University of Bremen
Faculty 3 – Mathematics and Computer Science
Master Program of Digital Media

Master Thesis

Balancing author and reader driven narrative structure in interactive data visualization

Name: Farnoosh Shirali

Matriculation number: 2593798

1st Supervisor: Prof. Dr. Gabriel Zachmann

2nd Supervisor: Dr.-Ing. Marc C. Lemmel

Mentor: Christopher Nigischer

Bremen, August 2014

Statutory Declaration

I declare that I have developed and written the enclosed Master Thesis completely by myself, and have not used sources or means without declaration in the text. Any thoughts from others or literal quotations are clearly marked. The Master Thesis was not used in the same or in a similar version to achieve an academic grading or is being published elsewhere.

Date: Signature:

Table of Content

List of Tables

List of Figures

List of Abbreviations

CPSC	Consumer product safety commission
CPSC	Consumer Product Safety Commission
ETL	Extract, Transform, Load
HCI	Human Computer Interaction
HIV	Human Immunodeficiency Virus
OECD	Organization for Economic Cooperation and Development
RAPEX	Rapid Exchange
RAPEX	Rapid Alert System for Non-Food Consumer Products (EU)
SME	Small and medium enterprises
YTD	Year to Date

Acknowledgment

Foremost, I would like to thank my advisor Prof. Dr: Gabriel Zachmann, for the continuous support of my master thesis study and for his patience, motivation, enthusiasm, immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

Special thanks go to Dr.-Ing. Marc C. Lemmel for his cooperation, guidance and inspiration during my research.

I would also like to express my sincere gratitude to my mentor Mr. Christopher Nigischer for his encouragement, motivation and his unlimited support. He provided me with direction, technical support through this journey. I doubt that I will ever be able to convey my appreciation fully, but I owe him my eternal gratitude.

I would also like to thank my family for the support they provided me through my entire life. Without the support of my father and my mother I would not have achieve to this very point.

Abstract

Ubiquitously available information has increased dramatically and has led to the quest of dealing with the threat of information overload. Understanding how humans process information and by what means we can leverage our natural strengths in this area is a key success factor for anyone who has something to communicate. Current state-of-the-art concepts that are analyzed and elaborated in detail in this study include methods like telling stories based on data using narrative structures and interactivity. Also interactive data visualizations play a relevant role in this field and promote information intake and learning processes with the audience. The specific question of interest that was examined with a user study is how different author- and reader-driven approaches influence how much information is understood and remembered by the audience. Users that were directed to the main areas of interest by providing them some questions upfront showed significantly better results in answering questions after the exposure to data than less guided users. This insight can be used to create even better suited visualizations for different purposes in the future.

1. Motivation

In the 21st century each of us is dealing with enormous amounts of data every day. It can vary from basic activities such as reading emails to very complex data exchange and transactions. “Perhaps we have not even realized, but in the last decade, “the world contains an unimaginably vast amount of digital information which is getting ever vaster ever more rapidly”(Kenneth 2010). Some scholars refer to this development as information pollution.

The trend towards further increasing amounts of data is evolving rapidly. But does all that data really provide useful information? What can be done with all that data? How much time is needed to extract the relevant content and turn data into useful information?

The austerity today lies in the ability to present data in convenient ways. The message has to be presented in a consumable way which is easy to understand and only takes the minimum necessary time to convey the message. Data visualization enables the reader to first preserve an overview as well as the option to dig dipper therefore So presenting „the right“ data at „the right time“ gives further depth to presenting it.

Research on human visual perception has come a long way and is providing techniques to successfully present data today. “A picture is worth a thousand words!" and data visualization techniques enable data analysts, scientists and visual designers to create meaningful visuals based on available data. Automation and interaction on the other hand are assisting technologies to bring static visuals to life.

Today’s info graphics and visualizations go beyond just presenting data. Visual designers have just started to figure out how to use narrative techniques in combination with interaction design to reach stronger effects with particular messages.

Based on those developments, journalists, computer graphic experts, visual designers and data analysts are starting to find approaches to present data as stories in the form of interactive visualizations.

1.1 Why visualization?

Data, information and their visualization are at the heart of this paper’s topic. Therefore it makes sense to examine the meaning of those terms.

According to the theory of Nathan Shedroff: ”Data alone are not enough to establish a communicative process. To give meaning to this data, they must first be processed, organized and presented in a suitable format. This transformation and manipulation of the data produces information” (Mazza 2009, p. 9).

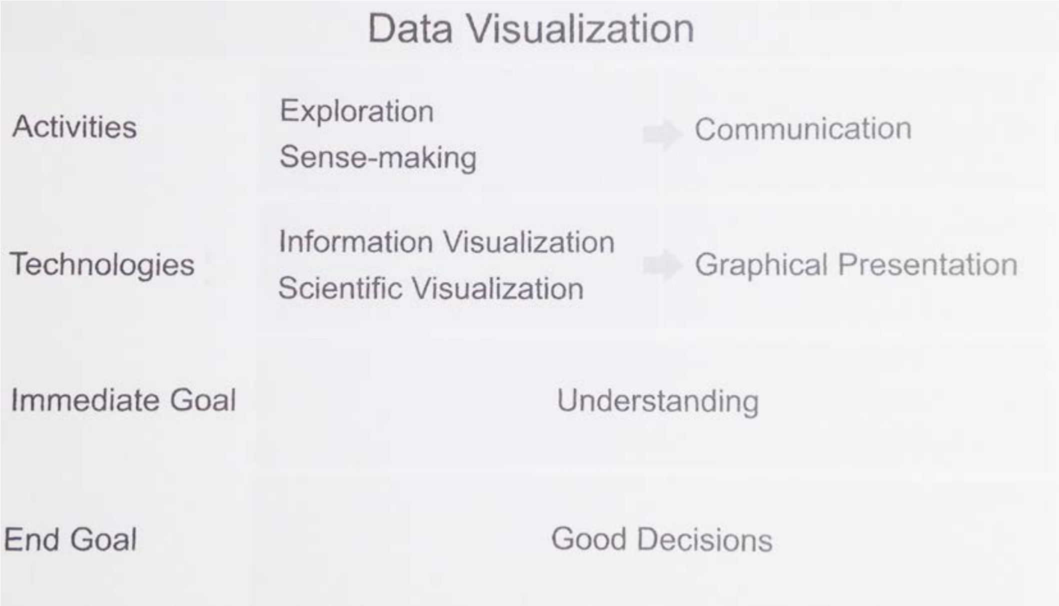
Each day of our life is filled with information. Every day we read newspapers, get the weather forecast or stay up-to-date with developments on the stock market. Regardless of what media we use to get information, they are most likely featured with some kinds of visualizations like pictures, maps, charts and so on. And beside the intense usage of data on daily basis, data itself is massively growing in current age. For example the rapid growth of medical information including health and clinical data leads medical experts and researchers to seek for larger data bank. This reference to our every-day lives shows that visualizing data in one way or the other is ubiquitous. But what exactly does visualization mean? The next chapter will provide some definitions and lay the basis for the further usage of the correct terms.

1.1.1 Definitions

In his book *Data Visualization for Human Perception* Stephen Few defines data visualization as “the graphical display of abstract information for two purposes: sense-making (also called data analysis) and communication” (Few 2013). According to Few’s definition, visual representations of abstract information are a general characteristic of any data visualization. Friendly, another scholar, gives a similar definition for it. “Data visualization, the science of visual representation of ‘data’, defined as information which has been abstracted in some schematic form, including attributes or variables for the units of information”(Friendly, 2006, p. 2).

In his book *Now you see it* Stephen Few divided the field of visualization into three sub categories: data visualization, information visualization and scientific visualization. Few considers data visualization as a general term which covers all categories of visual representation. It does not matter what it represents, as long as it has the two main characteristics of visual and information it can be called data visualization (Few, 2009, p. 12).

In 1999 the book *reading in information visualization using vision to think* introduced the new terms known as: *information visualization* and *scientific visualization*. They defined the information visualization as the visual representation of abstract data (information such as quantitative data, processes, or relationships) whereas scientific visualization focuses on the type of representation of objects which contains a physical form (the connection between data and visual characteristics such as shape and colors) (Few, 2009). The figure 1 presents the division of Few for visual representation.



Data Visualization

Activities	Exploration Sense-making	➡ Communication
Technologies	Information Visualization Scientific Visualization	➡ Graphical Presentation
Immediate Goal	Understanding	
End Goal	Good Decisions	

Figure 1. Data visualization scope

Card, Mackinlay, and Shneiderman define information visualization as: ” the use of computer–supported interactive, visual representation of abstract data to amplify cognition”(Card, Mackinlay, & Shneiderman, 1999, p. 7).

According to the tools and techniques that i use in this study, it is important to understand why we focus on information visualization. To support that, we will follow the foot prints of Stephan's statements.

In this thesis, the visualization displays with computer and it enables the user to manipulate the data by different types of interactions such as selection or drilldown. It uses different attributes such as color, shape or length to present the data. The information we have is quantitative data without any physical form and the last but not least, the interaction with the application in this study aids the user in a process of thinking, comparison and memorizing.

According to explanation above, from now on when we use the term information visualization, we only focus on the process of data transformation with characteristics explained above, whereas when use the term data visualization we are pointing out on a broader dimension of visualization which covers all type of visualization.

As pointed out earlier, we made ourselves aware of the fact that visualizations are a ubiquitous part of our every-day lives and that they are one of the answers to the explosion of available data in general. But what are the exact reasons for using information visualizations? To answer this question we will look at the effectiveness of presenting data in the form of graphics in comparison to the single usage of numbers and textual information. It will make clear how visualizing information helps the reader to receive, understand and store data.

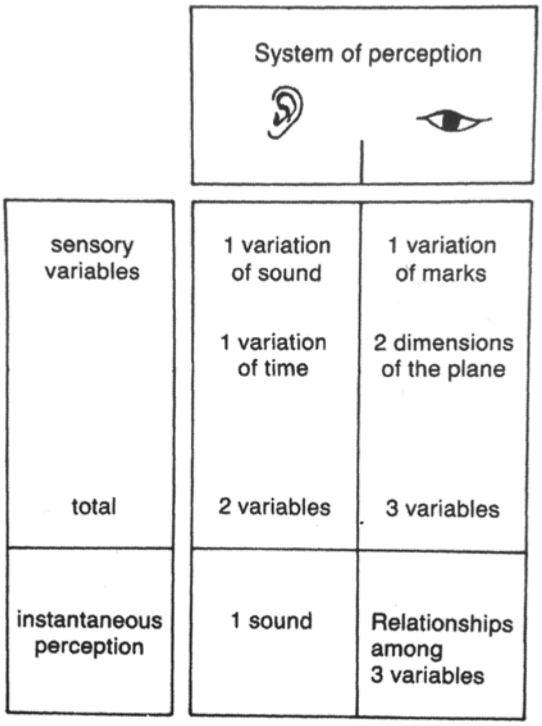
1.1.2 Human Perception and Data Visualization

Information can be presented in a textual form as well as in a visual form. It could be a table or a report or just an illustration in the form of a pie chart. But which approach is more effective and why do scholars agree on the fact that visualization of data in most cases is more effective than textual form?

How can we perceive information at all? In his book *Semiology of Graphics* Jacques Bertin explains how the human *sign-system* works with information**.** Sign-systems are a key concept in the studies of sense-making and describe any system of signs and the relations between them. Examples can be language, non-verbal communication, body-language but also man-machine communication (Beynon-Davies, 2010). Usually the ways how sense and meaning are being perceived is a key differentiator between different forms of sign-systems. Here we can very well differentiate between auditory and visual ways of perception. This perception means to decode the relationships between the signs and the meaning of the signs themselves. Also the context of meaning plays a role in this process (Bertin 1983, p. 2).

As far as human's auditory system is concerned, Bertin explains that this system is limited to working with only the two variables of sound and time. As also visualized in the following graphic (figure 2) the sign-system of hearing can also be described as being linear and time based.

In comparison visual perception has three sensory variables which do not involve time. Therefore, the auditory system is able to only communicate either with sound or with sign, whereas our visual system can perceive the following three variables simultaneously: the variation of marks and the two dimensions of the plane. As a result the visual system has a much higher bandwidth in perceiving information and can process larger volumes in much shorter time (Bertin 1983, p. 3). The further course of this thesis will therefore focus on visual representations of information.



System of perception

	👂	👁
sensory variables	1 variation of sound 1 variation of time	1 variation of marks 2 dimensions of the plane
total	2 variables	3 variables
instantaneous perception	1 sound	Relationships among 3 variables

Figure 2. Perceptual properties of linear and spatial system

Also within the category of visual representations of information there are huge differences in how fast human beings can process input. Ricardo Mazza in his book *Introduction to information visualization* makes a very simple example to show how visual representation is often more effective than written text (Mazza 2009, p. 2). The figure 3 is extracted from his book and explains how featuring numbers with a few simple lines aid the reader to remember the minimum and maximum of the numerical values.

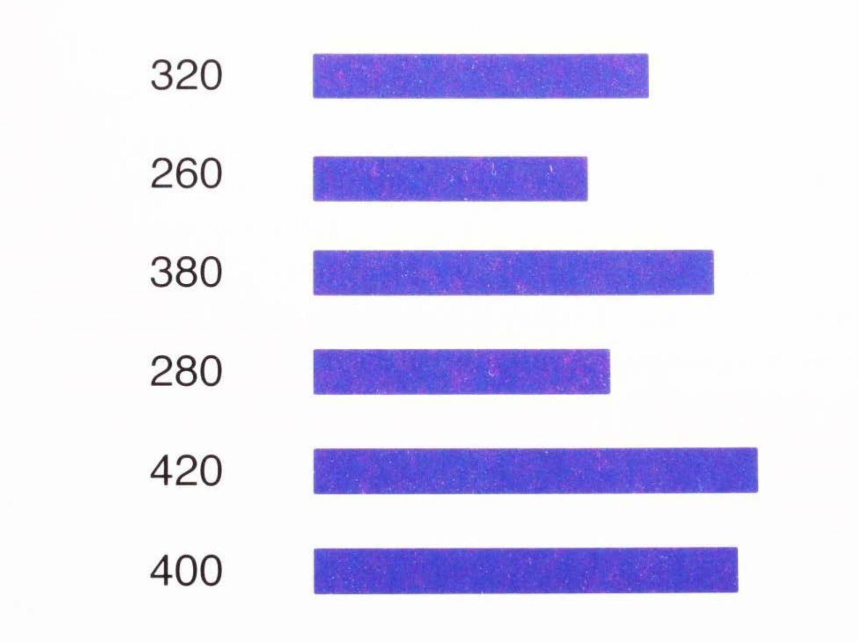


Figure 3. Example of human perception in relation between numbers and shapes

Mazza is pointing out that humans perceive visual attributes very well. Therefore a large amount of information can be conveyed using different visual attributes. An important aspect is to also define clear links between the underlying data and the corresponding visualization. That work is also referred to as “mapping”. Mapping techniques are helpful for designers and enable them to even use various different data types in one visualization (Mazza 2009, p. 4).

Bertin is also contributing an interesting example with a similar message: “It would take at least 20.000 successive instants of perception to compare two data tables of 100 rows by 100 columns. If the data are transcribed graphically, comparison becomes easy; it can even be instantaneous”(Bertin 1983, p. 3).

Overall the human visual system is the most powerful tool we have for ”identifying patterns as well as communicating the relationship and meaning of information” (Iliinsky & Steele, 2011, p. 1). It has the capability to process large amounts of data in short periods of time. Therefore illustration of information assists the designers, journalists and authors to convey their massage to the audience in the best possible way.

Besides the fact that every author has to have the knowledge to choose the most suitable and understandable data illustration, it is also necessary to transform the raw data into the structured, organized and meaningful form of information.

Regardless of type, technique and category, data visualization enables the reader to investigate and interact with graphics in order to extract more information. Although the level of interactivity can vary based on the aim of the visual designer, it is always sending the reader a message as well as giving him the opportunity of further exploration on his own. The next two sub-chapters on interactive and narrative visualization contain information about methods and approaches in those continuously evolving fields.

1.2 Interactive Visualization

When we use the term interaction in this study, it only refers to the interaction between machine and man: “Human-computer interaction (HCI) is the study of interaction between people (user) and computers. It is often regarded as intersection of computer science, behavioral sciences, design and several other fields of study”(Tripathi, 2011, p. 1).

With the rapid growth of data, the role of data visualization becomes more and more important. Finding new ways to mine the data and teaching computers how to deal with the data becomes an important task for computer scientists and data analyzers. In this matter, machine and human abilities such as visual perception, creativity and general knowledge must concur to deal with large scale data sets (Iwata, Houlsby, & Ghahramani, 2013).

Human-computer interaction is a new scientific field. It started 25 years ago with the revolution of electronic devices and rapid development of computer science. With the birth of interaction the identity of graphical display has changed. Today it is no longer the end product as the visuals on screen turned into temporary tools that can be manipulated by the user. This opens a new chapter in displaying information (Wilhelm 2005, p. 437). Today the designer is no longer limited to a certain amount of space to present the data and highlight the information. Different interaction enables the visual designer to present the data in different levels and only by demand of the user. This leads the new generation of visualization become more reader driven. It transfers a passive reader into an active participant. The user can choose from a variety of choices to investigate through his or her own subject of interest. This creates a balance between author and reader in terms of involvement. It involves the user to play and explore the data and find his own interesting highlights and create his own story through the journey of exploration.

1.3 Narrative Visualization

The Oxford English dictionary defines narrative as “an account of a series of events, fact, etc… given in order and with the establishing of connections between them” (Segel & Heer, 2010b, p. 1139).

Storytelling as a way of communication and conveying information is probably as old as human mankind itself. Good stories can convey huge amounts of information using relatively few words. They are easy to listen to or view and understanding the information integrated in stories is usually easier than understanding information that is coded in lists. Despite the weakness of lists in comparison to stories they are still widely used especially in presentations – e.g. bullet points in overhead slides. In their article *What Storytelling Can Do for Information Visualization* Gershon and Page make the point very clear. According to them a story is worth a thousand pictures and each picture is worth a thousand words. Sticking to that logic means that a story is worth ten thousand words and the value and effectiveness of stories go way beyond simple infographics (Gershon & Page, 2001, pp. 31–33).

Especially media and news corporations have understood the storytelling potential of data visualizations. As well-known publications as the New York Times, Washington Post and the Guardian have done a lot to build data visualization capabilities among their editorial staff (Segel & Heer, 2010a, p. 1139).

These organizations create persuasive interactive visualizations that convey the message to their audience. Jonathan Harris, the visual designer and creator of we feel fine considers himself a story teller and a visualization designer in the first place. He defines a story as: “to me, a story can be as small as a gesture or as large as a life. But the basic element of a story can probably be summed up with the well-worn who/what/where/when/why/how.” (Segel & Heer, 2010a, p. 1140)

To tell a story with data visualization, a journalist a designer or anyone who has a story that need to be told, must know how to choose the best interaction. Data analysts have to know how to choose the best headline. And a visual designer must know how to analyze and discipline the data according to the message he wants to send. Gershon and Page call this “a show business“. Therefore to produce such media a designer, data analyst and journalist has to overlap each other’s skills and capabilities.

2. Previous work

2.1 Principles of a graphical display

Before we dive into the history of data visualization, I will have a look at the principles of graphical display and see what the characteristics of graphical excellence are. Edward Tufte explains that every graphical display should have particular specifications as following:

- It has to show the data.
- The design of the visualization must attract the reader and focus on the content, and not on the aesthetics and production process of an image.
- It is very important that the graphic does not cause any misinterpretation.
- Big data needs to be integrated in small space and form a perfect combination! Graphical displays persuade the user to compare variables. It has the capability to present different levels of detail.
- The goal has to be clear by means of description, explanations, extra information or by applying narrative techniques.

In the following, what makes graphical visualization superior is when the data is interesting and well designed. The work is done when the complexity of information is shaped to clarity and simplicity. An excellent visualization usually is multivariate and honest to the audience (Tufte, 1986, p. 13).

2.2 History of data visualization

This chapter will give an overview on the history of data visualization and it will take a look at some significant examples. The reason for choosing particular examples is that they show the most relevant characteristics of graphical excellence and are repeatedly being referred to in literature.

In the paper Milestones in the History of Data Visualization: A Case Study in Statistical Historiography Michael Friendly provides a very interesting division and breakdown of the history of visualization. This sub-chapter presents the history of visualization based on Friendly’s categorization.

In 17th century the illustration of time, distance and space was one of the most relevant concerns in the need for more accurate illustration of geographical and physical measurements (Friendly, 2006, p. 4).

According to existing literature one of the most representative illustrations in this century was drawn by Michael Florent van Langren in 1644. Edward Tufte explains that this was “one of the earliest visual representations of statistic data“ (Tufte, 1997, p. 15).

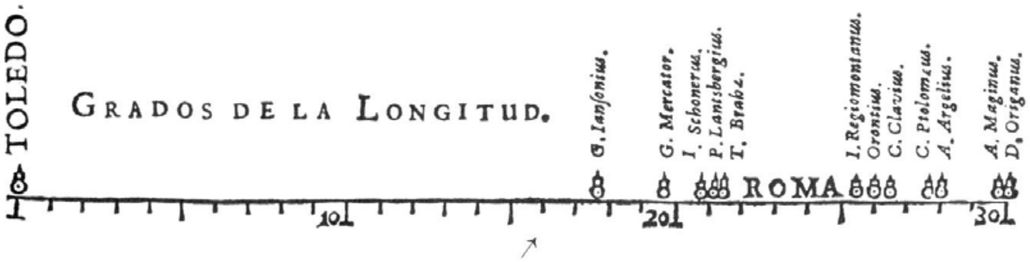


Figure 4. Early one-dimensional map

The figure 4 shows 12 various estimates representing the distance between Toledo and Rome. It contains the difference in longitudes between the two mentioned cities including names of astronomers who provided each observation.

The purpose of the graph was to advance Langren’s own method for the determination of longitude. “Although the graph is a one-dimensional map of data, the chart is remarkably advanced for its time.” (Tufte, 1997, p. 15) Therefore Tufte and a several other scholars refer to the illustration as one of the pioneer samples in 17th century.

The liberation of two dimensional space was one of the most important events in 18th century. In 1765, J.H.lambert explained a general graphical grid for depicting systematic relation between measured quantities:

“We have in general two variable quantities, x, y which will be collated with one another by observation, so that we can determine for each value of x, which may be considered as an abscissa, the corresponding ordinate y. were the experiment or observations completely accurate, these ordinates would give a number of points through which a straight or curved line should be drawn” (Hyrkäs 2000).

It was also during the 18th century when William Playfair invented the bar chart (1786), the pie chart and the circle graph (1801). The following illustration gives an impression on the cutting edge way of illustration by Playfair in the 18th century.

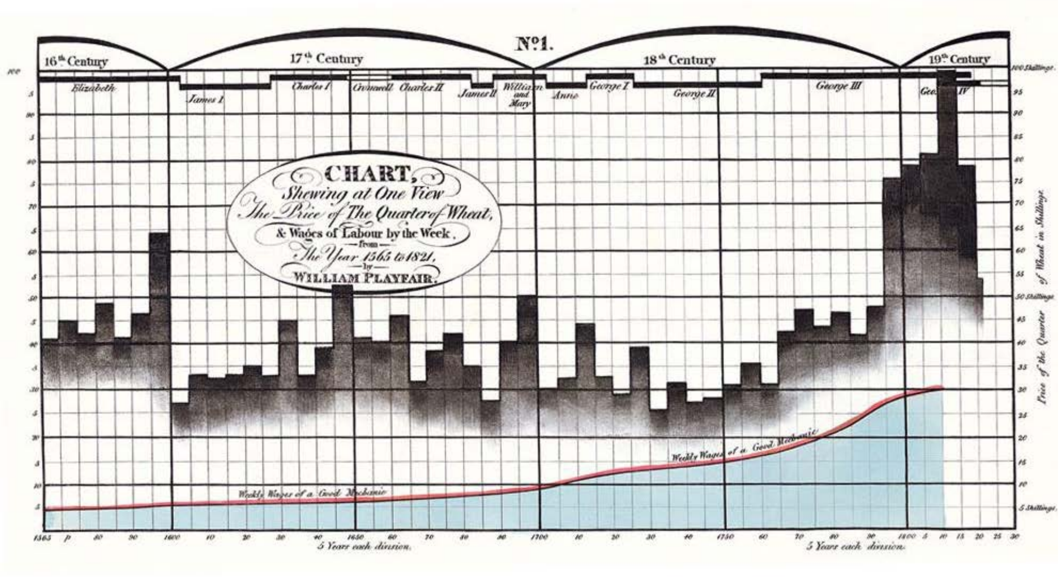


Figure 5. Price, wages and regions illustration in time by Playfair

In figure 5, Playfair presents the relation between the price of wheat and wages in the course of time. He plotted three parameters: price, wages and reigns of British Kings and Queens between 1565 and 1820 (Tufte, 1986, p. 34).

His unique approach to integrate different dimensions and attributes in one single image is very similar to today’s approach of presenting data. It resembles modern design in the way the author divides the screen into smaller sections. Thereby he helps the audience to compare different parameters in order to understand the meaning.

The 18th century also was a starting point for thematic mappings entering the field of visualization. Geologic, economic and medical data were mapped for different purposes. However, no earlier than in the second half of the 19th century John Snow plotted the locations of deaths from cholera in the center of London (Tufte, 1986, p. 24). See the figure 6.

Figure 6. Maps of the 1854 Broad Street Pump Outbreak

Snow marked the deaths with dots. He set them in relation to the water pumps which he marked with crosses. The interesting point in this example is the story that the reader can extract from the map. The observation shows that cholera occurred mostly among those who lived nearby the Broad street water pump and most probably used that pump for drinking. This is a great example for a visual data analysis.

Another great example of this time is the work of Charles Joseph Minard, a French civil engineer. In 1869, he illustrated the flow map of Napoleon’s army in Russia (Tufte, 1986, p. 40). See the figure 7.

