Measuring the user experience of data visualization

Author: *Tijmen van Willigen*

Supervisor: *Dr. Arjan van Hessen (University of Twente)*

Second Reader: *Dr. Mariët Theune (University of Twente)*

Direct Supervisor Company: *Drs. Ward Venrooij (TNO Soesterberg)*

Second Supervisor Company: *Dr. Alexander Toet (TNO Soesterberg)*

UNIVERSITEIT TWENTE.

TNO

March 2019

Graduation Project Interaction Technology

Abstract

In a world where data is increasingly being collected and used, it is important to develop ways to explore the data as well. This can be done by visualizing the data. The quality of such information visualizations is often measured with the usability metrics effectiveness and efficiency, which misses hedonic factors such as joy-of-use and aesthetic quality. The concept of user experience (UX) does include these factors and is a good predictor for the overall evaluation of an information visualization by the user.

However, there are numerous ways of measuring UX as the field is still young and has many definitions. This exploratory research examined if the CUE (components of user experience) model and its measurement tool meCUE, which were found to be promising candidates for measuring the UX of data visualization, are indeed suitable for the domain of data visualization. In specific, this research measured the UX of information visualizations which had small deviations in terms of animated elements, to see if measures of UX could explain the preferences of users.

The meCUE method could not measure the subtle differences in the experiment and in this case qualitative research seems superior to it. The results show the subjectivity of UX and outline the importance to specify a user group. The results also suggest that the evaluation of UX might benefit from not purely relying on a user's self-report and involve research interpretation and objectiveness.

Table of Contents

1

INTRODUCTION

1 Introduction

The amount of data is growing rapidly in society. There are numerous sources of data such as social media platforms, transactional information and internet activity. According to IBM every day 250 million gigabytes of data are produced (Winans et al., 2016). IDC predicts that that in 2025, 163 trillion GB of data will be created (Reinsel, Gantz, & Rydning, 2018), as shown in Figure 1. The question remains how to transform this enormous amount of data into insights and knowledge.

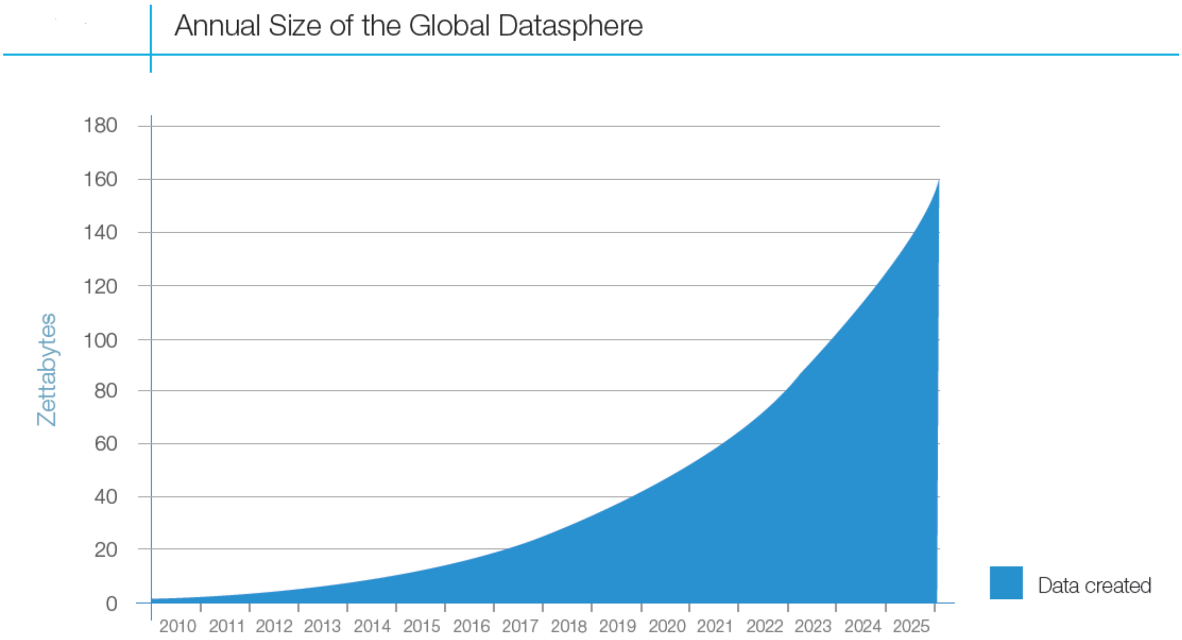

Figure 1 – Total predicted amount of data created according to IDC (Reinsel et al., 2018)

There is a lot of focus on techniques for collecting and managing this data, rapidly evolving the technology. However, there is little focus on human skills and ability to interpret the data (Few, 2009). Data visualization has potential to make the interpretation of abstract data easier, supporting the human skills. It shifts the balance between perception and cognition, taking fuller advantage of the brain's abilities (Few, 2006). When complex data is properly visualized, hidden messages can be revealed (Tukey & Wilk, 1966). However, it is important to design information visualizations in a way that they suit the abilities of the human perception and respect its limitations, as users could easily be overwhelmed with the available data (Few, 2009).

Therefore, it is important to be able to measure how well an information visualization is designed. This can be measured with objective usability measures such as the effectiveness (are users able to achieve their goals?) and the efficiency (how efficiently can those goals be achieved?) of an information visualization. Both the industry and academia have a high interest in the **subjective experience** of the users (Vermeeren et al., 2010), in an effort to fulfil the users' needs and wishes. This user experience (UX) is what will lead to the users' judgement (Thüring & Mahlke, 2007): a well-designed information visualization should thus give the user a positive experience. Logically, this experience is highly dependent on the usability of the visualization. It is however also dependent on factors that usability research often does not consider, such as aesthetics and joy of use. In data visualization, the current focus in evaluation frequently lies on usability aspects rather than the whole user experience (Cawthon & Moere, 2006), and little research has been done as to how to measure the UX of data visualization.

1.1 Scope and research questions

In a literature study concerning UX and data visualization (chapter 2), several models of UX are compared and assessed on their applicability in the domain of data visualization. The components of user experience (CUE) model was found to be a promising candidate, as it generally captures the most important aspects of data visualization and because the model allows for customization per specific domain. By means of experiment, this exploratory research will examine if the CUE model using its measurement tool meCUE are indeed suitable for measuring the UX of information visualizations in a quantitative manner, even if the differences between the conditions are small. The quantitative results will be compared to the preference of the user and a qualitative assessment of the UX. This experiment will specifically focus on animations within information visualizations, as they can both have hedonic and pragmatic value (explained in chapter 2). Several animations are manipulated to see if and how they influence the user experience of a task-based interaction with an information visualization. The same animations will be assessed using qualitative research to put the added value of the quantitative method in perspective.

Scope of the animations

Animations were chosen as independent variables as they can have an influence on both the non-instrumental quality and the instrumental quality of a visualization. Two types of animations will be used in this experiment. First, loading animations will be used as they mainly influence the non-instrumental quality of the visualization, e.g. the aesthetics. Loading animations are the animations that play when the graph is loaded and can be interpreted as the process of the data being loaded into the graph. These animations can be related to the application area 'functional description' of Bartram's taxonomy of application areas for motion (Bartram, 1998). Two standard loading animations from the d3 libraries Amcharts (www.amcharts.com) and Highcharts (www.highcharts.com) will be used for the loading animations. Second, transition animations will be used, being able to both influence the instrumental and non-instrumental quality of the visualization. Two different animations suggested by *Heer & Robertson (2007)* will be used for the transition between a 'stacked bar graph' and a 'grouped bar graph', giving the participant a better understanding of the relation between the two charts.

Scope of the context, system and user

As context, system and user are very extensive concepts that influence the UX, it is important to scope and specify them. This research will focus on task driven interactions with information visualizations; interactions where the goals shape the activities. In this case both hedonic and pragmatic qualities play a substantial role in the UX, according to Hassenzahl et al. (Hassenzahl, Kekez, & Burmester, 2002). Another contextual influence that can be accounted for is the screen that the visualizations will be presented on; for consistency reasons this experiment will only be allowed to run on a desktop screen. The visualization itself will only deviate in terms of animation; making the animations the only aspect influencing the UX. The data will deviate per visualization but have the same nature and cardinality, trying not to influence the UX. Other aspects, such as the colour of the visualization, will be kept the same for all conditions. Considering the users, this research aims at the naïve users that see the particular visualization for the first time and have an information need; in this case caused by the tasks of the experiment. Their emotional state is out of scope for this research even though it influences the UX. The amount of experience with information visualizations can be very different amongst users and will not be specified or researched, as this research tries to make generalizable conclusions over any type of user with an information need. Not overestimating the capabilities of the human perception and respect the limits, relatively easy types of graphs are used: bar graphs, stacked bar graphs and grouped bar graphs.

Research questions

As the concept of UX is associated with a wide variety of meanings, a literature study was first conducted to compare models and definitions of UX and relate them to the domain of data visualization. To find UX models that apply well to the field of data visualization, data visualization was also investigated in more depth. The first research question is;

RQ1 – *How can the UX of an information visualization be measured in a quantitative manner?*

> RQ1.1 – *How is UX defined and modelled?*
>
> RQ1.2 – *Which UX models are the most suitable for data visualization?*

By means of a literature study, the CUE model with the measurement tool meCUE were found to be promising candidates to measure UX in a quantitative manner. The second research question is therefore;

RQ2 – *What aspects of the CUE model and its measurement tool meCUE can be used for the domain of information visualization?*

> RQ2.1 – *What differences in UX can be measured using the CUE model and its measurement tool meCUE, and how does it relate to a qualitative evaluation of the UX?*
>
> RQ2.2 – *Can a visualization preference be explained using the scores of the meCUE questionnaire constructs?*

To answer these questions, the CUE model and the meCUE questionnaire were used to measure the UX of different versions of the same visualization. The visualization deviated in terms of animations, as animations can have both an instrumental value and a non-instrumental value (see chapter 2). In addition to the meCUE results, other forms of assessment of UX were evaluated. Qualitative feedback about the conditions was gathered to be able to compare the quantitative UX assessment to a qualitative form. Further, a questionnaire was conducted after comparing and explaining the differences in the conditions. The following research questions will guide in answering the questions above;

RQ3 – *How can loading animations and transition animations influence the UX of information visualizations?*

> RQ3.1 – *How does a bouncy loading animation in a bar graph (enlarging the bars and elastically bouncing around their value before reaching the static point of their value) affect the UX in a goal-driven interaction with an information visualization compared to the same visualization without loading animation?*
>
> RQ3.2 – *How does a calm loading animation in a bar graph (gradually enlarging the bars) affect the UX in a goal-driven interaction with an information visualization compared to the same visualization without loading animation?*
>
> RQ3.3 – *How does a direct transition animation (gradually moving from one chart into another using a direct animation, directly interpolating between start and end state. A representation of such a transition (Heer & Robertson, 2007)) affect the UX in a goal-driven interaction with an information visualization compared to the same visualization without transition animation?*
>
> RQ3.4 – *How does a staged transition animation (using two animation stages, where the first stage changes the widths and x-coordinates of the bars and the second stage drops the bars down to the baseline (Heer & Robertson, 2007)) affect the UX in a goal-driven interaction with an information visualization compared to the same visualization without transition animation?*

Considering the literature reviewed in chapter 2, it is expected that the calm loading animation will increase the perception of the aesthetic quality of the visualization as opposed to no animation. The bouncy loading animation could be perceived as distracting, even though some participants might find it attractive and stylish. The transition animations are both expected to have a positive effect on the UX, as participants might better understand the relation between the two views of the graph (Heer & Robertson, 2007). Besides, the transitions might increase the perception of the aesthetic quality.

For all conditions, the animations are likely to increase the engagement. The differences between no animation and animation are expected to be larger than the differences between the animations, as these differences are very subtle. The visualizations that are perceived to have a higher aesthetic quality, might be perceived as being more usable and useful too. This would support the claim: 'what is beautiful is usable' by Tractinsky et al. (Tractinsky, Katz, & Ikar, 2000).

The results of the meCUE measurements are expected to portray significant differences for most conditions, even though the differences between the conditions are small. For conditions which are very much alike, for example the two transition animations, there might be no measurable difference at all. It is expected that preferences participants have for a certain condition can be explained by differences on specific constructs in the meCUE results between conditions, as suggested by the CUE model.

2

Measuring the UX of data visualization

2 Measuring the UX of data visualization

2.1 Data visualization

The field of data visualization has gained momentum since the digital age, as more and more data became available. Even though the tabular representation was already used in the 2nd century to store for example astronomical information, the first time that quantitative data was presented in two dimensional graphs was much later, around the 17th century. Rene Descartes, a French philosopher and mathematician, invented the two-dimensional graphs using X and Y axis. After that, in the late 18th and early 19th century, the graphs known today were invented or improved. For example, William Playfair invented charts as the bar chart and the pie chart. In the 19th century universities started to recognize the field of data graphing. The statistics professor John Tukey recognized the power of visualization and introduced a new approach to analysing data called exploratory data analysis in 1977. A few years later Tufte (1983) wrote the ground-breaking "the visual display of quantitative information", which showed effective ways of displaying data visually.

Since we are living in a society which makes increasingly use of data intensive technologies, we have lots of data at our disposal. The large amount shows the potential it has, yet the question often rises how we should explore it. There is great focus on technology, for example the tremendous progress in technologies allowing us to collect, store and access data. However, there is little focus on human skills to interpret the data (Few, 2009), and we need human skills in order to make sense of data. Even though there are a lot of tools allowing us to explore and visualize data, the results depend on how skilled humans are in employing them (Few, 2009). According to Few, good data analysis will help us:

- *To better understand what is going on now*
- *To better predict what will likely happen under particular conditions in the future, so opportunities can be created and problems can be prevented*

Information visualizations are often used in the form dashboards. Few (2006) defines a dashboard as a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance. Dashboards usually display insights from different perspectives and shows the relations between these perspectives.

2.1.1 Data, information and knowledge

In visualization, data, information and knowledge are three terms used extensively, often to indicate different levels of abstraction, understanding or truthfulness (Chen et al., 2009). Data visualization and information visualization are often used as synonyms, generally referring to the techniques used to communicate data by encoding it in visual objects. The data-information-knowledge-wisdom (DIKW) hierarchy is a common model for humans' understanding in perceptual and cognitive space (Figure 2) (Rowley, 2007), explaining the difference between data and information. According to the original theory by Ackoff (1989):

- **Data** consists of raw symbols;
- **Information** is data that is given meaning, providing answers to 'who, what, where and when questions';
- **Knowledge** is the application of data and information; providing answers to the 'how questions', giving context to the information;
- **Wisdom** is the understanding of the knowledge, being integrated and actionable.

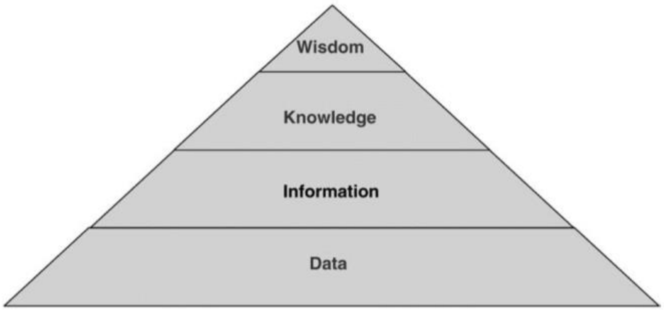

Figure 2 – The data-information-knowledge-wisdom (DIKW) hierarchy is a common model for human's understanding in perceptual and cognitive space, explaining the difference between data and information (Rowley, 2007)

Few (2009) describes data visualization as an umbrella term, where data visualization entails the communication, graphical representation and understanding of data with as end goal making good decisions. Information visualization can be seen as a specific form of graphical representation. Card et al. define information visualization as *"The use of computer supported, interactive, visual representations of abstract data to amplify cognition"* (Card, Mackinlay, & Shneiderman, 1999). These computer-supported and interactive visualizations can be contrasted to info graphics, where the visualization is usually a static graphical representation.

2.1.2 Understanding the data

2.1.2.1 Why data visualization?

Visual representations help to understand and explore the data. In contrast to complex statistical analysis, which is usually only for trained specialists, they are broadly accessible. Tukey & Wilk (1966) point out that: *"One great virtue of good graphical representation is that it can serve to display clearly and effectively a message carried by quantities whose calculation or observation is far from simple"*. Next to that Few (2009) states that visual representations help us to see more at once and remember the message better. This can be illustrated by comparing the table and graph in Figure 3. The graph portrays trends and peaks immediately, whereas the table should be examined thoroughly in order to find the same characteristics.

Revenue

Region	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
West	28,384	30,288	34,302	32,039	32,938	34,392	33,923	33,092	34,934	30,384	33,923	37,834	396,433
Central	15,934	16,934	17,173	16,394	17,345	16,384	15,302	14,939	14,039	12,304	11,033	9,283	177,064
East	11,293	12,384	12,938	12,034	11,034	13,983	12,384	12,374	12,384	13,374	14,394	19,283	157,859
Total	55,611	59,606	64,413	60,467	61,317	64,759	61,609	60,405	61,357	56,062	59,350	66,400	731,356

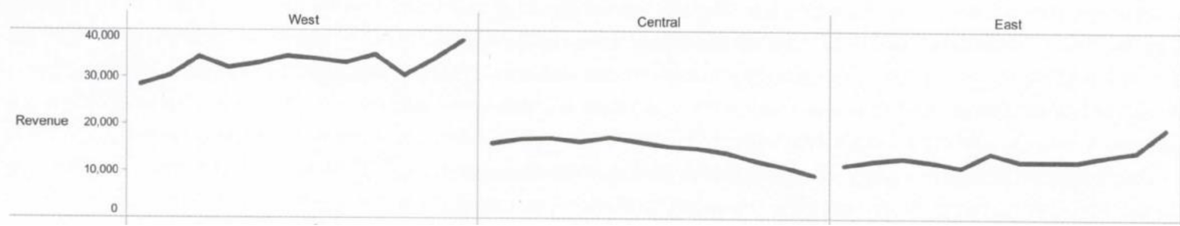

Figure 3 - Table versus graphs, where visual representations help us to see more at once and remember the message better. Edited from (Few, 2009)

Visual representations are suitable in emphasizing certain aspects of the data and telling a story, making it important to decide what to communicate. The same data can tell many different stories, depending on the size and cardinality of the data. A certain value could for example display an increase over time, but also a value per region. What a visualization should tell depends largely on the target group; what information do they need, how do they process visual information in general and how do they want to see the information they need? Knowing the user and their level of skills is important in designing any user interface (Zeng, 2005). Efforts have been made in defining and developing frameworks around the literacy of data visualization (DVL) (Börner, Bueckle, & Ginda, 2019), helping to determine the specific skill levels and guiding the design of data visualization.

Different kinds of visualizations serve different purposes, and it is important to choose the correct visualization for the right purpose. There are numerous types of visualizations, a useful collection of them can be found on https://datavizcatalogue.com. Two very basic visualizations, the line and bar charts, have for example very different uses (Sas, 2013). On one hand line charts are often used to track changes over time and are useful when comparing multiple items over the same time. On the other hand, bar charts are used to compare different quantities of categories or groups. When used in a wrong manner, the message can get lost. An example of this is shown in Figure 4 and Figure 5, where the graphs in Figure 4 match the data, whereas the graphs in Figure 5 feel less intuitive or even odd.

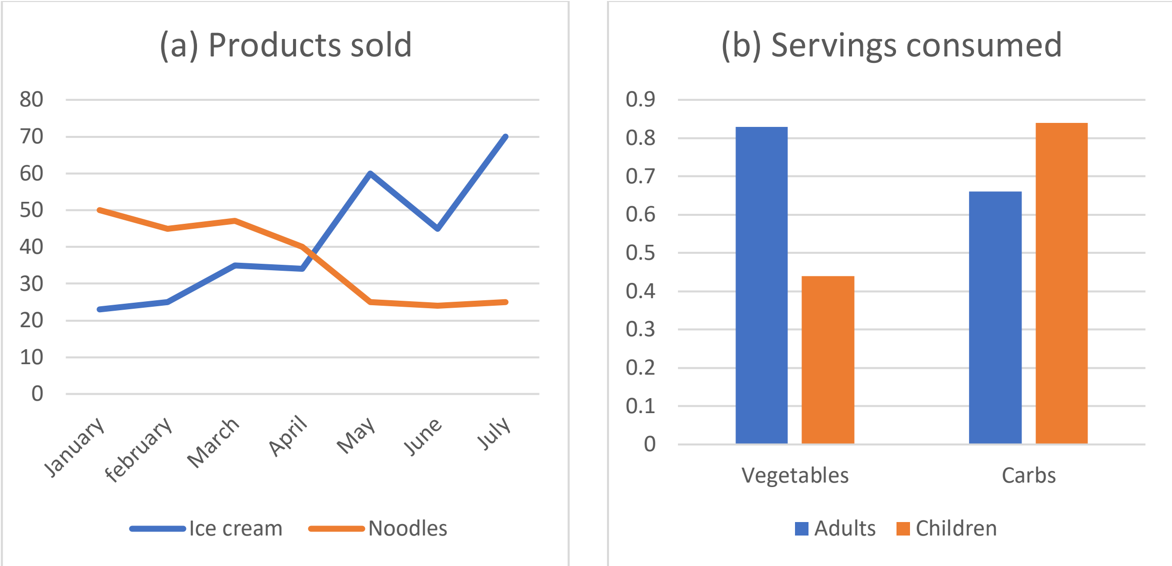

Figure 4 – (a) Line chart and (b) bar chart that match the type of their data (created in excel with fictional data)

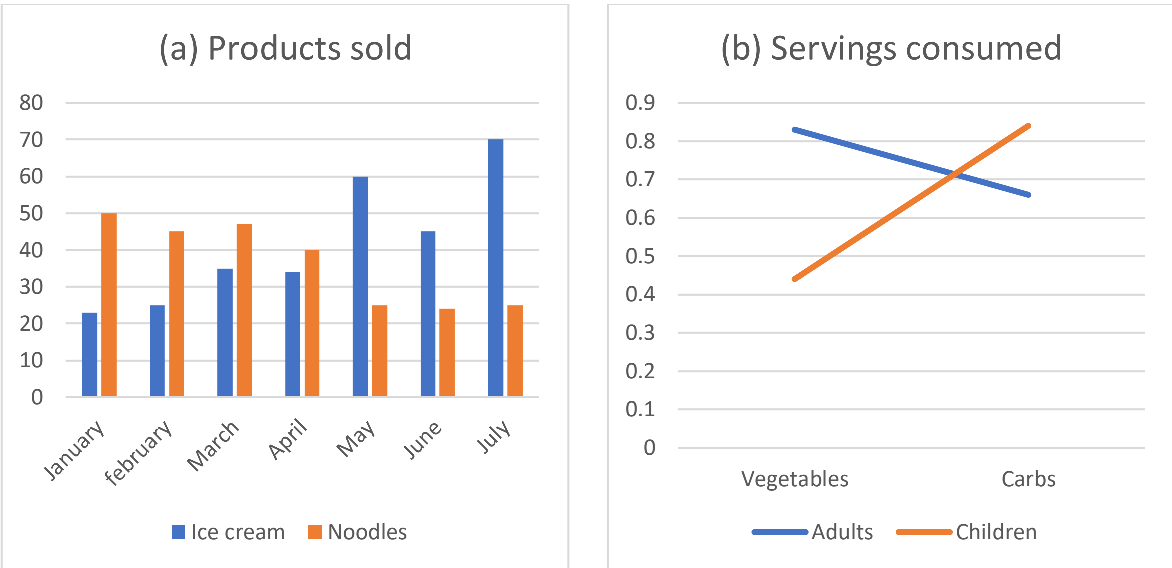

Figure 5 – (a) Bar chart and (b) line chart not matching the type of their data, resulting in unclear visualizations (created in Excel with fictional data)

2.1.2.2 'Misleading' with data visualization

There is a thin line between a visualization with a strong emphasis and 'misleading' visualizations. Visualizations can get misleading when someone is for example too eager to convey a certain message leaving out important context, and/or because someone is unknowing and does not use conventions. Visualization of data can be seen as storytelling, where someone can freely express their interpretation of a dataset. One dataset can have multiple interpretations and therefore convey many different stories. It is therefore useful to be aware of common ways that are used to 'lie with statistics' (Herne & Huff, 2006), some examples are listed below.

