Usability evaluation of focus & blur highlighting

GEO 511 Master Thesis
GIScience Center GIVA
University of Zurich
31.01.2013

Kaspar Fischer
07-714-199

Advisor

Dr. Arzu Çöltekin

Faculty Member

Prof. Dr. Sara Irina Fabrikant

Co-Advisors

Dr. Amy Griffin
School of Physical, Environmental and
Mathematical Sciences
UNSW
Canberra BC 2610, Australia
a.griffin@adfa.edu.au

&

Dr. Anthony C. Robinson
Pennsylvania State University
430 Earth and Engineering Sciences Building
John A. Dutton e-Education Institute
University Park, PA 16802, USA
arobinson@psu.edu

Abstract

This thesis presents a usability evaluation of the highlighting method focus & blur. Based on the recommendations of Robinson (2011) and the works of Griffin and Robinson (2010), a user study has been carried out to find out whether focus & blur highlighting is significantly different from color highlighting in efficiency, effectiveness and satisfaction. 31 participants were asked to perform four tasks in a two-view linked display with a map and a graph. The user study was performed under the recording of an eye-tracking device to provide AOI based analysis. Additionally, an ideal level of blur has been assessed by using a short visual search task. The visual variable blur has proved to be a salient (Ware 2012), preattentive (Kosara et al. 2002b), and a probable feature for guidance of attention (Wolfe & Horowitz 2004). The findings of this study are in line with recent research; focus & blur highlighting is not significantly different from color highlighting in terms of efficiency, effectiveness and satisfaction. Participants who were shown the focus & blur highlighting stimuli found the method visually pleasing and felt confident about using it for simple visual analytics tasks. A first basis for an ideal blur level has also been defined and can be used for further research.

Table of Contents

List of Figures

List of Tables

1 Introduction

Motivation

Highlighting is a visualization method which aims to emphasize and guide a viewer's attention to the highlighted object. We experience highlighting in our daily lives, be it at the supermarket, where the 30% off sign is visualized with bright red and conspicuously oversized letters, or the travel advertisement sign at your local bus stop, where the mind-numbingly low air travel fares are emphasized with bright yellow brushstrokes. We are so ubiquitously surrounded by these highlighted objects that we are not even consciously aware of them anymore. Nonetheless, many of us use highlighting techniques even ourselves, or at least have at some point, when reading books for school or university. The Stabilo[1], marker for example, is a highlighting tool which is used to emphasize important words or passages in a text and makes them immediately pop out whenever we have a look at them again. Adding high contrast color to an object is simple enough. So, what is the big deal about highlighting?

Highlighting is used for the communication and analysis of information. In times where data is so abundant, it is crucial that we have systems in place that are capable of handling the information overflow. There are of course several solutions. Automated analyzing tools, for example, will apply statistical analysis to find correlations. However, they only work when the problem is understood beforehand (Keim et al. 2010). Keim et al. argue that visual analytics is another way to cope with the aforementioned problem. It is the method of visualizing complex data sets in different representations and multiple views (Kerren et al. 2007). This

1 www.stabilo.com/pages-com/highlighting

allows for better understanding and analysis of data and supports decision making. A fundamental interactive behavior of such linking systems is the use of a transient visual link: highlighting. Selecting parts or clusters of data in one visualization will highlight the corresponding information in all the other views. Highlighting, in other words, gives the process of linking information a visual feedback (Robinson 2011).

The highlighting approach in these information visualization systems is currently most commonly limited to single color and value overlays (Robinson 2006). However, with more complex and multivariate data (Thomas & Cook 2005), color as the only highlighting approach may not always be good enough. For example, color used as a classification variable (Bertin 1983/2010) for an additional attribute would give the visualization yet another dimension. Color as the main highlighting method would not be the best solution then. Not only would the guidance of attention be reduced when color is also used for classification (Ware 2012, Wolfe & Horowitz 2004), but it would moreover not preserve the classification symbology of the highlighted object. To expand the toolbox for highlighting and to support visualization systems with different approaches to link information, alternative highlighting methods should, therefore, be sought after and empirically assessed for their efficiency, effectiveness and satisfaction.

Based on the criteria for highlighting methods in Robinson's work (2011) and the list of variables that draw the viewer's attention (Wolfe and Horowitz 2004), blur is a probable candidate for highlighting. Early empirical research by Kosara et al. (2002b) showed that blur, or semantic depth of field (SDOF), performed as efficiently as color for visual search tasks. Because blur as a highlighting method imitates how the human visual system (HVS) works, by having a sharp foveal and a blurred pararfoveal field of view (Ware 2012), it is a natural choice for a highlighting method to emphasize important areas or objects. Since the object of importance is visualized in focus while the surrounding context is blurred, the highlighting method is hereafter named «focus & blur» in this thesis. Efficiency, effectiveness and satisfaction are measured in this study, as they represent important factors for a usability assessment (Nielsen 1993). The user study's stimuli are based on the ones that Griffin and Robinson (2010) used to test a leader line highlighting approach. However, in this study, the stimuli are not static but allow for limited interactivity by enabling the highlighting method on a mouseover event. Additionally, the study will be recorded by an eye-tracking sensor to evaluate the effectiveness of the highlighting method's guidance of attention. Eye-tracking data can give insight into why the highlighting approach works or not, by revealing transition times and overall fixation lengths or counts on regions of interest (Duchowsky 2007).

The focus & blur highlighting method uses a Gaussian blur filter in this study. In order to be effective, the right level of blur needs to be used in the highlighting method. Kosara et al. (2002a,b) noted that a simple pixel value parameter for blur level is not universally applicable due to differing display values and technologies. In this thesis, the level of blur is therefore measured in relation to the thickness of the object's outline that is to be highlighted. The ideal level of blur is thus assessed by a small exploratory test.

Research Questions

In order to tackle the question of whether focus & blur highlighting is indeed a valuable alternative to color highlighting, the following research questions are proposed:

- **RQ1:** What is the ideal level of blur for focus & blur highlighting with regard to efficiency and effectiveness of object identification and context awareness?
- **RQ2:** Is there a significant difference in efficiency and effectiveness between focus & blur and color highlighting to support visual analytics tasks in coordinated displays?
- **RQ3:** Is there a significant difference in satisfaction between focus & blur and color highlighting when used in visual analytics tasks?

Based on the earlier findings of Kosara et al. (2002a,b), the working hypothesis in this thesis is that there is no significant difference between the two highlighting methods in terms of efficiency, effectiveness and satisfaction.

Thesis Structure

Section 2 introduces the research framework by discussing the human visual system and theory of attention. Moreover, it presents the subject of highlighting in general and the state of the art in highlighting research with a focus on the focus & blur approach. Section 3 describes the methods and details of the user study. Section 4 discusses the general results, while section 5 reveals the findings gained from the eye-tracking device. Section 6 discusses the results of these two sections 4 and 5, with a focus on linking the results with the working hypothesis, research questions and research framework. Finally, section 7 will sum up the thesis. Stimuli and questionnaires from the user study can be found at the very end of the thesis in the appendix.

2 Research Framework

This section presents the research framework. First, the human visual system (HVS) and theories of attention are described. Then, the state of the art in recent highlighting research is presented. Finally, we will have a closer look at the visual variable *blur* for highlighting.