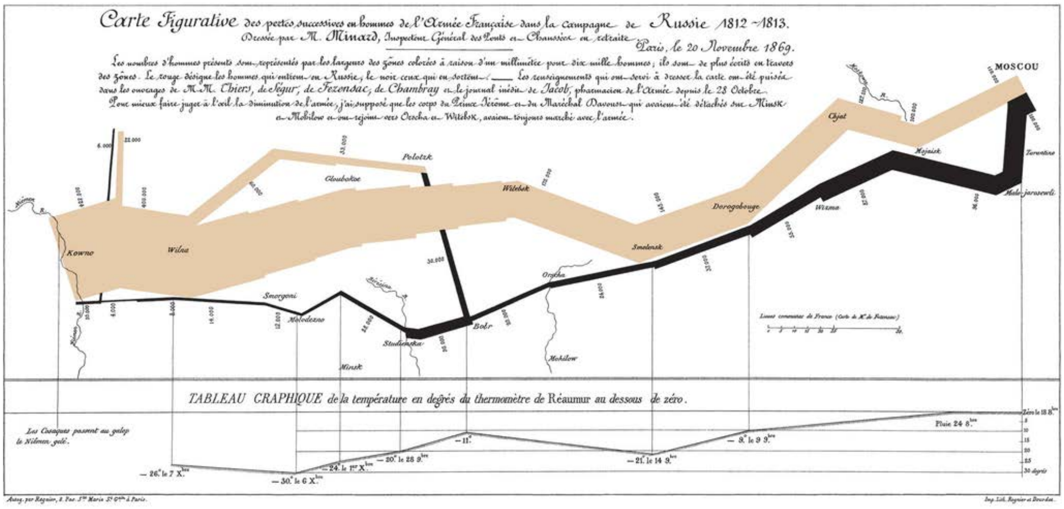


Figure 7. Charles Minard's flow map of Napoleon's March

The thick tan line in left side of the map illustrates the number of Napoleon's army (422.000 men) in June 1812. The width of the tan band shows the size of the army. The observation shows that after 4 months Napoleon's army arrived in Moscow with only 100.000 men left which is less than a quarter. The path of retreat of that army is shown by the black band and it is connected to temperature scale and date. As the illustration shows, Napoleon came back to Poland with only 10.000 men left alive – a tremendous blood spell (Tufte, 1986, p. 40).

One of the remarkable characteristics of the march of Napoleon is the implementation of time-series into spatial dimensions, so that the data is moving over the time. It does follow the basic rules of story-telling. The illustration answers the four major questions of who, when, where and what happened. In 1812 Napoleon's army started its march to Russia and because of the weather condition (coldness) they lost 412.000 men until arriving back in Poland in 1813. According to Friendly, the time period between 1850 and 1900 is called the *Golden Age* (Friendly, 2006, p. 5).

"A number of factors contributed to this "Golden Age" of statistical graphing: the industrial revolution, which created the modern business; official government statistical offices, to support an increasingly aware and global populace; and a growing recognition for the importance of numerical data in social planning, medicine, military, industrialization, commerce, and transportation. Statistical Theory also provided the means to make sense of large datasets.” (Jon Hazell, n.d.)

The 20th century brought important developments for the field of data visualization. During the 1930s, statistical graphics started to be widely adopted. They entered textbooks and educational materials where they helped to make complex matters more understandable. Beyond that data visualizations were used for presentations in the business and science world as well as in the public sector. (Friendly, 2006, p. 6)

Michael Friendly refers to the years between 1950 and 1975 as the time when data visualization was reborn. He claims that three major developments propelled data visualization in to a new phase:

- The invention of exploratory data analysis by John W.Tukey in the USA.
- The publication of the book *Semiologie Graphique* by Jacques Bertin in France.
- The Data processing with computers which have developed the high resolution graphics.

He continued that the development in computer science research, data analysis and display and input technology brought a new models, languages and software packages. These tools spurred a large growth in data visualization. (Friendly, 2006, p. 7)

In 1984, Apple offered the first affordable computer for private households which enabled the user to interact with graphics as shown on screen. Steve Jobs himself presented this innovation in the form of a dialogue with the machine that demonstrated the dawning power of interaction: “You’ve just seen some pictures of Macintosh. Now I’d like to show you Macintosh in person. All of the images you are about to see on the large screen, will be generated by what’s in that bag. Now we’ve done a lot of talking about Macintosh recently. But today, for the first time ever, I’d like to let Macintosh speak for himself.” And Macintosh answered with its synthetic voice: “Hello, I am Macintosh. It sure is great to get out of that bag.(Anon n.d.)”

Back to examples I would like to point out one of illustrations that Edward Tufte included in his collection. The name of that image is New reduction of the Lick Catalog of galaxies in 1974 (Tufte, 1986, p. 26). The illustration as shown in figure 8 represents the distribution of 1.3 million galaxies. It encoded the sky into 1024×2222 rectangles. The darker the tone is the greater the number of galaxies counted in that specific part of the universe.

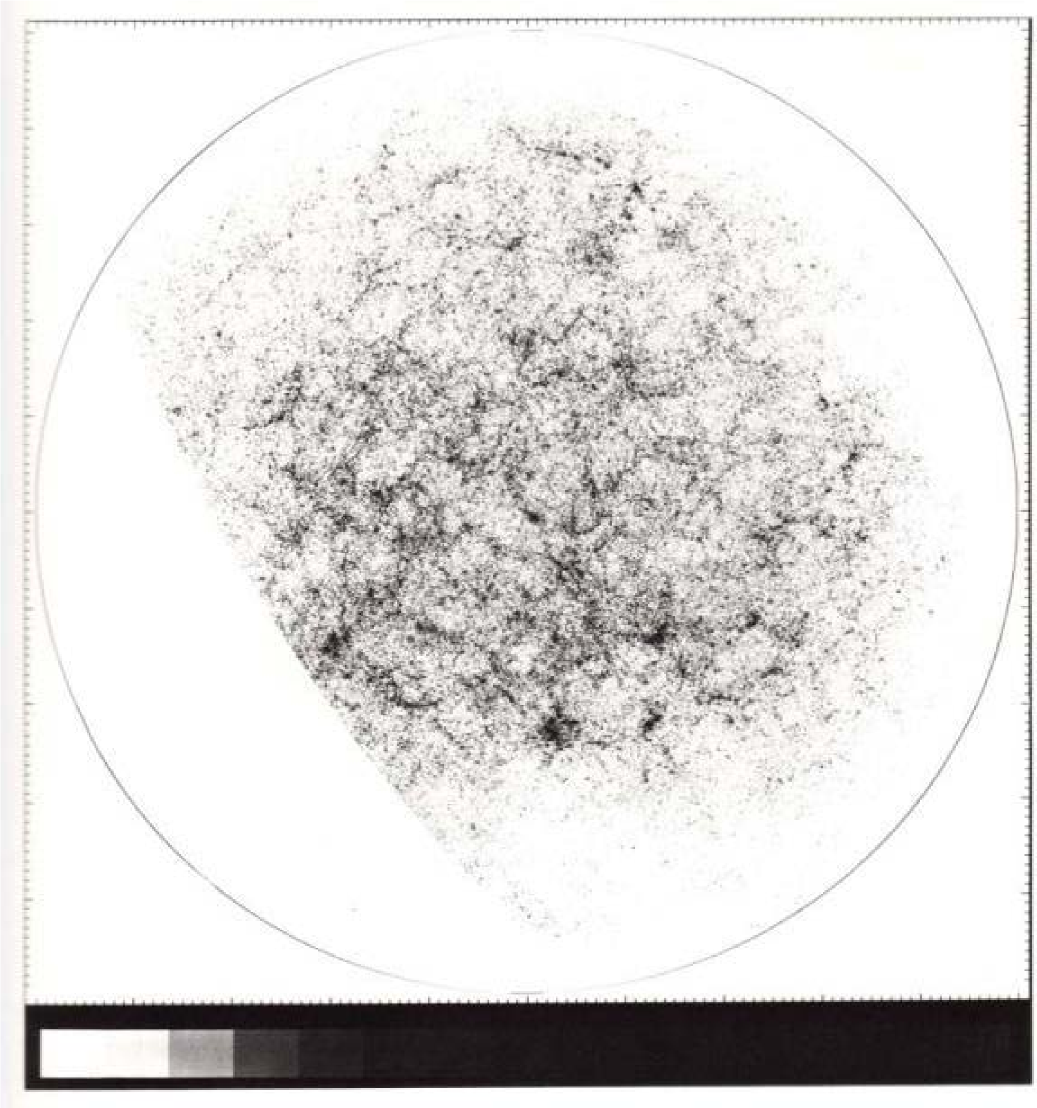


Figure 8. New reduction of the Lick catalog of galaxies

The figure 8 Is the representation of galaxies by Michael Seldner, B. H. Siebers, Edward J.Groth and P. James Peebles. What makes this representation of data unique is the implementation of a very large amount of information in a very limited space resulting in a high density of information with regards to space needed.

The 20th century New York Times was one the pioneers in presenting data in a form of visualization. The figure below presents the daily report on the maximum and minimum temperature over a one year period (1980) in New York city.

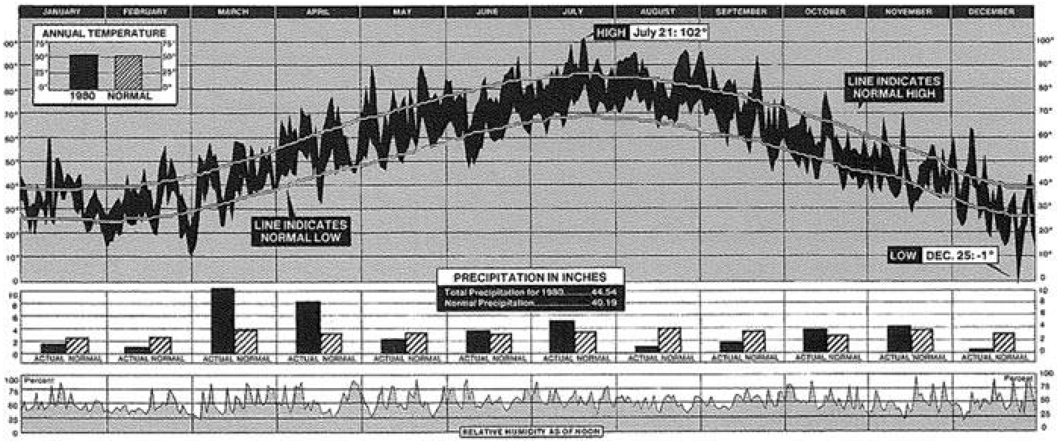


Figure 9. New York City weather report for 1980

This visual (figure 9) is a good example of an accurate explanation of events with time series. Although it contains a significant amount of data, the user can easily understand the message and make comparisons.

Today we are living in a century where the amount of data which is generated in a three years' period equals the amount of data that has been generated in the complete history of human mankind altogether (Keim, 2002, p. 1).

In order to extract enough useful information in a feasible amount of time, each designer must be aware of a wide variety of illustration techniques. The next sub-chapter introduces the principles and classification of different available visualizations.

2.3 Classification of visualization

As mentioned in the previous chapter, data visualization is a broad field of study. And information visualization is one of the sub-groups of the data visualization family which focuses on the illustration of non-physical information. This section will give an overview of how information can be represented in a visual from. It will introduce the characteristics of different visualizations as well as mapping and encoding techniques. Illinsky & Steele in *Designing Data Visualizations* created a very interesting classification for visualizations. First we will take a look at their point of view.

2.3.1 Infographics vs. Data Visualization

Illinsky & Steele stated that there are a few characteristic which distinguish data visualizations from infographics. The authors claim that infographics are mostly drawn manually while data visualizations are in most cases drawn algorithmically. Infographics are exclusively designed for a particular dataset and usually they cannot be updated or adopted for another dataset. On the other hand data visualization is more flexible with different data sets and refreshing the data is not a hard task. Another point which Illinsky and Steele pointed out is the aesthetical aspect of the two different visualizations: infographics are usually aesthetically rich and more appealing in comparison to data visualization but the richness of data volume in data visualization is bigger than infographics. Figure 10 shows the theory of the authors and makes the distinction between infographics and data visualization apparent (Iliinsky & Steele, 2011, p. 5).

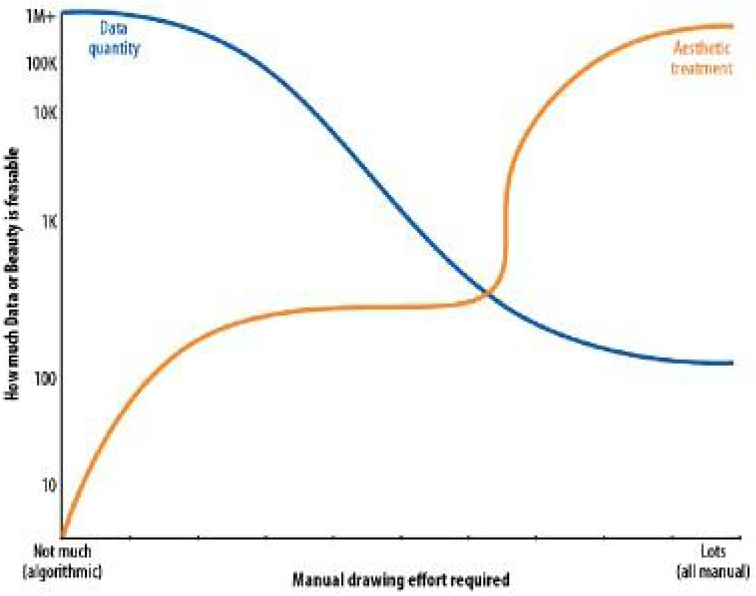


Figure 10. The relation between the amount of data, algorithmic effort and aesthetic treatment.

The figure below (figure 11) is an overview of classification of a visualization based on Illinsky and Steele's theory.

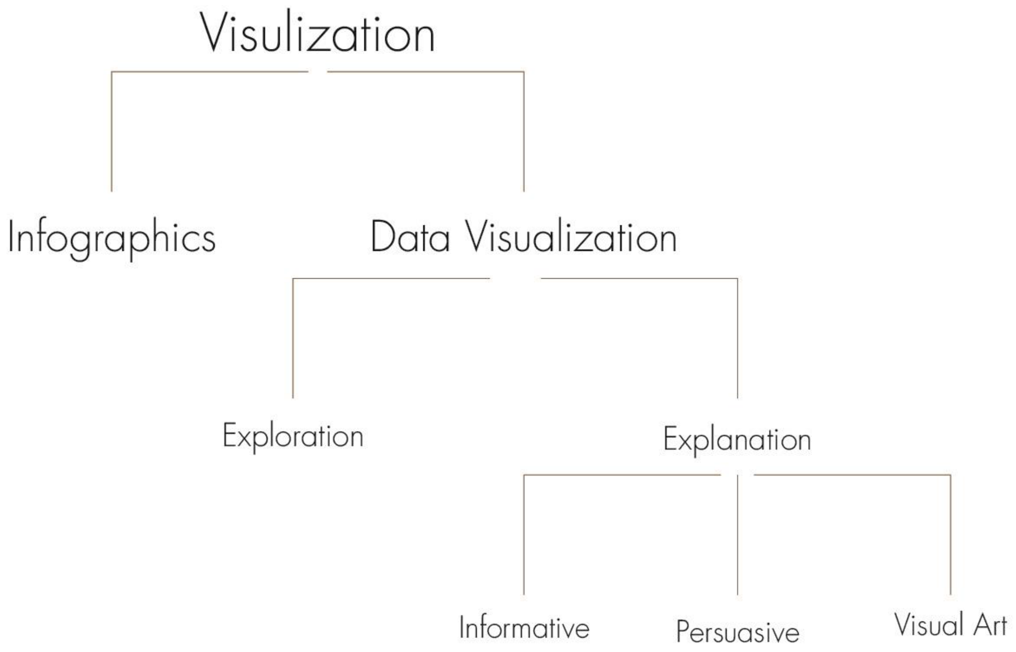


Figure 11. Classification of visualization based on Illinsky and Steele

2.3.2 Exploration vs. Explanation

It is important for the visual designer to know what kind of visualization he wants to present to the user. According to the type he or she chooses the right tools and techniques for the presentation of the information must be found.

A common challenge for designers is to have data but not the ready story yet. In that case the challenge is to analyze the data to come up with the right message to bring across to the user. In that context scholars also talk about exploratory visualization. The designer tries to discover the highlights in the data and therefore he translates the data into visualizations. By that the highlights of the data and the most important message and story behind the data appear. Explanation visualization on the other hand is used when the designer already has a story to tell. Therefore the data needs to be more refined and organized in comparison to the exploration approach. The data is particularly accommodated and highlights are clear. This approach needs more effort from the designer's editorial decision-making. He is not only responsible for analyzing the data and considering graphical aspects, he must also be familiar with authorship techniques to convey the message. Illinsky & Steele add one more category called Hybrids: Exploratory Explanation. This type of visualization is not only limited to the designer, it allows the reader to explore the data by means of interaction. That way the user can discover new insights and take an individual story without being too much directed, inspired or led by the designer (Iliinsky & Steele, 2011, pp. 7–8).

2.3.3 Informative vs. Persuasive vs. Visual Art

Each design requires three resources to really work: the designer, the reader and the data. But what makes a visualization an excellent piece of work is the special relationship between those three resources. As shown in figure 12, Illinsky and Steele present the mentioned relationships as follows: informative, persuasive and visual art.

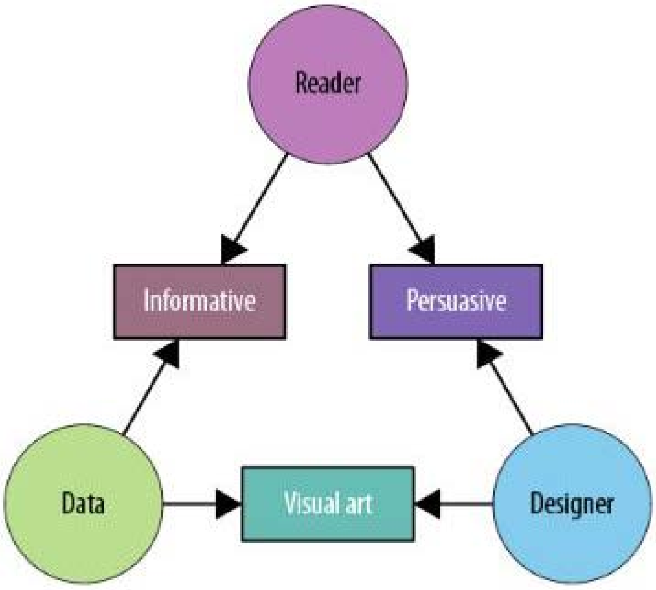


Figure 12. Depend on the nature of the components the two of the three visualization is more dominant.

The three relationships can be explained in the following ways:

1. Informative visualization serves the relationship between the reader and the data, it usually is a summarized version of a big data set, with a direct message which aims to inform and educate the reader.

2. Persuasive visualization serves the relationship between designer and reader. The designer usually uses this approach in order to convince the

user of something. Persuasive visualization is used when the designer has a particular point of view.

3. Finally, visual art - as you can see in the figure above - is located between the designer and the data. And it mainly serves the purpose of translating data into a form of visuals. With this aspect the designer pays more attention to the aesthetics of the design and does not concentrate so much on conveying a message (Iliinsky & Steele, 2011, p. 10).

2.4 Encoding

The term “information visualization” is self-explanatory. In other words the reader by hearing the term knows that there is something – information - which he can extract out of the form of graphics.

When we speak about visualizing information we directly want to present the analysis of the data in the form of graphics which is easy and efficient to understand. What distinguishes information visualization from other illustrations is the message that the author wants to address. Therefore presenting information has to be combined with explanations, labels or instruction which enables the ease of understanding.

In his book *Beautiful Evidence* Edward Tufte points out that: “explanatory, journalistic and scientific images should nearly always be mapped, contextualized and placed on the universal grid. He defines mapped images as “representational images with scales, diagrams, overlays, numbers, words and images … “(Tufte, 2006, p. 11)

In a previous chapter we reviewed how the visual perception works to aid the reader gain information. To use the advantage of our knowledge on human perception we must keep the fact in mind that visual perception can intake only a limited bandwidth of data. Going beyond that limit quickly leads to an overload situation and perception fails. Visual mapping – also referred to as visual encoding - defines which visual structure has to be used to map the data. In other words it means to choose the visual representation that best corresponds to the data.

Based on Ricardo Mazza the process of visual mapping consists of three components:

1. Spatial substrate (the dimensions in physical space where the visual representation is created). The space that the designer draw an object on them.It can be either analog or digital. Although both follow the same rules, they use different techniques.

2. Graphical elements (everything visible that appears in the space). Existing elements that present information on spatial subtracts. It is also known as *form.* The type and shapes of graphical element might be varied according to different information which will be explained further more as following chapter titled as form.

3. And graphical properties (properties of the graphical elements to which the retina of human eye is sensitive).these are the characteristics of each graphical elements such as *color*. It is also can be count as the object's *motion* behavior (Mazza 2009, p. 20).

It is very important that the designer pays attention to the data type and chooses the best practice for the visualization. Usually there is more than one option for one data type and selecting the best visualization technique can be very challenging. To make clear what is meant with this statement let's take a look at two examples of Stephen Few.

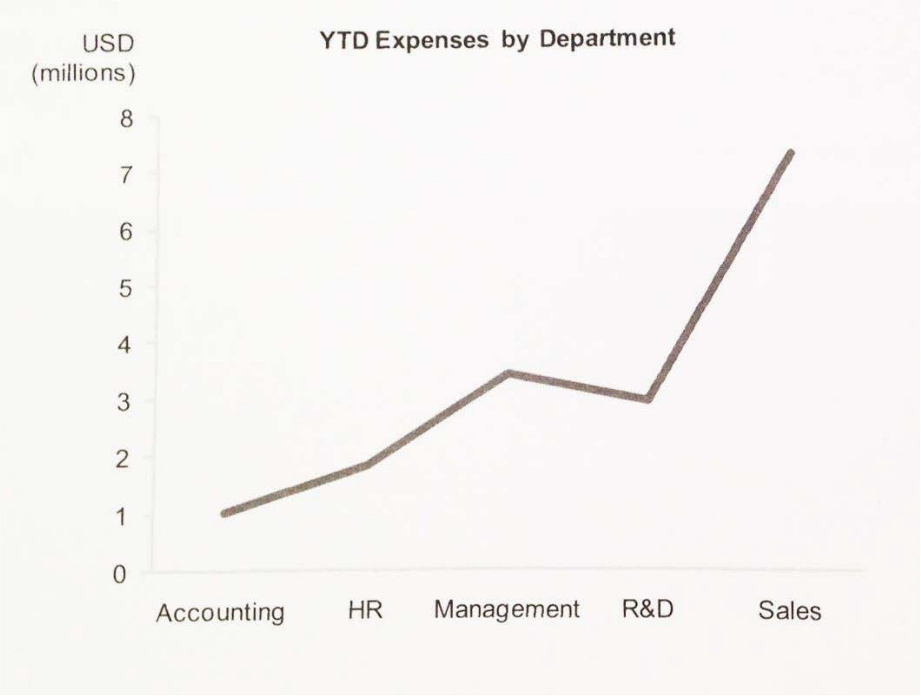


Figure 13. line chart and data type

Figure 13 presents the YTD expenses of different departments. I am sure after a few seconds of observation you will ask yourself, what is the relation between different departments? In fact there is none! Few intentionally chose this misleading way of presenting data. He wanted to show how choosing a wrong visualization style can easily corrupt the message. When we see a line our brain immediately looks for relations. “We are used to interpreting a line like this as indicating an increase or decrease in some variables.” In this particular case we have departments which are categories, and using lines in independent categorical data sets is not a very smart choice, whereas lines work perfectly with relational and time based data sets (Few, 2009, p. 36). Figure 14 presents an independent bar chart with the same data set. As you can see, it is clear that each individual department is shown with its individual YTD expenses. The reader can compare different bars and the comparison is really convenient. In total all information that the designer wants to say can be perceived in a few minute. That can be called effective data visualization.

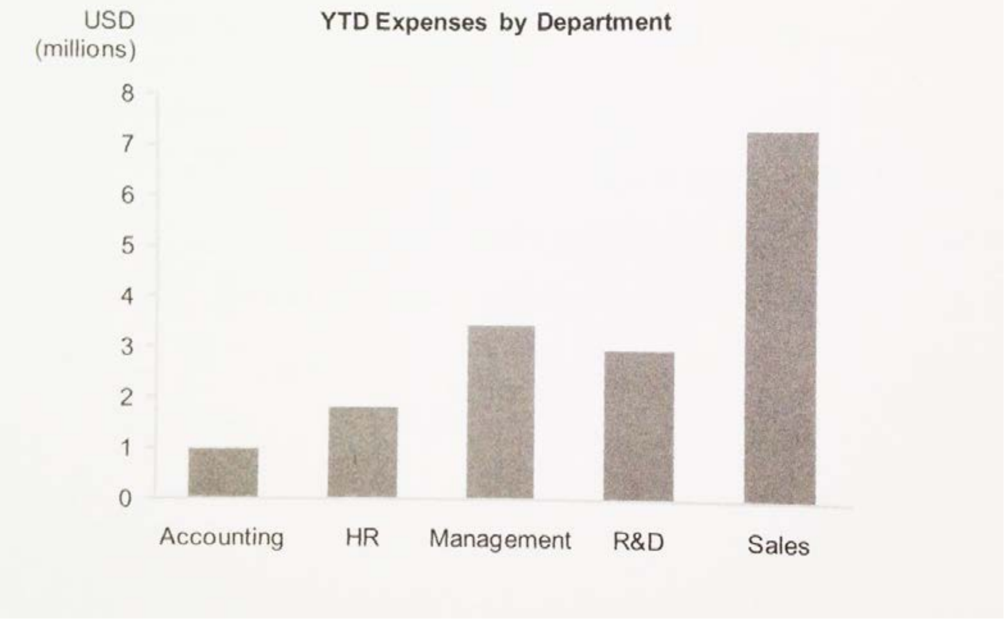


Figure 14. Bar chart and data type

In the next example Few asked his audiences: “Which one of the examples is better to rank and compare the performance of the 10 products?”(Few, 2009, p. 37) See figure 15.