A **truncated Y axis** is a classic way to visually 'mislead'. It occurs when graph's producers ignore conventions and manipulate the y-axis. Conventionally, the y-axis starts at 0 going up to the highest point of the data. A y-axis might however be truncated on purpose to show a very small difference. An example of this phenomenon is shown in Figure 6a and b, where the increase in figure seems quite large, whereas the actual increase as shown in figure b is almost unnoticeable. In this case the bottom line is thicker than the other lines, which also visually suggests that the bottom line is 0.

Omitting data is another process prone to misleading; by leaving out certain data points, trends that might not actually exist can become visible. By omitting data, there is a risk of crucial information. An example is shown in Figure 7 a and b, where figure b displays half of the data of figure a, creating a trend which cannot be seen in figure a. A related way of misleading with a visualization is to **crop the X or Y axis**, or both. Usually this is just a perfectly fine way of zooming in to the data and leave out unnecessary context. However, the intention should not be to make a story better than it actually is. By only cropping out an increase for example, data can suggest a trend to be more positive than it actually is.

A final example is the **correlation – causation** issue. Correlation does not imply causation. Nevertheless, a correlation is often seen as a causation, for example by internet articles with headers such as *"People drinking beer live longer; drinking beer is healthy!"*.

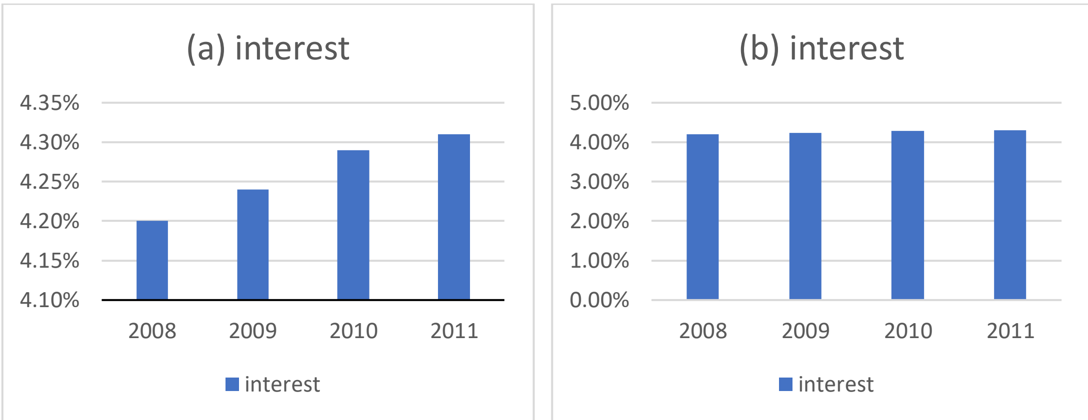

Figure 6 – Example of a truncated y axis, where the increase in interest in (a) seems quite large, whereas the actual increase (b) is almost unnoticeable (graphs created in excel)

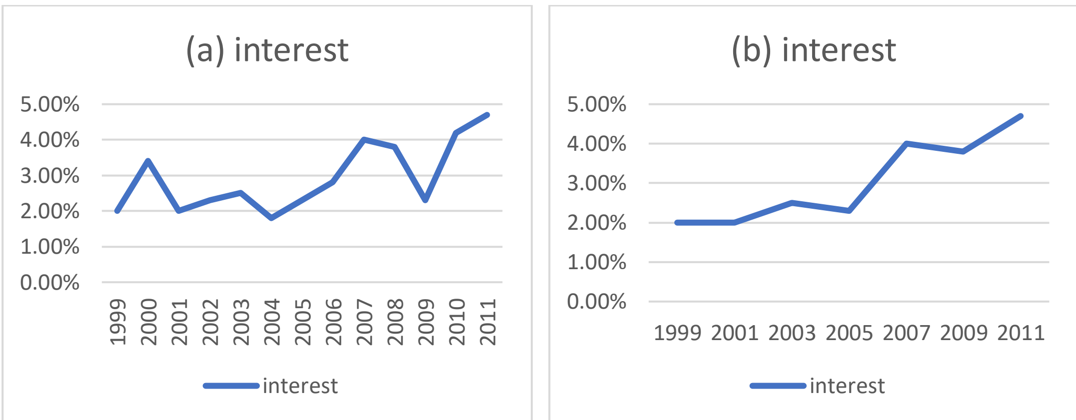

Figure 7 – Example of omitting data, where (a) shows the original data source, and (b) shows half of the data of (a), creating a trend that cannot be seen in (a) (graphs created in excel)

2.1.2.3 *Human perception & visual encodings*

Technology has rapidly evolved, and this has not been met with a parallel evolution of the human being. It is therefore important to design information visualizations that suit the abilities of the human perception and respect the limitations (Few, 2009).

Bertin (1983) argued that visual perception operates according to rules that can be followed in order to clearly visualize information. He makes a case that several basic attributes of visuals are perceived pre-attentively; these are the visual features that are perceived before conscious awareness. Also Ware (2004) emphasizes the importance of using the pre-attentive features when creating visual representations of abstract information. He states that certain simple shapes and colours pop out from their surroundings; they can be visually identified, even after very brief exposure.

Few (2009) sets out several facts about knowledge of perception:

- We do not attend to everything that we see, logically since awareness of everything that we see would overwhelm us. In visualizations we should therefore strive to let meaningful information stand out in contrast to what's not worth our attention.
- Our eyes are drawn to familiar patterns; we see what we know and expect. Information visualization should therefore also be rooted in an understanding of how people think.
- Working memory plays an important role in human cognition but is extremely limited. We only remember the elements to which we attend. Information visualizations should therefore serve as an external aid to augment working memory.

Following perception based rules, data can be presented in such a way that the important and informative patterns stand out (Ware, 2004). Because abstract data has no natural physical form, it must be visualized using colours and shapes that represent the data in perceptible and meaningful ways (Few, 2009). Originally provided by Ware (2004), Few has listed the most relevant pre-attentive attributes which are most useful in information visualization (Few, 2009), as can be found in Figure 8.

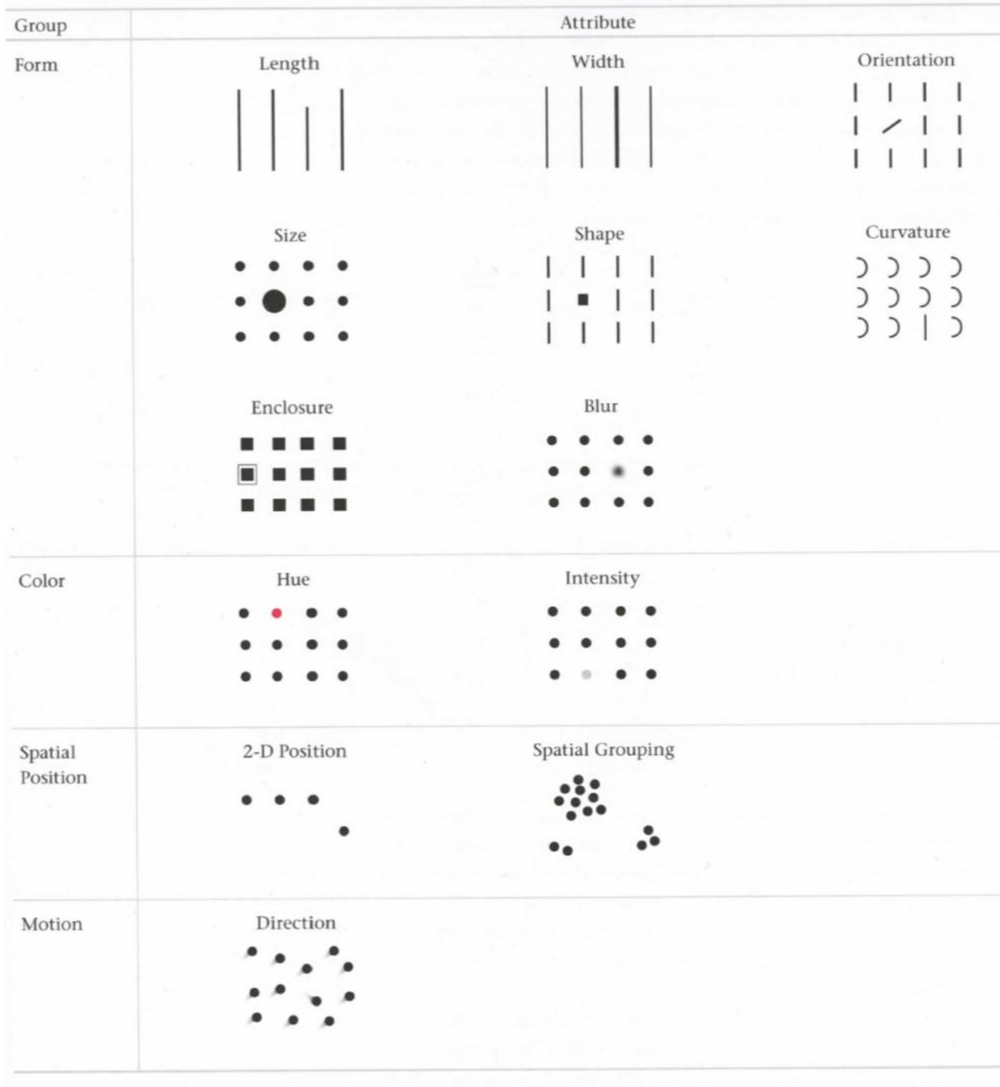

Figure 8 – Selection of pre-attentive attributes of visual perception (Few, 2009) that are most useful in information visualization for encoding

From the attributes depicted in Figure 8, Few describes how length and 2D position are perceived very precise, whereas width, size, intensity and blur are not. As a consequence, most common graphs use these the features 2D position and length (for example bar graphs or scatter plots). This claim is supported by a research from Stanford University (Heer & Bostock, 2010) showing the accuracy of visual decoding with the expected error rates for different encoding types in Figure 9. The different forms of position encoding that were measured are depicted in Figure 10, illustrating that users can more precisely estimate length when the items to compare are closer to each other. Looking at the huge difference in expected error between position and area, it is no surprise that pie charts are often criticized form of visualization (Wilkinson, 2010), as it is based on comparing areas and angles.

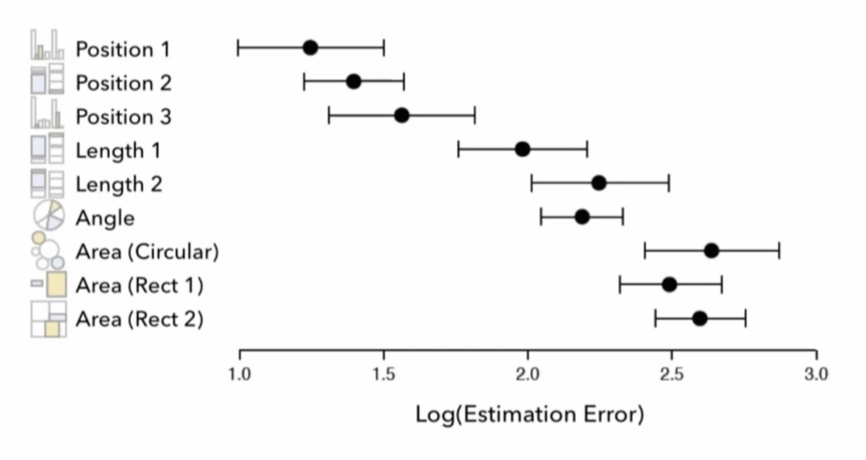

Figure 9 – The accuracy of visual decoding; results from a research from Stanford University showing the accuracy of visual decoding with the expected error rates for different encoding (Heer & Bostock, 2010)

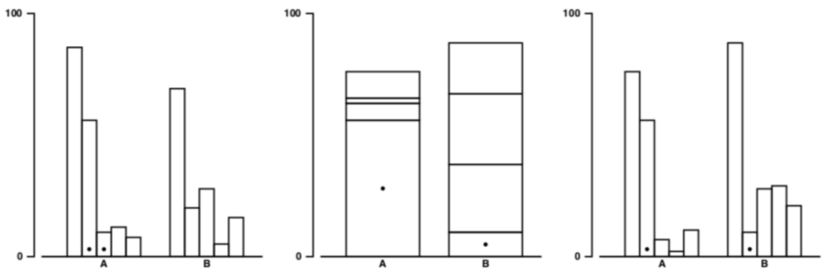

Figure 10 - The different forms of position that were measured, illustrating that its users can more precisely estimate length when the items to compare are closer to each other. From left to right: position 1, position 2, position 3 from Figure 9 (Heer & Bostock, 2010)

2.1.2.3.1 Gestalt principles of perception

The gestalt psychology tries to understand the ability to acquire and maintain perceptions in an apparently chaotic world. The central idea is to view information as a whole rather than the sum of its parts. Applying gestalt principles on the design of information visualizations has a positive effect on the understandability. Lemon, Allen, Carver, & Bradshaw (2007) for example outlined how gestalt principles of similarity, proximity and continuity influence diagram comprehension while Rusu, Fabian, Jianu, & Rusu (2011) show how using the gestalt principle of closure can improve graph readability.

The key ideas of the gestalt psychology are the principles of emergence, reification, multi-stability and invariance. The principle of **emergence** addresses the process where humans usually first identify the whole and then the parts. The principle of **reification** addresses the aspect of perception in which the objects are perceived to have more spatial information than what is actually present; human perception seems to fill in the gaps. The principle of **multi-stability** describes the tendency of ambiguous perceptual experiences to switch between alternative interpretations, not being able to see two interpretations at once in an effort to avoid visual uncertainty. The principle of **invariance** addresses the fact that similar and different objects can be identified independent of the scale, rotation or translation.

Figure 11 demonstrates an example of emergence, where the dog can only be seen by looking at the image as a whole. Figure 12 illustrates an example of reification where a sphere can clearly be identified in the centre, even though there is none. Figure 13 illustrates an example of multi-stability, where humans can see the cube in two ways. Figure 14 demonstrates an example of invariance, where similar objects can be identified, even though the orientations are very different.

Figure 11 – Example of Emergence, where the dog can only be seen by looking at the image as a whole (source: thatbrandguy.com)

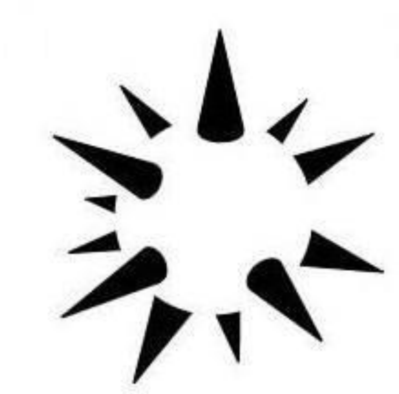

Figure 12 – Example of Reification, where a sphere can clearly be seen in the centre, even though there is none (source: study.com)

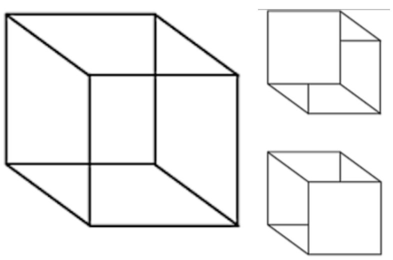

Figure 13 – Example of Multi-stability, where the cube can be seen in two ways (source: geoff-hart.com)

Figure 14 – Example of Invariance, where similar objects can be identified even though the orientations are very different. (source: cns-alumni.bu.edu)

In addition to these key ideas, several laws of the gestalt psychology exist. The most interesting ones in relation to information visualization are listed below.

- Law of Similarity: Items that are similar are grouped together by the brain, as is shown in the example in Figure 15.

- Law of Pragnanz: People will perceive and interpret ambiguous or complex images as the simplest form(s) possible. In the example in Figure 16 the form is interpreted as two circles, whereas one could also distinguish two mirrored half-moons.
- Law of Proximity: objects that are close are grouped together, as shown in the example in Figure 17. On the left all circles are equally close together and thus seen as one group, whereas at the right the circles are grouped into separate smaller groups.
- Law of Continuity: lines are seen as following the smoothest path. Figure 18 shows an example of this principle, where a straight and a curved line crossing are seen, instead of two similar mirrored lines next to each other.
- Law of Closure: objects that are grouped together are seen as a whole, and the mind is filling the missing information. Figure 19 shows three examples of this principle, where all examples do not explicitly show a square, but the mind sees a square in all examples.
- Law of uniform connectedness: items that are visually connected are perceived as more related. Figure 20 shows are different shapes are connected by a line, forming connectedness between the different shapes, rather than within the same shapes.
- Law of common regions: Elements that are located in the same closed region are perceived as part of a group. Figure 21 shows this principle, where the closed regions alter the way the groups are perceived.
- Law of focal points: Elements with a point of interest, emphasis or difference will capture and hold the viewer's attention, as shown in Figure 22.
- Law of past experiences: elements can be perceived according to an observer's past experience. Most of the times this is very subjective, but humans also have a lot of past experiences in common, as the familiar colours shown in Figure 23.

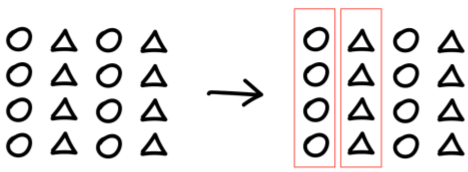

Figure 15 - Gestalt law of similarity: items that are similar are grouped together by the brain.

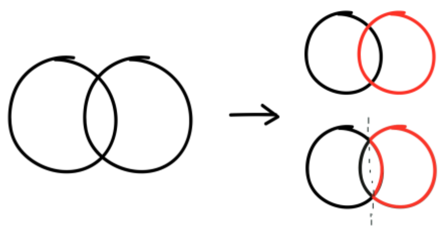

Figure 16 - Gestalt law of pragnanz: people will perceive and interpret ambiguous or complex images as the simplest form(s) possible. In this example the form is interpreted as two circles, whereas one could also distinguish two mirrored half-moons.

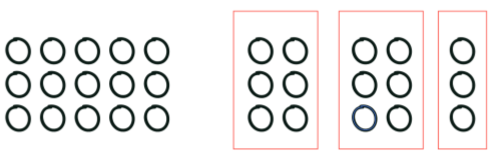

Figure 17 - Gestalt law of proximity: objects that are close are grouped together. On the left all circles are equally close together and thus seen as one group, whereas at the right the circles are grouped into separate smaller groups.

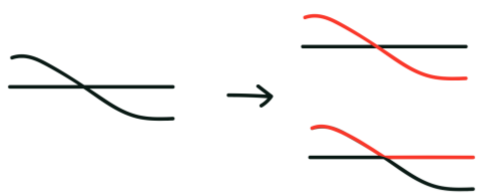

Figure 18 – Example of the gestalt law of continuity, where a straight and a curved line crossing are seen, instead of two similar mirrored lines next to each other

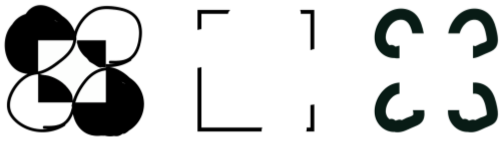

Figure 19 – Examples of the gestalt law of closure, where all examples do not explicitly show a square, but the mind sees a square in all examples.

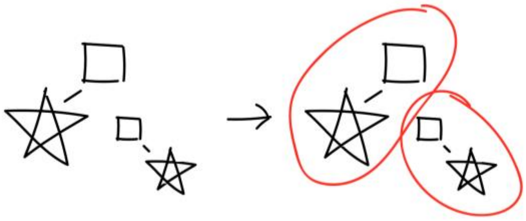

Figure 20 – Example of the gestalt law of uniform connectedness, where different shapes are connected by a line, forming connectedness between the different shapes, rather than within the same shapes.

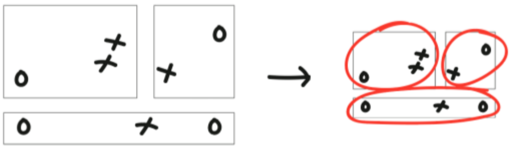

Figure 21 - Gestalt law of common regions: The closed regions alter the way the groups are perceived.

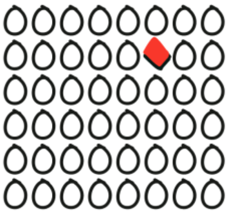

Figure 22 - Gestalt law of focal points: Elements with a point of interest, emphasis or difference will capture and hold the viewer's attention

Figure 23 - Gestalt law of past experiences: Humans have a lot of past experiences in common, such as the colours red, orange and green from for example a traffic light.

2.1.3 Aspects influencing the hedonic quality of data visualization

2.1.3.1 Aesthetics

There is a debate on the importance of aesthetic quality in information visualization. Some see it as an added bonus (Skog, Ljungblad, & Holmquist, 2003), whereas others show how aesthetics can have a positive influence on the usability (Sonderegger, Uebelbacher, Pugliese, & Sauer, 2014) (Kurosu & Kashimura, 1995), and more specific the effectiveness, efficiency and rate of task abandonment (Cawthon & Moere, 2007). Tractinsky et al. even argue that "what is beautiful is usable" (Tractinsky et al., 2000). They show that if something is more beautiful it is also perceived as more usable, called the aesthetic-usability effect. Cawthon & Moere (2006) argue that a user centred evaluation method not solely centred around task efficiency metrics is imperative.

Norman (2004) argues that by experiencing emotions humans unravel problems, as the human emotional system is intertwined with cognitive abilities. Even though this was originally aimed at the context of industrial products, it could as well be applicable to information visualizations. Also Sheldon et al. (2001) note that satisfaction of human needs is seen as a driver of experiences. Lachner, Naegelein, Kowalski, Spann, & Butz (2016) however suggest that such psychological needs are rather applicable to macro perspective (i.e. products overall purpose), and micro perspectives (e.g. visual characteristics) should be analysed in detail.

Cawthon & Moere (2007) show how high aesthetic quality can lead to a positive influence task abandonment rate. By looking at a visualization for a longer time, the interaction becomes more efficient and effective as less people abandon their task, even though the less aesthetic visualization would probably be more effective and efficient if people didn't abandon their task as fast. This finding suggests that the importance of aesthetics also largely depends on the kind of application the information visualization is used in. How strong is the information need from the user? Is the initiative of the information transmission taken by the user or the information provider?

For first time use, aesthetic quality has an even larger impact on the user. Jiang, Wang, Tan, & Yu (2016) have shown that in the context of websites, during a first encounter aesthetics have a larger impact on the attitude towards a website than perceived utility. The same likely holds for data visualization, meaning that especially during first time use of an information visualization the aesthetics are extremely important; possibly even more important that the perceived utility.

On the other hand, aesthetically appealing elements can reduce the effectiveness of the visualization when used without care, by obscuring the intended message (Tufte, 1983) (Brath, Peters, & Senior, 2005). Sonderegger et al. (2014) also warn that there may be a risk to overestimate usability of a product if relying only on subjective measures of a highly appealing product. Aspects like unnecessary colours, 3D elements, gradients and textures are often referred to as 'chart junk' (Tufte, 1983). One could however be specifically aiming at the memorability of the graph or incorporate it as part of artistic expression; accepting the loss in effectiveness, efficiency or readability. Examples of such visualization are shown in Figure 24.

Figure 24 – Examples of Chart junk (source: eagereyes.org)

Researchers trying to find indicators for perceived aesthetics, often mention visual complexity as the biggest influencer (Reinecke et al., 2013) (Michailidou, Harper, & Bechhofer, 2008), indicating that it is in particular important to keep designs as simple as possible. Colourfulness and harmony in colours also have an influence, but it is not as large as the influence of visual complexity. Colourfulness is the perceived intensity of the colours, measured with a function of saturation of different colours. Research also suggest to use personalized models, as age seemed to correlate with the influence of visual complexity on perceived aesthetics, and education level with colourfulness (Reinecke et al., 2013).

2.1.3.2 Interaction

Wimmer, Weishapl, Grechenig, & Kappel (2011) proposes to incorporate interaction specifically as an aesthetic quality in models for UX. In their study Wimmer et al. (2011) show that physical interaction affects the perceived aesthetic quality and hypothesize that this same holds for any other interaction characteristics. They emphasize that the concepts beauty and aesthetics are different from each other, as the physical behaviour in their research had no significant effect on the beauty (Wimmer et al., 2011).