2.1 Human Visual System and Attention

Every human being is, as long as there are no physical deficiencies present, equipped with two eyes. The eye is a complicated organ. Its most important parts are the cornea, pupil, iris, lens and retina (see Figure 1). The retina is the part where the photo-receptors are located, which are sensitive to light. Moreover, the retina itself contains two major photo-receptor cells: the cones and rods. Cones are responsible for color-vision, while rods are responsible for low-light, monochrome vision. (Liebowitch 2005)

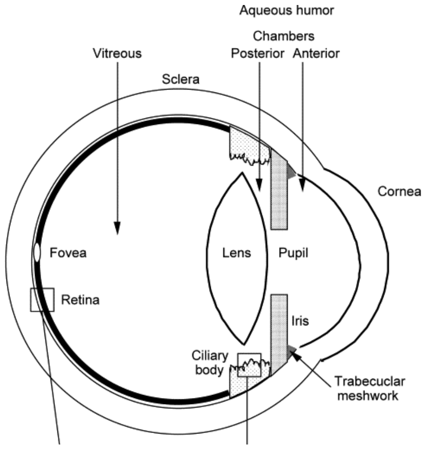


Figure 1: The Eye. Source: Illustration from Liebovitch (2005).

The cones are mostly present within a small area of the retina which is called the *fovea*. In this region the density of cones is highest, while rods are almost non existent. This is where the human eye has the best visual acuity, the ability to resolve fine detail. This ability is measured by arcminute or arcseconds of our visual field, which is roughly 180 degrees. Visual acuity can be more closely defined by point acuity (1 minute arc), grating acuity (1-2 arcminutes), letter acuity (5 arcminutes), stereo acuity (10 arcseconds) and vernier acuity (10 arcseconds). The visual acuity diminishes continually from the fovea to the edges of the retina, the parafovea. This is where mostly rods are present. The rods, albeit not capable of resolving fine detail and color, are very sensitive to low light. What is more, the sensitivity to movement is higher in the peripheral regions of our vision. (Ware 2012, Russ 2006)

The fact that the region of high resolution is constrained to the fovea, which is a very small region of our field of view, leaves the human being with a very limited optical tool, compared to a present day digital camera, for example. Nonetheless, in our everyday lives, humans are hardly aware of these limitations. For example, the image that is projected to the retina is inverted, but we do not wander around the world with an upside-down view. This is where the brain comes into play. Humans do not perceive what is on the retina, but what the brain makes of it (Ware 2012). Wearing the inverter glasses for longer periods of time will lead to normal view again. The sensory information that is gathered by the eye is just one part of the process to generate an internal representation of the visual world. In fact, the sensory information is transmitted to our brain and undergoes complex computations before we actively become aware of what we see. Now this process is of course quite fast. People can process a complex natural image and decide whether an object is present or not in roughly 150 ms, as research has shown (Thorpe et al. 1996, Vanrullen & Thorpe 2001). The human brain's capabilities of computation fools us into believing that we can see everything quite clearly and amiably resolved, when, in fact, we can only see the two central degrees of our field of view in full clarity. This limitation of information submitted to the brain is of utmost importance. The wealth of information that the surrounding environment provides makes it impossible for the human being to process everything everywhere at any given time (Wolfe & Horowitz 2004). The aforementioned physical limitations of our visual system, such as the constrained visual acuity, are one way to cope with the problem of information processing. But these «front-end» reductions of information fail to solve the problem, as Wolfe and Horowitz (2004) mention in their paper. To deal with the excess flow of information, the human visual system has attentional mechanisms to decide which subset of stimuli are processed more extensively, while the rest remains subject to minimal analysis (Wolfe & Horowitz 2004). But how do we decide what is worth of our attention, or what inevitably draws our attention?

Desimone and Duncen (1995) insinuate:

> «At some point (or several points) between input and response, objects in the visual input compete for representation, analysis or control.»

These objects compete for processing within a network of around 30 or more cortical visual areas (V1, V2, V3, V4 and so on). From the first cortical visual area V1 to the higher stages like V4, the receptive field size and complexity increase. If attention is drawn to a single object, all the cortical areas cooperate to represent the attended object and enter awareness (Yantis 2005). The competition is either won by «bottom-up» or «top-down» or a combination of the two processes. The first way is what geographic information science calls «perception», the «bottom-up» or «low-level» form of processing. In this case, the attention is spontaneously oriented towards an oncoming stimulus. The second is when attention is oriented towards a target by knowledge, expectation and current goals (Egeth & Yantis 1997). This is defined by the term «cognition», a «top-down» or «high-level» form of processing. (Svienty 2008)

Of great interest for this thesis is the «bottom-up» form of attention. In order for the focus & blur highlighting method to be suitable, it needs to draw the attention of the observer. In an early theory about attention, the feature integration theory of Treisman and Gelade (1980) proposes a two stage architecture of attention, where some stimuli are processed in a parallel «preattentive» stage, while others require serial processing in a second stage. Visuals that are processed preattentively are considered «pop-out» or «figure-ground» segregation features and can be recognized in less than 10 milliseconds per item (Treisman & Gormican 1988).

a)	12344863147996321476321214879351321548632156586313546987962134
b)	12344863147996321476321214879351321548632156586313546987962134

Table 1: Example of the preattentive visual variable color. Source: Adapted by Ware (2012).

The search for all 6s in the upper part (a) of Table 1 will take considerate time, while in the lower part (b) all 6s can be processed relatively fast. The color red guides our attention so that we easily find our next fixation point, which is where the next 6 is to be found. The color red, therefore, is a preattentive attribute. The theory of preattentiveness has been revised and improved by many following researchers (Treisman 1985, Wolfe 1994, Egeth & Yantis 1997), who all sticked to the two-stage architecture of attention but proposed that the preattentive stage could guide the attention in the second stage. The feature integration theory of Treisman and Gelade (1980) has been a central work for many years and is still being referred to today. However, as Ware (2003) notes in his review about the feature integration theory, newer research has provided alternative theories to the preattentive proposition, such as the feedback or «re-entrant» pathway theory by Di Lollo et al. (2000) and Di Lollo et al. (2001). The authors discard the notion of a linear processing system and propose that information is passed back and forth between early vision stages and later vision stages. All in all, the theory of attention is still subject of heavy research, where a multitude of obscurities exist (Ware 2012). Also, for purposes of designing attention guiding visuals, the theories of visual processing and attention sometimes fail to provide explanations. Wolfe and Horowitz (2004) enunciate the fact that some attributes are processed in early vision as well as in attentive vision, but do not guide the user's attention (as well as the other way around). Some visual attributes guide the attention more than others and provide a higher degree of «pop-out» (Ware 2012). Therefore, recent research tried to evaluate attention guiding visual attributes by means of empirical studies. Wolfe and Horowitz (2004) for example, asserted which visual attributes guide our attention. They report findings of various empirical studies from different researches in five categories, based on the probability of guidance.

Undoubted attributes	Probable attributes	Possible attributes	Doubtful cases	Probable non-attributes
- Color - Motion - Orientation - Size	- Luminance - Vernier offset - Stereoscopic depth - Pictoral depth cues - Shape - Line termination - Closure - Topological status - Curvature	- Lightening direction - Glossiness - Expansion - Number - Aspect ratio	- Novelty - Letter identity - Alphanumeric category	- Intersection - Optic flow - Color change - Three dimensional volumes - Faces - Your name - Semantic category

Table 2: List of attention-guiding attributes. Source: Adapted by Wolfe & Horowitz (2004).