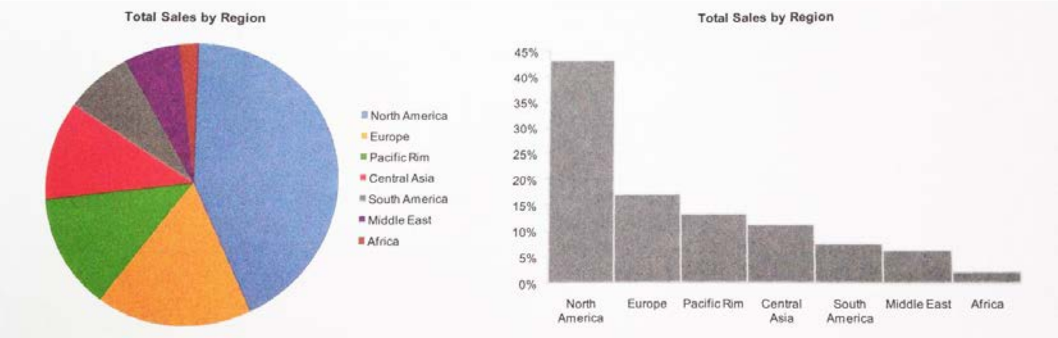


Figure 15. comparison of one data type in two different visualization

He explains that because in pie charts we have to compare 2-D areas and estimate the angel of each piece, it is not a very good choice to present this particular type of data with a pie chart. Presenting the data as a bar chart is a much better choice because our brain can decode the length of the objects very well (Few 2009, pp. 37-38). The question that needs to be raised in this context is why the author claims the comparison of 2-D areas and angel estimation is more difficult than comparing the length of a bar chart? The next paragraphs will give an answer to that enigma.

In data visualization there are general rules to follow to achieve the right results. Those rules can be compared to the grammar of languages. In visualization those rules need to be applied to data properties. Stephan Few presents a table which shows the most accurate and useful attributes. He distinguishes four categories: form, color, spatial position and motion as shown in figure 16 (Few, 2009, p. 39).

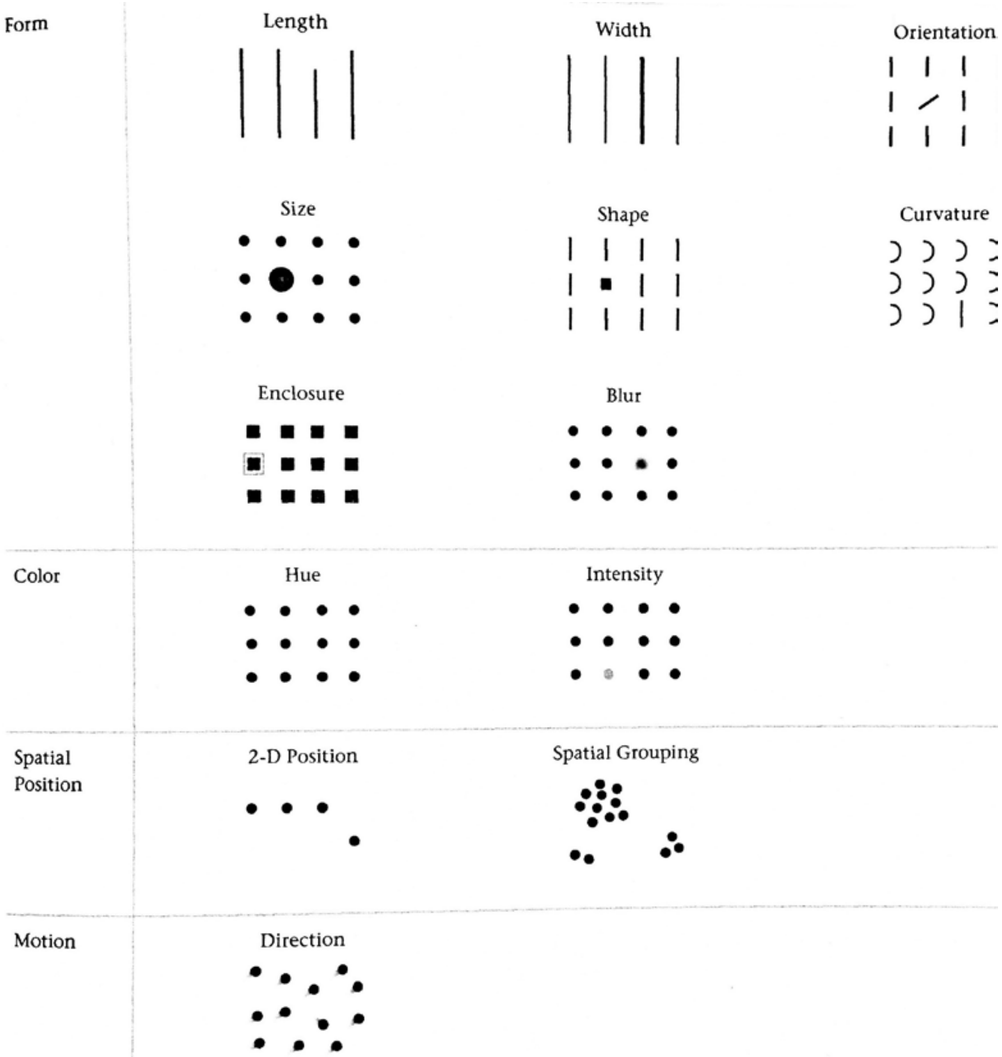


Figure 16. Most accurate attributes for visualization based on Few´s statement

Although it is necessary for the designer to know what attributes he can use in his visualization, accuracy in perception is higher for some attributes than for others. Ricardo Mazza illustrated the perception accuracy for the most relevant graphical elements (Mazza 2009, p. 22) as presented in the figure below (figure 17).

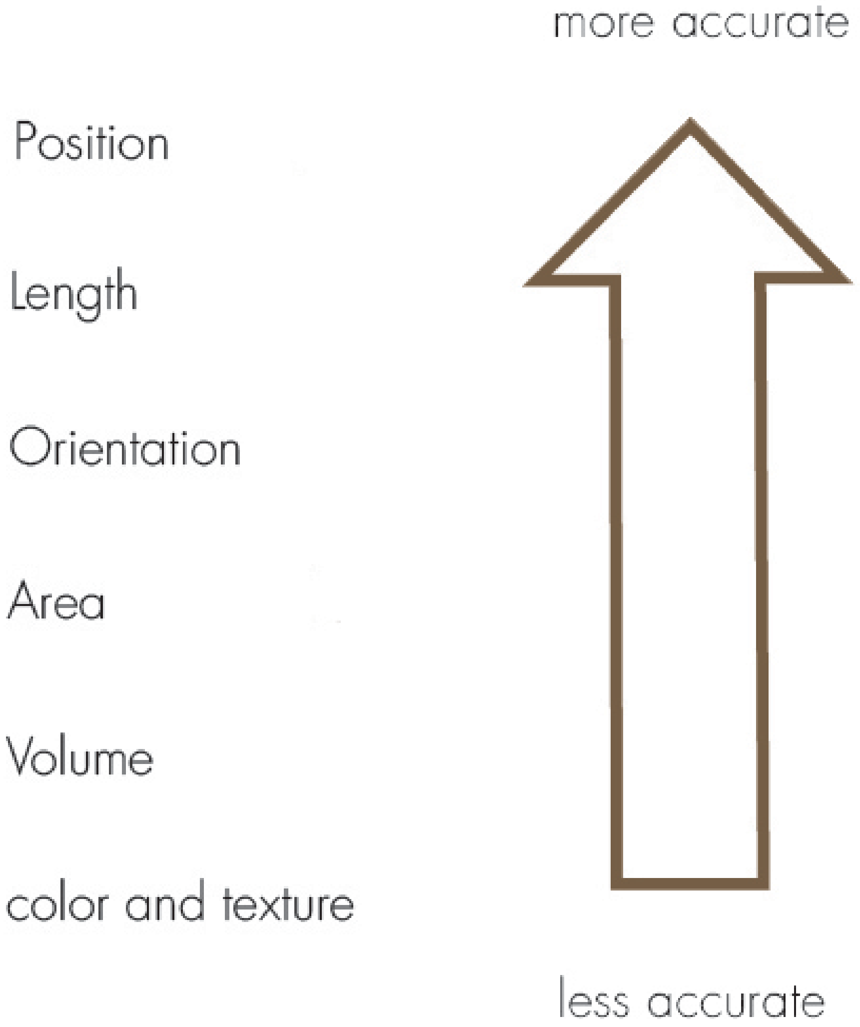


Figure 17. The level of accuracy in different attributes

Now we can answer the question we asked about figure 15. Why is a bar chart more effective than a pie chart? If the message we want to bring across is about the comparison of the sizes of different categories. First of all, it is because the process of recognition of an object's length is more accurate than orientation. The second reason is that in order to read a pie chart the reader must relate each segment of the pie chart with the corresponding color and the legend. While in a bar chart there is no linking process between two dimensional spaces necessary. The reader only has to compare the sizes of the bars and relate them to the legend.

In general, presenting a good visualization demands a great deal of attention from a designer. In figure 16 by Stephen Few we learned about the four major categories: spatial potion, form, color, motion. The Designer needs to be aware during the work of visualizing data. In the next sub-chapter we will take a look at each individual encoding property in more detail.

2.4.1 Spatial position

To understand the basic principles of the spatial position, let's first take a look at the definition of it: "An AREA signifies something on the plane that has a measurable size. This signification applies to the entire area covered by visible mark" (Bertin 1983, p. 44). Bertin continues that area can also vary in position.

In order to visual any data, first thing we need to set up is the space and position that visualization occupies. " Spatial position is often the most important visual encoding you'll have to select" (Iliinsky & Steele, 2011, p. 47). After that the designer has to decide how to position the object on the available area according to the level of organization of the plane. Before we dive into the details we must get familiar with some different levels of variables based on Bertin's theory:

- **Selective variable**: isolation of all correspondences belonging to the same category.
- **Associative variable**: grouping of all correspondences by another variable.
- **Ordered variable**: Assigning an order or ranking to different values.
- **Quantitative variable**: expression of a component by numerical ratio.

Bertin concluded that variation in position is applicable in all the variables mentioned above (Bertin 1983, p. 48). We will review Bertin's example to have a clear idea about the author's point of view.

Bertin explains that positioning objects on a plane can be seen as a category. Figure 18 shows how the eye perceives the dots in different positions as a group of elements without even having a label. Despite the fact that each group has a different number of elements (point, line) first thing which our brain sees in this example is the existence of three major groups in the right top, left top and middle bottom. This clearly demonstrates the power of selective positioning.

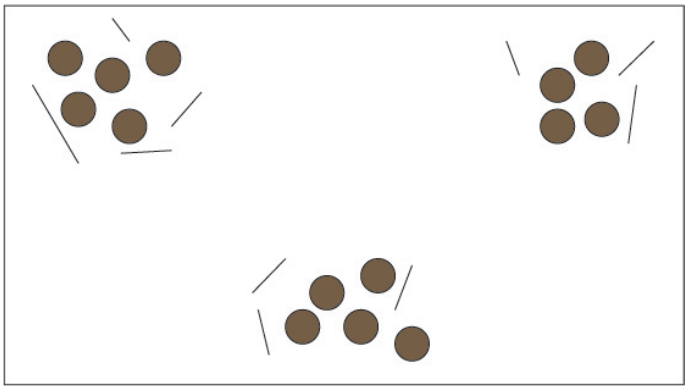


Figure 18. Position attribute

Position can be associative as well. Let’s take a look at Figure 18 once more - what other information does it provide? The answer is that we can tell that there are two different elements (lines and points) in each group. This is the process of grouping the objects by another property (shape).

Another capability of positioning in visualization is ordering. To explain this statement let’s take a look at Figure 19. Bertin explains that this illustration can universally be seen either as A-B-C or C-B-A but never as B-A-C or C-A-B. When we specify the direction then everything is perceived in the same order that authors want their readers to perceive it.

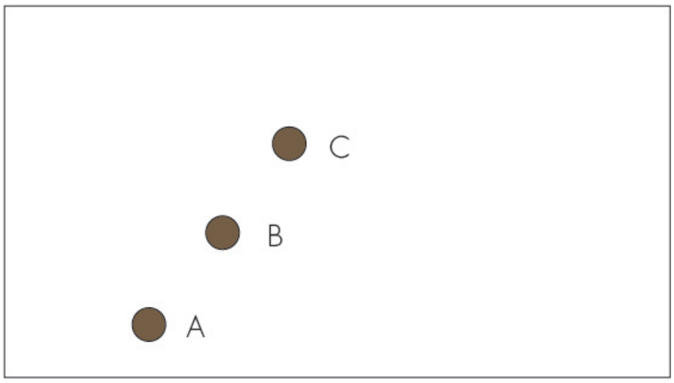


Figure 19. Order attribute

The last variable which is very important is quantitative. Ratio properties in a plane make the user able to estimate and calculate the proportions by numbers. Figure 20 shows three lines with different proportions. Although the reader can not exactly estimate the numbers there is some information which can be easily extracted from the graphic: for example the fact that A > C > B (Bertin 1983, p. 49).

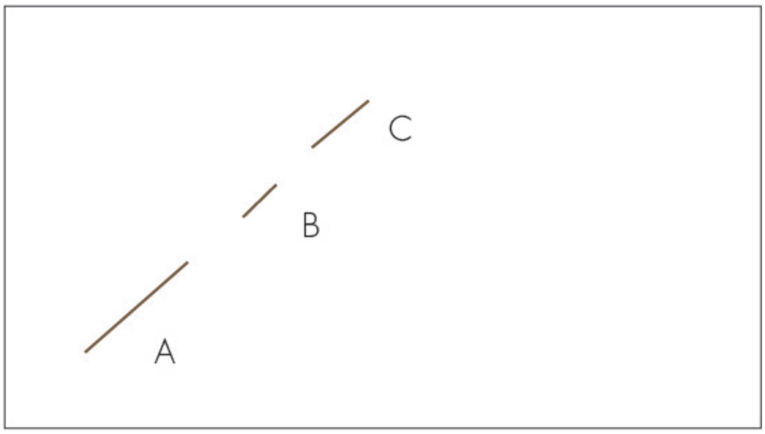


Figure 20. Size attribute

Now we understand why the position on a plane is a powerful encoding tool. Ricardo Mazza explains “the spatial substrate can be defined in terms of axes. In Cartesian space, the spatial substrate corresponds to X- and Y-axes” (Mazza 2009, p. 20) He continues that depending on the type of data axes can be placed differently. Axes play an important role in data visualization. Axes and layouts are also among the members of the positioning family. “A global importance of axes has to be taken into account, which is suitable for most users as well as an individual assessment need to be enabled”(Riehmann et al. 2012, p. 425).

In his book “designing data visualization Illinsky states that axes can contribute values in two ways: top down and bottom up. With the top down approach the reader can limit his search to a specific sub group or subset by only looking at the specific range on the axes. On the other hand if the reader wants to know some data according to one specific entity, he can easily figure out the value of an object in space by looking at the corresponding axes. This is called bottom up approach (Iliinsky & Steele, 2011, p. 48).

Techniques for placements

This subchapter introduces a few techniques that designers have to take into consideration while placing elements on a page.

The majority of readers are reading the screen, page (any physical display which presents an illustration) from top to bottom and the usual assumption is that the elements on the top of the display are more important than the elements in the bottom of the screen (Iliinsky & Steele, 2011, p. 50).

Each designer has to take the logical relationships as serious as the physical relationships, and depending on the situation the designer has to make a decision on which approach (order, size, and position) creates a stronger message to his reader (Iliinsky & Steele, 2011, p. 51).

It is important for the designer to choose the best practice for illustration of his data, although knowing the principles and basic rules aids the designer to have a strong foundation, the importance of cultural, and social background of the audience must not be neglected.

2.4.2 Form

Each illustration is a mixture of different objects with some specific characteristics which appear on the screen. Forms are usually the combination of one or more figures illustrated on the screen. The term form can have different attributes to illustrate the information.

As mentioned before, Ricardo Mazza explains that the process of visual mapping consists of three phases:

1. Spatial substrate

2. Graphical elements

3. Graphical properties

The spatial substrate is already explained in the previous sub chapters. See 2.4.1 Section.

Mazza introduces graphical elements as “everything visible that appears in the space. There are four possible types of visual elements: points, line, surface and volumes”(Mazza 2009, p. 20). See figure 17.

Graphical properties are “properties of graphical elements to which the retina of human eye is very sensitive. They are independent of the position occupied by a visual element in spatial substrate”(Mazza 2009, p. 20).

Based on the categorization of Stephen Few, the graphical properties categorization of form in visualization can be seen as different categories that will be mentioned further. I must stress the point that this categorization is limited to the existing examples based on Few's explanation and it does not cover all possibilities.

- **Length**: length is one of the powerful elements after positioning. The human eye can easily measure the dimensions of objects with different length.
- **Width**: the thickness of an element.
- **Size**: proportion of an object.
- **Shape**: different angels and attribute which particularly belong to one graphic.
- **Curvature**: “the amount by which a geometric object deviates from being flat, or straight”(Gielis 2011).
- **Enclosure**: the surrounding element around an object.
- **Blur**: the loose sharpness of an object. Bullring can be used for highlighting information. The less important information can be blurred and the important one stands out. (Ware, 2004, p. 157).

Each of these graphical properties mentioned above have their specific advantages and disadvantages. ”It is natural to ask which visual dimensions are preattentively stronger and therefore more salient. Unfortunately, this question cannot be answered, because it a ways depends on the strength of the particular feature and the context” (Ware, 2004, p. 152). Based on that, it is necessary that the designer is aware of the utility of the property to choose the best practice.

2.4.3 Color

“Color is normally a surface attribute of an object. The XYZ tristimulus values of a patch of light physically define a color” (Ware, 2004, p. 116). Observation shows the usage of color in data visualization as very common, but using color for illustration demands a great deal of attention from a designer.

Color is not naturally ordered; therefore it cannot be used for any kind of ranking. If a designer wants to sort some elements, the usage of brightness or saturation would be a good choice . There is evidence that certain colors are better choices for labeling different categories than others. In an experiment done by Kawai, M., Uchikawa, K., & Ujike, H. in 1995. They asked subjects to find a chip among many other elements. The surrounding elements in one group were coded in a way that the colors were different but belonged to the same category and in the other group the elements were coded with different color categories. The result was that the group which was looking for an object in the same color category needed more time to find the chip (Ware, 2004, p. 113).

Another interesting experiment mentioned by Colin Ware is the examination which Post and Greene wrote about in 1987. They generated 210 different colors for their audiences. Only eight colors in addition to white were mentioned by the subjects - see Figure 21 (Ware, 2004, p. 114). The experiment shows that only a limited number of colors can be identified as different categories. Therefore using colors for categorization of a data set which has more than 10 color categories might not be as effective as the one with fewer categories.

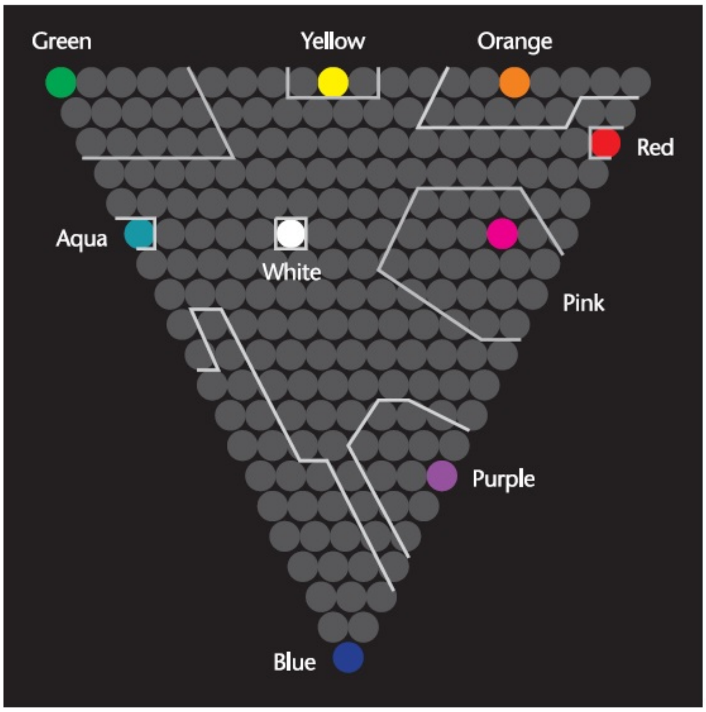


Figure 21. Color range identification by subjects

Color convention is another aspect that has to be mentioned. The perception of a color depends on external factors such as linguistics and psychology (Mazza 2009, 21). Color-coding conventions must be taken into account, like red = hot, red = danger, blue = cold, green = life, green = go. However, it is important to keep in mind that these conventions are not necessarily valid across cultural borders. In China, for example, red means life and good fortune, and green means death (Ware, 2004, p. 125). But how can we transfer the cultural background of humans into a useful way for data presentation? In 1969 Berlin and Kay made an interesting experiment with many cases with different cultural background in order to understand what are the most common colors mentioned by the subjects. Despite the fact that they examined a sample of 98 languages (Berlin & Kay, 1969), the result showed that there are certain colors which were better remembered compared to some other colors and those colors are basically where white, black and red. Green or yellow where equally mentioned and continuously blue, brown and then the four pink, purple, orange and gray were remembered. See figure 22 below (Ware, 2004, p. 112).

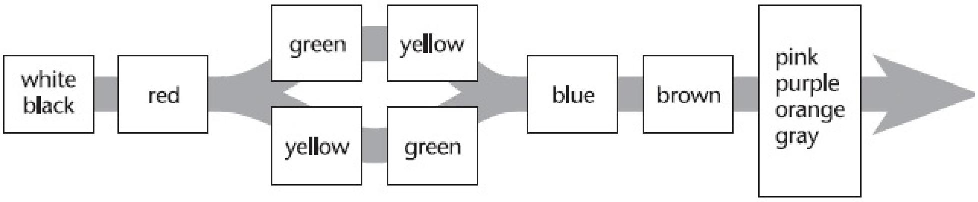


Figure 22. The priority of remembered colors by subjects

Color blindness

A good design is a design that is universal and useful for every user. Based on the target group, the designer must know how to structure his illustration but there are details that must also be taken into consideration. If the designer visualizes a flight simulation application for pilots for example then there is no reason to discuss further! The pilots are already tested and most probably they are not color blind. But the designer always has to pays attention to his users and target group. The following information is extracted from the research about colorblindness which was taken into consideration in this study.

There are different types of blindness: *Tritanopia*, *Protanopia* and *Deuteranopia*(Daniel, n.d., p. 10). Tritanopia is also called blue-yellow color blindness. People suffering from tritanopia confuse blue with green and yellow with violet. Actually it would be more precise to call this blue-green color blindness because usually the colors blue and yellow are not mixed up by affected people. See the figure 23.

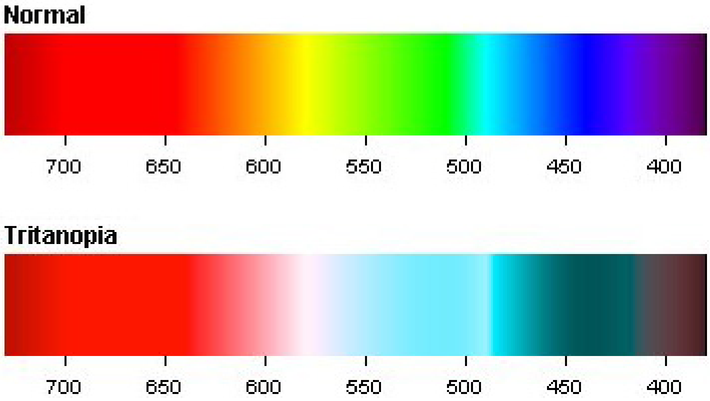


Figure 23. Blue-yellow colorblindness

People affected by Protanopia have difficulties to distinguish blue and green (figure 24) colors and also green and red. The figure below shows how a Patons see the color scheme.

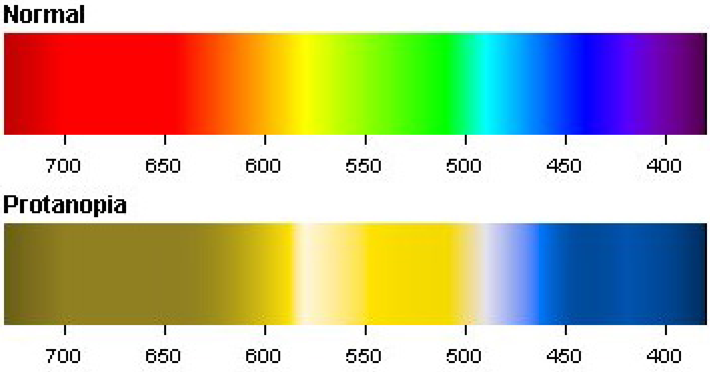


Figure 24. Blue- green color blindness spectrum

The last and most common type is Deuteranopia.it is also calls red-green color blindness. People affected by it can’t distinguish green from red. To them the world is made up of the colors shown in the following figure.(Daniel, n.d., p. 13)

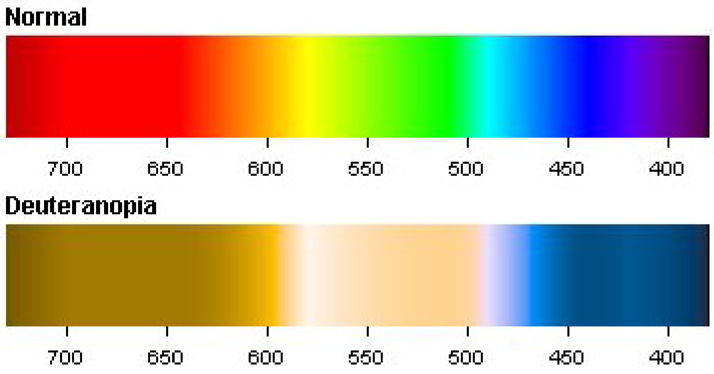


Figure 25. Red-green color blindness spectrum

Based on the explanation above, it is necessary that the designer looks at his or her product and considers the most suitable approach to avoid losing or confusing his potential users. Despite the fact that aesthetical aspects of visualization are very important, functionality is crucial in visualization too. Not to mention, the approach of a designer to choose the right colors is mostly depending on the user groups he wants to reach and the data type he aims to visualize. This means if the designer is dealing with the big number of audience which might have the potential of being colorblind; he either has to avoid using colors which is Indistinguishable or applying other technical approaches such as labeling in order to avoid losing the audience.

2.4.4 Motion

The Role of Motion in Attracting Attention A study by Peterson and Dugas (1972) they showed that subjects can respond in less than 1 second to targets 20 degrees from the line of sight, if the targets are moving.

