009, 2008), he uses his presence in front of the graphic as part of that persuation. When the designer is not there presonally to present his/her data story, other rhetoric elements could be used. Hullman & Diakopoulos (2011) analysed a number of information visualizations that had storytelling elements, using concepts from other domains such as literacy studies, and tried to provide a theoretical framework of the rhetorical or persuasive techniques such as information access and mapping used at four stages of visualization design (data, visual representation, annotations, and interactivity), and how these techniques might influence interpretation.

As the field of narrative visualization grows, new studies continue to focus on different aspects of data storytelling with information visualization. A recent study by Hullman et al. (2013) analysed a sample of linear slide-show presentations and looked for visualization-to-visualization transitions between slides and users' preferences in terms of the order of visualizations used in an individual presentation. They found that in slide-show presentations, users are more comfortable and capable of retreiving information if the change between two visualizations/slides occurs on one factor. For example, this is effective when the visualizations in two subsequent slides are identical except that each shows the data for a different time period. This is also true to some degree when a particular transition between any two subsequent visualizations/slides is repeated throughout the presentation. As the focus of the study is related to understanding transitions in narrative visualization, further discussions on this work will be postponed until section 2.3.

Information visualization is usually concerned with patterns and relationships in data. Hence, reasoning and comparisons are very common and important to any visualization design, including those that represent narrative or storytelling visualization. Gleicher et al. (2011) focused on how comparisons are made in information visualization designs. They identified three different approaches to comparisons in any visual design: juxtaposition ('show the objects to be compared separately'), superposition ('show the objects to be compared in the same space'), and explicit encodings ('the relationships between objects are shown explicitly by providing a visual encoding of them') (Gleicher et al. 2011). It can be argued that the use of these three approaches to comparisons in visualization design has an effect on the flow of the data story and how it unfolds, as well as on the ordering, amount, and type of information communicated at any single step in the story, whether this step is provided in a slide, by using some interactivity, told directly by a speaker, or annotations on a visual design. It also has an effect on the effort provided by the audience, viewer, or user to understand the story, as well as on the ability to recall this information or these comparisons.

Finally, a recent paper by Kosara and Mackinlay (2013) summarises and reviews the work done on this area and highlights the need for more empirical work and further research to be done to characterise this emerging area within data visualization. They also encourage attempts to benefit from more mature storytelling disciplines to facilitate understanding and characterisation of this emerging area of information visualization. This is one of the main arguments of the study in hand, where utilising theories and techniques from other visual storytelling mediums should enhance our understanding of some aspects of our domain.

2.2 Narrative Structure

Elements of stories vary across different domains. For example, film stories should have a character as a main component while in data visualization or journalism the story normally consists of different kinds of data. However, common characteristics have been identified. All stories should have a structure that controls the progression and transition between events and builds a unitary, coherent story (Block 2001, Gershon & Page 2001, Quesenbery & Brooks 2010). Without this progression and structure, there is no story but a series of facts.

According to Tobin (2007), whichever format is used, the structure of every story follows the same sequence: a beginning, middle, and end. However, Gershon & Page (2001) state that in information visualization this format is difficult to apply and that 'all stories should have beginning, middle, and end but not necessarily in the same order'. This is also related to the widely known five-acts classic narrative structure of Gustav Freytag (1863): Exposition, Rising Action, Climax/Conflict, Falling Action, and Denouement/Resolution (Figure 2.2). This is common in novels and films where the Exposition initiates and introduces the main characters and setting to begin the story, after which the story builds and the action rises, until getting to the point in the story when the tension highest- the Climax/Conflict. From there, the story proceeds with falling action and resolving problems and finally we reach the Resolution/Denouement stage where the problem/conflict is resolved and conclusion arrived. However, this does not always happen in this order as Gershon and Page (2001) pointed out. Some storytellers may withhold some exposition or information and reveal them later in the narrative (Richardson 2002, Chapter 12). An example of this reversal is the inverted pyramid structure used in journalism (below).

However, the 5-acts structure and its pyramid shape as developed by Freytag is usually applied to novels and screenplays where there is a main protagonist, and this 5-acts structure relates to the journey of this protagonist in the narrative. Duarte (2010) developed what she called *The Contour of Communication* that represents the narrative structure in oral presentations. Figure 2.3 shows this contour of communication as illustrated by Duarte (2010). The

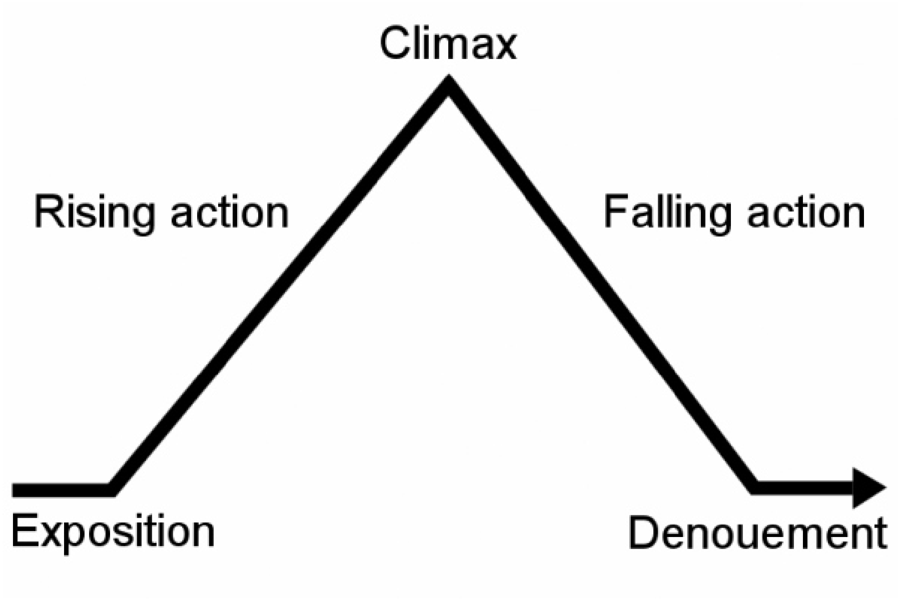


Figure 2.2: Freytag's (1863) pyramid 5-acts narrative structure

contour shows Duarte's argument that presentations have two main turning points that mark the change between the beginning, middle, and end of the story. As can be seen in Figure 2.3, there might be also other turning points between the beginning and end. These can be thought of as a repetition of Freytag's pyramid to cope with the complex nature of presentations that may extend beyond only one protagonist, topic, or storyline.

As stated above, stories should contain change or progression through a specific path. An obvious factor controlling this progression is change over time. While this is a common factor in data stories, of which many entail a change over time, it also has some limitations as regards storytelling in information visualization. The reference literature offers different opinions regarding the role of time in storytelling and visualization of data (Kuchar et al. 2006, Hullman et al. 2013). For example, time is used in facilitating understanding of story events, and in the structuring or authoring of stories. Different aspects of time are used to convey such things as absolute and relative time, and story and discourse time (Alexander 2011, Block 2001, Kuchar et al. 2006). Some studies have proven that a chronological order of events in

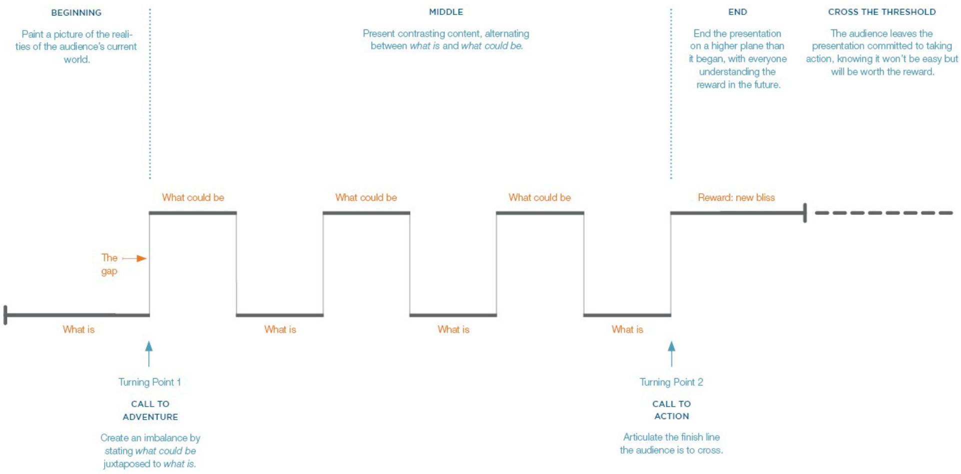


Figure 2.3: The Contour of Communication as illustrated by Duarte (2010)

stories enhances comprehension and makes it more memorable (Kuchar et al. 2006, Hullman et al. 2013). Moreover, one of the main arguments for the role of time in structuring and holding a story together is that storytelling is based on *events*, and events are successive actions performed during a period of time. Hence, without time, no storytelling is possible. This argument suggests that only time-series data can be effectively used in storytelling, but this is clearly not the case in all narrative visualizations. Therefore, alternative solutions for initiating a progression in story events, using information visualization in order to provide a coherent unitary narrative without explicitly using time or chronology to order or structure these events, are essential. According to (Kearney 2002, P. 4), 'storytelling may be said to *humanise* time by transforming it from an impersonal passing of fragmented moments into a pattern, a plot, a *mythos*'. Heath & Heath (2007) also discuss the possibility of structuring a story as a myth and then chasing a clue or an answer to a question. This may happen in film stories such as the crime genre as well as in information visualization. For example, a design study exploring the effect of the position and ethnicity of a candidate's name on the likelihood of their being elected in local elections in the UK studied only one dataset in a specific year, with no change over time in the data (Wood et al. 2011). It can be argued that this visualization told a story, as there was a clue or an answer for which the designers searched for as they investigated the problem in a series of steps, finally achieving a solution to the problem by the end of the story.

This may lead us to the general definition of a design study in information visualization and to what extent a design study constitutes a data story. Sedlmair et al. (2012) defined design study as, 'A project in which visualization researchers analyse a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines'. There is a huge overlap between design studies and data stories through information visualization. First, they are both considered problem-driven. This means the data story communicated through information visualization will depend on the nature of the problem being addressed in the design study, and obviously, this problem does not necessarily have to be chronoligcal or involve time-series data. Additionally, according to Sedlmair et al. (2012), it is very important to have a strong problem characterisation and abstraction, and this may involve dividing a high-level task into a number of low-level tasks to address the problem or the question being answered by the design study. It can be argued this hierarchy of the task may imply some kind of hierarchy in the solution (the visualization design or the data story), hence affecting the narrative structure of the data story. Sedlmair et al. (2012) also proposed a nine-stage framework summarising the process of conducting design study research. These stages are: learn, winnow, cast, discover, design, implement, deploy, reflect, and write. While designing a narrative visualization or telling a data story through information visualization may involve some or all of these stages, it can be argued the *telling* of the story might be more related to some specific stages in this framework. According to Sedlmair et al. (2012), during the final stage, *write*, the aim is 'to construct an interesting and useful story from the set of events that constitute a design study'. Although the authors are referring to constructing or telling a story in a form of a design study research paper, it can be argued this is relevant to the construction of a data story in any form. This story may involve a combination of some details from the analysis process along with the relevant insights from the data. In such cases, the selected details about the analysis or the design study process may help in adding a structure or chronology to the story that may not involve any time-series data. In other words, there might be no change over time in the data, but change over time in the representation and the analytical technique used. The data story here is about the researcher uncovering insights shaping the design. While this is not a story in the same way that Hans Rosling tells a data story in his various videos (Rosling 2009, 2008), it can be seen in other kinds of communicative visualizations such as the VAST Challenge videos.

Whatever the story time or story structure, stories can be told in many different ways in terms of the order of the events or narratives. In news stories for example, journalists often use an 'inverted pyramid' structure in their reports where story events are told in descending order from the most important to least important (Grunwald 2005) . However, it can be argued that this structure aims mainly to attract the reader's attention without making it necessary to read the whole story. In this case, important considerations should be taken into account such as the effect of the storyteller's subjectivity in judging the importance of events and the significance of not misleading the audience by using unrepresentative, less important cases at the top of the pyramid. Another example of the dicussion of time in storytelling is the common separation between actual time and story discourse time in films (Alexander 2011, Block 2001, Pramaggiore & Wallis 2011). The actual story time is the chronology of story events in the real world, while the discourse time is the period during which the story events unfold when the story is told in a film. This concept can be applied to storytelling presentations using information visualization such as those by Hans Rosling (Rosling 2009, 2008) and VAST Challenge videos that provided and reported answers for specific tasks using information visualization (*Visual Analytics Benchmark Repository* n.d.). Specifically, there might be no change over time in the data (actual time) but there is always change over time in the representation, and the analytical technique uses (discourse time). It can be argued that this sequential order and accumulation of pieces of information in order to form the big picture is what makes a story, and that this is a common characteristic of storytelling regardless of the medium.

As stated in section 2.1, narrative structures in storytelling visualization have been examined from a different perspective by Segel & Heer (2010). This study presents three structures of narrative visualization that are used via an author- or reader-driven approach. The unique stylisation of the story provides for variable experiences during the data navigation process, suggesting that in any presentation there is one best-fit structure that could conceivably enhance the viewer experience over the others. Firstly, the **martini glass** structure of a narrative visualization involves a tight narrative path early in the presentation which evolves into an open informational vessel of free, place-marked exploration and information review (Segel & Heer 2010). Alternatively, an **interactive slideshow** structure involves user-driven content in which key, visually coded information is actively transitioned via forward and back buttons towards the primary objectives of the presentation, highlighting key points via amorphous charts, graphs, and visual effects (Segel & Heer 2010). Finally, the **drill down story** structure involves the presentation of a general theme that is supplemented by individual (or grouped) details and backstories contained within the categorical distinction of the overarching theme and the study objectives (Segel & Heer 2010).

The relevance of this discussion of narrative structure to storytelling in information visualization is that different strategies of storytelling with information visualization impact on the structure of the stories perceived by audiences. This in turn alters the insights audience members gain and use to shape their mental model about the data story. There are other aspects of storytelling that can be studied along with narrative structure, one of which is the micro elements of the story structure or the transitions between smaller segments of the story. This is discussed in more detail in section 2.3.

2.3 Transitions and Sequences of Events in Storytelling

A transition simply means 'the process or a period of changing from one state or condition to another' (*Oxford English Dictionary* 2012). This basic definition of transition can be applied to the use of this word in several contexts, including storytelling and information visualization. As discussed above, all stories contain some form of change in which the narrative structure adopted controls these changes. There are micro-level elements of the changes in each story. These can be called transitions. Transitions control the more detailed changes and progression in any story, for example scene transitions in films (Bordwell 2008, Ganti 2004) and panel-to-panel transitions in comics (McCloud 1994, 2006). Some may argue that the meaning and use of transitions in information visualization is obvious. For instance, in a time-varying story that is based on change over time, the transition occurs from T1 (point of time) to T2 (another point of time). In stories that describe a process, the transition occurs between the steps of this process, showing the cumulative outcome of these steps. These transitions are popular in small multiples and slide-show presentations. However, as communicative visualization and storytelling is usually used to make a specific point, as it has a purpose, whether to persuade or inform, the effective use of transitions to achieve the purpose, keep the flow and continuity of the story, and hold the audience's attention extends beyond this simple interpretation of the meaning and role of transitions and may comprise more than one type or axis of transition at the same point. For example, this may involve changing the visual representation and also introducing or focusing on a new data point/variable.

Storytelling through information visualization, whether the story is delivered by a direct narrative or a speaker/presenter, or explored interactively by users, consists of a chain of actions, each triggering a transition from one event or scene to another. As a result, the story unfolds smoothly and gives a sense of continuity. This is similar to continuity editing in films, where there are a number of techniques to ensure the smooth transitions between shots and scenes (Bordwell 2008). Tufte's (2001) discussion of narratives of time and place, as stated in section 2.1, states that time and place are the two elements that should be used to create this continuity. However, with the variety of data that can be visualized and used to tell coherent stories, it can be argued that time and place are not the only factors that can be used to tell a story through information visualization. The designer may choose a different path and transitions to make their point depending on the task they want to achieve and the message they want to deliver. This also depends on the complexity of data and contents of the story, and sometimes on the storytelling genre (Segel & Heer 2010). While there are some helpful studies and rules of the graphical representations suitable for different data types and patterns usually represented and communicated through the use of information visualization, such as temporal patterns, network data, and geographical patterns, the choices of transitions the author has to communicate a message or make a point and tell a story through information visualization are much more complex. It sometimes requires using several visualization techniques, a combination of conflicting views or facts, etc.

There are good and valuable studies on transitions in information visualization in general. Heer & Robertson (2007) provide a taxonomy of animated transitions in statistical data graphics. They conducted a user study on different aspects of using animated transitions between different statistical charts using object tracking tasks. This study has been followed by Dragicevic et al. (2011), who also conducted a user study on animated transitions, particularly on temporal distortion or pacing of animation. Although these are useful and valuable user studies, they are limited to animated transitions and used object tracking tasks to look for time and error metrics. As stated in section 2.1, a recent study has been done to fill in this gap; this study examines visualization-to-visualization transitions in authored slide-show presentations and provides some guidance to the designers of this form of narrative visualization (Hullman et al. 2013). Based on the definition of a transition provided above, a transition occurs between two states or story units. Therefore, coding and identifying transitions in any type of narrative requires defining these states/units and identifying their boundaries. Hullman et al. (2013) consider each slide in a slide-show presentation as a state, which is a very straightforward approach that can be applied to this type of narrative visualization. However, they also state that each slide or static visualization is likely to be processed sequentially, but coding this sequence is more challenging due to the difficulty of dividing the slide or static visualization into smaller states (Hullman et al. 2013). Despite this difficulty, it can be argued that this can still be done, especially with other forms of narrative visualization, such as those that have an accompanied oral narration that guides the sequential analysis and reasoning process. Hence, further work that addresses the issue of transitions in narrative visualization from a different perspective, using different genre(s) and different theories to tackle this issue, may be highly beneficial and will definitely extend the work done on this area. This is what the research in hand will look at in Chapter 4.

There are a number of factors that may aid understanding, controlling, and carrying out the transitions in any data-driven story delivered or told through information visualization, particularly when stories are interactively explored and constructed. One factor is the use of annotations (Hullman & Diakopoulos 2011). It can be argued that annotations are essential not just as legends or axis-labels, but also in leading the users through the story path to its main message or point in static, and dynamic or interactive visualizations. This contributes to the process of narrative construction and the formation of a mental model by users and audiences. Kosara et al. (2003) discuss the issues of context and focus and how the transitions occurred between them. An example of this could be an interactive step such as distortion or zooming into a specific area on a map, which entails a transition. However, the discussion of these techniques is beyond the scope of the research in hand.

The characteristics and taxonomies of transitions in storytelling vary across domains. In comics, McCloud (1994, 2006) proposes taxonomy of transition types from one panel to another. His taxonomy consists of six types of transition: (1) *Moment-to-moment:* 'a single action portrayed in a series of moments'. (2) *Action-to-action:* 'a single subject (person, object, etc.) in a series of actions'. (3) *Subject-to-subject:* 'a series of changing subjects within a single scene'. (4) *Scene-to-scene:* 'transitions across significant distance of time and/or space'. (5) *Aspect-to-aspect:* 'transitions from one aspect of a place, idea, or mood to another'. (6) *Non-sequitur:* 'a series of seemingly nonsensical, unrelated images and/or words'. Figure 2.4 shows this taxonomy as represented in McCloud (1994).

This taxonomy of transitions is an important theory in storytellig in comics for a number of reasons. It provides a detailed explanation and analysis of the process of constructing and telling a story in a visual medium through the use of a combination of pictures and words. It can be argued that communicating data stories through information visualization consists of that same combination of pictures (the visual representation) and words (such as annotations, labels and oral narration). However, information visualization has one more important component that influences the story development: it is highly an interactive medium. Despite this fact, it can be argued that McCloud's (1994) taxonomy of transition types is still valid and applicable to storytelling through information visualization as this interactivity still occurs in a sequential manner, with one action at a time leading to the next event or panel. This is agrees with McCloud's (1994, 2006) definition of comics stories as a sequential art form. In addition, storytelling in general and comics in particular entail the selection of key appropriate images/panels, organising them in a specific period of time or number of words, pages, panels, or shots to make a coherent and meaningful story/narrative. All these strategies can also be easily related and applied to data stories told with information visualization.

Before McCloud's (1994, 2006) taxonomy is adopted and applied to information visualization case studies, this model should be refined and each type of transition should be defined in the context of information visualization. This refinement and development of a new framework for transition types used in narrative visualization is described in detail in Chapter 4, but before discussing the refinement and application stage, it is important to highlight some issues and limitations of this taxonomy that should be taken into account when dealing with other forms of storytelling such as narrative visualization and

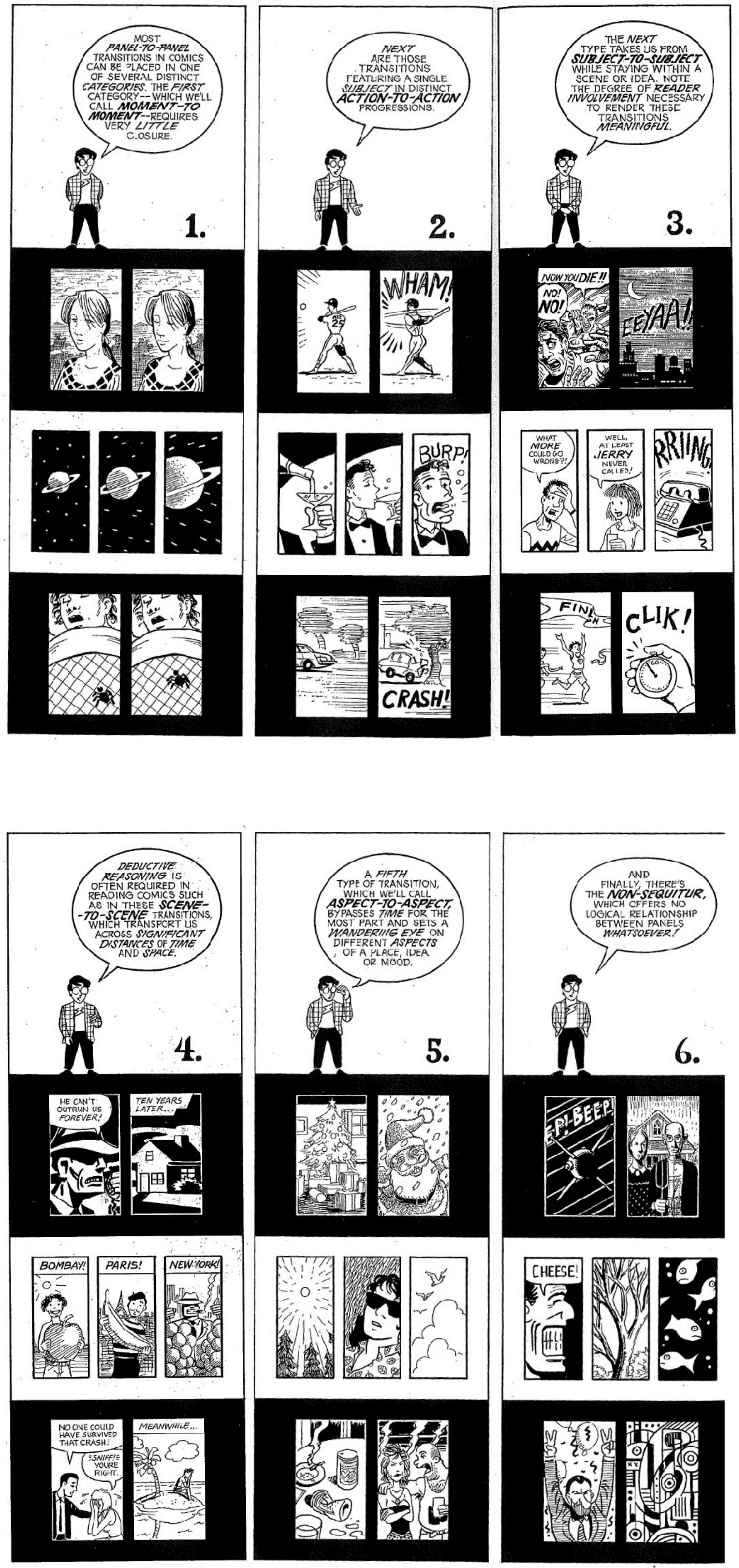


Figure 2.4: Panel-to-panel transitions in comics as defined and illustrated by McCloud (1994, 2006).

sometimes comics as well. Cohn et al. (2012) propose that visual narratives in comics are organised with a hierarchic narrative grammar. This means that a number of panels can be grouped together into segments depending on their contents and place in the whole narrative, and these segments of panels can be further grouped into higher level segments. All these together contribute to the global narrative structure (Cohn et al. 2012). This does not eliminate panel-to-panel transitions. The same also happen in various types of narratives. For example, it occurs in grouping of a series of shots into one episode of a TV show. In written narratives, in addition to sentence-to-sentence relationships/transitions, there is also a relationship/transition between paragraphs, sections, chapters, etc. It can be easily argued that this hierarchy exists in many narrative data visualizations where the structural segments may vary based on the topic or sub-topic addressed, the visual representation used, the question or sub-question being answered. Additionally, the transition is not necessarily limited to transitions from a sentence directly to the next one, or from a panel directly to the next one; it might be delayed or occur between a sentence and another sentence further on in the narrative. This is also applicable to narrative data visualization. For example, in a slide-show presentation, there might be a relationship between slide 1 and slide 4 in the presentation (delayed transition). There might be also more than one factor that governs or specifies the transition between two states, slides, or steps in the process of approaching a narrative visualization. For example, this might be seen in a change of time as well as a change in the data. These factors and characteristics of narratives will be taken into account when looking at transitions in narrative visualization in Chapter 4.

2.4 Communication Modes and Story Delivery

Some may argue that the data stories are created and structured in peoples' minds and not in the visualization design itself, as the latter only supports them in constructing this mental model. A similar point has been raised by Kosara (2010) where he suggests that 'stories don't tell themselves', and that visualization only sets a background or context, but that stories are told by people, including the users of the data.

Chen (2006) explains that different users may create different mental models even when using the same system. Hullman et al. (2011) also explain that small changes in representing information visualization may provide different results or insights. In the same way, it can be argued that different storytelling strategies with information visualization may result in differences in the way people perceive, understand, and construct narratives from them. There are various factors that can be examined, and one of them is the story delivery model, which will be addressed in the experiments reported in Chapter 3.

Various modes of communication exist, regardless of what is being communicated. These include scientific data in educational settings or stories in novels, films, or children's books. Each of these communication modes has its own nature and characteristics. Each may also have a different impact on how people (whether they are readers, listeners, users, etc.) comprehend what is being communicated. Before detailing the procedures and findings of the conducted experiments/user study, it is beneficial to review some modes of communication and story delivery techniques from other fields to enrich the scope of the study in hand and explain the relevance of these techniques and how each is expected to influence stories and comperehension on the part of the recipients.

Speech and writing are two communication modes that can be used in education, in telling stories and in communicating many other forms of content (Schallert et al. 1977). Although some well-crafted information visualization studies have assessed differences in various techniques or features in user tasks such as the use of animation and animated transitions (for example, Heer & Robertson (2007)), no studies have been found that examine differences between communication modes. For example, there is little or no study on the relationship between writing and speech or between direct narrative by a speaker and interactive exploration of the data. The differences between written and spoken communication can be thought of in relation to different communication modes of data stories through information visualization. One of the main differences between the nature of speech and that of writing is that speech has some prosodic features that are missing in writing, such as rhythm and stress. The lack of this information in writing requires the reader to have enough skills to interpret the writer's message (Schallert et al. 1977). It can be argued that, in the same way, watching a speaker who is using information visualization to tell a story (closer to speech mode) differs from interactive exploration of the data (closer to writing mode). Although both use information visualization to help deliver a data story, each of them has their own characteristics, advantages, disadvantages and effect on comprehension.

In films and documentaries, a technique called 'direct narration' or 'voice-over narration' is sometimes used to tell stories (Pramaggiore & Wallis 2011). Voice-over narration means that there is a narrator, who might be a character in the film, who narrates and directs the viewers through the scenes in the film. The role the narrator has in the story they describe may be made explicit at the outset or revealed as part of the storytelling process itself. The extent to which this technique is used varies from extensively narrated events to restricted narration, where the events in the story just unfold without direct narration. An interesting critique of the voice-over narration technique is that it replicates what the visual narrative already shows (Henderson 1983). Hence, it is suggested that this technique should be used to convey something different, and in some cases to highlight or emphasise some information (Henderson 1983). A similar delivery model to the voice-over narration technique in information visualization is employed when a speaker uses information visualization to tell a story. An example of this model of delivery is Rosling's talks on various issues in global health and economics using an animated bubble chart (Rosling 2009, 2008). It is difficult to avoid narrating what is shown in the visualization in such talks, but it should provide something different and add depth to the data stories and insights already shown in the visualization. Another consideration relevant to this story delivery technique is the effect of the personal and subjective interpretation and/or recommendations provided by the narrator on the viewers' engagement in the story, and on their acceptance of its message (Pramaggiore & Wallis 2011).

The lack of empirical evidence as to what effect an authored narration has on the interpretation of data-rich visualization was the motivation for the work carried out as part of this research. In the experiments reported in Chapter 3, the effect of the story delivery and narration model on the narratives constructed by audiences will be explored.

2.5 Summary, Emerging Issues, and Need for Empirical Research

This chapter reviewed the work done on storytelling with information visualization. It also highlighted some of the general theories and aspects of storytelling and how they relate to data storytelling with information visualization.

Table 2.1 summarises the differences between traditional storytelling and storytelling with information visualization.

Table 2.1: A summary of the differences between traditional storytelling and data storytelling using information visualization.

Traditional Storytelling	**Data Storytelling using Information Visualization**
In classic narratives, it is possible to have science fiction, fantasy, and/or things that are apparently removed from real life.	Data visualization is used to explore real things, facts, or data.
In classic narratives, there is usually a protagonist. Therefore, a classic narrative structure such as Freytag's (1863) 5-act structure that relates to the journey of a single protagonist in the narrative can be followed more easily. This does not interfere with the possible existence of many other characters, but the main focus is on this protagonist.	Data storytelling using information visualization usually lacks an obvious protagonist and extends beyond only one protagonist, topic, or storyline. Hence, more recent structures such as Duarte's (2010) contour of communication, which allows many turning points (i.e. not just those marking the separation between the beginning, middle, and end), could be more suitable to the nature of this form of storytelling.
Stories usually follow a strictly authored ordering of events, and mostly the audience cannot change the order of the narratives in the story.	In addition to the author-driven approach to storytelling, a more reader-driven approach, which allows for free interactivity with the presentation (Segel & Heer 2010) is also possible and common. The interaction between the audience/user and the story may imply changes in the story path and therefore, may also change the overall outcome of the story.
In traditional storytelling, there are sets of conventions, theories, and devices that can be used to enhance the storytelling experience. These are well known and well understood by people who consume these stories.	Because storytelling using information visualization is relatively new, there is a different set of conventions. Scholars in this domain are still exploring new ones.

Furthermore, the literature review reveals that storytelling by information visualization has been a topic of interest in recent years and although some useful studies have been conducted on this topic, the majority of these studies were designed as case studies and theoretical frameworks on specific aspects of storytelling. As a result, there is a need for empirical evaluation study to compare different approaches of storytelling with information visualization. The lack of these studies was the motivation for conducting empirical experiments comparing two different models of story delivery; the first is a direct narrative by a speaker using information visualization software, and the second is interactive exploration of a dataset using information visualization software. These experiments focus on comprehension/understanding rather than usability and metrics such as error and time commonly used in information visualization evaluation studies. The experiments conducted, as well as their procedures and contributing factors are detailed in Chapter 3.

The literature review reveals that much of the work on transitions in information visualization has focused more on technical or graphical techniques used than on placement of these transitions within a story, their impact on the meaning and comprehension, the ways in which these transitions are used, and their role in narrative construction. The extent to which different types of transitions have been used to tell stories with data through information visualization is still under-represented in the literature. Hence, further study is required to cover this aspect of narrative visualization. Analysis of some characteristics and taxonomies of transitions across other more matured domains of visual storytelling, such as films and comics, could be highly beneficial in addressing the issue of transitions in storytelling visualization. This research is concerned with the use of transitions, changes, or shifts within data-driven stories. The second part of the reseach in hand involves qualitatively coding a dataset of storytelling visualizations that use different transitions to tell a story based on the McCloud (1994, 2006) taxonomy of transition types after refinement and adjustment of the model. A dataset of VAST Challenge videos (*Visual Analytics Benchmark Repository* n.d.) has been chosen for this task for a number of reasons. Firstly, it will provide a visual summary of these stories to enable exploration of the patterns and styles of storytelling and the use of transitions within this dataset. Subsequently, it will explore the patterns of the use of these transitions within the chosen case studies. The details of the work done in this part are presented in Chapter 4.

The next two chapters (Chapters 3 and 4) detail the experimental work done to explore the effect of information visualization delivery on narrative construction, and the proposal of a transitions framework for data stories based on or inspired by McCloud's (1994, 2006) taxonomy of panel-to-panel transitions in comics. The subsequent two chapters (Chapters 5 and 6) summarise the analysis and findings from the experiment data, and themes of storytelling based on the proposed transitions framework.

Chapter 3

Research Methods I: Information Visualization Delivery

The literature review chapter above touches on some aspects of storytelling and the previous work done on this sub-field of information visualization. The literature review has shown that little empirical evaluation exists to compare different approaches of storytelling visualization. This chapter is mainly related to Objective 2, *To investigate the effect of the information visualization delivery on narrative construction and development.* It is divided into two main sections. Section 3.1 defines the two story delivery models whose effects on narrative construction are explored, and describes the research strategy adopted to achieve the research objective. Section 3.2 details the experimental design and procedure of the two sets of experiments conducted to examine the effect of information visualization delivery on narrative construction. To avoid any confusion, it is important to note that both sets of experiments address the same research objective, with the first set designed as within-subject and the second set as between-subject experiment. Each has its own advantages and disadvantages and in the case of the research in hand, using both designs to address the research question is expected to outweigh the advantages and disadvantages of each and provide richer explanations of some findings. More details are provided in the subsequent sections.

3.1 Research Strategy and Questions

The primary aim of this research is to gain insight into the area of narrative visualization, to understand and characterise some of its aspects, and to conduct some empirical work to fill in the gaps and expand the work in this area. It starts with a general question as to how different models of information visualization delivery differ and impact on narrative construction and development (Objective 2). More specifically, this research compares the effects of two different information visualization delivery models on people-constructed narratives:

1. Narrative model: Direct narratives by a speaker using information visualization software to tell a data story to the audiences.

2. Software model: Lets the users explore the data interactively using the visualization software to construct data stories.

Saunders et al. (2009) conclude that it is important to choose the appropriate research strategy for the research question. To achieve this objective (Objective 2), an experimental research strategy will be adopted where the two delivery models represent the main experiment factors (or independent variables) whose effect we aim to explore. To explore this effect, we identify a number of questions based on the expected differences between the two models. These are mainly based on the literature review and the extent each of the delivery models guides the audience's attention through the data story presented in the visualization. The following questions have been identified:

1. To what extent does the information visualization delivery model affect reporting/spotting outlier insights in the constructed narratives?

2. To what extent does the information visualization delivery model affect the depth of the constructed narratives?

3. To what extent does the information visualization delivery model impact the level of curiosity about the data story?

4. To what extent does the information visualization delivery model impact the level of difficulty of constructing narratives or telling a data story?

Some may assume that specific research strategies are assigned to specific research approaches. For example, case studies are always considered as qualitative research and experiments are always considered quantitative approach. This assumption is not always true. Also, the assumption that qualitative data (non-numeric data such as text, images, etc.) cannot be explored using statistical techniques and measures is not always valid. Some scholars provide strong arguments for separation between the data type or the approach to data collection (whether quantitative or qualitative) and the analysis performed on it to best answer the research question (Guest et al. 2011). The relevance of this discussion to the research strategy and methodology adopted in this part of the research in hand is that while we have a number of research questions that aim to measure the magnitude of the effect of the delivery models, the research objective addressed here is not to measure the usability of the delivery models. That is, measures such as time and accuracy as obtained from close-ended questions/tasks are not appropriate. Open-ended questioning as a data collection method will be important in the experiments. The questions to be used will be further detailed in section 3.2.2, but the important thing to note here is that using an open-ended questionnaire, which will obviously generate qualitative data, does not interfere with the ability to statistically describing the magnitude of the expected effects of the delivery models. This can be achieved by first performing a qualitative thematic content analysis on the data in order to categorise the data, and a quantification of occurrences of these categories will then be used to measure the size of the effect using estimation techniques (Guest et al. 2011, Cumming 2014). The mixed methods approach to dealing with and analysing the data generated from the experiments, with both quantitative and qualitative results, will provide an opportunity for an in-depth analysis of the user response. Thomas (2003) and Creswell (2009) argue that in a complex, modernised sociological environment, mixed-method research provides academics with a best-fit model for capturing multiple, interrelated streams of data.