Figueiras (2015) proposes eleven categories of interaction techniques for information visualization:

- Filtering – Only showing data in which the user is interested
- Selecting – The ability to mark or track items
- Abstract/elaborate – The ability to adjust the level of abstraction
- Overview and explore – Having an overview first, then zoom and filter and details on demand.
- Connect/relate – The ability to show the user how data is related
- History – Allowing the user to retrace steps in the exploration of the data
- Extraction of features – Allowing the user to extract data
- Reconfigure – Giving the user different arrangements of the data
- Encode – Giving the user a different representation of the data
- Participation/collaboration – Allowing the user to contribute to the data
- Gamification – Showing the data in a more playful way

2.1.3.3 Animation

An animation is a sequence of images that is characterized by subtle but highly structured changes between consecutive frame over space and over time; which create the illusion of movement in the human brain (Friedrich, 2002). Animations have a strong visual impact, and not all users like it (Bederson & Boltman, 2007), considering a user group is therefore important. Animation or motion can both be viewed from a pragmatic or hedonic perspective in information visualization. From a pragmatic point of view, animation is often seen as a promising candidate to increase the dimensionality of visualizations (Bartram, 1997), especially now hard- and software have grown to support it. Next to that, motion is also pre-attentively perceived and is therefore used to shift some of the users cognitive load to the human perceptual system (Robertson, Mackinlay, & Card, 1991). From a hedonic perspective motion can be useful by enhancing the perception of aesthetic quality (Bacigalupi, 1998) (Bartram & Nakatani, 2010).

Motion and animation in information visualization can both help and hurt the visualization (Heer & Robertson, 2007). Motion can for example attract the attention; being even more powerful than colour or shape (Bartram, 1998). On the other hand, motion can distract from the actual message when used without care (Hong, Thong, & Tam, 2004). Motion can be effective for object constancy; where users can track changes, for example with the scale of a graph or in graph transitions (Heer & Robertson, 2007) (Friedrich, 2002). When used without care, motion could however suggest false relations. Motion can enhance engagement (Bartram & Nakatani, 2010) but also be perceived as chart junk.

Bartram (1998) proposed a taxonomy of application areas for motion:

- **Awareness**: providing contextual information outside the specific area of attention or task
- **Transition**: process of smoothly guiding the user between different view or states
- **Functional description**: related to the behaviour of what the animated object or process represents. (e.g. 'scrolling paper through printer')
- **Emphasis**: uses motion to draw attention to a particular visual element or process.
- **Expression**: usually involves character-based animation and uses motions to enhance or enrich the user's sense of involvement with the task or application
- **Representation of change:** relates to indicating time-based behaviour and how objects and processes transform over some defined time frame.
- **Direct visualization**: maps motion attributes such as phase or frequency to actual data variables.
- **Association**: uses groups and/or sequences of motion to convey relationships between groups of information objects.

The animation duration is important yet very dependent on the application and context. Animations too slow may prove boring, while those that are too fast may result in increased errors. Optimal animation time may be hard to predict and subject to both the complexity of the system and the familiarity of the viewer (Heer & Robertson, 2007). Bartram (1998) argues that participants tend to wait for an animation to stop before they respond; therefore, longer animation times can impede search while the motion is active making short durations therefore often beneficial. Bederson & Boltman (2007) however argue that the time spend for animating does not seem to hurt the UX.

2.1.4 Discussion

Data visualization can help in the accessibility of the data, as it broadens the user group of data to more than just trained professionals. As initiated by Few (2009), this thesis will view data visualization as an umbrella term for all processes including the communication, graphical representation and understanding of the data with as end goal making good decisions. Information visualization, the main topic of this thesis, is a specific form of graphical representation.

A good information visualization should match the capabilities of the human perception and respect its limits. For that reason, it is also important to know the target group as good as possible, knowing the user needs and characteristics. This way a better choice can be made in how a story should be told, and what specific graphs should be used. In choosing the graphs the pre-attentive attributes described by Few form a useful help, as well as the gestalt psychology principles which can help us understand the differences between presentation and perception.

2.2 User experience

2.2.1 Defining UX

User Experience (UX) is associated with a wide variety of meanings (Forlizzi & Battarbee, 2004), varying between disciplines and experts. The meanings range from usability to beauty, hedonic, effective or experiential aspects of technology use (Hassenzahl & Tractinsky, 2006). A collection of definitions is gather on the website of allaboutUX.[1] UX covers many research fields and each discipline has a different view on UX (Alves, Valente, & Nunes, 2014). Most of the definitions seem to agree on the fact that UX is about the experience of *an interaction*. Some definitions have a business perspective and a marketing oriented focus[2,3], whereas others purely focus on the UX of interactive products and have a HCI perspective (McNamara & Kirakowski, 2006) (Sutcliffe, 2009).

UX became more important in recent years, mostly as a countermovement to the dominant, task- and work-related 'usability' paradigm (Hassenzahl & Tractinsky, 2006). The terms are overlapping; according to the ISO standards (ISO 9241-11, 2017), both usability and UX are outcomes of use. From a UX point of view, usability can be seen as a product aspect, influencing the UX. Usability criteria can therefore be used to assess aspects of UX, but UX includes other important aspects that traditional usability research does not consider like aesthetic qualities and emotional experiences, shown to be important in explaining why users prefer some systems over others (Thüring & Mahlke, 2007). It is also important to note that UX is not something one can design, UX can only be designed for. The context and the user will always influence the experience. A first-time use could have a whole different experience than a 10th time use, suggesting that the UX evolves over time (Karapanos, 2013) (Minge, 2008).

As the term UX became more popular it seemed that UX was used as a buzzword for a variety of aspects that didn't fit the usability paradigm, making the term fuzzy as there was no standard definition available. According to the UX whitepaper by Roto, Law, & Vermeeren (2011), UX is often used as a synonym for *"usability, user interface, interaction experience, interaction design, customer experience, web site appeal, emotion, 'wow effect', general experience, or as an umbrella term incorporating all or many of these concepts."* Each of these terms might be closely related to UX but has a different meaning.

There are several definitions of UX that seem to fit the area of data visualization. In 1996 Alben presented an early but broad definition of UX which is still often referred to:

> *All the aspects of how people use an interactive product: the way it feels in their hands, how well they understand how it works, how they feel about it while they're using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it (Alben, 1996).*

Hassenzahl defined UX using three main factors; the user state, the characteristics of the design and the context. In his definition he also describes how UX is related to usability, by treating usability as a characteristic of the system. Hassenzahl defined UX as:

> *A consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g.*

[1] UX definitions by allaboutux.org: http://www.allaboutux.org/ux-definitions

[2] The User Experience Professionals' Association (UXPA) defition of UX: http://www.usabilitybok.org/glossary

[3] UX defenition by Nielson Norman Group: https://www.nngroup.com/articles/definition-user-experience

complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.) (Hassenzahl & Tractinsky, 2006)

Roto however points out that the key difference between UX and Usability is that UX is a personal, subjective feeling about the product (Roto, 2007), which many definitions fail to address. The International Standardization Organization (ISO) made an effort to find a standard in the definition of User Experience. The ISO defines UX as:

The user's perceptions and responses that result from the use and/or anticipated use of a system, product or service (ISO 9241-11, 2017).

An additional note is made by the ISO that these perceptions and responses include the user's emotions and physical and psychological responses that occur before, during or after use. According to Law et al. (Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009) the ISO definition is a very promising one, but they note that some terms will need further explanation.

In this thesis the ISO definition will be used, with an extension from Hassenzahl's definition that these user perceptions and responses are a consequence of the user's internal state, the system, and the context.

2.2.2 UX evaluation methods

Roto et al. (Roto, Obrist, & Väänänen-Vainio-Mattila, 2009) evaluated UX evaluation methods and distinguished five main evaluation method categories: lab studies, field studies, surveys, expert evaluation and mixed methods.

Lab studies are very applicable for evaluation during an early phase of a prototype. Participants get a task and carry them out with one or several UI's. A 'think out aloud' method is often used. The analyst observes the participant and aims to understand the mental models. This is similar to a usability test; but also paying attention to experiential aspects. Since UX is so context dependent, **field studies** are often useful and recommended because they are examining in real life situations. Field studies include either prototype test sessions in context or observing and interviewing participants in context. **Surveys** can provide feedback in short time-frame, and they are easy to get to a large and international scale. In early prototype phase it's common to have usability experts go over a design. Running **expert evaluations** before the user study can avoid ruining an expensive user study. It is however also challenging because UX has no set heuristics. A way to do an expert evaluation is to use perspective-based inspection in the evaluation to let experts focus on one specific experiential aspect (such as fun, aesthetics or comfort).

It's important to use several methods to collect richer data, therefore **mixed methods** are often used. Examples could be observations followed by interviews. Observations should usually be mixed with another data collection method as it is hard to see subjective feelings from a plain observation. This is also noted by Mao et al., (Mao, Vredenburg, Smith, & Carey, 2005): *"A note of caution when interpreting these findings, which are based on perceptions of user experience evaluators, rather than hard fact."*

Psycho-physiological measurements are a objective form of evaluating UX. Examples are measuring heart rate, skin perspiration or facial muscles. Especially facial muscles are a promising domain to measure positive or negative emotions (Ganglbauer, Schrammel, Deutsch, & Tscheligi, 2009). The great advantage of psycho-physiological measurements is that they allow the researcher to measure momentary experiences without intervening the user in the interaction. On the other hand, with the current technology it still requires quite invasive measurement equipment, influencing the experience of the user and making the research more expensive. Also, the momentary emotions are only important in some domains.

2.2.3 User experience models

The different views on UX as well as the broadness of the term lead to diversity in the evaluation of UX as a whole. Different evaluation methods focus on different aspects of UX: They range from analysis of psychological needs to task oriented user goals or guidelines (Alves et al., 2014). Contributing to this diversity is a gap between academia and practice, partly caused by a lack of uniform evaluation tools (Väänänen-Vainio-Mattila, Roto, & Hassenzahl, 2008), that are publicly available (Roto et al., 2009). In practice the UX evaluations are often still based on usability methods as the R&D departments traditionally focused on usability, whereas marketing departments were responsible to communicate a certain experience (Väänänen-Vainio-Mattila et al., 2008). With a shift from usability-focused to experience-oriented perspective on product interactions, a shift in evaluation methods should take place (Väänänen-Vainio-Mattila et al., 2008).

2.2.3.1 Hassenzahl's UX model

Hassenzahl and Tractinsky describe UX in essence as a consequence of a combination of the following three factors (Hassenzahl & Tractinsky, 2006) as depicted in Figure 25:

- The user's state and previous experiences (**user**)
- The system properties (**system**)
- The usage context and situation (**context**)

This idea is further developed and described by a whitepaper by Roto et al. (2011). The *user state* for example refers to the willingness of the user to use the product, the expectations the user has and previous experiences with the product. *The system* refers to the user's perception of the system such as aesthetics, functionality, usability, but for example also the user's image of the brand sustainability. The context refers to several contextual influences: social context (for example working with other people), physical context (for example using a product on a bumpy road versus on a desk), task context (the surrounding task that also require attention) and technical and information context (for example connection to network services, other products).

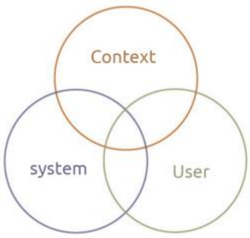

Figure 25 – The three factors of UX according to Hassenzahl

In modelling UX, Hassenzahl further distinguishes between pragmatic qualities and hedonic qualities (Hassenzahl, Platz, Burmester, & Lehner, 2000), highlighting that pragmatic qualities help users achieve hedonic goals. Hassenzahl (2003) describes the main product qualities belonging to either pragmatic or hedonic qualities as follows:

- **The pragmatic qualities** which are strongly related to the traditional usability measures such as learnability, efficiency and effectiveness.
    - **Manipulation** of the environment requires relevant functionality (utility) and ways to asses this functionality (usability).
- **The hedonic qualities** are the non-instrumental aspects appreciated by the user. Examples are product aspects that attract on a visual, behavioural or reflective level.

- **Stimulation** for the development of skills and the proliferation of knowledge. Trying to give the user insights and surprises, by for example given an unexpected hint which is still welcome.
- **Identification**, where the product has to be able to communicate identity, as individuals express their self through physical objects. An example of identification are personal homepages, which can be used to present the self to others. Often the possession is deliberately shaped to communicate advantageous identities to others.
- **Evocation**, where product can evoke memories. Also, previous experiences with products play a role here, as this experience influences the way an individual sees the product.

Hassezahl further makes a clear distinction between objective quality aspects and the user perception of these aspects. The perceived qualities are dependent on the objective qualities but also on the user characteristics and usage situation. These perceived qualities in the end lead to a judgement of appeal. Hassenzahl emphasizes that instead of providing a one size fits al UX model, UX should be approached as a subjective and dynamic concept, as the context plays such an important role (Hassenzahl, 2008b). Usability measures are seen as product features by this model. A schematic overview of the model is shown in Figure 26.

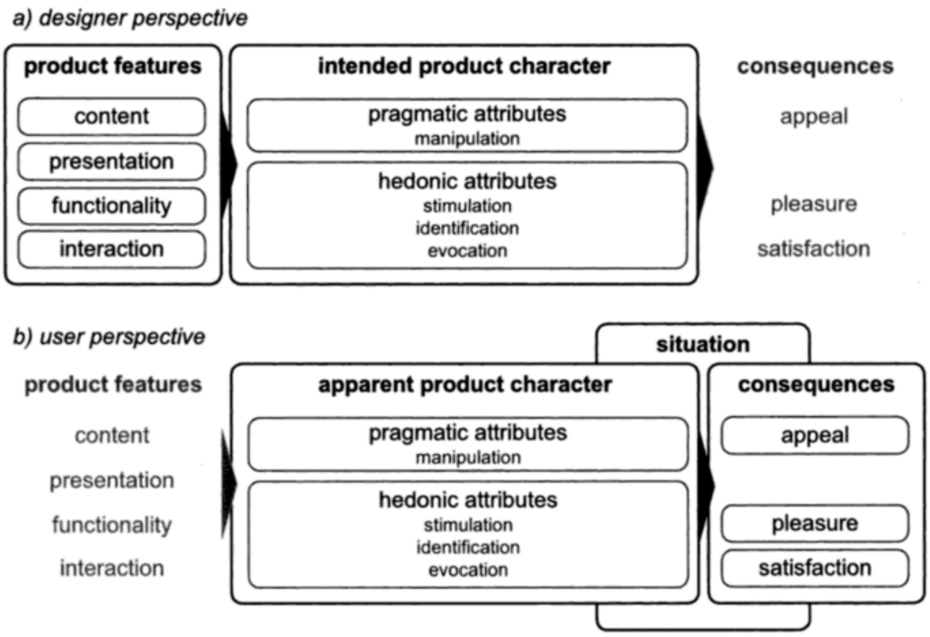

Figure 26 – Hassenzahl's UX model, from (Marc Hassenzahl, 2003).

To apply the theory of Hassenzahl in practice, the Attrakdiff questionnaire[4] was introduced. Based on measures of pragmatic and hedonic quality, the product can be placed in 2-dimensional space ranging from superfluous (low pragmatic and low hedonic quality) to desired (high pragmatic and high hedonic quality). In Figure 27 two systems A and B are plotted in the Attrakdiff space. The questions leading to this classification are based on 28 pairs of seven-step bipolar items (e.g. confusing – clear, good – bad). This evaluation method covers the pragmatic quality, hedonic quality (both identification (HQ-I) and stimulation (HQ-S)) and attractiveness. Hassenzahl & Monk (2010) have also created a shorter version of the Attrakdiff questionnaire, using just 10 word pairs.

[4] Attrakdiff questionnaire: http://attrakdiff.de/index-en.html

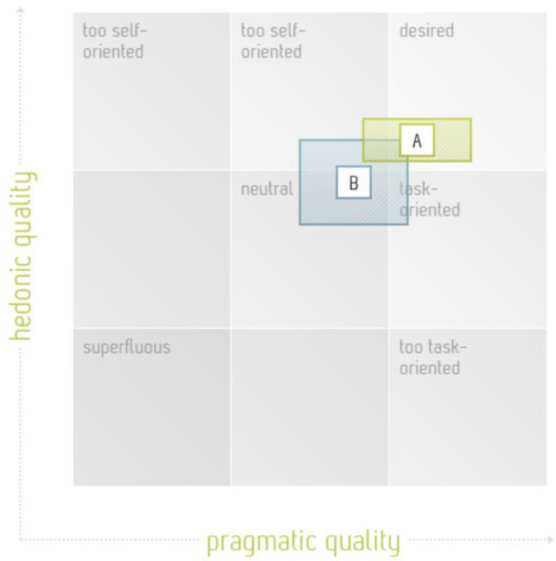

Figure 27 – 2d space of Attrakdiff, where two systems A and B are plotted in the space [5]

2.2.3.2 Components of user experience (CUE) Model

Mahlke & Thüring (2007) presented the Components of User Experience (CUE) model (Figure 28), which has a similar setup as Hassenzahl's model. The model defines instrumental and non-instrumental characteristics (which can be related to the pragmatic and hedonic qualities of Hassenzahl), which are likely to influence the emotional responses of the user, being the third central component of the UX. The interaction and its characteristics impact the UX, mainly depending on system properties, but also on the user characteristics and the context/task. Other research suggested that interaction characteristics can directly lead to emotions, and that emotions are not exclusively evoked by the product qualities (Aranyi & van Schaik, 2016).

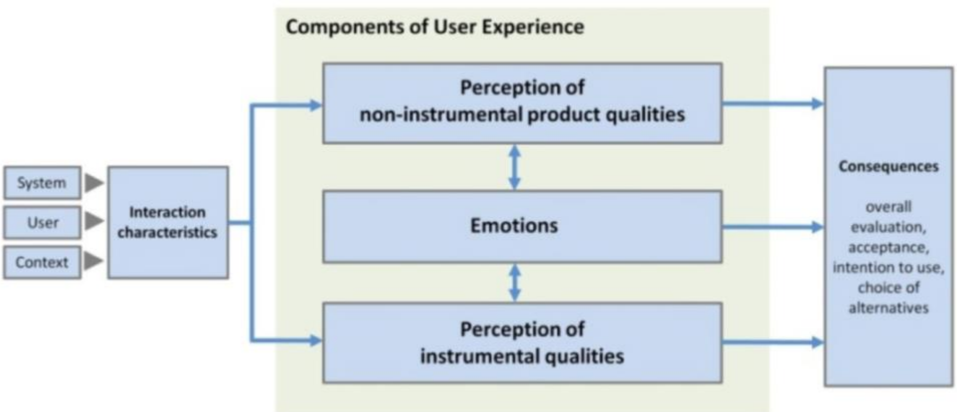

Figure 28 – CUE model, showing how UX consists of the perception of instrumental qualities, non-instrumental qualities and emotional reaction (Mahlke & Thüring, 2007).

[5] From http://attrakdiff.de/index-en.html

In a validation study of the CUE model, Mahlke (2008) showed that there was no direct link between instrumental and non-instrumental qualities. In this validation study they measured the usability (instrumental qualities) with a selection of the SUMI questionnaire (Kirakowski & Corbett, 1993) (controllability, effectiveness, helpfulness, learnability) and the aesthetics (the non-instrumental qualities) with the classical aesthetics questionnaire by Lavie & Tractinsky (2004), which is one of the most validated approaches to measure aesthetics for websites. This aesthetics questionnaire distinguishes classical aesthetics from expressive aesthetics. Classical aesthetics are for example orderly and clear design. The expressive aesthetics are about the creativity and originality of the design and its ability to break conventions. The expressive part of this aesthetics questionnaire is however criticized by Hassenzahl, arguing that it measures more symbolic or motivational aspects that are conveyed by visual attributes of an interactive product than directly focusing on aesthetic aspects (Hassenzahl, 2008a). The emotional user reaction was measured after each task with the SAM questionnaire (Bradley & Lang, 1994) and physiological measurements such as heart rate. Participants were also asked to rank the systems on preference. The physiological measurements were not very reliable and significant in this study.

From the CUE model the meCUE ('modular evaluation of user experience') questionnaire (Minge, Thüring, Wagner, & Kuhr, 2017) was developed in effort to create a standardized measurement of UX. For this evaluation three modules where constructed and validated separately. The first module consists of the product qualities (instrumental and non-instrumental), the second module consists of the emotion of the user, and the last module consists of the consequences that follow from the user experience. The modular nature of this evaluation model makes it possible to adjust the meCUE evaluation method to meet specific research goals by choosing the required modules (Minge & Thüring, 2018).

2.2.3.3 Honeycomb UX model

A well-known model for user experience is the UX Honeycomb Model (Morville, 2004). The idea of this model was to move beyond usability and help people to understand the need to define priorities within the design process. The value of the product to the user is central in this model and is caused by the surrounding 6 factors as shown in Figure 29; the balance of these factors depends on the unique balance of context, content and users. The UX honeycomb model is not often mentioned or used in scientific literature, but in practice the model is often used as it explains the concept in a simple way. The six factors are described as follows:

- Useful: The content should be original and fulfil a need
- Usable: The product must be easy to use, the learning curve should not be too steep
- Desirable: Image, identity, brand, and other design elements are used to evoke emotion and appreciation
- Findable: Content need to be navigable and locatable onsite and offsite
- Accessible: Content needs to be accessible to people with disabilities
- Credible: Users must trust and believe what they are told

Figure 29 - The seven factors of Morville's Honeycomb Model (Morville, 2004)

Efforts have been made to apply the honeycomb model to data visualization. Veeneklaas (2018) found that a visualization should indeed at least be usable, desirable, credible, aesthetically approved, technically adequate and consist of useful data in order for an information visualization to be successful on user experience.

2.2.3.4 Quantified UX model

In an effort to bridge the gap between theory and practice in the field of UX as described earlier, (Lachner et al. (2016) proposed a model for quantification of UX (QUX). In this model the scope was narrowed to product-oriented UX and they defined UX as the result of enjoyable interactions and/or anticipated interactions with a product. Lachner et al. created the model by analysing a sample of UX characteristics from literature found with a systematic search. Using theory and expert interviews, 285 UX characteristics were brought down to 9 UX dimensions. Lachner et al. also proposed a scoring tool where every aspect of the UX can be rated, resulting in a radar diagram with the 9 found UX dimensions as shown in Figure 30.

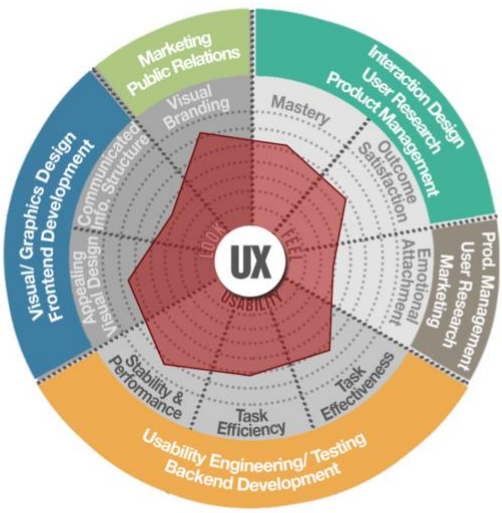

Figure 30 – QUX model, a quantifiable model for UX, created by analysing a sample of UX characteristics from literature found with a systematic search (Lachner et al., 2016)

2.2.4 Contextual influences

Most UX models focus on the perception of the system qualities, as these models try to guide designers to improve a design and the product quality. The contextual influences are however very important to understand as UX is a highly subjective, dynamic and context dependent domain. Research describes four main categories of influences of UX: system, users, context and temporal aspects (Ahsanullaha, Suziah Sulaiman, Ahmad K B Mahmood, & Muzafar Khan, 2015). The system influences are extensively described by the models above; this section will focus on the user, the context and the temporal aspects.

The user

As technology advanced the past years, the focus of HCI shifted from what computers could do to what people can do, illustrating the importance of understanding the user. Users can roughly be categorized in novice, experienced and expert users (Ahsanullaha et al., 2015). Whether a user is novice, experienced or expert is affected by the physical and cognitive abilities and disabilities of users, which are associated with user personality, (emotional) status, demography and functional and affective needs and goals (Ahsanullaha et al., 2015). These abilities are influenced by the previous experiences of the user with the product.

Context

The context generally refers to the physical environment or location where the system is used, and the conditions in this environment, for example if there is enough light, good internet, etc. It can however be seen from different perspectives such as the socio-cultural context (i.e. users self-image, attitude, values, life style, previous experience), the market context (i.e. product novelty, product comparison), the historic context (i.e. attachment, storytelling, memories), and use context (i.e. actions performed to achieve tasks and goals) (Ahsanullaha et al., 2015).