Although applying motion can be useful for attracting attention, this does not mean it necessarily increases the level of learning. The designer must be careful when and how features still graphics with motion. If the motion such as flickering or plenty number of movements in different directions appears on screen it might cause distraction and becomes annoying (Mazza, 2009, p. 38).

2.5 Data exploration with interaction

As human computer interaction (HCI) was introduced in earlier chapters, the invention of interactivity has opened a new opportunity for visual designers in order to convey their messages. “A good visualization is something that allows us to drill down and find more data about anything that seems important”(Ware, 2004, p. 345).

But why is interaction so important? The answer is simple: it is all about learning. Although this term cannot come under scrutiny in this study, it is important that we consider learning as a necessary factor. If the audience doesn’t extract some information from visualizations or if they don’t remember what they are about, what would be the point of data visualizations? Of course, visualization wants to say something or present some information. Before we dive into detail and possibilities for giving a life to static visualization we might want to take a look at the possibilities that visualization technology donated to us. As Daniel A. Keim has indicated as more as the number of the dimension of data increases the effort for visualization for particular data set increases. He made a classification using visualization techniques according to relations between data that has to visualize and interaction & distortion technique. See figure 26.

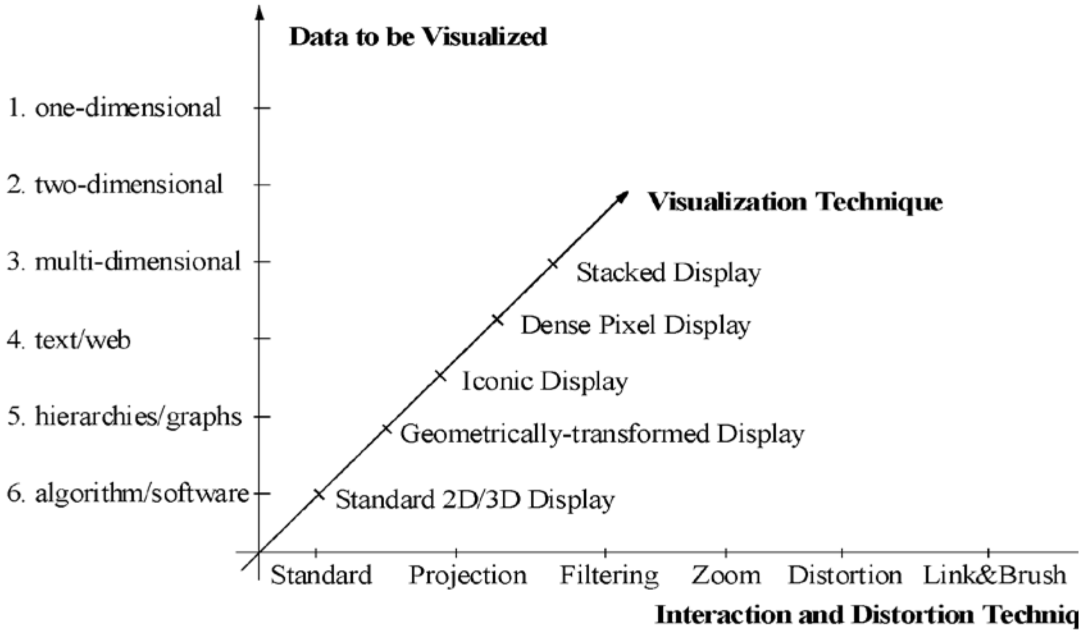


Figure 26. Classification according to relations between and interaction

2.5.1 Classification of interactivity in data visualization

As Stephen Few commented “the effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what information means” (Few, 2009, p. 55). He then introduces a list of possible interaction:

- Comparing
- Sorting
- Adding variables
- Filtering
- Highlighting
- Aggregating
- Re-expressing
- Re-visualizing
- Zooming
- Re-scaling
- Access to details on demand
- Annotating
- Bookmarking

Regardless of what media is used (print, web page, TV or others), the reader applies one or more approaches to receive some information out of visualization. As Ben Shneiderman perceptively states data exploration is a process of flowing actions: Overview on a general information, then zoom and detail and at the end details on demand (Shneiderman, 1996). As this study needs to focus its scope, we will not review all aspects of interaction in detail. But it is necessary to explain the most common and important ones.

Comparison “is the beating heart of data analysis”(Few, 2009, p. 55). Imagine you have a simple bar chart which shows the sales in Northern Germany in 2013. We can show the information in a way like in the figure 27. But what we can extract from the data is very limited information which can be told even in a very simple sentence: the amount of sales in 2013 in northern Germany was 350 euro.

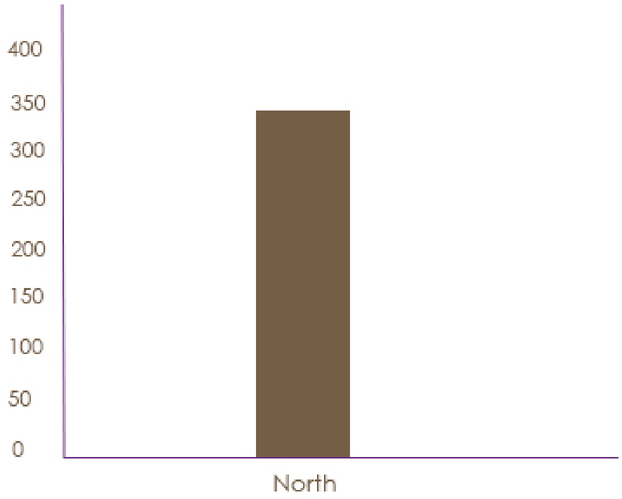


Figure 27. The sales rate in north of Germany

See figure 28 for the overview of sales in Germany as a whole.

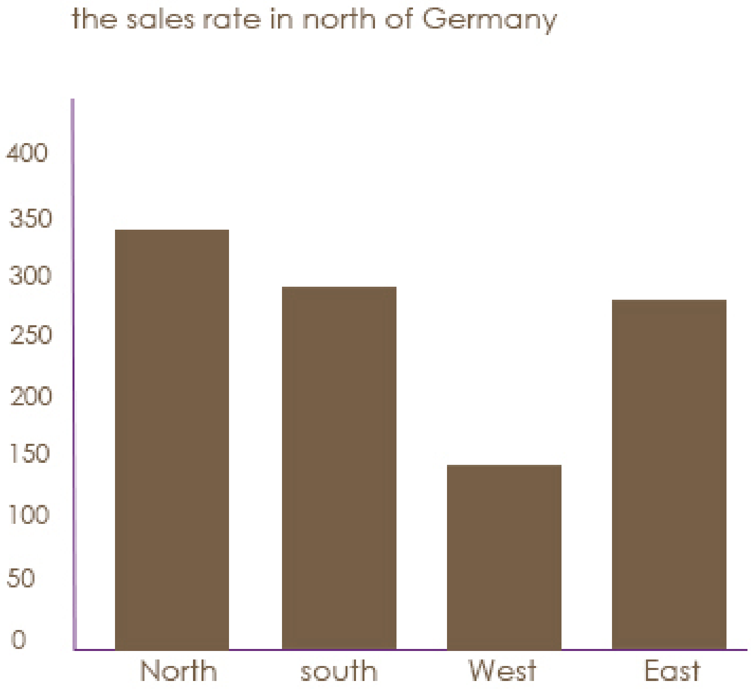


Figure 28. The sales rate in north of Germany

What we can say in this example is way more than in the last example. Northern Germany had the highest amount of sales in comparison to the other regions and south and east almost had the same amount. Meanwhile the west had the lowest amount of sales. You see with only adding 3 more regions (south, west, and east) we found out the north of Germany is leading region. Data visualization is most of the time has greater meaning when there are a few elements to compare!

Few distinguish five different ways of how comparison can work:

- Nominal
- Ranking
- Part-to-whole
- Deviation
- Time-series

Depending on the data the right way has to be chosen carefully.

Filtering is the action of narrowing down of a data into smaller data sets, to see less information. It limits the information into specific category. Filtering "This can be done by a direct selection of the desired subset (browsing) or by a specification of properties of the desired subset (querying)"(Keim, 2002, p. 104). It is the process of Choosing a particular region or check marking one category.

Zooming is the action of having a closer look. It is a very popular technique. Usually it can be done by zooming on specific part of content to enlarge the information. It helps the designer to include the information but not in first layer of the visualization. This means that the reader first focuses on general information and then if he needs more information for one particular case, has this option to dig deeper. Zooming does not only mean to display the data objects larger but it also means that the data representation Automatically changes to present more details on higher zoom levels" (Keim, 2002, p. 104).

Filtering and zooming are quite similar in practices: zooming changes the view to a more detail level while filtering omits the irrelevant part of the content for more concentration.

Accessing detail on demand is a powerful tool to avoid having too much information on actual and primary screen. It only appears when the user needs more information on a specific element (line, bar, circle) which stands for some data amount. It can be done by a pop-up box which appears by hovering on an object (Few, 2009, p. 79).

Re-visualizing is the change of current visualization into another form, for example, switching from bar chart to line chart with selecting the radio button. It can also only change the data according to category in a same chart type for instance the changing view of a pie chart by selecting different years (Few, 2009, p. 70).

2.6 Using narrative structures in interactive data visualization

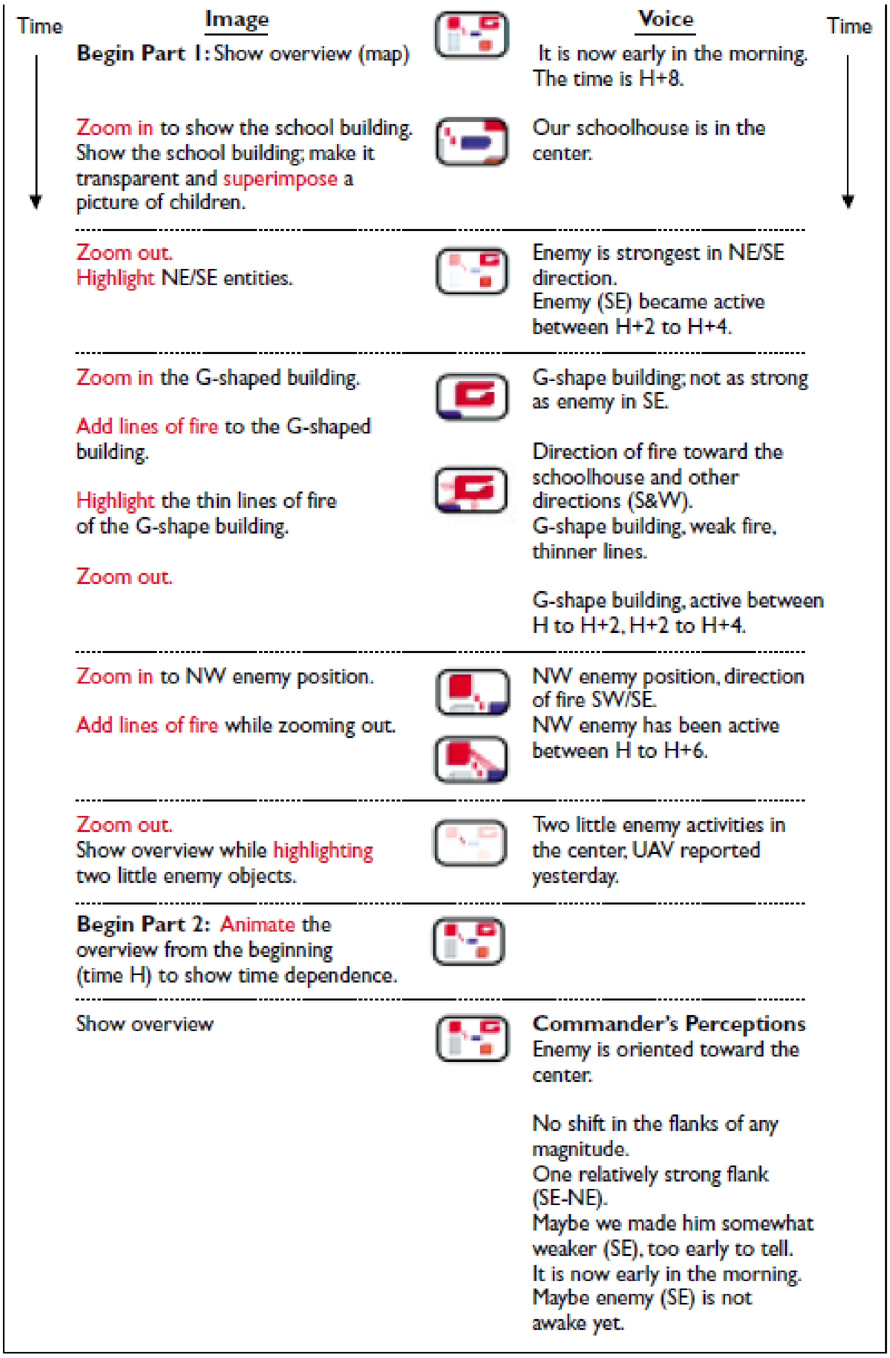
As we discussed in the first chapter, to tell a story with data, the designer needs to define the flow of the story first. He must know who is doing what and where and when it happened. These questions are the foundation of building a story. This study in particular focuses on applying narrative structures to interactive visualizations. In order to create a narrative visualization, first thing we need to create a narrative structure. “Stories are very compelling. We have always used stories to convey information, experiences, ideas and cultural values. Since the invention of writing, printing press, telegraph, radio, movies, and the computer, these technologies and culture have constantly provided us with new and increasingly sophisticated means to tell stories”(Wojtkowski & Wojtkowski, 2001, p. 1).

2.6.1 Principals and structure

In any narrative structure there are series of events that normally consist of beginning, middle and end. The reader follows the journey of the character and experiences the events in order to see the causes of the events and conflicts and at the end the reader will perceive the general message by following the story to the end (Segel & Heer, 2010b, p. 1139). But how to use narrative in an illustration? Visual narrative is the creation of an interesting flow that follows the story using visual language. When it comes to visualization the narratives show their power even stronger. Here, the skills of designers are needed to create and manipulate the display in a way that more important elements stand out and less necessary objects still exist in a scene but in a way that they don’t cause distraction. Using different visual features such as usage of size, color and motion help the designer to create the right illustration. It usually begins with a clear starting point and it contains a series of different scenes. To make a rich visualization in a Story-like style a few actions have to be taken into consideration.

Gershon and Page made a very interesting example to explain how visualization can adopt the structure of a story. The example is the illustration of a hypothetical shooting script. The story is about the attack of criminals against a friendly school. The children are still in the building and surrounded by the enemies. See figure 29.

The authors then explain that to present visualization in a narrative structure the script must be divided into two sections: building the picture and animating the events (Gershon & Page, 2001, p. 34).



Time	**Image**	**Voice**	Time
	Begin Part 1: Show overview (map)	It is now early in the morning. The time is H+8.	
	Zoom in to show the school building. Show the school building; make it transparent and superimpose a picture of children.	Our schoolhouse is in the center.	
	Zoom out. Highlight NE/SE entities.	Enemy is strongest in NE/SE direction. Enemy (SE) became active between H+2 to H+4.	
	Zoom in the G-shaped building.	G-shape building; not as strong as enemy in SE.	
	Add lines of fire to the G-shaped building.	Direction of fire toward the schoolhouse and other directions (S&W).	
	Highlight the thin lines of fire of the G-shape building.	G-shape building, weak fire, thinner lines.	
	Zoom out.	G-shape building, active between H to H+2, H+2 to H+4.	
	Zoom in to NW enemy position.	NW enemy position, direction of fire SW/SE.	
	Add lines of fire while zooming out.	NW enemy has been active between H to H+6.	
	Zoom out. Show overview while highlighting two little enemy objects.	Two little enemy activities in the center, UAV reported yesterday.	
	Begin Part 2: Animate the overview from the beginning (time H) to show time dependence.		
	Show overview	**Commander's Perceptions** Enemy is oriented toward the center. No shift in the flanks of any magnitude. One relatively strong flank (SE-NE). Maybe we made him somewhat weaker (SE), too early to tell. It is now early in the morning. Maybe enemy (SE) is not awake yet.	

Figure 29. Two part script sorted by location, visual operation are in red.

In order to build an image, first an overview of the general situation has to be illustrated. Then to vitalize the story factors such as time have to be added to the visuals (Gershon & Page, 2001, p. 35).

In addition to that also other elements can be added to the structure of the story, but these elements are more or less depending on the style the designer wants to use for telling his story. The following list contains story telling actions which a good visual narrative usually has.

- **Setting place and time**: it means the approach to explain the story and describe the situation.
- **Continuity**: different scenes of visualization must not be disconnected from each other, using an appealing approach to make the transaction between different frames more smoothly and in a continuous tempo is the goal of this technique.
- **Filling gaps**: in order to raise attention and emotion of the spectator the designer uses the characters of the story with superimposed images.
- **Increasing attention**: adding extra elements to raise the attention for example playing a sound or a blinking line.
- **Effective redundancy**: producing a timeline showing the series of events of the story in a same order as they happened. It functions as a more direct way to convey the message.
- **Conflict and ambiguity resolution**: the zooming on specific parts of the general view in order to get more detail and zooming out to reduce the conflict between two different frames. It helps the viewer to make a connection between scenes (Gershon & Page, 2001, p. 35)

Visual narratives are very common in our today’s life. The movies we watch or the comics we read all follow the same order of narrative visualization (Segel & Heer, 2010a, p. 1340). As Gershon and Ward Page both state, the visual designer is using the same visual language which is also used in comic books. Today visualizations even if they are presented on different displays, are using the same metaphors that are also used in comic books: dynamic individual panels which present a central object which is surrounded by other objects that stands for extra information (Gershon & Page, 2001, p. 36).

2.6.2 Design space classification

In order to be more precise about the categorization and according to the limitation of this study, I would like to draw up a general classification of narrative visualization. We will follow footprints of the work of Segel & Heer. The figure below is the overview of their study and their categorization for narrative visualization see figure30.

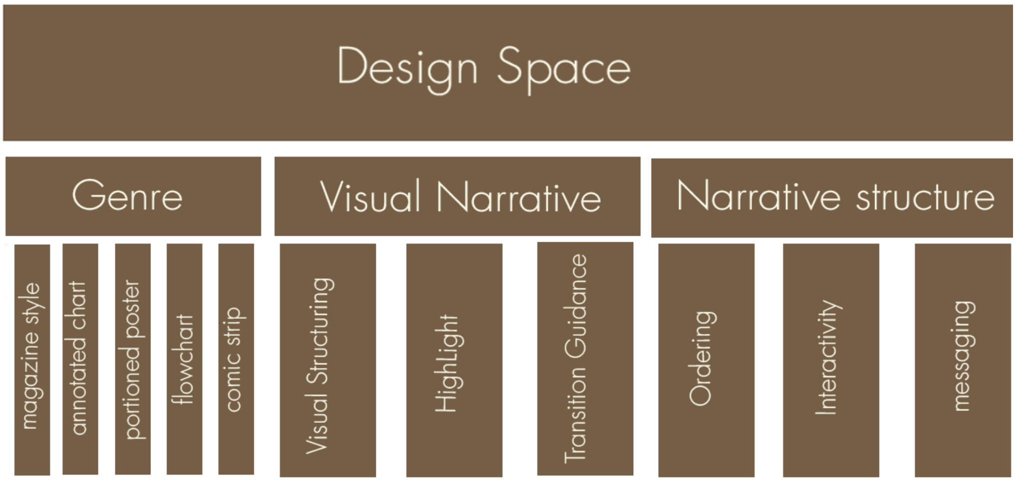


Figure 30. Design space classification in data visualization

Based on the authors the organization of the design space contains three divisions:

- Genre
- Visual narrative tactics
- Narrative structure

The genre of narrative visualization can be divided into 7 categories: magazine style, annotated chart, portioned poster, flowchart, comic strip, slide show and film/video/animation.

As you can observe in figure 30 narrative structure is divided into three sub mechanisms: visual structuring, highlighting, transition and guidance (Segel & Heer, 2010a, p. 1144).

Visual structuring: is a mechanism which guides the user through her journey. It provides the viewer the position that user is currently in and help her to have overview of where she was and where she is going.it is similar to functionality of a map. For example of a progress bar or a checklist are in this category.

On the other hand the mechanism of **highlighting** helps the user to pay attention to a specific object. For example this can be done by using motion or voice effects.

Transition guidance is a tactic that is used in order to make a smoother transition between scenes. As we discussed earlier, it is very important that the author keeps the flow constant. Choosing different viewing angel and using animated transactions are examples for this approach.

And the last design space category is the narrative structure. This category is divided into sub sections by the authors: ordering, interactivity, messaging.

Ordering mostly depends on the structure of the visualization and the characteristics of the design. The designer has the option to define a path that the reader must navigate through (author-driven) or he can give the reader the opportunity to surf through the information without any specific instruction or order (reader-driven). It is also possible that the designer balances the order by giving the reader the opportunity to choose her or his path (mixture of both techniques).

Interactivity is another technique for building a narrative visualization. For building a meaningful and comprehensive visualization there are plenty of options. Depending on the designer’s goal he must choose carefully. possible interactions were explained in section 2.5.1

Finally another important tactic is the **messaging**. Messaging is one of the most powerful tools for creating a story. Most journals and web applications are using this approach to persuade their audience to interact further with data. Small adaptions such as labeling or having a headline or annotation have a great effect in terms of persuasion.

Overall Segel & Heer have drawn our attention to the fact that three aspects are very important in order to have a narrative structure. The ordering structure, the consistency of interaction and the narrative messaging (Segel & Heer, 2010a, p. 1145).

2.6.2.1 Author-driven vs. reader-driven

In the previous section we have argued that using different ordering, is a choice that has to be made according to the goal of visualization. This means that if the designer does not aim to send a specific message to his user then he does not necessarily have to use any messaging and it has a high degree of interactivity in his visualization. He lets the reader freedom to navigate and manipulate the data in his own discipline and order. This kind of visualization is known as reader-driven. On the other hand if the designer for example is a journalist who has to present specific information and he strongly wants to stress his point of view then choosing a reader-driven approach is not a smart choice. The author-driven approach would be more suitable in such situations and provides a linear path with a lot of messaging and no interactivity.

However, in reality we see that existing examples of visualizations are neither absolutely reader-driven nor purely author-driven. Segel and Heer presented three interesting approaches in order to balance the author-driven and reader-driven stories.

It is very important for this study to have an overview on these three approaches.

- **Martini glass**: this approach starts with an author-driven structure where usually no text is involved other than in the form of annotations. It can also be presented as a video or animation. This part is the narrow part of the martini glass. Once the designer sent his message, he allows the user to interact freely with the visualization.
- **Interactive slideshow**: obviously this structure follows the rules of a basic slideshow. The difference between this approach and traditional slide show is interactivity. The user can interact with individual slides while following the flow of the story. This approach is a more balanced approach between reader-driven and author-driven. One of the strengths of this approach is that, it gives the user enough time to manipulate and analyze the information. So the user continues when she is ready with current slide.
- **Drill down story**: this approach starts with a general overview of the content and gives the user the opportunity to choose his or her desire detail aspect category among all the possibilities and drill down into more details. In this ap-

proach the order of choosing the specific theme controlled by the user and the level of freedom is more than slideshow.

In general, visualizations can divided based on the three mentioned sub structures but it is also possible that visualizations integrate a mixture of them (Segel & Heer, 2010a).

2.6.3 Case study on narrative data visualization

After we reviewed the principles of story-like visualization, it is now time to take a look at a few examples which use the same principles in order to visualize their data in the form of a story.

2.6.3.1 Gapminder

One of the examples in interactive data visualization is the work of Hans Rosling. *Gapminder.* It is an interactive slideshow. It contains different topics such as wealth and health of the nations, the trend of HIV, the mortality and life expectancy and many more topics. The figure 31 presents a screen shot of the latest version of the software. To have a closer look at the features of this application, we will analyze the data from two points of view: interactivity and narrative structure.

Figure 31. Gapminder interface

Interactivity: As you can see in the figure 31, at the bottom of the screen there is a progress bar labeled with years and a play button so the user can control the animation. There are two tabs at the top right of the screen, one shows the map view and the other one shows a view of the chart so the user can switch between the two views. The most interesting interaction in Gapminder happens on the axes. Usually axes present the most outstanding information in data visualization and are usually chosen by the designer (Tominski & Abello, n.d.). But Gapminder goes one step further in this matter and gives the audience the power to manipulate the information and find their own interest in the data while choosing from a menu. In figure 32 you can see that both axes can be changed to visualize various data entries.

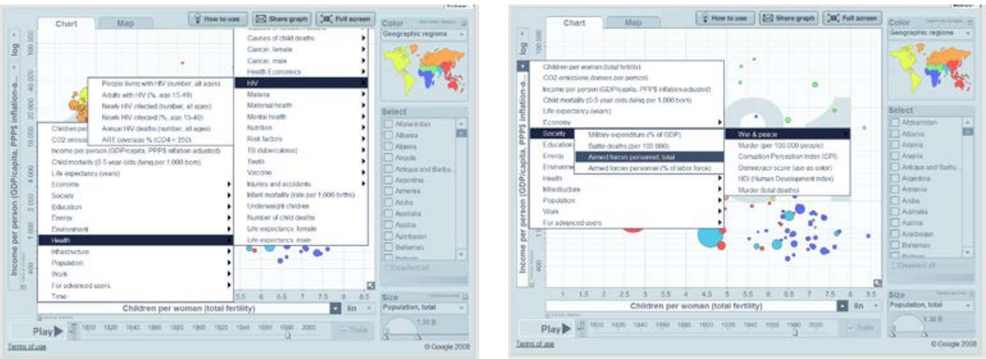


Figure 32. The view of changing axes attributes in Gapminder

Check list is another interaction that Gapminder provides to its users. It enables the user to highlight specific countries. Mouse over is another possible interaction as well as detail on demand. By hovering over bubbles the names of countries pop up. Another manipulation option is the ability to change the size of the bubbles so in case the size of data is not visible enough users can scale the objects.

Narrative structure: As observed Gapminder supports a high level of interactivity which further promotes the narrative structure. It also contains several messaging techniques such as giving a short description of the chart (introductory text). It also has interesting annotations such as labeling the country names on bubbles. From an ordering point of view, Gapminder does not particularly follow a linear trend but it is also possible that the user plays the animation across the time line. However, the user is still free to navigate with the play and stop buttons trough the time line as well.