3.2 The Experiments

In this section, two sets of experiments conducted with a total of 45 participants (22 females and 23 males) are described. These experiments explore and compare the effects of two different information visualization delivery models on how people construct narratives, comprehend and tell stories about data, using both within- and between-subject experimental designs. Using two different experimental designs for the same scientific question has been tested in studies in psychology and economics, and it can improve the reliability of results and enable testing for confounding factors such as the ordering effect (Charness et al. 2012). Using two different experimental designs will also enable discussion of the effect of the dataset used in each set of experiments. Due to the nature of the within-subject experimental design, whereby each participant is exposed to both delivery models, the dataset should be different in each model. While this may limit our ability to judge the effect of the models themselves, we should be able to compare the effects of the dataset and its underlying data story on the participants' constructed narratives by combining the results of this round of experiments with those of the between-subjects tests. In this regard, it is possible to use the same dataset for both delivery models, as each participant is exposed to one model.

Both sets of experiments examined the effect of the two delivery models: the narrative model and the software model. The same open-ended high-level questions were used to explore the differences in the users constructed narratives from these two models. Hence, the technique for the analysis of qualitative data used in both rounds was also the same; however, of course, the statistical techniques and measures appropriate for each experimental design were used. Due to the overlap between the two sets of experiments, those identical elements such as the experimental factors were summarised in one sub-section for both rounds. As the main difference between the two sets is the experimental procedure, these were described in two separate subsections, one for each round of experiments. The following sub-sections are as follows: Sub-section 3.2.1 describes the experimental factors, and sub-section 3.2.2 describes the questionnaire used. Finally, sub-section 3.2.3 details the experiment procedure for each set separately.

3.2.1 Experimental Factors

As stated, the aim of this study is to explore and compare the effects of two information visualization delivery models on people-constructed narratives: the narrative model and the software model (as defined in section 3.1).

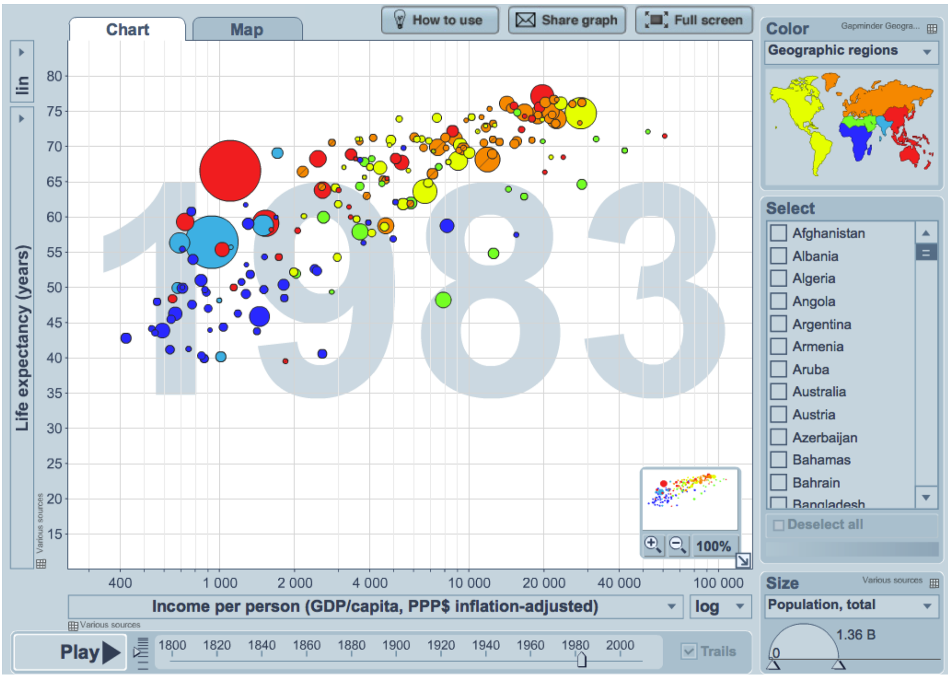


Figure 3.1: A screenshot of the interface of Gapminder World software.

The information visualization used to explore the difference between the effect of the two delivery models and to answer the questions above is Gapminder World software (*Gapminder* n.d.). This is an animated bubble chart with x- and y-axes that allow user-selected variables to be compared, and the bubbles represent countries. These bubbles are coloured by continent, and an animation and/or timeline slider can be used to show how the bubbles move over time (Figure 3.1 shows a screenshot of the interface of Gapminder World software). Within this context, the two delivery models of information visualization, the narrative and software models, were examined.

For the narrative model, two videos of Hans Rosling using Gapminder software to give a talk were chosen, including one video for each set of experiments (Rosling 2009, 2008). Rosling's model of storytelling with information visualization is one of the most famous in the field. He used Gapminder's animated bubble chart to give talks on several topics. The video chosen for the within-subject experiments concerns the HIV epidemic. The x-axis is income-per-person in USD, and y-axis is the percentage of adults infected by HIV (Rosling 2009). For the second set of experiments, between-subject, the video chosen concerns child mortality. The x-axis is the income-per-person in USD, and the y-axis is the child mortality rate (Rosling 2008).

For the software model, Gapminder World software was used (*Gapminder* n.d.). Using Gapminder, the participants interactively explored a dataset on child mortality (y-axis) and fertility rate (x-axis) in the within-subject experiments, and a dataset on child mortality (y-axis) and income-per-person in USD (x-axis) in the between-subject experiments.

3.2.2 Open-Ended Questionnaire and Subjective Feedback Questions

In both sets of experiments, participants were required to answer five questions after watching a video (narrative model) and/or equivalent questions after exploring data on Gapminder (software model). For each question, an estimation of the time required to answer the question was given. This helped control the overall experiment time and assisted participants in estimating the amount of emphasis and focus needed for each question.

As the aim of the experiments is not to measure the usability of the delivery models, measures such as time and accuracy would not be appropriate; instead, the experiments aim to understand and explore how and to what extent each model affects users in constructing narratives and telling data stories. Hence, open-ended questions are important. Moreover, open-ended questions in which participants tell data stories in writing help them to formulate their mental models about the story and to produce unitary narratives (North et al. 2011). This is in contrast to think-aloud techniques used in insight-based evaluation, which generate a series of insights gained in the order they were discovered (North 2006, North et al. 2011). An open-ended question was asked to trigger the re-telling of a story. Re-telling is a widely used task in education to assess comprehension (Fulkerson et al. 2009). The five open-ended questions on each delivery model are shown in Table 3.1. As can be seen from the table, both sets of questions (those on the narrative model and those on the software model) had exactly the same layout and approximate time needed for each question. The only difference was the precise wording of the questions. This was changed slightly to refer to the correspondence model.

At the end of the experiment (in both sets), participants were asked to answer two five-point Likert-scale questions on each model. These Likert-scale questions are shown in Table 3.2. The answers ranged from 'easy' to 'difficult' for the first question and from 'not at all' to 'very curious' for the second.

Table 3.1: Questions used for each delivery model in both sets of experiments

Narrative Model	**Software Model**
1. What was the video mostly about? (Approx. 1-2 min)	1. What was the data you explored in Gapminder mostly about? (Approx. 1-2 min)
2. Re-tell the story you gained from the video in as much detail as you can. Try to write a story that makes sense to someone who is not familiar with the story/topic. (Approx. 6-8 min)	2. Re-tell the story you gained from Gapminder in as much detail as you can. Try to write a story that makes sense to someone who is not familiar with the story/topic. (Approx. 6-8 min)
3. What did you learn that you did not already know? In other words, describe new information/knowledge you gained from the video. (Approx. 2-3 min)	3. What did you learn that you did not already know? In other words, describe the new information/knowledge you gained from Gapminder? (Approx. 2-3 min)
4. Did you learn something that contradicts what you already know about the topic? What is it? (Approx. 2-3 min)	4. Did you learn something that contradicts what you already know about the topic? What is it? (Approx. 2-3 min)
5. What do you think the speaker's purpose was in producing this video? (Approx. 2-3 min)	5. What do you think the purpose was in providing this data in Gapminder? (Approx. 2-3 min)

A copy of the full questionnaire as used in the within-subject experiment is attached in Appendix A.

Table 3.2: The Likert-scale questions on each delivery model in both sets of experiments

Narrative Model	**Software Model**
1. How easy or difficult did you find telling a story after watching the video?	1. How easy or difficult did you find telling a story after exploring the data in Gapminder?
2. How curious were you about the data/story in the video?	2. How curious were you about the data/story in Gapminder?

3.2.3 Participants and Experiments Procedures

3.2.3.1 Within-Subject Experiment

Thirteen students (9 females and 4 males) from a local university were recruited. The age of the participants ranged from 23 to 48. Three main selection criteria were identified: participants had not taken any data visualization course/module, they did not have advanced knowledge in information visualization, and they were not professional data analysts. In other words, this part of the research was aimed at educated but non-expert information visualization users. Participants came from varied academic backgrounds where some wanted to take a data visualization course but had not yet started one, and the others came from courses where data visualization was not covered. As it is difficult to recruit subjects from the general public, students who were likely to be motivated about the scope of the study but also did not have advanced knowledge about the topic have been chosen. The entire experiment was carried out in a single session for each participant. The total participation time for a single participant was about an hour.

Initially, each participant was briefed on the purpose of the experiment and experimental procedure and asked to sign the consent form, which was sent to them by email in advance. Next, a counterbalancing technique had been used to assign participants to two groups. Each group was shown the two delivery models in a different order to account for any ordering effect.

Participants were allowed to take notes while watching the video or exploring Gapminder if they wished, and they were given notepaper for this purpose. Group I watched a ten-minute video of Hans Rosling presenting data on the global HIV epidemic using an animated bubble chart (Rosling 2009). Then they answered Part I of the questionnaire, which included the five open-ended questions about the video (narrative model). Next, they were briefed about Gapminder and they interactively explored some data for 10 minutes. The dataset the participants explored using Gapminder in this set of experiments was about child mortality and fertility rates. Participants were asked not to change the indicators (x- and y-axes) when exploring the data in order to control the number of indicators they had to work with in both delivery models. After exploring the data, they answered Part II of the questionnaire, which included the five open-ended questions on the data they explored in Gapminder (see Table 3.1). Finally, after finishing the experiment, they were asked to answer the four Likert-scale questions (two on each delivery model) shown in Table 3.2.

3.2.3.2 Between-Subject Experiment

A between-subject design was used in this round. Thirty-two subjects (13 females and 19 males) ranging from 22 to 56 years old were assigned to two groups, with one group for each delivery model. Each group had 16 subjects balanced by gender. The main selection criteria for participants in this set of experiments were the same as in the within-subject experiments. The entire experiment was carried out in a single session for each participant. The total time for a single participant was about half an hour. Initially, each participant was briefed on the purpose of the experiment and the experimental procedure and was asked to sign the consent form.

Group I watched a ten-minute video of Hans Rosling presenting data on child mortality using an animated bubble chart (Rosling 2008). Then they answered the five open-ended questions on the video (narrative model) shown in Table 3.2. After the experiment, they were asked to answer the two Likert-scale questions on the narrative model shown in Table 3.2 and to comment on the overall experiment.

Group II were briefed about Gapminder and interactively explored a dataset on child mortality (software model), which was the same dataset used in the video in the first delivery model. Participants were asked not to change the indicators (x and y-axes) when exploring the data in order to control the number of indicators the participants had to work with in both groups/delivery models. Then they answered the five open-ended questions on the software model shown in Table 3.1. After the experiment, they were asked to answer the two Likert-scale questions on the software model shown in Table 3.2 and to comment on the overall experiment.

Data analysis and findings of both sets of experiments will be detailed in Chapter 5. The next chapter (Chapter 4) addresses the second part of the research in hand on transition types in narrative visualization and the refinement and application of McCloud's taxonomy of transition types in comics to narrative visualization.

Chapter 4

Research Methods II: Transitions in Narrative Visualization

This chapter describes the process of refining the taxonomy of panel-to-panel transitions in comics (McCloud 1994, 2006) and applying the refined framework on the VAST Challenge videos. It mainly relates to Objective 3 and Objective 4 of the study: *3. To propose and develop a framework of transition types used to tell data stories through information visualization based on theories in other storytelling mediums*, and: *4. To propose a visual language to characterise transitions in narrative visualization based on the proposed framework of transitions types*. It is divided into two main sections. Section 4.1 provides an overview of the research strategy adopted in general and the selection of VAST Challenge videos. Section 4.2 details the transitions framework proposed, how this framework is used to code transitions in the case of VAST Challenge videos, and describes the visual encoding of the transitions diagrams that summarises the contents and transitions in each individual data story/video.

4.1 An Overview of the Research Strategy and Data-Stories Selection

As stated above, the research in hand aims to explore and model transition types used in telling data stories with information visualization. A dataset of a number of data-driven stories is needed to address this objective, and to explore and analyse the transitions in the stories. This dataset should be suitable for the comparison and modelling of transition types. To accomplish this, a systematic review of the VAST Challenge videos dataset has been conducted.

The dataset chosen to explore transition types is the set of VAST Challenge videos from 2009 to 2011 (*Visual Analytics Benchmark Repository* n.d.). VAST Challenge is a series of tasks released annually as part of IEEE VIS conference on data visualization, where a large dataset is provided and a number of tasks specified. There are typically two or three mini-challenges or tasks and a grand challenge. Every year a different dataset is released with different task types and a compulsory video submission is required to address these tasks. These videos provide a valuable and suitable dataset because the videos of any specific year are based on the same dataset and address the same tasks, and they each have roughly the same duration and format, thereby providing suitable data for valid comparison of the storytelling transition types. This is also a new perspective on the issue of storytelling, as the previous work done on this area looked at samples of completely different datasets of information visualization addressing various types of data and tasks/questions (Hullman et al. 2013). Despite the fact that VAST Challenge videos are created to be presented to professionals or experts in data visualization as the main audience, they are still very communicative and based on scenarios that can easily be related to real life and are of interest to many audiences.

The proposed framework for characterising transitions in data stories in the VAST Challenge videos dataset was inspired by McCloud's (1994, 2006) taxonomy discussed in section 2.3 in the literature review chapter above, and shown in Figure 2.4. This taxonomy consists of six types of transitions: Moment-to-moment, Action-to-action, Subject-to-subject, Scene-to-scene, Aspect-to-aspect, and Non-Sequitur.

4.2 Coding and Representing Transitions in VAST Challenge Videos

This section details the coding process and the transitions framework proposed and used in coding VAST Challenge videos. It is divided into two sub-sections. First, sub-section 4.2.1 provides an overview of the steps of the coding process, and includes details about terms that represent some characteristics of data-stories and that will be needed to code transitions in VAST Challenge videos. These terms will also be used throughout the remaining chapters of this thesis. Next, sub-section 4.2.2 details the transitions framework proposed based on McCloud's (1994,2006) taxonomy of panel-to-panel transitions in comics after refining the definition of its components to reflect the nature of data-stories in information visualization. In order to present the proposed framework clearly and avoid any confusion, an informative definition or description of each of the five transition components/types is provided in a separate box, and more details, justifications, and discussion of the decisions made when coding transitions are provided underneath. Reading through these five boxes should provide a clear overview of the framework.

4.2.1 Coding Transitions

Before adopting the categories in McCloud's (1994, 2006) taxonomy and applying them to the VAST Challenge videos, this model should be refined and each type of transition should be defined in the context of data visualization. The coding steps are first summarised in this sub-section, and the categories of components and transitions proposed are summarised in the next sub-section (sub-section 4.2.2).

The coding process started with the task description of VAST Challenge. The purpose of going through the task descriptions before approaching the videos was to be familiar with the scenario and to identify an initial set of likely main components or elements of the story and how these components can be assigned a category (subjects, moments, etc.) based on the definitions of these categories (as in sub-section 4.2.2). Thus, one will have a consistent understanding of some initial story components among the video submissions for each challenge. The next step was to watch and transcribe the videos. All video submissions to the mini-challenges from 2009 to 2011 were transcribed and coded except those submissions with broken links/URLs or where the sound/audio is incomprehensible, or when the quality of the video is poor to the degree that transcription and coding were not possible. This left us with 70 videos in total, each lasting about four minutes. The transcription took into account various different details in addition to speech or verbal narration, such as periods of silence to show interactions with the visualization or to change the topic/focus, the use of explicit sub-topics (usually by using a title slide) to organise different parts of the submission, such as analysis stages (for example, Data Preprocessing, Visualization) or parts of the task/question being answered (for example, MC1.1, MC1.2), and some notes about visuals used (for example, changing of views or visualization techniques). Transcribing the videos in this way made it easier and more efficient to code transitions and story development/construction using mainly these annotated transcripts. This allowed us only to refer to the videos when more details were needed, or when we needed to revise or check a point. The semantics revealed by the oral narration that accompanies each video represent the progression of the data story in this type of narrative visualization (i.e., video presentations). The next step was to work through each video transcript and to identify the components of the transition types by colour coding them. As we went through each transcript, we segmented the story into smaller units. Each story unit contains a piece of information. A story unit is usually represented by one or more sentences, depending on the completion of the meaning that accompanies or follows a visual. In some cases, a sentence might be segmented into more than a single story unit. The story unit may also contain one or more transition components. Next, the transition type(s) between the story units and/or the groups of story units is identified based on which of the components has been changed or held constant. Finally, each story is visually represented by a transition diagram. The framework used in coding transition types and their components in the context of information visualization is detailed in the next sub-section; however, before moving to the framework and applying it to the VAST Challenge videos, several issues/details should be taken into account and covered in this sub-section. These issues relate to the nature and characteristics of storytelling in general, as well as to specific types of narrative visualization such as video presentations. These are also important when transferring/translating the taxonomy and its main blocks/definitions from comics to data visualization and in visually representing each story by a transition diagram.

First, based on the definition of a 'transition' provided in Chapter 2, a transition occurs between two states or story units. Therefore, coding and identifying transition types in any type of narrative requires the definition of what is considered a ***Story Unit*** and the identification of boundaries between different story units. In some domains, identifying story units is very straightforward. In the comics domain, by which the proposed framework is isnpired, and on which it is built, each individual panel is considered a story unit. Each panel consists of components that represent part of the categories/types of the transition between panels. For example, a panel showing a *subject* or character who does something (i.e., *action*), and shown from a specific angle or specific part of the room/place (i.e., *aspect*). The transition between this panel and the next depends on the changed component(s) in the next panel, and the transition is based on this change. For instance, if the character (*subject*) is changed, it will be a subject-to-subject transition, while if the same character is there but now doing something else (i.e., a different *action*), it will be an action-to-action transition. To be able to apply the same process/concept, some guidelines of what constitute a story unit when coding a VAST Challenge video are needed.

As seen in the literature review chapter above, previous work on transitions in narrative visualization focus mainly on slide-show presentations where an individual slide represents a single state or story unit (Hullman et al. 2013). Identifying story unit boundaries in video presentations that use different types of visualization to communicate a data story is not as straightforward as in slide-show presentations. Some previous work addresses the issue of automatically segmenting and identifying story unit boundaries in films/videos by developing algorithms to accomplish this task and to help in transferring a video into a storyboard or a number of shots/units (for example, Deardorff et al. (1994)). These techniques can be used with professionally produced videos, such as those in news stories, TV programs or films. However, these techniques and algorithms are not suitable to apply to VAST Challenge video submissions as these videos are not produced professionally using advanced production technologies and do not usually consist of many shots. Story units and the boundaries between them cannot be automatically detected using these algorithms. In addition, the main focus of this part of the research is to look at the progression of these stories depending on the meaning and pieces of information they delivered, as this is what contributes to the gradual/cumulative building of data stories. Hence, the task of segmenting a story into smaller units is a bit more challenging. A review of the literature reveals that people are quite consistent in segmenting goal-driven, everyday activities/events, even with little instructions (Hard et al. 2011). The literature also shows that 'breakpoints' that mark boundaries between each segment, unit, or task are distinctive (Hard et al. 2011). Hence, relative flexibility in judging or identifying story units is important and should not negatively affect the consistency of segmentation.

Another important point to take into account when coding and visually representing transition in each video is narrative hierarchy or what we can call ***Nested Transitions***. The literature shows that people construct hierarchical structure when attempting to segment events (Zacks et al. 2001). This hierarchy is a common characteristic between various types of narratives. It can be seen in comics and in TV shows where an episode consists of a number of shots and there is a transition between shots as well as between episodes. Similarly, it can be argued that this hierarchy is widely used in the sample of VAST Challenge videos and in most narrative visualization, as they are also goal-driven stories that aim to answer a particular question or deliver a particular message by going through a number of tasks or steps. A number of steps might be grouped together (on different levels) in order to achive a specific goal/sub-goal or to communicate a message. In VAST Challenge videos, this hierarchy can be seen, for example, in the use of sub-topics (usually represented by title slides in the video), and where a number of steps or story units are listed under each sub-topic.

Furthermore, it is not usually true that only one type of transition occurs at a time, especially in the case of narrative visualization. ***Multiple Transitions*** may occur between two story units. For example, if a story unit consists of a subject (a data point/variable), an action done on this subject (e.g. filtering or calculating something), and an aspect (a visual representation to look at this subject, e.g. a timeline or bar chart), and in the next story unit another subject is addressed and a different action is performed. In this case, we have two types of transitions between these two story units, a Subject-to-Subject and Action-to-Action transition. Finally, the transitions between story units do not only occur between juxtaposed/subsequent story units. In some cases we can see ***Delayed Transitions***.

In order to assess the consistency/reliability of coding transition types in the VAST Challenge videos, a detailed coding guidance has been produced by the main coder as a result of the iterative coding of videos and the refining of the framework. This coding guidance included detailed guidelines, expressions that are found to be commonly used to mark story unit boundaries as coded by the main coder, and some worked examples. This coding guidance is provided in Appendix D and it was used by a second coder to code a sample of the data. As can be seen from the description presented above, the coding process is relatively long and entails segmenting a data story until a transition diagram is created that summarises and represents the transition types used in the story. Hence, when assessing the reliability of coding, the coding process was divided into two main parts/tasks in order to be able to assess/identify the source of any inconsistency in the coding of transitions between coders. The first part, or coding task, is to identify the story units. For this part, the second coder segmented a corpus from three annotated data story transcripts into story units. The second part, or coding task, is to identify or code different components of the transition types in some of the already-segmented corpus of the data stories, judging the transition types between the story units or group of story units, and then visually representing this coding in a transition diagram. For this task, the second coder coded a corpus from six data stories. The percent agreement was calculated based on the agreement/disagreement regarding the breakpoints in the first task as well as the representation of all the components in any single story unit, the judgment of the transition type(s) between the story units and the hierarchy of the data story in the second task. The percentage of agreement between the two coders was generally good for both coding tasks (88.89% in the first task and 92.04% in the second task). Although the inter-coder agreement was good, the differences in the coding completed by each coder were discussed between them, and things that were missed by the second coder were identified (for example, skipping a line while creating the diagram, which resulted in a missing story unit in the diagram even though it was coded correctly in the transcript). Points that should clarify the coding instructions provided in the coding guidance were also agreed on. Specifically, two points were added to the guidance. The first point was that a finding always closes the scene in the transition diagram (see section 4.3 on the visual encoding and creation of transition diagrams). The second point was that sometimes the analyst may not explicitly say that he or she does something when he or she obviously did, as this affects the coding of Actions (see Actions and Action-to-Action transitions in section 4.2.2 below).

4.2.2 Framework of Transition Types in Narrative Visualization

This sub-section details the proposed framework of transition types in narrative visualization. This framework was proposed mainly by looking at the dataset of VAST Challenge videos. Therefore, the coding process and some details are only applicable to this type of narrative visualization (i.e. video presentations). However, it can be argued that the framework itself is still useful and applicable to other types of narrative visualization, e.g. slide-show presentations. More details on other contexts or areas into which the framework can be extended are discussed in the Discussion and Conclusion chapter (Chapter 7). As stated in the beginning of this chapter, an abstract/informative definition of each type/component of the transitions is provided in a box, while some more details/examples about the qualitative coding of transitions in VAST Challenge videos are provided underneath. Generally, the terms used to describe the components and types of panel-to-panel transitions by (McCloud 1994, 2006) have been maintained. However, the last type of transitions, *Non-Sequitur*, is not part of the proposed framework. This is because it is assumed that there is always a relationship between the information used in any data story even though this relationship does not necessarily have to be from one piece of information or story unit directly to the next one. It is also important to note that, when proposing these definitions, the main role of each transition type in its original domain (comics domain) was taken into consideration. By thinking of equivalent roles in the context of data stories through information visualization, the terms and categories of transition were redefined in the context of information visualization. A summary of these roles is provided at the end of this sub-section.

In the context of narrative data visualization, the transition terms and categories are redefined as follows:

Subject and Subject-to-Subject Transition:

A subject is an entity/actor that does something (an action) in the story or that presents the main relevant concepts of the scenario. This is similar to values or a variable in statistics. An example here might be an employee, or an IP address. A subject-to-subject transition occurs from a subject or type of subjects in one story unit to another subject or type of subjects in the next story unit (if the transition is not delayed).

When identifying subjects and subject-to-subject transitions in VAST Challenge videos, the following guidelines were applied:

- The noun/verb analysis may sometimes be helpful in identifying subjects, which are usually nouns or noun expressions, and actions, which are usually verbs (although neither every noun is a subject, nor every verb is an action).

- Going from whole to part or from a set to a sub-set is considered a subject-to-subject transition. For example, if story unit 1 involves 'seeing *all ID records*' and story unit 2 involves 'examine *16-source IDs only*', it is said that there is a subject-to-subject transition between these two story units.

- Some subjects may result from a process or an action that was done to another subject. This is similar to the previous point on going from whole to part, but the part in this case is not a direct sub-set from the whole, it is derived by performing an action/process on the whole. For example, story unit 1 might be 'I generated a report of ... of *all IPs*', and story unit 2 *"The destination IP which had the largest ... 37.170.100.75"*. In this case, there is a subject-to-subject transition between these two story units.

- In the video transcripts (the oral narration), nouns or noun expressions referring to the analysts are not considered as subjects (for example, *'I/we did ...'*). This is to avoid confusion when coming to the visual representation of these components. If these are considered as subjects and the scenario/domain principles are also subjects, it will not be appropriate to use the same descriptive term for them, nor to use the same visual encoding. Also, the nouns that refer to the person who did the analysis/interactions are always followed by some sort of actions (verbs), which are represented by a different type of component/transition. This will also makes it possible and easier to apply the framework to other types of narrative visualization where there is no oral narration provided, and to compare the construction and progression of data-stories between these different narrative genres.

- Examples of subjects include: events, activities, sequences, people, devices/computers, messages, cities, areas, a node(s) in a network of people etc., if these are what constitute the main concepts of the scenario. In many cases, the subjects are found to be countable. Additionally, the subjects might be referred to by using a pronoun.

- The whole dataset provided for a given VAST Challenge is not a subject on its own. For example, 'we loaded *the data* onto our system' or '*the challenge data* is downloaded'. However, something like *'the remaining data ...'* or a defined set from the data is considered a subject (for example, 'First, we looked at *the IP data*').

- When the same subject is found in two subsequent story units, there is obviously no subject-to-subject transition between these units. However, in some cases a subject may exist in two story units and at the same time we can have a subject-to-subject transition. This can be demonstrated in comic panels. If we have two comic panels and a subject exists in both of them but a new subject is introduced in the second panel (while the first is still there), then we have a subject-to-subject transition while retaining the first subject. In this case, the focus changes from the first subject to the second subject (the accompanying text may help to show this change, and in the same way the annotations or oral narration can show this change in narrative visualization).

Action and Action-to-Action Transition:

Here, we have differentiated between two types of actions. Type 1 relates to how the participant/analyst(s) interacted with the data/tool, for example, '*filter out* IP addresses', or decisions or tasks/processes made with the data, for example, 'we now *identify* who is the suspicious person'. Type 2 actions represent part of the data/scenario. For example, 'employee A *left* his office'. An action-to-action transition occurs from an action in one story unit to another in the next story unit. Sometimes the action-to-action transition is delayed by two or more story units, and sometimes the series of actions continues after reporting or revealing a piece of finding between them.

When identifying actions and action-to-action transitions in VAST Challenge videos, the following guidelines were applied:

- As stated above, noun and verb analysis of the annotated video transcripts is useful when identifying subjects and actions. Actions of type 1 are usually represented by verbs such as 'filter', 'compare', 'investigate', and 'select'. Actions of type 2 usually accompany a subject and sometimes also a moment (definition below); examples could be something like, '*The handyman sent a text message*' or '*Employee a closed the door and turn on the computer*'.

- In the language of data visualization, action type 1 is any visualization task done by interacting with the visualization or examining it (for more examples of these tasks, see Brehmer & Munzner (2013)).

- There is a difference between action type 2 and some action expressions used to describe a subject. For example, 'employee with ID40 *contacted* the employee with ID30', is an action type 2; however, 'now we will examine *IDs who are making more than 100 calls*' is not an action type 2, it is a subset or group of subjects.

- In many cases, after a finding is revealed, a new story unit with new action(s) begins. In this case, there is a delayed action-to-action transition from the story unit before the finding and the story unit after the finding.

- In some cases, the analyst does not explicitly say that he or she does something. For example, 'when the syndrome chart *is arranged* by number of deaths ...'. This should be coded in the same way as 'we *arranged* the syndrome chart by number of deaths ...' because even though the analyst does not say that he or she does something or interacts with the tool/data, he or she obviously did.

Scenes and Scene-to-Scene Transition:

> Scenes in different visual storytelling mediums in general (for example, comics and films) provide some kind of compression, conclusion or grouping based on an idea such as time or place. For example, in films, a scene is a complete unit that usually consists of a series of shots that form a coherent unit of the narration. Therefore, if the story in a given VAST Challenge video is divided explicitly into a number of sub-topics (usually represented by a title slide), each sub-topic is considered a scene. A scene-to-scene transition thus occurs between any two sub-topics, which in this case is a high-level transition (i.e., a series of different types of transitions may occur between smaller story units within each of those scenes). A group of story units that lead to a finding is also considered a scene, as is the finding or conclusion. A scene-to-scene transition in this case occurs between a number of steps or story units that are grouped based on a common idea/topic and the finding or conclusion drawn from this group.

When identifying scenes and scene-to-scene transitions in VAST Challenges video, the following guidelines were applied:

- As stated above, if the story in a given VAST Challenge video is explicitly divided into sub-topics, which are usually represented by a title slide, each sub-topic is considered to be a scene, and a scene-to-scene transition occurs between any two sub-topics.

- If the story in a given VAST Challenge video is not explicitly divided into sub-topics by title slides, but the narration shows a clear high-level segmentation of the story (for example, narration of the analysis and findings drawn using first tool A, then the analysis and findings drawn using tool B), each segment is also considered as a scene.

- After testing and revising the coding protocols several times and working back and forth between the coding and the corresponding visual representation, it was decided that the grouping of several story units will always be considered to be a *scene*. This practice simplifies the process of coding transitions and creating the transitions diagram (the visual encoding of the transitions in any single story). For example, a story may consist of a number of sub-topics, each of which communicates the analysis process and findings using a different visualization technique/application (i.e., an *aspect*). One may think about grouping the story units into aspects rather than scenes. However, this would complicate the process and sometimes makes it confusing when creating the transition diagram, especially if the grouping can be judged based on more than one type of transition component. For example, if these aspects/applications were done based on the subject/variable with which they deal with, it would be confusing when creating the transition diagram as to what colour the group as a whole should take: grey (based on aspects) or green (based on subjects). Grouping into scenes suggests a common idea or theme between a number of story units and simplifies presentation of scene-to-scene transitions between a group of story units and a finding scene in the transition diagram.

- Some expressions/terms are commonly used to report/narrate a finding, such as 'Hence' and 'concluded'. A list of these has been provided in Appendix D.

Aspect and Aspect-to-Aspect Transition:

An aspect-to-aspect transition occurs when the applications, views, techniques, or visual representations are changed from one story unit to another in order to look at the data from different angles (similar to having a 'wandering eye' around a place or object in comics (McCloud 1994). For example, this may occur between a map view and a timeline.

When identifying aspects and aspect-to-aspect transitions in VAST Challenge videos, the following guidelines were applied:

- Any description of a visual encoding/representation was colour-coded as an aspect (for example, '*red dots* represent poor countries'), but if we moved from something like *"industrialised countries represented by green bubbles"* in one story unit to *"developing countries represented by red bubbles"* in another story unit, this would not be considered an aspect-to-aspect transition if the two story units were both addressing the same graph. On the other hand, going from one graph to another graph, one view to another, or one visualization technique to another to investigate something else is considered an aspect-to-aspect transition. This explains why in some transition diagrams we might find two story units with an aspect component in each of them but no aspect-to-aspect transition between them.

- An aspect can be the result of an action on a different aspect (in other words, an action triggers the aspect-to-aspect transition.

- Sometimes a story (or part of it) starts with a general description of the visualization tool or the application used. This application may have two or more views and the first statement starts with a description of both views (hence, both views are from the same level and are components of the interface or application; therefore, the statement as a whole is considered to be one story unit). Then, when a specific step/investigation is carried out using one of these views, an aspect-to-aspect transition will occur from the visualization tool as a whole to one of its components/views. The same may occur between different components of a specific view. For example, if the interface has a chart view that consists of a number of charts, the transitions between these charts are considered aspect-to-aspect transitions even though they might be from the same type (e.g., bar charts) but involve addressing or looking at different parts of the screen or looking at something else (just like camera movement).

Moment and Moment-to-Moment Transition:

A moment is a point of time in the story/data scenario. For example, 'at 10am on 17th May, ...'. A moment-to-moment transition occurs between two story units if each story unit addresses a different point of time.

When identifying moments and moment-to-moment transitions in VAST Challenge videos, the following guidelines were applied:

- Expressions or use of time as an estimation of the analysts/participants' effort spent on a specific task is not considered as a moment. For example, '*we analysed the data for four hours*', or '*we built a visualization tool in six hours*'. This is particularly important when representing the components and transition types in each video visually in an abstract way, as in the transition diagrams (section 4.3 below), as moments should refer to the same thing to be visually encoded in the same way.

- Moments sometimes are very specific. For example, 'employee A was in the room from 1 p.m. to 3 p.m. on 15th May' or 'the first week of August'. In other cases, it might not be as explicit. In other cases it might not be as explicit, such as when it refers to a relative period in the data scenario. For example, 'we address/compared all events *before and after the truck accident happens'*.

This part of the present research proposed a framework/taxonomy in which the various components that compose a data story are related to one another so as to form and to be perceived as a unified whole, in the same way that comic panels are perceived as a unified whole. The components of story units and the various transitions between them, as defined above, work together to make the story unfold. In information visualization, the main components of a data story are: 1) the main principles of the data scenario (the subjects); 2) how the designer/viewer/user interacts with the data to move forward in the data story (actions Type I); 3) how the main principles of the scenario or the subjects interact and/or what they do (actions Type II); 4) what time frame that is covered or the time related to the phenomenon under investigation (moments); 5) the type of visual encoding/representation that is used during the course of the story or to tell the story, and how/when the designer/viewer/user explores the data story by navigating or changing between these visual encodings/representations (aspects); 6) the point at which the user applies analytical reasoning while being told the story, and how the story can be divided or segmented into high-level segments/sub-topics (scenes). Therefore, if a data story in information visualization is a collection of identifiable components or ingredients, it is important to identify how these components are related to each other. This is what the taxonomy of transition types in narrative visualization provides. It helps the designer/storyteller move the narrative forward while maintaining and maximising the continuity of the flow of narratives while connecting the different parts/components to form a coherent unified data story.

Additionally, it can be argued that in information visualization as in comics, each type of transition requires a different level of reader or audience involvement and a different amount of effort to fill in the gaps/missing pieces between the story units or the steps. Consequently, each type of transition plays a different role in the storytelling. For example, in describing a moment-moment transition Mccloud (1994, 2006) notes: 'it slows the action down' and that it 'requires little closure'. Therefore, it can be argued that, in the context of information visualization, a moment-moment transition, as defined above, may also require the audience to exert relatively little effort to fill in the missing pieces or the gap between two story units in which a moment-to-moment transition occurs. According to McCloud (1994, 2006), scene-to-scene transitions involve 'transitions across significant distances of time and/or space'; hence, this requires more effort from people who read comics. The same might well occur in the scene-to-scene transitions in the context of information visualization, as defined above. Reasoning is required to understand how a story unit or group of story units lead to a particular piece of finding; reasoning is also needed to work out the relationship between the high-level scenes or sub-topics in a data story in order to cumulatively build a unified narrative. In comics, subject-to-subject transitions show different perspectives/subjects within the same idea or scene. Similarly, in information visualization, subject-to-subject transitions occur between subjects that are related in terms of idea or by forming different categories or entities. Hence, this type of transition may encourage comparisons. Action-to-action transition helps move the narrative forward. The same also applies to action-to-action transitions in information visualization, as can be seen from the diagrams in Chapter 6. However, in information visualization, action-to-action transitions involve interacting with the data/visualization, and while the designer/storyteller can choose which interactions to select and include in his/her telling of the data story, the decision of which action/interaction to follow is more challenging when it is left to the user. Hence, as Segel & Heer (2010) note, a more author-driven approach might be needed to guide the user/audience through the desired set of interactions (actions). Aspect-to-aspect transitions, which in comics are usually described as 'having a wandering eye', can also be applied to aspect-to-aspect transitions in information visualization. Holding the time still and changing the way of looking at the data or the visual representation of the data, which may add some more factors, encourage contemplation and may open up a new direction or storyline. While McCloud's (1994, 2006) claims have been seen more explicitly in comics, there is scope for future study to empirically explore the degree to which different transitions encourage these particular characteristics in the context of information visualization. Finally, it is important to note that the occurrence of multiple transitions between any two story units may increase the amount of effort/involvement needed from the reader/audience of the data story; it may also require the designer to combine different ways/approaches in order to make the story unfold.