Temporal aspects

Time logically has a big effect on the UX. A system with a cool animation might attract attention and deliver a good first experience, but once a user uses the system very often, an animation could get distracting and too time consuming. Wilamowitz-Moellendorff et al. (2006) show how the importance of pragmatic quality (mainly usability) increases over time, whereas the importance of the hedonic quality decreases over time. Also Roto et al. (2011) describe how UX is affected by temporal aspects, i.e. how the UX evolves over time. They stress that people can have indirect experiences before their first encounter with a system based on expectations formed from related technology, brand, advertisements, presentations, demonstrations or others' opinions. Roto et al. call this anticipated UX (Figure 31). Similarly, users can have indirect experiences after use, for example through reflection on previous usage, or through changes in people's appraisals of use, called Episodic UX. The experience during use of a product is called momentary UX. Finally, Roto et al. distinguish cumulative UX, which is described as the views on a system as a whole, after having used it for a while. A focus on momentary experiences places different demands on design and evaluation that has a focus on use over longer time (Roto et al., 2011). So far, UX evaluation is mostly based on short term, momentary UX; measuring the UX of a first encounter with a system. Only a few published studies are focusing on long-term UX (Vermeeren et al., 2010). Research also notes that the principle of temporality is often overlooked and that positive initial experiences are not as crucial for motivating prolonged use (Karapanos, Zimmerman, Forlizzi, & Martens, 2009).

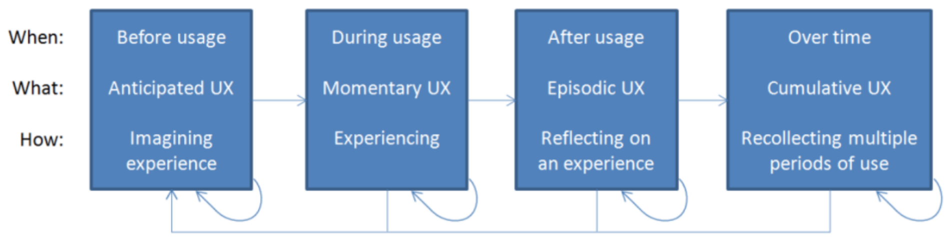

Figure 31 – Temporality of UX from (Roto et al., 2011), distinguishing four temporal phases that affect the UX; anticipated UX (before use), momentary UX (during use), Episodic UX (after use) and cumulative UX (after multiple periods of use).

2.2.5 Discussion

Most UX models incorporate usability or its measures such as efficiency, effectiveness and satisfaction directly as a factor of influence. Where usability is concerned with the ease of use and achieving the goals, UX is about the way users perceive their interaction. It is therefore no surprise that the user's perception of the product is strongly influenced by all usability aspects and that most UX models incorporate these aspects directly.

Some models see UX as a sum of particular factors on different interaction levels, like the QUX model (Lachner et al., 2016). The model has a total of 9 factors that are easily quantifiable, including product specific factors, contextual influences and even emotional user reactions such as satisfaction. The factors of this model were found with a systematic search in the literature; incorporating the factors that are often discussed in the literature and not being based on a solid UX theory or framework. The QUX factors match the factors described by the ISO partly, but deviate on usability aspects and context or user related aspects.

Other models describe UX by making a distinction between pragmatic or instrumental qualities and hedonic or non-instrumental qualities. Both Hassenzahl's UX model and the CUE model make this distinction, describing how these two qualities lead to emotional user responses. Morville's honeycomb model can be related to this distinction, where usefulness, usability, findability and credibility are instrumental attributes. Value and desirability on the other hand are non-instrumental qualities (Table 1). Desirability could however also be viewed as an emotional user reaction instead of a product quality. The three models emphasize how context has a big influence on the UX, but do not specifically describe contextual influencers, in contrast to some other models.

Table 1 – Hassenzahl UX model, Honeycomb UX model and CUE model divided in pragmatic and hedonic qualities.

	Hassenzahl's UX Model	CUE Model	Honeycomb UX Model
Hedonic / non-instrumental qualities	Stimulation Identification Evocation	Aesthetic qualities Symbolic qualities Motivational qualities	Value (result of all factors) Desirability
Pragmatic / instrumental qualities	Manipulation	Usability Usefulness	Usefulness Usability Findability Accessibility Credibility

Hassenzahl's UX model focusses on the hedonic qualities, whereas the honeycomb UX model focusses on pragmatic qualities. Hassenzahl does note that even though the perceived hedonic quality will in the end lead to the judgement of satisfaction and pleasure, the hedonics should not overshadow pragmatic aspects (Hassenzahl, 2003), which was also found to have negative consequences (Minge, 2008). The CUE model focusses both on instrumental and non-instrumental qualities and covers most of the factors mentioned by the other models with their broader descriptions. All models stress that the weight of the factors depends on the specific use case and context.

2.3 Evaluating data visualization on UX

Few (2017) has listed important aspects of data visualization in the data visualization effectiveness profile, which could be useful in assessing the aspects of the UX models that are applicable to information visualizations. Few (2017) divided the aspects in informative aspects and emotive aspects, which can be related to the pragmatic/instrumental and hedonic/non-instrumental subdivision made by several UX models.

Informative (product understanding)

- Usefulness: Satisfy the user's needs; the data should be useful.
- Completeness: right information and right amount, also context to understand
- Perceptibility: display information in a manner that the human can perceive with minimal effort and appropriate precision.
- Truthfulness: validity, accuracy and precision of the information visualization.
- Intuitiveness: an information visualization should be intuitive to the degree that it is familiar and easily understood. User dependent.

Emotive (produces useful emotional response)

- Aesthetics: The visualization should be pleasing to look at
- Engagement: The user must be invited to examine the information.

Approaching UX as a consequence of hedonic and pragmatic qualities, the balance of these qualities is dependent on the mode of use, suggested by Hassenzahl et al. (Hassenzahl et al., 2002). In the context of websites, they have shown that during "activity mode", where the activity is important and goals are defined on the fly, appeal was solely determined by hedonic qualities. In "goal mode" on the other hand, where the goals shape the activities, both hedonic and pragmatic qualities played a substantial role. As information visualizations are usually used in goal mode driven by an information need, it is especially important that in this context a UX model also clearly addresses usability measures.

In determining a suitable UX model for data visualization, it is useful to look at the differences between websites and information visualizations, as UX models are often based on websites. One main difference is that information visualizations are graphics based, whereas websites are usually based on text. In addition, websites usually display qualitative data (e.g. videos or photos), making the content more important than the way it is displayed. Information visualizations on the other hand display quantitative data, relying on the way it is displayed. It is therefore extra important that the UX model specifically incorporates visual aspects. Other UX models are clearly based on physical products and focus on status and identification. Identification and status are however not very applicable to data visualization as an information visualization is not something a user owns or tends to be identified with.

For this thesis, some UX models are thus more suitable than others. Hassenzahl's model and the measurement tools based on it make a useful distinction between pragmatic and hedonic system qualities but focus too much on the hedonic qualities for the context of information visualization, since the usability delivers an important share in the UX for information visualization. The honeycomb UX model on the other hand does not seem to cover all aspects in the hedonic domain, and the weights of the factors seem uneven, for example accessibility and usability having an even share in the model. The CUE model seems suitable for information visualization, as it addresses visual and aesthetic aspects but gives sufficient weight to the non-instrumental qualities. In addition, the model and its measurement tool leave room for interpretation and different applications (Minge & Thüring, 2018). Engagement, which is mentioned by Few as an important quality of data visualization, could for example be added as a non-instrumental quality of information visualization. Other constructs that are less applicable to data visualization, such as status, could be left out given the modular nature of the evaluation model.

3

Method

3 Method

3.1 Participants

Participants were recruited through **Prolific** (www.prolific.ac). This is an online platform that helps researchers to find participants for online experiments, using micro payments and exact research-participant matching using an extensive participant information database. To make sure that the sample represents adults that have most likely been in contact with (digital) graphs, respondents were chosen to be older than 18 and younger than 65. Participants were also restricted to be fluent at English, as they should be able to understand the English graphs and explanations. No further restrictions for participation were set, keeping the experiment as generalizable as possible. Via prolific, participants received an incentive of £2 for participation in the study with an estimated duration of 20 minutes. To stimulate participants to answer truthfully and not just click randomly, participants were awarded a bonus of £1 if all questions are answered correctly. First the experiment was conducted with 50 participants. After changing the UX questionnaire to measure the meCUE construct 'emotions' as described in 'measures', another 50 participants completed the experiment. Participants could only participate once, so the experiment had a total of 100 unique participants. The effect size was hypothesized to be small, since the differences between the conditions are small.

To ensure the quality of the online crowdsourced data, several criteria for inclusion of data were set. First, participants must have seen the animation. Therefore, they should have switched between the different views. Second, the number of correct answers for each task should be higher than 2, as the questions are simple questions that can be read directly from the graphs. This criterium is allowing a maximum of 1 out of 3 answers to be wrong. Finally, the completion time for the first UX questionnaire should be long enough, to filter out the participants that just randomly clicked and not actually answered the questions. The following thresholds for completion time are based on a study by Slattery & Yates (2018), in which they show how their fastest reading participant reads 246 words per minute, while still being accurate. In addition, participants are estimated to need at least half a second to click the answer on the 7-point Likert scale.

All excluded participants and accompanying reasons and calculations can be found in appendix IV. 4 participants were excluded due to incomplete responses, 23 participants were excluded due to the criteria below. In total 27% of the participants was excluded (27 out of 100). A total of n = 73 participants remained (n = 38 for product qualities iteration, n = 35 for emotions iteration, n = 73 for the general questionnaires). The **gender** of the participants for both iterations is shown in Figure 32.

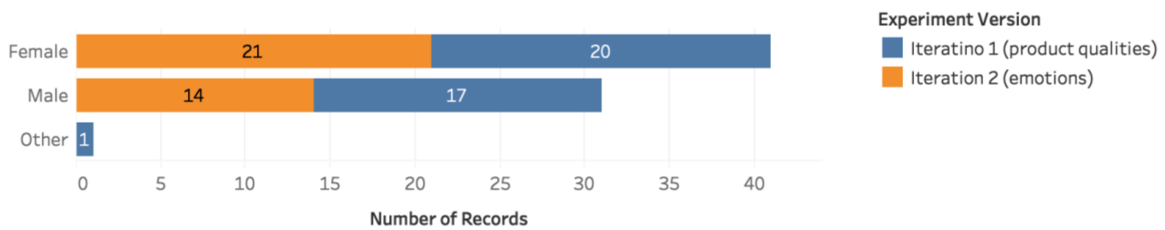

Figure 32 – The gender of the participants in the experiment

The characteristics of **participant age** in iteration 1 are M = 32, SD = 9.39. The characteristics of participant age in iteration 2 are M = 29, SD = 8.05. The characteristics of participant age in general (iteration 1 and 2 together): M = 30, SD = 8.84, also see Figure 33.

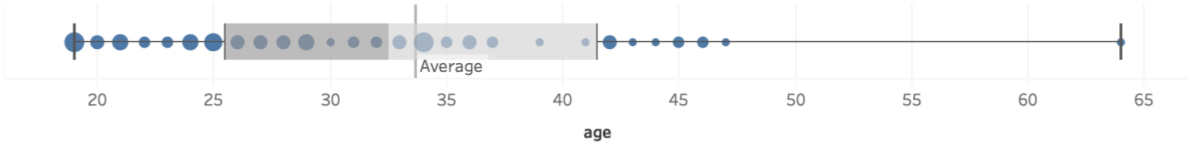

Figure 33 - Boxplot of the participant age with the average

Participants from all over the world participated in the experiment. Participants had the following **nationalities**: Belgium (1), Finland (1), India (1), Latvia (1), Slovenia (1), Canada (2), Hungary (2), Ireland (2), Mexico (2), Australia (3), Italy (4), France (5), Spain (5), United States (5), Poland (6), Portugal (6) United Kingdom (25).

3.2 Measures

To measure the UX of the interaction between user and graph in a quantitative manner, the **meCUE** questionnaire (Minge et al., 2017), based on the CUE model (Thüring & Mahlke, 2007), was used.

For the first iteration of the experiment, the constructs emotions, status, commitment, intention to use and product loyalty were left out as they are less applicable to data visualization. 'Status' is not applicable to data visualization as an information visualization is not something a user owns or tends to be identified with. 'Commitment', 'intention to use' and 'product loyalty' are less applicable as the interaction in this experiment is goal based, making these constructs rather rely on the goal and the experiment itself. The constructs of 'emotions' (both positive and negative) were left out as the differences in the conditions were assumed to be too small to have a measurable effect on emotions.

In the first iteration, the UX is measured in a quantitative way using a selection of the meCUE questionnaire (Minge et al., 2017) and qualitative data is recorded with an open question. Then the conditions are compared, and a preferred condition is requested. Participant characteristics are collected via prolific, and objective data about the experiment is collected. A more detailed explanation of the procedure can be found in chapter 3.5. In this iteration the following variables are measured (see appendix III for the full questionnaires):

Quantitative UX questions first iteration (Minge et al., 2017)

Construct	Source	Questions	Input
Usefulness	meCUE module I	3 questions	7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree)
Usability	meCUE module I	3 questions	
Visual aesthetics	meCUE module II	3 questions	
Overall Evaluation	meCUE module V	1 question	10-point scale from -5 (bad) to 5 (good)

Qualitative question (asked before comparing the conditions)

Question	Input
Remarks about the visualization (optional)	Text Field

Comparison questions (after comparing and explaining the conditions)

Questions	Input
Questions on the added value of the loading animation (to the beauty and the understanding)	7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree)
Questions on the added value of the transition animation (to the beauty and the understanding)	
Preferred visualization (transition animation)	1 of the three conditions
Preferred visualization (loading animation)	1 of the three conditions

Participant characteristics and experiment data

Item
Age
Gender
Education level
Nationality
Employment status
Total time taken for the experiment
Date experiment
Answer speed per questionnaire
Number of clicks (switches between views) per condition
Correct answers per condition

In a second iteration the same experiment was repeated, but instead of measuring the instrumental and non-instrumental product qualities of the visualizations, the negative and positive emotions were measured with module III of the meCUE questionnaire. In first place the emotion constructs of the meCUE questionnaire were left out as the conditions were assumed to have no effect on it and to reduce the length of the questionnaire. From the feedback of the first iteration it however seemed that some conditions are actually evoking emotions. It was therefore chosen to measure emotions in a second iteration of the experiment. In this iteration the quantitative UX questions part is replaced with the variables listed below (see appendix III for the full questionnaires), the rest of the measures were kept the same.

Quantitative UX questions second iteration (Minge et al., 2017)

Construct	Source	Questions	Input
Positive Emotions	*meCUE module III*	6 questions	7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree)
Negative Emotions	*meCUE module III*	6 questions	
Overall Evaluation	*meCUE module V*	1 question	10-point scale from -5 (bad) to 5 (good)

3.3 Stimuli

There were two graphs with three conditions each. **Gorilla** Experiment Builder (www.gorilla.sc) was used to create and host the online experiment (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2019). Data was collected between 20-11-2018 and 06-12-2018. For the creation of the graphs, Javascript was used with the libraries **D3** (www.d3js.org), Highchars (www.highcharts.com) and Amcharts (www.amcharts.com).

Part 1: In this part loading animations are added purely for aesthetic reasons, not increasing the efficiency or effectiveness. In this case an animation will be used that is executed when loading a graph (a 'loading animation'). These animations have no added use other than aesthetics. One could even argue that the animations decrease efficiency as the animations take approximately 1 second to execute, as there is no actual loading time. From a usability point of view these animations could harm the visualization's efficiency. For consistency between the conditions, dates were chosen for the x-axis and similar numerical values on the y-axis. The full datasets behind the graphs can be found in appendix I. All datasets are about some drug and a number of people using it in different years. The three conditions of **part 1** are as follows:

- **Condition 1a** – **Static** - No loading animation, directly displaying the graphs.
  Example: http://graphs.tvanwilligen.nl/bar_static.html

- **Condition 1b** – **Calm loading Animation**: Two graphs with an animation on load, using a standard loading animation from Highcharts D3 library (www.highcharts.com), which gradually enlarges the bars until they reach their value. Every time the user switches between graphs in this condition, the new graph is loaded again with the same animation. *Duration of the animation: 1s*.
  Example: http://graphs.tvanwilligen.nl/bar_sine.html

- **Condition 1c** – **Bouncy loading animation**: Two graphs with an animation on load, using the "elastic" animation of Amcharts D3 library (www.amcharts.com), which enlarges the bars and elastically bounces around their value before reaching the static point of their value. Every time the user switches form graph in this condition, the new graph is loaded again with the same animation. *Duration of the animation: 1.7s*.
  Example: http://graphs.tvanwilligen.nl/bar_elas.html

Part 2: In this part, transition animations were added, aiming to give the user a better understanding of the relation between the graphs. These animations focus mainly on increasing the efficiency and effectiveness of the graph, not specifically aiming on increasing the aesthetic quality. An animation for the transition between a 'stacked bar graph' and a 'grouped bar graph' will be used, giving the participant a better understanding of the relation between the who charts *(Heer & Robertson, 2007)*, which is useful in completing the tasks accompanying the visualizations. The full datasets behind the graphs can be found in appendix I. The datasets were all three about the consumption of a group of categories per country; meat *(beef, pork, poultry, sheep)*, alcohol *(beer, wine, spirits, other)* and energy *(thermal, nuclear, hydro, other)*. The three conditions of **part 2** are as follows:

- **Condition 2a** – **Static transition** - a non-animated direct transition from one view to another.
  Example: http://graphs.tvanwilligen.nl/transition_static.html

- **Condition 2b** – **Direct transition animation**: A transition from one chart to another on user request (button press), gradually moving from one chart into another using a direct animation, directly interpolating between start and end state. A representation of such a transition (Heer & Robertson, 2007) can be seen in the top row of Figure 34. *Duration of the animation: 0.5s*
  Example: http://graphs.tvanwilligen.nl/transition_direct.html

- **Condition 2c** – **Staged transition animation**: A transition from one chart to another on user request (button press), moving from one chart to another using a staged animation, using multiple animation stages (Heer & Robertson, 2007). The first stage changes the widths and x-coordinates of the bars, the second stage drops the bars down to the baseline. A representation of this transition (Heer & Robertson, 2007) can be seen in the bottom row of Figure 34. *Duration of the animation: 1s.*
  Example: http://graphs.tvanwilligen.nl/transition_staged.html

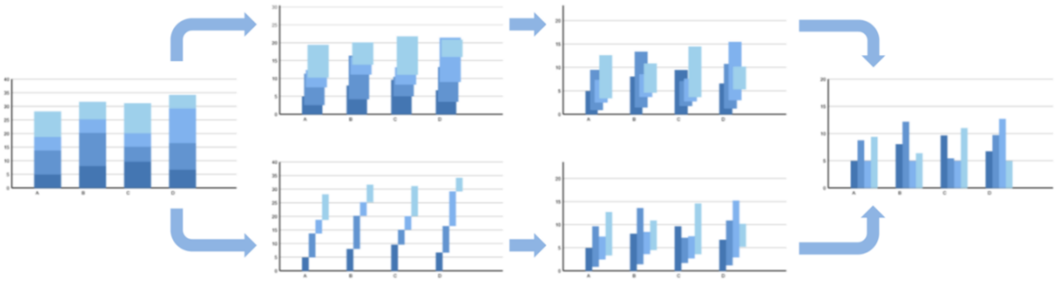

Figure 34 - Examples of direct and staged animations by (Heer & Robertson, 2007). The top path shows a direct animation, the bottom path a staged animation where the first stage changes the widths and x-coordinates of the bars, the second stage drops the bars down to the baseline.

3.4 Task

To get the participant to interact with the graph, they had to complete tasks. The tasks consisted of three multiple choice questions per visualization. These questions and corresponding answers for every condition can be found in appendix II. They were based on the following set of low-level analysis tasks that Amar et al. identified as largely capturing people's activities with visualizations (Amar, Eagan, & Stasko, 2005):

Part 1 (the graphs with loading animation):

1. *Retrieve value*
2. *Filter operation*
3. *Find extremum*

Part 2 (the graphs with transition animation):

1. *Find extremum*
2. *Retrieve 2 values and compare*
3. *Retrieve value*

3.5 Procedure

The study consists of two parts, each part having three conditions. One condition for each part has no animation, the other two conditions have an animation with a subtle difference between them. The participants assessed all six visualizations in a within-subjects research on their experience. A within-subjects design was chosen because it has greater power and less variability, and are considered by some to be the best design for subjective judgements (Lambdin & Shaffer, 2009). Latin square randomization was applied within part 1 and part 2, to overcome within-subject research problems such as attitude formation, fatigue, learning and carryover effects. Figure 35 shows a schematic representation of the research procedure.

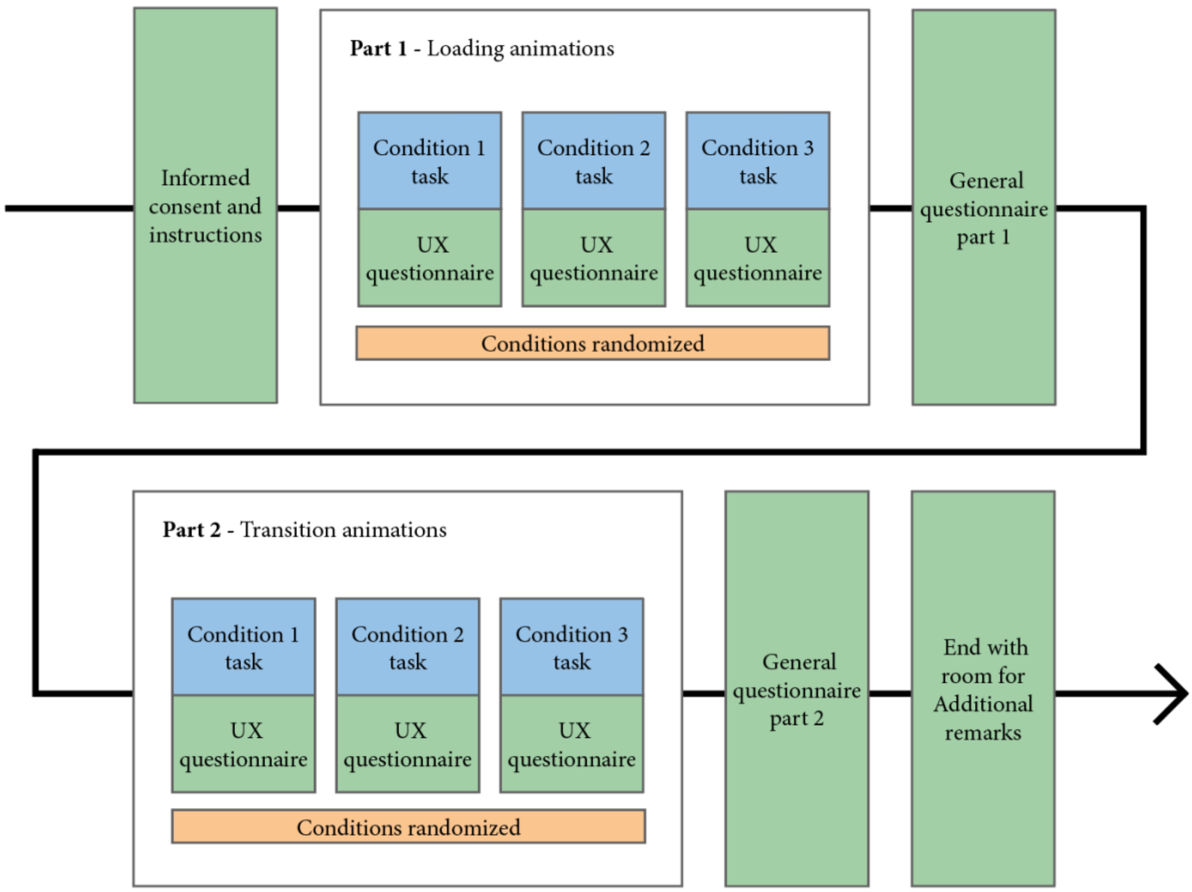

Figure 35 - Schematic representation of the research procedure

Participants were first recruited through **Prolific** (www.prolific.ac), and then linked through to the experiment on Gorilla experiment builder (Anwyl-Irvine et al., 2019). Before the experiment started the participant had to agree with the informed consent (see appendix V). Before every task, the participants were shown information telling them what they would see, how they could switch between the two views and that they should note the transition. After clicking next the visualization was shown, and the participant was asked to complete a task using the visualization. This task consisted of answering three questions about the content of the visualization, giving the participant a task-based interaction with the graph. One of the questions for each visualization in part 2 was chosen in a manner that it is useful to understand the relation between the two charts, making the animation actually useful. The questions about the data were asked in an order that required the participant to switch between views for every other question. An example of a task screen is shown in Figure 36.

