Human creativity in the data visualisation pipeline
A systematic literature review

Coral Featherstone[1] and Etienne Van der Poel[2]
[1]*Meraka Institute, CSIR, Gauteng, South Africa*
[2]*University of South Africa, Gauteng, South Africa*
cfeatherstone@csir.co.za, evdpoel@unisa.ac.za

Keywords: Creativity, Data Visualisation, Data Visualization, Visualisation Pipeline

Abstract: There are some aspects of visualisation that are uniquely human. This is because data visualisation is heavily influenced by the science of human visual perception, much of which emerged from the *Gestalt School of Psychology*. Humans can recognise why certain aspects of the data matter, are aware of background information and use intuition, purpose and storytelling when choosing what to visualise. Intuition and visual perception are used in an iterative manner to craft the final visualisation. Whilst computer algorithms can create visualisations from data in a brute force combinatorial manner, humans are still much better at quickly determining what context and which aspects of the data, at what granularity, will successfully highlight what the data represents. Visual analytics, that combines machine learning, graphic user-interfaces and human interaction, is a popular way of addressing the shortcomings of fully automated computer generated visualisations. This paper is part of a larger project that will be exploring the development of a non-interactive computational algorithm that enhances the process of computer produced visualisations by introducing criteria and techniques from the theories of computational creativity, which is sub-field within the artificial intelligence domain. One of the objectives of the larger project is to identify the parts of the visualisation pipeline and also to identify what aspects of visualisation generation process humans are better at than computers – specifically with respect to human creativity. This literature review aims to address these identification objectives by means of a critical review of the parts of the visualisation pipeline.

1 INTRODUCTION

This review attempts to identify – by means of a critical literature review – the process that occur between the point where the data to visualise is supplied, up until the point where a visualisation has been created from the data, in order to identify where in the visualisation pipeline, human creativity and other distinctly human traits occur. The following questions are investigated:

1. Which stages in the visualisation pipeline require human creativity?
2. How are computers facilitating creativity in the visualisation pipeline?
3. Which stages in the visualisation pipeline are affected by storytelling and purpose?

The literature is reviewed from three angles. The visualisation process that a human would follow is investigated first. Thereafter, the visualisation process that a computer algorithm or machine learning solution would follow is investigated. Finally, the visualisation pipeline of an area of research known is *visual analytics* is explored. Visual analytics is a method of automating visualisations with the help of computers, while keeping humans involved in the process. The alteration in the pipeline between human generated, computer generated and visual analytics is then compared to identify the primary differences between the same task executed with and without a human.

Various books and papers have different terminology that is being used to describe the same process. The words *pipeline* (Wickham et al., 2009; North et al., 2009; Wu et al., 2014; Chi and Riedl, 1998), *data flow* (Muehlenhaus, 2012; Chi and Riedl, 1998) and *visualisation process* (Wickham et al., 2009; Yau, 2013) appear to be used interchangeably. The intended discussion concerns the activities that occur between the point where the data is supplied, up until the point where a visualisation has been created. The term *visualisation pipeline* will be used here.

Inclusion criteria	Exclusion criteria
Mention of any human factors	Anything restricted to a specific domain (such as healthcare or genetics)
Mention of human visual perception	Literature only describing types of visualisations
Mention of human inference or intuition	Internet of things, sensor data or geographic information systems
Mention of human creativity	Graphic software, drawing or art unrelated to computers
Mention of automated generation by computers	3D Modelling or Image processing
Mention of formal grammars	Marketing, presentations or adverts
Mention of storytelling	Cloud computing

Table 1: Inclusion and exclusion criteria

2 METHODOLOGY

A critical literature review, focusing on the data visualisation pipeline will be performed with the intention of identifying potential locations within the visualisation pipeline that are still best performed by a human. As per the recommended procedure for this style of research (Fink, 2013; Kitchenham, 2004; Pickering and Byrne, 2014), the search procedure (Section 2.1), keywords (Tables 2, 3 and 4), inclusion and exclusion criteria (Section 2.2) as well as screening criteria (Section 2.3) will be described, before the literature is presented. Attributes used to screen the literature for appropriateness are laid out in Section 2.3. An indication of the most relevant literature is presented alongside the keywords in Tables 2, 3 and 4 and is presented in Section 3. Lastly, Section 4 summarises how the identified pipelines differ and in which stages of the pipeline, human creativity, cognition and human traits are most evident.

2.1 Search procedure

The starting point for the literature survey on the visualisation pipeline was a search of electronic books. The academic library, *Ebrary* was used as a starting point for the e-book search (ProQuest, 2016). The relevance of the books were identified from the inclusion and exclusion criteria. They were also ranked as less or more relevant by how many of the topics, related to the keywords, were covered. Once relevant books were identified, additional key phrases, citations and frequently cited authors were used to extend the search onto *Google Scholar* and references from highly relevant literature was followed, forward and backward in time. Both American and UK English spelling was used for the word "*Visualisation*". The keywords were chosen in order to identify literature specifically discussing the visualisation pipeline and excluding literature that was focussed too narrowly on one specific visualisation problem or tool. *Big data* and *data analytics* were included because visualisation is one of the tools used to investigate unstructured data since it is powerful at helping with the problem of data overload since human vision is adept at detecting patterns (Myatt and Johnson, 2011; Simon, 2014). Storytelling was included because of the number of books mentioning the importance of a visualisation's purpose and storyline (Katz, 2012; Kirk, 2012; Kosslyn, 2006; Meirelles, 2013; Pereira, 2007).

The keywords used to extend the literature search to computer automated visualisations and visual analytics are presented in Tables 3 and 4.

Exclusion criteria were designed to eliminate advanced visualisations such as those specific to 3D modelling, sensor data or geographic information systems. Visualisation that focussed on a particular domain's data were also excluded because the described pipelines are frequently specific to that data set.

2.2 Inclusion and exclusion criteria

Once potentially pertinent literature had been identified, they were once again assessed in order to isolate those describing the following things: the visualisation pipeline, computer generated visualisation, tasks during visualisation that are difficult for a computer to automate, how visualisations relate to human visual perception, and the role of human inference with regard to generating effective visualisations.

Specific criteria were used to further refine the results thereby refining what results were included. These inclusion and exclusion criteria are listed in Table 1.

Several visualisation tools are built on formal grammars known as a context free grammars. These grammars identify specific processes used to generate visualisations. Any literature mentioning grammars was also prioritised (Bostock and Heer, 2009; Wickham, 2010; Wilkinson, 2006).