This section covered the coding process and the proposed framework of transition types in narrative visualization. The next section will summarise the visual representation and creation of the transitions diagrams as a result of this coding.

4.3 The Visual Encoding and Creation of Transition Diagrams

Going through each data story either in its original form (video presentation) or as an annotated transcript with highlighted transition components, is not an easy task nor an effective way to obtain an overview of the use of these transition components/types. Visually representing the transition components/types used in each data story would be beneficial for looking at the nature of storytelling in the VAST Challenges video samples.

While creating a visual representation of the transition types and components used in a single data story is not as complex as some design studies in information visualization in which the problem domain is more complicated and far from the background of the information visualization designer, the process of creating a visualization design that shows the structure, components, and transitions used in any data story involves roughly the same process/steps of any design study. The potential errors and validations are also applied. A number of papers provide guidance for the visualization design process (for example, Sedlmair et al. (2012) and Munzner (2009)). While some of the steps provided in prior papers cannot be applied to the task at hand, for example working with the domain experts, the main steps and requirements can still be applied. According to Munzner (2009), the first stage in the visualization design process is characterising the problem and defining the goal of the visualization design. The general/main goal here is creating a visual representation of the transition types in each data story. However, an accurate/adequate representation of the data story would not be possible if the transition types were the only components that were addressed. Hence, a visual representation in which the nestedness or hierarchical nature of storytelling is also maintained/shown is important. After characterising the problem and specifying the goal of the visualization design, the next step is to describe the data using 'the language of computer science' or 'the language of information visualization' (Munzner 2009). As stated above, in addition to taking the transition types into account, the visualization design must also show the structure and hierarchy of the data story and the components that are used (not just those associated with the transition types). After data description/abstraction, the next step is to choose the appropriate visual encoding/representation to the data at hand based on the perceptual and cognitive process in order to create an effective visualization design (Munzner 2009).

Before describing the visual representation of the components and transition types used in each VAST Challenge video (i.e. the transition diagram), it is important to note that the refining and proposing of the framework of transition types in narrative visualization that was summarised in the previous section, and the process of developing an appropriate visual representation, has been done simultaneously by working back and forth between them. This is because a refinement or a storytelling characteristic discovered while coding the videos should be reflected in the transition diagram, and any difficulty or ambiguity found in a transition diagram should be solved in the coding process and the framework as well as in the visual representation.

As stated above, understanding what data we are aiming to represent is an important step toward choosing the best visual encoding and creating a suitable transition diagram. The main aspects that a transition diagram is expected to show is the components and the transition types as summarised in the previous section. The transition types and their components (i.e., subject, action, etc.) are categorical data. Different colour hues should be used to represent these components and transitions types, avoiding combinations of colours indistinguishable to those with colour blindness, such as red and green (Ware 2004). Data can also be classified into entities and relationships (Ware 2004). In the case of the proposed framework, the story units and their components can be considered the entities, and the transitions between them are the relationships. Using visual encodings such as nodes/cells with lines between them can be useful in representing this kind of data. Additionally, the order or progression of story units as they were in the data story can be shown by their spatial position from left to right. Finally, the hierarchy, such as in high level scenes and groups of story units inside a scene, can be represented by using nested-boxes and using colour saturation to differentiate between different levels within each story (Ware 2004).

After several iterations and refinements, a design in which the transitions used to tell a data story in each VAST Challenge video (or in a single data story in general) are fused into a single diagram is proposed. This will be called a ***transition diagram***. Each type of the transition components is represented by a small colour-coded node/cell. The same colour codes used in identifying these components in the videos' annotated transcripts are also used in the diagram. The nodes representing components of one story unit are positioned vertically on top of each other. By looking at the transition(s) made, we may draw a line from the corresponding cell in the first story unit from which the transition was made to the cell in the destination story unit (this could be the next story unit or it could be delayed by one or more story units). The design also differentiates the finding scenes with bold borders. Also, no transition can be made from a finding or its contents. A finding always ends the scene (i.e. the story unit(s) after the finding will always be represented in a new scene or orange box). The visual encoding of each transition diagram is summarised in Figure 4.1. As can be seen, this visual representation can handle the various narrative visualization transitions issues that were discussed above. Rendered using this visual encoding, a completed transition diagram would look like the one in Figure 4.2. *yEd Graph Editor software* (n.d.) was used to create the hierarchical transition diagrams for the dataset of VAST Challenge videos. Looking at the transition diagrams representing the VAST Challenges videos enables comparisons and classifications of some of the data stories that share a specific style(s) or feature(s).

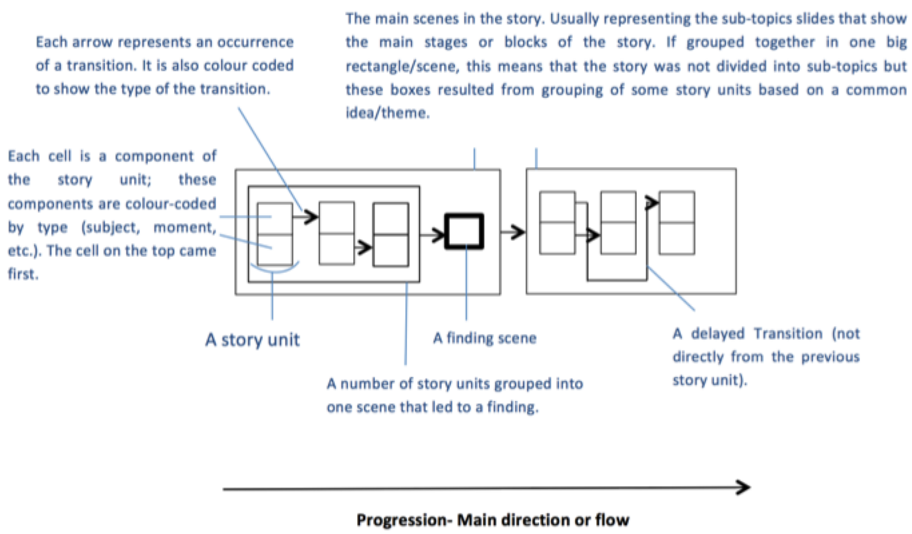


Figure 4.1: The visual encoding of the story components and transition types in each VAST Challenge video.

This chapter provides a taxonomy/framework of transition types in nar-

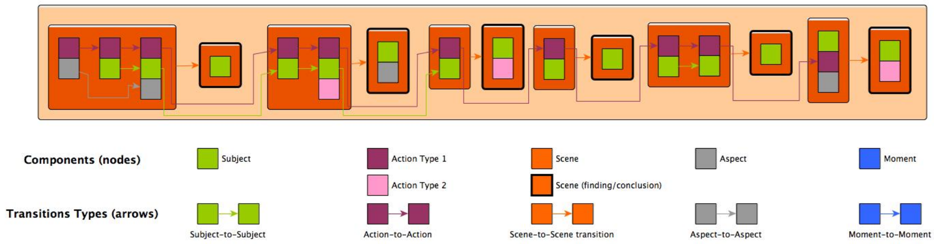


Figure 4.2: A completed transition diagram summarising the contents of the data story and transition types used in a video submission to VAST Challenge. (Submission: VAST 2009-MC2-Palantir Technologies)

rative visualization, and a visual design that summarises the components and transition types used in the VAST Challenge videos. The next two chapters are the analysis and results chapters. First, Chapter 5 summarises the mixed methods of analysis of the experiments conducted on the effect of information visualization delivery on narrative construction and development. Next, Chapter 6 summarises a number of styles of storytelling in VAST Challenge videos by looking at the transition diagrams created for these videos.

Chapter 5

The Effect of Information Visualization Delivery on Narrative Construction and Development

The preceding two chapters described the methodology used in the present research. Chapters 5 and 6 present the data analysis and findings. Chapter 5 analyses the the results of the experiments described in Chapter 3 to compare the effects of direct narratives and the interactive exploration of data stories on narrative construction (objective 2). This chapter is divided into five main sections. First, section 5.1 provides an overview of the analytical process. Then, sections 5.2 to 5.5 report the analysis, themes, and findings from every questionnaire item used in both sets of experiments. Finally, section 5.6 summarises and collates the findings.

5.1 Analysis of Qualitative Data

In this section, the process of analysis and coding of the qualitative data is described. Essentially, a mix of a priori and open coding was used when examining and analysing the data. As Q2 (re-telling the story) in the questionnaire was expected to generate the most insights; the researcher had identified a priori themes to code the answers, with some preliminary codes under these themes. The themes were mainly generated from the definition and explanation of stories, narrative, and storytelling through information visualization in the first two chapters of the thesis. To be specific, the explanation was that stories consist of some contents (such as pieces of information, insights, and so on) that are accumulated and structured in a particular way to form the overall narrative structure. Accordingly, two themes were identified to code the participants' written narratives. The first theme insight-type was used to assess the categories and types of the pieces of information in the written narrarive. The second theme, narrative structure, was deployed to examine the accumulation and structure of these insights in the narrative. Open coding was used with the other questions, such as Q1 and Q5. As the answers to Q2 were the longest and least straightforward, it was necessary to test the reliability of coding these answers. Therefore, the main coder iteratively coded the data until all codes were generated and defined. When all the themes and categories were identified and most of the data coded, a coding guidebook was created. Thereafter, a sample of ten participants' stories was randomly chosen and distributed, along with the coding guidance, to two other coders, who independently coded them. The percentage of agreement among all coders was generally good, ranging from 81.48% to 100%. The main coder then used the codebook to code the rest of the data. The coding guidance and the sheet used to record the coding results are included in Appendix C.

Before moving on to the analysis and findings, it is important to note that the analysis and coding of the qualitative data detailed in the subsequent sections is applicable to both within-subject and between-subject experiments as the open-ended questions used were the same in both sets of experiments.

However, suitable effect size estimators to the experimental design in each round were used to answer the research questions. Specifically, Phi or Cramer's V in the between-subject experiments, and the odds ratio or r for the within-subject experiments were used. The research questions were addressed first by analysing and coding the qualitative data and then by quantifying the codes generated from this analysis to examine differences and patterns in more detail.

5.2 Q1: Reported Impression

5.2.1 Analysis of Qualitative Data

A completely data-driven coding approach was followed when analysing participants' answers to this question. This is because there were no previous expectations, nor specific a priori codes, about the answers to this question. It was originally set to introduce the story and to gradually direct the participants to the more important and longer question that followed. Upon looking closely at the variety of the participants' answers to this question, both within and between the narrative models, a theme began to arise. The emerging theme here was the participants' impressions about the topic and the main factors for comparison in the story. Answers to this question were coded into four categories: 1) Geographical, 2) Temporal, 3) Both, Geographical and Temporal, and 4) Neither. Before we define these categories in greater detail and provide quotes from the participants' answers, it is important to note that these were not topics or titles provided explicitly to the respondents, nor themes suggested by asking the users to focus on one visualization design or view over the others, especially with the second delivery model (the software model).

- **Geographical**: some participants only reported spatial change as the main topic of the story. Certain terms were commonly used in this category such as *in different countries and globally*, and sometimes specific areas were named. The following are some examples of participants' answers that were coded as Geographical:

"*HIV epidemic in different countries.*"

"*Explaining the HIV/AIDS contraction globally but attempting to separate different locations have different reasons for infection rate.*"

"*Child mortality/fertility rates in Spain, Denmark, and Chile mainly.*"

"*Distribution of Aids Victims by Geography.*"

- **Temporal**: if the participant only reported change over time as the main topic of the story. The following is an example of an answer that was coded as Temporal:

  "*How child mortality changed over time.*"

- **Both Geographical and Temporal**: some participants reported both patterns in their answers to this question.

  "*About child mortality related to the women's fertility rate per nation on a chronological scale.*"

  "*Comparing countries' fertility and child mortality rates, over time.*"

  "*It was about fertility rates and child mortality rates in different countries over time.*"

"Total fertility, children per woman across the world in the last 200 years."

- The fourth category/code was **Neither.** This code was assigned to answers that could not be described as Geographical, Temporal, or both.

  "Number of children with respect to child mortality."

  "The video was an attempt to prove, via data visualization, that achieving the UN millennium goals is feasible."

5.2.2 Quantitative Summary of Findings

The results point to a notable difference between the two delivery models in terms of how the participants reported the main pattern or factor of comparison in each model. From the frequencies shown in Figure 5.1 and 5.2, it can be seen that more participants emphasised the *Geographical* pattern after watching Rosling's video (narrative model) in both sets of experiments. In contrast, after exploring the Gapminder data (software model), more participants emphasised the *Geographical* and *Temporal* patterns.

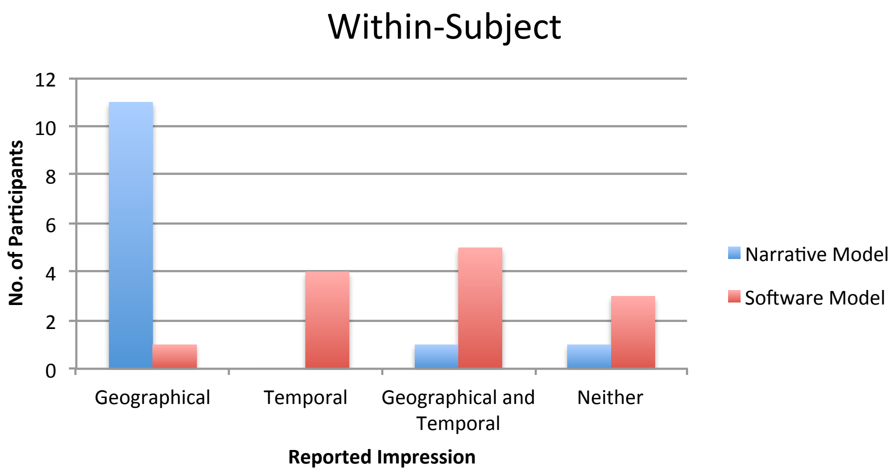


Figure 5.1: Reported impression of geographical and temporal patterns (within-subject experiment)

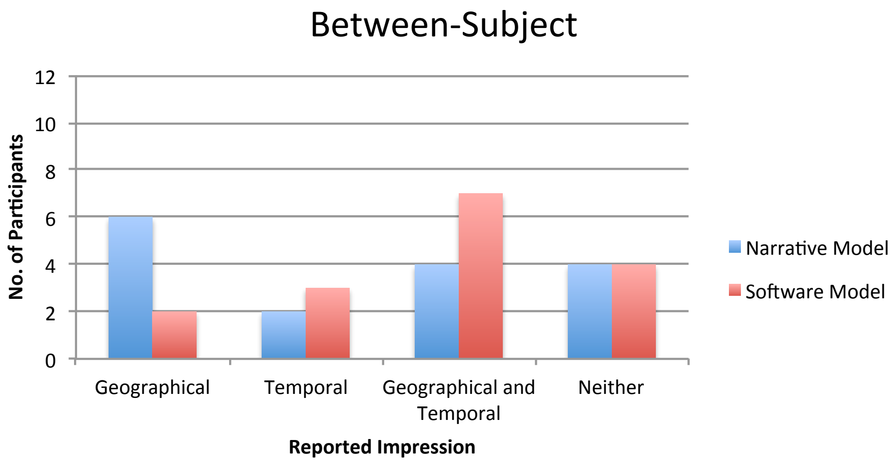


Figure 5.2: Reported impression of geographical and temporal patterns (between-subject experiment)

5.3 Q2: Re-telling the Data Story

Participants were encouraged to spend the most time answering the main question, '*Q2: Re-tell the story you gained from the video/Gapminder in as much detail as you can. Try to write a story that makes sense to someone who is not familiar with the story/topic*'. In the analysis of the answers to this question, participants' stories were coded based on two themes: *Insight Type* and *Narrative Structure*. These two themes helped answer the research questions identified at the beginning of the study:

1. To what extent does the information visualization delivery model affect reporting/spotting outlier insights in the constructed narratives?

2. To what extent does the information visualization delivery model affect the depth of the constructed narratives?

5.3.1 Insight Type

5.3.1.1 Analysis of Qualitative Data

As stated in section 5.1, we planned to code the contents of the participants' written narratives and their accumulation into a narrative structure in order to analyse Q2. Accordingly, the insight type and narrative structure themes were used. However, while some a priori codes were initially defined as per expectations, the exact codes and definitions were developed more explicitly and to a greater degree of clarity after data collection.

Researchers in various domains have defined insight, its characteristics, and how to arrive at a particular insight in different ways. One of the most well known historical models is Wallas's (1926) model of the creative process, which attempted to explain the experience of illumination and insight. Wallas (1926) identified four stages of developing or arriving at an insight or solution to a problem: preparation, incubation, illumination, and verification. In the field of cognitive psychology, Klein (2014) defined insight as an 'unexpected shift in the way we understand things'. This unpredictability contradicts Wallas's process, which starts with preliminary work and preparation and then offers success through verification. However, both models, although originating in different domains, could easily be related to the process of gaining an insight in the context of information visualization. This could either be a result of gathering requirements and specifying a problem, through working with and investigating various assumptions, to arriving at an insight, or experiencing the 'Aha moment' as an unexpected flash of insight. Klein (2014) also stated that effective insights have an impact on behaviours, feelings, and/or goals as a result of this new understanding. He also identified three 'insight-hunting techniques': look for connections, look for contradictions, and look for creative desperation. This is also related to the way in which people approach information visualization when they make connections between different factors and verify any inconsistencies between the information they have uncovered and their background knowledge about the topic or problem. It can also be argued, as Klein (2014) explained, that good insight has an impact and may alter goals or behaviours; insights generated as a result of exploration of certain data, using information visualization, may also change the consequent insights, the path of analysing and exploring the data interactively, and the overall goal of the user.

In the field of information visualization and visual analytics, it is usually said that the purpose of visualization is to gain insight. However, the definition of what constitutes an insight is still a hot topic of debate, with some recent studies zeroing on this topic/term (Saraiya et al. 2005, North 2006, Yi et al. 2008, Chang et al. 2009). North (2006) described an insight as: 'complex, deep, qualitative, unexpected, and relevant'. Meanwhile, Saraiya et al. (2005) defined insight as 'an individual observation about the data by the participant, a unit of discovery'. Similarly, Chang et al. (2009) stated that one possible definition of insight is 'a broader term to mean an advance in knowledge or piece of information'. Yi et al. (2008) stated that although they agreed with North (2006) that insight is complex and deep, the definition provided by Saraiya et al. (2005) was more appropriate in the context of evaluation studies and where some quantifiable measures are required. Accordingly, this definition was adopted in the present research when analysing the participants' answers/narratives based on the insight-type theme.

As stated above, the analysis of the participants' stories based on this theme was inspired by the insight-based evaluation (North 2006, North et al. 2011, North 2005, Saraiya et al. 2005). The insight types reported by each participant were identified but their numbers were not quantified, as it is difficult to count insight occurrences in written stories where a sentence may contain more than one insight type. For example, '*the Asian countries have the highest child mortality rate while the European countries have the lowest child mortality rate*', can be characterised as a Tradeoff insight comparing the maximum and minimum, and at the same time it can be considered as a Grouping insight, as it groups Asian countries together and European countries together.

The following insight types that emerged from the data, were identified from the literature, and were used to code the data based on the *Insight Type* theme:

- **General Pattern**: the general trend or pattern of most countries. For example, the general rise or decline in an aspect or a general relationship between two factors.

  "Child mortality decreases while fewer children are born."

  "The overall trend is that child mortality falls as income increases over the years."

  "Overall, over the past 60 years, the world has seen a general increase in income per person and this has correlated/been associated with a decrease in child mortality rates."

- **Detailed Pattern**: description of details on specific points of time or instances in general patterns. For example, the general pattern is a decline in child mortality while the detailed pattern is the average mortality rate in different periods of time to show how this happened.

  "By the time the gapminder reaches 2010 China and India have fallen in

the mortality rate and the income has increased, Sub Africa have moved up in the mortality rate and the income has fallen per person."

- **Outlier**: maximum, minimum or anything outside the general pattern; in other words, an exception.

  "In the last 50 years, child mortality has dropped and GDP/capita has increased. The main exception to this is Sub-Saharan Africa."

- **Trade-Off**: a combination of minimum and maximum or making comparisons between most and least in terms of one or more specific factors.

  "In general, child mortality rates have fallen over time in all geographical regions. However, it is at its highest in African countries, and at its lowest in European countries."

- **Grouping**: to group different things in one category based on one or more specific criteria. In other words, to define a subset or category of data. For example, western and eastern countries have different figures, or oil-rich and western countries have lower child mortalities, or something about developing/industrial countries as a group.

  "Industrial countries have decreased their mortality rate faster than developing countries."

- **Anomalies**: to identify data errors if any are present, such as missing countries at specific periods.

"Only data for a few central European and Asian countries was available initially."

"A lot of countries only have numbers from 1960."

5.3.1.2 Quantitative Summary of Findings

A summary of the insight types reported in the participants' stories constructed with each delivery model for both sets of experiments is shown in Figures 5.3 and 5.4. Looking at the existence of outlier insights in the participants' stories, we can answer the first research question identified prior to the experiment and measure the corresponding effect size:

1. To what extent does the information visualization delivery model affect reporting/spotting outlier insights in the constructed narratives?

It is noted that in the between-subject experiments where the dataset used in each delivery model was the same, seven participants out of the sixteen who used the software model to construct data stories reported an outlier insight in their stories, while none of the group who used the narrative model reported any outliers. The size of the effect was large in the between-subject experiments (Phi=0.53, with .95 CI of [0.0165, 0.8585]). However, the effect size was small in the within-subject experiments (Odds Ratio= 1, with .95 CI of [0.2018, 4.9547]). The wide confidence interval in the within-subject experiments indicates the small sample size used in that round of experiments.

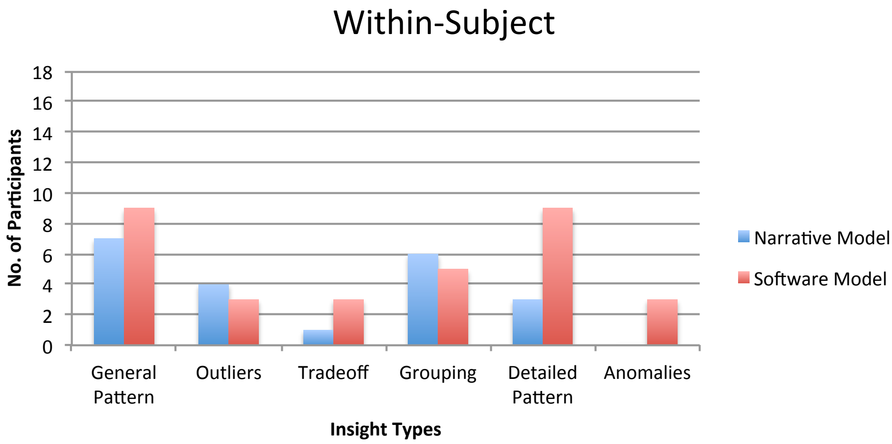


Figure 5.3: The number of participants (out of 13) who reported each insight type in their stories for each delivery model (within-subject experiment)

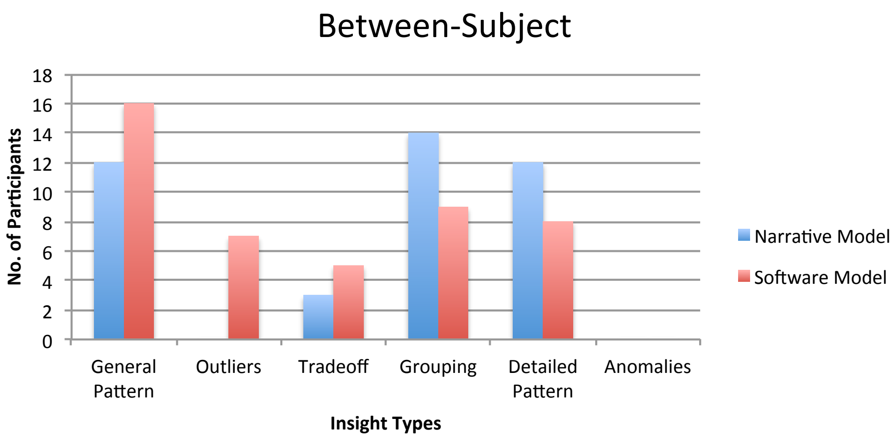


Figure 5.4: The number of participants who reported each insight type in their stories for each delivery model (between-subject experiment)

5.3.2 Narrative Structure

5.3.2.1 Analysis of Qualitative Data

The second theme used to analyse Q2 answers was the narrative structure. The structure is simply how the story progressed, i.e., the order of sentences or information blocks in the story. A range of educational literature on assessing learning progress and comprehension through re-telling as well as literature on written news stories were consulted (Fulkerson et al. 2009, Grunwald 2005, Spiegel 1981) to identify common narrative structures and develop appropriate codes for them. In addition, some narrative structures became obvious when reading the participants' stories. In the analysis of the participants' stories based on the narrative structure theme, up to two structures for each story were recorded. All stories were assigned a *main narrative structure* that was the most appropriate to explain the narrative flow in the story. Some also had a *sub-structure* that was the structure used within the main story theme. For example, the main structure might have been a chronological one, but within this chronology, it may have demonstrated a cause-and-effect over a specific period. Some of the stories written by the participants had a clear sub-structure, while others had only a main structure. Each of these main and sub-structures was assigned one of the following five types:

- **General-to-Specific**: starting from the general trend to more specific instances, details, outliers, maximum, minimum, and more specific insights and relationships.

  > *The Gapminder data visualization tells me that both child mortality and the fertility rate (births per woman) have largely gone down in most of the world since 1800. However, I also infer that availability of this data has greatly increased since 1800. Huge regional disparities currently exist in both the fertility rate and child mortality- for instance, compare African countries with European countries. African countries - Congo, Sierra Leone and several others in particular- have much higher child mortality and birth rates than European countries, where both have fallen considerably over time. I was surprised to discover that America has a lower birth rate (2.1) and higher mortality rate than many other places. Israel for instance has a higher birth rate (3.1) and lower mortality rate.*

- **Specific-to-General**: starting from outliers, maximum, minimum, etc.

to general trends or the big picture.

> *The speaker shows a graph with a child mortality ratio in terms of death v/s income. He makes his arguments by showing us the relationship for different countries. He used Norway as an example, as it is a successful country. He compares the history of Norway to three different countries: Bangladesh, Egypt, and Brazil. He showed us that for a span of time, each of the countries had a faster reduction than Norway. Afterwards, he showed us a graph of countries with no reduction and took Tanzania as an example. Tanzania has no reduction from 1990-1998 and then witnessed a reduction from 1998-2005 because the government provided and invested in new facilities. As a conclusion, he has provided us with a predicted global goal from 2005-2015 and how we can improve the reduction if we invest and improve the facilities worldwide.*

This is an example of an occasion where the Specific-to-General structure is being used. Although this participant started with a statement about the graph (Gapminder software used by Hans Rosling) and what it shows, they only stated that a graph was used to show the relationship between certain factors. Therefore, this statement did not deliver a general pattern or idea and the subsequent information in the narratives was about specific cases in particular countries, with a focus on Norway. It is evident from the flow of narratives that this emphasis on Norway and a number of other countires was a result of the participant being encouraged by seeing Hans Rosling focusing on those areas in his video. Moreover, as could be seen from this participant's answer to Q3 in the questionnaire regarding the new information she learnt, the answer was also about specific countries and not the overall idea/conclusion.

- **Chronological**: starting from past to present, present to past or using time points/intervals to control story progression.

  Over the last 200 years, child mortality overall in the world has reduced dramatically. Only 100 years ago, mortality was about 30% for babies 0-5 years old. Our days is about 3%. Only 40 years ago, on average there were about 6 babies born per woman. This was reduced to about 2 children over a 30 years period from 1970 till 2000, and for the last 10 years remained at this level.

  This is an example of where the temporal flow of the story is being strongly emphasised by this particular user. It was written after exploring the data on Gapminder (i.e. Software model), and the user actually felt that it was evident that temporal flow was encouraged by providing a timeline and animation to explore the data interactively. The user also commented that she thought that, as she does not have a specific idea or problem, using time to structure both her exploration and written narrative was the best way to do it.

- **Problem-Solution**: emphasising the problem and suggesting solutions, whether from external information used by the narrator in the first delivery model or from personal knowledge about the topic in either the first or the second model.

  A study has shown that from 1959 until now, there has been as impact of countries' economy on child mortality whereas successful industrial countries have less child mortality than *countries that are not as successful. But if resources such as beds, vaccines, and mother care are provided, then poor countries could save millions of children every year globally. An example of this was seen where these aids were provided to Tanzania where the country immediately start reducing child mortality numbers.*

- **Cause-and-Effect**: describing a figure, pattern or insight and providing the cause(s) of this insight. This structure is beyond the simple correlation between two factors (x and y-axes) and involves richer explanations of causes. For example, this could include figures in years when there was a war or a special event in a country or an explanation of situations in Africa or a specific country that caused some patterns.

  There is a higher proportion of aids victims in Africa than in any other areas of the world as a % of population and overll. It has also stabilised at a higher level; however the distribution within Africa is not even. It is not related particularly to war or poverty, but most significantly with death rates and frequency of sex, particularly multiple partners within one month, also age differential between partners is significant. So sexual behaviour is a key factor in countries without good medical systems.

The example above is a story written by a participant after watching the video in the narrative delivery model. In the video, Hans Rosling discussed the problem and its causes, as well as some solutions, looking at various examples around the world. This participant wrote about the causes of the AIDS problem in Africa, which was the most prominent case discussed in the video. It could be argued that this structure was also encouraged by observing Rosling doing similar things in his presentation and providing interpretations beyond the simple correlations seen in the visualization software employed.

The following points have also been taken into account when coding the data based on the narrative structure theme:

First, in both the General-to-Specific and Specific-to-General structures, participants may have written a statement to conclude the overall message for delivering this kind of information. This does not change the main narrative structure of the story.

Second, in the Chronological structure, participants may start with a general statement describing the aim of the visualization or the general pattern in the story as an introduction. However, the remaining details are told as change over time. Also, in coding a story as a chronological structure, differentiation was not made between chronology, reverse chronology, and picking up different points of time to structure and control story progression.

5.3.2.2 Quantitative Summary of Findings

The most common narrative structure in the participants' stories in both delivery models and in both sets of experiments was the *General-to-Specific* structure. Figures 5.5 and 5.6 summarise the main narrative structures used in the participants' stories in both rounds.

Furthermore, it can be argued that both the Problem-Solution and Cause-and-Effect structures provide more depth to the data story than other structures. They both involve going beyond simple correlation between x and y-axes either to explain causes for specific patterns or to discuss the problem and suggest solutions. For this reason, the second research question identified prior to the experiment can be answered by quantifying the qualitative codes generated as a result of coding the participants' stories based on the narrative structure theme. The question was previously worded as follows:

2. To what extent does the information visualization delivery model affect the depth of the constructed narratives?

To clarify, this question can be re-worded as follows:

2. What is the magnitude of the association between the information visualization delivery model and the use of Problem-Solution and/or Cause-and-Effect narrative structures (either as main or sub-structures)?

In the within-subject experiments, this association was moderately large (Odds Ratio= 4, with .95 CI of [0.4471, 35.789]. The wide confidence interval here is due to the small sample size used in this round of experiments. However, in the between-subject experiments it is found that 13 out of the 16 participants who used the narrative model used the Problem-Solution and/or Cause-and-Effect narrative structures in their stories, while only three out of the 16 who used the software model used these structures in their stories. The association between the delivery model and the existence of these two structures is considered to be very strong (Phi= 0.63, with .95 CI of [0.3249, 0.9251]).

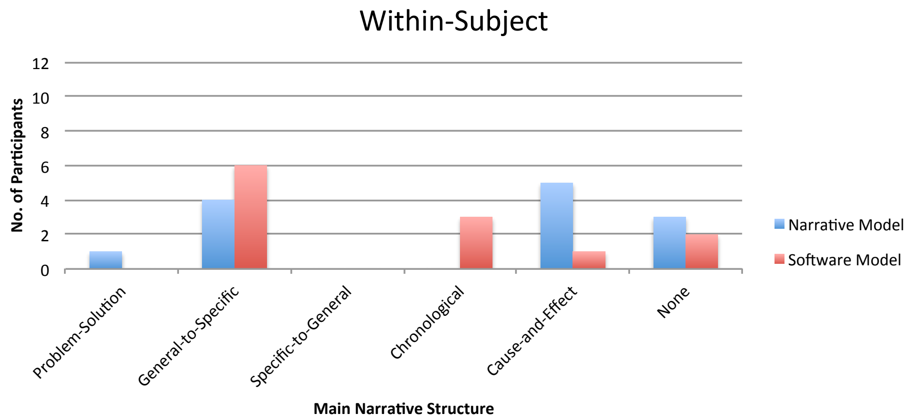


Figure 5.5: Main narrative structures used by participants for each delivery model (within-subject experiment)

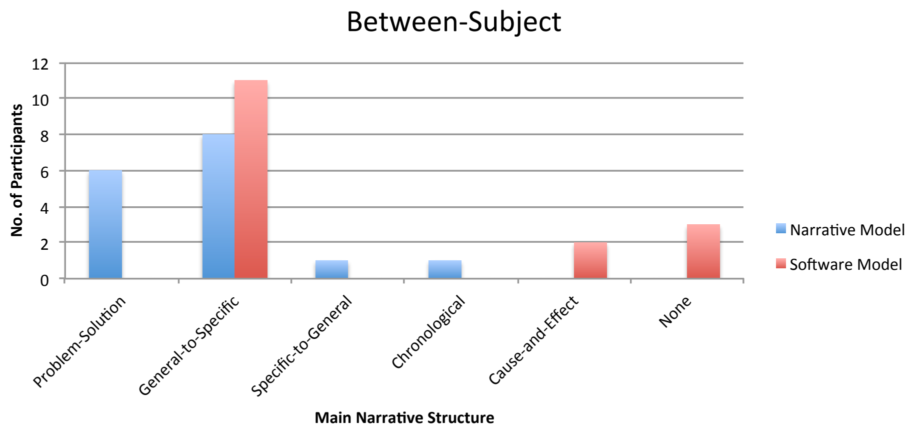


Figure 5.6: Main narrative structures used by participants for each delivery model (between-subject experiment)

5.4 Participant Learning (Q3 and Q4) and Narrative Purpose (Q5)

A content analysis was performed for Q3 and Q4 in terms of what insight types were reported by participants as new or contradictory information, and recurring codes were sought out. Klein (2014) identified three things to look for when hunting for insights: connections, contradictions, and creative desperation. As the terms suggest, connections occur when new information or knowledge is connected to or added to what is already possessed by the user; contradictions eventuate when new information or knowledge does not agree or align with existing learning, and creative desperation 'requires finding a way out of a trap that seems unescapable' (Klein 2014). The answers to these two questions did not show any pattern in the results in terms of the insight types reported by the participants, as they depended heavily on the individual background of each topic. However, some qualitative observations have been made, which will be summarised in section 5.6 below.

Although no specific themes were used in Q5, some observations were made based on the participants' answers to this question. For both delivery models and in both rounds of experiments, most answers to the fifth question about the purpose of providing the data story in each model included raising awareness about the topic, urgency in solving the problem and providing the data in a more compelling way that was easier to understand than a tabular format. That is, the purposes identified were mostly relevant to the data and visual representations. However, something different was found in the answers provided by two participants in the first round of experiments and two participants in the second round of experiments about the speaker's purpose in producing the video in the narrative model. These participants thought that a personal or political purpose was present in the story. For example, one of them said that the purpose was '*political posturing, perhaps to seek funding*'.