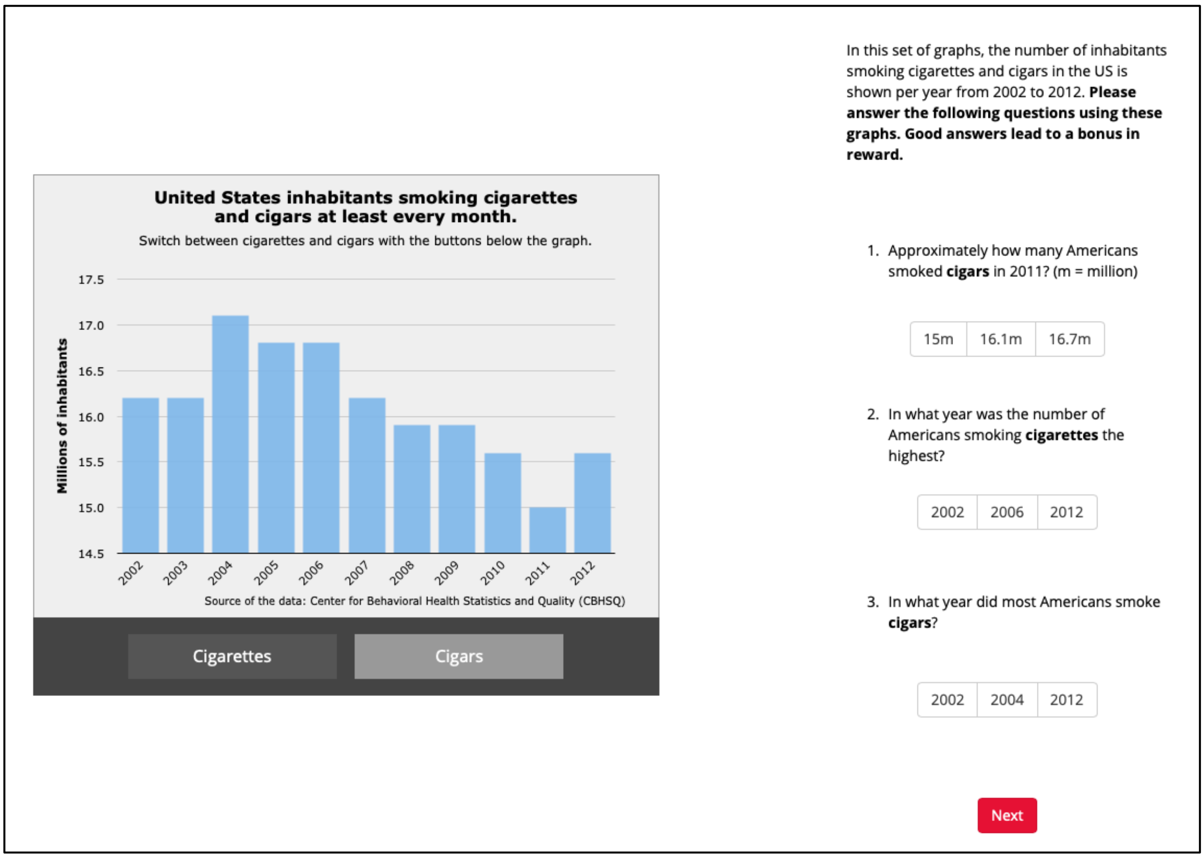

Figure 36 - An example of a task screen in the experiment, with the graph on the left and the task questions on the right

Straight after each visualization, the respondent was asked to answer questions about his/her experience with the visualization. After completing all three visualizations and the corresponding three UX questionnaires of part 1, the participant was asked to pick a preferred visualization, indicate if the animation added to the beauty of the graph and/or helped in understanding the graph. Then the participant could continue to the second part, repeating the process. After the second part the participant has room to leave additional remarks before the experiment is finished and the participant is sent back to Prolific. The response is then reviewed and if the response is complete the participant is paid the reward via prolific.

The mean time participants needed to complete the experiment was 953s = 15.8m (Figure 37). In this calculation and graph one outlier was left out, which was a participant that had spent 178767s (= 50 hours) on the experiment. This participant did not complete the experiment in one go.

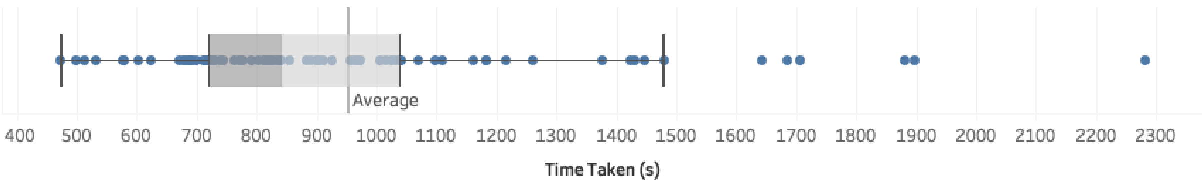

Figure 37 - The time participants needed to complete the experiment

3.6 Data analysis

The data that was collected during the experiment partly consists of participant information from Prolific, which can be linked to the experiments. For each participant, general data like the randomization groups they belonged to and which visualization from both parts they preferred. For each participant task-specific data for all six tasks was also collected, containing the objective interaction characteristics and self-reported UX Likert scores. A schematic representation of this data is shown in Figure 38.

Figure 38 - The data that was collected during the experiment, and how different components are linked

To analyse the **meCUE data**, the mean scale scores of the different conditions were compared using a repeated measures ANOVA. The Bonferroni-corrected *p*-values are reported. The SPSS code is added in appendix VI.

The **qualitative responses** in the additional remarks section were analysed in more depth, to see what remarks participants have about specific visualizations, and how these remarks differ per condition. The answers are blindly rated on a scale from -2 (negative) to 2 (positive). The ratings for all responses can be found in appendix VII. The definitions are as follows:

- **-2** - Clearly negative remarks such as ‘irritating’, ‘terrible’, ‘stressful’, etc.
- **-1** - Somewhat negative description such as ‘somewhat off-putting’, ‘I wasn’t a big fan of’, etc.
- **0** - When discussing matter not related to the UX of the visualization, when both positive and negative remarks cancel each other out (e.g. ‘boring but effective’) and neutral remarks.
- **1** - Somewhat positive such as ‘easy to understand’, ‘nice’, etc.
- **2** - Clearly positive remarks such as ‘love it’, ‘I really appreciated the animation’, etc.

The **comparison questionnaire data**, comparing the animations after explaining the differences between the conditions, was analysed using paired samples t-tests, as the two different animations were compared here rather than three conditions where one condition had no animation.

The **preferred graphs** for both parts were assessed, to see if this preference can be explained by the differences in UX measured with the meCUE questionnaire.

4

Results

4 Results

4.1 Results meCUE questionnaire

The results of the Likert scales (from strongly disagree to strongly agree) are shown using divergent stacked bar graphs (Wexler, Shaffer, & Cotgreave, 2017), an insightful way to visualize Likert scales.

4.1.1 Results meCUE iteration 1 - product qualities (*n = 38*)

In the first iteration, the product qualities aesthetics, functionality and usability were measured. Table 2 shows the descriptive statistics of the answers to Likert scale questions on product qualities of the meCUE questionnaire.

Table 2 - Descriptive statistics of the product quality constructs of meCUE

			Mean	SD	N
Loading animations	Aesthetics	Bouncy Loading	5.28	1.45	19
		Calm Loading	5.37	1.35	19
		Static Loading	5.31	1.48	19
	Functionality	Bouncy Loading	6.46	0.98	19
		Calm Loading	6.65	0.84	19
		Static Loading	6.64	0.86	19
	Usability	Bouncy Loading	6.62	0.86	19
		Calm Loading	6.75	0.61	19
		Static Loading	6.70	0.61	19
Transition animations	Aesthetics	Direct Transition	5.38	1.59	19
		Staged Transition	5.35	1.76	19
		Static Transition	4.96	1.74	19
	Functionality	Direct Transition	5.83	1.50	19
		Staged Transition	5.56	1.60	19
		Static Transition	5.78	1.53	19
	Usability	Direct Transition	5.56	1.47	19
		Staged Transition	5.51	1.51	19
		Static Transition	5.62	1.45	19

4.1.1.1 Loading animations

For the conditions in iteration 1, no significant differences in aesthetics, usability and functionality were found.

Aesthetics

The Huynh-Feldt estimate of departure from sphericity was ε = 0.882. According to the meCUE results, the construct aesthetics of the UX **was not significantly affected** by the type of loading animation, $F(1.764, 62.280) = .229$, $p = .768$. These results are shown in Figure 39.

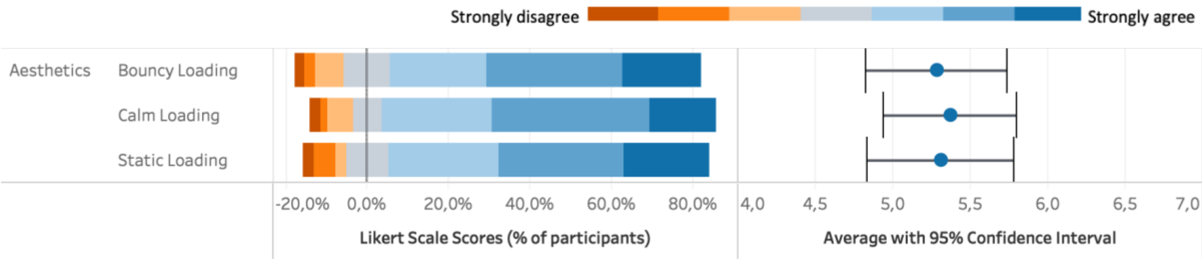

*Figure 39 – Likert Scale results from the **aesthetics** construct of meCUE of the loading animations, with left the scores from strongly disagree (red) to strongly agree (blue), and right the average scores and their 95% Confidence Intervals.*

Functionality

The Huynh-Feldt estimate of departure from sphericity was ε = 0.874. According to the meCUE results, the construct functionality of the UX **was not significantly affected** by the type of loading animation, $F(1.748, 64.665) = 2.752$, $p = .078$. These results are shown in Figure 40.

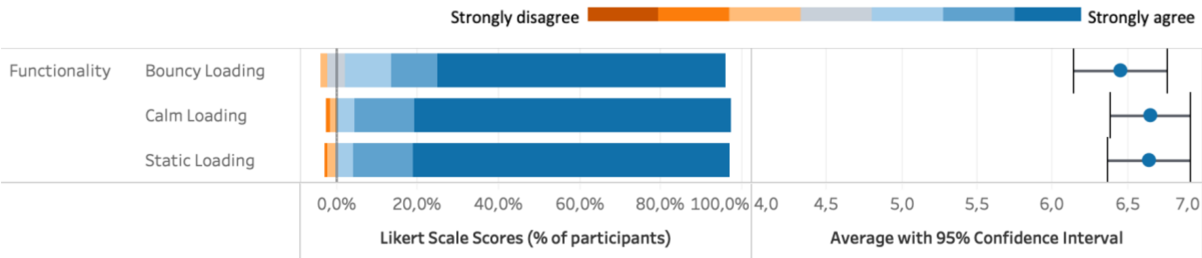

*Figure 40 – Likert Scale results from the **functionality** construct of meCUE of the loading animations, with left the scores from strongly disagree (red) to strongly agree (blue), and right the average scores and their 95% Confidence Intervals.*

Usability

The Huynh-Feldt estimate of departure from sphericity was ε = 0.853. According to the meCUE results, the construct usability of the UX **was not significantly affected** by the type of loading animation, $F(1.706, 63.110) = 1.934$, $p = .159$. These results are shown in Figure 41.

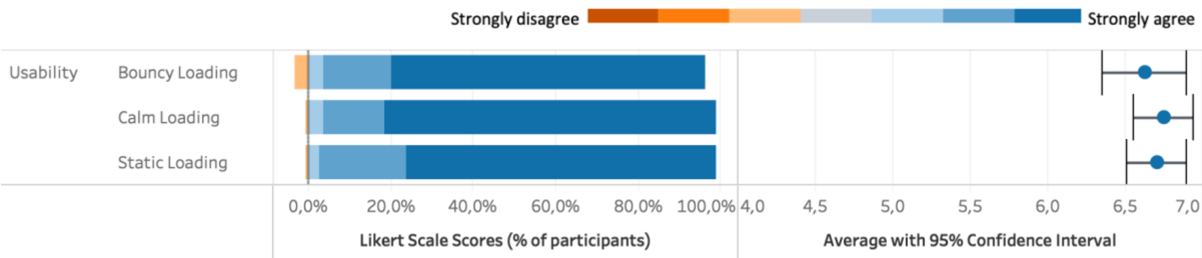

*Figure 41 – Likert Scale results from the **usability** construct of meCUE of the loading animations, with left the scores from strongly disagree (red) to strongly agree (blue), and right the average scores and their 95% Confidence Intervals.*

4.1.1.2 *Transition animations*

Aesthetics

The Huynh-Feldt estimate of departure from sphericity was $\varepsilon = 0.795$. According to the meCUE results, the construct aesthetics of the UX **was not significantly affected** by the type of transition animation, $F(1.590, 58.846) = 2.175$, $p = .133$. These results are shown in Figure 42.

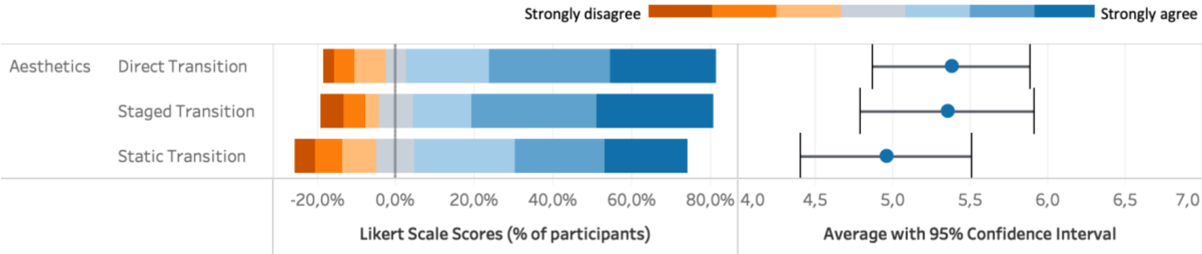

*Figure 42 – Likert Scale results from the **aesthetics** construct of meCUE of the transition animations, with left the scores from strongly disagree (red) to strongly agree (blue), and right the average scores and their 95% Confidence Intervals.*

Functionality

The Huynh-Feldt estimate of departure from sphericity was $\varepsilon = 1$, so sphericity can be assumed. According to the meCUE results, the construct functionality of the UX **was not significantly affected** by the type of transition animation, $F(1.581, 47.604) = 1.229$, $p = .299$. These results are shown in Figure 43.

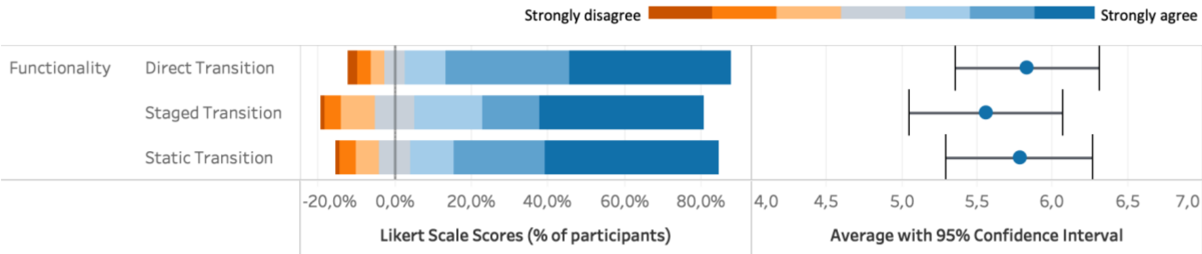

*Figure 43 – Likert Scale results from the **functionality** construct of meCUE of the transition animations, with left the scores from strongly disagree (red) to strongly agree (blue), and right the average scores and their 95% Confidence Intervals.*

Usability

The Huynh-Feldt estimate of departure from sphericity was $\varepsilon = 0.973$. According to the meCUE results, the construct usability of the UX **was not significantly affected** by the type of transition animation, $F(1.946, 43.975) = .208$, $p = .807$. These results are shown in Figure 44.

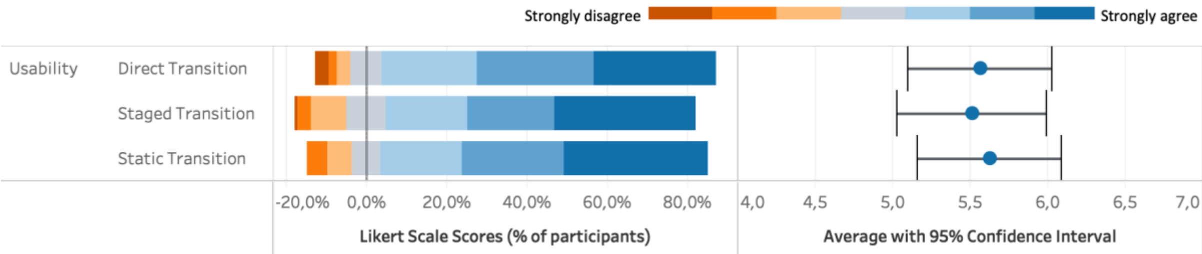

*Figure 44 – Likert Scale results from the **usability** construct of meCUE of the transition animations, with left the scores from strongly disagree (red) to strongly agree (blue), and right the average scores and their 95% Confidence Intervals.*

4.1.2 Results meCUE iteration 2 - emotions (n = 35)

In the second iteration, the positive and negative emotions were measured. In general, much more people disagree with all statements about emotions. Table 3 shows the descriptive statistics of the answers to Likert scale questions on product qualities of the meCUE questionnaire.

Table 3 - Descriptive statistics of the emotion constructs of meCUE

			Mean	SD	N
Loading animations	Negative Emotions	Bouncy Loading	2.60	1.94	35
		Calm Loading	2.29	1.68	35
		Static Loading	2.13	1.52	35
	Positive Emotions	Bouncy Loading	2.71	1.60	35
		Calm Loading	2.92	1.56	35
		Static Loading	2.88	1.56	35
Transition animations	Negative Emotions	Direct Transition	2.65	1.52	35
		Staged Transition	2.80	1.79	35
		Static Transition	2.71	1.71	35
	Positive Emotions	Direct Transition	2.81	1.59	35
		Staged Transition	2.70	1.59	35
		Static Transition	2.54	1.42	35

4.1.2.1 Loading animations

The results of the Likert scales from the constructs positive emotions and negative emotions of the meCUE questionnaire and their averages and 95% confidence intervals are shown in Figure 45.

Negative emotions

The Greenhouse-Geisser estimate of departure from sphericity was ε = 0.686. According to the meCUE results, the construct negative emotions of the UX **was not significantly affected** by the type of transition animation, F(1.372, 46.078) = 2.976, p = .079.

Positive emotions

The Huynh-Feldt estimate of departure from sphericity was ε = 0.936. According to the meCUE results, the construct positive emotions of the UX **was not significantly affected** by the type of loading animation, F(1.872, 39.994) = .729, p = .478.

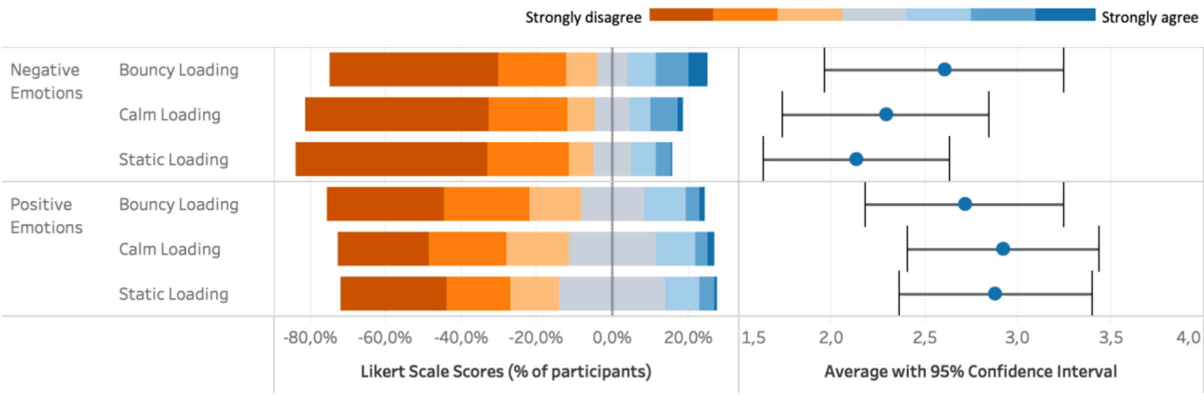

Figure 45 – Likert Scale results from the ***Positive emotions*** *and* ***Negative emotions*** *constructs of meCUE of the loading animations, with left the scores from strongly disagree (red) to strongly agree (blue), and right the average scores and their 95% Confidence Intervals.*

4.1.2.2 Transition animations

The results of the Likert scales from the constructs positive emotions and negative emotions of the meCUE questionnaire and their averages and 95% confidence intervals are shown in Figure 46.

Negative emotions

The Greenhouse-Geisser estimate of departure from sphericity was ε = 0.952. According to the meCUE results, the construct negative emotions of the UX **was not significantly affected** by the type of transition animation, $F(1.372, 64.761) = .377$, $p = .677$.

Positive emotions

The Huynh-Feldt estimate of departure from sphericity was ε = 0.859. According to the meCUE results, the construct positive emotions of the UX **was not significantly affected** by the type of loading animation, $F(1.717, 58.379) = 1.395$, $p = .255$.

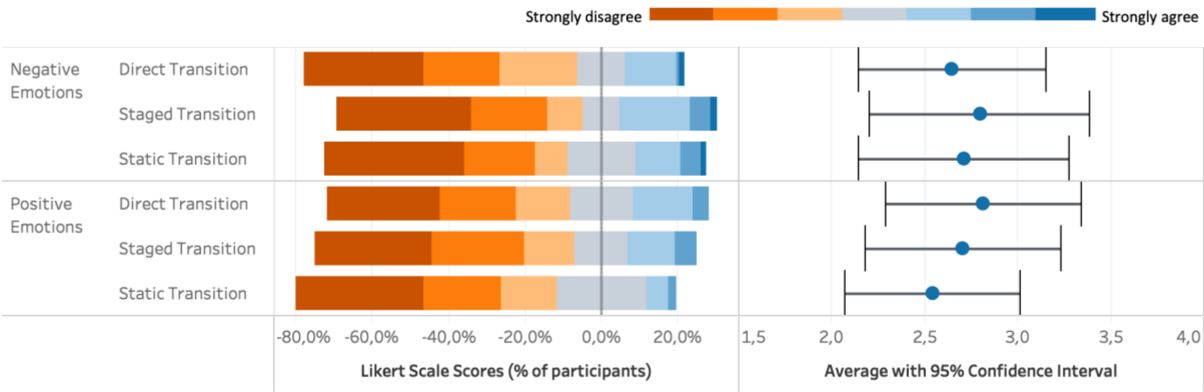

Figure 46 – Likert Scale results from the ***Positive emotions*** *and* ***Negative emotions*** *constructs of meCUE of the transition animations, with left the scores from strongly disagree (red) to strongly agree (blue), and right the average scores and their 95% Confidence Intervals.*

4.1.3 Overall UX (n = 73)

The descriptive statistics of the overall UX, measured by the meCUE questionnaire, are shown in Table 4.