2.6.3.2 OECD Better Life Index

The organization for economic co-operation and development (OEDC) also provides their data in the form of visualization. The *better life index* by Moritz Stefaner is an interesting visualization in the form of a slideshow. It represents each country in the shape of a flower and each category as petals for this visualization. “This Index allows you to compare well-being across countries, based on 11 topics the OECD has identified as essential, in the areas of material living conditions and quality of life. (“OECD Better Life Index,” n.d.)” These topics consist of subjects such as housing, income, jobs, community and others. Each topic is represented by a particular color and icon. See figure 33.

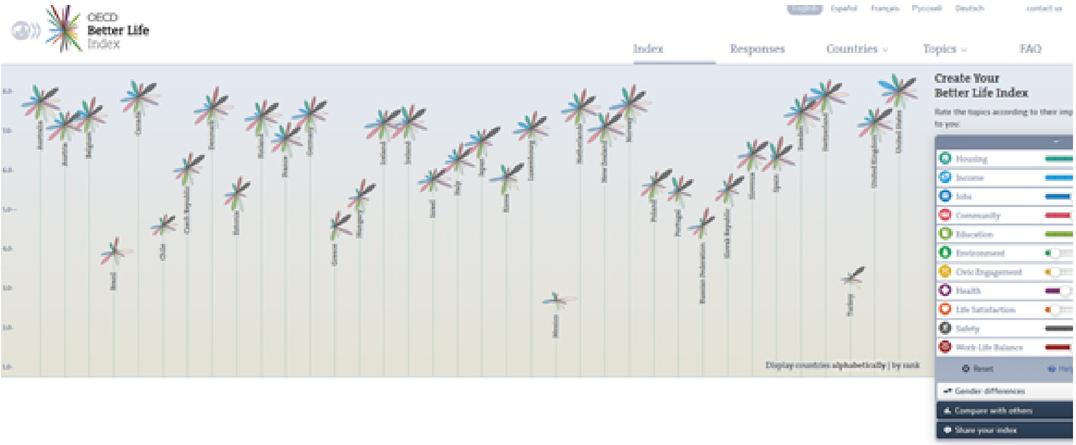


Figure 33. OECD Better Life Index interface

Interactivity: the user is able to manipulate each category by rating them with a scrollbar. The navigation box in this visualization covers a variety of interactivities. For instance the user is able to filter different topics only by sliding down the bars to zero. And it enables the user not only to select a number of categories but it also enables them to rank each category according to their interest. Another possible interaction is the detail on demand feature. It gives users the capability of seeing more details as vertical bar charts. In the next slide Stefaner illustrates a map with the results of 60.000 users which answered to a small form on a platform. And each category is visualized by its own unique color and represents the highest rated topic. The user gets more information when hovers the mouse over a specific bubble (detail on demand).

Narrative structure: from a narrative point of view, as we observed, this is a very powerful interaction technique to support narratives. The OECD Better Life Index is also empowered by using small messaging techniques in order to give a short starting point to the user. According to the level of interactivity this visualization is more directed to the user-driven approach. The smooth transformation of the petals while the user interacts with the slide bar is one of the strongest points of this visualization. It keeps the user moving along with the flow of the story while not really feeling the direct transformation between two data entries. This effect fills the gap between current and next input.

2.6.3.3 Violence and guns in best-selling video games

Another example is a visualization extracted from the Guardian library on 30th of April 2013. The Guardian had a “research into connection between gun violence and video games to violence in real life has been inconclusive.(Guardian US interactive team n.d.)” On top of the screen there is a short description about the investigation that the guardian did according to the top 50 games videos and next to the short summary there is a video giving almost the same information about the usage of violence in video games. In the lower part of the screen there is an illustration of an interactive Chord diagram which shows the relation between *the top 50 games* in 2013, *different weapons types* and *content labels.* See figure 34.

Violence and guns in best-selling video games

Research into connections between gun violence and video games to violence in real life has been inconclusive. As the hype for the latest in the Grand Theft Auto series builds up, a Guardian analysis of the top 50 video games sold in 2012 found more than half contain violent content labels, as assigned by an independent video game rating board based in the US (ESRB). Overall, one-third have weapons that depict real-life firearms

Violent content

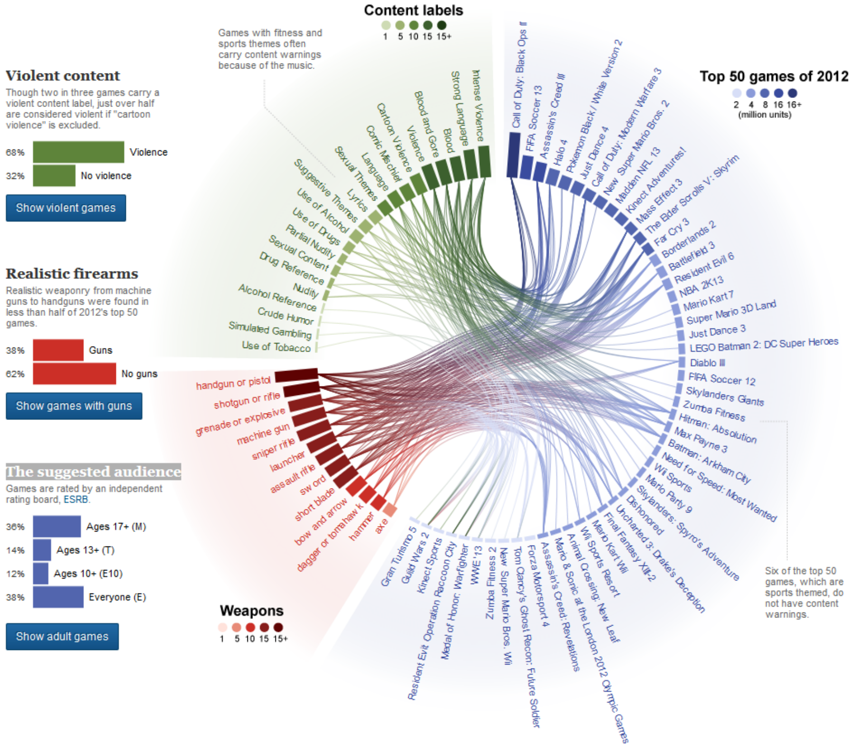
Though two in three games carry a violent content label, just over half are considered violent if "cartoon violence" is excluded

Games with fitness and sports themes often carry content warnings because of the music

Figure 34. Video as introductory part of the Gardian library

Interactivity: One of the significant and major interactions in the chord diagram is **highlighting.** Each subcategory in three major divisions will get highlighted by hovering on them. It is very helpful for the reader to see which video games are using specific weapons or content presenting the violence. Next to the chord diagram there is another box representing 3 bar charts with different colors with headlines including *violent content*, *Realistic firearms* and *the suggested audience*. By hovering on each bar chart the user highlights the related information on chord diagram. This technique is a very strong approach for **filtering**. See figure 35.

Narrative structure: Violence and guns in best-selling video games is an interesting example regards to its unique usage of visualization to tell a story. The first technique that the guardian interactive teams used was establishing a movie at the top of the screen. Although there is no obligation for the user to watch the video, the author uses the more or less martini glass approach to tell his story. The content of the movie is more focused on details of the data and pointing out the most important messages that the media wants to send. Meanwhile the chord chart presenting more general information again the author used the left corner to tell something extra by illustrating bar charts with different colors. As mentioned before the way that the author leads the user to receive the message is very smooth and **intangible**. This visualization is a good example for visualization which is particularly built to give a statement to its users and have strong point of view.



Though two in three games carry a violent content label, just over half are considered violent if "cartoon violence" is excluded.

Realistic weaponry from machine guns to handguns were found in less than half of 2012's top 50 games.

Games are rated by an independent rating board, ESRB.

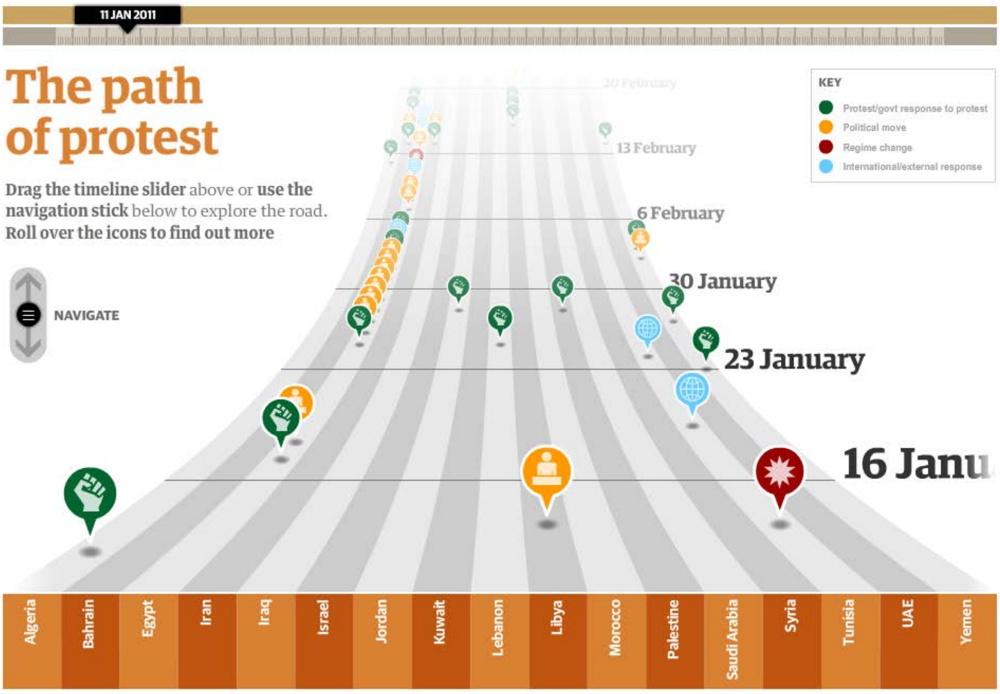
Games with fitness and sports themes often carry content warnings because of the music.

Six of the top 50 games, which are sports themed, do not have content warnings.

Figure 35. The Guardian visual representation of violence in best-selling video games

2.6.3.4 Arab spring: an interactive timeline of Middle East protests

Another interesting example is *The path of protest* by Garry Blight, Sheila Pulham and Paul Torpy for the Guardian newspaper. This interactive data visualization illustrates the pro-democracy rebellions erupted across the Middle East between late 2010 to end of 2011. See the figure 36.



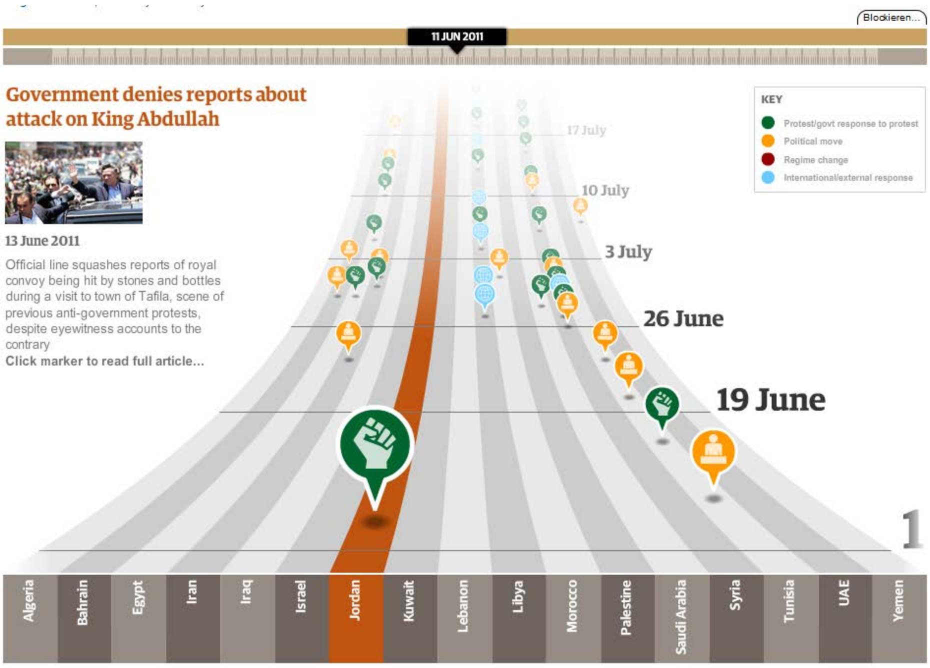
The path of protest

Drag the timeline slider above or use the navigation stick below to explore the road. Roll over the icons to find out more

Figure 36. Interface of Path of Protest by Guardian

In this particular example the visual designers are using the semantic background of humans to make the user more attached to their illustration. They simulated a road and placed the events in their own position according to the country and date of the events. Therefore when the user slides the timeline, the scene is similar to a moving in a path or a road.

Interactivity: the possible interactions in this data visualization consist of the user's **navigation** with the slider which stands for the time of the events. It is also possible to **mouse over** (details on demand) on each event and see a related headline, short summary with a small image (an abstract version of related news). Each event depends on the category (Protest/ government response to protest, political move, regime change and International / external response) has its own color and individual icon.



Government denies reports about attack on King Abdullah

13 June 2011

Official line squashes reports of royal convoy being hit by stones and bottles during a visit to town of Tafila, scene of previous anti-government protests, despite eyewitness accounts to the contrary

Click marker to read full article...

Figure 37. High light feature in interactive timeline

.

Narrative: this is an example of a very rich collection of data and in general will give a rich overview of what happened during 2011 in Arab countries. It answers to the questions such as who did what and where it happened. Narrative techniques such as messaging and highlighting (one click on each event) the road belonging to the particular country helps the user to gain information in flow of the story and time and whenever the user demands more information he or she will be directed to the specific topic. See figure 37.

3. Product safety organization and Content Analysis

3.1 Product Safety Management

This chapter particularity introduces the content and structure of the data used for visualization in this study. In order to understand the importance of topic first I will give an instruction to the chosen topic and the motivation of usage of the existing data set.

Let’s take a look at one short scenario of a person’s one hour of a normal day. It starts with a coffee of course! He goes to kitchen turns on the coffee machine and then bathroom to takes a shower and get ready. Turns on the TV to see what is going on in the world. Then get dressed and perhaps get on the bicycle, bus or car to go to work.

Now let’s make a list of devices that this person uses in this hour.

Let’s name them: coffee, coffee machine, cup, toothpaste, shampoo, bath gel, hairdryer, towels, TV, newspaper, clothes (any kind), bicycle, car or bus. Of course this is not all but let’s assume we use these devices and products in one hour. Although there are many other more possibilities, this product mentioned above is just the example of possible items that the person might uses in one hour period. Yet, plenty of device and very limited information about each product we are using? Are we sure that all of them are safe?

The point is that what we know about each product is very limited. Consumers (us) in some ways trust the organizations that are responsible for checking the safety of the products to test the devices before they enter to the market. Our knowledge about the products we use is limited. Producers are legally required to ensure the safety of the products. They have to make sure that products they produce are in conformity with all legal and standards requirements.

There is a large amount of products available in the market which potentially possesses a risk. These products in first step have to be identified and the risk has to be evaluated through risk assessment strategies. And the next step is to take further action if required. Klindt, Popp and Rösler stated the three main topics that *recall management* takes a look at are:

- Product observation (how the product is doing in the market as well as identification of abnormalities)

- Risk assessment which is systematic evaluation of the observed information on the risks that are related to the product, and if the there is a serious risk further actions have to be taken.

- Further actions include such actions as consumer warnings, redemption of the product and further activities such as correction, seizure, import rejected, ban of the marketing and the product recall (Thomas Klindt , Michael Popp 2008).

Why do recalls exist?

Product recalls have become more and more present in the life of consumers. Reason for that can be found in insufficiently manufactured products and the risks that evolve from them. Overall, there are four aspects which increase the risk potential of products

- The development of individual products including customer integration within the production process, is getting more and more complex(Schließmann 2012, p.9)
- Decreasing Quality of the products through competitive pressure of costs and human recourses(Ritz 2014).
- Globalization of research and development as well as the production process of product components(EUROPA - Consumer Affairs - Unsafe products n.d.).
- Increasing regularities and standards for product safety and market surveillance(Schlüter 1998, p. 4)

The increasing complexity of manufacturing processes as well as, competitive factors and the intensification of existing standards and regularities illustrate the higher possibility of products posing a risk.

According to the European Rapid Exchange System (RAPEX) which records number of notifications for consumer products in being related to a corrective action has raised by 300 percent between 2005 and 2013 in Europe. See figure 38.

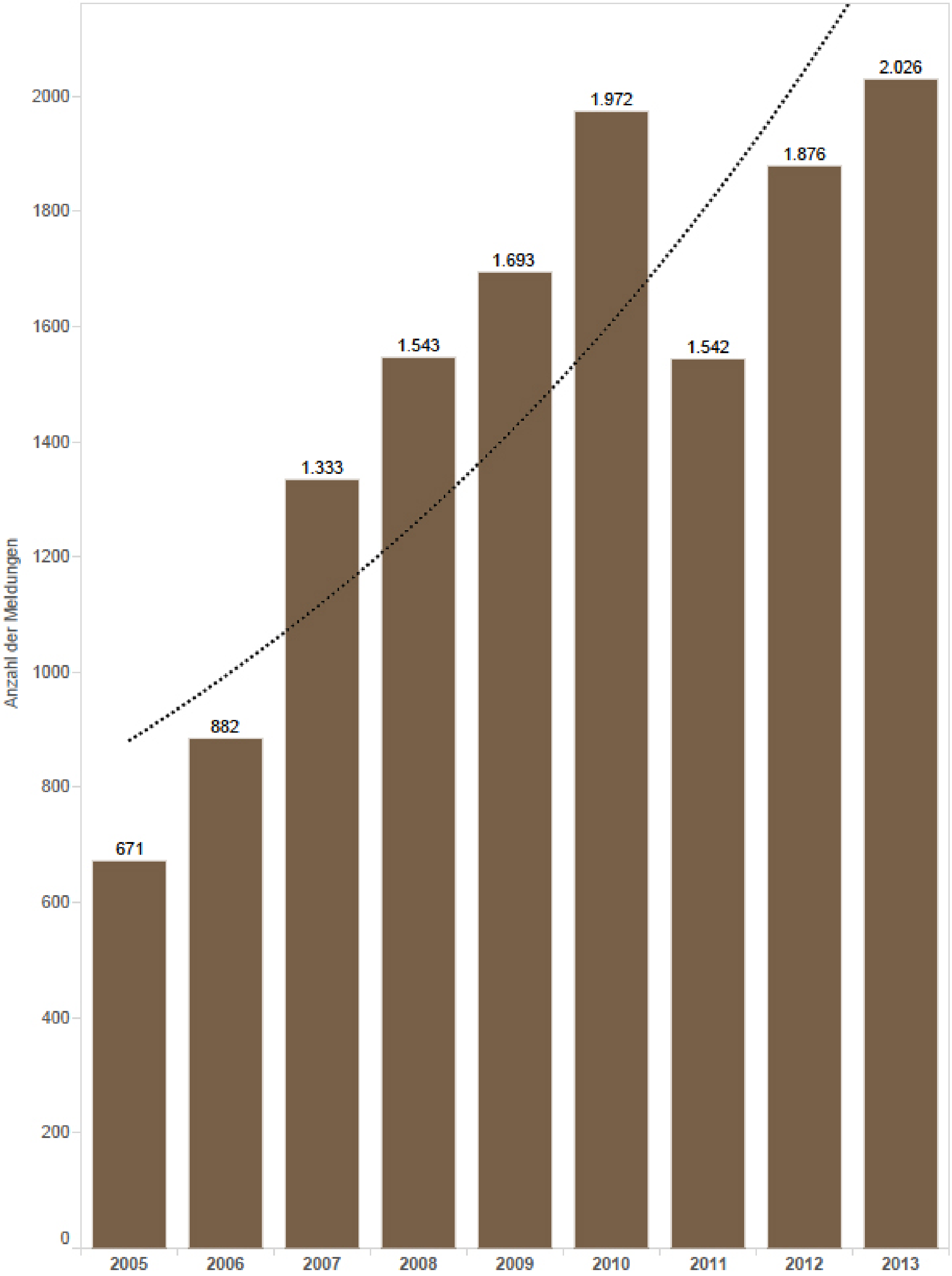


Figure 38. The number of corrective actions taken in past 8 years in Europe.

As a conclusion, not only companies, but also consumers have to understand the increased meaning of recall situations and have to deal with serious risks, by building their awareness of the subject.

Although with the help of automation and computers in general gathering data nowadays is an easy task, in general the information inside databases of the organization which are responsible for gathering information are quite large and usually very tumultuous and disordered. Commonly they are large tables with a significant number of rows and if this table presented to an ordinary consumer or someone who does not have background knowledge of the topic; the level of understanding the data in huge tables is too low or at least it takes too much time.

As mentioned before, the number of products that cause injuries or have a risk of any harm (physical, chemical, sociological) is increasing dramatically and organizations which are responsible to identify and report are sending out these large data sets regularly, organizations such as CPSC (consumer product safety commission). The European Commission has an active role in this field, too. Depending on the style, the process of data collection might be different.

3.2 Data source

For this study, I used the data provided by European Commission. “RAPEX is established as the EU rapid alert system that facilitates the rapid exchange of information between Member States and the Commission on measures taken to prevent or restrict the marketing or use of products posing a serious risk to the health and safety of consumers with the exception of food, pharmaceutical and medical devices, which are covered by other mechanisms”(EUROPA - Consumer Affairs - Unsafe products n.d.) RAPEX provides the list their weekly report for public and is available for download on the European commission website.

European commission has separated the non-food product to its individual alert system called as RAPEX. The organization started their activity since 2004.

Consumer products are all present in our life; they can pose high risk injuries, chemical risk, choking, and so on. Except the food and feed which plays an important role in our daily life and has a direct influence on health, non-food products are the most important objects that we are using and interacting with. It exists in all activities on daily bases. The electronic devices we use every day, the car we drive, the toys we buy for our children, cosmetics, decoration and many more.

3.3 Target group

As reviewed the RAPEX reports, the next question which rises here is who is interested in to these types of data? Based on the investigation, it is possible to align various potential target audiences and their related interests and questions along the product life cycle as follows:

- **Product Development**:
  - What are potential risks that need to be taken into consideration?
  - What kind of accidents has happened with these or similar products so far?
  - Which unintended but foreseeable uses of the products can be derived from the data and what kind of risks may result from it?
  - How can safety protection be integrated in the customer journey in an unobtrusive but effectual way?

- **Sales and Distribution**:
  - For retail-related decisions: How well prepared are crisis management systems of suppliers and which negative occurrences happened with those companies in the past?
  - Market Surveillance and social media monitoring: Where are indicative reports that may refer to possible product crisis situations? How can an early warning system be built up?
  - Which customer complaints have led to product crisis situations in the past?

- **Product Crisis**:
  - Which recall strategies work best in crisis situations?
  - What measures proved to be effective so far and which did not?
  - How could simulations support management teams to come up with the right decisions?
  - What can be learned from crisis situations for the future?

- **Product related damages & compensations**:
  - What can be learned from accident and damage related data that are linked to product usage?
  - Which trends and accumulations can be detected and what are the reasons behind them?
  - Which insurance claims may actually be related to product harms and how can insurance companies find the correct root causes to be only made liable for damages they really need to cover based on the contracts with their clients?

In earlier chapters we stated that when it comes to big data visualization is very handy and further concludes with the assistance of narrative techniques the visual designer is able to improve the level of learning and understanding of large data sets. Next chapter defines goals and aims of this study as well as explanation of the design process from gathering data to creation of the application.

4. Designing an interactive data visualization framework

4.1 The scope of the study and design

Before diving into the practical part of the study, the goal of this study has to be defined. *Consider it* GmbH is a consulting firm that specialized in product recall management and its operational institution. In general when product crisis occur they are incidents that require a lot of attention. In this situation companies often times are not capable to accumulating enough human recourses. Another important point is that the top management of SMEs do not have the knowledge of dealing with recall situations because it generally happens only once in a company's life. In this scenario Consider it GmbH offers a wide service range from crisis preventions to the strategic and operational recall execution. In order to be effective in dealing with product recall situations, companies have to be aware of their economic and reputational damage which can be caused. In order to raise the awareness and provide a better service which suits to companies with no special skills in recall product management. Consider it, is focusing on creation of tools and techniques which provides more information to its audience and customers. One of these tools is illustration of big data in interactive data visualization. The idea of the project is to illustrate trends and developments all around product safety in Europe. The aim is to provide guidance, as well as assistance to manufacturers and distributors. The result shall be increased efficiency of actions taken. The following analysis covers the past 9 years of European alerts on consumer non-food products.

The aim of this study is to create interactive web based visualization and evaluate the effectiveness of using different narrative techniques. In earlier chapters we discussed that the designer's goal is a very important fact in choosing different visualization structures. Sometimes the designer has an important message to tell and his only goal is to give his message via illustration (author-driven) and sometimes the goal is to motivate the audience to think over the topics without having any instructions from the visual designer and raise their own questions and find their own answers (reader-driven). It is also possible that the designer to pursue both aims in his visualization. By that he wants to state an important message as well as let the user create his own story and make his own conclusion.

In general the implementation of the application in this study can be divided into two phases:

- the creation of visual elements and possible interaction
- Using narrative structures to investigate the effectiveness of different methods for the balancing the author-driven and user-driven techniques based on this master thesis goal.

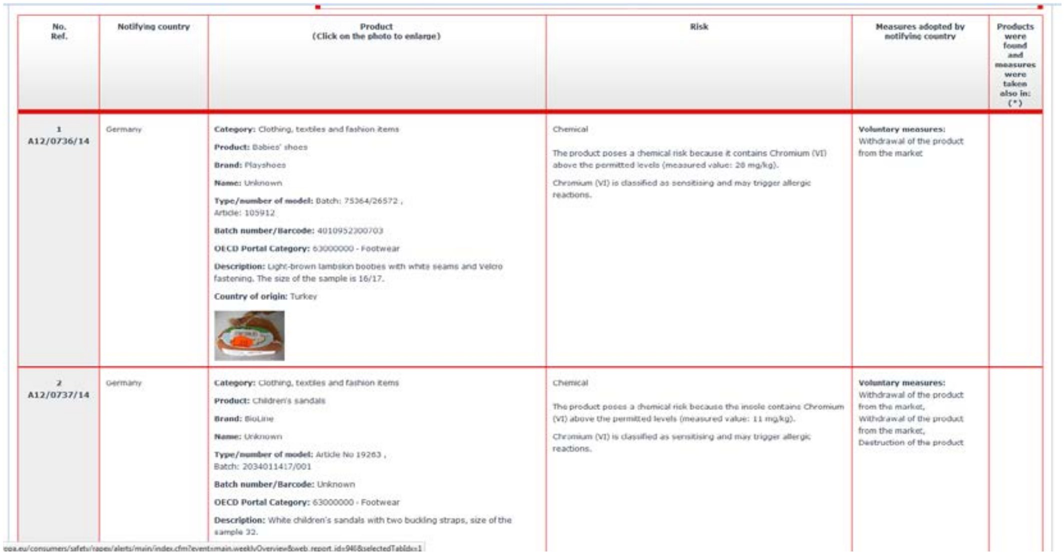
The creation of visual elements and interaction

For the creation of graphics we used basic principles of data visualization which will be further explained in the sub-chapter 4.2.3. To create interaction, according to case study and literature review I used interactions which are usually used to create narratives. This will be further explained.