5.5 Participants' Subjective Feedback

Participants' answers to the subjective feedback questions in both sets of experiments are summarised in Figures 5.7 to 5.10. In the within-subject experiments, it was found that the delivery model has no effect on the level of curiosity about the data/story (the effect size r= 0.04, indicating a small effect size using Cohen (1988) criteria of 0.1=small effect, 0.3= medium effect, 0.5= large effect), while it has a moderate effect on the difficulty of telling a

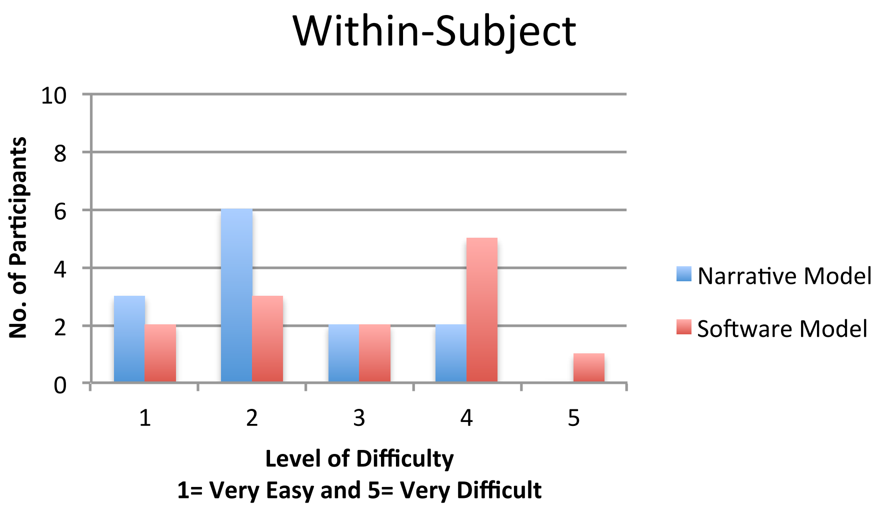


Figure 5.7: Participants' subjective feedback on the level of difficulty of telling a story for each delivery model (within-subject)

story (r= 0.36). Similarly, in the between-subject experiments, there was a little effect of the two delivery models on the curiosity about the data/story and the difficulty of telling a story (Cramer's V effect size= 0.26 and 0.23 respectively).

5.6 Summary of the Differences Between the Two Delivery Models

Table 5.1 below displays some excerpts from participants' answers to the questionnaire. These answers illustrate some of the main differences noticed between the two delivery models (the narrative and software models), as discussed in this chapter, including some qualitative differences that are not quantitatively verifiable due to the sample size.

For Q1, it may be noted that without the direct use of a map as the main visual representation, the users tended to focus on other details, such as the chronology and change over time, rather than treating the narrative as a geographical comparison between locations. However, the presence of a narrator

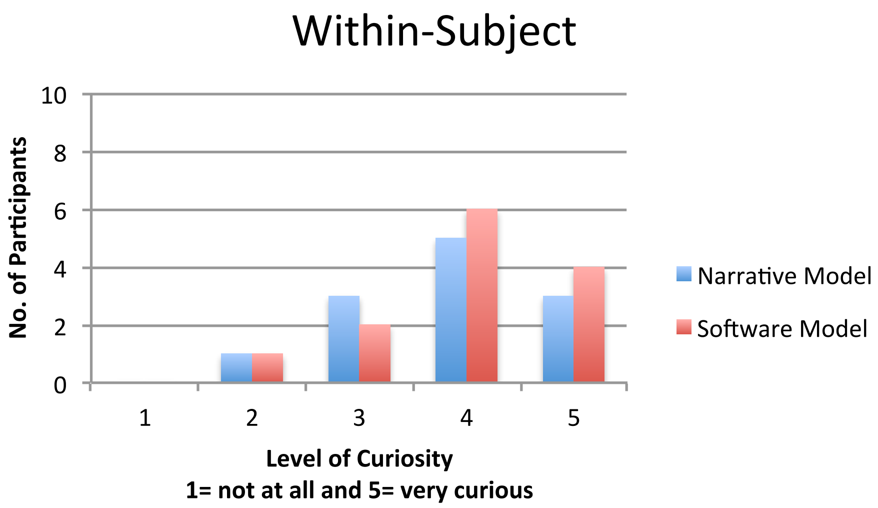


Figure 5.8: Participants' subjective feedback on the level of curiosity about the data/story for each delivery model (within-subject)

altered this impression and the construction of the written narratives. Therefore, it could be argued that the designer/narrator could divert the audience's attention to geography, even though a map was not used or presented to them. Table 5.1 shows excerpts from participants' answers to this question to demonstrate how they differed.

As stated in the previous sections of this chapter, Q2 was the main question in the study and the participants were encouraged to spend the greatest amount of time answering it. From the excerpts in Table 5.1, it can be seen that the narratives written after watching the videos (narrative model) discussed and interpreted, in greater detail, the simple correlations and statistics/facts shown on the visual representation. They provided more detailed answers or solutions to the problem (such as the use of bed nets and vaccines to help reduce HIV rates). These solutions were not represented on the screen/visualization, but rather were provided in the oral narration by Hans Rosling. However, an excerpt from an answer to this question, after exploring the data interactively using Gapminder software, shows that the participant explained the data as provided in the visualization. In the example provided,

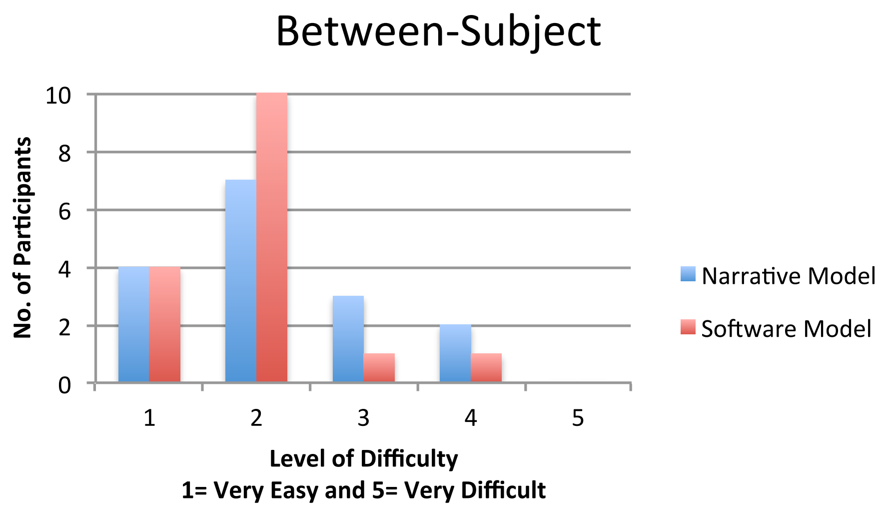


Figure 5.9: Participants' subjective feedback on the level of difficulty of telling a story for each delivery model (between-subject experiment)

the chronology or timeline of the narratives was emphasised more clearly. No connections were made between the information represented in the visualization and external knowledge. However, although it became apparent that this was the case with most of the participants' written narratives upon exploration of the data using Gapminder, very few of them made connections between the facts shown in the visualization and outside knowledge, either about the topic itself or as part of their reasoning; for example, relating or making a connection between certain figures and the period of the World Wars, or between specific figures in a given country and particular historical events, such as a genocide. This shows how the formation of connections, as described by Klein (2014), could be used as an insight-hunting technique in the context of information visualization. In addition, although the overall pattern or picture was picked up and reported in the participants' answers to this question in both delivery models, there was a degree of variation in the examples used to prove a specific point with the software model. Outlier insights also appeared to be more dominant when there was no narrator or speaker to guide or divert participants' attention to other details.

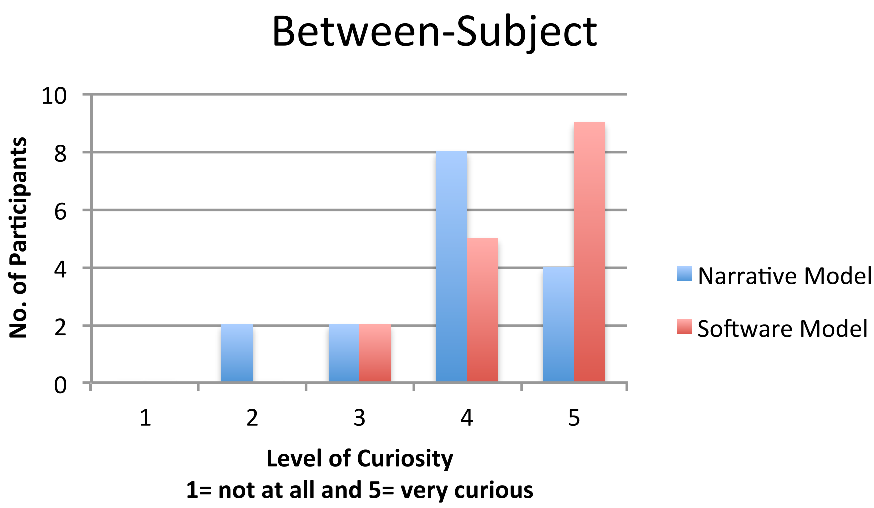


Figure 5.10: Participants' subjective feedback on the level of curiosity about the data/story for each delivery model (between-subject experiment)

Q3 and Q4 asked the participants more explicitly about the new information they had learnt and any apparent contradictions with their existing knowledge. As stated in section 5.4, when looking at each of these two questions separately, no specific pattern emerged from the answers from the various participants. However, if we look at the answers to both questions by each individual participant, we may reach some interesting conclusions. For example, it was noticed that when most of the data was new to the participant, either with the narrative or software model, there tended to be no or very few contradictions. One participant answered '*I didn't know too much about the topic so I'm not sure*' to Q3. This participant was not confounded by any contradictions. Conversely, respondents who were more knowledgeable about the topic were better placed to answer the question on contradictions and were also more critical about the emphases and omissions, especially those made by the speaker in the narrative model. For example, one participant answered '*I was rather surprised that the presenter chose to only talk about % of adults with HIV. Children are a huge demographic in Africa and Asia.*' to Q4. In other cases where the participants were less knowledgeable about the topic but still knew a little about it, they merely reported corrections of some previous stereotypical beliefs they had held regarding the topic, for instance that HIV rates are high throughout the whole of Africa. Scepticism and critiquing were generally more evident in answers to Q5, which featured the narrative model. Table 5.1 shows some examples.

This chapter provided details on the qualitative and quantitative analysis performed on the experiments data, and the main emergent findings. The next chapter provides a number of styles/themes of storytelling in the VAST Challenge videos as emerged from the transition diagrams.

Table 5.1: Excerpts from transcripts of participants' answers, illustrating the differences between the narrative and software models.

	Q1	Q2	Q3	Q4	Q5
Narrative Model	*Distribution of AIDS victims by geography.*	*A study has shown that from 1959 until now, there has been as impact of countries' economy on child mortality whereas successful industrial countries have less child mortality than countries that are not as successful. But if resources such as beds, vaccines, and mother care are provided, then poor countries could save millions of children every year globally. An example of this was seen where these aids were provided to Tanzania where the country immediately start reducing child mortality numbers.*	*Nothing* *Most of the data was new to me. I could name for example that countries with high infection levels seem to be ...*	*Not all Africa, but only the south of the continent is heavily affected by HIV.* *I didn't know too much about the topic so I'm not sure.* *I was rather surprised that the presenter chose to only talk about % of adults with HIV. Children are a huge demographic in Africa and Asia.*	*Political posturing, perhaps to seek funding.*
Software Model	*How child mortality changed over time.*	*Over the last 200 years, child mortality overall in the world reduced dramatically. Only 100 years ago, mortality was about 30% for babies 0-5 years old. Our days is about 3%. Only 40 years ago, on average there were about 6 babies born per woman. This was reduced to about 2 children over a 30 years period from 1970 till 2000 and for the last 10 years remained at this level.*	*Nothing* *I learned that assumptions are not always correct! For instance, Israel has a lower mortality rate and higher fertility rate than America.*	*It doesn't contradict anything specific, except my incorrect belief that Afghanistan held the top position for child mortality rate.* *Not all Africa, but only the south of the continent is heavily affected by HIV.*	*To show others what can affect mortality rates and raise awareness of that.*

Chapter 6

Interpretation and Analysis of the Transition Diagrams

This chapter relates to Objective 5 of this thesis, *To use the visual representation of transitions in narrative visualization to summarise and provide a number of storytelling styles in narrative visualization.* Following the proposal of the framework of transition types in narrative visualization, coding transitions in VAST Challenge videos, and representing the contents and transition types used in each video graphically with a transition diagram (Chapter 4), the next step is to take a deeper look at the transition diagrams created in order to characterise and work out some storytelling styles in narrative visualization, and examining how narrative construction can be applied to data visualization. This is the aim of this chapter.

This chapter is divided into two main sections. Section 6.1 revisits the transition diagrams, how the transition types can be read and interpreted, and what should be taken into account when interpreting the transitions. Section 6.2 summarises and discusses some of the storytelling styles identified using the transition diagrams. It is divided into four sub-sections. Sub-section 6.2.1 presents styles related to the balance between analysis/exploration and messaging in data storytelling. Sub-section 6.2.2 presents and describes the episodic and continuous approaches to storytelling. Sub-section 6.2.3 presents and discusses storytelling styles related to narrative structure and messaging paradigms, and sub-section 6.2.4 summarises and provides examples of storytelling with a rhythm. It is important to note that this chapter covers only the identification of storytelling styles, but the discussion of these styles will be presented in the next chapter (Chapter 7).

6.1 Reading Transition Diagrams and Interpreting Transition Types

Before moving to the storytelling styles found and identified in the case study of VAST Challenge videos, it is useful to look at the representation of each type of transition, and their role in the progression of the data story. This will provide a richer scope and explanation of the ways of looking at these transition types and diagrams. Going back to comics, each transition type emphasises something and helps the story to progress in a certain way. For example, subject-to-subject transition is a change of perspective in order to make comparisons (McCloud 1994, 2006). In the same way a subject-to-subject transition in the context of data visualization is about changing perspective in the data set. Moment-to-Moment increases suspense, slows the action down, and catches small changes. A series of action-to-action transitions helps to move the narrative forward, creates a temporal dimension to the story, and places the narrative in the history or order of the exploration journey. Scene-to-Scene transition encourages deductive reasoning to provide an explanation for the link between scenes. Finally, aspect-to-aspect transition usually holds time still, and encourages contemplation. To better read and interpret transition diagrams, it is important to note that the presence or existence of many aspect-to-aspect transitions in a data story does not necessarily mean that the analysts have been using too many visualization techniques, views, or tools. There might be only two visualization techniques, such as in the data story in Figure 6.10 but there is lots of back-and-forth between them to help us look at the data from a different angle, and derive an answer or conclusion (or in other words, the main message of the story). Knowing how each type of transition may affect storytelling and how it can be interpreted should provide richer insights when examining a transition diagram.

Before examining the diagrams and figuring out some themes and styles of storytelling, we should not ignore the influence of the requirements of the VAST Challenge in order to avoid overstating some observations. For example, the instructions for the challenges ask for some details of how the analyst (the storyteller) arrived at the answer. However, although these videos are created to be presented to professionals or experts in data visualization, as stated in Chapter 4, and one may think that the analysis process and/or description of the visual representation is only of interest to these target audience, this is not true. Looking at some more casual narrative visualization videos targeted at a more general audience (for example, Rosling's videos used in the experiments in Chapter 3), describing the visual representation so the audience knows how to interpret the data is a common characteristic of any narrative visualization, no matter the target audience. Similarly, a description of the analysis process that helps the storyteller to make a specific point or reveal a finding, regardless of how detailed or advanced this description is, will still be found in any narrative visualization and in many cases complements the description of the visual data representation. It might also serve as a way of providing stronger evidence for the main message of the story represented by the findings of this analysis. In which case, the audience will be more willing to accept this message. Therefore, while the use of action-to-action transition is found to be common as will be seen in the diagrams below, this was not a result of the way these stories were coded, yet it is characteristic of any narrative visualization. Additionally, VAST Challenge instructions discourage spending the whole time (the 4 minutes allowed for the mini-challenges) on showcasing a specific visualization tool/application and how it was built. So, the use of aspects and aspect-to-aspect transitions as part of the oral narration in the videos is important to understand and interpret the data and the answer(s) presented. Given this, the nature of aspects and aspect-to-aspect transitions in VAST Challenge videos, and the corresponding transition diagrams, is comparable to other types of narrative visualization aimed at more general audiences and delivered by different means (e.g., slide-show presentation or an interactive visualization).

Having learned how a transition diagram can be read and interpreted, the next step is to look at the created transition diagrams and figure out some similarities and differences between data stories. While each story has its own diagram, there are many recurring characteristics and styles, although there are no two transition diagrams that are completely identical. By looking at things like the structure and hierarchy shown in each diagram, the transitions made, the grouping of several steps or story units that lead to a finding/conclusion, and how this has been used in the data stories, a number of storytelling styles have been identified. Section 6.2 below presents some of these styles.

6.2 Storytelling Styles

As stated in Chapter 4, visually representing the components and transition types used in the case of VAST Challenge videos enables comparisons to be made between different storytelling styles, and shows how the narrative is constructed through the progression or transition from one step or story unit to another. This chapter examines the created transition diagrams for the case studies of VAST Challenge videos and, as a result, identifies several storytelling styles which existed in the case studies and emerged more clearly from the transition diagrams. It is divided into four sub-sections that represent storytelling themes, under each of which a number of storytelling styles have been identified. One or more examples are also provided to help explain and clarify how these styles were identified using the transition diagrams.

6.2.1 The Balance between Exploration and Goal/Message in Data Stories

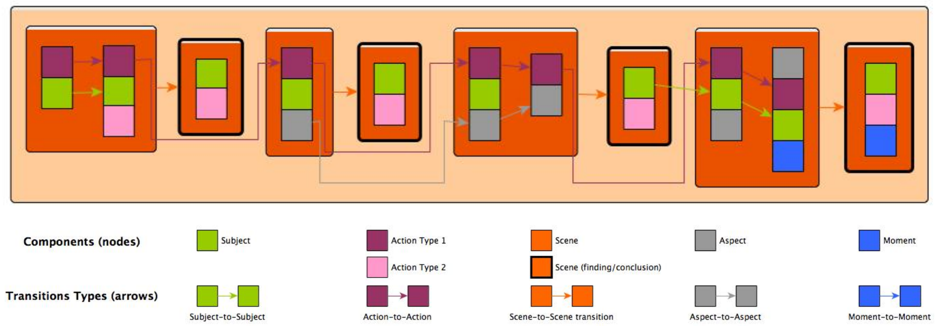


Figure 6.1: An example of a story that achieved or represents a balance between analysis/exploration and the answers/message (Submission: VAST 2009- MC1- University of Maryland- Mindlab).

Data stories in VAST Challenge videos can be considered goal-driven stories, aiming to achieve a goal or answer a particular question. By examining the transition diagrams, it can be seen that there are different ways to balance the content of the story between the data exploration journey on one hand, and the answer(s) or the goal achieved through taking this journey on the other. The transition diagram in Figure 6.1 shows a story that achieved this balance and delivered a *combination* of both the exploration steps and the answers or conclusions that represent the message(s) of the story. The exploration or analysis journey is represented in the diagram by the series of *actions* and *action-to-action* transitions along with some *subject-to-subject* and *aspect-to-aspect* transitions, while the message(s) or the point(s) is represented as findings scenes in the story (*scene-to-scene* transitions).

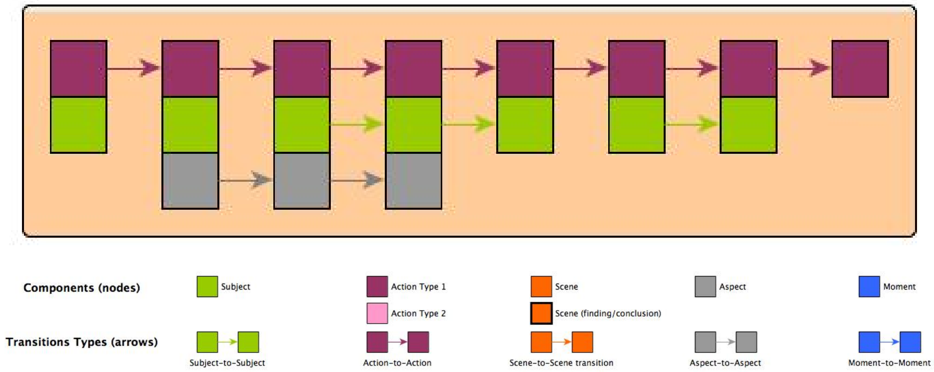


Figure 6.2: An example of a story on the exploration/analysis process which does not provide any findings/conclusions (Submission: VAST 2010- MC1-giCentre).

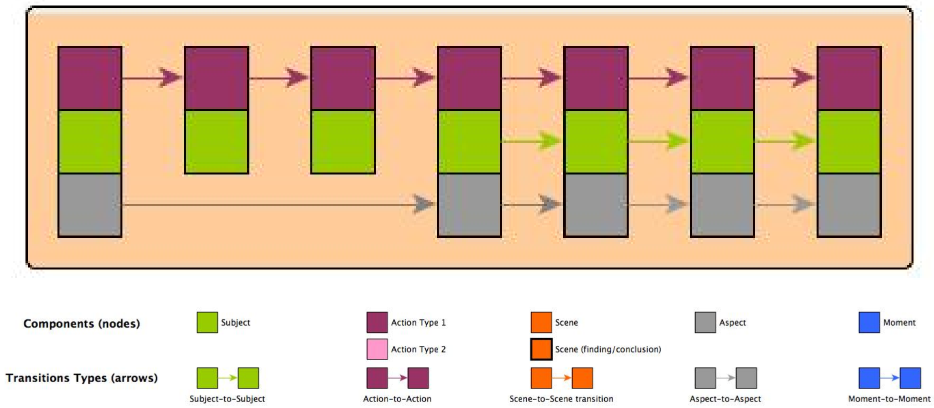


Figure 6.3: An example of a story on the exploration/analysis process which does not provide any findings/conclusions (Submission: VAST 2010- MC1-Georgia Institute of Technology- Jigsaw).

In contrast, in some data stories, the analysts/participants chose to only tell a story of how they arrived at their goal/answer without reporting or telling the answer itself. We can call this an *exploration story*. Examples of this kind of data story are those represented by the transition diagrams in Figures 6.2 and 6.3. We can see that in these two examples, no findings scenes exist and, consequently, there are no *scene-to-scene* transitions between a group of story units and a finding. The stories in this case progressed by a chain of *action-to-action*, *aspect-to-aspect*, and *subject-to-subject* transitions. It is important to note that while only telling a story of the exploration or analysis process, this does not mean that these stories did not find an answer. It is the choice of the analysts/participants to only tell this story and leave the findings and answers to the accompanying written summary.

Some stories were found to not fully achieve the intended goal/answer despite delivering some findings in the video. This can sometimes also be seen in the transition diagram representing this story. An example can be seen in the story represented by the transition diagram in Figure 6.4; this story reported a few findings in the beginning, after a relatively short sequence of *action-to-action* transitions (they suspected some in the scenario), but after that the story did not provide any more findings. This is shown by the relatively long series of *action-to-action* transitions after the initial findings, with no more findings revealed. In comparison to all the other encoded stories, the unusual structure of this story, as seen in its transition diagram, suggests that no more evidence was found and that the story was unable to confirm or build on the initial finding.

It is important to note that not all stories end with a finding, there might be one or two story units that leave an open ending to the story and represent how further exploration and analysis might reveal more insights/findings using the same approach presented throughout the story. However, the very long series of actions and action-to-action transitions in the data story in Figure 6.4 is more than a simple note at the end of the story to tell the audience what they can explore by themselves, which is usually provided in one or two story units at most.

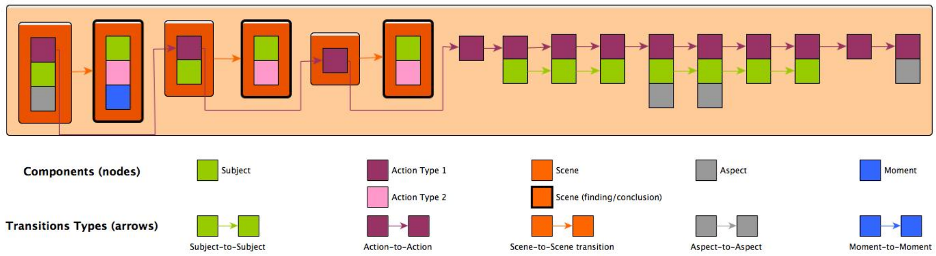


Figure 6.4: An example of a story that delivered some findings in the beginning (the orange nodes/scenes with bold-faced edges) and then did not deliver more findings or provide more evidence (ended with a long series of Action-to-Action and Subject-to-Subject transitions)- (Submission: VAST 2009- MC1- Leonard-EAKOS 2009).

6.2.2 Episodic and Continuous Approaches to Storytelling

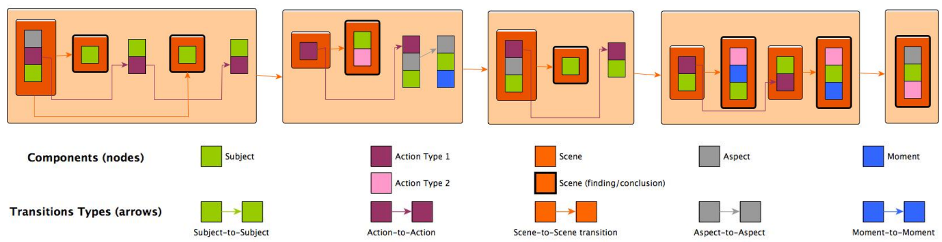


Figure 6.5: An example of an episodic story that is divided explicitly into a number of sub-topics or high-level scenes (Submission: VAST 2009- MC1-GAMI: International Institute of Information Technologies).

The explicit division of some data stories into a number of sub-topics was already recognised from the coding process and the creation of the transition diagrams. As stated in the definition of the *scene-to-scene* transition in Chapter 4, a high-level scene-to-scene transition occurs between sub-topics in the case of data stories that are explicitly divided into a number of them. This was also differentiated in the transition diagrams as can be seen in the examples (below).

First, the *Episodic* approach to storytelling is a topical way of organising the story and its contents. It is similar to book chapters. This approach to storytelling breaks the chronology and continuity of the story for the sake of organisation. For example, a data story may be divided into 3 episodes, one for data processing and preparation, the second for building a visualization tool/representation, and the third for main insights that can be revealed. These three topics, or as we can call them *episodes*, are not necessarily narrated in the same order in which they occurred. Figures 6.5 and 6.6 show examples of *episodic* data stories. It is important to note that, as in any other medium, dividing a story into a number of episodes/sub-topics can be done based on different factors. For example, these episodes may represent stages in the analysis or addressing of the question, they can also represent the analysis done by each participant/analyst, or can represent different data variables and/or visualization applications.

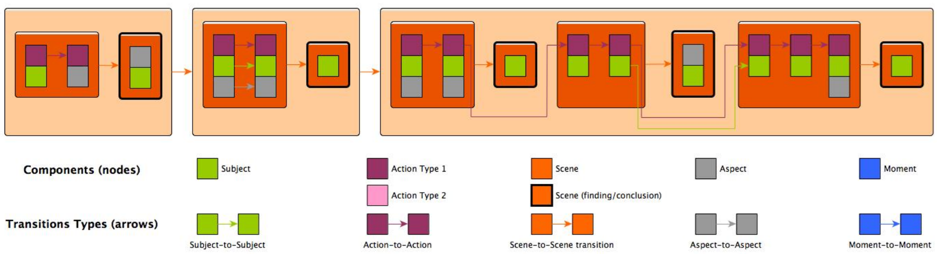


Figure 6.6: An example of an episodic story that is divided explicitly into a number of sub-topics or high-level scenes (Submission: VAST 2010- MC3-Team Stuttgart and Tubingen).

Continuous approach to storytelling, on the other hand, is a non-topical way of organising the story and its contents. It is like telling the whole story in one continuous episode. In this approach the relationship between the series of story events and components is more apparent in the diagram, as in most cases an event usually builds on the previous one. In this type of data story the existence of longer series of *action-to-action* transitions shows the temporal or chronological order of the analysis. Figures 6.7 and 6.8 show examples of *continuous* data stories.

However, there are some stories which were not explicitly divided into sub-

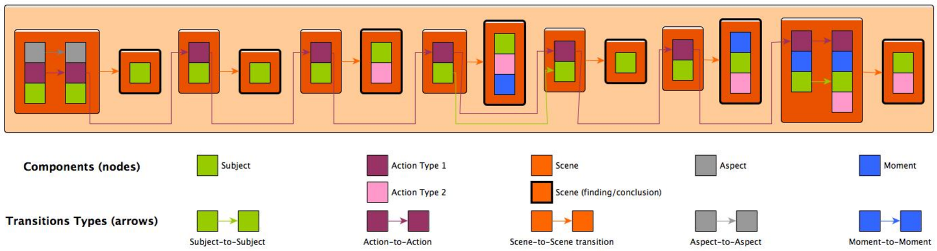


Figure 6.7: An example of a continuous approach to storytelling, i.e. telling a story in one continuous episode/scene shown in the diagram with the outer high-level box/scene (Submission: VAST 2009- MC1- Georgia Institute of Technology).

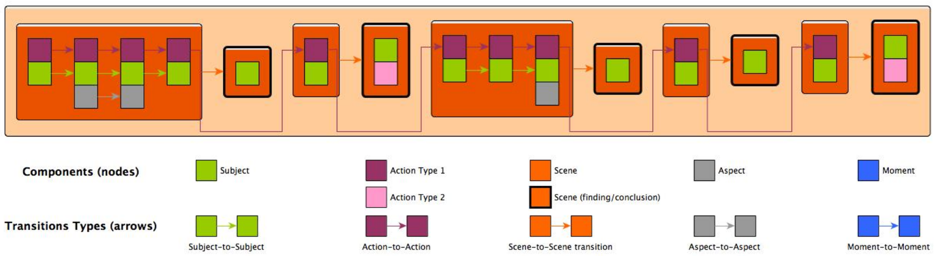


Figure 6.8: An example of a continuous approach to storytelling, i.e. telling a story in one continuous episode/scene shown in the diagram with the outer high-level box/scene (Submission: VAST 2011- MC3- Middlesex University/ University of Leeds).

topics, but whose narration suggests a very clear separation between different parts of the story, despite not using title slides to separate these parts. These were also represented in the transition diagrams by maintaining the outer high-level box/scene that shows the continuity of the story or the lack of explicit sub-topics, while inside this single high-level scene/episode, the story is also divided into a number of scenes. Figure 6.9 shows a story of this kind; there were no sub-topics but there is a clear separation in the narration between different parts of the question being answered in this video submission. Figure 6.10 provides another example of this kind of story; in this story the narrator clearly distinguishes the steps of the analysis done by each of the participants in this VAST Challenge video submission.

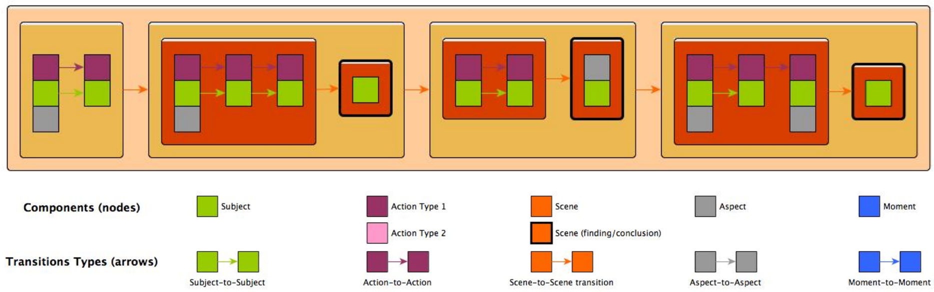


Figure 6.9: An example of a continuous story with a clear separation between different scenes/parts of the story. In this example, these scenes represent parts of the question being answered (Submission: VAST 2010- MC3- Georgia Institute of Technology).

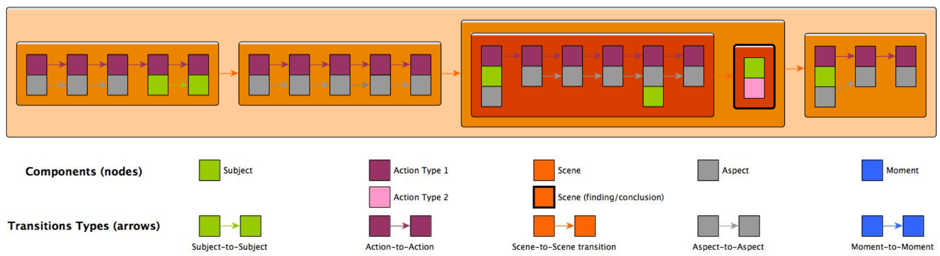


Figure 6.10: An example of a continuous story with a clear separation between different scenes/parts of the story. In this example, these scenes represent the analysis done by each participant (Submission: VAST 2009- MC1- Virginia Tech).

6.2.3 Narrative Structure and Messaging Paradigm

Sub-section 6.2.1 addressed whether the storyteller provides a specific answer/message, or just discusses how they arrived at the answer, leaving it to the audience to find this in a written report or summary, such as that accompany VAST Challenge video submissions. Sub-section 6.2.2 defined episodic and continuous approaches to storytelling. In this sub-section, we look more deeply into the narrative structure and ordering of findings and analysis steps, as well as the messaging paradigm utilised in data stories. Findings and answers to questions provided in data stories can be considered the main elements or blocks that build the overall message of the story. Therefore, we can examine messaging paradigms used in the case study of VAST Challenge videos by looking at the created transition diagrams and examine how the findings are presented in these data stories.

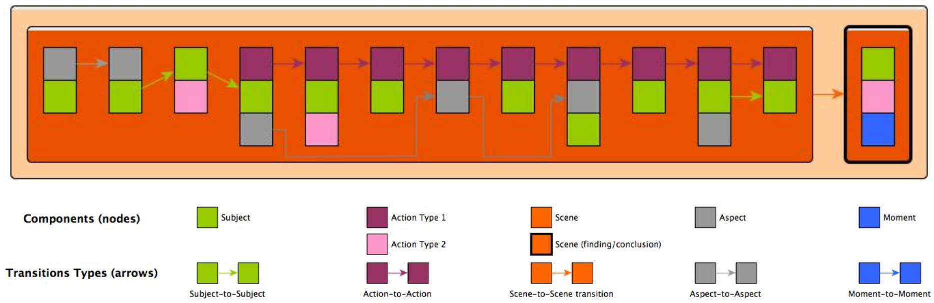


Figure 6.11: An example of a continuous story (with one episode) that consists of a long series of story units or actions and provides the findings at the end (Submission: VAST 2009- MC1- VIDI Surveillance).

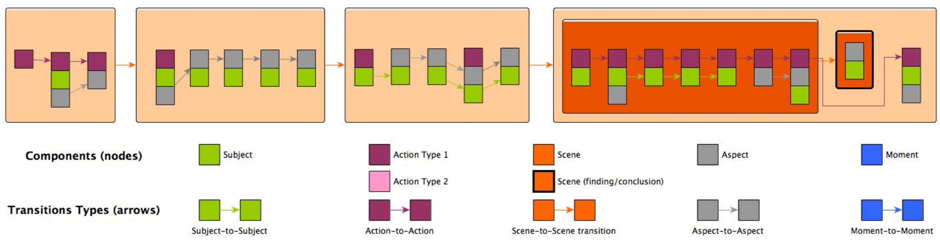


Figure 6.12: An example of an episodic story that provides the findings at the end. (Submission: VAST 2010- MC1- VRVis-ComVis

With regards to the messaging paradigm, a number of styles were identified by examining the transition diagrams of the case study of VAST Challenge videos. First, some stories/videos were found to be reporting all the work and analysis steps in addressing a question, and then reporting the final answer/message at the very end. Examples of this messaging paradigm are the data stories represented by the transition diagrams in Figures 6.11 and 6.12, which show that the *findings or whole message comes at the end*. In such stories, the finding at the end is mostly a complex detailed finding.

Another approach is to break down the findings of the story, but also start presenting these findings toward the end of the story. In this case, the first finding usually comes after a relatively long series of story units, after which the pace of the story increases, with subsequent findings coming after shorter steps. Examples of this approach are shown in Figures 6.13 and 6.14.

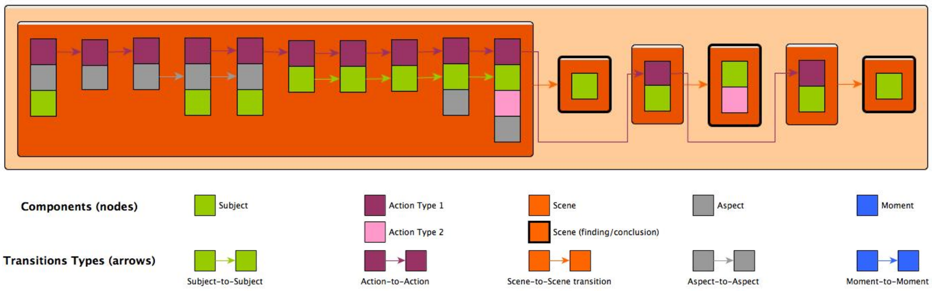


Figure 6.13: An example of a data story where the first finding comes after a relatively long series of story units and then subsequent findings are revealed after shorter steps toward the end of the story. (Submission: VAST 2009-MC2- University of Ulm and derive- VIScover)

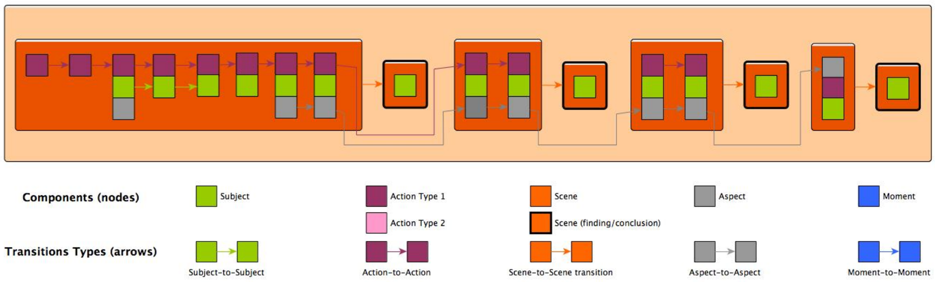


Figure 6.14: An example of a data story where the first finding comes after a relatively long series of story units and then subsequent findings are revealed after shorter steps toward the end of the story. (Submission: VAST 2010-MC2- University of Constance- Applied Visual Analytics)

In contrast, in few cases, this narrative structure that starts with exploration and ends with findings, delivering the message at or toward the end of the data story was completely reversed. The story started with a finding or answer, and then proceeded to the exploration and presenting more findings. An example of this type of data story is provided in Figure 6.15.