2.3 Literature screening criteria

The following features were considered when prioritising and screening the literature. Literature mentioning the inclusion criteria in the title, chapters or sections. Multiple pages or entire chapters dedicated to the desired criteria or coverage of more than one of the desired criteria. Newer literature was prioritised;

Keyword/phrases	Most relevant References
Data visualisation *and* Visualisation pipeline	(Dos Santos and Brodlie, 2004; Keim et al., 2008; Myatt and Johnson, 2011)
Infographics	(Alexander et al., 2014; Justin Beegel, 2014; Lankow et al., 2012; Simon, 2014; Krum, 2013)
Infographics *and* Visualisation process	(Alexander et al., 2014; Byrne et al., 2016; Spence, 2001; Wickham et al., 2009)
Infographics *and* Storytelling	(Alexander et al., 2014; Baer and Vacarra, 2008; Kosara and Mackinlay, 2013; Krum, 2013; Lankow et al., 2012)
Infographics *and* Storytelling *and* Creativity *and* Visual perception	(Kirk, 2012)
Visualisation pipeline	(North et al., 2009)
Visualisation process	(Alexander et al., 2014; Chi, 2000; Muehlenhaus, 2012; Teller, 2013; Ware, 2004; Yau, 2013; Zhu, 2013; Jankun-Kelly et al., 2002)
Visualisation process *and* Storytelling *and* Creativity	(Gershon and Page, 2001; Ware, 2004)
Narrative visualisation *and* Storytelling	(Figueiras, 2014; Hullman and Diakopoulos, 2011; Khataei and Lau, 2013; Segel and Heer, 2010)
Visualisation data pipeline	(Chi, 2000; Deshpande et al., 2015; North et al., 2009)

Table 2: Keywords and phrases used in the initial human only visualisation pipeline search and the most relevant references

Keyword/phrases	Most relevant References
Automatic generation information graphics presentation	(Mackinlay, 1986)
Grammar of graphics	(Wilkinson, 2006; Wickham, 2010)
Information visualisation *and* Grammar	(Bostock and Heer, 2009; Conti, 2007; Meirelles, 2013; Myatt and Johnson, 2009; Myatt and Johnson, 2011; Redström et al., 2000; Ware, 2004; Wilkinson et al., 2000; Yau, 2013; Zhu, 2013)
Automated visualisation	(Gahegan, 1999; Tatu et al., 2009; Sun et al., 2009; Hooi-Ten Wong and Ramadass, 2010; Tatu et al., 2009; Wills and Wilkinson, 2010; Hooi-Ten Wong and Ramadass, 2010)
Visualization system	(Chi and Riedl, 1998)
Visualization system *and* automated	(Cho et al., 2016; Ryabinin and Chuprina, 2015; Hanna, 2015)
Machine learning visualisation pipeline	(Bouali et al., 2015; van der Maaten and Hinton, 2008)

Table 3: Keywords and phrases used in the automated computer visualisation pipeline review

however, older literature was not excluded, since the visualisation process prior to visual analytics is relevant.

3 LITERATURE

The literature is reviewed in three parts. Section 3.1 discusses the basic visualisation pipeline that a human would follow when designing a visualisation from a provided set of data. Specific attention is given to where in this process, human visual perception and human understanding occur. Iteration tied to purpose and storytelling, emerge as distinctly human aspects of this process and they are discussed in more detail.

Section 3.2 reviews automated visualisation tasks relevant for machine learning or computer algorithms. Some of the tools used to generate visualisations are mentioned. A formalised model of the visualisation pipeline is discussed, since it formalises the visualisation pipeline into a reproducible model, and the shortcomings of that model are presented. Techniques of producing visualisations, with no human involvement, are not very prevalent in the literature using the search keywords chosen.

The final section (Section 3.3) focusses on computer applications that incorporate a human in the process of producing a visualisation. Specific focus is given to the visual analytic pipeline, which facilitates this human/computer collaboration. Also discussed is how these computer applications foster human creativity.

Out of scope in this discourse is the exploration of knowledge versus information in visualisations (Chen et al., 2009), the controversial dispute on when a visualisation becomes art (Kosara, 2007; Yau, 2013), what makes any visualisation better that any other (since there is more than one way to visualise the same data (Yau, 2013), or reviews of the types of visualisation (Cleveland, 1985). Specific detail of individual parts of the process, such as statistical techniques, how to clean the data, how to convert from data to meta-data, and choose the scale and axis are not explored – but are described very briefly for context.

3.1 Human only visualisation pipeline

There are some aspects of visualisation that are uniquely human. This is because data visualisation is heavily influenced by the science of human visual perception, much of which emerged from the Gestalt School of Psychology (Kirk, 2012). The *Gestalt laws* – such as the *Law of Similarity*, *Law of Closure* and the *Law of Proximity* – emerged from studies of how our brains form a global sense of pattern. Critical to this understanding is the fact that our visual perception is faster and more efficient than our cognitive processes. Human eyes can process information in parallel (Wang et al., 2016). Exploiting these features of visual perception has significant influence on how well a visualisation is interpreted (Kirk, 2012; Kosslyn, 2006; Meirelles, 2013).

The data is transformed, in stages, until it can

Keyword/phrases	Most relevant References
Interactive visualisation pipeline	(Bavoil et al., 2005)
Visual analytics *and* Analytic provenance	(Sacha et al., 2016b; Rodrigues-Jr et al., 2015)
Sense making cycle	(Bradel et al., 2015; Erwin et al., 2015; Heer and Agrawala, 2008; Sacha et al., 2016c; Wang et al., 2016; Lee et al., 2016; Ottley, 2016; Reda et al., 2016; Bradel et al., 2015)
Visual analytics process	(Grammel et al., 2010; Shrinivasan and van Wijk, 2008; Mueller et al., 2011; Keim et al., 2008)
Visual analytic pipeline	(Grammel et al., 2010; Wang et al., 2016; Bradel et al., 2015; Kohlhammer et al., 2011)
Visual analytics *and* feedback loop	(Keim et al., 2008; Sun et al., 2013; Bradel et al., 2015; Hermann and Klein, 2015; Makonin et al., 2016; Karami, 2015)
Visual analytics *and* Insight provenance	(Xu et al., 2015; Sacha et al., 2016a)

Table 4: Keywords and phrases used in the initial human and computer visualisation pipeline review and the most relevant references

be rendered (Wickham et al., 2009). Specific phases within the pipeline showing the stages of the process have been identified and acknowledged (Myatt and Johnson, 2011; Ware, 2004; Wickham et al., 2009; Wilkinson, 2006).