The list of Wolfe and Horowitz (2004) connotes that focus & blur is a probable attribute of the guidance of attention (pictoral depth cues). Visual isolation and visual levels also play a crucial role in terms attention guidance. In his works about visual variables, Bertin (1967 / 2010) proposed the principles of selectivity and associativity (visual isolation and visual levels). A visual variable is selective when it allows for visual isolation and a visual variable is associative or dissociative when it allows for grouping of sufficiently similar or for segregation of sufficiently different objects (MacEachren 2004). On the one hand, selective visual variables will isolate all objects of this category and let the viewer disregard all other objects with a different category. On the other hand, associativity is needed when a representation contains two components. If a visual variable is associative it will group objects with the same visual variable easily together, despite a visual variation of another attribute. Shape, orientation, color and texture are associative, while value and size are not. Variation of size, which is dissociative, will dominate all other visual variables combined with size, so that grouping within the other visual variable is limited or even prohibited. (Bertin 1967 / 2010)

Isolation and associativity are important factors when two visual attributes are conjoined. One can easily discern the red objects from the black objects in the left view of Figure 2, but the search for red squares is a very slow process. The viewer must serially search all red objects and determine whether they are squares or circles. In the right view of Figure 2, the red squares are quite easily found. Nakayama and Silverman (1986) analyzed conjuncted visual variables and found – in contradiction to the attention theory of Treisman and Gelade (1980) – that stereoscopic depth and color in conjunction may be preattentively processed.

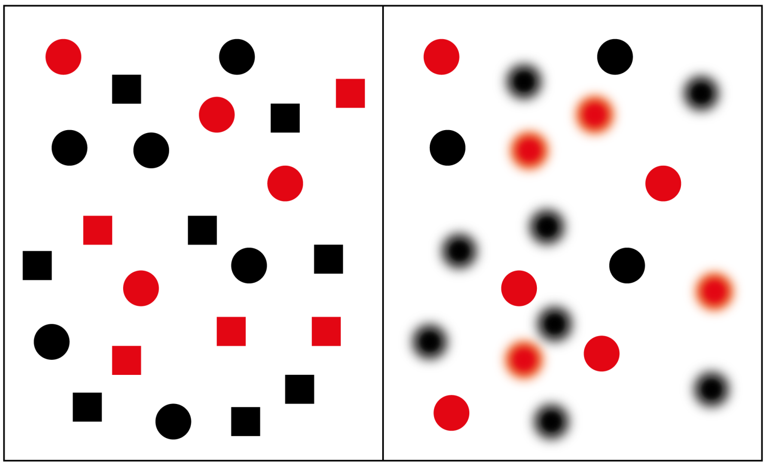


Figure 2: Conjunction of visual variables. Source: Adapted from Ware (2012).

The fact that focus & blur may be processed in an early stage of vision, allows for capable visualizations. Using the focus & blur method for highlighting, allows simultaneous use of color for other visualization means, like the classification of an other attribute.

2.2 Highlighting

What is highlighting?

Highlighting is a commonly used term in computer graphics. However, there seems little agreement on the understanding of highlighting across professionals and non-professionals (Liang & Huang 2010). So then, what is highlighting? This is quite a difficult question to answer, since it means many different things across different domains of research. Becker and Cleveland (1987) called it brushing, which is a way of applying transient paint with a brush on scatterplots. The system worked in the way that when in a scatterplot matrix certain value points are «brushed» with the highlighting method, the corresponding points in the rest of the matrix will be highlighted too. This is an early version of linking data throughout multiple representation dimensions. Later, Liston et al. (2000) defined highlighting as the process of emphasizing by use of visual annotations on either a single display or across multiple displays. The use of highlighting to link semantically paired pieces of information on one or across multiple displays has been utilized and further defined by MacEachren et al. (2003) with the use of the GeoVista studio software[2] and Ware and Bobrow (2005) with node-link diagrams. The use of highlighting in order to link pertinent pieces of information is closely related to the visual information seeking mantra taxonomy that Schneidermann (1996) proposed. In his taxonomy he listed the seven crucial tasks of information seeking: *Overview, Zoom, Filter, Details-on-demand, Relate, History, Extract*. The task *relate* represents the searching for relationships among pieces of information: when the user selects an item the corresponding pieces of information are highlighted (Schneidermann 1996). Others used highlighting as a way of guiding users through graphical interfaces (GUI) and controlling the flow of GUI interactions (Zhai et al. 1997).

2 Hardisty & Robinson (2001)

In conclusion, highlighting is a method to emphasize certain objects, to link these objects with corresponding data on a same or different display, and to guide or control user interaction with a graphical interface system. Similarly to this conclusion, Liang and Huang (2010) tried to define and classify highlighting in a new three way model of highlighting:

1. Viewing control to reduce information overloading and reduce complexity
2. Interaction control to guide information seeking process
3. Visual recommendation for decision making

The classification from Liang and Huang (2010) not only describe the three different fields of application, but also the reasons why highlighting should be used. This leads us to the next question:

Why use highlighting?

The three categories by Liang and Huang (2010) provide a useful framework to distinguish the different purposes and applications of highlighting. It also tackles the question of why highlighting is required or perhaps supportive in visualizations. The first statement of their taxonomy describes the use of highlighting in order to reduce information overload and complexity. Highlighting is used to specifically guide the user's attention (see section 2.1) to the important parts of a visualization and does so by visually enhancing the saliency of the **important** objects. For example, the word important in the last sentence pops out from the rest of the text and may very well be the first word that you have read when turning to this page. The second and third statements refer to the information seeking and decision making process. Highlighting should support the user in gaining information, the ability for inference making and decision making. When using multiple linked information displays, which are often used in visual analytics (Keim et al. 2010), highlighting acts as a linking mechanism between different data dimensions (space, time, and attribute, for example). With regard to our visual short term memory (VSTM) the concept of highlighting may overcome the capacity limitations. The VSTM is thought to be very limited in how many objects can be stored at any given time (Sperling 1960, Lee & Chun 2001, Todd & Marois 2004). While some think the limitation lies around 3 to 5 objects which can be stored simultaneously, also called the 4-chunk theory (Cowan 2000), others suggest that the limitation is not solely based on object count but also on how many features these objects have (Olson & Jiang 2002). Olson and Jiang (2002) contend that not only the count of objects plays a role but that the objects' features play an important part. In their findings, the *weak-object* theory [3] was thought to be the best explanation to our VSTM limitations. The highlighting mechanism, therefore, should alleviate the VSTM capacity limitation by improving the saliency (attention guiding, isolation and visual level) of important objects and linking pertinent sets of data. Hence, the user may externalize the memory (Barkowsky et al. 2005) of important and related pieces of information through highlighting, which should lend to better performances. People would not need to repetitively go back and forth on different data visualizations, because the moment they look at another data representation, they will already have forgotten the original representation. In terms of visual search tasks, the VSTM does not impact the performance. Horowitz and Wolfe (1998) empirically tested how the search performance varies if the scene changes while the search is in progress (simple absence and presence search task). While participants made more mistakes and needed more time when the scene was dynamically changed, the slope from fewer to more distractions suggests that visual search is amnesic and hence has no memory. Nonetheless, highlighting, if preattentively designed, will reduce the time until the highlighted target is spotted (Ware 2012).In the work of Philipsen (1994), the highlighting methods such as reverse value, reverse color hue, the color red and opposite colors were empirically tested for their influence on search performance tasks. The results showed that all highlighting methods performed better than having no highlighting in terms of efficiency (response time) and effectiveness (target miss-rate). Opposite color highlighting was the top performer, but Philipsen (1994) reminds us that the relatively high frequency of color deficiency should not be disregarded when designing visualizations. Second best performer was the use of the color

3 The weak-object theory implies that objects can be grouped into higher order chunks, depending on how their features are conjoined and how different they are from one another (Olson & Jiang 2002).

red, followed by opposite color value.