Table 4 – descriptive statistics of the overall UX measured by the meCUE questionnaire

Task Name	Mean	SD	N
Bouncy loading animation	2.75	2.79	73
Calm loading animation	3.38	2.08	73
Non-animated	3.38	1.65	73
Direct transition animation	2.62	2.26	73
Non-animated transition	2.62	2.13	73
Staged transition animati..	2.55	2.22	73

Loading animations

The Huynh-Feldt estimate of departure from sphericity was ε = 0.927. According to the meCUE results, the measure of overall UX **was significantly affected** by the type of transition animation, $F(1.854, 133.458)$ = 4.294, p = .018. Post hoc tests using the Bonferroni correction revealed that the mean of the bouncy loading condition was lower than both other means. It was **significantly lower** than the calm loading condition (p = .045) but **not significantly lower** than the static loading condition (p = .081). The average overall UX scores and their 95% confidence intervals for the loading animations are shown in Figure 47.

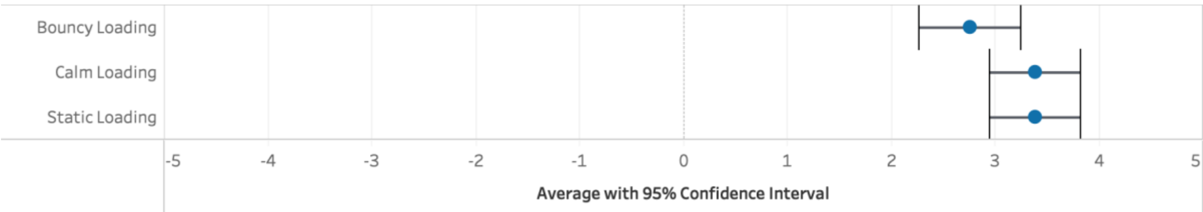

Figure 47 – Averages with 95% confidence interval of the ratings on the overall UX of the loading animations, rated from -5 (bad) to 5 (good)

Transition animations

The Huynh-Feldt estimate of departure from sphericity was ε = 0.925. According to the meCUE results, the measure of overall UX was not significantly affected by the type of transition animation, $F(1.849, 133.155)$ = .066, p = .925.

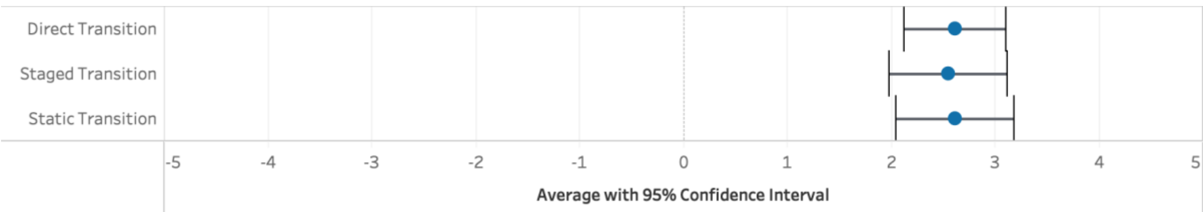

Figure 48 – Averages with 95% confidence interval of the ratings on the overall UX of the loading animations, rated from -5 (bad) to 5 (good)

4.2 Results comparison

4.2.1 Comparison questionnaire (*n* = 73)

After explaining and showing the difference between the animations, participants answered questions about the animations. Comparing the two loading animations (bouncy and calm), the calm animation scored way higher both on added value in beauty and added value in understanding the graph. The results are shown in Figure 49.

A paired samples t-test was conducted to compare the means of the Likert scale scores for both animations on the questions if the animation added to the understanding of the graph. There was a **significant difference** in the scores for the bouncy animations (M=2.85, SD=1.75) and the calm animation (M=4.30, SD=2.08); t(80)=-5.81, p=0.00. Same was done for the Likert scale scores for both animations on the questions if the animations added to the beauty of the graph. There was a **significant difference** in the scores of the bouncy animation (M=3.62, SD=2.79) and the calm animation (M=5.21, SD=1.60); t(80)=-6.34, p=0.00.

Interestingly, on average people tended to agree (M=4.30) that the calm animation added to the understanding of the graph, even though it was just an animation with aesthetic purpose.

In general, all the animations added more to the aesthetics than to the understanding of the graphs.

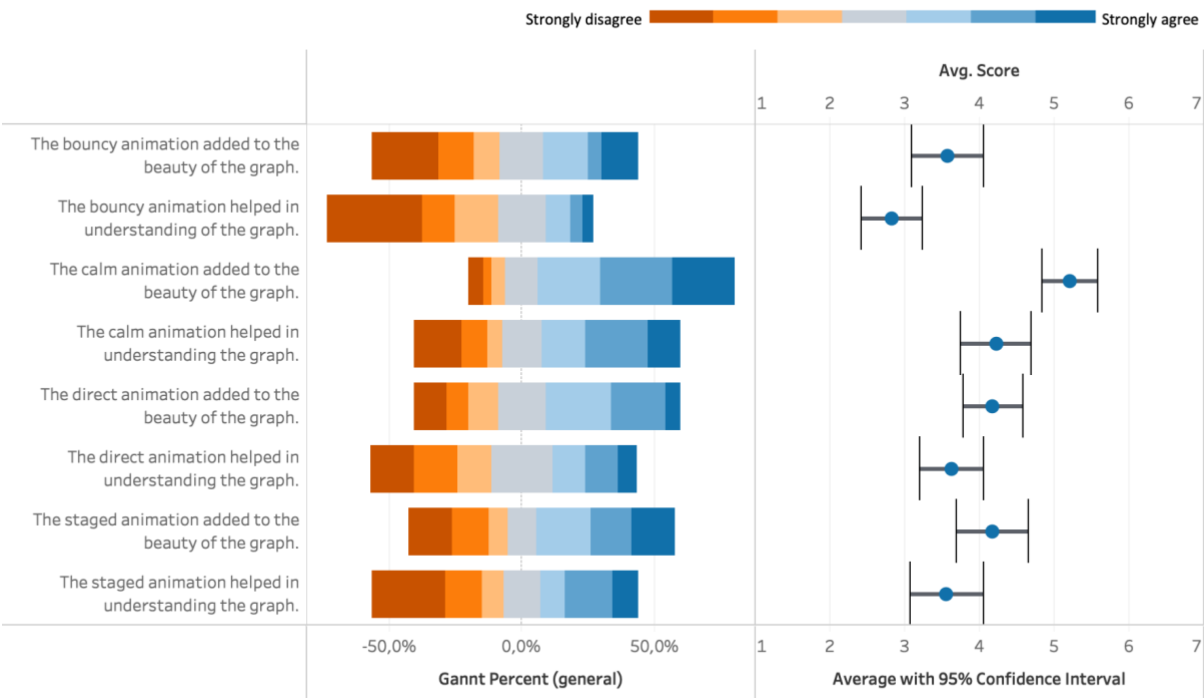

Figure 49 – Answers to the Likert scales of the comparison questions for all animations

4.2.2 Preferred graphs (*n* = 73)

For both parts of the experiments, participants were asked which condition of the graphs they preferred. For the loading animations, there is a big preference for the calm animation (*n* = 47) as opposed to the bouncy (*n* = 13) and non-animated (*n* = 13) loading of the graph. For the transition animations, there is a relatively small preference for both animated graphs (*n* = 27 and *n* = 28) as opposed to the non-animated graph (*n* = 23). These results are shown in Figure 50.

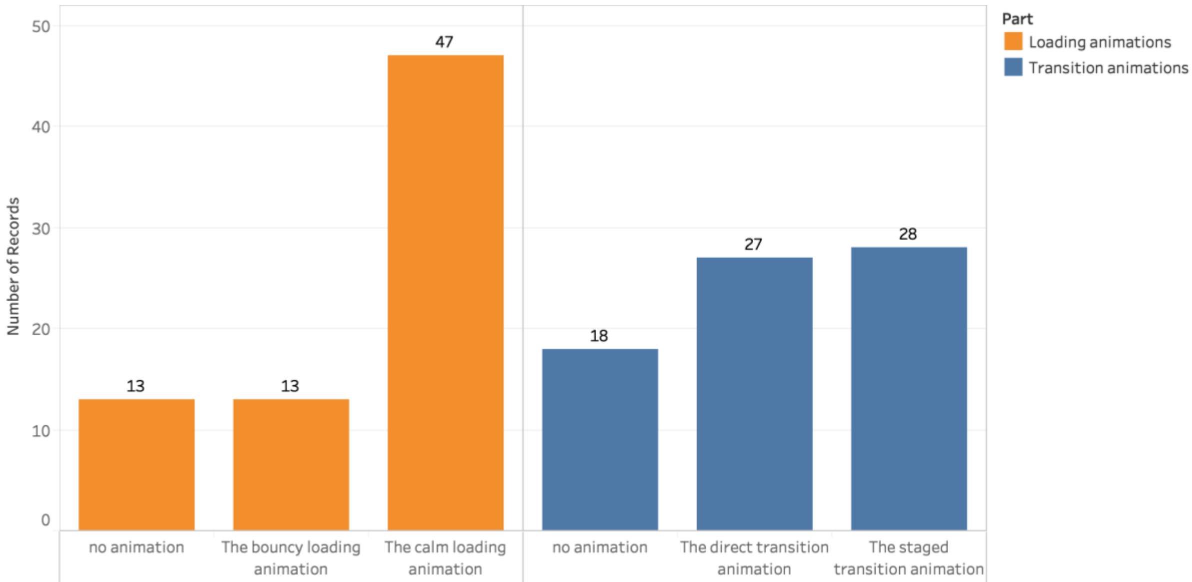

Figure 50 - Preferred conditions with right the loading animation conditions and left the transition animation conditions

Looking at the age of the participants (Figure 51), it can be observed that the average age of the participants preferring animated conditions is lower that the participants preferring the non-animated conditions. This seems to be influenced by an outlier, but when plotting the median and the 95% CI, it can be observed that still more younger participants prefer the animated conditions, and more older participants prefer the non-animated conditions (Figure 52).

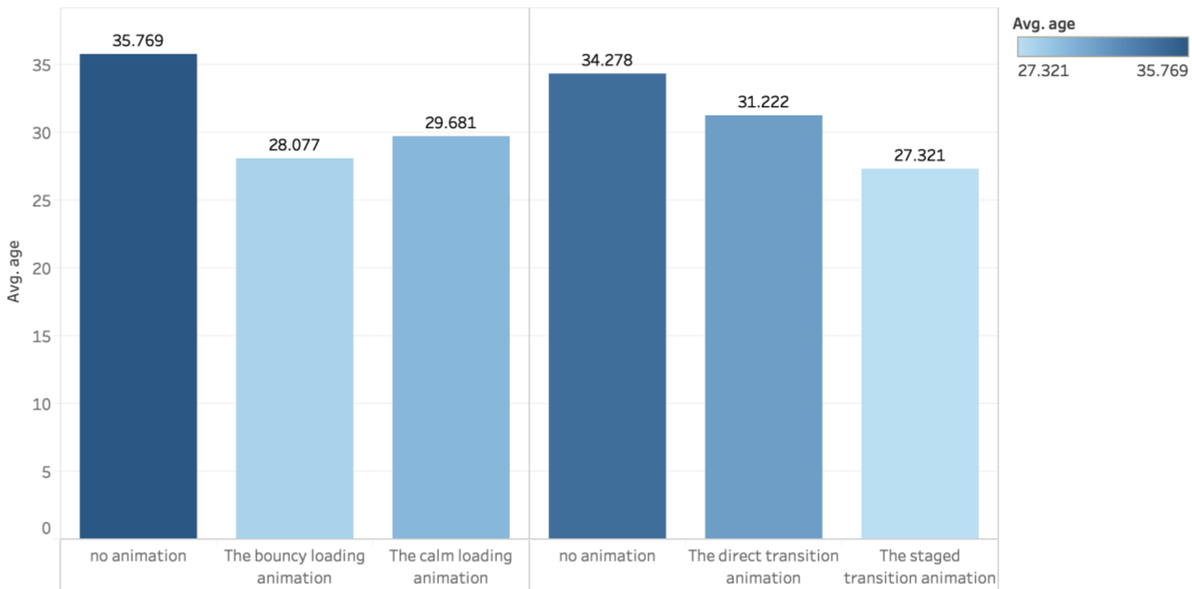

Figure 51 – Average age of the groups of participants preferring a specific condition

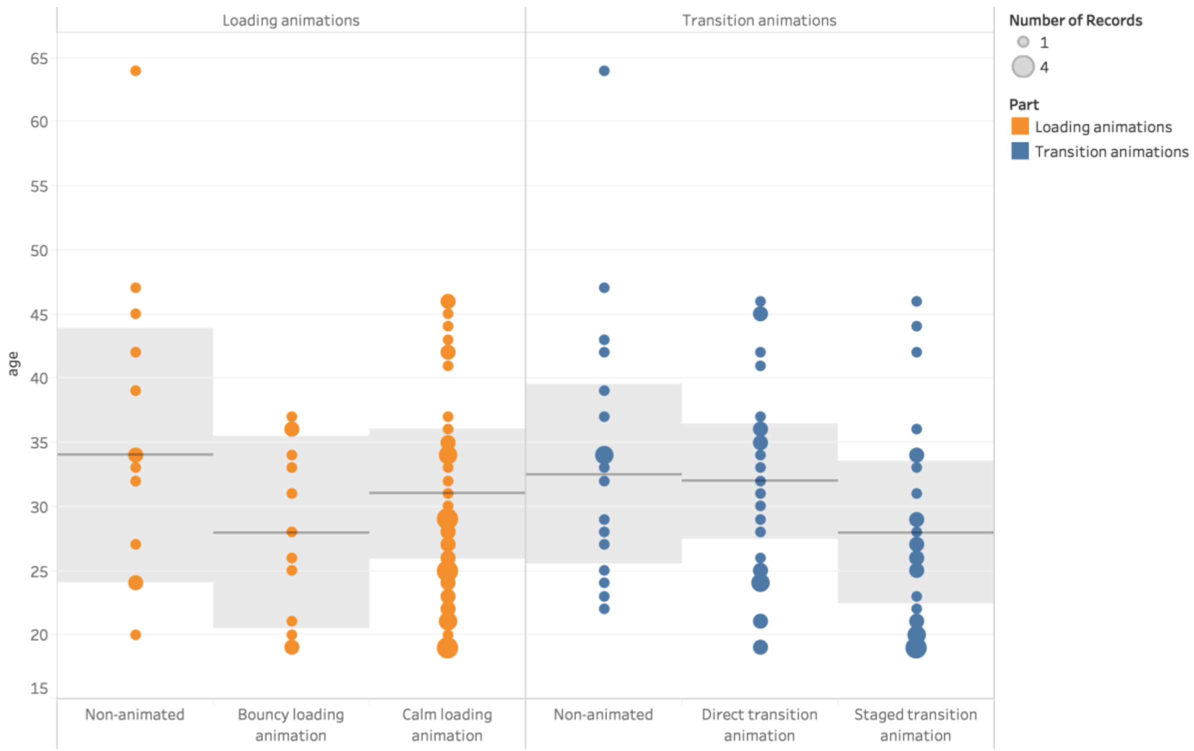

*Figure 52 – **Median** and 95% CI - Age of the groups of participants preferring a specific condition*

4.3 Qualitative results (optional from n = 73)

Number of qualitative reactions are shown in Table 5. In both parts, the animated conditions evoked slightly more (optional) reactions.

Table 5 – Number of qualitative reactions on the conditions

Task Name	
Bouncy loading animation	19
calm loading animation	19
no loading animation	12
Direct animated transition	11
Staged animated transition	16
non-animated transition	9

The qualitative reactions to the conditions of the loading animations were mapped from -2 (negative) to 2 (positive). The results of this mapping are shown in Figure 53. For the loading animations it can be observed the bouncy animation has more negative remarks associated with it compared to the calm and non-animated conditions. The calm animation has the most positive remarks. For the transition animations, both animated conditions have both more negative and positive remarks associated with them. They elicit more reactions, both negative and positive.

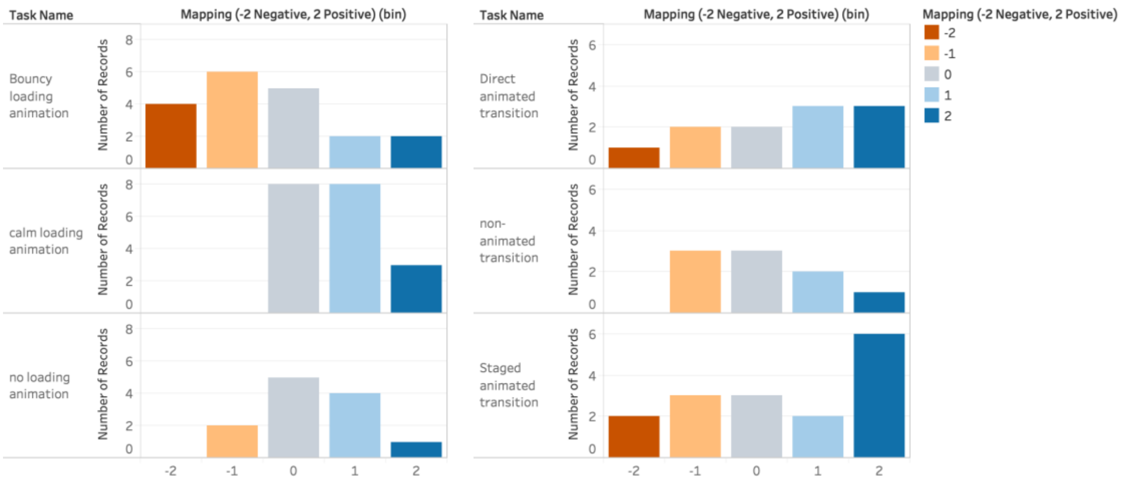

Figure 53 – Summary of the mappings of the qualitative data from -2 (negative) to 2 (positive)

4.4 Secondary measurements

4.4.1 Average number of clicks per condition (*n* = 73)

The number of times participants switched between views in all conditions were measured. The conditions in both parts of the experiment did however not seem to have an influence on the number of clicks. The results are shown in Figure 54.

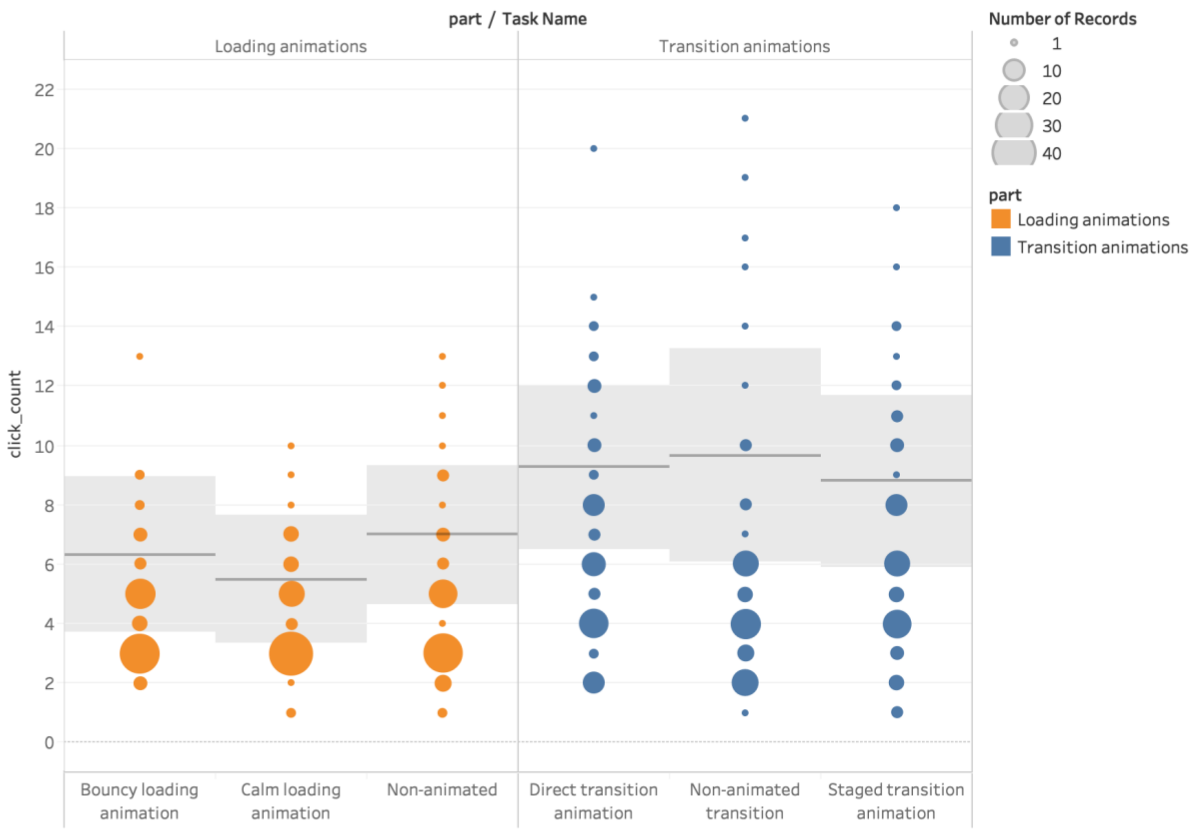

Figure 54 – The average number of clicks (switches between the 2 views) per condition for the loading animations and the transition animations, with their 95% confidence intervals.

5

Discussion

5 Discussion

Participants had very **diverse experiences** with the different conditions for both the loading animations and transition animations. Especially considering the transition animations, preferences differ a lot between participants. Where one participant clearly sees the staged animation as an added value opposed to not having an animation, the other participant finds it distracting and unnecessary. This difference clearly outlines how users can have different preferences and wishes, showing how important it is to know who you are designing an application for. This is emphasized by the observation that older participants seem to be more likely to prefer the non-animated condition whereas younger participants seem to be more likely to prefer animations. In general, the differences between the conditions are very small, which is also reflected in the very similar results of the meCUE questionnaire. Especially for the loading animations there is however a big preference for one condition, which cannot be clearly concluded from the meCUE questionnaire results.

5.1 Loading animations

Participants have a clear preference for the calm loading animation. The overall UX measured by meCUE also shows how the UX of the calm animation is rated highest, closely followed by the condition without animation with a nearly similar score. The overall UX of the condition with the bouncy animation is clearly rated lower than both other conditions. The number of clicks (switches between the 2 views) shows that all conditions have approximately the same level of interaction: in all conditions, participants switch on average about *6* times between the two views. It is surprising that the overall UX of the non-animated condition is similar to the calm animated condition, as much more people prefer the condition with the calm animation.

The meCUE results are similar for all loading animation conditions in all constructs of product qualities (*usability, functionality, aesthetics*) and emotions (*negative emotions, positive emotions*). Even though the bouncy animation seems to elicit more negative emotions and less positive emotions, none of the results is significant. These results can thus not explain the big difference in preference nor the difference in overall UX. Apparently there is a content validity problem: the questions are not sensitive enough to measure the subtle differences or the constructs of product qualities do not cover the relevant condition differences. The power validity is also questionable; a larger number of participants would have most probably led to a significant effect on the negative emotions for the bouncy animation. In general, the statements about emotions are all rated very low, showing how participants have difficulty relating emotions to graphs. This is also reflected in the qualitative feedback participants give, such as *"do not try to attach visualizations to emotions"* and *"I think the statements are odd, i don't have a particular feeling with graphs"*.

The fact that the calm animation scores higher on overall UX and in preference compared to the bouncy animation can also be seen in the answers on the comparison questionnaire. The calm animation scores on average 1.5 points higher both on '*aesthetic quality*' as well as '*helping in the understanding*' compared to the bouncy animation, leading to a significant difference. By specifically naming the difference between the conditions and making participants aware of it, participants seem to feel the urge to form an opinion. This could have been influenced by the steering terms '*bouncy*' and '*calm*', which were used in explaining the different animations.

The qualitative feedback shows results that are more in line with the preferences of participants. This feedback was given before the comparison and animation explanation. Respondents also often call the bouncy loading animation "*distracting*" or "*irritating*". This is a clear indication why participants prefer the calm animation over the bouncy animation and suggests an emotional reaction, even though this was not significantly measured with the meCUE questionnaire. In the mapping of the qualitative feedback the bouncy animation has much more negative remarks associated with it (n = 10 negative) as compared to the calm animation (n = 0 negative). Participants refer to the non-animated graph as: "*not interesting but illustrating the data well*' and "*simple but effective*". The big difference in preference between the calm animated condition (preferred by n=51) and the non-animated condition (preferred by n=16), even though their measured UX was approximately the same, can be seen in the qualitative feedback too. The calm animation has twice as much positive remarks (n = 19 positive) as compared to the non-animated version (n = 10), even though the remarks for both conditions are generally positive.