4.2 From data to interactive data visualization in practice

4.2.1 ETL (Extract, Transform, Load)

The first thing which had to be done was the collection of the data and downloading them from the website. There are three available formats for download (printable version, Excel and XML) and I have chosen XML format because there was a need to adjust the structure and format of the data. And with XML format I was able to program the process and speedup the processes. The data consists of the weekly RAPEX report between 2005 and 2013. Each report in average has about 50 notifications with details on each product including the country of origin, an image, product specifications the risk type and a short explanation as well as the measure that had been taken in the reported case. Figure 39 is screen captured from the RAPEX alert system. The product column consists of category, product, brand, name, batch number, OECD portal category description and country of origin. These categories had to be transferred to individual columns. Therefore i had to program a transaction which put the related information in separated columns.



No. Ref.	Notifying country	Product (Click on the photo to enlarge)	Risk	Measures adopted by notifying country	Products were found and measures were taken also in: (*)
1 A12/0736/14	Germany	**Category:** Clothing, textiles and fashion items **Product:** Babies' shoes **Brand:** Playshoes **Name:** Unknown **Type/number of model:** Batch: 75364/26572 , Article: 105912 **Batch number/Barcode:** 4010952300703 **OECD Portal Category:** 63000000 - Footwear **Description:** Light-brown lambskin boobies with white seams and Velcro fastening. The size of the sample is 16/17. **Country of origin:** Turkey	Chemical The product poses a chemical risk because it contains Chromium (VI) above the permitted levels (measured value: 28 mg/kg). Chromium (VI) is classified as sensitising and may trigger allergic reactions.	**Voluntary measures:** Withdrawal of the product from the market	
2 A12/0737/14	Germany	**Category:** Clothing, textiles and fashion items **Product:** Children's sandals **Brand:** BioLine **Name:** Unknown **Type/number of model:** Article No 19263 , Batch: 2034011417/001 **Batch number/Barcode:** Unknown **OECD Portal Category:** 63000000 - Footwear **Description:** White children's sandals with two buckling straps, size of the sample 32.	Chemical The product poses a chemical risk because the insole contains Chromium (VI) above the permitted levels (measured value: 11 mg/kg). Chromium (VI) is classified as sensitising and may trigger allergic reactions.	**Voluntary measures:** Withdrawal of the product from the market, Withdrawal of the product from the market, Destruction of the product	

opa.eu/consumers/safety/rapex/alerts/main/index.cfm?event=main.weeklyOverview&web_report_id=946&selectedTabIdx=1

Figure 39. Primary data structure in RAPEX online data base

After structuring the data we loaded them into a data base ready for analysis and extract the relevant information.

4.2.2 Content Analysis Methodology

The next step after cleaning the data was to define the goals and extract the most relevant information for the audience. Due to the large volume of the data, extracting relevant information was a difficult task. At this stage of the process I decided to use visualization software to analyze the data and collect the most relevant stories for the visualization.

4.2.2.1 Tableau

In this study i used Tableau software to in to assist in analyzing the big data set. Tableau is a software which gives the user the opportunity to create interactive data visualization based on existing data. This software allows the user to import different formats of data (Excel, access, text file) and based on the data type (dimensions and measures) it gives the user a variety of different options to illustrate the data in the form of significant charts and graphics (pie chart, heat map, Geo map, bar chart, bubble chart, scatterplot and many more…). Figure 40 shows a preview of Tableau software.

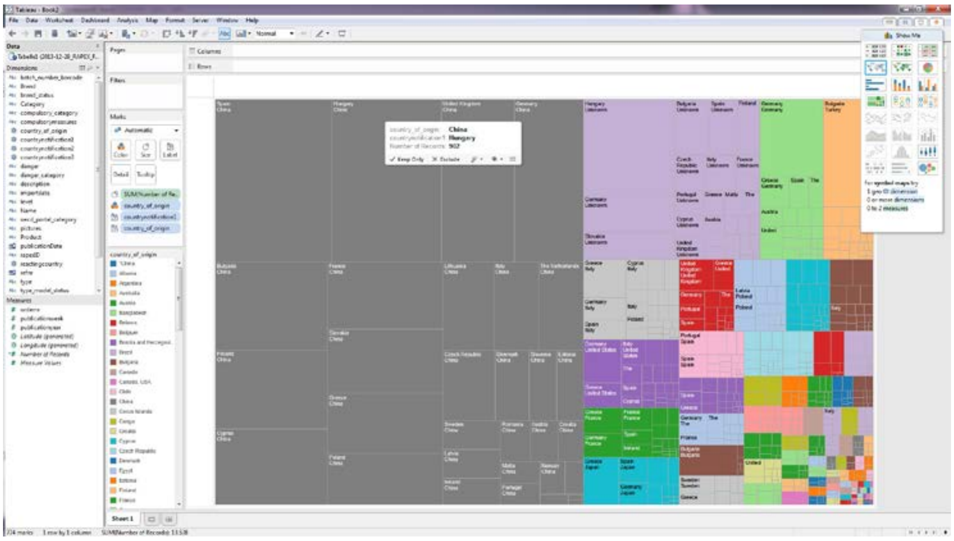


Figure 40. Tableau software interface

4.2.2.2 Finding relevant information

In the next step this study was focusing on using different data sets to generate different visualizations via Tableau software. The aim was to find the most appropriate and understandable visualization in order to show the data. To create a story, selected data should have an interesting and meaningful content which persuades the user to discover further details. It basically should raise questions in the audience's mind. Then they are motivated enough to interact with the application (Yang et al. 2014, pp. 6-8). After exploring the data, and defining the most important story to tell, the best visualization for the specific data types had to be chosen. As discussed before, data visualizations can be created in many different ways but the aim of this study is to find effective ways to send the right message. According to the analysis plenty of interesting and potentially relevant information has been extracted from the data. The next paragraph is the summary of what has been derived from the analysis.

There are a number of products which have certain issues with regards to their safety. Before extracting interesting stories story, certain terms which will be frequently used is defined as following:

Countries of origin are the countries which export products that have safety issues. This category of products basically has a high chance of being harmful and cause danger and for the consumer.

Notifying countries are the countries which found the dangerous products and report them as an object with a high risk of danger. Usually the content of the report includes the type of danger, for example if it is chemical or if it causes allergy and explains which part of the product is harmful.

Product categories are the division of the non-food product types. It varies from motor vehicles to toys and clothes.

Measure taken is basically the action ordered by authorities. There are different types of actions which were already explained in chapter 3.

Number of records is the number of the reported product with safety issues which was reported during the past 9 years.

In order to find relevant stories I created a question and answer section which will be presented as following .these stories uncovered using Tableau software. It is featured with related figure which was screen captured from the Tableau according to the visualization, related question were created.

What is the trend of number of notification in different product categories through the years? The data shows the number of records in different product categories in the last 7 years. It is very interesting that in some categories the number of notifications decreased and in some increased. It is also possible to observe that the number of countries which started to become more active in the matter of product safety was increasing through the last years.

What kinds of actions are taken after notifying products with safety issues? There are different types of reactions that can be taken which will be further explained in the practical part of the study.

Who are the most active countries in terms of notifying dangerous and harmful products? The next interesting point of view is to see which countries are the most caring countries in product safety activities.

Who produces the unsafe product and who notify it? Analysis shows that numerous countries produce unsafe products and many countries are reporting about these products. It is also possible to see what countries have the highest number of notifications and which countries are most active in terms of reporting. The annual numbers of records are very significant for some countries which are worth to be illustrated.

After founding the related story I have created the icons related to the topic shown in table below:

	**What is the trend of number of notification in different product categories through the years?**
	**What kinds of actions are taken after notifying products with safety issues?**
	**Who are the most active countries in terms of notifying dangerous and harmful products?**
	**Who produces the unsafe product and who notify it?**

Table 1. Icons created according to each story

After this phase of the study and defining the intended message that needs to be told to the audience, it is time to dive into the technical part of the work.

4.2.3 Visualization Design Methodology

4.2.3.1 D3 & Adobe Edge

D3

In order to create an interactive visualization, specific skills and tools are needed. According to the fact that I wanted to create a web base application I found D3 one the most powerful tools for data visualization. This study uses D3 to create interactive visualization. D3 is a java script library written by Mike Bostock. The D3 is a tool that let designers manipulate the elements of a web page in a context of a data set. Elements such as SVG, HTML or canvas can be easily created and are editable at any time. Mike Dewar stated that “a huge benefit of how D3 exposes the designer directly to the webpage is that the existing technology in the browser cab be leveraged without having to create a whole new plotting language.(Dewar 2011, p. 2)” Overall, D3 fills the large gap between the designer and technical barriers which traditional visualization toolkits used to have. Instead D3 is focusing on providing functions in order to fulfill the designer’s expectation. The designer does not longer need to be a debugger and programmer to produce an interactive visualization; instead he can focus more on the creation of graphical elements and positioning them on a plane as well as finding ways to best gives life and interactivity to them.

Adobe Edge

Another technology which is used in this study is Adobe edge software. Adobe edge is a new multimedia authoring tool which is based on the foundation of JavaScript, HTML5, jQuery and CSS3. This software enables users to either create or import graphical elements such as SVG files and define different interactions based on possible functionalities.

4.2.3.2 Design of the visualization software architecture

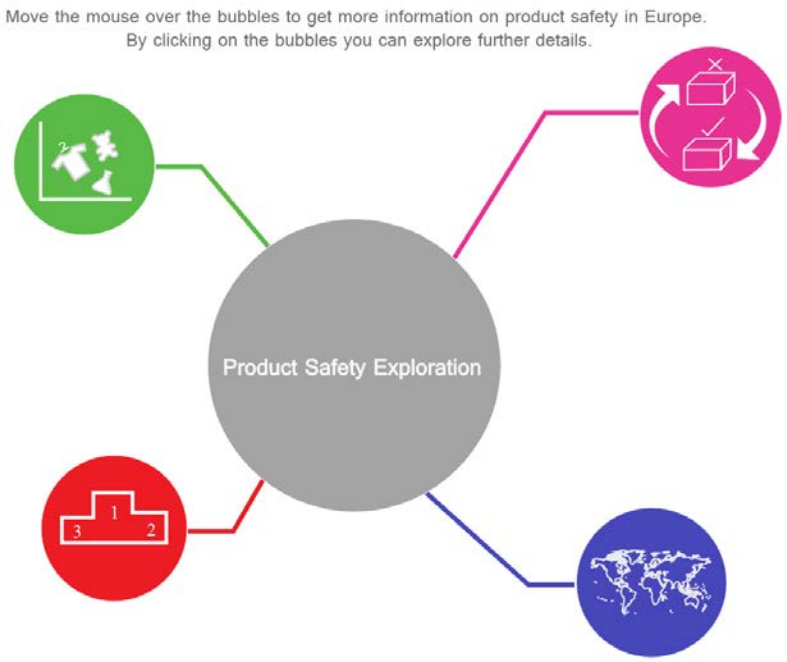
The overall structure of the application is inspired by a drill-down narrative structure. As discussed in a chapter 2.6 this narrative structure gives the user more freedom to navigate. And it guides the user to explore more when he or she felt that there is more need for detail.

The following sub-chapter is an analysis of source code for interface (creation of graphic elements) and interaction design. To provide an interactive visualization, the graphical element in this study provides the following interactions. The overview below does not cover entire functions in the application. I chose the most important functions related to the narrative structure.

Index

There are 4 major categories which are related to their own slide. According to the subject of the slide, each circle has its own particular icon with individual color code. The icons have been designed based on the story that extracted from the analysis of the data. See the section 4.2.2.2. The reason to design each circle both with individual color and icon is to increase the user attention as well as help the user to separate the different topic. Another reason of designing the icons was to avoid losing potential users which suffer from color blindness.

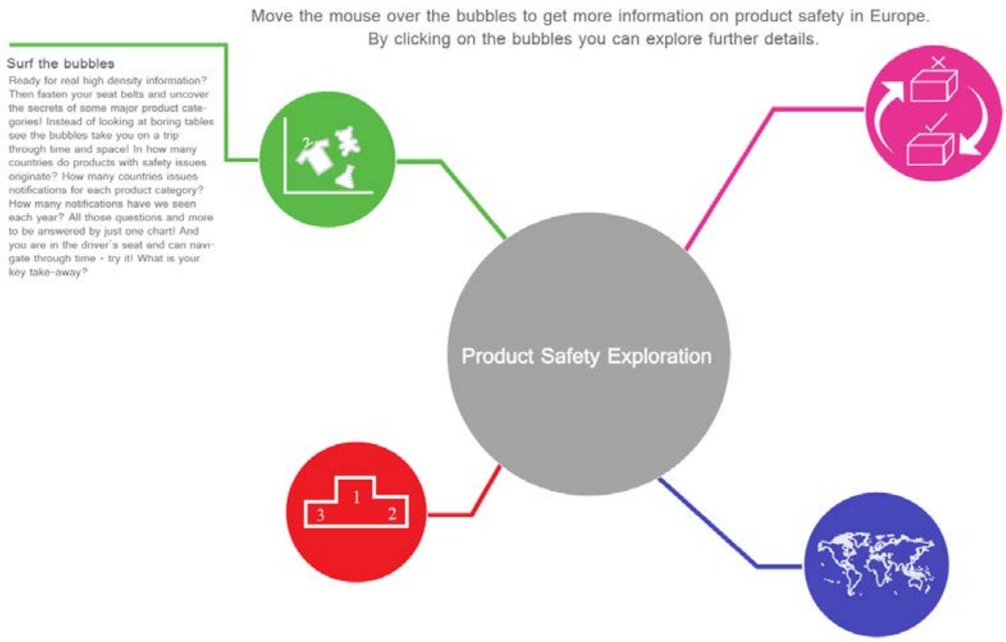
The application starts with an overview of four different stories. See the Figure 41.



Move the mouse over the bubbles to get more information on product safety in Europe.
By clicking on the bubbles you can explore further details.

Figure 41. The application index page interface

The initial page is featured with hover effect for more detail. Detail on demand technique enables the user to have an access to more detail when it is needed. See figure 42 below.



Move the mouse over the bubbles to get more information on product safety in Europe.
By clicking on the bubbles you can explore further details.

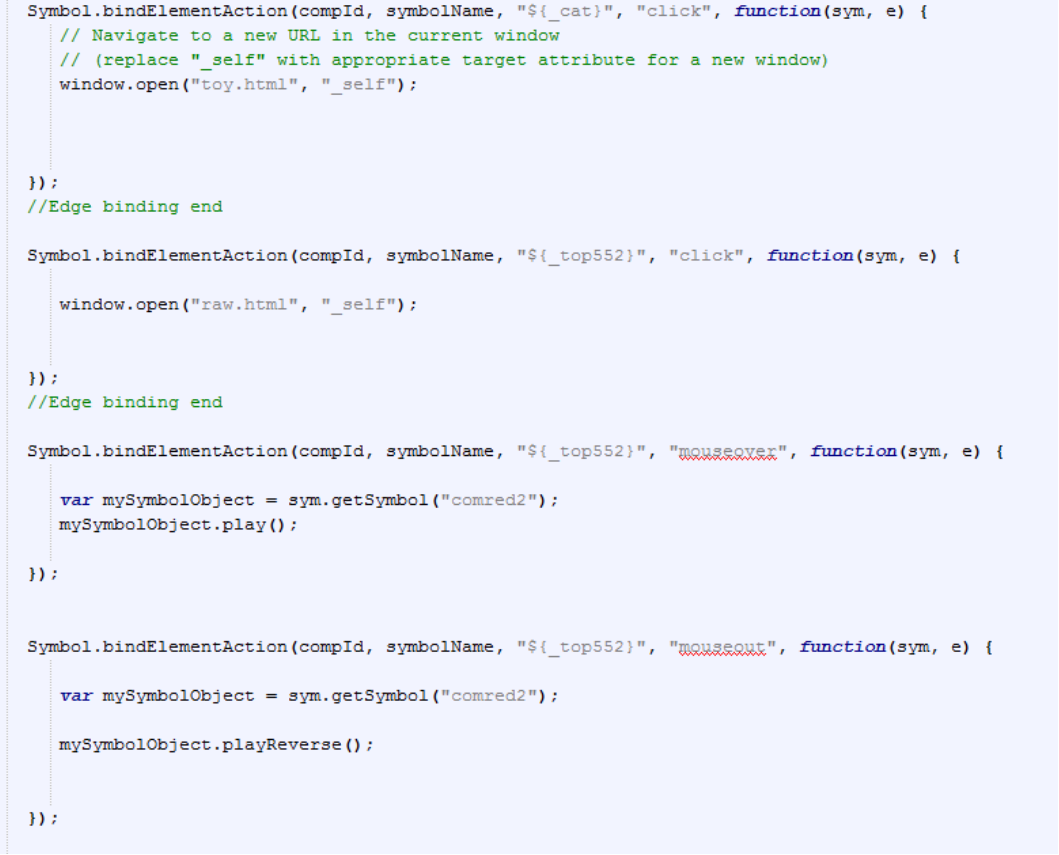
Surf the bubbles

Ready for real high density information? Then fasten your seat belts and uncover the secrets of some major product categories! Instead of looking at boring tables see the bubbles take you on a trip through time and space! In how many countries do products with safety issues originate? How many countries issues notifications for each product category? How many notifications have we seen each year? All those questions and more to be answered by just one chart! And you are in the driver´s seat and can navigate through time - try it! What is your key take-away?

Figure 42. The application mouse over interface

the

The development of the index page is done by Adobe Edge Animate software. To provide the interactions and animation, the simple java functions such as mouse over, mouse out and click are used. The figure 43 is an example of the java functions for click and mouse over and mouse out on one of the elements. The same code structure applies for the rest of elements on index page.



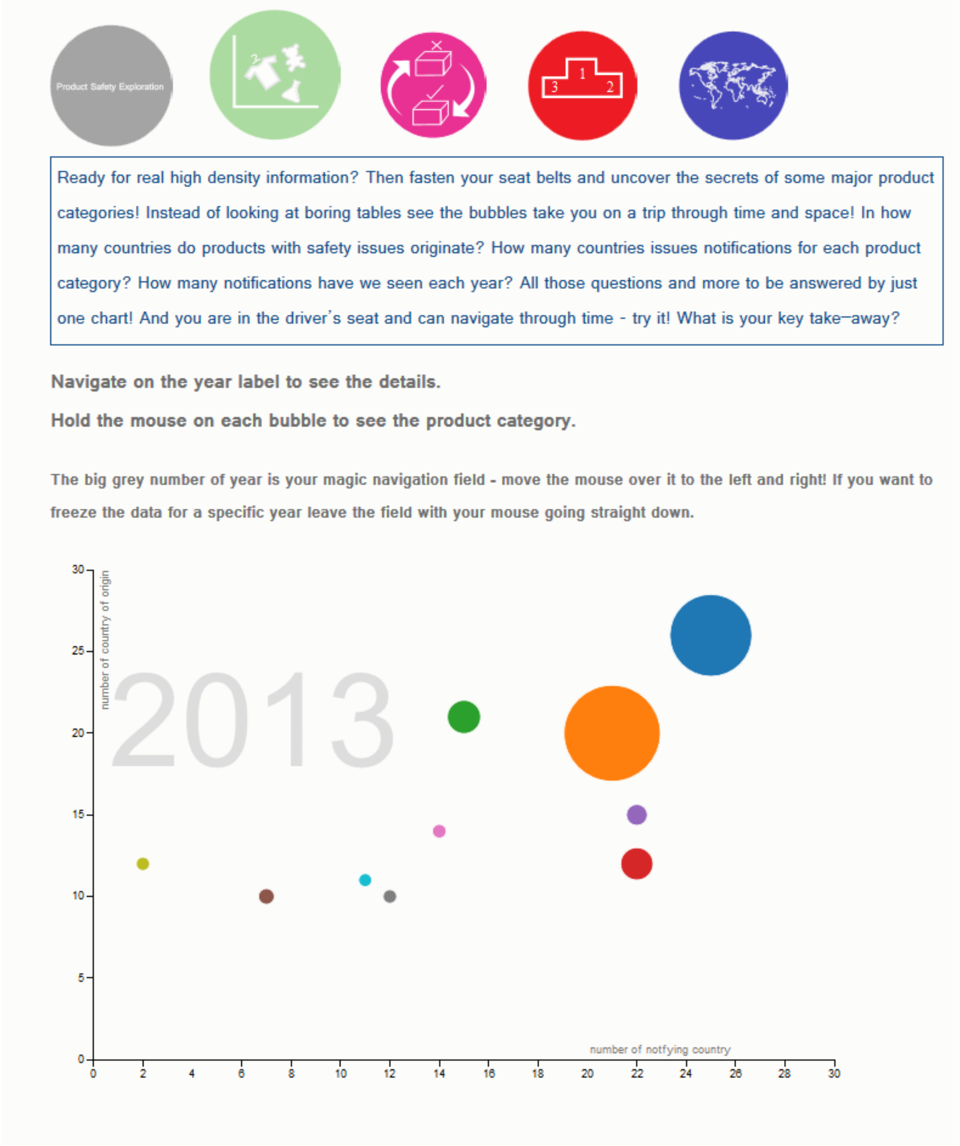
```
Symbol.bindElementAction(compId, symbolName, "${_cat}", "click", function(sym, e) {
   // Navigate to a new URL in the current window
   // (replace "_self" with appropriate target attribute for a new window)
   window.open("toy.html", "_self");



});
//Edge binding end

Symbol.bindElementAction(compId, symbolName, "${_top552}", "click", function(sym, e) {

   window.open("raw.html", "_self");


});
//Edge binding end

Symbol.bindElementAction(compId, symbolName, "${_top552}", "mouseover", function(sym, e) {

   var mySymbolObject = sym.getSymbol("comred2");
   mySymbolObject.play();

});


Symbol.bindElementAction(compId, symbolName, "${_top552}", "mouseout", function(sym, e) {

   var mySymbolObject = sym.getSymbol("comred2");

   mySymbolObject.playReverse();


});
```

Figure 43. Interaction source code example

The user has the opportunity to choose each bubble and get directed to the related topic.

What is the trend of number of notification in different product categories through the years?

The figure 44 presents the outlook of the slide. There is a navigation bar in upper part, stands for each slide. The navigation bar also includes one button (gray circle) stands for index. This directs the user to go back to the index page. The navigation is universal for other slides as well.



Ready for real high density information? Then fasten your seat belts and uncover the secrets of some major product categories! Instead of looking at boring tables see the bubbles take you on a trip through time and space! In how many countries do products with safety issues originate? How many countries issues notifications for each product category? How many notifications have we seen each year? All those questions and more to be answered by just one chart! And you are in the driver's seat and can navigate through time - try it! What is your key take–away?

Navigate on the year label to see the details.

Hold the mouse on each bubble to see the product category.

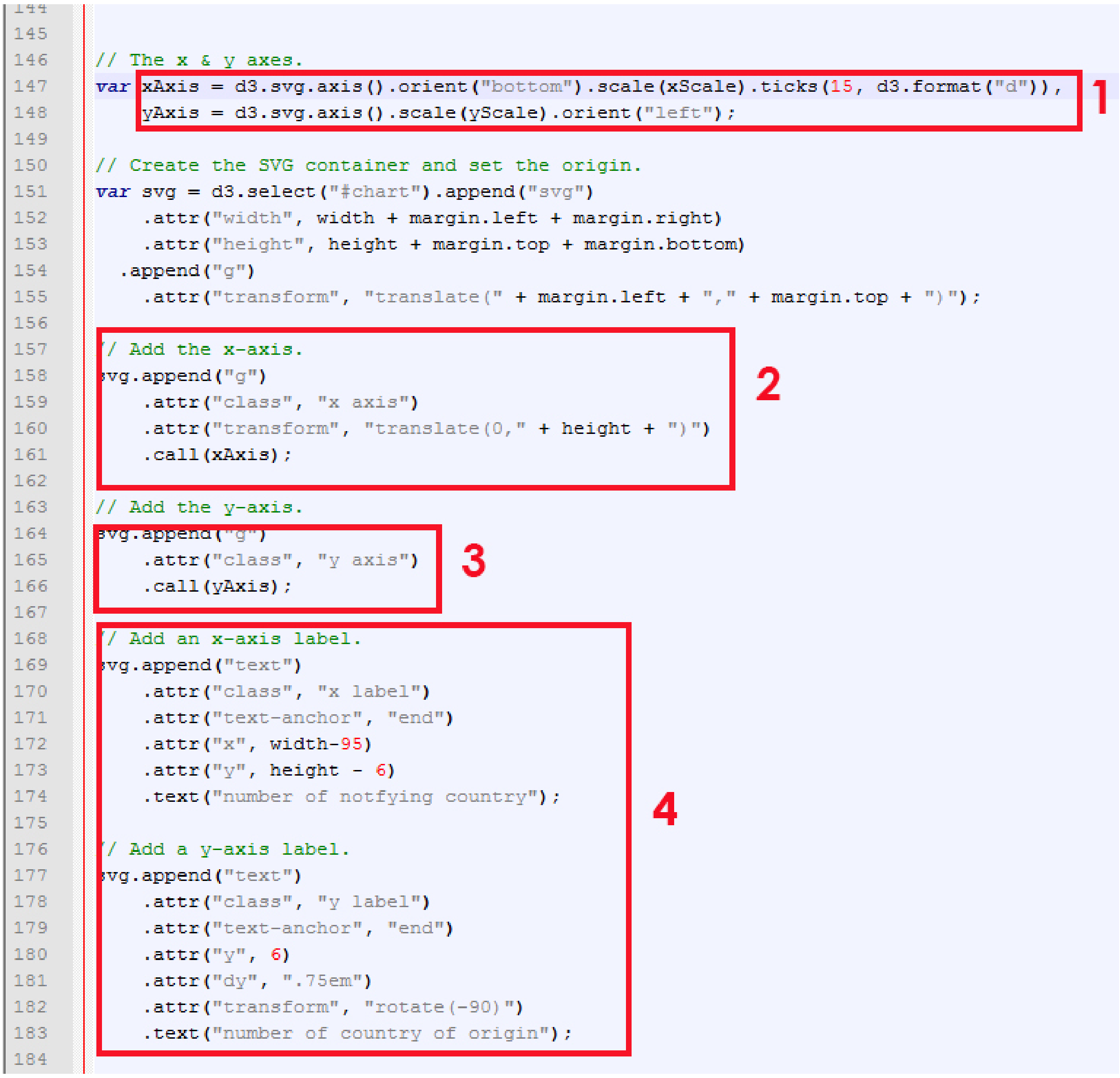
The big grey number of year is your magic navigation field - move the mouse over it to the left and right! If you want to freeze the data for a specific year leave the field with your mouse going straight down.