A different approach or paradigm of messaging and revealing findings in a data story is what we can call *gradual messaging*, in which the findings are

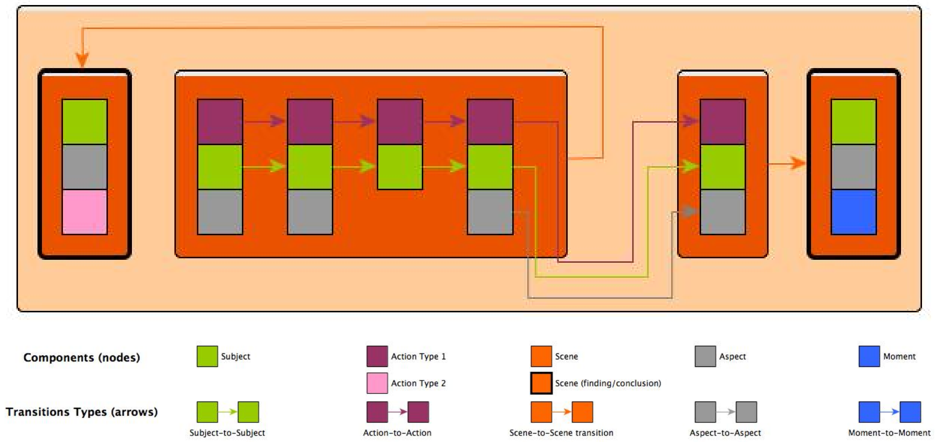


Figure 6.15: A transition diagram shows a data story with reversed structure. A finding is represented first and then the exploration process that lead to this finding. The exploration process then unfolds and more findings revealed (Submission: VAST 2011- MC1- Pennsylvania State University).

broken into smaller pieces that are revealed after shorter analysis steps from the beginning throughout the whole story. Examples of this paradigm are the data stories represented by the transition diagrams in Figures 6.16 and 6.17.

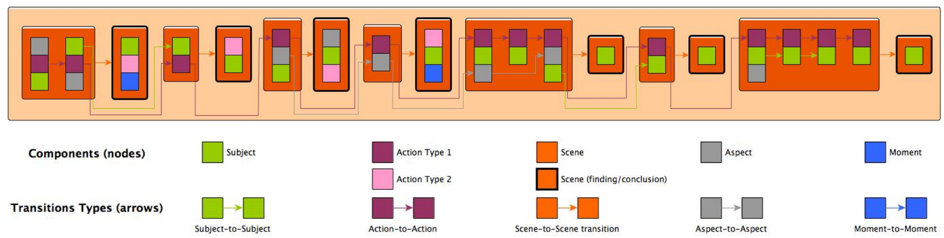


Figure 6.16: An example of a continuous story that adopts gradual messaging, i.e. providing pieces of finding/conclusion after relatively short steps throughout the whole story (Submission: VAST 2009- MC1- Palantir Technologies).

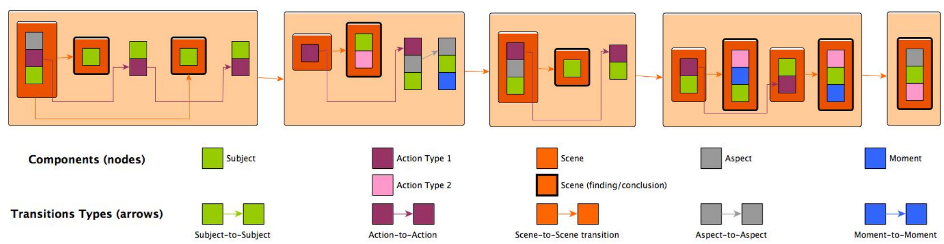


Figure 6.17: An example of an episodic story that adopts gradual messaging, i.e. providing pieces of finding/conclusion after relatively short steps throughout the whole story (Submission: VAST 2009- MC1- GAMI- International Institute of Information Technologies).

6.2.4 Storytelling with a Rhythm

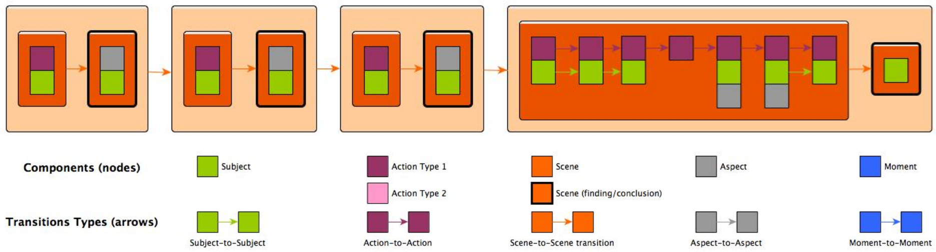


Figure 6.18: A transition diagram that shows a rhythm in the first three episodes of the data story. (Submission: VAST 2010- MC3- University of Konstanz)

According to Block (2001), 'every rhythm is made up of three subcomponents: alternation, repetition, and tempo'. Alternation may happen between two or more things, such as sound and silence, red and blue, and up and down. Repetition of an alternation creates a rhythm. For example, red-blue-red-blue. A rhythm cannot be recognised if something happens only once (Block 2001). The last subcomponent of a rhythm, Tempo, is the rate between the states, colours, sounds, etc. This last subcomponent of rhythms cannot be clearly seen in the transition diagrams but looking at the alternation and repetition in the transition diagrams created for the case study of VAST Challenge videos, we can identify some data stories that have some form of rhythm. Some examples are provided below in this sub-section.

Some episodic data stories have a rhythm between the episodes when something is repeated in each of them. For example, the story in Figure 6.18, which is an episodic story, has a clear rhythm in the first three episodes. This rhythm involves a story unit with an action and subject, followed by a finding. The same thing is then repeated in the next two episodes. However, while the last episode is not typically the same as the first three, we can still argue that the rhythm is maintained by providing a finding at the end of each episode.

Figure 6.19 shows a continuous data story which did not use explicit subtopics, but there is a clear separation between the parts of the problem in the narration. Each part is answered with some analysis before moving to the next part. While not completely identical, the last three episodes that represent parts of the problem are very much the same, particularly the last two story units before revealing a finding at the end of each part. This is also a form of rhythm in the storytelling. This very structured format of storytelling shown in the transition diagram of this particular data story is also recognised in the commentary of the award given to this submission which is for *excellent process explanation*, which is mostly related to the adoption of this rhythm in the story making the revelation of findings and messages predictable.

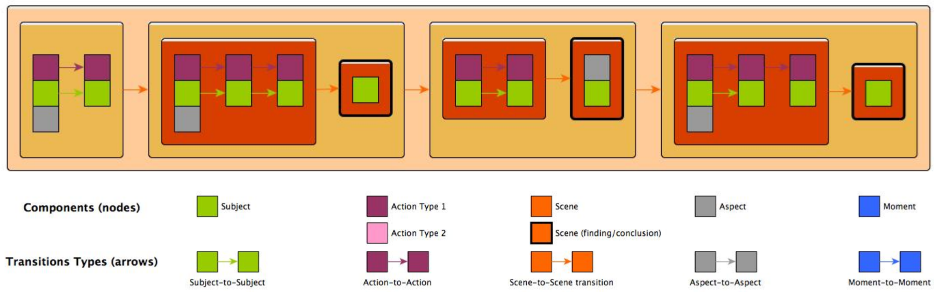


Figure 6.19: A transition diagram of a continuous data story that has a rhythm in the last three episodes. (Submission: VAST 2010- MC3- Georgia Institute of Technology)

This chapter examined the transition diagrams created for the dataset of VAST Challenge videos, and discussed some storytelling styles found in the case study. The next chapter concludes and discusses the findings of this thesis, and provides suggestions for future research.

Chapter 7

Discussions, Conclusions, and Future Work

The goal of this thesis was to contribute to the field of narrative visualization. This chapter discusses and concludes the findings of the experimental work, the framework of transition types proposed and the storytelling styles defined in Chapter 6. It is divided into three sections. Section 7.1 revisits the research objectives specified in section 1.3 in the Introduction. It also summarises the findings of this research work, offers conclusions and suggests the implications regarding storytelling through information visualization based on the findings and the proposed work. section 7.2 reviews the contributions of this research work. The limitations of the study and the recommendations for future research will be discussed in section 7.3.

7.1 Research Objectives: Summary of Findings and Conclusions

7.1.1 Objective 1: To identify the characteristics of storytelling in general and as an emerging field within information visualization

Chapter 1 introduced the topic of the thesis, touched on some different perspectives of storytelling, and defined the terms narrative, story, storytelling through information visualization, and data story, which are then used throughout the thesis. It also summarised the contribution of this research.

Chapter 2 expanded on the discussion of some storytelling perspectives, reviewed the previous work on storytelling in information visualization, and discussed the theories and characteristics of storytelling in a number of domains, such as films and critical theory, and their relevance to storytelling through information visualization. By summarising and reviewing the previous work completed in this area of information visualization, and on related issues in other storytelling and communication domains, emerging issues and aspects that require further research were identified, which in turn shaped the scope of the research at hand. Looking at storytelling in general, regardless of the domain, and at storytelling in more mature domains, helped to open up new methods of looking at and addressing storytelling in information visualization. A number of storytelling characteristics were covered in the first two chapters of the thesis, including what the nature of a story is, the narrative structure and progression of events/details in a story, and how the story can be delivered and communicated. There are different perspectives to storytelling, one of which is critical theory. Storytelling from the point of view of critical theory is concerned with myths and the 'intimitation of an action'

(Kearney 2002). However, although this differ from the perspective of the work reported in this thesis, which relates to storytelling with data through information visualization, some concepts and ideas are still relevant to some extent, for example, the idea of structuring narratives as a myth. Understanding the literature on storytelling in general, and on data storytelling in data visualization, informed and contributed to the subsequent objectives and work completed in this research.

Although Chapter 1 and Chapter 2 introduced the topic/area and provided context for the work that has been completed, this first objective extends through all the chapters, as more details and characteristics were identified and examined in the context of information visualization, such as the nature of hierarchical or nested transitions in narrative visualization.

7.1.2 Objective 2: To investigate the effect of the information visualization delivery on narrative construction and development

In Chapter 3, the effect of information visualization delivery on narrative construction and development was explored. This involved conducting within- and between-subject experiments to examine two models: 1) a direct narrative from a speaker using visualization software to tell a data story, and 2) allowing the users to explore the data interactively through the software. The results of the experiments were presented in Chapter 5.

It is found that the delivery models had a strong impact on the use of outlier insights in participants' stories in the between-subject experiment. However, this impact was small in the within-subject experiment. A possible interpretation of this variation refers to what Hans Rosling (the narrator/speaker) focused on in the videos used in each set of experiments. In the video used in the within-subject experiments (Rosling 2009), which was on the HIV epidemic, the case of an outlier was the most important (Africa) and Rosling focused on it and spent a considerable time explaining the patterns and causes for the epidemic there. On the other hand, in the video on child mortality (Rosling 2008) that was used in the between-subject experiments, the concentration was not on an extreme case but on the variation between different countries/regions and improvement over time. Hence, when the dataset in the two delivery models was neutral and did not focus on outliers, outlier insight was used more often by the participants who constructed and wrote data stories by exploring the data in Gapminer (software model) than those who watched the video (narrative model). The fact that outliers caught the audiences' attention is important, particularly in the absence of a narrator who directs audiences' attention and emphasises the desired events in the story. The danger in this approach is that outliers, by definition, are not representative of the data as a whole, so care in needed to lead users from attention-grabbing outliers to the core message implied by the data.

As it is more likely that participants will use a Problem-Solution or Cause-and-Effect narrative structure with the narrative model, one should pay special attention to such details in the absence of a narrator who can justify, explain, or provide background information. Annotations and choice of labels play an important role in this case. Moreover, it is important to note that using a specific narrative structure by a participant, such as General-to-Specific, does not necessarily mean that the general insights in the beginning of the participants' story are more important than the following insights. Similarly, reporting specific insight types does not necessarily mean that these are the only insight types the participant could gain. Storytelling is selective by nature, and the reported insights represent those that contributed to the overall mental model of the audience rather than a quantification of what they gained.

As shown in section 5.4, in both rounds of experiments, some participants were skeptical about the purpose of delivering a story in the narrative model. They suspected a personal or political purpose behind the story. While this was not a question in the study and there was a relatively small number of participants who reported such an impression in the narrative model (two in the first round and two in the second round), it is still interesting, especially as it occurred with two different videos/datasets. This may indicate that the objectivity offered in the software model (by Gapminder software) did not prompt such questioning.

7.1.3 Objective 3: To propose and develop a framework of transition types used to tell data stories through information visualization based on theories in other storytelling mediums

The second part of this thesis looked at transition types in narrative visualization. Inspired by McCloud's (1994, 2006) work in comics, which defines six types of panel-to-panel transition, each with a role in story progression, a framework of transition types for telling data stories through information visualization was proposed and developed in Chapter 4. Examples of these transitions, and a discussion of their roles in story progression, were also provided.

The taxonomy consists of five types of transitions that can be used to tell data stories through information visualization: subject-to-subject, action-to-action, scene-to-scene, aspect-to-aspect, and moment-to-moment (all defined in section 4.2.2). This framework or taxonomy provides a new way of understanding how data stories function in information visualization by using a defined set of elements that represent the ingredients of any data story told through information visualization. It also defines transition types between the elements that contribute to the progression of the story in a certain way. These transitions show how one story unit or event may lead to the next; hence, they can be considered as ways to maintain or enhance the continuity of the story by building upon its parts and connecting them in a certain way to ease the story progression. While the viewer/user/audience might not perceive or closely examine all the transitions that occur during the data story, he or she will know that they encountered some of them. For the designers of narrative visualization or data storytellers, knowing the transition types is important in order to create or tell a smooth, seamless data story.

The proposed taxonomy is valuable for the designers or authors of data stories in many ways, whether they are presenting their stories in an interactive visualization or directly narrating them. The taxonomy defines five types of possible transitions. A clear understanding of these transition types and their role in story progression can help the designer in choosing his or her next step in the story as well as shaping the overall story structure and experience that he or she wants the audience to get. The role of each transition type may also help in the visual design of interactive narrative visualizations where the user or audience has more freedom to explore the story and choose his or her own path through it. In this case, the designer should choose the design elements and guidelines that maximise the likelihood that the user or audience will have the same story experience that the designer had in mind. For example, if the designer aims to trigger a comparison, he or she may use a subject-to-subject transition and choose design elements or guidelines that encourage the user to make this kind of transition when exploring the story. While this may require further investigation to provide a definite and precise answer as to what design elements encourage specific types of transitions over others, some simple design elements can be provided. For example, annotations, direct guidance and posing specific questions about the presented data in a way that is likely to motivate the user to choose a specific direction and make a specific transition may be helpful.

7.1.4 Objective 4: To propose a visual language to characterise transitions in narrative visualization based on the proposed framework of transition types

Following the framework of transition types used to tell data stories through information visualization, a visual representation summarising the use of transition types in data stories in the case study of VAST Challenge videos was also provided (Chapter 4).

A transition diagram that represents a data story captures not only the low-level story units, but also the higher-level sub-topics and parts of the story and the connectivity of these low- and high-level parts in the data story. The use of the transition diagrams helps in gaining more insights about the nature of storytelling in data visualization, such as the hierarchy of the story, what components have been used and how the story progressed on various levels. This is not limited to a single story, as transition diagrams enable the comparison and classification of data stories that share a specific feature(s). A transition diagram is not just a unique picture of one story that is unrelated to other pictures of other stories. There are some common features. For example, there are a group of stories that have a broadly similar structure, while there are rare cases with an inverted structure (Chapter 6). A good validation of the approach is that it was possible to spot these common storytelling styles and structure. The presentation and coding of data stories in VAST Challenge videos revealed common visual elements which reflected these characteristic structures and types (see Objective 5 below for more details on these styles).

7.1.5 Objective 5: To use the visual representation of transitions in narrative visualization to summarise a number of storytelling styles in narrative visualization

This objective is also related to the second part of the thesis on transition types. Here, the created transition diagrams were examined and used to identify a number of storytelling styles in narrative visualization (Chapter 6).

As stated in sub-section 6.2.1 in Chapter 6, not all stories end with a finding. Instead, there might be one or two story units at the end which provide an *open ending* to the story and represent how the audience may explore further to reveal more insights using the same approach presented by the storyteller throughout the story. Data stories with open endings can be thought of as having a martini-glass structure explained by Segel & Heer (2010), although the further exploration or the reader-driven stage of the martini-glass may entail switching to a different genre of narrative visualization other than video presentation which is a very authored approach to data storytelling. Also, the relevance of the discussion of the balance between the exploration/analysis process, and the message/insights the story aims to deliver (sub-section 6.2.1 in Chapter 6) is not that one of these approaches is better than the other. the important thing for any story, whatever the domain or narrative genre used, is to reach its audience. One may think that for a successful data story, a finding or a message should be delivered and proved. However, while the data stories in Figure 6.2 and 6.3 did not provide an answer or finding, one of them was successful in receiving an award from the VAST Challenge committee. However, this does not mean that this storytelling style would also be successful with a more general audience, as it leaves them without a message to take from the story and may overwhelm them with details that are not necessarily of interest.

While it is difficult to argue that the episodic or continuous approach to storytelling (sub-section 6.2.2 in the previous chapter) is generally more effective than the other, we can suggest when each can be more effective to use. For example, having some episodes in the story may be helpful in complex or long data stories, while the continuous approach to storytelling might be more engaging with brief stories (with shorter durations) or stories with less complex contents. This is not just applicable to video presentations. We can see these approaches to storytelling in other types of narrative visualization, where the user can interactively navigate through a data story where headings or titles are used to label different parts of the story. The continuous approach to storytelling, on the other hand, could be more helpful when the goal of the narrative is to hold the audience attention and deliver an overall message, without the need for them to remember specific numbers or details. While in the VAST Challenge videos the duration of the data stories was the same, there are variations in using these styles/approaches to storytelling. By looking at the mini-challenge questions and the use of episodic and continuous approaches to storytelling, we can notice that participants are more likely to organise their data stories into episodes or scenes when the question is explicitly divided into a number of sub-questions. In some other cases, the question may consist of just one part, but the participants chose to use an episodic approach in their data stories to better organise the contents. It is also important to note that the number of episodes in a data story does not usually reflect the number of sub-questions in a given mini-challenge, but those episodes are expected or should help in addressing complex long questions.

The messaging paradigms and narrative structures found in the case studies (sub-section 6.2.3 in Chapter 6), are relevant to previous work on theories of narrative structure discussed in the literature review chapter (Chapter 2), such as Freytag's 5-acts structure (Figure 2.2), and the discussion of what comes in the beginning, middle, and end of the story. For example, the *Exposition* might be thought of as the main data variables and the description of the visual representation and how it can be interpreted. This might suggest the existence of *aspects* and *subjects* in the beginning of the data story, which clearly occurred in most cases. The rising action and climax/conflict might be reached when or before the first piece of finding or evidence is revealed, and the resolution/denouement is reached with the last finding. Stories that begin with a finding without going through the explanation of data variables and the visual representation might be thought of as a story that has a delayed exposition. This agrees with the statement of Gershon & Page (2001) that every story should have a beginning, middle, and end but not necessarily in the same order, and that *Exposition* can sometimes be delayed and the storyteller may withhold some information and start with something more tense (Richardson 2002). In this case, we can say that the order of the three acts, Exposition-Climax-Resolution, is changed or reversed to some extent. This storytelling style is also similar to the use of inverted pyramid structure in journalism where the most important information or message is found at the beginning of the story, or where the title of an article or a data visualization reveals the main finding of the data story before going through more details and explorations. It is also similar to police mysteries where we first see a crime, so we already know something, and then the reasoning process starts. By examining the transition diagrams of data stories that adopt a *gradual messaging* paradigm, it can be argued that these data stories mostly follow the contour of communication developed by Duarte (2010), with findings in the middle can be considered as turning points that influence the flow of narrative.

Having a rhythm in the data story between the analysis and conclusions (sub-section 6.2.4 in Chapter 6) should let the audience knows when to expect a message or a finding. This may make them more prepared and alert when the time comes to reveal or deliver a message. Hullman et al. (2013) discussed the concept of 'parallel transition structure' which means repetition of a transition structure in slide-show presentations. They found that this repetition in slide-show presentations enhances memorability of the order or sequence of the slides. It can be argued that having a rhythm in a data story in the case of video presentations, such as in VAST Challenge videos, is very similar to the concept of parallel transition structure in slide-show presentations. Therefore, some form of rhythm may also enhance memorability of the details and order of information in the video's data story.

Providing these different styles of storytelling, the visualization designer may choose one of these styles to tell a data story. Together with the understanding of the transition elements and types and how to use them, these styles provide a different way of looking at the narrative structure in data storytelling through information visualization. The discussion of storytelling styles based on the transition types used in Chapter 6 and in this section provide an insight into what flavour each of these styles gives to data storytelling. It is important to note that it is not about which flavour is generally better than another, however, it could be argued that each flavour is optimised toward something different as discussed above. For example, having a rhythm in the use of transition types in the data story is expected to enhance memorability, and revealing the insights/points of the story gradually throughout the narratives helps to avoid overwhelming general audiences with complex insights and details at the end of the story and encourages deductive reasoning throughout the data exploration journey. However, it is important to note that despite the adoption of a storytelling style based on which flavour it is optimised toward, the contents of the narratives are equally important. Representing the wrong relationships or focusing on describing the wrong concepts can negate the benefits of the style. Furthermore, what has been done in the research in hand gives future researchers a basis on which they can use these styles and then evaluate how successful each style is based on the purpose, target audience, task, and so on.

7.2 Review of Thesis Contributions

This research contributed to the field of storytelling in information visualization. As stated in section 1.4, the contribution of this research applies mainly to two aspects of data storytelling through information visualization. The first contribution is to the information visualization delivery model, and the second is to the transition types used to tell stories with information visualization.

It can be argued that the contribution of this research starts with the literature review, which relates storytelling in information visualization to other, more mature domains and theories of storytelling in more depth and detail than previous work. This contributes to a wider understanding of storytelling and its characteristics, and it provides new opportunities for research in this area.

As shown in the literature review, there have recently been several research the subject of narrative visualization. However, there is little empirical work, and the question of whether a visualization tells a story, or whether the accompanying direct narrative does, has been posed on different occasions. The experimental work in this research provides empirical evidence comparing the effects of direct narrative with those of the interactive exploration of visualization. The results of the experiments and the subsequent discussion (in Chapter 5 and section 7.1.2 in this chapter) provided a number of implications and recommendations that the visual data storyteller or designer should consider when using each delivery model.

This research extends our knowledge regarding the progression of events in data stories by looking at theories from other domains and how they can be applied to the context of storytelling in information visualization. Specifically, McCloud's (1994, 2006) taxonomy of panel-to-panel transitions in comics was utilised. It provided a framework of transition types in narrative visualization as well as details on how this framework can be used to code transitions in narrative visualization and to help storytellers or designers to tell smooth stories. Furthermore, this research provided a form visual summary that can summarise the transitions used in any data story, including the hierarchy and structure of the story, and the use of this visual summary identified a number of storytelling styles that can be adopted by storytellers or designers in their stories.

Taken together, the literature review, the experimental work on the effects of information visualization delivery on narrative construction, and the framework of transition types in narrative visualization provided in this thesis demonstrate a new angle for looking at data storytelling in information visualization and facilitates understanding and development in this area.

7.3 Limitations, Implications, and Recommendations for Further Research

Any research, regardless of how well it is, should have some limitations. Good research also usually leads to further questions or areas to be examined. In this section, some of the limitations of the work reported in this thesis, and some suggestions for possible future work are presented.

Although the difference between the two delivery models (the narrative and software models) in terms of the main pattern or impression reported and whether it is *Geographical*, *Temporal*, or *Both* was not a question in the study, the difference was notable, especially in the within-subject experiment (section 5.2). Possible future work related to this could examine to what extent people perceive geography and the nature of geography - categorical or continuous, hierarchical, etc.- with and without explicit use of maps as a visualization technique when no speaker is present to guide their attention to geographical information. Also, although the difference between participants in terms of perception of chronology as part of the main topic or pattern in the data story was not very large, more participants perceived the chronology and temporal pattern as part of the main topic or pattern when they explored the data interactively. Possible future work could further examine the role of time in storytelling through information visualization while varying the level of guidance throughout the sequence of narratives.

One limitation of the experimental work done for this thesis could be the use of a student sample in the experiments for convenience reasons. Although a number of criteria were identified prior to the study in order to minimise the sampling effect on the results, and some students had backgrounds which were interesting and relevant for the purpose of the study, future work may examine some specific users in more detail, such as professional journalists, and ask them to use an inverted pyramid structure to write a data story and see how the delivery model or other factors might affect the narrative structure used and the insights reported in the data story.

With regards to transition types in narrative visualization, a possible limitation that can be looked at in subsequent work is that at this stage the approach of coding and representing transitions is not fully automated. Future work could look at or develop an automatic method of constructing these diagrams. Future research could also scale the various parts of each transition diagram and make them proportional to their durations in the original videos, particularly if this is something that might help studying this area from a different angle and/or to help obtain more details by only looking at the diagrams.

Other limitations include the lack of empirical evaluation of transition types. Although a discussion of the efficiency of some storytelling styles and how they differ was provided, no empirical or experimental work was done on transition types. Possible future work could evaluate the visual summary using the transitions diagrams. This can be done by testing whether people can distinguish between two different groups using the diagrams, as well as the effect of the use of specific types of transition on comprehension, memorability of specific details, etc.

By observing how people have chosen to tell their stories with transitions, we understand how stories are told in different ways. Given this understanding, a number of recommendations for how information visualization/visual analytics systems could be designed in future to facilitate explanation/storytelling could be provided. Before discussing some of these recommendations, it is important to understand that the type of storytelling in the VAST Challenge videos was basically done in two separate steps by the analysts/designer. The first involves exploring the data, building or capturing views, deductive reasoning, and confirming or excluding/negating some assumptions. This step also involves one or more of the insight-hunting techniques as defined by Klein (2014) from looking for connections, to looking for contradictions, and creative desperation to find an answer to the challenge or the task. The second step is to construct a narrative taking into account the purpose, the audience, the story delivery model, and how many details from the exploration/discovery step to include in the narratives. However, it is important to note here that there is a difference between an information visualizaiton or visual analytics system that facilitates storytelling by letting the user analyse the data and extract their own narratives and probably save these narratives to be communicated with others, and an information visualization that is designed with the intention of telling a story in and of itself to the user. The first is similar to storytelling in the VAST Challenge videos as described above. The latter is more like providing an information visualization in a news website to convey a particular narrative. The design of each of these needs to deploy different techniques, tools, and/or strategies.

In the storytelling style in Figure 6.1 in Chapter 6, the narrative is about deductive reasoning, showing the anaylsis steps, or providing evidence that leads to a specific conclusion throughout the narrative. A future system could allow those kinds of deductive reasoning to be captured so people could refer to the reason they use it for a specific insight. An insight management/recording tool can be developed and integrated to achieve that and allow users, whether data analysts who want to explore and tell a story about the data to some audiences as in the VAST Challenge videos, or a general user exploring some data using the visualization design and trying to assemble his own narrative, to retrieve, annotate, arrange and re-arrange their narratives and evidences.

Furthermore, omissions and emphases are an important factor in all forms of storytelling. With the existence of a presenter/narrator, there could be more omissions of some information from the visualization design (e.g. extensive annotations or text guidance) and this information is left to the presenter to reveal. However, such omissions should still be done with caution, as it was found that people were more skeptical and critical about omissions done by the narrator/presenter. Without the existence of a presenter, the task of telling the story by interactive exploration of the data by the user/audience becomes more challenging. Particularly, if the designer has a very specific path/storyline in mind that s/he wants the user to take. Emphases and omissions can be used as powerful techniques to emphasise specific patterns or insights in such cases. Additionally, emphasising specific patterns in interactive visualization may affect the subsequent exploration and interaction by the users and could be used to encourage some analyses/cognitive processes, for example, comparing two things or data points. Providing a feature that allows the recording of ongoing insights and showing the recorded insights to the analysts/designer might help in revisiting these insights and interactions made, which will then affect the final product or impact, and assist in choosing the details to keep or omit from the narratives.

When designing an information visualization or a visual analytics system that facilitates storytelling, the designer needs to provide the user/audience with all the data necessary to assemble a desired insight. At the same time, overwhelming the audience may make it more difficult and complicate the process of arriving at that insight. It is important to re-emphasise that achieving a good balance between free or full interactivity and control over the visualization and the controlled/authored interactivity is the key. Additionally, the importance of annotations in information visualization in general and in narrative visualization in particular has been acknowledged by many scholars (Segel & Heer 2010, Hullman & Diakopoulos 2011). The use of annotations in any visualization design is very important to guide users' discovery of the narratives the visualization tells. Furthermore, if the path of the story needs to be recorded and shared with others, a unifying approach to annotations will be helpful. An overview window that draws and overview of the user path (in a similar form to the transition diagram) somewhere on the screen might also be useful. Educating the user about the possible structures and paradigms of exploring or navigating through a data story beforehand, before starting the exploration and before building his exploration path is also an option. Therefore, he will know which approach or storytelling style s/he wants to adopt. Sense making and narrative construction require interpretation and explanation. Providing full control to the user to freely explore the data without directing his interpretation of this data might result in a completely different narrative. It can be seen from the participants' narratives written after exploring the data on Gapminder that although there was a degree of control over what they could explore, such as the indicators they were allowed to use, they made sense of the data and interpreted it in different ways that were highly affected by the individual's background. For example, one participant referred the reason for the mortality rate to the genocide in Gana, while another thought that health services were the cause. Both of these interpretations were not provided in Gapminder. Therefore, if the designer of a narrative visualization has a very sepcific intention or narrative in mind, it would not be appropriate to allow the audiences full control or free interactivity. However, annotations are one way to balance or overcome this point by offering the desired interpretation directly to the user while allowing free interactivity with the data.

In his book, McCloud (1994) stated that action-to-action, subject-to-subject, and scene-to-scene transitions are the most commonly used transition types in Western (American and European) comics. In contrast, in Japanese comics (or as usually called Manga) fewer scene-to-scene and action-to-action transitions are used, while they use moment-to-moment and aspect-to-aspect transitions much more than in Western comics (McCloud 1994). An interesting area to examine could be the existence of this difference in the context of narrative visualization. Specifically, one could examine how transition types differ in data stories between Western and Eastern narrative visualization. This would be more interesting and meaningful if it is done on data stories produced in the language and aimed for audiences of these cultures, rather than, for example, comparing data stories in an English language environment produced by Japanese designers for American or European audiences.

In addition to VAST challenge videos and video presentations, it would be interesting to apply the framework of transition types to other types of narrative visualization. For example, the remainder of the seven genres of narrative visualization defined by Segel & Heer (2010) might be an interesting area to extend the proposed framework to. The transition types will mostly be the same, and the expected difference when using the framework and applying the approach is the definition of story units and their boundaries.

As stated above, in interactive storytelling visualization, where the user should navigate through the data story interactively, the storyteller or narrative visualization designer should employ design elements and techniques that increase the likelihood that the audience or user of the visualization will follow the same story or path that the designer had in mind. Returning to Klein's (2014) definition of insight in Chapter 5 and that insights change how one understands, acts, sees, and feels, and his/her desires, one can incorporate these characteristics of insight in designing information visualization that facilitates explanation and storytelling. Future work could look in more detail at how different design elements and decisions encourage the use of some specific transition types over others. It can also look at how specific types of insights and understanding change how the user/audience *act* in terms of his/her interaction with the software, and how the new desires as a result of an insight may change where the user wants to go or creates a new goal, which as aresult, implies a different data exploration path and a different narrative.

Deciding on or choosing a specific storytelling style or structure is difficult using a general approach/criteria as it varies considerably among data sets, applications, analysts, and the user/audience goal. While the kind of facts or pieces of knowledge or insights might be predictable from a specific data set and/or application, the choice of the collection of facts/insights and the connection(s) between them may vary based on the audience goal, background knowledge, and beliefs. This may result in a different approach to storytelling. Therefore, it is challenging to provide a definite answer as to what structure is absolutely better. Despite this difficulty, some guidelines and design ideas have been suggested based on the findings and work conducted as part of the research in hand.

To conclude, ever since human beings have told stories. The traditional form of storytelling has many characteristics such as usually having a single protagonist, and being fiction, fantasy, or removed from real life. It also has a well-known set of theories and conventions. The term 'storytelling' has been used with data visualization for some time. However, despite this match between storytelling and data visualization, the relationship between them is rarely articulated (Segel & Heer 2010). The stronger link between the storytelling outside of the information visualization domain, and data stories in this domain in the first two chapters, as well as looking closely at the similarities and differences between them, inspired the work done in this thesis. Using theories and techniques from other more mature domains of storytelling was beneficial in understanding how people could tell data stories with information visualization. This new understanding of the area will have an effect on how future research into this topic should progress and help people tell more effective stories with data visualization and communicate facts, relationships, and messages in an accessible and understandable way. The transition types and storytelling styles defined and explained also provide a standard language among information visualization designers and authors to describe their data stories.

Bibliography

Alexander, B. (2011), *The New Digital Storytelling: Creating Narratives with New Media*, Praeger Publishers Inc.

Block, B. A. (2001), *The visual story: seeing the structure of film, TV, and new media*, Focal Press.

Bordwell, D. (2008), 'The hook: Scene transitions in classical cinema', http://www.davidbordwell.net/essays/hook.php. Accessed 2014-06-15.

Brehmer, M. & Munzner, T. (2013), 'A multi-level typology of abstract visualization tasks', *Visualization and Computer Graphics, IEEE Transactions on* **19**(12), 2376–2385.

Card, S. K., Mackinlay, J. D. & Shneiderman, B. (1999), *Readings in information visualization: using vision to think*, Morgan Kaufmann.

Chang, R., Ziemkiewicz, C., Green, T. M. & Ribarsky, W. (2009), 'Defining insight for visual analytics', *IEEE Computer Graphics and Applications* **29**(2), 14–17.

Charness, G., Gneezy, U. & Kuhn, M. A. (2012), 'Experimental methods: Between-subject and within-subject design', *Journal of Economic Behavior & Organization* **81**(1), 1 – 8.

Chen, C. (2006), *Information visualization: beyond the horizon*, Springer.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences (2nd Edition)*, 2 edn, Routledge.

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J. & Kuperberg, G. R. (2012), '(pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension', *Cognitive Psychology* **65**(1), 1 – 38.

Creswell, J. (2009), *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, SAGE Publications.

Cumming, G. (2014), 'The new statistics: why and how.', *Psychological science* **25**(1), 7–29.
URL: *http://dx.doi.org/10.1177/0956797613504966*

Deardorff, E., Little, T. D. C., Marshall, J. D., Venkatesh, D. & Walzer, R. (1994), Video scene decomposition with the motion picture parser, *in* 'IS&T/SPIE Symposium on Electronic Imaging Science and Technology', pp. 44–55.

Dragicevic, P., Bezerianos, A., Javed, W., Elmqvist, N. & Fekete, J.-D. (2011), Temporal distortion for animated transitions, *in* 'Proceedings of the 2011 annual conference on Human factors in computing systems', CHI '11, ACM, New York, NY, USA, pp. 2009–2018.
URL: *http://doi.acm.org/10.1145/1978942.1979233*

Duarte, N. (2010), *Resonate: Present Visual Stories that Transform Audiences*, Wiley.

Fulkerson, D., Nichols, P., Haynie, K. & Mislevy, R. (2009), Narrative structures in the development of scenario-based science assessments (large-scale assessment technical report 3), Technical report, SRI International, Menlo Park, CA.
URL: *http://ecd.sri.com/publications.html*

Ganti, K. (2004), 'In conversation with walter murch', http://filmsound.org/murch/interview-with-walter-murch.htm. Accessed 2014-06-15.

Gapminder (n.d.), http://www.gapminder.org. Accessed: 2013-11-30.

Gershon, N. & Page, W. (2001), 'What storytelling can do for information visualization', *Commun. ACM* **44**(8), 3137.
URL: *http://doi.acm.org/10.1145/381641.381653*

Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D. & Roberts, J. C. (2011), 'Visual comparison for information visualization', *Information Visualization* **10**(4), 289–309.
URL: *http://ivi.sagepub.com/content/10/4/289*

Grammel, L., Tory, M. & Storey, M. (2010), 'How information visualization novices construct visualizations', *IEEE Transactions on Visualization and Computer Graphics* **16**(6), 943 –952.