Visualisations have both a purpose and a storyline. Humans are aware of the purpose (underlying intention) and storyline of a particular visualisation (Grammel et al., 2010; Kirk, 2012; Kosslyn, 2006; Munzner, 2009; Yau, 2013); the reason for its existence (Kirk, 2012). Whether the intention is to convince the viewer of something, such as visualisations used for advertising, or summarise results, the purpose will affect the decisions made during the design (such as the variables chosen, or whether to introduce artistic elements) as well as the final outcome (Kosslyn, 2006; Meirelles, 2013). Storylines compress information in the same way that visualisations do, conveying large amounts of inferred information with very few words (Gershon and Page, 2001). Marrying facts to the story behind them makes the data memorable (Kosara and Mackinlay, 2013) and allows viewers to relate to the information (Figueiras, 2014). Adding story-like features to a visualisation, such as continuity and context, can also help guide a user's focus and highlight the intended purpose of the visualisation (Gershon and Page, 2001). Starting a visualisation without first clarifying why it matters to the audience is a recipe for failure (Sykes et al., 2012). Visual designers should know what the design needs to achieve and should aspire to emulate journalists propensity of uncovering important and relevant information (Kirk, 2012). What one knows about one's data can drive elements of the visualisation (Pereira, 2007) and combining data visualisation with domain-specific knowledge is considered to be challenging even for a human (Kalogerakis et al., 2006). Storytelling can make sure that the author's objective message is delivered (Khataei and Lau, 2013).

Stories are found in comparisons (range, ranking), measurements (magnitude), context (deviation, forecasts, averages), trends (direction, rate of change, fluctuations), intersections, relationships (exceptions, correlations, association, gaps) and in hierarchies (Katz, 2012). Information overload needs to be avoided since it obscures purpose (Katz, 2012). A story can be defined as an ordered sequence of steps and can be aligned with the visualisation pipeline (Kosara and Mackinlay, 2013). In fact order (examples of which include time and causality) is a key feature of stories (Kosara and Mackinlay, 2013; Segel and Heer, 2010). Techniques from film making are used (Hullman and Diakopoulos, 2011; Khataei and Lau, 2013). Stories told in visualisations are generally not interactive (Gershon and Page, 2001). The three main ways of calling attention to storytelling in visualisations is the use of genres, visual clues that direct attention or orient (called anchoring) and tactics such as ordering (Hullman and Diakopoulos, 2011; Segel and Heer, 2010).

A successful visualisation is able to make smart comparisons, show causality and present multivariate data in a manner that exposes useful information such as size, direction or position. Ideally a visualisation should be able to highlight aspects of the data that were not visible before (Yau, 2013). It also needs to retain the integrity if data and be respectful of the credibility of the data (Tufte, 2006). There are different ways of visualising data, such as data maps, bar graphs, time series and many others (Tufte and Graves-Morris, 1983). There can be multiple attributes, such as who, what, where and when, that can be viewed within the same dataset (Yau, 2013). The connection between the data and its meaning is very important for a successful visualisation (Yau, 2013). Labelling and context is critical without which a visualisation is meaningless (Tufte, 2006; Yau, 2013).

At its most basic, the visualisation process only has four stages namely collection of data, data preprocessing, standardising into something that can be charted and rendering (See Figures 1 and 2). The process involves cleaning the data, understanding what the data is about, deciding on the context and what aspects of the data can (or should) be presented and then choosing a graphic representation for display (Kirk,

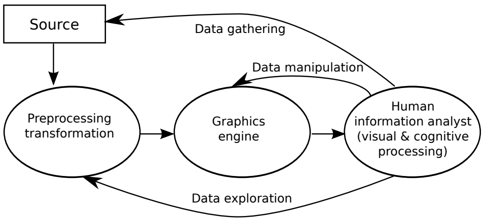

Figure 1: The graphics pipeline according to Ware (Ware, 2004)

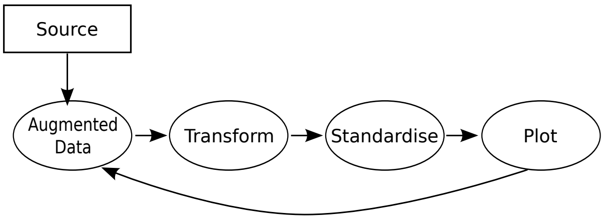

Figure 2: The simple pipeline according to Wickham (Wickham et al., 2009)

2012; Spence, 2001; Yau, 2013). The purpose of this work-flow is to make the data useful by distilling and highlighting patterns within the data that are useful to humans (Myatt and Johnson, 2011; Spence, 2001; Yau, 2013).

As indicated in Ware's diagram (Figure 1), this process is iterative. The iteration exists because humans may expect to find certain things in the data, but they also learn about their data as they go along; adapting the visualisation according to what they see (Sugiyama, 2002). Choosing which data will produce a useful visualisation can require an exploration phase (Yau, 2013). Iteration and experimentation are important because some aspects in the data only show shortcomings when graphed and one successful graph can suggest an idea for a better one (Cleveland, 1985). Visualisation designers infer and induct information about the data as they work and make adjustments (Kirk, 2012). These adjustments could be based on the realisation that another axis scale would be better or that the data is showing outliers or has gaps in the information (Kirk, 2012). A visualisation is not constrained to one set of data, and could include other data such as predictions from a statistical model (Wickham, 2010).

Much of this iterative design of visualisations is entirely due to the close link between visualisation and human visual skills (Myatt and Johnson, 2011). Humans can easily detect things in data, such as outliers and symmetry, but some artificial intelligence methods and implementations find these patterns hard to detect (Boden, 2004; Myatt and Johnson, 2011).

In addition to background knowledge and understanding of purpose, humans are also aware about background to the data as well as vocabulary in the data that emerges from the particular domain (such as e-commerce or politics) in which the data exists (Munzner, 2009). Humans are aware of relationships between concepts, data attributes and data values and know how to calculate some values from others (Grammel et al., 2010).

3.2 Computer only visualisation pipeline

Various data visualisation tools exist. Gephi (Bastian et al., 2009), Treemap (Shneiderman, 2015) and D3.js (Teller, 2013) combined with a vector graphics program are frequently used. They still require some low level programming on the part of the user and are not fully automatic (Wu et al., 2014). Some visualisation tools use a specialised visualisation tool on top of a database. Activities such as aggregation and filtering get duplicated with this approach (Wu et al., 2014). In order to create an algorithm that produces a visualisation the pipeline needs to be abstracted and modelled (Munzner, 2009).