Which highlighting method should be used?

Up to this day, in most software solutions, the only highlighting method implemented is the use of color hue or brightness value. Software packages such as ArcGIS[4], QGIS[5] or GeoMedia[6] uniformly use color highlighting which can be adjusted to whatever color the user prefers. Additionally, software packages from Adobe, such as Illustrator[7] also provide color highlighting (adjustable hue) but no alternative. Visualizations on the web mostly make use of color highlighting, too.

The efficiency and effectiveness rely heavily on the saliency of the highlighting method. Ware (2012) states that the choice of highlighting is relatively easy for homogeneous displays. Using color highlighting in such a visualization will provide enough saliency in terms of isolation and attention. However, as the visualizations and the data displays become more complex and go from singular view to multiple views handling multivariate data (Thomas & Cook 2005), the choice of highlighting becomes much more difficult to answer. In general, Ware (2012) recommends using the graphical dimension, which is least used otherwise in the design. The commonly used highlighting approach of color might just not be the ideal solution anymore. The visualizations might use color as a visual variable for classification of single or clustered pieces of information (Bertin 1967 / 2010), which would render the use of another color for highlighting less powerful in terms of visual isolation. Context is thus a determining factor when choosing the appropriate highlighting method.

In order to have the possibility to choose from a variety of highlighting methods when designing visualizations, other forms of highlighting approaches should be considered, more closely analyzed and empirically assessed. In Figure 3 alternative highlighting approaches are depicted. Aside from color hue and value, there are saturation, focus & blur, transparency, and

4 www.esri.com/software/arcgis
5 www.qgis.org
6 http://geospatial.intergraph.com/products/GeoMedia/Details.aspx
7 www.adobe.com/products/illustrator.html

texture. In his paper, Robinson (2011) additionally lists resolution, shape, arrangement, orientation, size, location, leader line, contour and style reduction. He evaluates the possible highlighting approaches based on his proposed design criteria. The highlighting should be visually salient (isolation, visual levels and attention), usable across a broad range of representational forms (i.e. point, line, shape and text) and suitable for implementation without performance penalties. The method should also preserve observation size and shape, observation symbolization and nearby observations for context. In this theoretical assessment, focus & blur achieves good marks for all design criteria other than saliency (guidance of attention, based on Wolfe & Horowitz 2004) and context preservation, where Robinson (2011) thought it to be only marginally effective.

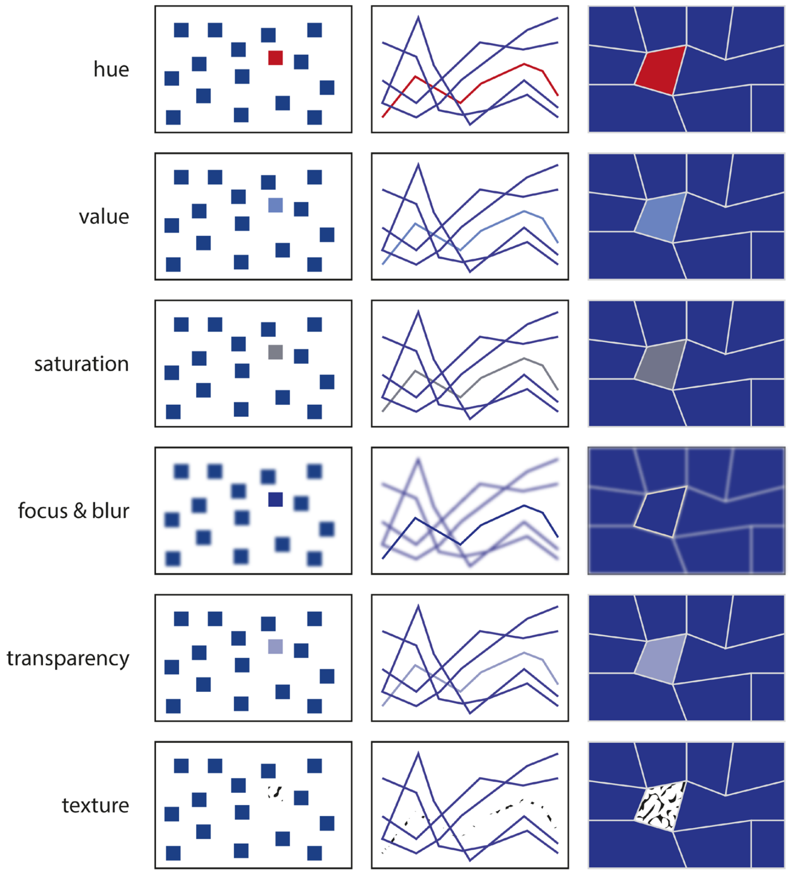


Figure 3: Highlighting methods based on common visual variables. Adapted from Robinson (2011).

Some of the highlighting methods have been empirically tested. Wickens et al. (2004) tested intensity coding (highlighting and lowlighting in their terms) for visual search through maps and found intensity to be useful as a guidance of attention, but not as a discriminator between classes. They also concluded that the intensity values do not have to be too large. Unfortunately, Wickens et al. (2004) did not compare their approach directly to the commonly used color highlighting or against the lack of highlighting.

A paper by Waldner et al. (2010), moreover, proposed visual links to connect related pieces of data across multiple windows in addition to targets highlighted by a yellow background. They shortly evaluated the proposed method by 7 participants with static stimuli and came to the conclusion that the visual links guide the user's attention and make relevant information in additional windows more distinct. In an updated version of their visual links Steinberger et al. (2011) proposed context-preserving visual links, which avoid overlapping with underlying information (if possible). Additionally, they also tested their design against simple leader lines and normal color highlighting (in red) with a small user study (N = 18) using static stimuli. They found the context-preserving links and the normal links to have a significant better response time performance than color. What is more, the context-preserving links were rated as the most visually pleasing design solution. As the test setup was composed of Google Maps[8] and Wikipedia[9] text, it remains to be seen whether the results of context-preserving leader lines are transferable to large data sets with map and multiple graphs (time-lines, scatter-plots and PCP graph). Another interesting application using actual visual links was proposed by Collins et al. (2007), where they implemented the data visualization in a 3D virtual desktop, called VisLink. The system looks incredibly fancy, however it has not yet been assessed for usability.

In another evaluation of alternative highlighting methods, Griffin and Robinson (2010) tested a leader line approach against color highlighting (red). The leader line approach connects the relevant information with a red line over multiple views. In their evaluation, the

8 www.maps.google.com
9 www.wikipedia.com

authors designed a two view static display using a map and either PCP or scatter-plot. In their preliminary results of the empirical study (N=8), participants had slightly lower mean absolute value estimation errors for the leader line approach and the mean response times were slightly faster, but both not significantly. Nonetheless, their report indicates that leader lines are comparable to color highlighting if not slightly better.

In a series of experiments Ware and Bobrow (2004, 2005) tested static colored highlighting against the use of motion or a combination of the two in node-link graphs. Participants needed to perform several tasks, such as answering whether a common node between two subgraphs existed or not, using the node-link graphs. The authors found that the motion highlighting was equally efficient as the static color highlighting and that the combination of the two was also a similarly viable option.

The highlighting method of using blur has received some attention lately. The authors of Giller et al. (2001) and Kosara et al. (2001) focused their research on the basics of how blur works (the concept of focus & blur is more closely looked at in section 2.3. Focus & Blur) and found that blur or DOF (depth of field) can be equally efficient as color highlighting.