Gray, J., Chambers, L. & Bounegru, L. (2012), *The Data Journalism Handbook*, O'Reilly Media, Inc.

Grunwald, E. (2005), 'Narrative norms in written news', *Nordicom Review* **26**(1), 63–79.

Guest, G., MacQueen, K. & Namey, E. (2011), *Applied Thematic Analysis*, SAGE Publications.

Hard, B. M., Recchia, G. & Tversky, B. (2011), 'The shape of action.', *Journal of Experimental Psychology: General* **140**(4), 586.

Heath, C. & Heath, D. (2007), *Made to stick: why some ideas survive and others die*, Random House.

Heer, J., Card, S. K. & Landay, J. A. (2005), Prefuse: A toolkit for interactive information visualization, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', CHI '05, ACM, New York, NY, USA, pp. 421–430.
URL: *http://doi.acm.org/10.1145/1054972.1055031*

Heer, J., Mackinlay, J., Stolte, C. & Agrawala, M. (2008), 'Graphical histories for visualization: Supporting analysis, communication, and evaluation', *Visualization and Computer Graphics, IEEE Transactions on* **14**(6), 1189–1196.

Heer, J. & Robertson, G. (2007), 'Animated transitions in statistical data graphics', *IEEE Transactions on Visualization and Computer Graphics* **13**(6), 12401247.
URL: *http://dx.doi.org/10.1109/TVCG.2007.70539*

Henderson, B. (1983), 'Tense, mood, and voice in film: Notes after genette', *Film Quarterly* **36**(4), 4–17.

Hullman, J., Adar, E. & Shah, P. (2011), 'Benefitting InfoVis with visual difficulties', *IEEE Transactions on Visualization and Computer Graphics* **17**(12), 2213–2222.

Hullman, J. & Diakopoulos, N. (2011), 'Visualization rhetoric: Framing effects in narrative visualization', *IEEE Transactions on Visualization and Computer Graphics* **17**(12), 2231–2240.

Hullman, J., Drucker, S., Riche, N., Lee, B., Fisher, D. & Adar, E. (2013), 'A deeper understanding of sequence in narrative visualization', *IEEE Transactions on Visualization and Computer Graphics* **19**(12), 2406–2415.

Kearney, R. (2002), *On Stories*, Thinking in action, Routledge.
URL: *https://books.google.co.uk/books?id=QqE6IErlrdEC*

Keim, D. (2002), 'Information visualization and visual data mining', *IEEE Transactions on Visualization and Computer Graphics* **7**(1), 100–107.

Klein, G. (2014), *Seeing what Others Don't: The Remarkable Ways We Gain Insights*, Nicholas Brealey Publishing.
URL: *https://books.google.co.uk/books?id=ececngEACAAJ*

Kosara, R. (2010), 'Stories don't tell themselves'.
URL: *http://eagereyes.org/blog/2010/stories-dont-tell-themselves*

Kosara, R., Healey, C. G., Interrante, V., Laidlaw, D. H. & Ware, C. (2003), 'Thoughts on user studies: Why, how, and when', *Computer Graphics and Applications* **23**(4), 20–25.

Kosara, R. & Mackinlay, J. (2013), 'Storytelling: The next step for visualization', *Computer* **46**(5), 44–50.

Kuchar, O. A., Hoeft, T. J., Havre, S. & Perrine, K. A. (2006), 'Isn't it about time? [data visualization]', *IEEE Computer Graphics and Applications* **26**(3), 80–83.

McCloud, S. (1994), *Understanding Comics: The Invisible Art*, 1st HarperPerennial ed edn, HarperCollins.

McCloud, S. (2006), *Making Comics: Storytelling Secrets of Comics, Manga and Graphic Novels*, HarperCollins.

Munzner, T. (2009), 'A nested model for visualization design and validation', *IEEE Transactions on Visualization and Computer Graphics* **15**(6), 921–928.
URL: *http://dx.doi.org/10.1109/TVCG.2009.111*

North, C. (2005), Information visualization, *in* 'Handbook of Human Factors and Ergonomics', 3rd edn, John Wiley & Sons, New York.

North, C. (2006), 'Toward measuring visualization insight', *IEEE Computer Graphics and Applications* **26**(3), 6–9.

North, C., Saraiya, P. & Duca, K. (2011), 'A comparison of benchmark task and insight evaluation methods for information visualization', *Information Visualization* **10**(3), 162 –181.
URL: *http://ivi.sagepub.com/content/10/3/162.abstract*

Oxford English Dictionary (2012), http://www.oxforddictionaries.com/definition/english/tra
Accessed 2014-06-15.

Pramaggiore, M. & Wallis, T. (2011), *Film: A Critical Introduction*, 3 edn, Laurence King.

Quesenbery, W. & Brooks, K. (2010), *Storytelling for user experience: crafting stories for better design*, O'Reilly Media, Inc.

Richardson, B. (2002), *Narrative Dynamics: Essays on Time, Plot, Closure, and Frames*, Theory and interpretation of narrative series, Ohio State University Press.

Rosling, H. (2008), 'Reducing child mortality'.
URL: *ttp://www.gapminder.org/videos/gapmindervideos/gapcast-11-reducing-child-mortality/*

Rosling, H. (2009), 'Hiv: New facts and stunning data visuals'.
URL: *http://www.gapminder.org/videos/ted-talk-2009-hans-rosling-hiv-facts/*

Saraiya, P., North, C. & Duca, K. (2005), 'An insight-based methodology for evaluating bioinformatics visualizations', *IEEE Transactions on Visualization and Computer Graphics* **11**(4), 443–456.

Satyanarayan, A. & Heer, J. (2014), 'Authoring narrative visualizations with ellipsis', *Computer Graphics Forum* **33**(3), 361–370.

Saunders, M., Lewis, P. & Thornhill, A. (2009), *Research Methods for Business Students*, Always learning, Financial Times Prentice Hall.

Schallert, D. L., Kleiman, G. M. & Rubin, A. D. (1977), Analyses of Differences between Written and Oral Language, Technical report, University of Illinois at Urban -Champain, Center for the Study of Reading.

Sedlmair, M., Meyer, M. & Munzner, T. (2012), 'Design study methodology: Reflections from the trenches and the stacks', *Visualization and Computer Graphics, IEEE Transactions on* **18**(12), 2431–2440.

Segel, E. & Heer, J. (2010), 'Narrative visualization: Telling stories with data', *IEEE Transactions on Visualization and Computer Graphics* **16**(6), 1139–1148.

Spiegel, D. L. (1981), 'Six alternatives to the directed reading activity.', *Reading Teacher* **34**(8), 914–20.
URL: *http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ245508*

Tablue (n.d.), http://www.tableausoftware.com/about/blog/2014/5/82-preview-tell-story-your-data-story-points-30761. Accessed: 2014-08-07.

Tergan, S.-O., Keller, T. & Burkhard, R. A. (2006), 'Integrating knowledge and information: digital concept maps as a bridging technology', *Information Visualization* **5**(3), 167–174.

Thomas, R. (2003), *Blending Qualitative and Quantitative Research Methods in Theses and Dissertations*, SAGE Publications.

Tobin, P. (2007), 'Teaching IT through storytelling', *South African Computer Journal* **38**, 51–61.

Tufte, E. R. (1986), *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT, USA.

Tufte, E. R. (2001), *The visual display of quantitative information*, Graphics Press.

Visual Analytics Benchmark Repository (n.d.), http://cs.umd.edu/hcil/varepository/. Accessed: 2013-11-30.

Wallas, G. (1926), *The Art of Thought*, Harcourt, Brace.
URL: *https://books.google.co.uk/books?id=ZlF9AAAAMAAJ*

Ware, C. (2004), *Information visualization: perception for design*, Morgan Kaufmann.

Wojtkowski, W. G. (2001), 'Storytelling : its role in information visualization', *Information Visualization* pp. 1–5.

Wood, J., Badawood, D., Dykes, J. & Slingsby, A. (2011), 'BallotMaps: detecting name bias in alphabetically ordered ballot papers', *IEEE Transactions on Visualization and Computer Graphics* **17**(12), 23842391.
URL: *http://dx.doi.org/10.1109/TVCG.2011.174*

Yau, N. (2011), *Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics*, John Wiley & Sons.

yEd Graph Editor software (n.d.), http://www.yworks.com/en/products_yed_about.html. Accessed: 2014-06-24.

Yi, J. S., Kang, Y.-a., Stasko, J. T. & Jacko, J. A. (2008), Understanding and characterizing insights: How do people gain insights using information visualization?, *in* 'Proceedings of the 2008 Workshop on BEyond Time and Errors: Novel evaLuation Methods for Information Visualization', BELIV '08, ACM, New York, NY, USA, pp. 4:1–4:6.
URL: *http://doi.acm.org/10.1145/1377966.1377971*

Zacks, J. M., Tversky, B. & Iyer, G. (2001), 'Perceiving, remembering, and communicating structure in events', *Journal of Experimental Psychology: General* **130**, 29–58.

Appendix A

The Questionnaire used in the Experiments

Note: The questionnaire tasks were the same in both sets of experiments. In the within-subject experiments, each participant answered both parts of the questionnaire. In the between-subject experiments, each participant answered one part of the questionnaire based on the group allocation of the participants to the two delivery models.

Part I

(HIV Epidemic Video)

Age: ___________

Gender: M F

1. **What was the video mostly about? (Approx. 1-2 min)**

2. **Re-tell the story you gained from the video in as much detail as you can. Try to write a story that makes sense to someone who is not familiar with the story/topic. (Approx. 6-8 min)**

Turn the page for more space>>

If you need more space please use the back of this page

3. **What did you learn that you did not already know? In other words, describe the new information/knowledge you gained from the video? (Approx. 2-3 min)**

4. **Did you learn something that contradicts what you already know about the topic? <u>What is it?</u> (Approx. 2-3 min)**

5. **What do you think the speaker's purpose was in producing this video? (Approx. 2-3 min)**

Part II

(Graphical Representation of Child Mortality)

1. **What was the data you explored in Gapminder mostly about? (Approx. 1-2 min)**

2. **Re-tell the story you gained from Gapminder in as much detail as you can. Try to write a story that makes sense to someone who is not familiar with the story. (Approx. 6-8 min)**

Turn the page for more space>>

If you need more space please use the back of this page

3. **What did you learn that you did not already know? In other words, describe the new information/knowledge you gained from Gapminder? (Approx. 2-3 min)**

4. **Did you learn something that contradicts what you already know about the topic? <u>What is it?</u> (Approx. 2-3 min)**

....

....

....

5. **What do you think the purpose was in providing this data in Gapminder? (Approx. 2-3 min)**

....

....

....

Appendix B

Explanatory Statement and Consent Form for Participants

CITY UNIVERSITY LONDON

Explanatory Statement and Consent Form for Participants

Project Title: A Research Experiment on User Interpretation of Graphical Data Representation

Explanatory Statement

Purpose of the Study:

The aim of this study is to examine how the general users (non-expert users) interpret different forms of data visualization (graphical data representation), construct meaning and make sense of them.

Suitable Participants:

A suitable participant for this experiment should satisfy the following criteria:

- Participant has not previously taken Data Visualization module (IN2393).
- Participant does not have advanced knowledge in data visualization.
- Participant is not a professional data analyst.

Benefits of the experiment in general and for participants:

This study helps understanding how people interpret and comprehend data represented graphically/visually. The outcomes of this experiment contributes to the body of the knowledge by providing information on how different forms of data visualization affect the way in which people interpret and make sense of it and therefore, provide useful design guidelines of data visualization.

As the participants who have been asked to take part in this study are students in the School of Informatics who have the Data Visualization module as an option for elective modules in the next semester, this experiment is a good opportunity for them to get a flavour of what is involved in data visualization.

Experiment Procedure:

Overall, the experiment will take approximately <u>half an hour</u> distributed as follows:

8-10 min: watching a video OR exploring an interactive graphical data representation (depending on group allocation)

15-18 min: answering a questionnaire.

2-3 min: chatting/interview to get a general feedback on the experiment.

What will happen to the results of the study:

This experiment represents part of the PhD project of the researcher (see below for contact details). The results will be reported in the PhD thesis and/or any published material based on it (e.g. a journal article or conference paper). Participants will be informed about the results of the study once it is completed by e-mail.

For any inquiries about the research contact:

Donia Badawood (the researcher) PhD Student giCentre School of Informatics- City University London E-mail: [redacted]	*Prof. Jo Wood* (the supervisor) giCentre School of Informatics- City University London E-mail[redacted]

If you need to complain

If you would like to complain about any aspect of the study, City University London has established a complaints procedure via the Secretary to Senate Research Ethics Committee. To complain about the study, you need to phone 020 7040 3040. You can then ask to speak to the Secretary to Senate Research Ethics Committee and inform them that the name of the project is: *A Research Experiment on User Interpretation of Graphical Data Representation*

You could also write to the Secretary at:

Anna Ramberg
Secretary to Senate Research Ethics Committee
Research Office, E214
City University London
Northampton Square
London
EC1V 0HB
Email[redacted]

CITY UNIVERSITY LONDON

Consent Form

I agree to take part in the above City University research project. I have had the project explained to me, and I have read the Explanatory Statement, which I may keep for my records. I understand that agreeing to take part means that I am willing to:

- Use a computer to watch a video.
- Use a computer to explore some graphical data representation.
- Complete questionnaires on both the video and the interactive graphical data representation.
- Be interviewed by the researcher if needed and allow the interview to be audiotaped.

Data Protection

This information will be held and processed for the following purpose:
To explore/examine how users/audience interpret and construct meaning from different forms of graphical data representation.
I understand that any information I provide is confidential, and that no information that could lead to the identification of any individual will be disclosed in any reports on the project, or to any other party. No identifiable personal data will be collected nor published.

Withdrawal from study

I understand that my participation is voluntary, and that I can withdraw at any stage of the project without being penalised or disadvantaged in any way.

Name: ....(please print)
Signature: ....Date: ....

Appendix C

Qualtitative Coding Guidance for Coding Experiments Data

Coding Guidance

You will receive a MS Word file containing the scripts (data-stories written by the participants in response to Q2 in the questionnaire, which is to re-tell the story they gained from each delivery model). You will also receive an example of a coded answer to help you code the data. Each story/answer should be coded based on two themes, the first theme is the insight types gained, and the second is the narrative structure used. It is recommended that you code each answer based on each theme separately.

Insight Types Theme

Six insight types were identified from the literature and by going through the data in the first round of experiments. You just need to identify the types of the insights you found while going through each answer/data-story. Please note that we are only looking at the insights types gained by each participant and not the number/quantity of them as this is very difficult to judge due to the complexity of the insight types and the fact that some insights types are very interrelated and may fill into more than one type. For example, *"the Asian Countries have the highest child mortality rate while the European countries have the lowest child mortality rate"*, you can say that this is a *Tradeoff* insight comparing the maximum and minimum, and at the same time it can also be considered as *Grouping* insight as it groups Asian countries together and European countries together.

The six insight types are:

- ***General Pattern:*** the general trend or pattern of most countries. For example, the general rise or decline in an aspect, or a general relationship between two factors.
- ***Detailed Pattern:*** description of details on specific points of time or instances in general patterns. For example, the general pattern is a decline in child mortality while the detailed pattern is the average mortality rate in different periods of time to show how this happened.
- ***Outlier:*** maximum, minimum, or anything outside of the general pattern. In other words, an exception.
- ***Tradeoff:*** a combination of minimum and maximum or making comparisons between most and least in terms of one or more specific factors.

Note on the Outlier and Tradeoff insights: As a Tradeoff insight usually consists of maximum and minimum (comparison between them or between different things), we can assume that in most cases the occurrence of a Tradeoff insight implies an occurrence of Outlier insight.

- ***Grouping:*** to group different things in one category based on one or more specific criteria. In other words, define a subset or category of data. For example, Western and Eastern countries have different figures, or oil-rich and Western countries have lower child mortality, or something about developing/industrial countries as a group.

- ***Anomalies:*** the ability to identify data errors if there are any. For example, missing countries at specific period of time.

To code a given script based on the insights types theme, you need to go through each answer sentence by sentence and just put ✓ under the insight type you found in the sentence. Note that there could be more than one insight type in one sentence.

Narrative Structure Theme

The structure is simply how the story progressed; the order of sentences or information blocks in the story.

In the analysis of participants' stories, up to two structures can be recorded for each story. All stories should be assigned a *main narrative structure* that is the most appropriate one to explain the flow of narratives in the story (or the big/main blocks of the story). Some stories/answers may additionally have a *sub-structure* which is the structure used within the main story structure (inside the big/main blocks). For example, the main structure might be a chronological one but within this chronology it may demonstrate a cause and effect over a specific period of time. Each of these main and sub-structures can be assigned one of the following five types:

- ***Problem-Solution:*** emphasizing the problem and suggesting solutions whether from external information used by the narrator in the first delivery model (Hans Rosling video) or from personal knowledge about the topic either in the first or second model.

- ***General-to-Specific:*** starting from the general trend to more specific instances, details, outliers, maximum, minimum, and more specific insights and relationships.

- ***Specific-to-General:*** starting from outliers, maximum, minimum, etc. to more general trends or the bigger picture.

- ***Chronological:*** starting from past to present, present to past, or using time points/intervals to control story progression.

- ***Cause and Effect:*** describe a figure, pattern, or insight and provide the cause(s) for this insight. This structure is beyond the simple correlation between two factors (x and y-axes), and involves richer explanations of causes. For example, figures in years when there was a war or a special event in a country or an explanation of situations in Africa or a specific country that caused some patterns.

You need to take the following points into account when coding the data based on the narrative structure theme:

- In both the *general-to-specific* and *specific-to-general* structures, participants may have written a statement to conclude the overall message for delivering this kind of information. This does not change the main narrative structure in the story.

- In the *chronological* structure, participants may start with a general statement describing the aim of the visualization or the general pattern in the story as an introduction. However, the remaining details are told as change over time (note that I did not differentiate here between chronology, reverse chronology, and picking up different points of time to structure and control story progression).

To code a given script based on the narrative structure theme, you may segment the story into two or three main/big parts as suitable, and look at how the story is progressed. For example, did the second part give more specific details than the first part/segment? Did the first part explain that there is a problem and the second or the subsequent parts explain the causes or the solutions to that problem? Did the story unfold by describing the situation in different intervals/points of time?

Insight Types Theme

Put ✓ if the insight type exists in the participant's answer/data-story.

Participant #	General Pattern	Outliers	Tradeoff	Grouping	Detailed Pattern	Anomalies

Narrative Structure Theme

Remember there are five narrative structures: Problem-Solution, General-to-Specific, Specific-to-General, Chronological, Cause-and-Effect. See the Coding Guidance for more details.

Participant#	Main Narrative Structure	Sub-Structure (if applicable)

Appendix D

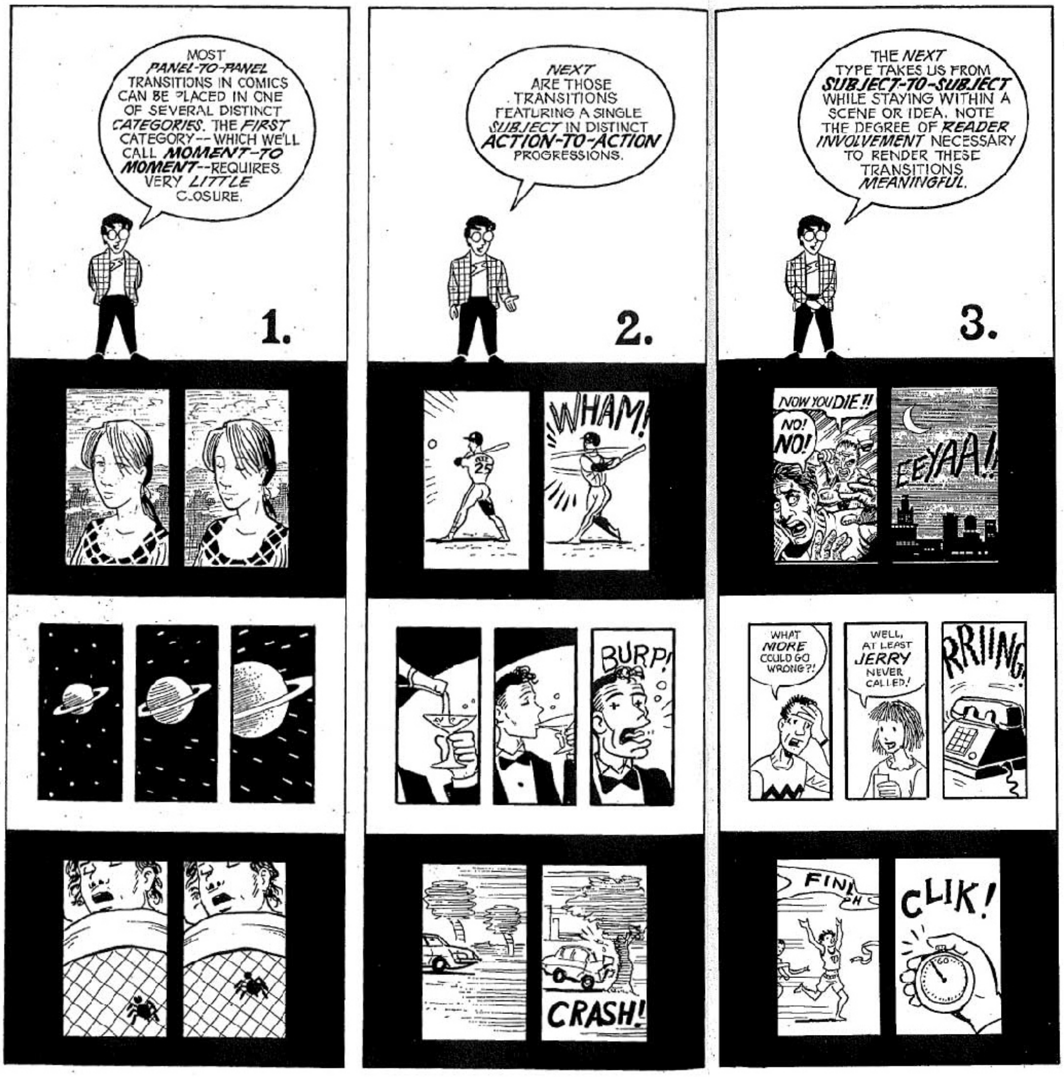
Transitions Coding Guidance

Transitions in Narrative Visualization: Coding Guidance

This guidance is produced to assess the consistency of coding transition types in narrative visualization. Before getting to the detailed instructions on the coding process, it is essential to provide some background information in order to be familiar with the context and aim of the work. The first section in this document provides a brief insight into the meaning of transitions in storytelling and summarises McCloud's taxonomy of panel-to-panel transitions in comics, which is the main taxonomy used to propose a framework of transition types in narrative visualization. The second section summarises the coding process/steps followed in general and some important terms/characteristics of data stories that you need to be aware of in order to code the data. Finally, the third section provides detailed instructions that you need to follow to code the sample data along with some examples (this is the main section that you need to use to code the sample data).

1. Brief Background

The characteristics and taxonomies of transitions in storytelling vary across domains. In comics, McCloud (1994, 2006) proposes taxonomy of transition types from one panel to another. His taxonomy consists of six types of transition: (1) Moment-to-moment: 'a single action portrayed in a series of moments'. (2) Action-to-action: 'a single subject (person, object, etc.) in a series of actions. (3) Subject-to-subject: 'a series of changing subjects within a single scene'. (4) Scene-to-scene: 'transitions across significant distance of time and/or space'. (5) Aspect-to-aspect: 'transitions from one aspect of a place, idea, or mood to another'. (6) Non-sequitur: 'a series of seemingly non-sensical, unrelated images and /or words'. Figure 1 below shows this taxonomy as represented in McCloud (1994).


MOST PANEL-TO-PANEL TRANSITIONS IN COMICS CAN BE PLACED IN ONE OF SEVERAL DISTINCT CATEGORIES. THE FIRST CATEGORY--WHICH WE'LL CALL MOMENT-TO MOMENT--REQUIRES VERY LITTLE CLOSURE.
1.
NEXT ARE THOSE TRANSITIONS FEATURING A SINGLE SUBJECT IN DISTINCT ACTION-TO-ACTION PROGRESSIONS.
2.
WHAM!
BURP!
CRASH!
THE NEXT TYPE TAKES US FROM SUBJECT-TO-SUBJECT WHILE STAYING WITHIN A SCENE OR IDEA. NOTE THE DEGREE OF READER INVOLVEMENT NECESSARY TO RENDER THESE TRANSITIONS MEANINGFUL.
3.
NOW YOU DIE!!
NO! NO!
EEYAA!!
WHAT MORE COULD GO WRONG?!
WELL, AT LEAST JERRY NEVER CALLED!
RRIING
CLIK!

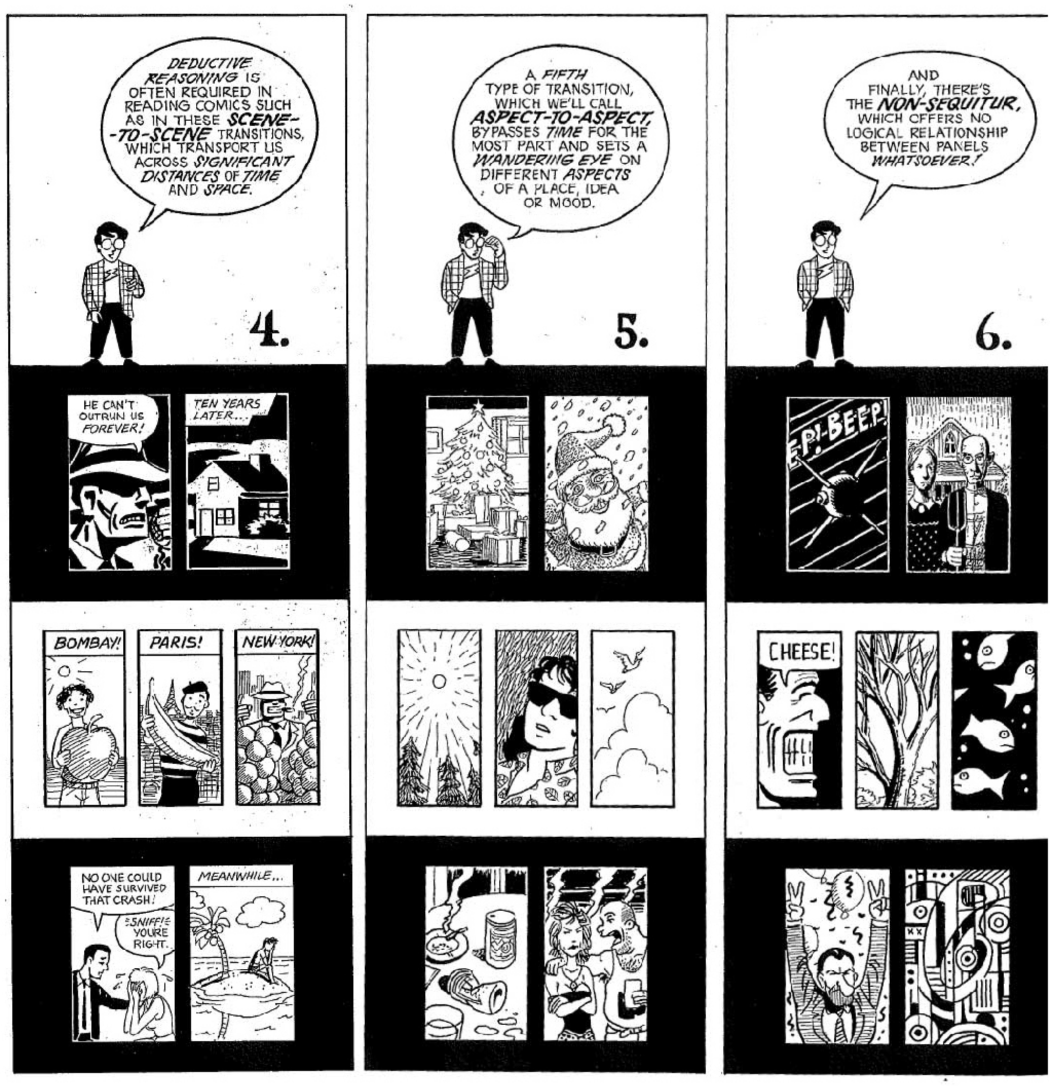


Figure 1 Panel-to-panel transitions in comics as defined and illustrated by McCloud (1994, 2006).

The research in hand will refine this taxonomy and build on it in order to:

1. Propose and develop a framework of transition types used to tell data stories through information visualization based on theories in other storytelling mediums; and
2. To propose a visual language to characterise transitions in narrative visualization based on the proposed framework of transition types.

The case study data chosen to explore transition types is the set of VAST Challenge videos from 2009 to 2011. These videos provide a valuable and suitable dataset because the videos of any specific year are based on the same dataset and address the same tasks, and they each have roughly the same duration and format, thereby providing suitable data for valid comparison of the storytelling transition types. This is also a new perspective on the issue of storytelling, as the previous work done on this area looked at samples of completely different datasets of information visualization addressing various types of data and tasks/questions.

2. Overview of the Coding Process

Before adopting the categories in McCloud's taxonomy and applying them to the VAST Challenge videos case study, this model should be refined and each type of transition should be defined in the context of data visualization. The coding steps are first summarised in this section, and then the qualitative coding tasks that you need to complete and the guidelines that you need to follow are summarised in the next section (section 3).

The coding process started with the task description of VAST Challenge. The purpose of going through the task descriptions before approaching the videos was to be familiar with the scenario and to identify an initial set of likely main components or elements of the story and how these components can be assigned a category (subjects, moments, etc.) based on the definitions of these categories (as in the next section). Thus, one will have a consistent understanding of some initial story components among the video submissions for each challenge. The next step was to watch and transcribe the videos. All video submissions to the mini-challenges from 2009 to 2011 were transcribed and coded except those submissions with broken links/URLs or where the sound/audio is incomprehensible, or when the quality of the video is poor to the degree that transcription and coding were not possible. The transcription took into account various different details in addition to speech or verbal narration, such as periods of silence/pauses to show interactions with the visualization or to change the topic/focus, the use of explicit sub-topics (usually by using a title slide) to organize different parts of the submission, such as analysis stages (for example, Data Preprocessing, Visualization, etc.) or parts of the task/question being answered (for example, MC1.1, MC1.2), and some notes about visuals used (for example, changing of views or visualization techniques). Transcribing the videos in this way made it easier and more efficient to code transitions and story development/construction using mainly these annotated transcripts. The semantics revealed by the oral narration that accompanies each video represent the progression of the data story in this type of narrative visualization (i.e., video presentations). The next step was to work through each video transcript and to identify the components of the transition types by colour coding them. As we went through each transcript, we segmented the story into smaller units. Each story unit contains a piece of information. A story unit is usually represented by a sentence or more, depending on the completion of the meaning that accompanies or follows a visual. A sentence might also be segmented into more than a single story unit. The story unit may also contain one or more of the transition components. Next, the transition type(s) between story units and/or groups of story units is identified based on what components has been changed or held constant. Finally, visually representing each story by a transition diagram.

3. Detailed Coding Instructions

In order to be able to assess the source of any inconsistency in coding transitions and creating transition diagrams, the coding of the transition types was divided into two main parts/tasks. The first part, or coding task, is to identify story units. For this part, the coder should follow the guidance and examples provided in order to segment a corpus of annotated data story transcripts into story units, taking into account the hierarchy in the data story (if applicable). The second part, or coding task, is to identify or code different components of transition types in some already-segmented corpus of data stories, judging the transition types between story units or group of story units, and then representing this coding visually in a transition diagram.

Task 1: Identifying Story Units (segmentation task)

A ‘transition’ occurs between two states or story units. Therefore, coding and identifying transition types in any type of narrative requires the definition of what is considered a ***Story Unit*** and the identification of boundaries between different story units. In some domains, identifying story units is very straightforward. In the comics domain, by which the proposed framework is inspired, and on which it is built, each individual panel is considered a story unit. Each panel consists of components that represent part of the categories/types of the transition between panels. For example, a panel showing a *subject* or character who does something (i.e., *action*), and shown from a specific angle or specific part of the room/place (i.e., *aspect*). The transition between this panel and the next depends on the changed component(s) in the next panel, and the transition is based on this change. For instance, if the character (*subject*) is changed, it will be a subject-to-subject transition, while if the same character is there but now doing something else (i.e., a different *action*), it will be an action-to-action transition. To be able to apply the same process/concept, some guidelines of what constitute a story unit when coding a VAST Challenge video are needed.

In some genres of narrative visualization, story unit identification might also be a bit straightforward. For example, in slide-show presentations, each slide can be considered, in most cases, a story unit. However, in video presentations (such as, VAST Challenges video submissions) that do not always use slides to present a data story, the task of segmenting a story into smaller units is more challenging. A review of the literature reveals that people are quite consistent in segmenting goal-driven, everyday activities/events, even with little instructions. The literature also shows, that 'breakpoints' that mark boundaries between each segment/unit/task are distinctive. Additionally, the literature reveals that people construct hierarchical structure when attempting to segment. It can be argued that data stories in the sample of VAST Challenges videos (and in most narrative visualization) are also goal-driven stories that aim to answer a particular question or deliver a particular message by going through a number of tasks or steps. A number of steps might be grouped together (on different levels) in order to achieve a specific goal or to communicate a message.

In this coding task, you will be given a corpus of data story transcripts that you will need to segment into steps or story units. Below are some cues and guidelines that will help you to accomplish this task.

***Some cues that indicate the start of a new story unit or show story unit boundaries*:**

Many of these cues are linguistic cues, or words and expressions that are commonly found to mark story units and story unit boundaries. For example,
starting a sentence or step with something like: 'Now, we will do …', usually indicates a beginning of a new story unit.

When the word 'then' is used in the beginning of a sentence ('We then …'), this indicates a start of a new story unit.

Transitional terms, such as 'First, … Second' or 'First, … Then, …', exist in some video's data stories. These are straightforward indicators of story unit boundaries.

Starting each sentence with a specific expression or style, for example, 'the analyst filter the data. The analyst represented the suspected IDs. …' indicates a new step or story unit.

Silence/pause periods while something is demonstrated or to make a transition to a new step, are also an indication of a story unit boundary. In most cases, a new story unit starts after this silence. The annotated video transcripts show when these periods of silence/pause occur.

Once a finding is started, this finding should be coded as a new story unit, even if it starts in the middle of a sentence. For example (note: the [ ] are not part of the quote, they mark the beginning and end of the story unit):
'[Using employee 30's office mates as a starting point, we will identify possible unauthorised computer use by searching for data transmissions events that occurred during classified space prox events.]
[Our search reveals an unauthorised data transmission to IP address 100.59.151.133]'
Or
'[A search around on the suspect's IP address] [reveals 18 data transmission to 12 different embassy computers.]'

Here, in the latter example, the finding starts in the middle of the sentence; therefore, we segmented it into two story units. More guidelines/cues are provided below to identify findings story units/scenes.

Sometimes colour-coding transitions components and looking at the transition types between story units may help to judge story unit boundaries. For example, if you did not find a common component(s) between two subsequent story units to make a match. (for example, to make an action-to-action or subject-to-subject transition), then reconsider your segmentation of these two story units in the transcript. You may find that you need to merge two story units together if they mostly represent the same thing, or segment a story unit into two. For this coding task, I will only look at the segmentation task in order to assess the consistency of coding. However, if you think that practice would help you perform better segmentation please do so (using the guidance provided in Task 2 below).

To identify story units in the sample provided under this step, please follow the guidelines and examples provided and use square brackets at the beginning and end of each story unit. You may also start each story unit in a new line.

<u>Some cues that help in identifying findings story units/scenes:</u>

Many of these cues are linguistic cues. i.e., words and expressions that are commonly used to indicate a finding story unit/scene. For example:

- '*Interesting enough*, employee ID#30 was responsible for ….'. The expression *interesting enough* indicates a beginning of a finding scene. *Employee ID#30* is a subject in this data story and therefore, should be represented by a green node inside the orange finding node in the transition diagram (more to follow on the visual representation).

- '*We found that …*'

- '*We notice that …*'

- '*We conclude …*'

- '*The answer found was* Nigeria B.' Nigeria B, in this example, is a subject and it should be represented by a green node inside the orange finding node.

- 'This phylogram *clearly shows* that sequence 123 is a direct descendent of sequence 583; whereas, sequence 51 is neither a direct ascendant or descendent of 583'. In this quote, the phylogram represents an aspect in this data story, while sequence 123 and sequence 51 represent subjects. Hence, these are represented by grey and green nodes inside the orange finding node in the transition diagram. In other words, this story unit as a whole is a finding scene. This finding scene consists of an aspect and subject.

- '*Hence, ….*'

- '*Here we see that …*'

- 'This search/filtering/process *reveals that …*'

- '*It is clear that …*'

In some cases, while no cue phrases that indicate a finding exist, the statement can clearly be judged as a finding. For example, 'ID30 has piggy-backed three times this month' or 'The top three mutations were …'. In this case, I recommend colour coding part of the statement in orange to make it clearer when it comes to visually representing this finding in the transition diagram.