The process of data visualisation has been formalised as a formal grammar known as a context free grammar (Chomsky, 1956). This particular context free grammar is called the "*grammar of graphics*" (Wilkinson, 2006). The grammar formalises the data pipeline into a formal reproducible process. The grammar identifies three graphical elements that make up a visualisation: the data, the scales and coordinates, and plot annotations – such as title and background (Wickham, 2010). The formalisation of the visualisation pipeline into a context free grammar allows the abstraction of the task in a manner that focusses on the structure of the graphic process rather than on specific representations (Wickham, 2010). The grammar facilitates moving away from specific chart types – such as pie chart or scatterplot – and focus instead on the underlying structure as well as composition; It also facilitates insight into how seemingly different graphics can be related to each other (Wickham, 2010).

A context free grammar consists of a set of rules for generating allowed transforms in a language; it can be thought of as a set of nested production rules (Wilkinson, 2006). This has specific meaning in linguistics. Formal grammars are also the basis of theories of computation (Chomsky, 1956; Cohen and Chibnik, 1991; Jurafsky and Martin, 2014). The "*grammar of graphics*" is a set of rules by which data can be transformed into graphics (Wilkinson, 2006). Wilkinson et al. (2000) also show that sections of

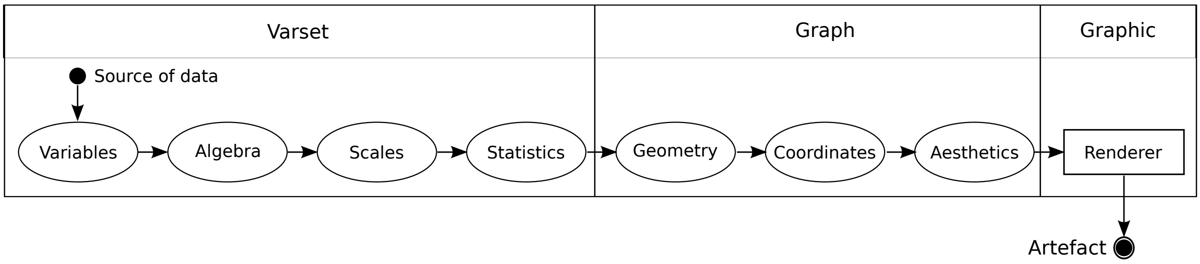

Figure 3: Diagram of grammar rules used for the visualisation of a pie chart as described by the *grammar of graphics* (Wilkinson, 2006).

Category	stage	brief description of activity
Varset	Variables	Converting the raw data into variables. Includes activities such as calculating ranges, assigning ordering and deciding how to assign size to non numeric data. Transforming the data so that statistical operations can be performed. Sorting (Patterns can be more obvious with sorted data (Yau, 2013)), aggregating and calculating statistics.
	Algebra	Various set theoretic operations to combine variables. These operations have equivalent database structured query language (SQL) statements.
	Scales	These are activities that decide how variables map to a scale or axis in the visualisation. Scales can be nominal, ordinal, intervals or ratios and they can have a measurement unit (seconds) but not always (towns). The choice of scale is critical to whether patterns emerge in the visualisation. Scales can also be transformed into other scales.
	Statistics	Statistical methods applied to the scales. Includes calculations such as average, smooth, sum, mean and so on.
Graph	Geometry	These are functions that convert the variables into geometric objects such as point, line, contour, area and polygon.
	Coordinates	Coordinates are functions that decide location in space (Cartesian, polar, spherical). They can also be transformed (dilate, stretch, rotate).
	Aesthetics	The application of various aesthetical elements such as choice of colour, hue, texture, brightness and labels.
Graphic	Renderer	The final conversion of all of the previous stages into the visualisation

Table 5: The types of activities that occur in each of the stage of the visualisation pipeline as per the "*grammar of graphics*" (Wilkinson, 2006)

the grammar can be executed in parallel (Wilkinson et al., 2000). Wickham (2010) refers to this as layering. Breaking the graphics pipeline into further layers – or individual plots – facilitates the plotting of multiple datasets against multiple types of geometries on the same visualisation (Wickham, 2010). The rules of the grammar are independent so any of the steps can be swapped out with an equivalent production rule without effecting the rest of the pipeline (Wickham, 2010). This allows for really complex visualisations to be produced – such as the reproduction of Minard's famous 1869 visualisation of Napoleon's march (Wickham, 2010). Formal grammars are not deterministic and there is more than one way to get from the data to a visualisation, which makes sense insomuch as there is more than one way to visualise a set of data (Cohen and Chibnik, 1991; Yau, 2013).

The *grammar of graphics* is widely recognised and is used by various visualisations tools (Wickham, 2010). Tools include, the Graphics Production Library (GPL) (Wilkinson et al., 2000), nViZn (Wilkinson, 2006), ggplot2 (Wickham, 2010), Protovis (Bostock and Heer, 2009), Grammar of graphics in D3 (Braşoveanu et al., 2009; Hunter, 2016), Vega (Braşoveanu et al., 2009), nViZn (Jones and Symanzik, 2001), VizQL (Mackinlay et al., 2007) and VizJSON (Malaika and Brunssen, 2015). Wilkinson does not claim that the *grammar of graphics* can be used to produce every visualisation possible (Wilkinson, 2006; Wickham, 2010), but his comprehensive book suggests that it covers most aspects of the process (Wilkinson, 2006). An activity diagram of the visualisation pipeline for a pie chart, as described by the *grammar of graphics*, is shown in Figure 3. An indication of the types of activities occurring in each of the stages of the "*grammar of graphics*" are shown in Table 5.

There are three groups within the grammar (See Figure 3). Production rules (the *Varset* group) are all related to preparing the variables in the dataset. A significant part of the pipeline involves choosing which variables to display. The second group (labelled *Graph* in Figure 3), contains rules pertaining to the choice of graph type and aspects of the graph such as axis and scale. The final group of rules belong to the *rendering* stage of the visualisation, and include things like the overall title and background images. The order of events in the grammar is fixed and does not change for other types of visualisations and therefore variables are always chosen before algebra; which is always completed before attending to scale (Wilkinson, 2006).

Although widely recognised, some authors are working on improvements with respect to the shortcomings of the "*grammar of graphics*". It lacks the ability to address abstract attributes such as information density and does not support strategies to come up with multiple representations of the same information (Redström et al., 2000). Control over graphical output needs to be addressed and graphic customisations are not supported (Bostock and Heer, 2009), although, Wilkinson mentions that this is by design (Wilkinson, 2006). Many visualisation systems propose methods on how to apply the graphics pipeline, but fail to address when a particular method should be chosen (Munzner, 2009) – in other words the process is non-deterministic. The grammar also requires a steep learning curve due to complexity (Bostock and Heer, 2009) and does not support dynamic graphics (Young et al., 2011).