In conclusion, there are many alternatives for highlighting other than color. Some early empirical studies showed promising results. Nonetheless, further research is clearly needed to evaluate alternative approaches and to find out how they perform and what advantages or disadvantages they bring along.

2.3 Focus & Blur

In this thesis the proposed highlighting method is focus & blur. Focus or depth of field (DOF) is a term that is widely used in photography. When the photographer uses the right combination of focal length and aperture, an object that is clearly separated from the background is sharp and in focus, while the background is blurred out, also called *bokeh*. The human visual system works similarly to a photo camera. Something that is aligned with the fovea is sharp while with radial distance everything becomes blurry (see section 2.1).

Figure 4: Depth of field in a photograph. Chili oil for Chengdu Hotpot, China.

Another form of application where blur is used is called *foveation*. Foveation is the term for biologically motivated image compression where only the relevant part is in focus while the rest is blurred. Çöltekin (2006) analysed foveation for 3D visualizations and stereo imaging and created a program called Foveaglyph, which allowed the creation of 3D or 2D foveated images by taking stereo pair images as input. Çöltekin (2006) also discussed different possible fields of use for the foveation model. One of these fields recommended, was the guidance of attention, as it could be used in 3D cinema. Foveation may also be used in visualizations as a means of reducing data load for better performance, which could possibly be used in conjunction with an eye-tracker for live view visualizations, where the focused area follows the users gaze pattern. Others propose the use of DOF to reduce data load for 3D terrain models in cartographic applications (Häberling 2003). Visualizing virtual environment requires high amounts of processing power, thus blurring objects based on their distance to the viewer can reduce the workload. This is a method that is already frequently used in 3D games.

Highlighting something with focus & blur produces a depth cue which makes the highlighted object pop out. It simulates the natural viewing behavior when we focus an object with our eyes. The surroundings are blurred while the object of attention is sharp. While the application of focus & blur is clearly a preattentive feature (Kosara 2001, Kosara et al. 2001), it is also a probable method to guide attention (see Table 2, Wolfe & Horowitz 2004).

The finding of Nakayama and Silverman (1986) that focus & blur can be used in conjunction with color (as in Figure 2) has been confirmed by newer studies from Kosara et al. (2001, 2002b). Later, Giller et al. (2001) and Kosara et al. (2002a, 2002b) used depth of field (blur) to provide «literal» focus+context[10] and renamed it semantic depth of field (SDOF). Objects are blurred based on their relevance rather than on their distance and thus represent a form of highlighting. The authors tested the performance of SDOF in terms of preattentiveness[11], the conjunction with color, and as a fully fledged visualization dimension. SDOF was reported to be preattentive and yielded no significantly less efficient results than color for search tasks. However, SDOF did not perform satisfyingly enough as a fully fledged visualization dimension. Participants could distinguish between objects of different blur levels, but were unable to cluster objects of the same blur level Kosara et al. (2002b). Also, participants disliked to look at images with different levels of blur, for it made them tired.

10 Focus+context is a visualization term that describes the methodology of providing a focused (typically enlarged, zoomed) view of an area of interest, while simultaneously providing context of the surroundings (Baudisch et al. 2001).

11 The notion of preattentiveness is still used in literature nowadays, even if the theory has been highly criticized (Di Lollo et al. 2001, Wolfe and Horowitz 2004).

3 Methods

3.1 Participants

For this user study, 36 participants have agreed to voluntarily take part in the experiment and were given a piece of Lindt chocolate as the only form of compensation. Unfortunately, out of these 36 participants only 31 recordings were suitable for analysis (N = 31, 9 females and 22 males). The five participants had to be rejected due to impairing visual corrections (N = 5) and one participant due to being cross-eyed. Visual corrections provided problems when the diopter rating was well above -6 and in connection with astigmatism.

	All participants	Focus & Blur	Color
Thematic maps	3.5	3.4	3.5
GIS systems	3.9	3.8	4.0
3D games	2.1	1.9	2.3
English	3.8	4.0	3.7

Table 3: Participant information with regard to their experience in related areas. Likert scale (1 to 5).

Participants were in average 25 years old and none of them had color vision deficiency. Furthermore, all participants were either students or doctorates from the Faculty of Science from University of Zurich. In general, they all had to some extent experience using maps and GIS software in their daily life. Participants were also asked about their video games consumption. Table 3 illustrates the findings of the participant information questionnaire, in which participants were asked to answer their experience in related areas to their best knowledge. A Likert scale (1 to 5) was used, where 1 means no experience and 5 means professional experience. In terms of games consumption, 1 means no gaming at all, while a 5 means daily use.

3.2 Apparatus

The user study was conducted using Tobii Studio analysis software 3.0.2[12] and the Tobii X120 infrared sensor at 60 Hz to recognize fixations. The Dell UltraSharp 2007FP display measured 20inch and the resolution was set to 1280x960 pixels. Tobii Studio handled the presentation of the stimuli and recorded the fixations as raw X,Y points. For the analysis of the the gaze data a ClearView fixation filter with 50 pixels radius and minimal fixation duration of 100 milliseconds was used. Additionally, video and audio were recorded in order to elicit the exact time of task completion for each of the participants.

Figure 5: Eye-tracking Lab at University of Zurich

12 Tobii Technology AB, Sweden

The eye-tracking device is situated in a controlled environment. For each participant the environmental factors were identical: location of sensor and display, illumination of the room, and position of screen and chair. This is important to ensure that the external effects remained constant throughout the whole user study phase and that the eye-tracking results were optimal and comparable (Duchowsky 2007).

3.3 Design

3.3.1 Test 1

In a quick exploratory experiment participants were shown one of four stimuli containing different levels of blur. The purpose of this user study was to identify an ideal level of blur (radius r in a Gaussian blur filter). In a controlled four to one factorial between-subject design, participants were asked to identify the highlighted object within a container and count how many of the same object were present. The task required participants to search and identify the highlighted symbol and make use of the context to count how many of the same symbol were present. The independent variables are the four levels of blur and the task. The dependent variables are efficiency and effectiveness.

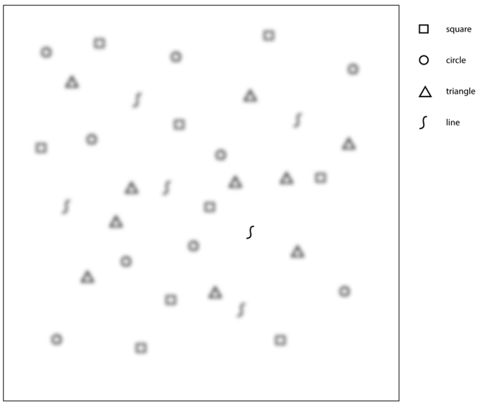


Figure 6: Example stimuli from Test 1. Blur level 3.

3.3.2 Test 2

This test was designed to evaluate the two different highlighting methods: focus & blur and color. In a controlled two by four factorial between-subject design, the efficiency, effectiveness and satisfaction (EES) of both highlighting approaches were assessed. Four different tasks needed to be completed on either highlighting method. The underlying data for the stimuli describe overall crime rate, measured in crimes per 1000 people, for the boroughs[13] of London in a timeline graph from the year 2000 to 2010 and five different crime rate categories in a PCP (parallel coordinate plot) graph. The stimuli were designed to be interactive in a limited form. Users may use the mouse to mouseover the boroughs of London and the graph plots, which will then highlight the corresponding piece of information in the other view. Without a mouseover event, no highlighting will be present. Users may also left-click on an area on the map to freeze the highlighting and display multiple instances of highlighting at the same time.