Task 2: Identifying Story Components, transition types, and creating the visual representation using segmented data stories

In this coding task, you will be given a corpus from some annotated video transcripts. These corpus/transcripts are already segmented into story units and the high-level scenes (sub-titles) are also shown where applicable. You need to go through each story unit and identify the components of the story units, group some of these units together if needed, judge the transition types and create a transition diagram that visually represents your coding of the components, transition types, and structure/hierarchy of the data story.

The framework that you will use in coding the components and transition types is provided below. In the framework, an abstract/informative definition of each type/component of the transitions is provided in a box, while some more details and guidelines about the qualitative coding of transitions in VAST Challenge videos are provided underneath. Generally, the terms used to describe the components and types of panel-to-panel transitions by McCloud (1994, 2006) have been maintained. However, the last type of transitions, *Non-Sequitur*, is not part of the proposed framework. This is because it is assumed that there is always a relationship between the information used in any data story even through this relationship does not have to be from one piece of information or story unit to directly the next one.

You will receive a number of video transcripts to code. You are not obliged to watch the videos for the purpose of coding; the annotated scripts along with the provided guidelines should be OK. However, if you would like to have a look at the task description of any VAST Challenge, please feel free to do so.
First, you will need to go through each story unit in a given transcript and identify and colour-code the five components as follows (based on the definitions and guidelines provided below):

Subject Action Type I Action Type II Scene Aspect Moment

Then, you need to consider the structure when you start creating the transition diagram. In some transcripts, you will find this very straightforward. For example, if there is a sub-title then colour-code it in orange in the transcript (you can use different saturations of orange to show different kinds of scenes in the transcript). When you finish coding the story units under this sub-title, it will be translated into an outer scene/box that contains these story units along with any hierarchy found in them. In other cases, you just need to code and judge the transitions step-by-step from one story unit to the next. If a finding emerges after a number of story units, these story units are grouped together into a scene (a box that contains all of them), as is the finding, and a line representing the scene-to-scene transition is drawn between the scene containing the group of story units to the finding scene. You will also need to colour-code the contents of any given finding if possible. For example, if a finding is about a subject or describes a Type 2 action that has been done by a subject, you need to show these details inside the finding scene. The finding, at one point in the story, is considered a scene and its contents are represented as one story unit regardless of how complex or long this piece of finding (more details to follow on the visual representation).

The terms and categories of transition in the context of narrative data visualization are redefined as follows:

Subject and Subject-to-Subject Transition:

A subject is an entity/actor that does something (an action) in the story or that presents the main relevant concepts of the scenario. This is similar to values or a variable in statistics. An example here might be an employee, or an IP address. A subject-to-subject transition occurs from a subject or type of subjects in one story unit to another subject or type of subjects in the next story unit (if the transition is not delayed).

To identify subjects and subject-to-subject transitions in VAST Challenge video transcripts, you can use the following guidelines:

The noun/verb analysis may sometimes be helpful in identifying subjects, which are usually nouns or noun expressions, and actions, which are usually verbs.

Going from whole to part or from a set to a sub-set is considered a subject-to-subject transition. For example, if story unit 1 involves 'seeing *all ID records*' and story unit 2 involves 'examine *16-source IDs only*', it is said that there is a subject-to-subject transition between these two story units.

Some subjects may result from a process or an action that was done to another subject. This is similar to the previous point on going from whole to part, but the part in this case is not a direct sub-set from the whole, it is derived by performing an action/process on the whole. For example, story unit 1 might be 'I generated a report of … of *all IPs*', and story unit 2 '*The destination IP which had the largest … 37.170.100.75*'. In this case, there is a subject-to-subject transition between these two story units.

In the video transcripts (the oral narration), nouns or noun expressions referring to the analysts are *not* considered as subjects (for example '*I/we* did …'). This is to avoid confusion when coming to the visual representation of these components. If these are considered as subjects and the scenario/domain principles are also subjects, it will not be appropriate to use the same descriptive term for them, nor to use the same visual encoding. Also the nouns that refer to the person who did the analysis/interactions are always followed by some sort of actions (verbs), which are represented by a different type of component/transition. This will also makes it possible and easier to apply the framework to other types of narrative visualization where there is no oral narration provided, and to compare the construction and progression of data-stories between these different narrative genres.

Examples of subjects might be events, activities, sequences, people, devices/computers, messages, cities, area, a node(s) in a network of people etc., if these are what constitute the main concepts of the scenario. In many cases, you will find that subjects are countable. Additionally, subjects might be referred to by a pronoun.

The whole dataset provided for a given VAST Challenge is not a subject on its own. For example, 'we loaded *the data* onto our system' or '*the challenge data* is downloaded'. However, something like '*the remaining data …*' or a defined set from the data is considered a subject (for example, 'First, we looked at *the IP data*').

When the same subject is found in two subsequent story units, there is obviously no subject-to-subject transition between these units. However, in some cases a subject may exist in two story units and at the same time we can have a subject-to-subject transition. This can be demonstrated in comic panels. If we have two comic panels and a subject exist in both of them but a new subject is introduced in the second panel (while the first is still there), then we have a subject-to-subject transition while retaining the first subject. In this case, the focus changes from the first subject to the second subject (the accompanying text may help to show this change, and in the same way the annotations or oral narration can show this change in narrative visualization).

Action and Action-to-Action Transition:

Here, we have differentiated between two types of actions. Type 1 re- lates to how the participant/analyst(s) interacted with the data/tool, for example, 'filter out IP addresses', or decisions or tasks/processes made with the data, for example, 'we now identify who is the suspicious per- son'. Type 2 actions represent part of the data/scenario. For example, 'employee A left his office'. An action-to-action transition occurs from an action in one story unit to another in the next story unit. Sometimes the action-to-action transition is delayed by two or more story units, and sometimes the series of actions continues after reporting or revealing a piece of finding between them.

To identify actions and action-to-action transitions in VAST Challenge video transcripts, you can use the following guidelines:

As stated above, noun and verb analysis of the annotated video transcripts is useful when identifying subjects and actions. Actions of type 1 are usually represented by verbs such as 'filter', 'compare', 'investigate', and 'select'. Actions of type 2 usually accompany a subject and sometimes also a moment (definition below); examples could be something like, 'The handyman *sent a text message*' or 'Employee a *closed the door and turn on the computer*'.

In the language of data visualization, action type 1 is any visualization task done by interacting with the visualization or examining it.

There is a difference between action type 2 and some action expressions used to describe a subject. For example, 'employee with ID40 *contacted* the employee with ID30', is an action type 2; however, 'now we will examine *IDs who are making more than 100 calls*' is not an action type 2, it is a subset or group of subjects (i.e., a subject).

In many cases, after a finding is revealed, a new story unit with new action(s) starts. In this case, there is a delayed action-to-action transition from the story unit before the finding and the story unit after the finding.

In some cases, the analyst does not explicitly say that he or she does something. For example, 'when the syndrome chart *is arranged* by number of deaths, …'. This should be coded in the same way as 'we *arranged* the syndrome chart by number of deaths …' because even though the analyst does not say that he or she does something or interacts with the tool/data, he or she obviously did.

Scenes and Scene-to-Scene Transition:

Scenes in different visual storytelling mediums in general (for example, comics and films) provide some kind of compression, conclusion or grouping based on an idea such as time or place. For example, in films, a scene is a complete unit that usually consists of a series of shots that form a coherent unit of the narration. Therefore, if the story in a given VAST Challenge video is divided explicitly into a number of sub-topics (usually represented by a title slide), each sub-topic is considered a scene. A scene-to-scene transition thus occurs between any two sub-topics, which in this case is a high-level transition (i.e., a series of different types of transitions may occur between smaller story units within each of those scenes). A group of story units that lead to a finding is also considered a scene, as is the finding or conclusion. A scene-to-scene transition in this case occurs between a number of steps or story units that are grouped based on a common idea/topic and the finding or conclusion drawn from this group.

To identify scenes and scene-to-scene transitions in VAST Challenge video transcripts, you can use the following guidelines:

As stated above, if the story in a given VAST Challenge video is explicitly divided into sub-topics, which are usually represented by a title slide (this will be clear in the transcript of the story), each sub-topic is considered a scene, and a scene-to-scene transition occurs between any two sub-topics.

If the story in a given VAST Challenge video is not explicitly divided into sub-topics by title slides, but the narration show a clear high-level segmentation of the story (for example, narration of the analysis and findings drawn using first tool A, then the analysis and findings drawn using tool B), each segment is also considered as a scene.

Groupings of several story units will always be considered a scene, regardless of the theme or component on which the grouping is based. For example, a story may consist of a number of sub-topics, each of which communicates the analysis process and the findings using a different visualization technique/application (i.e., an *aspect*). One may think about grouping the story units into aspects rather than scenes. However, this would complicate the process and sometimes makes it confusing when creating the transition diagram, especially if the grouping can be judged based on more than one type of transition component. Grouping into scenes suggests a common idea or theme between a number of story units and simplifies the presentation of scene-to-scene transitions between a group of story units and a finding scene in the transition diagram.

As stated above, there are some expressions/terms that are commonly used to report/narrate a finding, such as 'Hence' and 'concluded'. A list of these has been produced.

Aspect and Aspect-to-Aspect Transition:

An aspect-to-aspect transition occurs when the applications, views, techniques, or visual representations are changed from one story unit to another in order to look at the data from different angles (similar to having a 'wandering eye' around a place or object in comics (McCloud 1994). For example, this may occur between a map view and a timeline.

To identify aspects and aspect-to-aspect transitions in VAST Challenge video transcripts, you can use the following guidelines:

Any description of a visual encoding/representation is colour-coded as an aspect (for example, '*red dots* represent poor countries'), but if we moved from something like '*industrialised countries represented by green bubbles*' in one story unit to '*developing countries represented by red bubbles*' in another story unit, this would not be considered an aspect-to-aspect transition if the two story units were both addressing the same graph. On the other hand, going from one graph to another graph (even from the same type e.g., from a bar chart to another bar chart), one view to another, or one visualization technique to another to investigate something else is considered an aspect-to-aspect transition. This explains why in some transition diagrams we might find two story units with an aspect component in each of them but no aspect-to-aspect transition between them.

We can have an aspect as a result of an action on a different aspect (in other words, an action triggers the aspect-to-aspect transition).

Sometimes a story (or part of it) starts with a general description of the visualization tool or application used. This application may have two or more views and the first statement starts with a description of both views (here both views are from the same level and are components of the interface or application; therefore, the statement as a whole is considered as one story unit). Then, when a specific step/investigation is carried out using one of these views, an aspect-to-aspect transition will occur from the visualization tool as a whole to one of its components/views. The same may occur between different components of a specific view. For example, if the interface has a chart view that consists of a number of charts, transitions between these charts are considered aspect-to-aspect transitions even though they might be from the same type (e.g., bar charts) but involve addressing or looking at different parts of the screen or looking at something else (just like camera movement).

Moment and Moment-to-Moment Transition:

A moment is a point of time in the story/data scenario. For example, 'at 10am on 17th May, ...'. A moment-to-moment transition occurs between two story units if each story unit addresses a different point of time.

To identify moments and moment-to-moment transitions in VAST Challenge video transcripts, you can use the following guidelines:

Expressions or use of time as an estimation of the analysts/participants' effort spent on a specific task is not considered as a moment. For example, 'we analysed the data *for four hours*', or 'we built a visualization tool *in six hours*'.

Moments sometimes will be very specific. For example, 'employee A was in the room from 1 p.m. to 3 p.m. on 15th May' or 'the first week of August'. In other cases, it might not be as explicit. Instead, it may refer to a relative or a period in the data scenario. For example, 'we address/compared all events *before and after the truck accident happens*'.

Creating the Visual Representation (transition diagram) based on coding:

Before describing the visual representation of the components and transition types used in each VAST Challenge video (i.e., the transition diagram), it is important to note that you do not have to approach the coding of a data story and creation of a transition diagram as completely two separate steps/tasks. They can be done simultaneously by working back and forth between them and represent each story unit visually before continue coding the next one. In some cases, this might be easier and more effective. For example, when there are many findings revealed throughout the story and you need to group the story units before each finding together.

The transitions used to tell a data story in each VAST Challenge video (or in a single data story in general) are fused into a single diagram. This is called a ***transition diagram***.

Each type of the transition components (subject, action, etc.) is represented by a small colour-coded node/cell. The same colour codes used in identifying these components in the videos' annotated transcripts should also be used in the diagram.

The nodes representing components of one story unit should be positioned vertically on top of each other.

By looking at the transition(s) made, you may draw a line from the corresponding cell in the first story unit from which the transition was made to the cell in the destination story unit (this will mostly be the next story unit, but in some cases it could be delayed by one or more story units).

The design also differentiates the finding scenes with bold borders. Also, no transition can be made from a finding or its contents. A finding always ends the scene (i.e, the story unit(s) after the finding will always be represented in a new scene or orange box).

The visual encoding of each transition diagram is summarised in Figure 2. Rendered using this visual encoding, a completed transition diagram would look like those provided in the examples in section 4 in this document.

I used yEd Graph Editor software (http://www.yworks.com/en/products/yfiles/yed/) to create the hierarchical transition diagrams. It is an open source software, and works on both Windows and Mac OS. You can use the same software or any other software to create similar diagrams. You can also simply use MS Word to create the visual representation.

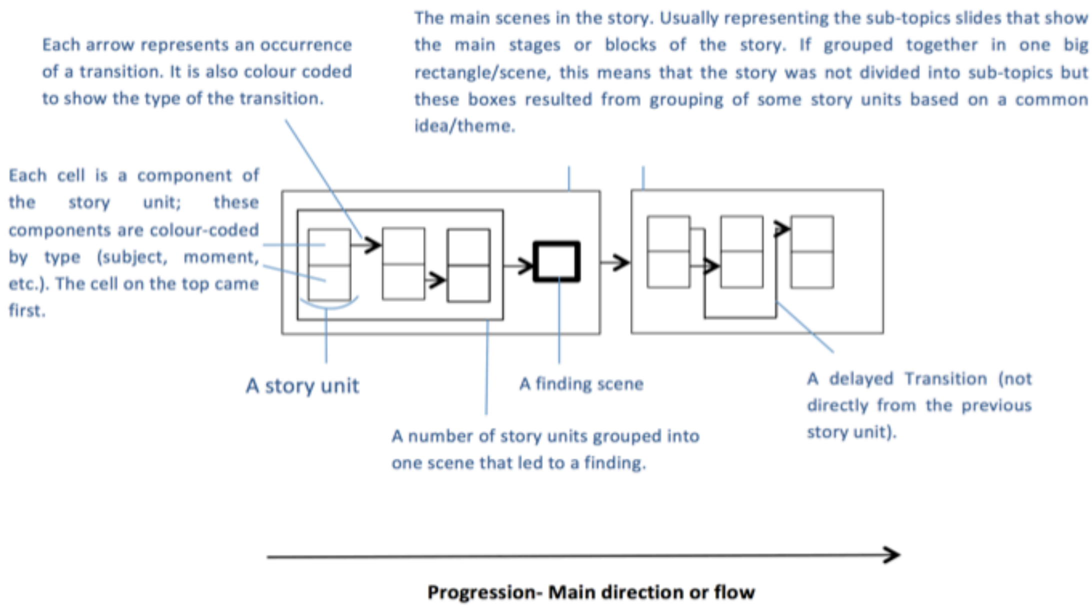


Figure 2 The visual encoding of the story components and transition types in each VAST Challenge video.

Notes on the delayed transitions:

Delayed transitions usually occur from a story unit before a finding and a story unit after that finding (to continue the story after the finding). Remember, no transition can be made from the content of the finding or the story unit inside the finding scene.

The most common transition types that occur from the story unit before the finding to the story unit after the finding are: action-to-action, and aspect-to-aspect. Sometimes there might also be a subject-to-subject transition. Therefore, watch out for these transitions if you have a finding story unit in the middle of the narratives.

Sometimes a story unit establishes an analysis process (that may last for a few subsequent story units) using a visualization tool/view. After a few story units, the analysts may transform to another view. In this case, a delayed aspect-to-aspect transition occurs from the first story unit to the story unit where there was a shift or change in view/tool.

General notes on the transition diagram and coding of transition types:

When a new sub-topic starts, the previous sub-topic (scene) is closed and a scene-to-scene transition between the two sub-topics is drawn. In some cases, you may feel like there is a relationship between the last story unit under the first sub-topic and the first story unit in the second sub-topic. To avoid confusion, we do not show these transitions if there is a higher-level scene-to-scene transition between two sub-topics.

Multiple Transitions may occur between two story units. For example, if a story unit consists of a subject (a data point/variable), an action done on this subject (e.g., filtering or calculating something), an aspect (a visual representation to look at this subject, e.g., a timeline or bar chart), and, in the next story unit another subject is addressed and a different action is performed. In this case, we have two types of transitions between these two story units: a subject-to-subject and action-to-action.

If only one story unit precedes a finding story unit/scene, we put this story unit inside a scene (an orange node) and a scene-to-scene transition is made between the scene containing the story unit and the finding scene.

If there is, for example, three subjects in one story unit, they are still represented by one green node in the transition diagram. The same applies to all other transition components (actions, moments, etc.).

When a new sub-topic starts, the previous sub-topic (scene) is closed and a scene-to-scene transition between the two sub-topics is drawn. In some cases, you may feel like there is a relationship between the last story unit under the first sub-topic and the first story unit in the second sub-topic. To avoid confusion, we do not show these transitions if there is a higher-level scene-to-scene transition between two sub-topics.

3. Worked Examples

Example 1

The Transcript

For further analysis we selected the symptoms with the highest frequencies among the death records. Additionally we selected the symptoms with high frequencies among all records that had not been selected before. We used the same routine to match the selected frequent symptoms to the text describing the syndromes and thereby transformed the text in to feature factors. We applied the projection to the feature factors in order to obtain a two-dimensional spatial layout over the syndromes. Then we created a map of syndromes to be used for utilisation.

1. Segmenting the transcript into story units

[For further analysis we selected the symptoms with the highest frequencies among the death records].

[Additionally we selected the symptoms with high frequencies among all records that had not been selected before].

[We used the same routine to match the selected frequent symptoms to the text describing the syndromes and thereby transformed the text in to feature factors].

[We applied the projection to the feature factors in order to obtain a two-dimensional spatial layout over the syndromes].

[Then we created a map of syndromes to be used for utilisation].

Notes:

Each story unit is shown between [] and starts on a new line for simplicity.

You may have noticed that starting each step in the analysis process or in the story with '*We….*' was the main linguistic cue that helps in segmenting this story into story units or steps. There are other cues, such as '*Additionally*, …..', and '*Then we* created….'.

2. *Colour-coding the components of transition types*

[For further analysis we selected the symptoms with the highest frequencies among the death records].

[Additionally we selected the symptoms with high frequencies among all records that had not been selected before].

[We used the same routine to match the selected frequent symptoms to the text describing the syndromes and thereby transformed the text in to feature factors].

[We applied the projection to the feature factors in order to obtain a two-dimensional spatial layout over the syndromes].

[Then created a map of syndromes to be used for utilisation].

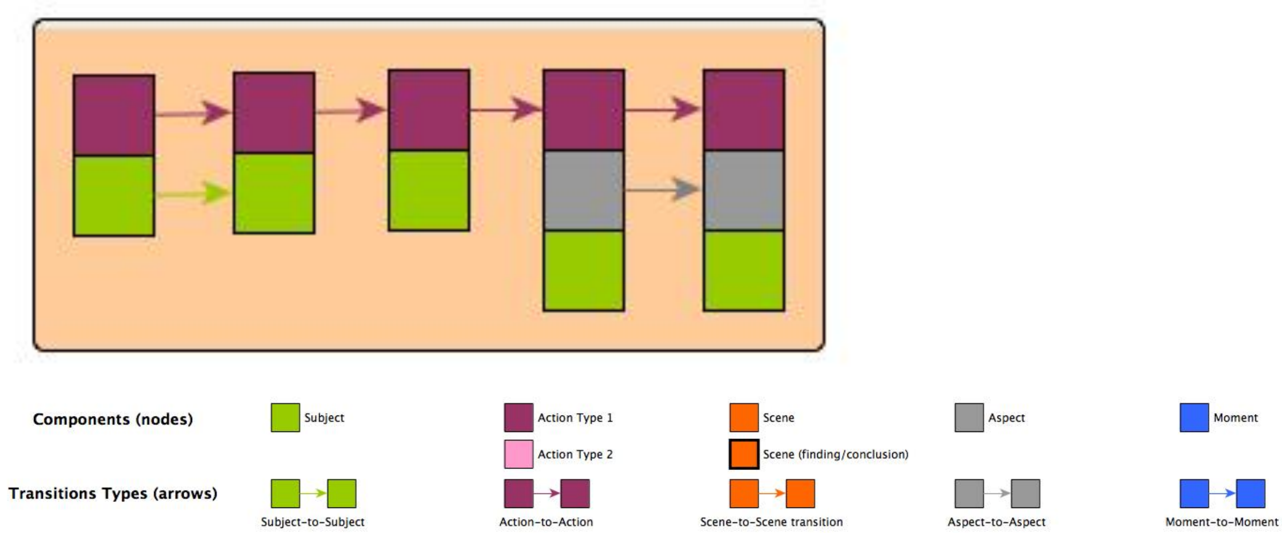
Notes:

Subjects here are the syndromes, symptoms of the disease (note that all symptoms are different than symptoms with high frequencies, they are two different subjects), and death records that refer to some patients.

Actions were all of type I (which is more common than Actions of Type 2 in the whole sample).

The noun/verb analysis is very helpful in identifying subjects and actions (not only in this example but in the whole dataset).

3. *Judging the transition types and creating the transition diagram*



Notes:

There are five story units. As there is no sub-topics used, all story units are contained in one high-level scene.

Example 2

The Transcript

Sub-title: Determine Spread

To investigate how the disease has spread, we analyse how the movement of people before and after they got sick. In the time graph representation of the trajectories, we compute the temporal distance to the first flu event. We then use this attribute (they mean the flu) to filter out data and show the part of the trajectories before and after this first occurrence of the event. We clearly see that many people were at the centre of Vastopolis before getting sick. After getting sick, people move in all districts and many visits to the hospitals seen at the yellow dots in the display.

1. Segmenting the transcript into story units

Sub-title: Determine Spread

[To investigate how the disease has spread, we analyse how the movement of people before and after they got sick.]

[In the time graph representation of the trajectories, we compute the temporal distance to the first flu event.]

[We then use this attribute (they mean the flu) to filter out data and show the part of the trajectories before and after this first occurrence of the event.]

[We clearly see that many people were at the centre of Vastopolis before getting sick. After getting sick, people move in all districts and many visits to the hospitals seen at the yellow dots in the display.]

Notes:

There are 4 story units under the sub-title. We started the segmentation after the sub-title. At this stage, sub-titles are not part of the segmentation task (they are already segmented as high-level scenes). They will be more important when colour-coding and representing them in a transition diagram.

2. ***Colour-coding the components of the transition types***

Sub-title: Determine Spread

[To investigate how the disease has spread, we analyse how the movement of people before and after they got sick.]
[In the time graph representation of the trajectories, we compute the temporal distance to the first flu event.]
[We then use this attribute probably they mean the flu to filter out data and show the part of the trajectories before and after this first occurrence of the event.]
[We clearly see that many people were at the centre of Vastopolis before getting sick. After getting sick, people move in all districts and many visits to the hospitals seen at the yellow dots in the display.]

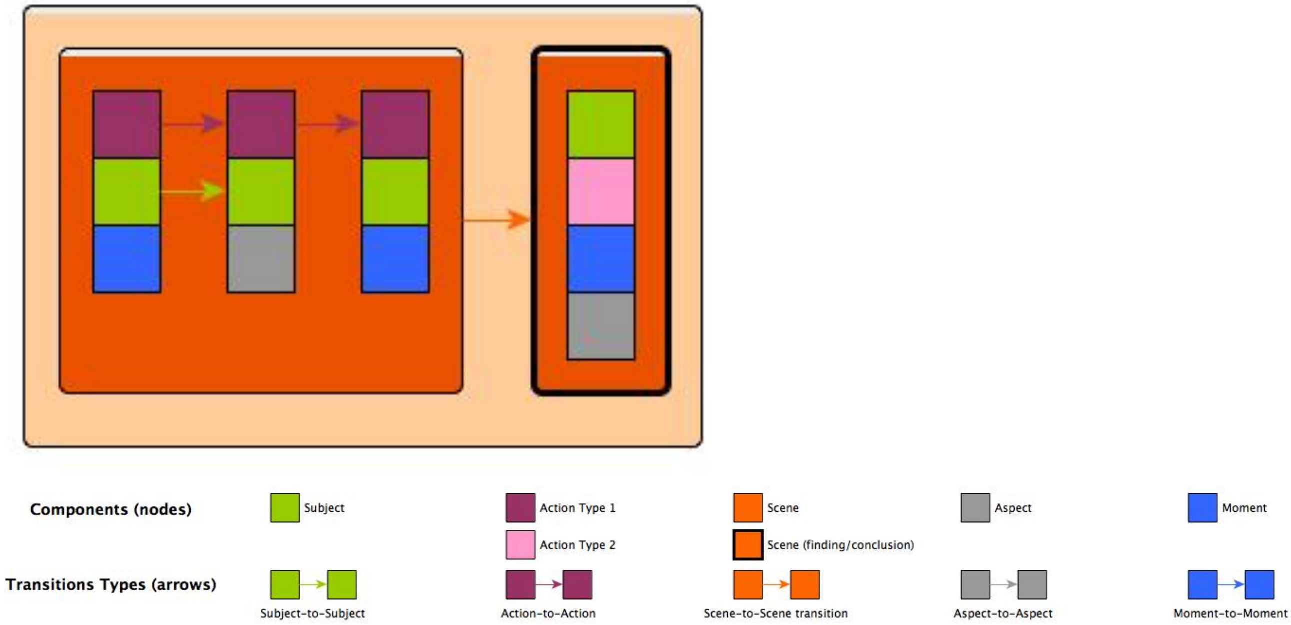
Notes

Before and after getting sick are both representing times in the scenario. So, they are coded as *moments*.

The term 'We clearly see that…' was colour-coded in orange to make it clear that this is a finding scene. Colour-coding the pharses that tell you that a story unit is a finding scene will help when creating the transition diagram.

Note the different shades of orange to show different levels of scenes (finding scenes are more saturated/darker and the explicit sub-topics are coded in slightly lighter orange).

3. Judging the transition types and creating the transition diagram



Notes:

Note the order of nodes in the second story units is changed in the diagram to avoid clutter (crosses of transition lines).
In a single story unit, if we have more than one subject or more than one action (or any other component), we only use a single node to represent the existence of this component in the story unit. For example, in the third story unit, we have 'filter out' and 'show', which are actions. In this case we only use one action node when representing this story unit visually in the transition diagram.

Note the grouping of the story units before the finding into one scene, and the coding/representation of the contents of the finding scene.

Remember that there is no transition type other than scene-to-scene transition can be made *to* a finding scene. Also, if we have more story units after the finding scene, we would have a delayed transition from the story unit before the finding to the story unit after the finding (but not from or to the components of the finding).

Example 3

The Transcript

Subtitle: Characterizing Suspicious Usage

To characterise the behaviour of suspicious computer use, the analyst uses Geotime link analysis to isolate all events that were directly to the suspicious IP. Using the charts feature, the analyst is able to chart events based on many characteristics. By charting only events visible to the viewer, the analyst can see that no computer was used more than 3 times to send information to the suspicious IP. By charting day of week the analyst can see that events only occurred on Tuesdays and Thursdays. Applying colour highlights this pattern. When charting by hour, the analyst observes that data was only sent between 8am and 6pm. In calendar view the analyst observes that employee 35 consistently comes in between 8am and 6pm, correlating with the suspicious IP traffic.

1. Segmenting the transcript into story units

Subtitle: Characterizing Suspicious Usage

[To characterise the behaviour of suspicious computer use, the analyst uses Geotime link analysis to isolate all events that were directly to the suspicious IP].

[Using the charts feature, the analyst is able to chart events based on many characteristics].

[By charting only events visible to the viewer],

[the analyst can see that no computer was used more than 3 times to send information to the suspicious IP].

[By charting day of week]

[the analyst can see that events only occurred on Tuesdays and Thursdays].

[Applying colour highlights this pattern. When charting by hour],

[the analyst observes that data was only sent between 8am and 6pm. In calendar view the analyst observes that employee 35 consistently comes in between 8am and 6pm, correlating with the suspicious IP traffic.]

Notes:

2. ***Colour-coding the components of the transition types***

Subtitle: Characterizing Suspicious Usage

[To characterise the behaviour of suspicious computer use, the analyst uses Geotime link analysis to isolate all events that were directly to the suspicious IP].

[Using the charts feature, the analyst is able to chart events based on many characteristics].

[By charting only events visible to the viewer],

[the analyst can see that no computer was used more than 3 times to send information to the suspicious IP].

[By charting day of week]

[the analyst can see that events only occurred on Tuesdays and Thursdays].

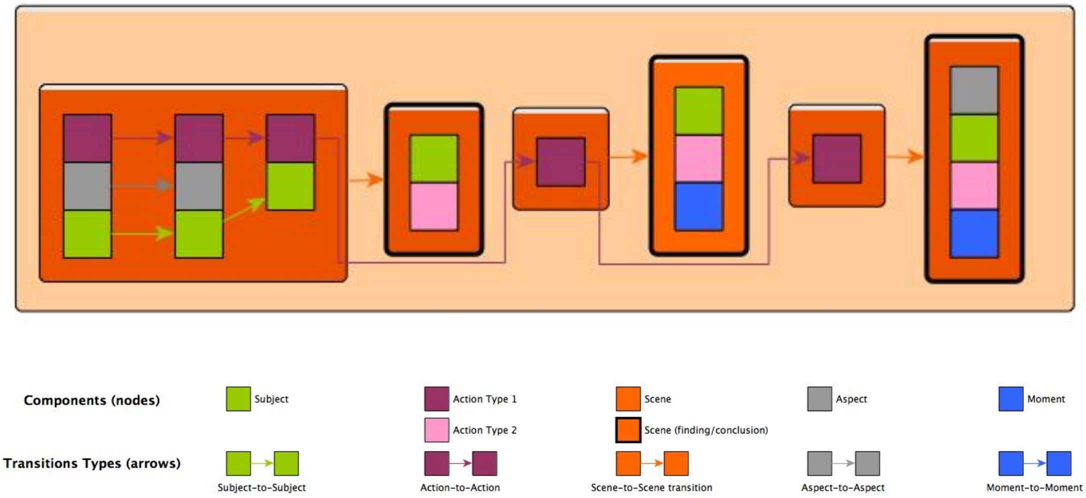
[Applying colour highlights this pattern. When charting by hour],

[the analyst observes that data was only sent between 8am and 6pm. In calendar view the analyst observes that employee 35 consistently comes in between 8am and 6pm, correlating with the suspicious IP traffic.]

Notes:

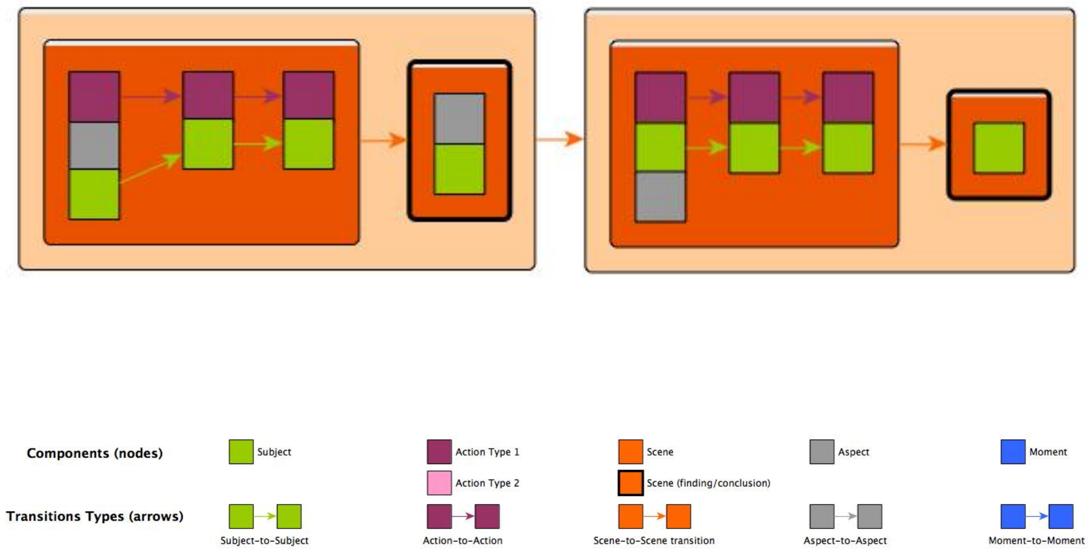
Note the finding scene at the end, that consists of more than one piece of finding or more than one statement. This is still coded as one story unit.

3. Judging the transition types and creating the transition diagram



Example 4

Here only a diagram representing the coding of a data story that has two explicit sub-topics. It is provided to show how we close the first sub-topic (scene) when a new one starts and how we draw a scene-to-scene transition between the two sub-topics. In some cases, you may feel like there is a relationship between the last story unit under the first sub-topic and the first story unit in the second sub-topic. To avoid confusion, we do not show these transitions if there is a higher-level scene-to-scene transition between two sub-topics. Additionally, when the story is explicitly divided into sub-topics, we do not draw an outer box/scene that contain all the sub-topics.