Other automation mechanisms include the use of dataflow diagrams (Senay and Ignatius, 1994), genetic algorithms (Bouali et al., 2015) and machine learning techniques such as t-distributed stochastic neighbour embedding (van der Maaten and Hinton, 2008). Many machine learning tools, which feature visualisation, facilitate the display of multiple dimensions, but still leave the interpretation to a human (van der Maaten and Hinton, 2008).

3.3 Human and computer visualisation pipeline

One research area that tries to address the, "*human factors limit*", in automated data visualisation is the field of *Visual Analytics*. *Visual Analytics* solves the limit by involving humans in the process using interactive visual interfaces (Cybulski et al., 2015; Keim et al., 2008; Myatt and Johnson, 2011). Visual analytics adds data analysis (informational, geospatial, scientific or statistical) and human factors to the process of data visualisations and combines three visualisation tasks, *reasoning* (sense making), *interactive visualisations* and *analytical processes* (such as statistics and data mining techniques) (Keim et al., 2008). This adds additional iterative cycles to the visualisation pipeline. The knowledge generation model for visual analytics is composed of a two parts. The first part concerns the data model and the second human part is specific to the view of the data, and facilitates hypothesis finding, insight and knowledge generation (Chi and Riedl, 1998; Grammel et al., 2010; Wang et al., 2016). Visual analytics shares some of the usual visualisation challenges, such as scalability, uncertainty and difficulty with evaluation; It also has its own challenges, including, hardware support and the difficulty in the design of intuitive graphic user-interfaces (Kohlhammer et al., 2011). One model of the visualisation pipeline in visual analytics – the data state reference model – also has four stages (Data, models, knowledge and visualization) (Wang et al., 2016).

The graphic user-interfaces for visual highlighting and interrogation of data, support seven manipulation tasks (Shneiderman, 1996). Shneiderman (1996) identifies these tasks as overview, zoom, filter, details-on-demand, relate, history and extract. The tasks help a user navigate data in a visual manner using dynamic feedback, thereby facilitating iteration, but also offers hints as to human perception (Shneiderman, 1996). Users choose where to zoom because they know what they are interested in seeing, and the visualisation gives them perceptual clues as to patterns in the data. Filtering out uninteresting items, requesting more detail and viewing relationships requires knowledge and intuition of what is important and what is un/interesting after getting a broad overview of the data. Some interfaces offer editing of the transformations (Jankun-Kelly et al., 2002). The interactive user-interface facilitates the incorporation of human intelligence into the process, but the combination of knowledge and visual feedback produces more satisfying and intuitive visualisations (Wang et al., 2016).

The visual analytics pipeline, like the non-interactive visualisation pipeline, also contains iterative feedback loops. The key feature is that the user controls the iteration (Wang et al., 2016). Since the graphic user-interface controls are attached to the visualisation transformation changes, the changes in visualisation transforms between iterations should probably be part of the pipeline (Jankun-Kelly et al., 2002). These iterations continually refine the visual towards the intended purpose of the visualisation (Grammel et al., 2010).

There are ways humans and computers can generate visualisations other than using a human driven, interactive graphic user-interface. Bouali et al. (2015) generated visualisations using a genetic algorithm, but using a human to assess and score the result. They started with a model of possible mappings between the visual and data attributes – encoded as a vector of weights. Standard operations, such as crossover and mutation, produced potential visualisations. The genetic algorithm used the human-supplied scores to iteratively produce another set of improved visualisations (Bouali et al., 2015). Visualisation can also be a part of machine learning tools (such as *WEKA*). These tools also facilitate some data pre-processing and filtering choices before learning the data and visualising the result (Hall et al., 2009).

One of the ways in which computers can support creativity in the visual analytics pipeline is by suggesting visual mappings and alerting the user to the advantages and features of the suggested mappings (Grammel et al., 2010). In the visual analytics environment, rapidly allowing the user to switch visualisation types and visual mappings and facilitating backtracking also facilitates creativity support while automated wizards stifle creativity (Grammel et al., 2010). Novel user-interfaces focussing on visual design, such as *Visualization-by-Sketching*, have also been suggested (Schroeder and Keefe, 2016) to facilitate creativity for artistic, but non visualisation-expert individuals.

Other visualisation pipelines include the information visualization data state reference (Chi and Riedl, 1998; Chi, 2000), the generic visualization model (Van Wijk, 2005), the reference model (Card et al., 1999) and the nested model of visualization creation (Munzner, 2009). They are similar to those discussed here and the reader is referred to Wang et al. (2016) for a comprehensive discussion of these pipelines.

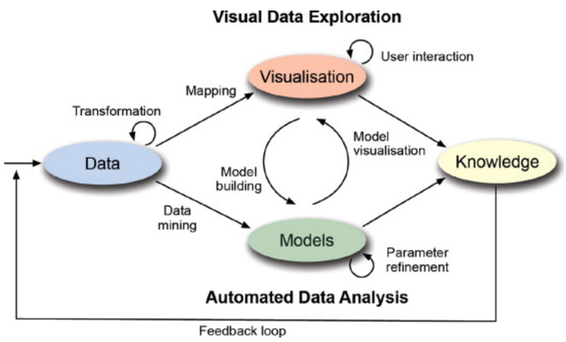

Figure 4: Visual analytics pipeline (Kohlhammer et al., 2011)

The interactive visualisation system, VisTrails, makes the pipeline visible in the graphic user-interface in the form of a breadcrumb trail, but also provides multiple possible views of that pipeline, the comparison of which provide insight into the data (Bavoil et al., 2005). Many interactive visualisation systems build the visualisation pipeline out of smaller modules – a process referred to as the data flow module (Bavoil et al., 2005) or flow networks.

Sense making is the process of foraging through information in order to generate and identify meaning and gain insight (Xu et al., 2015; Nguyen et al., 2016). Attempts to model the sense making process in visual analytics are referred to as analytic provenance (Nguyen et al., 2016). Proposed sense making pipelines include, the *sense-making model* (Pirolli and Card, 2005), the *knowledge generating model* (Sacha et al., 2014), the *data/frame theory of sense-making* (Klein et al., 2006), *human cognition model* (Green et al., 2009) and *a pipeline of the knowledge discovery in databases* (Han et al., 2011). Analytic provenance incorporates levels of semantic information. Sense making includes *tasks* (analyse the stock market for investment recommendations) and *sub-tasks* (identify performance trends of companies) as well as semantic elements such as *actions* (sort by stock price) and *events* (mouse click) (Nguyen et al., 2016; Gotz and Zhou, 2009).