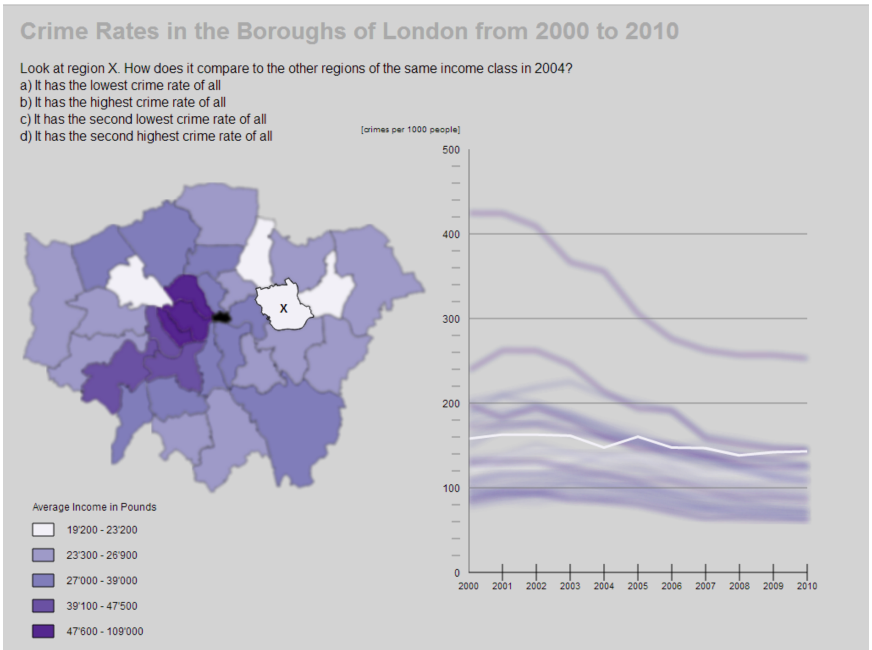


Figure 7: Example stimuli of Test 2. Map and timeline plot with focus & blur highlighting, task 2.

13 The boroughs of London were referred to as «regions» in the task instruction for better comprehension

The tasks

1. In which year did region X fall below a crime rate of 100 for the first time?
2. Look at region X. How does it compare to the other regions of the same income class in 2004?
3. Which region has the third highest crime rate for violence?
4. Compare regions A and B. How much is the difference for robbery crime rate?

Task 1 was designed as a simple value search question, which is a common task for a visual analytics situation (KEIM ET AL. 2010). Task 2, however, fared a higher complexity, for the participants needed to compare three regions based on the income classification and put the target region in context to these other regions. Task 2 looked therefore into the contextualization impairments of the highlighting approaches. Task 3 and 4 follow the scheme of task 1 and 2, albeit using a PCP instead of a time-line graph. In task 3 the value search question is reverted and the participants needed to go from graph to map.

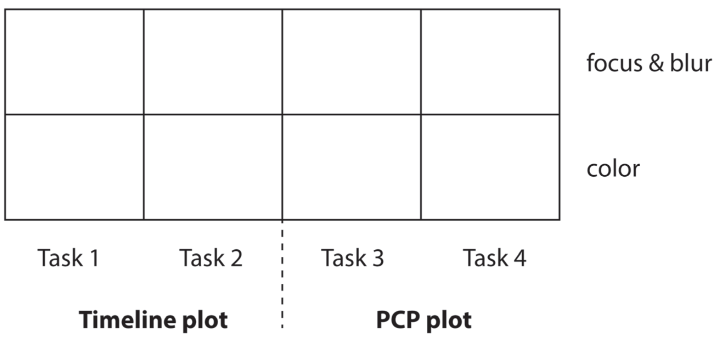


Figure 8: Test 2 experiment design matrix.

The independent variables are the four different tasks and the two highlighting methods: focus & blur and color. The dependent variables are efficiency, effectiveness and satisfaction. The four tasks were, moreover, randomly ordered to circumvent learning bias so that each of the tasks have been equally put in first, second, third or last place.

3.4 SUS Questionnaire

As a last step in the user study, participants were given an SUS (system usability scale) questionnaire to fill out. SUS is a likert-scale, where people indicate on ten standard questions whether they agree or disagree with a given statement (Brooke 1996). The answers are encoded and yield a final usability score from 0 to 100, where 100 is best. The contribution to the final score are the following: odd numbers yield scale position minus 1 and even numbers yield 5 minus the scale position. The sum of the scores are then multiplied by 2.5 for the final score (Brooke 1996). Five of the ten questions are negatively formulated and ordered in an alternating way with positively formulated statements. An SUS with negatively and positively worded items force participants to sometimes agree and sometimes disagree with the statements. This was designed to avoid the common occurrence of acquiescence bias (Nunally 1978). However, as recent research has shown, having negatively worded statements increases mistakes on the participants side and coding errors on the researchers side and thus cancels out the advantage of not having an acquiescence bias (Sauro & Lewis 2011). On the contrary, Sauro and Lewis also mention that when performing moderated experiments and users are briefed, fill-out mistakes can be reduced to a minimum, which have no to minimal impact on the overall SUS score. The SUS is foremost a reliable and low-cost usability scale, which has been widely used to test the usability of systems (Sauro & Lewis 2011, Brooke 1996). Therefore, in this thesis an adapted SUS (see Apendix) is used to measure the subjective usability, or satisfaction, of the two highlighting methods. The adaptation changed some statements in order to better describe the highlighting approaches, rather than full software solutions.

3.5 Materials

3.5.1 Test 1

The stimuli in the first short test was designed using Adobe Illustrator software[14], with which a Gaussian filter could be defined on a pixel basis. The four different blur levels were designed as a relative value to the highlighted objects. Kosara et al. (2002b) mention that defining blur levels on a pixel basis is not recommended, because varying display technologies and viewing circumstances can change the appearance of the applied blur level. Therefore, this approach tried to define blur level based on the thickness of the symbols' lines. The symbols' line thickness measured 2 pixels and the Gaussian blur filter was applied using either a 1, 2, 3 or 4 pixel radius. Thus, the blur filters strength is either 50%, 100%, 150% or 200% of the lines' thickness as long as the document's density is exactly 72ppi[15] which had been used for this experiment. This approach may be applied to all objects which have an outline that is visually separable from the object's filling. Otherwise the blur level should be measured with regard to its mean diameter.

3.5.2 Test 2

The second test contained a two view linked visualization with a map of the boroughs of London and a related time-line or PCP to describe the crime rates in each of the boroughs. The boroughs were, moreover, classified using the average income to discern more affluent from lacking regions. Although most participants knew London, they were not familiar enough with the geographies to have a substantial foreknowledge which would have biased the results. Most participants did not even know what political entity the boroughs of London are.

The stimuli were created using HTML 5 capabilities of SVG (scalable vector graphics) and Javascript. The data for the map of the boroughs of London was taken from the the Ordnance

14 www.adobe.com/products/illustrator
15 If ppi is higher, the blur radius is can be calculated like this: objects line thickness * blur level * (new-ppi / 72ppi)

Survey of England[16]. The crime rate data was provided from the Greater London Authority[17]. In order to have maximum control over the stimuli, the design made from scratch in SVG and Javascript provided to be useful. Complexity and distractions from the actual stimuli could be held to a minimum, which should allow for preciser evaluation of the highlighting methods.