In summary, participants clearly prefer the calm loading animation, even though their measured UX with the non-animated graph is approximately the same. The bouncy animation clearly results in a lower UX, which cannot be significantly measured with the meCUE questionnaire in any construct. The qualitative feedback also shows how some participants find the non-animated graph somewhat simple and uninteresting, being a possible reason for the big preference for the calm animated condition. However, also this cannot be seen in the quantitative survey results. This suggests that not all aspects of the UX were properly measured with the meCUE questionnaire.

5.2 Transition animations

Considering the transition animations, there is a small preference for the conditions with a transition animation as opposed to no transition animation. The overall UX measured with the meCUE questionnaire is however approximately the same for all conditions, not showing the small preference for the animated conditions. The number of clicks (switches between the 2 views) shows that all conditions have approximately the same level of interaction: in all conditions, participants switch on average about *9* times between the two views

The meCUE results are similar for all loading animation conditions in all constructs of product qualities (*usability, functionality, aesthetics*) and emotions (*negative emotions, positive emotions*). The aesthetics from the meCUE questionnaire score higher on the animated transitions as opposed to the non-animated transition, however not significant. Even though the difference is not significant, the meCUE results suggest that the difference in preference might thus mainly be caused by the difference in aesthetics. Objectively speaking, the transition animation could help the usability by giving participants an understanding of the relation between the two views of the graph (Heer & Robertson, 2007), this is however not at all seen in the meCUE results.

From the qualitative feedback very similar scores are observed compared to the preferences. Both animated conditions get more qualitative responses, both positive and negative. It can for example be seen that some people find the staged animation distracting and unnecessary whereas others mention how they really appreciate the animation. Both the very positive and very negative remarks are mainly given on the animated conditions.

Looking at the comparison questionnaire, the two animations score very similar, both on '*added aesthetics*' and '*helping in the understanding*'. It is however important to note that the duration of the staged animation was about twice as long as the direct animation. The bigger standard deviation compared to the loading animations conditions show how opinions are divided, logically resulting in a less clear average preference.

In summary, participants have different preferences considering the transition animations. There is a small average preference for both animated transitions, which seems to be attributed to the higher aesthetic quality according to the meCUE results. None of the meCUE scores were however significant. When looking at participants in specific, there are a lot of participants that have a clear preference, which are evened-out by participants with opposite experiences. This outlines the subjectivity of the concept of UX and suggests that a one-size-fits-all UX approach is not favourable.

5.3 Implications on the measurement of UX

All in all, the meCUE questionnaire results show very similar results for the conditions, probably as a result of the very subtle differences between the conditions. The meCUE questionnaire results do however show some differences in product qualities and emotion, but none are significant except for the 'overall UX'. It is therefore not surprising that these results do not always account for the differences in preference. To see why participants prefer a certain visualization with corresponding animation, researchers might have to compare the conditions and specifically ask for their opinion about the difference. This suggestion is strengthened by the fact that participants do have an opinion about the visualizations and the differences, which they often give as qualitative feedback and in the form of their preference. They however seem to be unable to catch this opinion in the current statements of the meCUE questionnaire.

A possible solution to improve the sensitivity of this questionnaire is adding extra constructs to the product qualities of the CUE model. Similar to leaving out certain constructs, which is possible because of the modular nature of the evaluation method, certain constructs could be added. Identification and status were left out in this experiment, as they are not applicable to information visualization. Also, the questions about the emotions should be carefully reconsidered in further research, as these constructs frustrate participants; they had a hard time associating specific emotions with charts. As described in the introduction, animation is an aspect that mainly influences the aesthetic quality of an information visualization, but can also affect aspects as relational understanding (Heer & Robertson, 2007) and engagement (Bartram & Nakatani, 2010), which are not well reflected by the current questions of the meCUE questionnaire. Future research with the meCUE model should thus try to find all product qualities for a domain and note that these qualities are different for every application.

Also, it might be beneficial to make a less strong separation between emotions and product qualities. Aranyi & van Schaik (2016) for example suggest a direct relation in the CUE model from interaction characteristics to emotions, not necessarily passing the interpretation of product qualities first. This way 'desirability' and 'credibility' could be added as product qualities of UX, even though these qualities are intertwined with emotions. By not making this strong separation, self-reporting a UX might be easier for a user, as users might not always know what product quality triggers their emotions.

Another interesting observation is that the conditions with higher rated aesthetics are also rated higher on usability, even though the loading animation does, objectively speaking, not improve anything other than aesthetics. This finding is in line with the aesthetic-usability effect as described by Tractinsky; that if something is more beautiful it is also perceived as more usable (Tractinsky et al., 2000). This shows that the perceived usability differs from the objective usability, but simultaneously raises the question how good humans are at self-reporting their UX on the basis of different product qualities. Do users always know why they prefer a system or have a better experience with a system? Some research argues that when designing a UX, one should not listen to the user but rather observe them (Nielsen & Levy, 1994). As the meCUE method of measuring UX solely depends on the self-report of the user, this is an important question to consider in future research with the meCUE questionnaire. This same fact also suggests that objective behavioural or physiological measurements form a promising alternative as measurement of UX.

Aside from the limitations mentioned above, a limitation involves the property of UX that it can change over time, and that a positive initial experience is not guaranteed to motivate prolonged use. How UX evolves over time is not often researched and one could imagine that animations are fun for a while, but after a user has seen it a couple of times, it might get distracting. Future research should prove if this is indeed the case when a visualization is more often used. At a first encounter hedonics are important, but over time aspects as usability will become more important (Wilamowitz-Moellendorff et al., 2006). This should also be considered and researched for the animations described in this experiment. Giving users the ability to turn on or off animations, could help in a positive UX over a longer time.

As this research found that the meCUE method is not as useful for small differences between conditions, also bigger differences in conditions should be researched with the meCUE method, to see if larger differences can be measured more accurately. Even though the objectivity of the ratings of qualitative feedback is debatable as it was just rated by one person, it indicates that qualitative methods can strongly outweigh quantitative methods of measuring UX in cases with small differences between conditions. Future research could benefit from specifying specific user groups; this research already shows how age has an influence on the preference for animated visualizations or non-animated visualizations. It is interesting to research how other user characteristics have an influence on UX.

6

Conclusion

6 Conclusion

The first research question was: *"How can the UX of an information visualization be measured in a quantitative manner?"*. An extensive literature review set out different definitions and models of UX and related these to the domain of data visualization. This literature review showed how the concept of UX is still vague with a wide variety of meanings. Different UX models have different areas of focus. There are many methods for measuring UX, both objective and subjective methods. As this experiment aimed to research the UX of an information visualization using it while having a goal, a model was chosen in which instrumental aspects such as the usability were well reflected. Since information visualizations are graphics based, also aesthetics plays an important role. The fact that information visualizations are not physical products made aspects as identification or status less important. From these criteria the CUE model seemed to match best with the domain of data visualization. This model was used in the experiments.

The second research question was: *"What aspects of the CUE model and its measurement tool meCUE can be used for the domain of information visualization?"*. The meCUE results show no significant differences between the conditions, but some small differences are observable. These small differences are not only a consequence of the subtleness of the differences between the conditions, they are also caused by the diverseness of opinions in the participants group. By averaging results over this group, contrasting opinions can cancel each other out. The only significant difference measured by the meCUE questionnaire was the 'overall UX', where the bouncy animation scored lower than the other two conditions. The preferences however show a clear preference for the calm animation, even though it's measured UX is approximately the same as the non-animated condition.

First, the constructs of the product qualities usability, functionality and aesthetics do clearly not cover all aspects of the UX of information visualization. The needed constructs to measure the product qualities are different for each application. It is thus important that future research determines what other constructs have to be added to the meCUE questionnaire, to be able to apply it to data visualization. As discussed in the background, identification and status are examples of constructs that do not apply to data visualization but rather apply to physical products. Other models and theories indicate that promising candidates for added constructs could be engagement, interaction aesthetics, credibility and desirability. Future research should however show if these constructs overlap too much with other constructs or with each other, and find suitable questions to measure them.

Further, it is conceptually adequate to separate emotion from product qualities, but impractical as humans cannot always relate emotions to product qualities causing them. In this research for example, people had negative emotions associated with the bouncy animation, but this was not reflected in the aesthetic construct. It might be practical to not separate emotions and qualities as strong as the CUE model currently does, as users might understand their emotions better than the product qualities causing them. By including constructs as "Joy of use" or "desirability", which are clearly qualities intertwined with user emotions, users might succeed better in self-reporting their UX.

Finally, compared to qualitative research, quantitative UX research using the meCUE method has a lot of disadvantages in the context of this research. First of all, the effect size is very small, requiring a very large number of participants to show significant differences. Second, qualitative research allows for finding more specific answers as to why a certain visualization is preferred if participants get to express their ideas freely.

The third research question was: *"How do loading animations and transition animations influence the UX of information visualizations?"*. The different conditions in the experiment have clearly shown that animations can both hurt and improve the UX of information visualizations. For the loading animations, there was a big preference amongst the participants for the calm loading animation. The bouncy loading animation was an example of an animation that hurt the UX of the information visualization; participants generally found it distracting and unnecessary which resulted in a lower overall UX. Assuming that the preference for the calm loading animation is a consequence of a positive UX (as suggested by the CUE model), animation did in this case improve the UX. The measured UX with the meCUE was however approximately the same for the non-animated condition, suggesting that not all aspects of the UX were properly measured. For the transition animations, there was a small preference for both animated transitions, but overall participants had very different opinions. Some participants really appreciated the animated transitions, whereas others found them distracting.

These different experiences emphasize the subjectivity of UX. For one participant the animated transitions improved the UX, for others the animated transitions were a distraction from the message of the graph. A one-size-fits-all UX design is therefore not favourable. By more specifically specifying a user group these deviations within the participants can be decreased. In addition to that, some participants indicated increased usability of functionality even if the animation only had aesthetic purpose. This shows raises questions about the tenability of self-reported UX and could be an argument for more objective measures of UX.

References

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *16*, 3–9. https://doi.org/citeulike-article-id:6930744

Ahsanullaha, Suziah Sulaiman, Ahmad K B Mahmood, & Muzafar Khan. (2015). Understanding factors influencing User Experience of interactive systems: A literature review. *ARPN Journal of Engineering and Applied Sciences*, *10*, 18175–18185.

Alben, L. (1996). Quality of experience. *Interactions*, *3*, 11–15. https://doi.org/10.1145/235008.235010

Alves, R., Valente, P., & Nunes, N. J. (2014). The state of user experience evaluation practice. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction Fun, Fast, Foundational - NordiCHI '14*, 93–102. https://doi.org/10.1145/2639189.2641208

Amar, R., Eagan, J., & Stasko, J. (2005). Low-level components of analytic activity in information visualization. In *Proceedings - IEEE Symposium on Information Visualization, INFO VIS* (pp. 111–117). https://doi.org/10.1109/INFVIS.2005.1532136

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2019). Gorilla in our Midst: An online behavioral experiment builder. *BioRxiv*, 438242. https://doi.org/10.1101/438242

Aranyi, G., & van Schaik, P. (2016). Testing a model of user-experience with news websites. *Journal of the Association for Information Science and Technology*, *66*, 2471–2493. https://doi.org/10.1002/asi.23462

Bacigalupi, M. (1998). The craft of movement in interaction design. In *Proceedings of the working conference on Advanced visual interfaces* (pp. 174–184). ACM. https://doi.org/10.1145/948496.948521

Bartram, L. (1997). Perceptual and interpretative properties of motion for information visualization. *Workshop on New Paradigms in Information Visualization and Manipulation*. https://doi.org/10.1145/275519.275520

Bartram, L. (1998). Enhancing Visualizations with Motion. In *Proceedings of Information Visualization* (pp. 3–7).

Bartram, L., & Nakatani, A. (2010). What makes motion meaningful? Affective properties of abstract motion. In *Proceedings of the 2010 Fourth Pacific-Rim Symposium on Image and Video Technology* (pp. 468–474). https://doi.org/10.1109/PSIVT.2010.85

Bederson, B. B., & Boltman, A. (2007). Does Animation Help Users Build Mental Maps of Spatial Information? In *The Craft of Information Visualization*. https://doi.org/10.1016/b978-155860915-0/50015-9

Bertin, J. (1983). *Semiology of Graphics*. University of Wisconsin Press. https://doi.org/10.1103/PhysRevLett.94.208902

Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences*, *116*(6), 201807180. https://doi.org/10.1073/pnas.1807180116

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Brath, R., Peters, M., & Senior, R. (2005). Visualization for communication: The importance of aesthetic sizzle. In *Proceedings of the International Conference on Information Visualisation* (pp. 724–729). https://doi.org/10.1109/IV.2005.145

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). Readings in Information Visualization: Using Vision to Think. In *Information Display*. Morgan Kaufmann Publishers Inc.

Cawthon, N., & Moere, A. V. (2006). A Conceptual Model for Evaluating Aesthetic Effect within the User Experience of Information Visualization. *Tenth International Conference on Information Visualisation (IV'06)*, 374–382. https://doi.org/10.1109/IV.2006.4

Cawthon, N., & Moere, A. Vande. (2007). The effect of aesthetic on the usability of data visualization. *Proceedings of the International Conference on Information Visualisation*, 637–645.

https://doi.org/10.1109/IV.2007.147

Chen, M., Ebert, D., Hagen, H., Laramee, R. S., Van Liere, R., Ma, K.-L., … Silver, D. (2009). Data, Information, and Knowledge in Visualization. *Computer Graphics and Applications, IEEE*, *29*(1), 12–19. https://doi.org/10.1109/MCG.2009.6

Few, S. (2006). *Information Dashboard Design*. O'Reilly Media, Inc.

Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis* (1st ed.). Analytics Press.

Few, S. (2017). Data Visualization Effectiveness Profile. *Perceptual Edge*.

Figueiras, A. (2015). Towards the Understanding of Interaction in Information Visualization. In *2015 19th International Conference on Information Visualisation* (pp. 140–147). https://doi.org/10.1109/iV.2015.34

Forlizzi, J., & Battarbee, K. (2004). Understanding experience in interactive systems. *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 261–268. https://doi.org/10.1145/1013115.1013152

Friedrich, C. (2002). *Animation in relational information visualization. (Unpublished doctoral dissertation)*. University of Sydney. Retrieved from https://ses.library.usyd.edu.au/bitstream/2123/9288/1/FriedrichThesis.pdf

Ganglbauer, E., Schrammel, J., Deutsch, S., & Tscheligi, M. (2009). Applying psychophysiological methods for measuring user experience: Possibilities, challenges and feasibility. *Interact'09 Conference*. https://doi.org/10.1007/978-3-642-23768-3

Hassenzahl, M. (2003). The Thing and I: Understanding The Relationship Between User and Product. In *Human–Computer Interaction Series, Funology* (pp. 31–42). Springer. https://doi.org/10.1007/1-4020-2967-5_4

Hassenzahl, M. (2008a). Aesthetics in interactive products: Correlates and consequences of beauty. In *Product Experience* (pp. 287–302). Elsevier. https://doi.org/10.1016/B978-008045089-6.50014-9

Hassenzahl, M. (2008b). User experience (UX): Towards an experiential perspective on product quality. *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine on - IHM '08*, 11–15. https://doi.org/10.1145/1512714.1512717

Hassenzahl, M., Kekez, R., & Burmester, M. (2002). The importance of a software's pragmatic quality depends on usage modes. *Proceedings of the 6th International Conference on Work With Display Units (WWDU 2002)*, 275–276.

Hassenzahl, M., & Monk, A. (2010). The inference of perceived usability from beauty. *Human-Computer Interaction*, *25*(3), 235–260. https://doi.org/10.1080/07370024.2010.500139

Hassenzahl, M., Platz, A., Burmester, M., & Lehner, K. (2000). Hedonic and ergonomic quality aspects determine a software's appeal. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI 00*, 201–208. https://doi.org/10.1145/332040.332432

Hassenzahl, M., & Tractinsky, N. (2006). User experience - A research agenda. *Behaviour and Information Technology*, *25*(2), 91–97. https://doi.org/10.1080/01449290500330331

Heer, J., & Bostock, M. (2010). Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 203–212). ACM. https://doi.org/10.1145/1753326.1753357

Heer, J., & Robertson, G. G. (2007). Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1240–1247. https://doi.org/10.1109/TVCG.2007.70539

Herne, H., & Huff, D. (2006). How to Lie with Statistics. *Applied Statistics*. https://doi.org/10.2307/2346789

Hong, W., Thong, J. Y. L., & Tam, K. Y. (2004). Does animation attract online users' attention? The effects of flash on information search performance and perceptions. *Information Systems Research*, *15*(1), 1–108. https://doi.org/10.1287/isre.1040.0017

ISO 9241-11. (2017). Part 11: Usability: Definitions and concepts. Retrieved January 5, 2019, from https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

Jiang, Z. (Jack), Wang, W., Tan, B. C. Y., & Yu, J. (2016). The Determinants and Impacts of Aesthetics in Users' First Interaction with Websites. *Journal of Management Information Systems*, *33*(1), 229–259. https://doi.org/10.1080/07421222.2016.1172443

Karapanos, E. (2013). *Modeling Users' Experiences with Interactive Systems*. *Studies in Computational Intelligence 436*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-31000-3

Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J. (2009). User experience over time: An initial framework. In *Proceedings of the 27th SIGCHI Conference on Human Factors in Computing Systems* (pp. 729–738). New York, NY, USA: ACM. https://doi.org/10.1145/1518701.1518814

Kirakowski, J., & Corbett, M. (1993). SUMI: the Software Usability Measurement Inventory. *British Journal of Educational Technology*, *24*, 210–212. https://doi.org/10.1111/j.1467-8535.1993.tb00076.x

Kurosu, M., & Kashimura, K. (1995). Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In *Conference Companion on Human Factors in Computing Systems (CHI '95)* (pp. 292–293). New York, NY, USA: ACM. https://doi.org/10.1145/223355.223680

Lachner, F., Naegelein, P., Kowalski, R., Spann, M., & Butz, A. (2016). Quantified UX: Towards a Common Organizational Understanding of User Experience. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction - NordiCHI '16* (pp. 1–10). New York, NY, USA: ACM. https://doi.org/10.1145/2971485.2971501

Lambdin, C., & Shaffer, V. A. (2009). Are Within-subjects designs transparent? *Judgment and Decision Making*, *4*(7), 554–566. https://doi.org/10.1037/e722352011-194

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human Computer Studies*, *60*(3), 269–298. https://doi.org/10.1016/j.ijhcs.2003.09.002

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P. O. S., & Kort, J. (2009). Understanding, scoping and defining user experience. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09* (pp. 719–728). New York, NY, USA: ACM. https://doi.org/10.1145/1518701.1518813

Lemon, K., Allen, E. B., Carver, J. C., & Bradshaw, G. L. (2007). An empirical study of the effects of Gestalt principles on diagram understandability. In *Proceedings - 1st International Symposium on Empirical Software Engineering and Measurement, ESEM 2007* (pp. 235–244). https://doi.org/10.1109/ESEM.2007.17

Mahlke, S. (2008). Visual aesthetics and the user experience. In *Proceedings of The Study of VisualAesthetics in Human-Computer Interaction* (pp. 1862–4405).

Mahlke, S., & Thüring, M. (2007). Studying antecedents of emotional experiences in interactive contexts. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* (pp. 915–918). New York, NY, USA: ACM. https://doi.org/10.1145/1240624.1240762

Mao, J. Y., Vredenburg, K., Smith, P. W., & Carey, T. (2005). The state of user-centered design practice. *Communications of the ACM*, *48*(3), 105–109. https://doi.org/10.1145/1047671.1047677

McNamara, N., & Kirakowski, J. (2006). Functionality, Usability, and User Experience: Three Areas of Concern. *Interactions*, *13*(6), 26–28. https://doi.org/10.1145/1167948.1167972

Michailidou, E., Harper, S., & Bechhofer, S. (2008). Visual complexity and aesthetic perception of web pages. In *Proceedings of the 26th annual ACM international conference on Design of communication - SIGDOC '08* (pp. 215–224). New York, NY, USA: ACM. https://doi.org/10.1145/1456536.1456581

Minge, M. (2008). Dynamics of user experience. In *Proceedings of the Workshop on Research Goals and Strategies for Studying User Experience and Emotion, NordiCHI*.

Minge, M., & Thüring, M. (2018). The MeCUE questionnaire (2.0): Meeting five basic requirements for lean and standardized UX assessment. *Design, User Experience, and Usability: Theory and Practice. DUXU 2018.*, *10918*, 451–469. https://doi.org/10.1007/978-3-319-91797-9_33

Minge, M., Thüring, M., Wagner, I., & Kuhr, C. V. (2017). The meCUE questionnaire: A modular tool for measuring user experience. In *Advances in Intelligent Systems and Computing* (pp. 115–128). https://doi.org/10.1007/978-3-319-41685-4_11

Morville, P. (2004). User Experience Design. Retrieved September 20, 2018, from http://semanticstudios.com/user_experience_design/

Nielsen, J., & Levy, J. (1994). Measuring usability: preference vs. performance. *Communications of the ACM*, *37*(4), 66–75. https://doi.org/10.1145/175276.175282

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things. Why We Love (or Hate) Everyday Things*. Basic Civitas Books.

Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., & Gajos, K. Z. (2013). Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (pp. 2049–2058). New York, NY, USA: ACM. https://doi.org/10.1145/2470654.2481281

Reinsel, D., Gantz, J., & Rydning, J. (2018). The Digitization of the World From Edge to Core, (November). Retrieved from https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

Robertson, G. G., Mackinlay, J. D., & Card, S. K. (1991). Cone trees: Animated 3D visualizations of hierarchical information. In *1991 SIGCHI Conference on Human Factors in Computing Systems, CHI 1991* (pp. 189–194). New York, NY, USA: ACM. https://doi.org/10.1145/108844.108883

Roto, V. (2007). User experience from product creation perspective. *Towards a UX Manifesto Workshop*, (Roto), 1–4. Retrieved from https://research.nokia.com/files/UXmanifesto-Roto.pdf

Roto, V., Law, E., & Vermeeren, A. P. O. S. (2011). User experience white paper. Retrieved from http://www.allaboutux.org/files/UX-WhitePaper.pdf

Roto, V., Obrist, M., & Väänänen-Vainio-Mattila, K. (2009). User experience evaluation methods in academic and industrial contexts. *User Experience Evaluation*. https://doi.org/10.1.1.150.1764

Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163–180. https://doi.org/10.1177/0165551506070706

Rusu, A., Fabian, A. J., Jianu, R., & Rusu, A. (2011). Using the gestalt principle of closure to alleviate the edge crossing problem in graph drawings. In *15th International Conference on Information Visualisation* (pp. 488–493). https://doi.org/10.1109/IV.2011.63

Sas. (2013). Data Visualization Techniques.

Sheldon, Kennon M.; Elliot, Andrew J.; Kim, Youngmee; Kasser, T. (2001). What is satisfying about satisfying events. *Journal of Personality and Social Psychology*, *80*(2), 325–339. https://doi.org/10.1037//O022-3514.80.2.325

Skog, T., Ljungblad, S., & Holmquist, L. E. (2003). Between aesthetics and utility: Designing ambient information visualizations. In *Proceedings - IEEE Symposium on Information Visualization, INFO VIS* (pp. 233–240). https://doi.org/10.1109/INFVIS.2003.1249031

Slattery, T. J., & Yates, M. (2018). Word skipping: Effects of word length, predictability, spelling and reading skill. *Quarterly Journal of Experimental Psychology*, *71*(1), 250–259. https://doi.org/10.1080/17470218.2017.1310264

Sonderegger, A., Uebelbacher, A., Pugliese, M., & Sauer, J. (2014). The influence of aesthetics in usability testing: The Case of Dual-Domain Products. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI*, 21–30. https://doi.org/10.1145/2556288.2557419

Sutcliffe, A. (2009). *Designing for User Engagement: Aesthetic and Attractive User Interfaces*. Morgan & Claypool Publishers. https://doi.org/10.2200/S00210ED1V01Y200910HCI005

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human-technology interaction. *International Journal of Psychology*, *42*(4), 253–264. https://doi.org/10.1080/00207590701396674

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, *13*(2), 127–145. https://doi.org/10.1016/S0953-5438(00)00031-X

Tufte, E. R. (1983). *The visual display of quantitative information*. *CT Graphics, Cheshire*.

https://doi.org/10.1119/1.14057

Tukey, J. W., & Wilk, M. B. (1966). Data analysis and statistics: an expository overview. In *Proceedings of SIGCHI '66 conference on Human Factors in computing systems* (pp. 695–709). New York, NY, USA: ACM. https://doi.org/10.1145/1464291.1464366

Väänänen-Vainio-Mattila, K., Roto, V., & Hassenzahl, M. (2008). Towards practical user experience evaluation methods. *EL-C. Law, N. Bevan, G. Christou, M. Springett & M. L{á}rusd{ó}ttir (Eds.) Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, 19–22.