4 DISCUSSION

The two additional stages of the visualisation pipeline added by visual analytics – the sense making and interactive visualisation stages – hint at where a fully automated computer generated visualisation algorithm could attempt to address the shortcomings of existing automated visualisation algorithms. All visualisation pipelines have a data pre-processing, clean-up and variable choice process; however, background knowledge of the data, purpose, intended storyline as well as recognition of outliers or elements of interest are human attributes.

The literature indicates that, while some stages in the pipeline are marginally better suited to either a computer or human, a significant use of visual perception, intuition, storytelling, purpose and creativity occur at the points within the pipeline where iterative loops occur – and stages of the pipeline are revisited. Examples of iteration include, returning to the variable choice stage to choose another aspect of the data to add to the visualisation or returning from a rendered visualisation to change the glyphs, scale or axis. The combination of exploration and iteration confer more satisfying visualisation results, than knowledge about the data does on its own. This iteration is strongly tied to the purpose of the visualisation. Choosing when and how to iterate could be motivated by background information, current culture or events, knowledge of relationships in the data, purpose or intent of the visualisation as well as knowledge of *Gestalt principals*. In other words, when computers are generating visualisations without the aid of humans, it may be beneficial to facilitate the addition of an automated sense making (or reasoning and information seeking) cycle and analytic provenance model into the algorithm; particularly if this model facilitates knowledge driven looping between the stages of the visualisation pipeline.

Computers facilitate creativity during the visualisation process by assisting with choices at every stage of the visualisation pipeline and aiding with filtering and transforming data, visual comparisons, backtracking, memory aids and quick access to high and low level detail of the visualisation.

Storytelling is one of the targets of existing computer programs attempting to emulate creativity and these existing techniques may come in use for computational data visualisation. Among the techniques used, are scripts and frames (which are used to provide background information and fill in missing data) or semantic networks (which are used to establish meaning and relationships between concepts) (Boden, 1998). These techniques are among those currently used to automate the creation of poetry (Colton et al., 2012) and computer generated stories (Gervás, 2009; Gervás and León, 2016; Riedl, 2016). The awareness of a story behind the data, or the intent of a visualisation, does not appear to be modelled in any computer-only visualisation generating algorithms and the computer generated story creation techniques could be explored and adapted for inclusion. Machine learning tools or computer algorithms that automate the generation of visualisations would likely benefit from fitness functions and heuristics that can drive the visualisation towards a storyline or purpose. Constraining a computer generated visualisation to a predefined grammar or wizard may be too limiting when attempting to emulate human creativity and information seeking.

5 CONCLUSION

Creativity in the data visualisation pipeline is tied to knowledge about the data, but only when combined with an iterative exploration process and insight. Incorporating story lines and purpose into the computational algorithms that generate visualisations, and incorporating models of insight provenance, are potentially good enhancements to existing visualisation algorithms that are attempting to incorporate techniques and theories from computational creativity. Particularly in light of the popularity of visual analytics and the known shortcomings of existing no-human-in-the-loop visualisation generating software.

REFERENCES

Alexander, M., Yuk, M., and Diamond, S. (2014). *Data visualization for dummies*. John Wiley & Sons.

Baer, K. and Vacarra, J. (2008). *Information design workbook: Graphic approaches, solutions, and inspiration+ 30 case studies*. Rockport Publishers.

Bastian, M., Heymann, S., Jacomy, M., et al. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362.

Bavoil, L., Callahan, S. P., Crossno, P. J., Freire, J., Scheidegger, C. E., Silva, C. T., and Vo, H. T. (2005). Vistrails: Enabling interactive multiple-view visualizations. In *VIS 05. IEEE Visualization, 2005.*, pages 135–142. IEEE.

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, 103(1):347–356.

Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. Psychology Press.

Bostock, M. and Heer, J. (2009). Protovis: A graphical toolkit for visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1121–1128.

Bouali, F., Guettala, A., and Venturini, G. (2015). Vizassist: an interactive user assistant for visual data mining. *The Visual Computer*, pages 1–17.

Bradel, L., Wycoff, N., House, L., and North, C. (2015). Big text visual analytics in sensemaking. In *Big Data Visual Analytics (BDVA), 2015*, pages 1–8. IEEE.

Braşoveanu, A. M., Sabou, M., Scharl, A., Hubmann-Haidvogel, A., and Fischl, D. (2009). Visualizing statistical linked knowledge for decision support. *Semantic Web*, (Preprint):1–25.

Byrne, L., Angus, D., and Wiles, J. (2016). Acquired codes of meaning in data visualization and infographics: Beyond perceptual primitives. *Visualization and Computer Graphics, IEEE Transactions on*, 22(1):509–518.

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

Chen, M., Ebert, D., Hagen, H., Laramee, R. S., Van Liere, R., Ma, K.-L., Ribarsky, W., Scheuermann, G., and Silver, D. (2009). Data, information, and knowledge in visualization. *Computer Graphics and Applications, IEEE*, 29(1):12–19.

Chi, E. H. (2000). A taxonomy of visualization techniques using the data state reference model. In *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, pages 69–75. IEEE.

Chi, E. H.-h. and Riedl, J. T. (1998). An operator interaction framework for visualization systems. In *Information Visualization, 1998. Proceedings. IEEE Symposium on*, pages 63–70. IEEE.

Cho, W.-J., Shim, J.-S., and Park, S.-C. (2016). Design and implementation of bigdata visualization system based on meta db automatic mapping algorithm. *International Information Institute (Tokyo). Information*, 19(2):673.

Chomsky, N. (1956). Three models for the description of language. *Information Theory, IRE Transactions on*, 2(3):113–124.

Cleveland, W. S. (1985). *The elements of graphing data*. Wadsworth Advanced Books and Software Monterey, CA.

Cohen, D. I. and Chibnik, M. (1991). *Introduction to computer theory*, volume 2. Wiley New York.

Colton, S., Goodwin, J., and Veale, T. (2012). Full face poetry generation. In *Proceedings of the Third International Conference on Computational Creativity*, pages 95–102.

Conti, G. (2007). *Security data visualization: graphical techniques for network analysis*. No Starch Press.

Cybulski, J. L., Keller, S., Nguyen, L., and Saundage, D. (2015). Creative problem solving in digital space using visual analytics. *Computers in Human Behavior*, 42:20 – 35. Digital Creativity: New Frontier for Research and Practice.