An important factor in creating the stimuli was the design of the blur filter. The World Wide Web Consortium (W3C)[18] provides recommendations for SVG filters. A commonly used blur filter in SVG is the "feGaussianBlur", which is a kernel that applies an approximation of the normalized convolution.

```
<filter id="f1" x="0" y="0">
    <feGaussianBlur in="SourceGraphic" stdDeviation="180" />
</filter>
```

The standard deviation describes the blur filters radius. The notion *in* characterizes which input object the filter will be applied upon. The blur filter is called when the user touches an object with the mouse pointer (mouseover). The mouseover call, furthermore, needs to be annotated in the SVG object where the mouseover is required. The annotation must additionally indicate what function is called when a mouseover is executed:

```
onmouseover="highlight(evt)"
```

The mouseover call refers to the function *highlight*, where the *(evt)* defines which variable is forwarded to the function. In this case it is the event(-target) which defines the object where the mouseover happened. The function then executes the blur filter, or for the color highlighting version a new color assignment. Because there are many data lines in the PCP or time-line plot which overlap, the function needs to clone the object before highlighting and append it at

16 www.ordnancesurvey.co.uk
17 data.london.gov.uk
18 www.w3.org

a higher draw hierarchy in the SVG file. This will ensure that every highlighted object is fully visible and not concealed or partially concealed by another object. A sample javascript code for the highlighting function is attached in the Appendix.

3.6 Procedure

The user study was performed in a special eye-movement laboratory room at the University of Zurich. Before every participant, hardware and settings of Tobii Studio were checked for proper handling. Upon arrival of a participant, they were greeted and asked to sit down at a separate table next to the eye-tracking device. Before starting the experiment, test users were given a consent form, in which they were informed about the test procedure, safety, privacy and their right to withdraw their consent at any time. Participants subsequently needed to fill out a personal questionnaire (see Appendix), where they were asked about their name, sex, age, visual impairments and field of study. Furthermore, participants were tested for color vision deficiency. They were also asked about their daily use of thematic maps, GIS, proficiency of the English language and their video games consumption. After having completed the latter steps, participants were asked to sit down in front of the eye-tracking device, where the calibration for the eye-tracking was performed. The calibration required from each test user to visually follow a colored circle on the screen with their eyes only. Granted that the calibration was successful, participants were informed that video and audio would be recorded and that the task completion time would be measured. Finally, the testing could begin.

In the first short test, participants had to identify a symbol, which was highlighted using the focus & blur method, within a frame containing different kinds of symbols. Ensuing the identification of the symbol, test users were asked to count how many of the same kind of symbol category were present within the frame. The answers were to be given verbally. Following an answer, participants had to indicate their confidence on a scale 1 to 5, where 1 was being least confident and 5 the most confident.

In the second part of the user study, participants had to perform four different tasks (see 3.3.2) using a two view (map and diagram) display. For each task the time needed for completion and the correctness was reported. All participants were shortly trained on a sample stimuli, so that everyone knew what interactions were possible (mouseover and klick to freeze highlighting) The tasks were loaded up in a fresh browser tab in Google Chrome[19] after participants read the introduction text in Tobii Studio . When they completed the task and verbally reported their answer, the browser needed to be closed by the participant in order to load up the confidence question (Likert scale, 1-5), which was integrated in Tobii Studio. Having filled out the confidence question, a mouse click brought the participant back to the browser and the next task. When all four tasks were completed, participants were asked to fill out the SUS questionnaire. The whole experiment session, beginning with the personal information to finishing the SUS questionnaire, took in average twenty to twenty five minutes.

3.7 Sequence analysis of eye movement data

An interesting way to explore eye-tracking data in a more detailed way than the regular eye-tracking metrics, such as time to first fixation, AOI fixation length and fixation counts, is the sequence analysis. Sequences are defined by the are of interests (AOIs). A possible sequence for a Task 2, for example, could be: task instruction - region x - target lines - x-axis - region x - target lines. Sequences can be analyzed by a theory - or data driven approach. In the theory driven approach (top-down) the eye movement data is analyzed for similarities to a beforehand defined ideal sequence, while the data driven approach (bottom-up) analyses the sequences for inherent patterns (Çöltekin et al. 2010). Patterns can define similarities or dissimilarites among groups. In this thesis, similarities in eye movement sequences for Task 2 will be analyzed in a top-down and bottom-up approach. Task 2 is chosen, because it is the most complex task, for participants had to compare four different regions of London, and is therefore ideally

19 www.google.com/chrome

suited for this kind of analysis. The other tasks are fairly straight forward and are not assumed to have significantly different sequences.

The data first needed to be adjusted before the sequences could be analyzed using the eyePatterns software (West et al. 2006). An ArcGIS implementation software Point Pattern Analyst (PoP) from Çöltekin et al. (2010) allows for the transformation of the Tobii Studio data into sequence lists, where each AOI is represented by a coded letter. The AOI title, for instance, is coded with a T, the AOI task instruction is a J and the region x an R. The coding can, of course, be manually adjusted. After PoP has semi-automatically transformed the eye movement data, the result is simply a string of AOI letters in text form:

> participant=01, highlighting=blur, task=2, speed=30.5

TTTTTTTTTTTTTJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJRRRRRRRRRRR-RRRRRRRRRRRRRRRRRRRRRRRRRGGGGGGGXXXXXRRRRRRRRRR-RJJJJJJJJJJJJJJJJJJJJJRRRRRRRRRRGGGGGGGGGGGGGGGGGGGGGG-GYYYYYYYYYGGGGGGGGGGGGGGGGGRRRRRRRRRRGGGGGGGGGGG-GG

The > symbol tells the eyePatterns software (West et al. 2006) that the sequence will start on the new line. Everything after the < on the same line can be used to define attributes, like participant=01. The sequence data in the example above uses full sequences. However, for this kind of analysis the order is more important then the counts within an AOI (Çöltekin et al. 2010). Repeating letters should therefore be accounted for and deleted. The above sequence in a *collapsed* state would thus more look like this: TJRGXRJRGYGRG. The eyePatterns software, fortunately, provides the option to either analyze the *extended* sequences or *collapsed* sequences.

The eyePatterns software from West et al. (2006) provides different methods to analyze the sequence data. The default similarity detection algorithm in eyePatterns is the Levenshtein dissimilarity measure (Levenshtein 1966) which counts necessary steps (i.e. delete, insert or substitute) to transform one sequence into the other. The other available similarity measure in eyePatterns is the Needleman and Wunsch's (1970) global sequence alignment algorithm. It uses a scoring scheme for matches, mismatches and gaps. Matches are rewarded with 1 point, gaps with 0 points and mismatches with -1 (standard scores in eyePatterns, West et al. 2006). The findings of these measures can then be visualized in either a distance matrix, average linkage clustering or cluster visualization using multidimensional scaling (MDS, Kruskal & Wish 1978).

4 General Results

The user study required participants to perform one task in Test 1 and four tasks in Test 2. This section reports the results from both tests. First, the qualitative and quantitative results from Test 1 are described. The results from Test 2 are then disclosed and thirdly, the results from the SUS questionnaire are provided. The findings from the eye-tracking device are separately presented in section 5.

4.1 Test 1

Efficiency

The different blur levels ranged from level 1 (50% line thickness as blur radius) to level 4 (200% line thickness as blur radius). Participants who were shown level 1, performed the least efficient (mean = 23.37 seconds), followed by the level 4 group (mean = 16.15 seconds). The two medium levels 2 and 3 performed best (13.24 and 13.29 seconds respectively). Figure 9 reveals the results in a boxplot.