Figure 44.The trend of number of notification visualization overall interface

The overall interaction and interface design of this slide is inspired by the Gapminder software. See the section 2.6.3.1.

As explained, the goal of visualization in this slide is to present the number of countries which notified unsafe products in relation to country of origin. It was important to divide different categories in order to have a clear vision of which products cause more danger as well as how many countries are more active in reporting different product categories; the *motion chart* also features an interactive year label. The size of each bubble is related to the number of records in that particular product category. There is an introductory text on top of the screen. More explanation the graphic elements in this slide consist of following:

The axes: the code below is the screen capture of the source code in order to create the axes. Figure45.



```
// The x & y axes.
var xAxis = d3.svg.axis().orient("bottom").scale(xScale).ticks(15, d3.format("d")),
    yAxis = d3.svg.axis().scale(yScale).orient("left");

// Create the SVG container and set the origin.
var svg = d3.select("#chart").append("svg")
    .attr("width", width + margin.left + margin.right)
    .attr("height", height + margin.top + margin.bottom)
  .append("g")
    .attr("transform", "translate(" + margin.left + "," + margin.top + ")");

/ Add the x-axis.
svg.append("g")
    .attr("class", "x axis")
    .attr("transform", "translate(0," + height + ")")
    .call(xAxis);

// Add the y-axis.
svg.append("g")
    .attr("class", "y axis")
    .call(yAxis);

/ Add an x-axis label.
svg.append("text")
    .attr("class", "x label")
    .attr("text-anchor", "end")
    .attr("x", width-95)
    .attr("y", height - 6)
    .text("number of notfying country");

/ Add a y-axis label.
svg.append("text")
    .attr("class", "y label")
    .attr("text-anchor", "end")
    .attr("y", 6)
    .attr("dy", ".75em")
    .attr("transform", "rotate(-90)")
    .text("number of country of origin");
```

Figure 45. The trend of number of notification source code example

1. The first code Creates the line of the axes and positioning and defining the number of ticks

2. The code attr("transform", "translate(0," + height + ")") transfers x – axes into the bottom of the screen. It is specified where x and y has to be placed on the plane. And with giving the “0” and “height+” we position the x axes at bottom of the height of the entire image.

3. For y axes, a class has been created and drawn the line by using append function.

4. Style the labels on the axes.

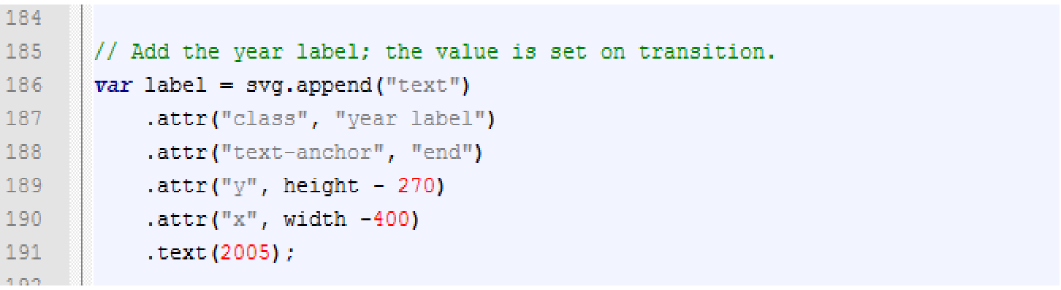
5. To create each individual bubble and append the related number of records. The code below is positioning each individual circle in relation to x and y axes.

```
function position(dot) {
   dot .attr("cx", function(d) { return xScale(x(d)); })
      .attr("cy", function(d) { return yScale(y(d)); })
      .attr("r", function(d) { return radiusScale(radius(d)); });
 }
```

The function “position” is a placement function which always draws the smallest dots on top of the bigger dots.

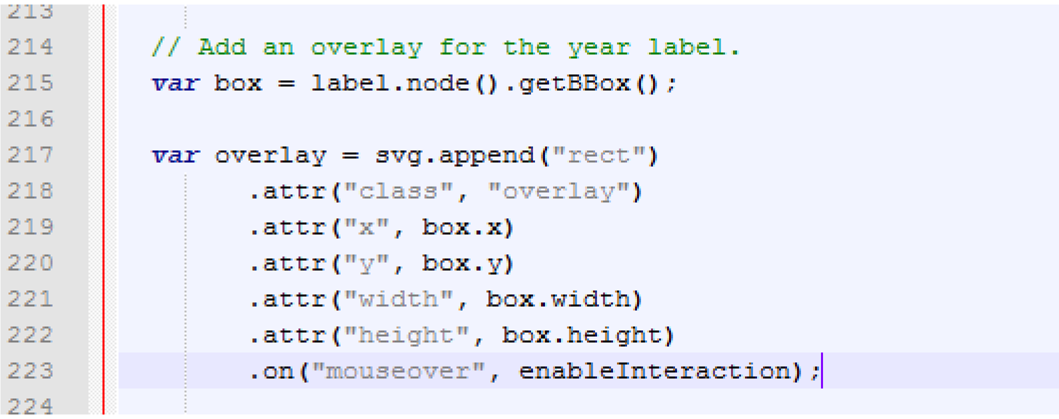
```
function order(a, b) {
   return radius(b) - radius(a);
 }
```

Possible interactions in this slide are as following: Year label slider: this slide enables the user to manipulate the data by sliding over and see the process through different years. The source code below is the function related to slider. See figure 46 and 47.



```

// Add the year label; the value is set on transition.
var label = svg.append("text")
    .attr("class", "year label")
    .attr("text-anchor", "end")
    .attr("y", height - 270)
    .attr("x", width -400)
    .text(2005);
```

Figure 46. slider function 1



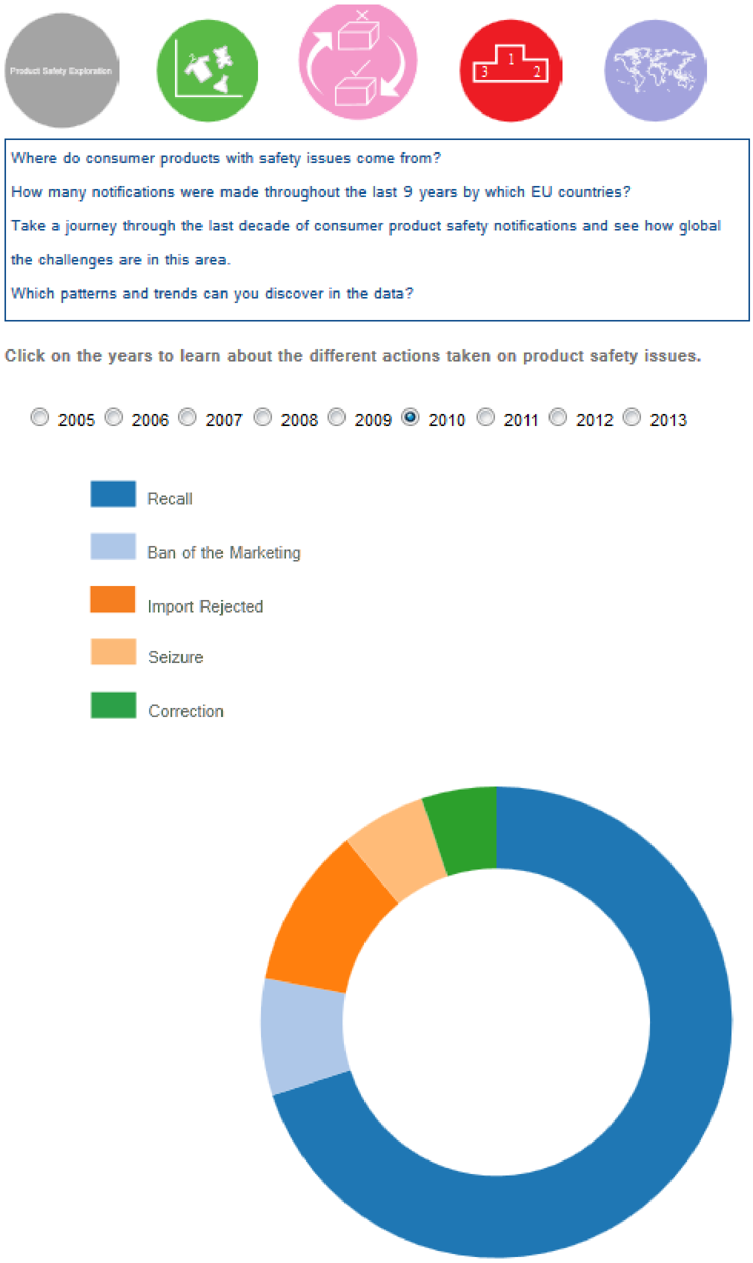
```
// Add an overlay for the year label.
var box = label.node().getBBox();

var overlay = svg.append("rect")
      .attr("class", "overlay")
      .attr("x", box.x)
      .attr("y", box.y)
      .attr("width", box.width)
      .attr("height", box.height)
      .on("mouseover", enableInteraction);
```

Figure 47. Slider function 2

Mouse hover: another interaction in this illustration is the hover effect. By holding the mouse on each bubble the user can observe the related name of the category.

What kinds of actions are taken after notifying products with safety issues?

Another slide presents the data related to the reactions taken. It is an illustration of an animated pie chart. In this visualization there is a guide-line with different colors which stands for different reactions taken towards the unsafe products. Figure 48.



Where do consumer products with safety issues come from?

How many notifications were made throughout the last 9 years by which EU countries?

Take a journey through the last decade of consumer product safety notifications and see how global the challenges are in this area.

Which patterns and trends can you discover in the data?

Click on the years to learn about the different actions taken on product safety issues.

○ 2005 ○ 2006 ○ 2007 ○ 2008 ○ 2009 ● 2010 ○ 2011 ○ 2012 ○ 2013

Figure 48. Action taken visualization overall interface

Similar to the previous slide, it features with navigation bar in the top of the screen as well as a short description of the content and interaction guideline. The graphical element source codes are as following:

The code below is the draw function of the pie chart.

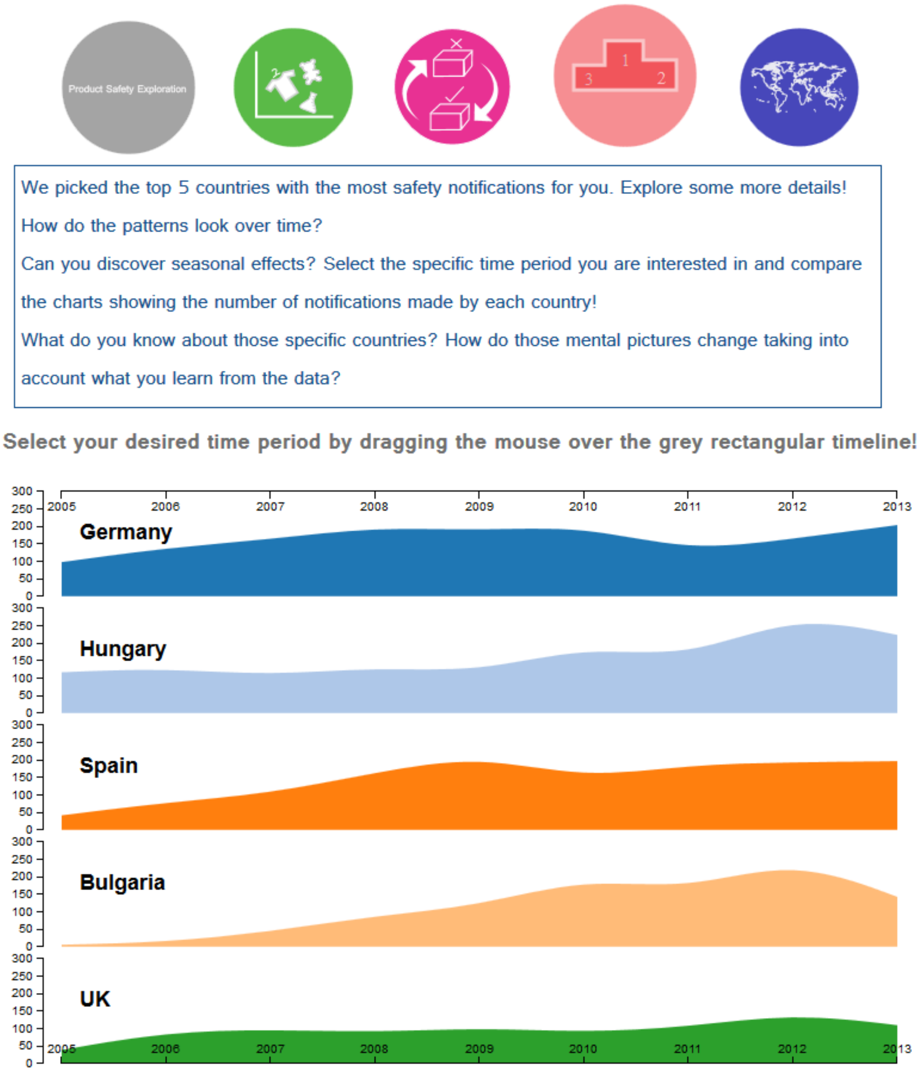
```
d3.tsv("data.tsv", type, function(error, data) {

var path = svg.datum(data).selectAll("path")
.data(pie)
.enter().append("path")
.attr("fill", function(d, i) { return color(i); })
.attr("d", arc)
.each(function(d) { this._current = d; }); // store the initial angles
 d3.selectAll("input")
.on("change", change);
```

The possible interaction in this slide is the ability to choose separated year by choosing the radio button. An interesting point in this design which is worth mentioning is the smooth transaction created by animation. This effect aids the user to connect to datasets into individual scene by smooth movements.

```
function change() {
var value = this.value;
clearTimeout(timeout);
pie.value(function(d) { return d[value]; }); // change the value function
path = path.data(pie); // compute the new angles
path.transition().duration(750).attrTween("d", arcTween); // redraw the arcs
}
});
```

Who are the most active countries in terms of notifying dangerous and harmful products?

The next slide is a graphical presentation of the top five countries that had the most active part in the matter of product safety. This slide follows the same structure of the pervious slide in terms of having the navigation bar, summary and interaction guidance. It has individual Area charts for each country. See Figure 49.



We picked the top 5 countries with the most safety notifications for you. Explore some more details!

How do the patterns look over time?

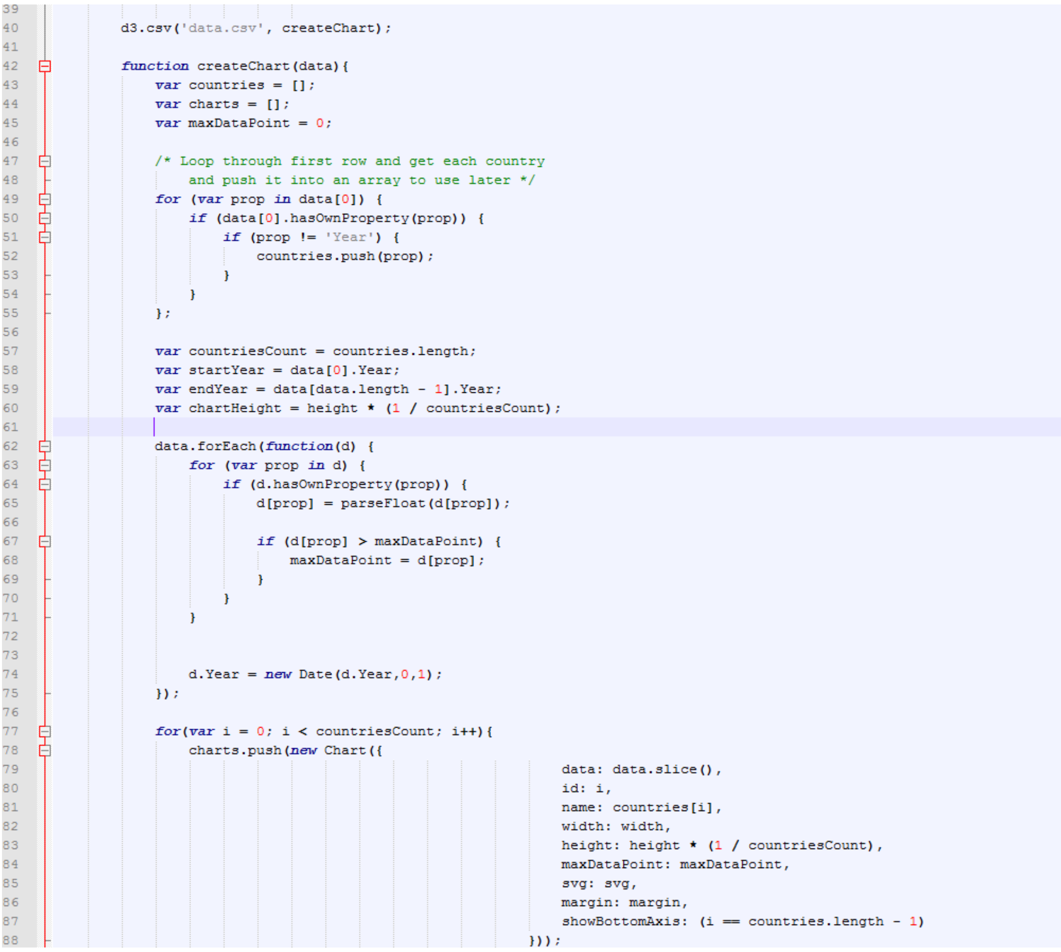
Can you discover seasonal effects? Select the specific time period you are interested in and compare the charts showing the number of notifications made by each country!

What do you know about those specific countries? How do those mental pictures change taking into account what you learn from the data?

Select your desired time period by dragging the mouse over the grey rectangular timeline!

Figure 49. Top five countries visualization overall interface

The code shown in Figure 50; it is a source code for creating the five chart areas.



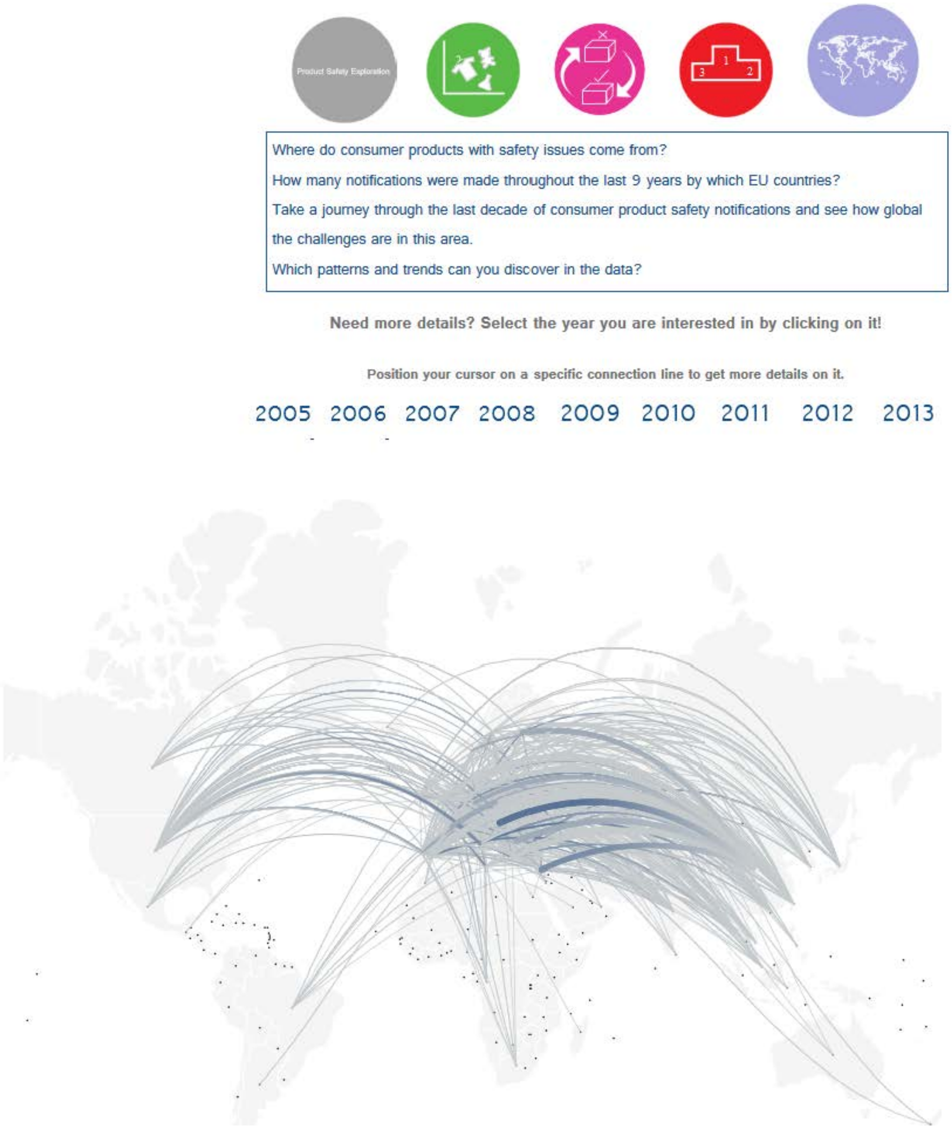
```

        d3.csv('data.csv', createChart);

        function createChart(data){
            var countries = [];
            var charts = [];
            var maxDataPoint = 0;

            /* Loop through first row and get each country
                and push it into an array to use later */
            for (var prop in data[0]) {
                if (data[0].hasOwnProperty(prop)) {
                    if (prop != 'Year') {
                        countries.push(prop);
                    }
                }
            };

            var countriesCount = countries.length;
            var startYear = data[0].Year;
            var endYear = data[data.length - 1].Year;
            var chartHeight = height * (1 / countriesCount);

            data.forEach(function(d) {
                for (var prop in d) {
                    if (d.hasOwnProperty(prop)) {
                        d[prop] = parseFloat(d[prop]);

                        if (d[prop] > maxDataPoint) {
                            maxDataPoint = d[prop];
                        }
                    }
                }


                d.Year = new Date(d.Year,0,1);
            });

            for(var i = 0; i < countriesCount; i++){
                charts.push(new Chart({
                                                            data: data.slice(),
                                                            id: i,
                                                            name: countries[i],
                                                            width: width,
                                                            height: height * (1 / countriesCount),
                                                            maxDataPoint: maxDataPoint,
                                                            svg: svg,
                                                            margin: margin,
                                                            showBottomAxis: (i == countries.length - 1)
                                                        }));
```

Figure 50. Top five countries source code for elements creation

Interactivity: the user is able to drag a certain period of time instead of having an overview. This zooming effect function provides more details information in case it is needed and for further investigation.

Who produces the unsafe product and who notifies it?

One of the outstanding data this study is focusing on, is the countries which produce the unsafe product and countries which have an active role in notifying them. Representation of this data on a map is a very helpful since the geographical data visualization on a map is a very effective way (Cartwright et al. 2004). Since the density of number of records in past seven years has been very high, on the first slide there is an overall overview of relations between countries is shown. See figure 51.



Where do consumer products with safety issues come from?
How many notifications were made throughout the last 9 years by which EU countries?
Take a journey through the last decade of consumer product safety notifications and see how global the challenges are in this area.
Which patterns and trends can you discover in the data?

Need more details? Select the year you are interested in by clicking on it!

Position your cursor on a specific connection line to get more details on it.

Figure 51. World map visualization overall interface

In case the user wants to extract more data, he has the opportunity to click on individual year links and gets directed to that specific year with more detail. This technique is a zooming technique, but it can also count as a detail on demand approach. The data in this illustration has been divided into separated years. Interaction in this slides include following:

Mouse hover: it is possible that the user hover on each arc and according to related arc the message box pops up and presents more information including: where the products come from, where tis their destination and the number of records in the related year. See figure52.

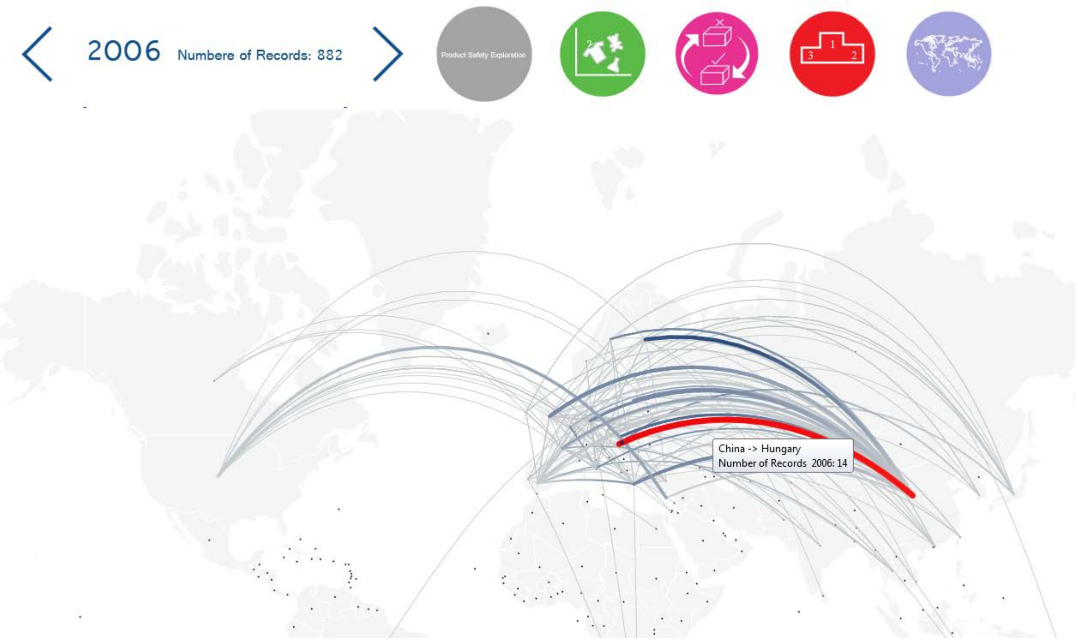


Figure 52. world map possible mouse over interaction

Step Backward and step forward general overview (home): this slide featured with navigation links for ordinal exploration. And direct link to general overview.