Deshpande, M., Ray, D., Dixit, S., and Agasti, A. (2015). Shareinsights: An unified approach to full-stack data processing. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1925–1940. ACM.

Dos Santos, S. and Brodlie, K. (2004). Gaining understanding of multivariate and multidimensional data through visualization. *Computers & Graphics*, 28(3):311–325.

Erwin, K., Bond, M., and Jain, A. (2015). Discovering the language of data: Personal pattern languages and the social construction of meaning from big data. *Interdisciplinary Science Reviews*, 40(1):44–60.

Figueiras, A. (2014). Narrative visualization: A case study of how to incorporate narrative elements in existing visualizations. In *Information Visualisation (IV), 2014 18th International Conference on*, pages 46–52. IEEE.

Fink, A. (2013). *Conducting Research Literature Reviews: From the Internet to Paper: From the Internet to Paper*. Sage Publications.

Gahegan, M. (1999). Four barriers to the development of effective exploratory visualisation tools for the geosciences. *International journal of geographical information science*, 13(4):289–309.

Gershon, N. and Page, W. (2001). What storytelling can do for information visualization. *Communications of the ACM*, 44(8):31–37.

Gervás, P. (2009). Computational approaches to storytelling and creativity.

Gervás, P. and León, C. (2016). Integrating purpose and revision into a computational model of literary generation. *Creativity and Universality in Language*, page 105.

Gotz, D. and Zhou, M. X. (2009). Characterizing users' visual analytic activity for insight provenance. *Information Visualization*, 8(1):42–55.

Grammel, L., Tory, M., and Storey, M.-A. (2010). How information visualization novices construct visualizations. *IEEE transactions on visualization and computer graphics*, 16(6):943–952.

Green, T. M., Ribarsky, W., and Fisher, B. (2009). Building and applying a human cognition model for visual analytics. *Information visualization*, 8(1):1–13.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.

Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Hanna, G. (2015). Automated data visualizations–a modular and extensible approach.

Heer, J. and Agrawala, M. (2008). Design considerations for collaborative visual analytics. *Information visualization*, 7(1):49–62.

Hermann, M. and Klein, R. (2015). A visual analytics perspective on shape analysis: state of the art and future prospects. *Computers & Graphics*, 53:63–71.

Hooi-Ten Wong, D. and Ramadass, S. (2010). An automated algorithm in data visualization for large network data: A review and an initial study. *International Journal of Simulation–Systems, Science & Technology*, 11(3).

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

Hunter, B. (2016). ggD3 - grammar of graphics in d3. Retrieved September 19, 2016, from http://benjh33.github.io/ggd3.

Jankun-Kelly, T., Ma, K.-L., and Gertz, M. (2002). A model for the visualization exploration process. In *Visualization, 2002. VIS 2002. IEEE*, pages 323–330. IEEE.

Jones, L. and Symanzik, J. (2001). Statistical visualization of environmental data on the web using nViZn. *Computing Science and Statistics*, 33.

Jurafsky, D. and Martin, J. (2014). *Speech and Language Processing*. Pearson custom library. Prentice Hall, Pearson Education International.

Justin Beegel, M. (2014). *Infographics for Dummies*. John Wiley & Sons.

Kalogerakis, E., Christodoulakis, S., and Moumoutzis, N. (2006). Coupling ontologies with graphics content for knowledge driven visualization. In *Virtual Reality Conference, 2006*, pages 43–50. IEEE.

Karami, A. (2015). A framework for uncertainty-aware visual analytics in big data. *AIC 2015*, page 146.

Katz, J. (2012). *Designing information: Human factors and common sense in information design*. John Wiley & Sons.

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., and Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In *Information Visualization*, pages 154–175. Springer-Verlag.

Khataei, A. and Lau, D. (2013). Recommender narrative visualization. In *Proceedings of the 2013 Conference of the Center for Advanced Studies on Collaborative Research*, pages 415–421. IBM Corp.

Kirk, A. (2012). *Data Visualization: a successful design process*. Packt Publishing Ltd.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26.

Klein, G., Moon, B., and Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, 21(5):88–92.

Kohlhammer, J., Keim, D., Pohl, M., Santucci, G., and Andrienko, G. (2011). Solving problems with visual analytics. *Procedia Computer Science*, 7:117–120.

Kosara, R. (2007). Visualization criticism-the missing link between information visualization and art. In *Information Visualization, 2007. IV'07. 11th International Conference*, pages 631–636. IEEE.

Kosara, R. and Mackinlay, J. (2013). Storytelling: The next step for visualization. *Computer*, pages 44–50.

Kosslyn, S. M. (2006). *Graph design for the eye and mind*. Oxford University Press.

Krum, R. (2013). *Cool infographics: Effective communication with data visualization and design*. John Wiley & Sons.

Lankow, J., Ritchie, J., and Crooks, R. (2012). *Infographics: The power of visual storytelling*. John Wiley & Sons.

Lee, S., Kim, S.-H., Hung, Y.-H., Lam, H., Kang, Y.-a., and Yi, J. S. (2016). How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking. *IEEE transactions on visualization and computer graphics*, 22(1):499–508.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions On Graphics (Tog)*, 5(2):110–141.

Mackinlay, J., Hanrahan, P., and Stolte, C. (2007). Show me: Automatic presentation for visual analysis. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1137–1144.

Makonin, S., McVeigh, D., Stuerzlinger, W., Tran, K., and Popowich, F. (2016). Mixed-initiative for big data: The intersection of human+ visual analytics+ prediction. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 1427–1436. IEEE.

Malaika, S. and Brunssen, V. C. (2015). VizJSON: The grammar of graphics in JSON. Retrieved September 19, 2016, from http://www.ibm.com/developerworks/library/bd-vizjson.

Meirelles, I. (2013). *Design for information: an introduction to the histories, theories, and best practices behind effective information visualizations*. Rockport publishers.

Muehlenhaus, I. (2012). Visualize this: The flowingdata guide to design, visualization, and statistics. *Cartography and Geographic Information Science*, 39(3):170–172.

Mueller, K., Garg, S., Nam, J. E., Berg, T., and McDonnell, K. T. (2011). Can computers master the art of communication?: A focus on visual analytics. *IEEE Computer Graphics and Applications*, 31(3):14–21.

Munzner, T. (2009). A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics*, 15(6):921–928.

Myatt, G. J. and Johnson, W. P. (2009). *Making sense of data II: A practical guide to data visualization, advanced data mining methods, and applications*. John Wiley & Sons.