Veeneklaas, J. (2018). *VisUX: a framework of user experience within data visualizations (masters thesis)*. Utrecht University.

Vermeeren, A., Law, E., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In *Proceedings: NordiCHI 2010* (pp. 521–530). New York, NY, USA: ACM. https://doi.org/10.1145/1868914.1868973

Ware, C. (2004). *Information Visualization: Perception for Design: Second Edition* (2nd ed.). Morgan Kauffmann Publishers. https://doi.org/10.1016/B978-1-55860-819-1.X5000-6

Wexler, S., Shaffer, J., & Cotgreave, A. (2017). The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios (p. chapter 16). John Wiley & Sons.

Wilamowitz-Moellendorff, M., Hassenzahl, M., & Platz, A. (2006). Dynamics of user experience : How the perceived quality of mobile phones changes over time. In *4th Nordic Conference on Human-Computer Interaction* (pp. 1560–1563). New York, NY, USA: ACM. https://doi.org/10.1109/IEMBS.2006.260296

Wilkinson, L. (2010). The grammar of graphics. *WIREs Comp Stat, 2*. https://doi.org/10.1002/wics.118

Wimmer, C., Weishapl, P., Grechenig, T., & Kappel, K. (2011). Aesthetic Quality of Physical Behavior and its Impact on User Experience. *MobileHCI 2011*.

Winans, M., Faupel, D., Armstrong, A., Henderson, J., Valentine, E., McDonald, L., … Magill, E. (2016). 10 Key Marketing Trends for 2017 Customer Expectations and Ideas for Exceeding Customer Expectations. *IBM Offering Information*. Retrieved from https://www.ibm.com/downloads/cas/XKBEABLN

Zeng, L. (2005). Designing the user interface: strategies for effective human-computer interaction. *International Journal of Human–Computer Interaction*, *25*(7), 707–708. https://doi.org/10.1080/10447310903187949

Appendixes

I. The data sets

United states drug use:

https://www.samhsa.gov/data/sites/default/files/NSDUHresults2012/NSDUHresults2012.pdf

US - Percentage of Smokers aged (among 12 years or older)

	1	2	3	4	5	6	7	8	9	10	11
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Cigarettes	78	76,2	74,7	74,7	75	72,9	72	69,9	69	66,3	66,3
Cigars	16,2	16,2	17,1	16,8	16,8	16,2	15,9	15,9	15,6	15	15,6

US - Past year non-medical usage of selected drugs (among 12 years or older)

	1	2	3	4	5	6	7	8	9	10	11
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Pain Relievers	5,7	6	5,4	5,7	6,3	6,3	5,7	6,3	6	5,1	5,7
Tranquilizers	2,4	2,4	2,1	2,1	2,1	2,1	2,1	2,4	2,7	2,1	2,4

US – Past month usage of selected drugs (among 12 year or older)

	1	2	3	4	5	6	7	8	9	10	11
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Cocaine	2,7	3	2,4	3	3	2,4	2,1	2,1	1,8	1,5	1,8
Marijuana	18,6	18,6	18,3	18	18	17,4	18,3	20,1	20,7	21	21,9

Pure alcohol consumption among persons (age 15+) in litres per capita per year, 2010. By WHO (World Health Organization) Data.

http://www.wikiwand.com/en/List_of_countries_by_alcohol_consumption_per_capita

	1	2	3	4	5	6	7	8
	Australia	France	Germany	Netherlands	Poland	Russia	UK	US
Beer	5,37	2A9	6,32	4,63	6,89	5,68	4,28	4,60
Wine	4,48	6,88	3,28	3,60	1,16	1,72	3,92	1,57
Spirits	1,53	2,82	2,19	1,67	4,44	7,70	2,53	3,01

Other	0,83	0,21	0	0	0	0	0,87	0

Meat consumption in Kilograms per capita, 2017.

https://data.oecd.org/agroutput/meat-consumption.htm

	1	2	3	4	5	6	7	8
	Australia	China	EU	Israel	Japan	Russia	Turkey	US
Beef and Veal	20,9	4,1	11,0	20,0	6,6	10,1	8,3	25,8
Pork	20,7	30,8	32,5	1,6	15,4	20,7	0,1	23,6
Poultry	44,5	12,3	24,2	56,9	14,3	28,7	17,9	48,8
Sheep	8,5	0,9	1,9	1,8	0,1	1,2	4,1	0,4

Energy consumption per capita per country, and the source (disregarding exchange of energy between countries) 2016 (in watts per person).

https://www.cia.gov/library/publications/the-world-factbook/

	1	2	3	4	5	6	7	8
	Australia	China	France	Germany	Netherlands	Russia	Spain	US
Thermal Energy (coal, gas, oil)	1000	414	63	452	603	576	275	944
Nuclear	0	6	584	135	27	140	110	264
Hydro power	73	75	59	22	1	134	58	102
Other sustainable	39	15	30	144	93	4	107	67

II. The questions for the tasks

The first answer mentioned is the correct answer.

Graph	Question Cat	Question and answers (correct in *bold*)
Bar Static	Retrieve value	Approximately how many Americans used cocaine in 2003? (m = million) Answers: **3m**, 2m, 2.4m
	Filter operation	In what year did more than 21 million Americans use marihuana? Answers: **2012**, 2010, 2007
	Find extremum	In what year did the lowest number of Americans use cocaine? Answers: **2011**, 2004, 2012
Bar Sine	Retrieve value	Approximately how many people used tranquilizers for non-medical purposes in 2010? (m = million) Answers: **2.7m**, 2.4m, 2.1m
	Filter operation	In what year was the number of Americans using pain-relievers for non-medical purposes lower than 5.2 million? Answers: **2011**, 2004, 2006
	Find extremum	In what year was the number of Americans using Tranquilizers for non-medical purposes higher than 2.5 million? Answers: **2010**, 2012, 2004
Bar Static	Retrieve value	Approximately how many Americans smoked cigars in 2011? (m = million) Answers: **15m**, 16.1m, 16.7m
	Filter operation	In what year was the number of Americans smoking cigarettes the highest? Answers: **2002**, 2006, 2012
	Find extremum	In what year did most Americans smoke cigars? Answers: **2004**, 2001, 2012

The first answer mentioned is the correct answer.

Graph	Question Cat	Question
Transition static	Find Extremum	Which of the selected countries has the highest consumption of nuclear energy per person? Answers: **France**, United States
	Retrieve 2 values and compare	What is the main source of energy of the country with the highest total energy consumption per person? Answers: **Thermal energy**, Nuclear energy
	Retrieve value	Approximately how much energy in watts does an inhabitant of Australia consume? Answers: **1100 watts**, 500 watts
Transition direct	Find Extremum	Which of the selected countries has the highest consumption of poultry per person?

		Answers: **Israel**, United States
	Retrieve 2 values and compare	What is the main kind of meat eaten by the country with the lowest total consumption of meat per person? Answers: **Poultry**, Beef and Veal
	Retrieve value	Approximately how many kilograms of meat does an inhabitant of Turkey eat in a year? Answers: **30 kilograms**, 60 kilograms
Transition staged	Find Extremum	Which of the selected countries has the highest consumption of wine per person? Answers: **France**, Australia
	Retrieve 2 values and compare	What is the main type of alcohol drank by the country with the highest total alcohol consumption per person? Answers: **Spirits**, Beer
	Retrieve value	Approximately how many litres of pure alcohol does an inhabitant of the Netherlands drink on average in a year? Answers: **10 litres**, 12 litres

III. Questionnaires

MeCUE questionnaire (iteration 2, emotions)

Experience Questionnaire

Based on the interaction with the previous visualization and corresponding questions, please indicate for every of the following statements to what extend you agree, with answers ranging from 1 (strongly disagree) to 7 is (strongly agree).

Emotions

1. The visualization exhilarates me.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

2. The visualization makes me tired.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

3. The visualization annoys me.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

4. The visualization relaxes me.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

5. When using this visualization i feel exhausted.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

6. The visualization makes me feel happy.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

7. The visualization frustrates me.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

8. The visualization makes me feel euphoric.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

9. The visualization makes me feel passive.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

10. The visualization calms me.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

11. When using this visualization, i feel cheerful.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

12. The visualization angers me.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

Overall Evaluation

10. How did you experience the visualization as a whole?

Bad	-5	-4	-3	-2	-1	0	1	2	3	4	5	**Good**

11. Remarks about this visualization (optional)

Next

MeCUE questionnaire (iteration 1, product qualities)

Presented in the same manner as above; with the questions about emotions replaced by the following questions:

Module I: Usefulness (F), Usability (U)

1	The product is easy to use.	U. 1
2	The functions of the product are exactly right for my goals.	F. 1
3	It is quickly apparent how to use the product.	U. 2
4	I consider the product extremely useful.	F. 2
5	The operating procedures of the product are simple to understand.	U. 3
6	With the help of this product I will achieve my goals.	F. 3

Module II: Visual aesthetics (A), Status (S), Commitment (C)

7	The product is creatively designed.	A. 1
8	The design looks attractive.	A. 2
9	The product is stylish	A. 3

General questionnaire part 1

The difference between the graphs was the animation when loading the data;

- One graph had no animation
- One graph had a calm animation (such as *Fig 1*)
- One graph had a bouncy animation (such as *Fig 2*)

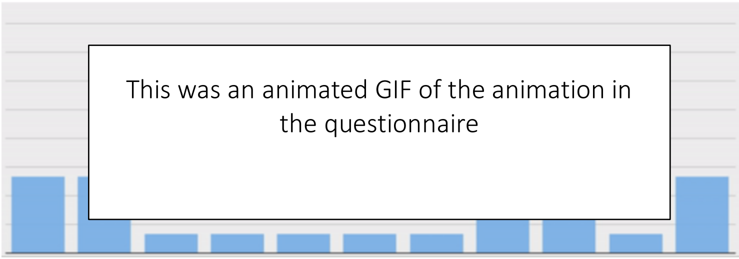

Fig 1 - Loop of the calm animation

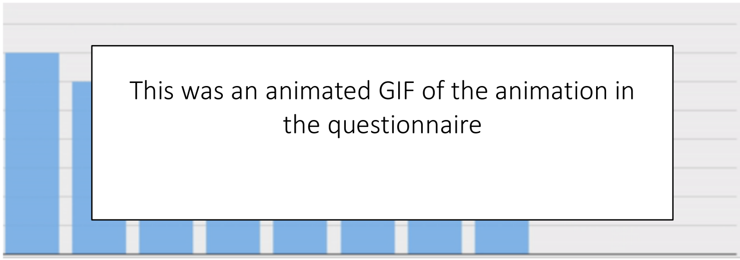

Fig 2 - Loop of the bouncy animation

Please indicate to what extend you agree with the following statements:

The calm animation (fig 1) added to the beauty of the graph.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

The calm animation (fig 1) helped in understanding the graph.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

The bouncy animation (fig 2) added to the beauty of the graph.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

The bouncy animation (fig 2) helped in understanding the graph.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

Which graph do you prefer?

Please Select...

Motivation (optional)

General questionnaire part 2

The difference between the graphs was the animation when switching between the views:

- One graph had no animation
- One graph had a direct animation (such as *Fig 1*)
- One graph had a staged animation (such as *Fig 2*)

Fig 1 - Loop of the direct animation

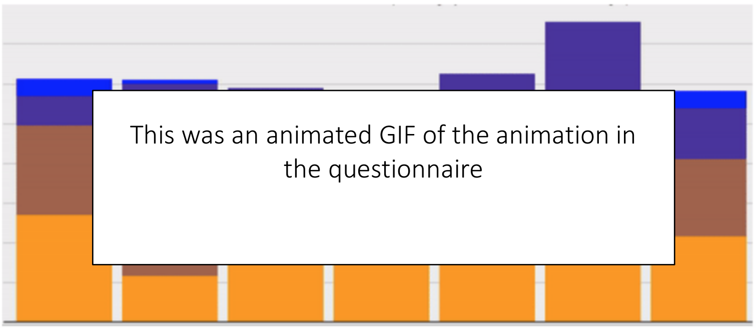

Fig 2 - Loop of the staged animation

Please indicate to what extend you agree with the following statements:

The direct animation (fig 1) added to the beauty of the graph.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

The direct animation (fig 1) helped in understanding the graph.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

The staged animation (fig 2) added to the beauty of the graph.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

The staged animation (fig 2) helped in understanding the graph.

Strongly disagree	1	2	3	4	5	6	7	**Strongly agree**

Which graph do you prefer?

Please Select...

Motivation (optional)

IV. Excluded participants

To assure the quality of the data, several criteria for inclusion of data were set.

Data quality criterium 1: Interaction

Participants must have seen the animation. Therefore, they should have switched between the different views. This criterium lead to 15 exclusions.

Participant private ID	Min. click-count
498524	0
498540	0
513885	0
498555	0
513815	0
513871	0
514001	0
498523	0
498546	0
513764	0
513790	0
513823	0
513865	0
513877	0
513883	0

Data quality criterium 2: Task

The number of correct answers for each task should be higher than 2, as the questions are simple questions that can be read of directly from the graphs. This criterium is allowing a maximum of 1 out of 3 answers to be wrong. Participants are expected not to have paid attention to the questions when they answered wrong. This criterium lead to 3 exclusions.

Participant private ID	Minimum number of correct answers
513904	0
514001	0
498456	1

Data quality criterium 3: Questionnaire completion time

The completion time for the first UX questionnaire should be long enough, to filter out the participants that just randomly clicked and not actually answered the questions. The following thresholds for completion time are based on a study by Slattery and Yates (Slattery & Yates, 2018), in which they show how their fastest reading participant reads 246 words per minute, while still being accurate. In addition, participants are estimated to need at least half a second to click the answer on the 7-point Likert scale. The calculations of these thresholds can be found in the appendix. In total, the first time the participants see the UX questionnaire, they should at least take:

- 28.68s to complete the UX questionnaire on product qualities (iteration 1), leading to 2 exclusions. *Iteration 1 (aesthetics, usability, utility): questions: 12, words: 93. 60 / (246 / 93) = 22.683 + 12*0.5 = **28.683***
- 26.02s to complete the UX questionnaire on emotions (iteration 2), leading to 3 exclusions. *Iteration 2 (emotions): questions: 14, words: 78. 60 / (246 / 78) = 19.024 + 14*0.5 = **26.024***

Participant private ID	UX questionnaire duration (ms)
498463	18001
498547	25633
514001	20473
513779	22470
513867	24273

After the first time that participants have seen the questionnaire, they didn't have to read every single word again in order to answer the questions. Therefore, another threshold for exclusion was calculated for the completion time of the UX questionnaire when participants had already seen it. This is based on identifying a question based on reading at least one word and answering a question in at least half a second. In total, they should at least take:

- 8.93s to complete the UX questionnaire on product qualities (iteration 1), leading to 3 exclusions. *60 / (246 / 12) = 2.927 + 12*0.5 = **8.927***
- 10.42s to complete the UX questionnaire on emotion (iteration 2), leading to 3 exclusions. *I2: 60 / (246 / 14) = 3.415 + 14*0.5 = **10.415***

Participant private ID	UX questionnaire duration (ms)
498463	6.610
513867	8.414
498567	8.662
498407	8.729
513809	9.183
513885	10.036

V. Informed consent

Animations in graphs

Welcome to this experiment, where we measure the user experience and quality of data visualizations, and how animation influence this. Before taking part in this study, please read the consent form below and check the box at the bottom of the page if you understand the statements and freely consent to participate in the study.

Consent form

This study involves a web-based experiment designed to measure how data visualizations effect user experience. The study is being conducted by Ward Venrooij and Tijmen van Willigen of TNO, and it has been approved by the TNO Review Board. No deception is involved, and the study involves no more than minimal risk to participants (i.e., the level of risk encountered in daily life). In this study you will be asked to read specific information from **6 data visualizations** (2 groups of 3 similar visualizations). The visualizations can represent any subject, for example the population numbers of countries. The tasks that need to be completed relate to this visualization. You will be asked to read the information from the visualization. The visualizations are accompanied with an explanation. After each time finishing the tasks of the visualization, you will receive the same user experience questionnaire regarding your interaction with the data visualizations and tasks. The questionnaire consists of **11 questions** that relate to your experience with the visualization. No personal or sensitive questions are asked; the data is therefore strictly anonymous. All responses are treated as confidential, and in no case will responses from individual participants be identified. Rather, all data will be pooled and published in aggregate form only.

The goal of this study is to asses a method of evaluation of User Experience in the context of data visualizations. With the results of this research, we will also be able to gain insight in the user experience of data visualizations, and how this experience can be affected by animations within a visualization. Participants of this study should be older than 18 and younger than 65. Participation in the study typically takes **20 minutes**.

Via Prolific, participants will receive an incentive of **£2**, with a bonus of **£1** if all questions about the data visualizations are answered correctly (they are simple questions). Participation is voluntary, refusal to take part in the study involves no penalty. Participants may withdraw from the study at any time but will in that case not receive the incentive as mentioned above. Incomplete or random submissions will also be rejected and in these cases no reward will be rendered to the participant.

If participants have further questions about this study or their rights, or if they wish to lodge a complaint or concern, they may contact the researchers at tijmen.vanwilligen@tno.nl

☐ **I am between 18 and 65 years of age, understand the statements above, and freely consent to participate in the study.**

Next

VI. SPSS CODE

Repeated measures ANOVA

```
GLM <condition 1> <condition 2> <condition 3>
  /WSFACTOR=<factor> 3 Polynomial
  /METHOD=SSTYPE(3)
  /PLOT=PROFILE(<factor>) TYPE=LINE ERRORBAR=CI MEANREFERENCE=NO YAXIS=AUTO
  /EMMEANS=TABLES(<factor>) COMPARE ADJ(BONFERRONI)
  /PRINT=DESCRIPTIVE ETASQ PARAMETER HOMOGENEITY
  /PLOT=RESIDUALS
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=<factor>.
```

VII. Qualitative data mappings

Task Name	Remarks Visualization	
no loading animation	-	0
	An option to visualize both cocaine and marihuana together would be nice, as it would show that as one decreased, the other increased; lines are also good to s..	1
	Change the colours of each differential and when asking questions mark the target diffential in bold in the same colour as the bar chart	0
	Easy to interpret. Simple. Not very creative or exciting - there may have been a better way of presenting the information that allowed both drugs to be present..	0
	Easy to read and understand	1
	I like the simple transition. It focuses on the data and isn't flashy. I don't think this is part of this study, but I didn't like that the range on the y-axis was so diffe..	1
	I think the spelling was wrong on marijuana	0
	Is a little bit boring if the bars do not move	-1
	It was easy to read the graphs and switch between the two. It didn't take long to find out the answers to the questions.	1
	Not very interesting to look at, but illustrates data well	-1
	The bars didn't hop around - that was much easier to see.	2
	Yep, that's two graphs.	0
Bouncy loading animation	-	0
	All previous remarks apply, but I'd like to add: it is not clear which button at the bottom is selected (they're both two shades of gray) - an outline or a more obvi..	0
	Animation was a bit too much	-1
	Change the colours of each differential and when asking questions mark the target diffential in bold in the same colour as the bar chart	0
	Easy to use, I liked that it moved slightly more than the others. More visual.	1
	Honestly the animation is actually banging, love it. But euphoric? I don't know whether you are Dave Gorman or not (Youtube him) but he's the only man I know..	2
	I didn't like the animation of the bars - the springy-ness was a little off putting	-1
	I prefered the first transition over the second one, I found the "pop-up" aspect slightly irritating	-2
	I wasn't a big fan of each of the bars popping up in a sequence	-1
	inability to see the data on a single chart combined with differing scales made use of these charts frustrating and required too much attention to glean useful c..	-1
	It is a great way of learning data and presenting it, great job	2
	once again I had to refresh the page over and over to view the visual.	0
	Quite stressful to look at	-2
	represented the data really well.	1
	same remark, both bars should be on the same chart, with differentiated axes	0
	Terrible - too jumpy horrid	-2
	The animation was annoying	-2
	The animation wasn't helpful. It was mostly distracting and a bit gimmicky.	-1
	The effect placed on the chart only delays its reading and adds nothing new.	-1
calm loading animation	-	0
	Again, easy to use and find the answers to the questions.	1
	Bit boring/very standard but effective	0
	Definite improvement on the static graphs, personally I prefer the first one but I think they have their uses. I see the first one as being more appropriate for me..	1
	I like this one better than the first, but the second was my preferred transition. Making the transition 'flashier' doesn't help me interact with the data.	1
	I prefer the animation on this one - a bit calmer especially as it suits the subject matter	2
	It was broken. I had to refresh the page several times to even view it.	0
	It's a bit unattractive to look at - literally doesn't attract you to it, but once you have concentrated and read it through, operationally it is great. the emboldene..	1
	it's so easy to understand the graph	1
	Not creative or particularly attractive, however it serves its intended purpose well	1
	Numbers could maybe be bigger	0
	Pleasant to look at.	2
	Same as before, easy to read and understand	1
	The animation was much more appealing	2
	The chart had no artifice, but it was objective in its content.	0
	The lettering size on the right could be bigger, I've used the browser zoom.	0
	The same remarks from the previous graph apply: lines would show progress and, if tranquilizers and pain meds were shown together, the relationship betwee..	1
	the text was very small in questions 2 and 3	0
	we should not have to switch between both sets, include both on the same graph for easier access to data	0

Direct animated transition	-	0
	As I read the data of this type of chart it becomes faster my understanding of its operation.	1
	Great. Meets the purpose of statistics: making complex data very readable.	2
	I didn't understand why there were 2 visualizations of the same thing but in answering the question I could see why it would be needed and I liked it.	1
	It was a bit harder to use at first, but easy once I figured out what I needed to be looking at.	1
	Not as good as the stacking animation, it felt very odd seeing the shapes mixing over each other and looking like they were just tweened together. Really didn't..	-2
	One thing i thought about after seeing the colours for a second time is that someone who is colour blind may have trouble distinguishing between the purple an..	0
	That was much easier to understand than the previous option.	2
	The fact that elements intersect makes it look a little less clean when transitioning to the different states, in particular the blue bar	-1
	The totals graph is a bit of an eyesore to look at.	-1
	This is my favorite transition of the three. It helps show how the graphs relate to each other, but is smoother/clearer than the second.	2
non-animated transition	-	0
	Easy to use, simple to understand, but transition is jarring and not attractive.	-1
	I hate the colors, I hate the visualization and you've spoiled me with animation now. Overall it just annoys me and while I can see you can condense the data do..	-1
	information is fragmented and requires more than a glance to understand	-1
	It was easy to understand how to go between the two and interact with it - probably not the colour scheme I would have gone for though.	0
	Nice use of colour	1
	Previous remarks apply, and let me add that this is all really interesting info. I like that it's "per person", so a comparison is possible despite big differences in ..	1
	The visualization was clear. I liked that the two graphs showed interesting, though different, ways to display information.	2
	You don't actually need to switch between the graphs for the answers. The first one told me what I needed to know.	0
Staged animated transition	-	0
	As I said in the previous one I think the data could be represented better and I hate the colours. But not gonna lie, I did a small "Oooo" when it did its fancy stac..	2
	Better after several utilisation, but at first, I couldn't understand it	0
	I didn't like the transition, a bit annoying and takes longer to see	-2
	I liked that this transition helped clearly show how the two charts relate to each other. This transition made it easy to see that both were displaying the same d..	2
	It was hard to tell the colours blue apart so had to keep referring to the key	-1
	It was hard to tell the difference between the graphs (even though one was a clustered bar)	-1
	Nice and clear and easy to understand.	2
	Ok in this one I really appreciated the animation. It might even help with the questions, as the bars are moved around to their new places and you can keep your..	2
	Simple to use once you've figured out where to look.	1
	The colors and animation are good.	2
	the colours used were not friendly on the eyes	-1
	The transition grew on me	2
	the use of the term "pure alcohol" was confusing	0
	too distracting	-2
	Very neat and useful animation. Colours could be more attractive, but the animations are smooth and easy to understand. More space between the countries is..	1