4.2.3.3 Design Limitation

Implementation tools limitation

Although using D3 and Adobe edge has many strong points for creation of interactive data visualization, I found number of difficulties during the design process.

One of the limitations which I was facing during implementation phase was difficulties for entering the data into the data sheet. Because D3 usually works with JSON or CSV files to load the data into a function, process of creation of such files was very time consuming and needs a lot of attention. Although there are plenty libraries provided by D3 support organization which helps to not entering the data manually, to create a new illustration with particular dataset the preparation of the datasheet was time consuming.

Another disadvantage that needs to be mentioned is the speed of loading data especially when it comes to large number of entries. To avoid this problem I have copies all the existing libraries on local host which prevent wasting time on loading library in real time. There was also performance limitation for drawing plenty number of SVG objects on screen.

4.3 Scientific question

To balance author-driven and reader driven stories there are three common schemas: Martini glass (prioritizes the author driven approach), interactive slideshow (the dialogue between the two approaches) and drill-down story (prioritize the reader-driven approach) (Segel & Heer 2010b, p.1146). In this study the most important motivation is to raise awareness to users (product consumers). In other words, the message that the author wants to convey primarily is to compare with other possible tasks. Therefore i narrowed down the study to investigate different approaches with the Martini Glass schema. I have chosen the martini glass approach because it enables the designer to convey the message before it allows the user the opportunity to investigate on the application with absolute freedom.

The goal of this master thesis is to develop interactive data visualization application (web-based) and evaluate how applying different narrative structure influences the learning effect and analyze the users visual perceptual behavior focused on the level preserving the author’s massage. Therefore the scientific question can define as following: *which approach in martini glass narrative structure in interactive data visualization is more effective?*

4.4 The sample group analysis

The samples for this study were randomly collected from solar service company called ADLER Solar. ADLER Solar is the main company which is the base of the employees who works for consider it GmbH. The reason I chose my test subjects from Adler was the ease of finding potential users as well as their interest in the topic. Another reason was the age distribution and Diversity in different experience and specialties. All of the samples are employees between 20 and 55, including of 13 males and 9 females were participating in the test.

Since this application is built for users with a certain Level of Proficiency in using of computers, the samples without having basic computer skills were discarded from the study after the test.

One day before the original test, I ran a pretest with four subjects. According to the pretest the design of the questions has been revised. The pretest has shown that users are getting confused without guideline to interaction. Therefore I have also added more details in the interaction instruction of each slide.

As mentioned, among 22 subjects that were tested two of them had to be discarded from the study. One subject could not continue the test because he had lack of IT skills and had difficulties to interact with the application. He stated that “I cannot interact with the application” and another subject finished the test but at the end he stated that he answered to the questions based on his assumption: “I was just guessing. I don’t understand the question.” The reason was lack of fluency in English language.

4.5 Hypothesis & Preparation of user test

The hypothesis of this study focuses on the different ways of presenting the narrow part (author-driven) of interactive data visualization application. To see the difference between different approaches the hypothesis and null hypothesis are as following:

a) Hypothesis: The participants, who observe the representation of the data as written text in a form of questions, have a higher level of understanding compared to the users who observe the data as a non-interactive linear animation.

b) Null hypothesis: There is no difference in level of efficiency and learning between those using written text as questions and people watching linier animation.

As decided to have the user study only focusing on the martini glass approach and based on the principles of the Martini glass structure the application has been divided into two phases:

- The tenuous phase or single path (author-driven).
- The wide phase or freely explore path (reader-driven)

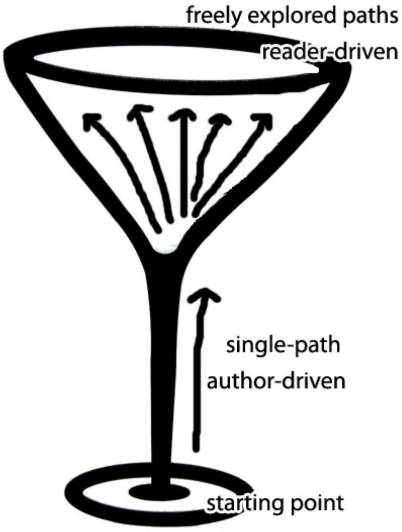


Figure 53. Martini glass guide in the process of media creation

The wide phase: The design of the main application and the details in the existing interface and interaction were explained in detail in the last chapter.

The tenuous phase: Although there is variety of choices to create the narrow author-driven phase of the martini glass approach such as questions, observations, or written articles, animation and so on, I have decided to focus on only two techniques to design the tenuous part of the glass.

- Video
- Question

Question design:

The given information is divided into two levels: intermediate and advanced. It starts with easy questions and it moves on to more details. The *appendix I* shows the first questionnaire given to the subjects. Each slide has been categorized by its individual icon.

Video design:

In order to create a video I have screen captured each individual sub-story and ordered them in a timeline. Each sub-story was introduced by its topic as a text. I highlighted the intended message by creating animated text.

4.6 Experiment design

Due to the fact that, in this study there was very strong potential of learning effect, I have chosen the between- subject design. I divided the participants into two groups: The participants who had to watch the video. We will call them *video group* from now on. And the group which received the questions, we will call this group the *question group.* The collection of subject was completely random. It means on a test day I was testing one subject with video material and the next with questionnaire and continuously with the same order. The test ran in an individual room. The room was a quiet room to avoid distraction by external noise. Each subject was tested separately and they were asked to sit in front of the computer. Two monitor were used in this test. It was also one mouse for navigation. A key board was available for more navigation. No paper and pen was provided for user to take notes during the test.

The test started with a short summary of the tasks. A short instruction were given to the subject Then both group had to observe either the animation or reading the questions (author-driven). And they had to switch to application and start navigating and interact with the application (reader-driven). The subjects had 5 minutes to surf and gain information from the web based application. In this part of the test no guidance or hint was given to the users. The test designed somewhat to give the user absolute freedom in order to make his own conclusion.

Video group: the control group was asked to watch a video. The duration of the animation was one minute and 23 seconds and it was presenting the short illustration of the application with highlighted words. After they finished the video the participants could switch to the application and freely interact with the application for 5 minutes. It was explained to the participants that the video had the same content as the main application. After the test the participant was asked to answer the questionnaire. The video was featured with background music.

Question group: a list consisting of 8 questions was given to the participants of this group. They had 40 seconds to read the questions. The reason for choosing 40 second was to give 5 second for each question and not give too much time for analyzing the questionnaire .the goal was to give the subject an opportunity to scan the keywords and not preserve the information. After reviewing the questions the participants could switch to the application and freely interact with it for 5 minutes. After exploring the application, a multiple choice questionnaire was given to the test subjects. The questions consist of the same questions that the subjects had to answer after exploring the main application.

Before starting the test, a short summary of the tasks was given to all the participants. The subjects were not informed that they will receive a questionnaire after the test. The time for answering the questionnaire after the test was unlimited. I also stressed that if they don’t remember any answer they should not guess the answer. This approach can help to reduce misinterpretation and it increases the level of accuracy in the test. The questionnaire was designed based on the goal of this study. I provided multiple questions as well as the questions that the subject had to rate the elements. *Appendix II* shows the final questionnaire that all the subjects had to fill in.

5. Result and analysis of the user study

According to the importance of each answer and the importance of analyzing the level of understanding, each individual right answer was counted as one point. In total the subjects had to answer 11 questions.

The table 2 shows the results of the test after analyzing the data from both groups.

Question group			
Subject #	Correct	incorrect	No answer
1	8	3	
2	7		4
3	9	2	
4	10	1	
5	8	3	
6	7	4	
7	7	4	
8	10	1	
9	9	2	
10	8	2	1
SUM	83	22	5

Video group			
Subject #	correct	incorrect	No answer
1	4	7	
2	6	5	
3	8	2	1
4	4	4	3
5	8	3	
6	6	3	2
7	8	2	1
8	6	5	
9	6	5	
10	3	6	2
SUM	59	42	9

Table 2. The score of each individual subject

The question group performed much better than the video group, with 83% correct answer, 22% incorrect answers and 5% no answers compared to the video group with 59% correct answers and 42% incorrect answers and 9% no answers.

To calculate the variation of existing data the following steps had to be taken - see table 3 for details of the calculation. Based on the results and the hypothesis of this study the result of the T-test is as following (Chambliss & k schutt n.d., pp.155–165):

To calculate that:

$$t = \frac{M_x - M_y}{\sqrt{\frac{S_x^{\ 2}}{n_x} + \frac{S_y^{\ 2}}{n_y}}}$$

In order to calculate the $S^{2:}$

$$S^2 = \frac{\sum (x - M)^2}{n - 1}$$

In order to calculate $M_{y \text{ and}}$ $M_{x\text{ :}}$

$$\overline{X} = \frac{\sum X}{N}$$

Question group correct answers	Video group correct answers
Mean = 8.3 Standard Deviation = 1.1595 Standard Error = 0.36667	Mean = 5.9 Standard Deviation = 1.79196 Standard Error = 0.56667

Table 3. The result of the calculation

After the calculation the T value equals to T=3.555827 and converting the value to *p* we have the result of *p*=0.001129. The p value in this test is 0.001≤ 0.01 which means there is a significant relationship between the narrow part of the glass and a very strong presumption against the neutral hypothesis.

Therefore the result of the study shows that there is a significant difference between the two groups – one of them having had the questionnaire before starting to interact with the application and the other one which had to watch the animation in the initial phase of the test. Therefore the data provided simply shows that the null hypothesis is false. And this study's hypothesis can be concluded by stating as statistically proven that the participants who read the questionnaire before free exploration have a higher learning effect than the users who watch the video.

Clearly the study answers to the research question that applying different narrative structures (animation vs. questionnaire) can obviously influence the learning effect in interactive data visualization applications.

Another interesting observation which could me made was the interaction behavior of the subjects. One of the most interesting behaviors observed was, that when the user wanted to start the navigation on the index page and chose one of the elements (bubbles) to see more details and information. 17 subjects out of 20 first choice was the green bubble (product category). It was placed at the top-left of the screen. That supports the theory of the Eyetrack III research of The Poynter Institute, the Estlow Center for Journalism & New Media, and Eyetools. According to the research done by Steve Outing and Laura Ruel In *What We Saw through Their Eyes*, the results show that The upper left quarter of the screen gets the most attention (Outing & Ruel 2004). See figure 54.

Figure 54. The most common regions which attract attention on the screen

Comments

After finishing the test I asked the subjects to comment on the application and what kind of improvements can be done in order to make the application more user-friendly. The observation and participant’s comments demonstrate that participants in the video group had difficulties to extract information from the video. ”The video was too fast.” And one subject suggested that the symbols (circles) of each slide could be shown in the video. ”If you could show the icons in the beginning of each part, then I could say ok this part is related to this topic”

One of the comments that were received many times was about the mouse over effect. “The mouse over interaction is too slow” or “the label appears too late.” Apparently the duration between the mouse over and label appearance was longer than it should be. This can be improved in the next version of application.

6. Conclusion

With the rapid growth of information in this century there is a high demand for finding new ways to manage big data and approaches to extract the most important message from the existing information. In other words the effort for the data analysis (from data collection to data presentation) has to be illustrated in some different ways than static analog versions which the public can benefit from them without spending the same amount of time and energy. Data mapping is one of the earliest techniques of presenting data in smaller proportion on paper or on a display. It goes back to 16th century and it is still a powerful technique to present data. Beside mapped images, storytelling is also an old tradition since ancient time. Human mankind are used to pass on information and gain knowledge to the next generation via storytelling. Thanks to advanced calculators (computers) and automation, receiving and saving data becomes more and more convenient. Every day the capacity of data storage becomes larger. But for the same (automation) reason data which is produced these days is significantly growing. But this is not an obstacle which cannot be solved, when the three fields of data visualization, storytelling and computer science meet. Computer scientists with the help of technology and new objective tools found ways to visualize big amounts of data on tiny screens. Technology allows the scientist to create objects with plenty of possible interactions based on big data. With one click big data can be spread out to the web. it is the spectator’s choice to decide and start interacting and gain information. But the computer scientists were not the only group that was amazingly impressed by the capabilities of human-computer interaction to present the data. Journalists, storytellers, authors and many more were also amazed by the new technology. This was the moment that data visualizers tried to create stories using interactions. Narrative data visualization, this new approach with very ancient roots in human culture gathers scientist, designers and authors together to walk through one path for using big data which is most of the time nothing but big tables with an enormous number of rows and columns.

The goal of this study was to investigate the deeper levels of narrative structures and find out the best practices when the author tends to send a specific message to the user as well as give the user the opportunity to interact with the application freely and make personal conclusions. This study focuses on the martini glass approach among the rest of other possible presentation modes. For the author-driven part of the martini glass, a presentation with questionnaire and video have been chosen (Segel & Heer 2010a). To evaluate and compare the two chosen approaches I have designed a user study. The result of the study showed that the group, which started the author-driven part of the application, significantly performed better and had a better understanding compared to the group which had the video as the introduction part. Therefore to design an application which is based on martini glass's narrative structure the introductory (thin part of the glass) part using questions is more effective than using videos.

7. Future Work

Due to the limitations of this study in timing and volume, not all the problems and issues, faced during the implementation phase and the user study design have been resolved. However it is worth mentioning in this chapter and to propose it for future work.

As the outcome of the research experiment showed using questions in martini glass approach improves the level of learning as well as performance. This study only limited the research to two approaches, the questions and video. However there are other different options to create the experiment with different introductory methods such as messaging and explanation or creating interactive animation.

Another issue that I was facing during the study was the language problem. Although the majority of participants had a high level of skills in English, in some cases the user could not understand some expressions related to the topic and it had to be translated or explained by the observer. This caused distraction and it was time consuming. This study has the potential to provide multi language options to avoid and solve the issue.

The current animation which was used for the user study was featured with background music. Other possible improvements could be featuring the animation with to a voice which explains the scenes and motion graphics and their relations according to the topic. This approach might be more helpful to transfer the message because it takes the advantage of auditory sensory as well as visual perception. I did not use that approach because this study’s goal was to measure the visual perception and did not to interfere with other sensory effects. Another suggestion for video is to redesign the elements more based on scientifically effective approaches. Redesigning the element in animation might improve the performance of the subjects.

The current version of the application is providing possible interaction instructions in the form of a box as a text. The next version of the application can provide different approaches. For instance it can be a ready animation showing the movements on a side of the screen. This avoids the screen to be loaded with too many text fields.

It was mentioned in the design limitation section that D3 technology is not very universal and flexible when it comes to the usage of different browsers. The next version of the current application must take a closer look on the issues and problems and resolve the problem to avoid losing potential spectators.

Furthermore the application has the capacity to improve in terms of interactions. More options for filtering and creation of smooth transaction between two slides can improve the flow of the story.

8. References

Berlin, B, and P Kay. *Basic Color Terms: Their Universality and Evolution*. University of California Press, 1969. http://books.google.de/books?id=xatoCJbhifkC.

Bertin, Jacques. *Semiology of Graphics*. *Physical Review Letters*. Vol. 94, 1983. doi:10.1103/PhysRevLett.208902.

Beynon-Davies, P. "Dances with Bees: Exploring the Relevance of the Study of Animal Communication to Informatics." *International Journal of Information Management* 30 (2010): 185–98.

Billmeyer, Fred W. "Basic Color Terms, Their Universality and Evolution, by Brent Berlin and Paul Kay, University of California, Berkeley, 1969; Reprinted in Paperback, 1991, 209 Pp. 12.95." *Color Research & Application* 17, no. 5 (1992): 368. doi:10.1002/col.5080170514.

Börner, Katy, William R Hazlewood, Sy-miaw Lin, Jordan Avenue, and Main Library. "Visualizing the Spatial and Temporal Distribution of User Interaction Data Collected in Three-Dimensional Virtual Worlds," 2001.

"Business Intelligence and Analytics | Tableau Software." Accessed May 27, 2014. http://www.tableausoftware.com/.

Card, Stuart k, Jock Mackinlay, and Ben Shneiderman. *Reading in Information Visualization Using Vision to Think*. 1st ed., 1999.

Cartwright, W., S. Miller, and C. Pettit. "Geographical Visualization: Past, Present and Future Development." *Journal of Spatial Science*, 2004. doi:10.1080/14498596.2004.9635003.

Chambliss, Daniel F, and Russel k schutt. *Making Sense of the Social World Methods of Investigation*. 4th ed., n.d.

"D3.js - Data-Driven Documents." Accessed May 28, 2014. http://d3js.org/.

Daniel, By. "C OLOR B LIND," n.d.

"Deuteranopia – Red-Green Color Blindness | Colblindor." Accessed May 13, 2014. http://www.color-blindness.com/deuteranopia-red-green-color-blindness/.

Dewar, mike. *Getting Started with D3*, 2011.

EUROPA - Consumer Affairs - Unsafe products. "EUROPA - Consumer Affairs - Unsafe Products." *RAPEX*. Accessed May 22, 2014. http://ec.europa.eu/consumers/safety/rapex/alerts/main/index.cfm?event=main.listNotifications&CFID=7054516&CFTOKEN=27133894&jsessionid=089cf37d7b28fc6588f8950207d146327e7b.

"EUROPA - Consumer Affairs - Unsafe Products." Accessed May 26, 2014. http://ec.europa.eu/consumers/safety/rapex/alerts/main/index.cfm?event=main.listNotifications&CFID=7118036&CFTOKEN=18926849&jsessionid=089c29afc611d2e6f13f1e587861676d6432.

Few, Stephen. *Data Visualization for Human Perception*, 2010.

———. *Now You See It Sample Visualization Techniques for Quantitative Analysis*. jonathan G. koomey, 2009.

———. *The Encyclopedia of Human-Computer Interaction, 2nd Ed*. Edited by Mads Soegaard and Rikke Friis Dam. The Interaction Design Foundation, 2013. http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html.

Franklin, Anna, Michael Pilling, and Ian Davies. "The Nature of Infant Color Categorization: Evidence from Eye Movements on a Target Detection Task." *Journal of Experimental Child Psychology* 91, no. 3 (July 2005): 227–48. doi:10.1016/j.jecp.2005.03.003.

Friendly, Michael. "Milestones in the History of Data Visualization : A Case Study in Statistical Historiography," 2006.

———. "Milestones in the History of Data Visualization: A Case Study in Statistical Historiography." In *Studies in Classification, Data Analysis, and Knowledge Organization*, 34–52, 2005. doi:10.1007/3-540-28084-7-4.

Friendly, Michael, and Ernest Kwan. *E Ect Ordering for Data Displays*. Vol. 43, 2003.

Gershon, Nahum, and Ward Page. "What Storytelling Can Do for Information Visualization." *Communicatoin of the ACM* 44, no. 8 (2001): 31–37.

Gielis, Johan. "The Geometry of Universal Natural Shapes." *Biological Shape Analysis - Proceedings of the 1st International Symposium*, 2011, 275–99. doi:10.1142/9789814355247_0016.

Guardian US interactive team. "Violence and Guns in Best-Selling Video Games – Interactive | World News | Theguardian.com." Accessed May 21, 2014. http://www.theguardian.com/world/interactive/2013/apr/30/violence-guns-best-selling-video-games.

Healey, Christopher G. "Building a Perceptual Visualisation Architecture," 1998, 1–20.

Heer, Jeffrey, and Ben Shneiderman. "Interactive Dynamics for Visual Analysis A Taxonomy of Tools That Support the Fluent and Flexible Use of Visualizations." *Acmqueue*, 2012, 1–26.

Hullman, Jessica, and Nicholas Diakopoulos. "Visualization Rhetoric: Framing Effects in Narrative Visualization." *IEEE Transactions on Visualization and*

Computer Graphics 17, no. 12 (December 2011): 2231–40. doi:10.1109/TVCG.2011.255.

Hyrkäs, Johanna. “Scientific Visualization Scientific Visualization in Art @ Science , Actually Concentrating More on Science @ Art,” 2000.

Iliinsky, Noah, and Julie Steele. *Designing Data Visualizations*, 2011.

Iwata, Tomoharu, Neil Houlsby, and Zoubin Ghahramani. “Active Learning for Interactive Visualization” 31 (2013): 342–50.

Jon Hazell. “A Brief History of Data Visualization | Dundas Data Visualization.” Accessed April 29, 2014. http://www.dundas.com/blog-post/a-brief-history-of-data-visualization/.

Keim, D.a. “Information Visualization and Visual Data Mining.” *IEEE Transactions on Visualization and Computer Graphics* 7, no. 1 (2002): 1–8. doi:10.1109/2945.981847.

Kenneth, Cukier. “Data, Data Everywhere | The Economist.” *The Economist*, 2010. http://www.economist.com/node/15557443.

Mazza, R. *Introduction to Information Visualization*. *October*. Vol. 1, 2009. doi:10.1007/978--84800-219-7.

Nations, United, Economic Commission, and F O R Europe. “No Title,” n.d.

“OECD Better Life Index.” Accessed May 20, 2014. http://www.oecdbetterlifeindex.org/.

Outing, By Steve, and Laura Ruel. “What We Saw Through Their Eyes Welcome to Eyetrack III,” 2004.

Riehmann, Patrick, Jens Opolka, and Bernd Froehlich. “The Product Explorer: Decision Making with Ease.” In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 423–32. AVI ’12. New York, NY, USA: ACM, 2012. doi:10.1145/2254556.2254638.

Ritz, Achim. “Rückruf-Aktionen Und Kostendruck Nehmen Zu - Service,” 2014. http://motor-exclusive.de/news.php?newsid=219533.

Schließmann, Christoph Philipp. “Bachelorstudium Recht Und Wirtschaft Bachelorarbeit Rückrufaktionen in Produktionsunternehmen . Eine Ganzheitliche Analyse Und Inhaltsverzeichnis,” no. 500 (2012).

Schlüter, Alexander. “Produktrückrufe Als Marketingproblem,” 1998, 58.

Segel, Edward, and Jeffrey Heer. “Narrative Visualization : Telling Stories with Data.” *IEEE Transactions on Visualization and Computer Graphics* 16, no. 6 (2010): 1139–48.

———. “Narrative Visualization: Telling Stories with Data.” *IEEE Transactions on Visualization and Computer Graphics* 16, no. 6 (2010): 1139–48. doi:10.1109/TVCG.2010.179.

Shneiderman, B. “The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations.” *Proceedings 1996 IEEE Symposium on Visual Languages*, 1996. doi:10.1109/VL.1996.545307.

“Steve Jobs Introduces the Macintosh (1984) - SpeakingFrog.” Accessed July 14, 2014. http://speakingfrog.com/?p=537.

Ten, Chapter. “Interacting with Visualizations,” 1998. doi:10.1016/B978-0-12-381464-7.00010-7.

“The Startup Universe.” Accessed May 15, 2014. http://visual.ly/vizbox/startup-universe/.

Thomas Klindt , Michael Popp, Matthias Rösler. *Rückrufmanagement, Ein Leitfaden Für Die Professionelle Abwicklung von Krisenfällen*, 2008.

Tominski, Christian, and James Abello. “Interactive Poster : Axes-Based Visualizations for Time Series Data,” n.d., 2–3.

Tripathi, K P. “A Study of Interactivity in Human Computer Interaction” 16, no. 6 (2011): 4–6.

Tufte, Edward R. *Beautiful Evidence*. 1St Editio. Graphics Press, 2006. http://www.worldcat.org/isbn/0961392177.

———. *The Visual Display of Quantitative Information*. Cheshire, CT, USA: Graphics Press, 1986.

———. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT, USA: Graphics Press, 1997.

Ware, Colin. *Information Visualization: Perception for Design*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2004.

Wilhelm, Adalbert. “Interactive Statistical Graphics : The Paradigm of Linked Views” 24, no. 04 (2005): 24016. doi:10.1016/S0169-7161(04)24016-0.

Wojtkowski, Wita, and W.Gregory wojtkowski. “Storytelling : Its Role in Information Visualization,” 2001, 1–5.

Yang, Huahai, Yunyao Li, and Michelle X Zhou. “Understand Users ’ Comprehension and Preferences for Composing” 21, no. 1 (2014).

———. “Understand Users&Rsquo; Comprehension and Preferences for Composing Information Visualizations.” *ACM Trans. Comput.-Hum. Interact.* 21, no. 1 (2014): 6:1–6:30. doi:10.1145/2541288.

Appendix I

What is the trend for the number of records in the past 7 years?

Where did most unsafe products come from?

Which 3 countries report the most unsafe products?

In which country did the number of notifications grow fastest throughout the last years?

What is the most common reaction taken on product safety issues?

Which reaction was not taken in the past two years?

Which are the two product categories with the most records?

Which year has the highest number of records?

Appendix II

What is the trend for the number of records in the past 7 years?

- Increasing
- Decreasing
- I don’t know

Where did most unsafe products come from?

- Germany
- USA
- China
- Taiwan

Which 3 countries report the most unsafe products?(please rate with number next to the country)

- Hungary
- Germany
- Uk
- Bulgaria
- Spain

In which country did the number of notifications grow fastest between 2005 to 2013 ?

- Spain
- Germany
- Bulgaria
- Hungary

What is the most common reaction taken on product safety issues?

- Recall
- Ban of Marketing
- Correction
- Import rejected

Which reaction was not taken in the past two years?

- Ban of marketing
- Seizure
- Correction
- Recall

Which are the two product categories with the most records?(please check mark)

- Motor Vehicles
- Cosmetics
- Toys
- Electrical appliances
- Clothing and textile

Which year has the highest number of records?

- 2013
- 2010
- 2012
- 2011