Myatt, G. J. and Johnson, W. P. (2011). *Making Sense of Data III: A Practical Guide to Designing Interactive Data Visualizations*. John Wiley & Sons.

Nguyen, P. H., Xu, K., Wheat, A., Wong, B. W., Attfield, S., and Fields, B. (2016). Sensepath: understanding the sensemaking process through analytic provenance. *IEEE transactions on visualization and computer graphics*, 22(1):41–50.

North, C., Endert, A., Andrews, C., and Fink, G. (2009). The visualization pipeline is broken. In *Proceedings of the 2009 IEEE Information Visualization Conference (IEEE InfoVis)*.

Ottley, A. (2016). Toward personalized visualizations.

Pereira, F. C. (2007). *Creativity and artificial intelligence: a conceptual blending approach*, volume 4. Walter de Gruyter.

Pickering, C. and Byrne, J. (2014). The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *Higher Education Research & Development*, 33(3):534–548.

Pirolli, P. and Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, volume 5, pages 2–4.

ProQuest (2016). Ebrary - academic electronic books. Retrieved September 19, 2016, from http://proquest.libguides.com/ebrary.

Reda, K., Johnson, A. E., Papka, M. E., and Leigh, J. (2016). Modeling and evaluating user behavior in exploratory visual analysis. *Information Visualization*, page 1473871616638546.

Redström, J., Skog, T., and Hallnäs, L. (2000). Informative art: using amplified artworks as information displays. In *Proceedings of DARE 2000 on Designing augmented reality environments*, pages 103–114. ACM.

Riedl, M. O. (2016). Computational narrative intelligence: A human-centered goal for artificial intelligence. *arXiv preprint arXiv:1602.06484*.

Rodrigues-Jr, J., Zaina, L., Oliveira, M., Brandoli, B., and Traina, A. (2015). A survey on information visualization in light of vision and cognitive sciences. *arXiv preprint arXiv:1505.07079*.

Ryabinin, K. and Chuprina, S. (2015). Development of ontology-based multiplatform adaptive scientific visualization system. *Journal of Computational Science*, 10:370–381.

Sacha, D., Boesecke, I., Fuchs, J., and Keim, D. A. (2016a). Analytic behavior and trust building in visual analytics.

Sacha, D., Sedlmair, M., Zhang, L., Lee, J. A., Weiskopf, D., North, S., and Keim, D. (2016b). Human-centered machine learning through interactive visualization: Review and open challenges. In *Proceedings of the 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*.

Sacha, D., Senaratne, H., Kwon, B. C., Ellis, G., and Keim, D. A. (2016c). The role of uncertainty, awareness, and trust in visual analytics. *IEEE transactions on visualization and computer graphics*, 22(1):240–249.

Sacha, D., Stoffel, A., Stoffel, F., Kwon, B. C., Ellis, G., and Keim, D. A. (2014). Knowledge generation model for visual analytics. *IEEE transactions on visualization and computer graphics*, 20(12):1604–1613.

Schroeder, D. and Keefe, D. F. (2016). Visualization-by-sketching: An artist's interface for creating multivariate time-varying data visualizations. *IEEE transactions on visualization and computer graphics*, 22(1):877–885.

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

Senay, H. and Ignatius, E. (1994). A knowledge-based system for visualization design. *IEEE Computer Graphics and Applications*, 14(6):36–47.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE.

Shneiderman, B. (2015). Treemaps for space-constrained visualization of hierarchies.

Shrinivasan, Y. B. and van Wijk, J. J. (2008). Supporting the analytical reasoning process in information visualization. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1237–1246. ACM.

Simon, P. (2014). *The visual organization: data visualization, Big Data, and the quest for better decisions*. John Wiley & Sons.

Spence, R. (2001). *Information visualization*, volume 1. Springer.

Sugiyama, K. (2002). *Graph drawing and applications for software and knowledge engineers*, volume 11. World Scientific.

Sun, G.-D., Wu, Y.-C., Liang, R.-H., and Liu, S.-X. (2013). A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology*, 28(5):852–867.

Sun, Y., Leigh, J., Johnson, A., and Chau, D. (2009). Articulate: a conversational interface for visual analytics. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 233–234. IEEE.

Sykes, M., Malik, A. N., and West, M. D. (2012). *Stories that Move Mountains: Storytelling and Visual Design for Persuasive Presentations*. John Wiley & Sons.

Tatu, A., Albuquerque, G., Eisemann, M., Schneidewind, J., Theisel, H., Magnor, M., and Keim, D. (2009). Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 59–66. IEEE.

Teller, S. (2013). *Data Visualization with D3.js*. Packt Publishing Ltd.

Tufte, E. R. (2006). Beautiful evidence. *New York*.

Tufte, E. R. and Graves-Morris, P. (1983). *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT.

van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605.

Van Wijk, J. J. (2005). The value of visualization. In *VIS 05. IEEE Visualization, 2005.*, pages 79–86. IEEE.

Wang, X.-M., Zhang, T.-Y., Ma, Y.-X., Xia, J., and Chen, W. (2016). A survey of visual analytic pipelines. *Journal of Computer Science and Technology*, 31(4):787–804.

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

Wickham, H. (2010). A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, 19(1):3–28.

Wickham, H., Lawrence, M., Cook, D., Buja, A., Hofmann, H., and Swayne, D. F. (2009). The plumbing of interactive graphics. *Computational Statistics*, 24(2):207–215.

Wilkinson, L. (2006). *The grammar of graphics*. Springer Science & Business Media.

Wilkinson, L., Rope, D. J., Carr, D. B., and Rubin, M. A. (2000). The language of graphics. *Journal of Computational and Graphical Statistics*, 9(3):530–543.

Wills, G. and Wilkinson, L. (2010). Autovis: automatic visualization. *Information Visualization*, 9(1):47–69.

Wu, E., Battle, L., and Madden, S. R. (2014). The case for data visualization management systems: vision paper. *Proceedings of the VLDB Endowment*, 7(10):903–906.

Xu, K., Attfield, S., Jankun-Kelly, T., Wheat, A., Nguyen, P. H., and Selvaraj, N. (2015). Analytic provenance for sensemaking: A research agenda. *IEEE computer graphics and applications*, 35(3):56–64.

Yau, N. (2013). *Data points: visualization that means something*. John Wiley & Sons.

Young, F. W., Valero-Mora, P. M., and Friendly, M. (2011). *Visual statistics: seeing data with dynamic interactive graphics*, volume 914. John Wiley & Sons.

Zhu, N. Q. (2013). *Data Visualization with D3.js Cookbook*. Packt Publishing Ltd.