Appendix E

The Transitions Diagrams for the Dataset of VAST Challenge Videos for 2009, 2010, and 2011

E.1 The Transitions Diagrams for VAST Challenge 2009 Videos

E.1.1 VAST Challenge 2009- MC1

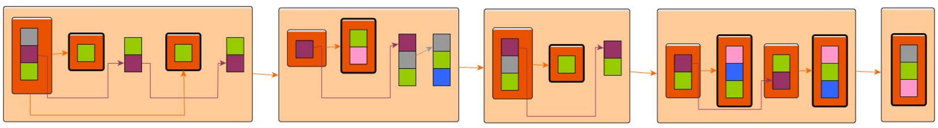


Figure E.1: GAMI- International Institute of Information Technologies

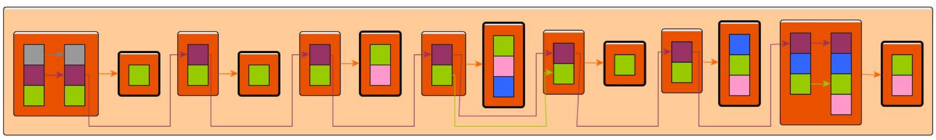


Figure E.2: Georgia Institute of Technology

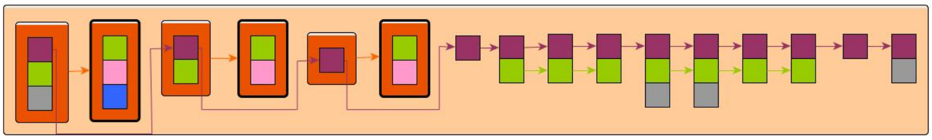


Figure E.3: Leonard- EAKOS 2009

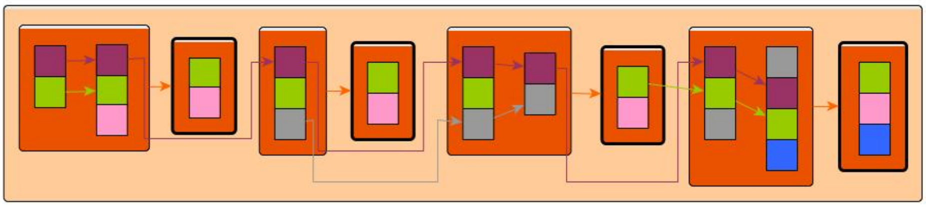


Figure E.4: University of Maryland- Mindlab

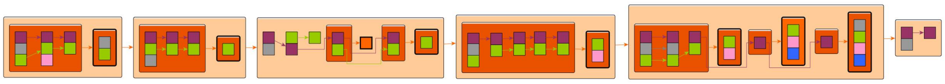


Figure E.5: Oculus Info Inc

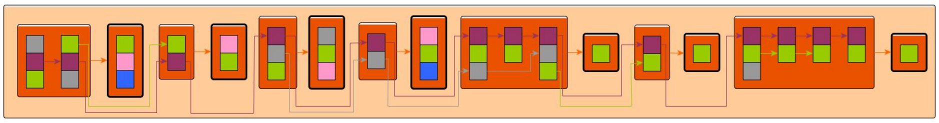


Figure E.6: Palantir Technologies

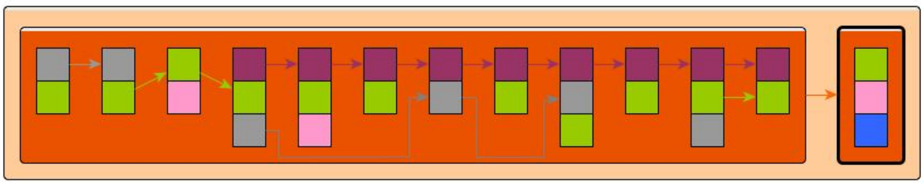


Figure E.7: VIDI Surveilance

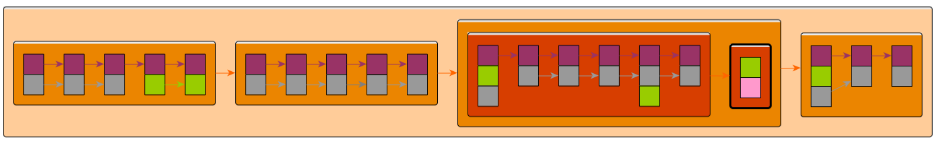


Figure E.8: Virginia Tech

E.1.2 VAST Challenge 2009- MC2

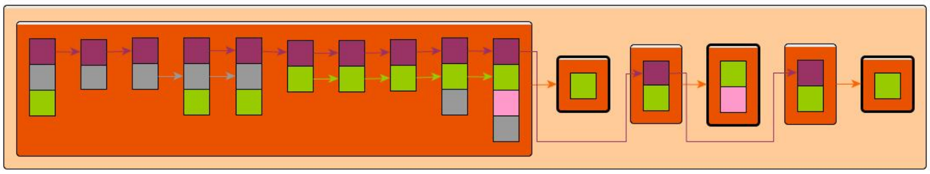


Figure E.9: University of Ulm and derive- VIScover

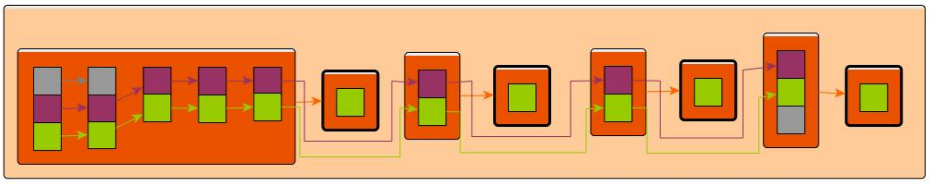


Figure E.10: Georgia Institute of Technology

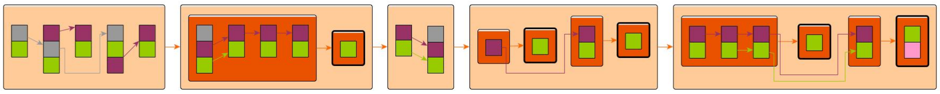


Figure E.11: giCentre- City University London

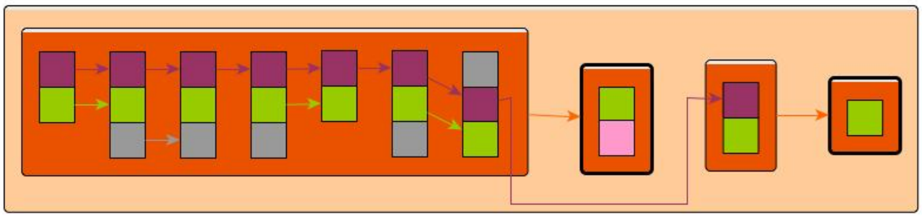


Figure E.12: University of Michigan- Guava

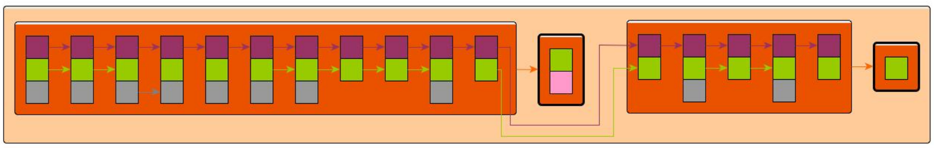


Figure E.13: LaBRI, INRIA Bordeaux Sud- Quest

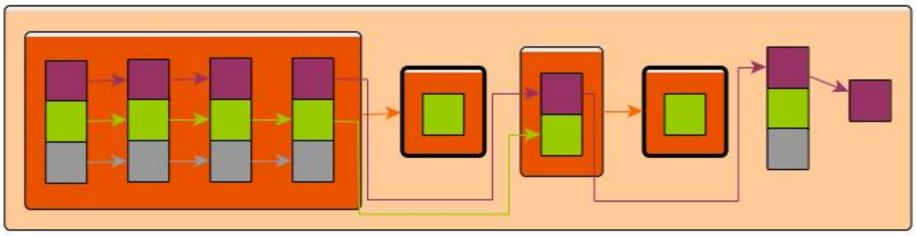


Figure E.14: Leonard- EAKOS 2009

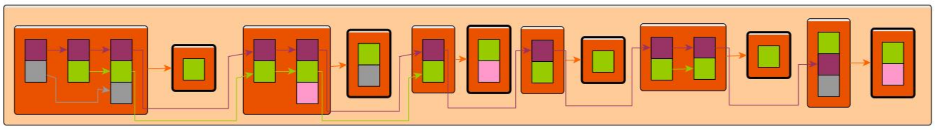


Figure E.15: Palantir Technologies

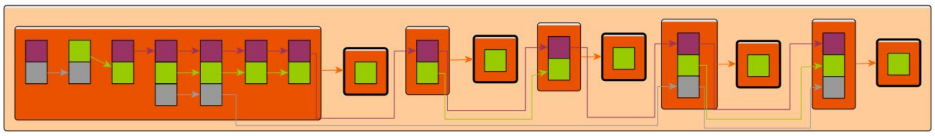


Figure E.16: VIS(US) Stuttgart

E.1.3 VAST Challenge 2009- MC3

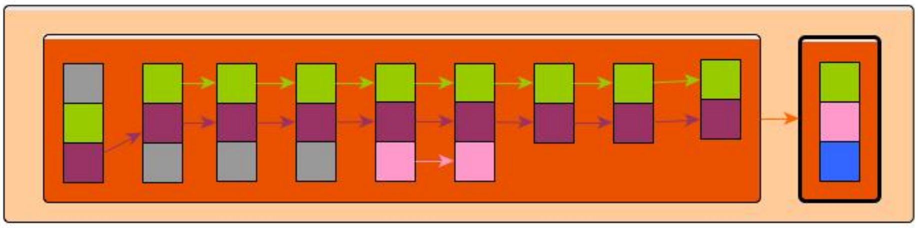


Figure E.17: LaBRI, INRIA Bordeaux Sud-Ouest

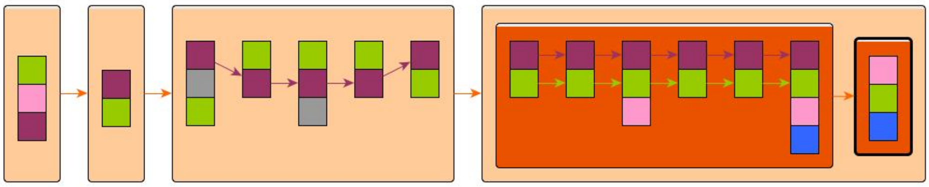


Figure E.18: US-Spray3D, Universitat Stuttgart

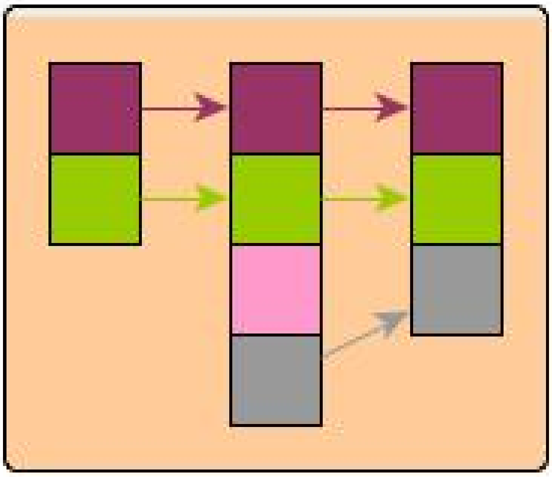


Figure E.19: Leonard- EAKOS 2009

E.2 The Transitions Diagrams for VAST Challenge 2010 Videos

E.2.1 VAST Challenge 2010- MC1

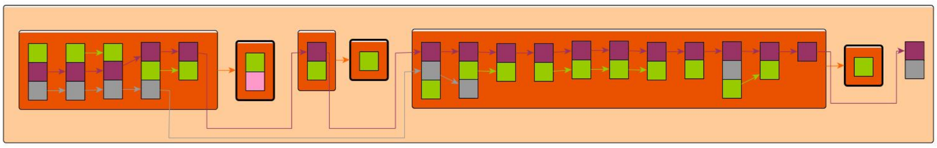


Figure E.20: Xanalys Ltd

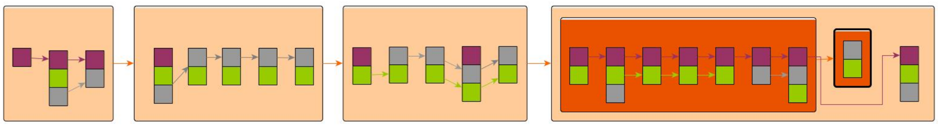


Figure E.21: VRVis-ComVis

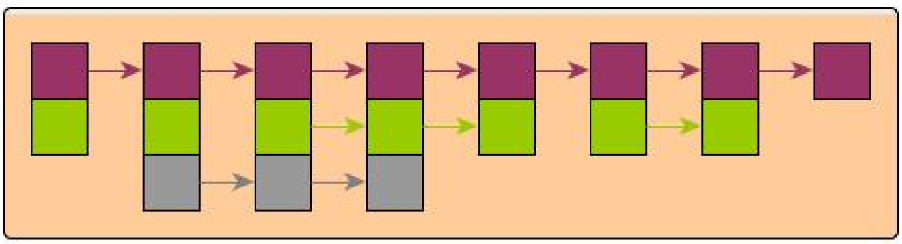


Figure E.22: giCentre

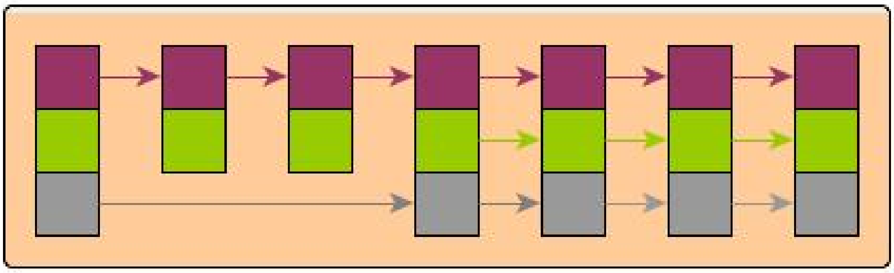


Figure E.23: Georgia Institute of Technology- Jigsaw

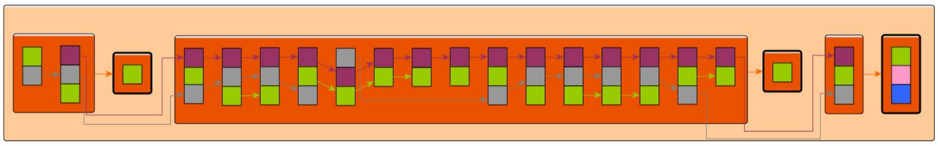


Figure E.24: Palantir Technologies

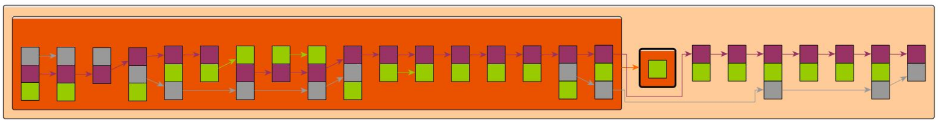


Figure E.25: Simon Fraser University

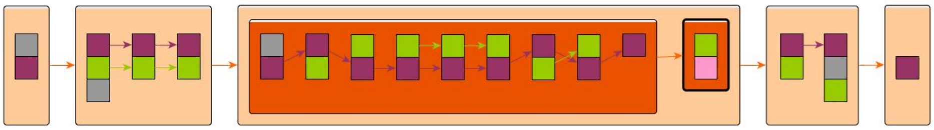


Figure E.26: Virginia Tech

E.2.2 VAST Challenge 2010- MC2

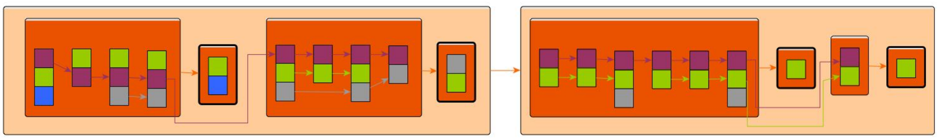


Figure E.27: VRVis- ComVis

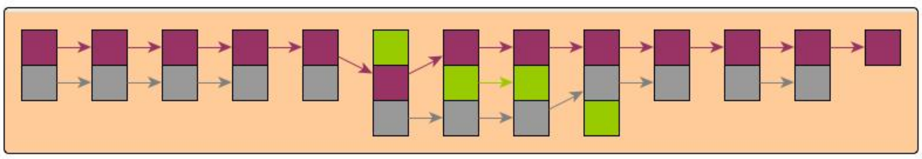


Figure E.28: Stottler Henke

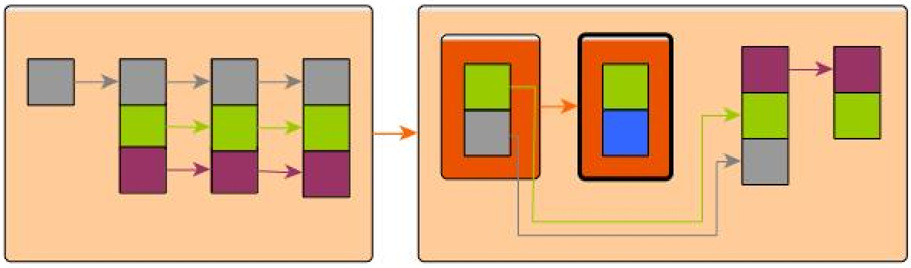


Figure E.29: giCentre- PandemView

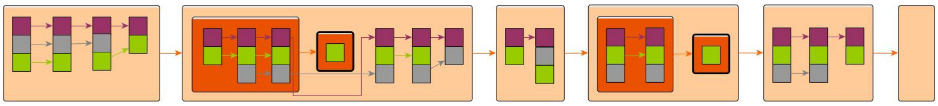


Figure E.30: Bangor University

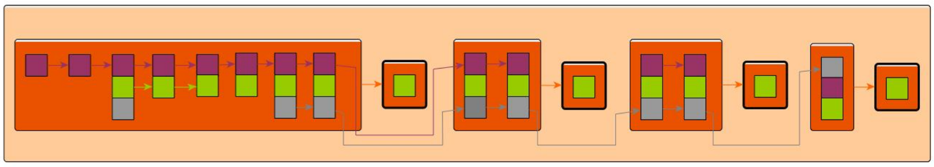


Figure E.31: University of Constance- Applied Visual Analytics

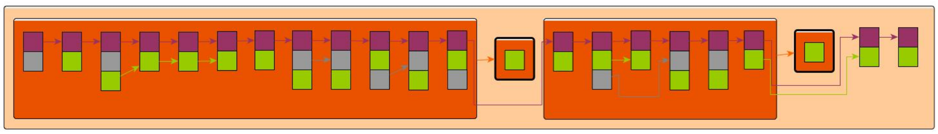


Figure E.32: Fraunhofer Institute IAIS

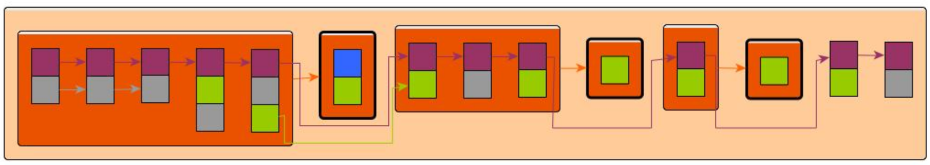


Figure E.33: Rice University

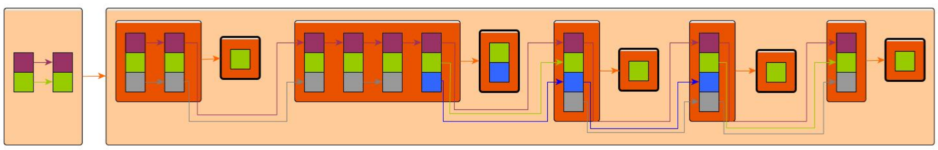


Figure E.34: College of Engineering Pune

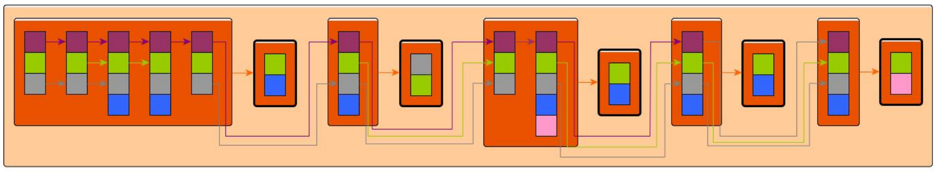


Figure E.35: Purdue University

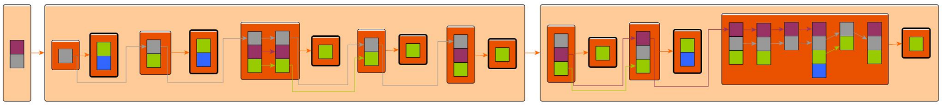


Figure E.36: PennState

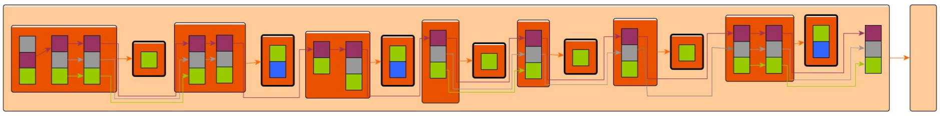


Figure E.37: Palantir Technologies

E.2.3 VAST Challenge 2010- MC3

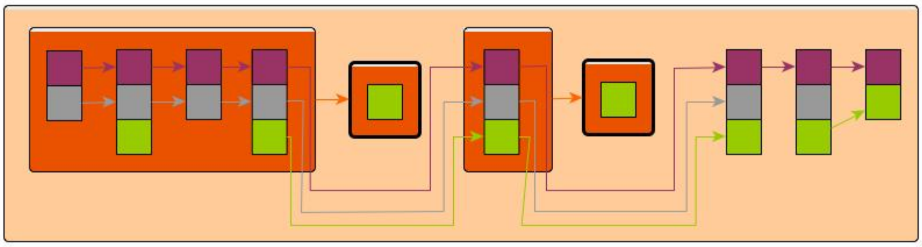


Figure E.38: Stottler Henke

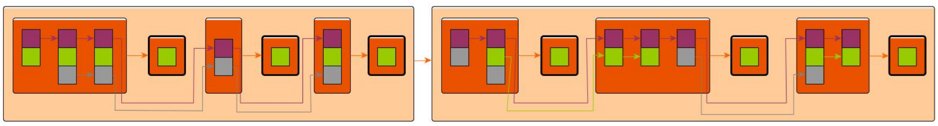


Figure E.39: Universidad Autonoma de Madrid and University of Maryland

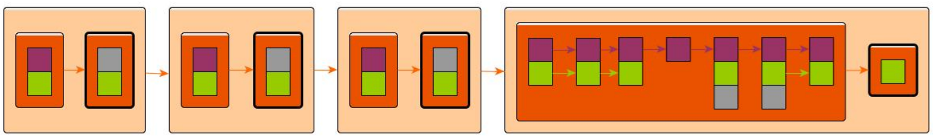


Figure E.40: University of Konstanz

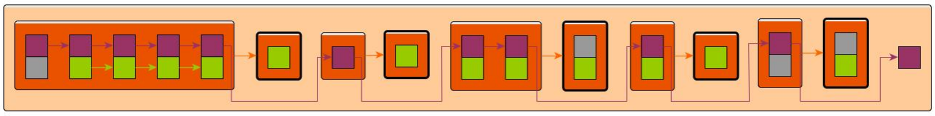


Figure E.41: giCentre- SequenceView

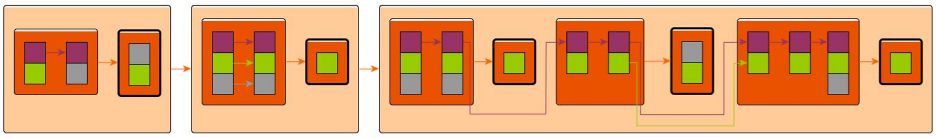


Figure E.42: Team Stuttgart and Tubingen

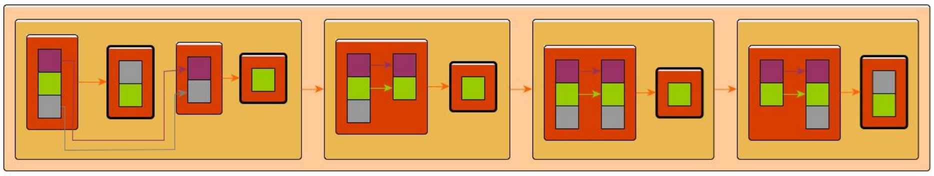


Figure E.43: CERTH/ITI

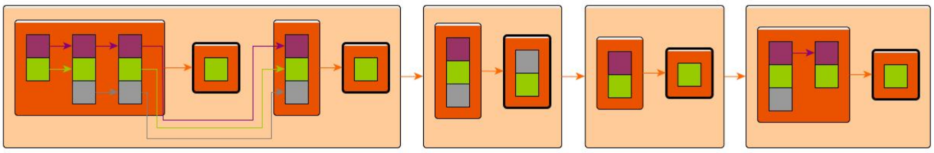


Figure E.44: Middlesex University

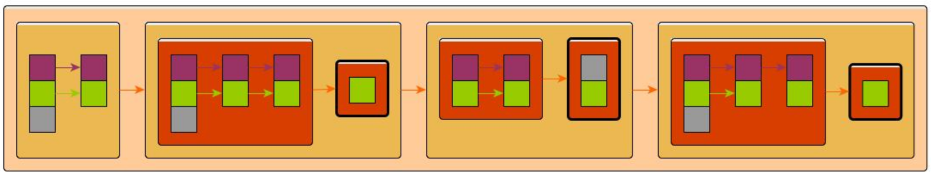


Figure E.45: Georgia Institute of Technology

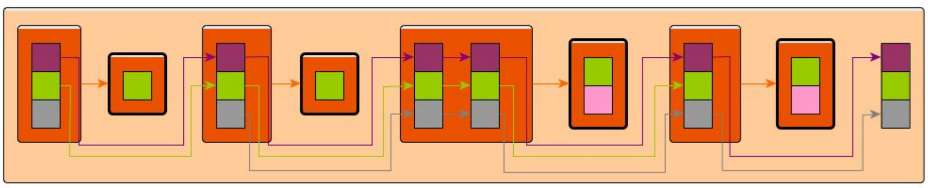


Figure E.46: UrsaManor

E.3 The Transitions Diagrams for VAST Challenge 2011 Videos

E.3.1 VAST Challenge 2011- MC1

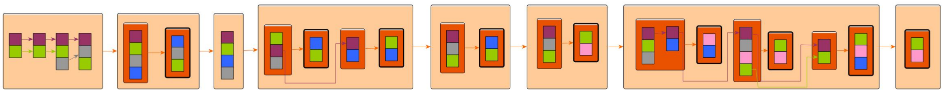


Figure E.47: University of Konstanz

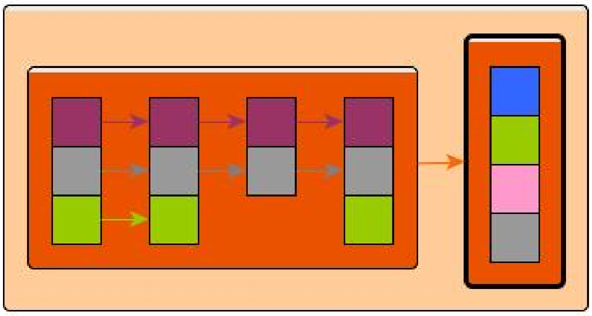


Figure E.48: UBA- Python

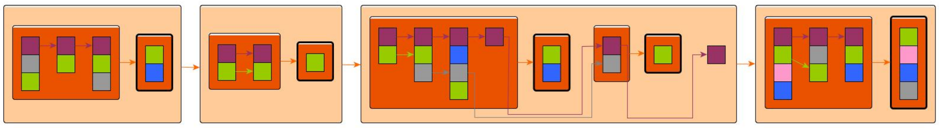


Figure E.49: Fraunhofer IAIS- V-Analytics

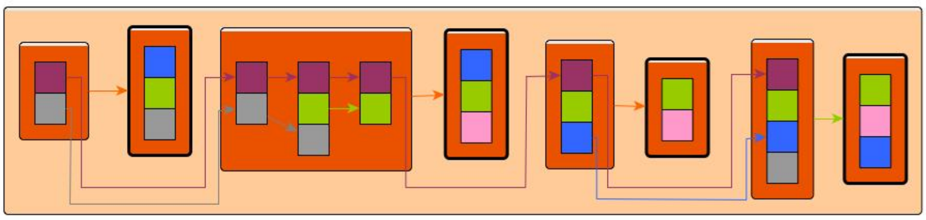


Figure E.50: Purdue University

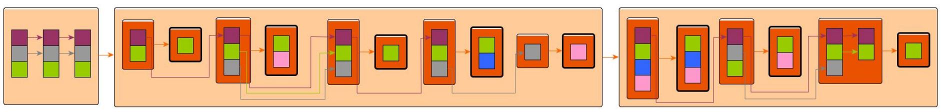


Figure E.51: VIS Stuttgart- ScatterBlogs

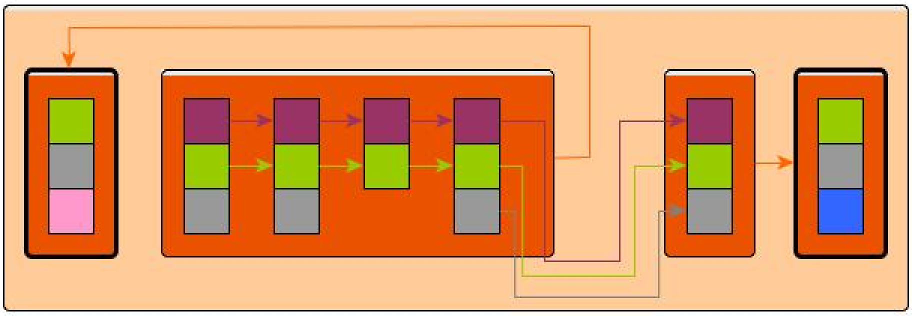


Figure E.56: Pennsylvania State University

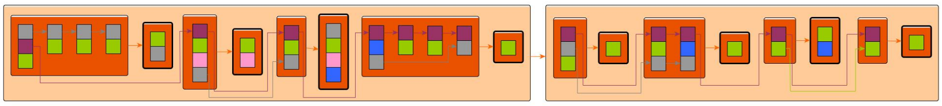


Figure E.52: giCentre, City University London- BlogViewer

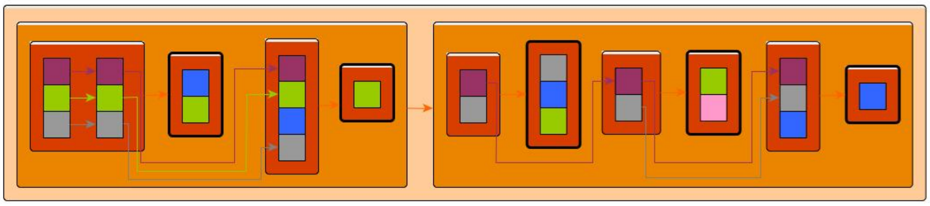


Figure E.53: SAS

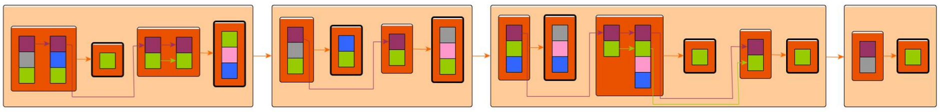


Figure E.54: Bangor University- epSpread

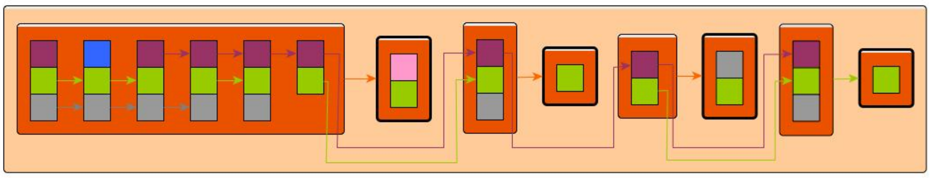


Figure E.55: MTA-SZTAKI

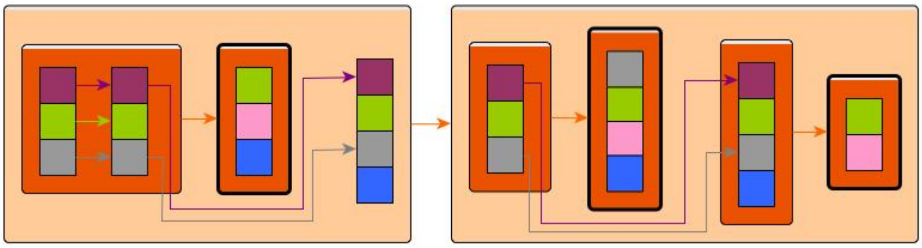


Figure E.57: UNCC

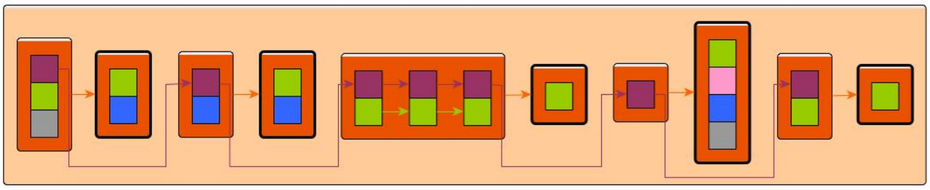


Figure E.58: Middlesex University and University of Leeds

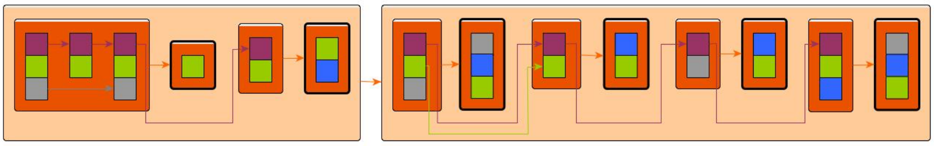


Figure E.59: University of Maryland

E.3.2 VAST Challenge 2011- MC2

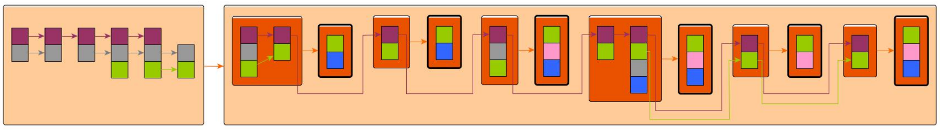


Figure E.60: University of Konstanz

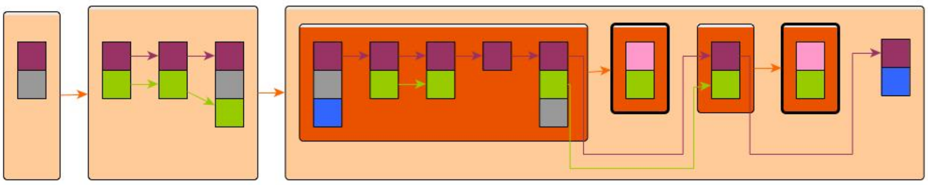


Figure E.61: CMU/IBM- Network Security and Anomaly Visualization

E.3.3 VAST Challenge 2011- MC3

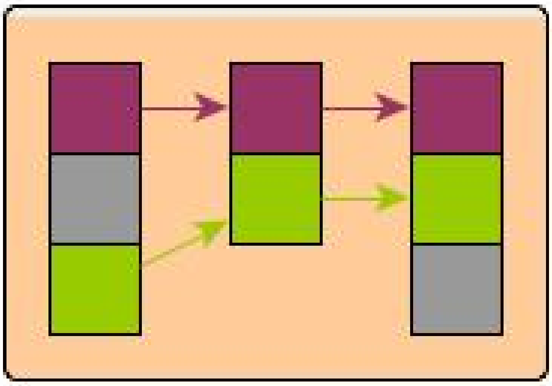


Figure E.62: Old Dominion University

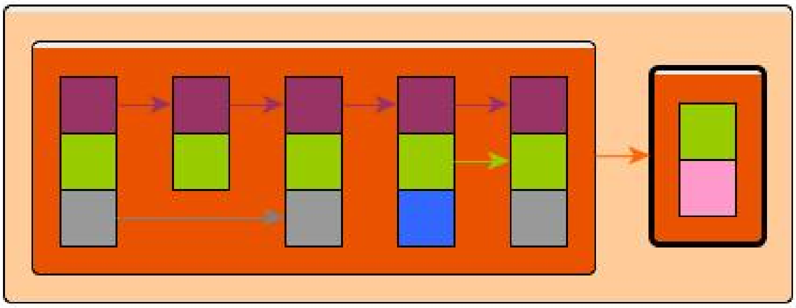


Figure E.63: SAS

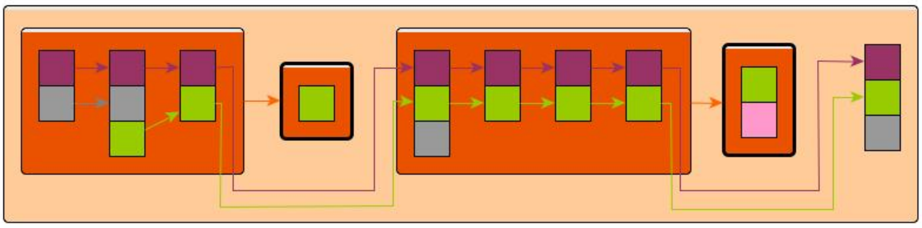


Figure E.64: VSTI

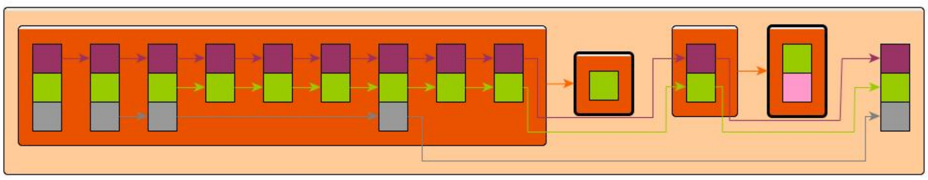


Figure E.65: Georgia Tech- Jigsaw

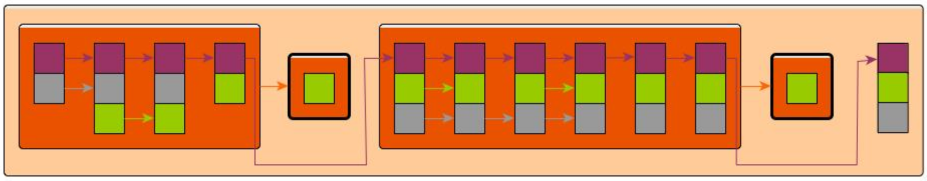


Figure E.66: Oculus Info. Inc.- nSpace2

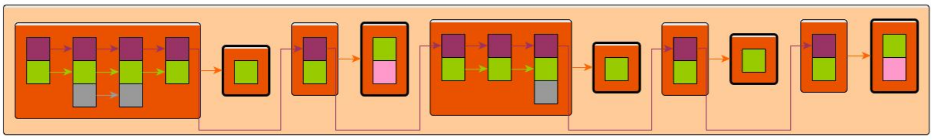


Figure E.67: Middlesex University/ University of Leeds

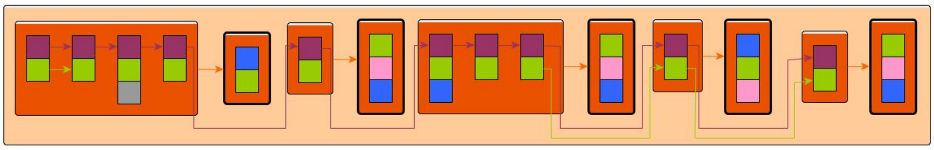


Figure E.68: UNCC

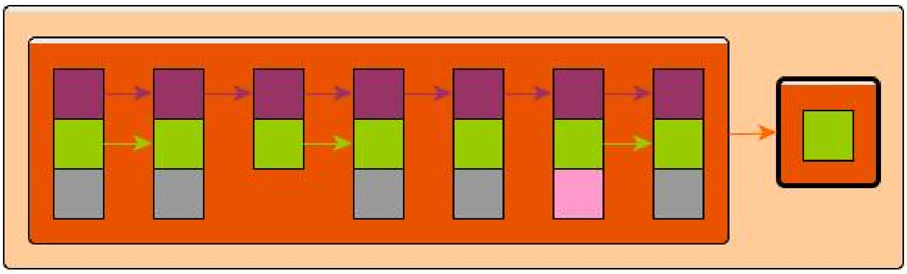


Figure E.69: University of Maryland