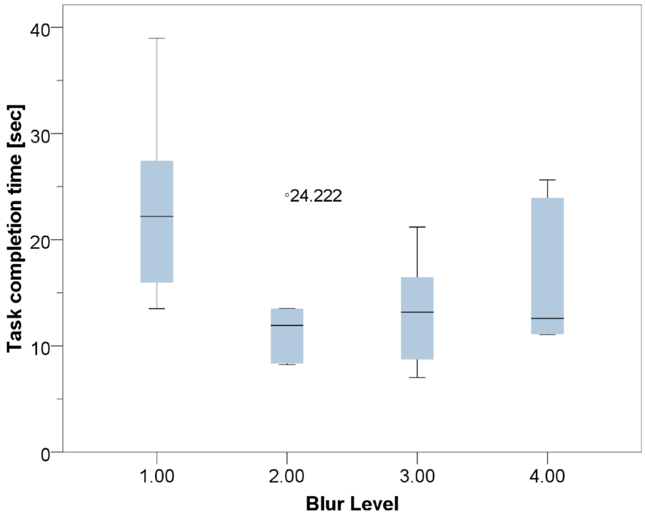


Figure 9: Task completion time for the four different blur levels.

The median values for level 2, 3 and 4 are very similar, while level 1 has a clearly higher median. A one way ANOVA, however, reveals no significant differences among the four levels of blur ($F(3,132.806) = 2.613$, $p = 0.081 > 0.05$). It should be noted that due to the errors that have been made, the sample sizes ranged between 5 and 6 values, which is rather low. Nonetheless, it can be assumed that the blur levels 2 and 3 tended to perform slightly better than the other two.

Effectiveness

The only level of blur, where no errors were made was level 3. In the other three groups, the error rates were either 37.5% for level 2 or 25% for level 1 and 4. In the group that viewed level 1, two participants thought that the highlighted symbol was a triangle instead of a line symbol. In the other groups (2 and 4) participants falsely counted the wrong quantity of same symbols within the frame.

4.2 Test 2

Participants in this test had to perform four tasks. Before looking into the individual tasks, the overall performances of the highlighting methods are provided.

4.2.1 Overall Highlighting

Efficiency

Figure 10 presents the average task completion times for focus & blur (41.60 seconds) and color (49.34 seconds). The average task completion times for this efficiency analysis include only correct answers.

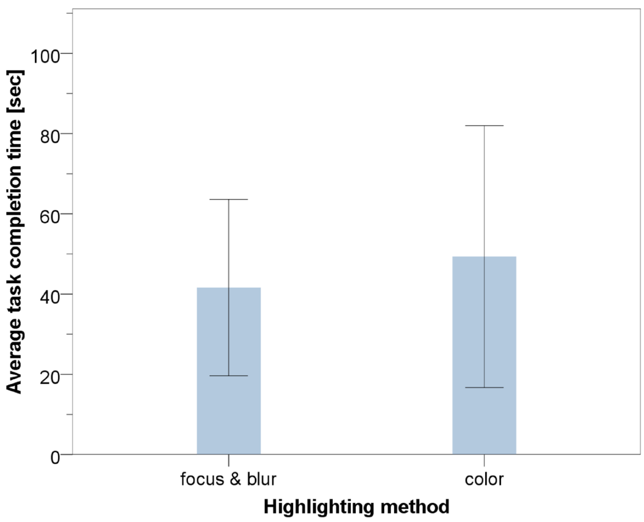


Figure 10: Average task completion times, regardless of task. (Error bars = σ)

The overall distribution of the response times are not normally distributed, even if outliers are accounted for. Therefore, the Mann-Whitney's U test is used to compare the response times for focus & blur and color, which reveals no significant effect of group ($U = 1139.0$, $Z = -0.967$, $p = 0.333 > 0.05$).

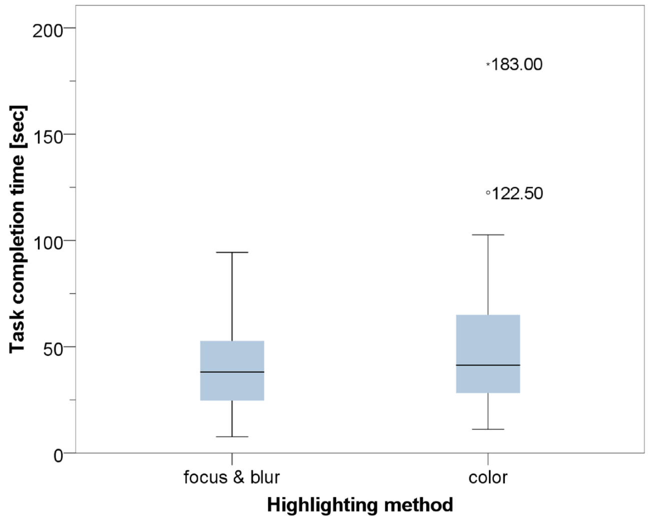


Figure 11: Overall performance of focus & blur and color, regardless of task.

The box-plot in Figure 11 reveals that the color highlighting method has one outlier and an extreme value which is more than three times larger than the mean value. Both highlighting approaches show similar variances in the response times, with minimum values of under 10 seconds and maximum values far above 100 seconds. Moreover, the median values with 38.0 seconds and 41.3 seconds are almost identical.

Effectiveness

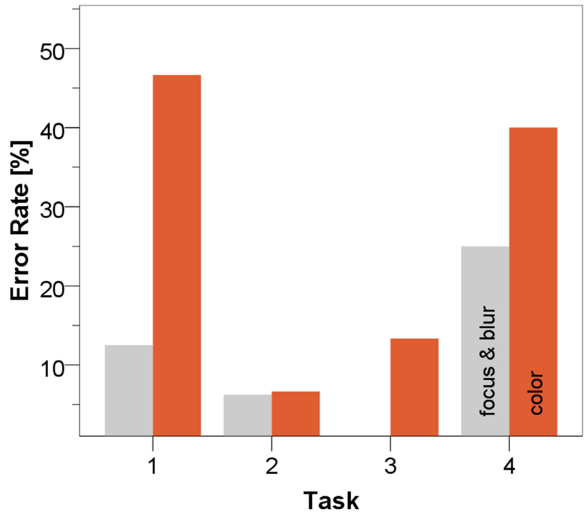


Figure 12: Error distribution among the four tasks for the two highlighting methods.

Overall, the participants who used color highlighting have given a wrong answer in 26.6%, of the time, while the participants using focus & blur fared an error rate of 10.9%. Figure 12 depicts the error rates for each of the four tasks. Of all the answers in Task 1, the color highlighting users gave 46.6% wrong answers (N = 15), while the focus & blur users made errors in 12.5% times (N = 16). The errors have occurred evenly distributed with regard to the order of appearance in the user study for this task. The errors in Task 4 could be related to the fact that participants had to distract one value from the other. At least some participants gave an answer that suggests that they did sum up the two values instead. The standard deviation from the right answer will be discussed later when Task 4 is more closely looked into. The other two tasks faired relatively good, with 2 errors for the color group in Task 3 and one error each in Task 2. Interestingly, Task 2 which was the most complex task was solved best of all.

4.2.2 Overall Tasks

The results from the four tasks are now analyzed regardless of the highlighting method. Task 1 and 3 were simple value search tasks and should have yielded faster response times. Task 2 and 4 required comparison and context analysis and should have required more time to solve than the first two. The more simple tasks were indeed completed in less time than the more complex tasks. Task 1 and Task 3 scored an average completion time of 31.70 seconds and 37.63 seconds respectively. Task 4 was in average completed in 42.17 seconds and Task 2 took the longest with an average of 65.03 seconds. The box-plot visualization Figure 13 will provide further information.

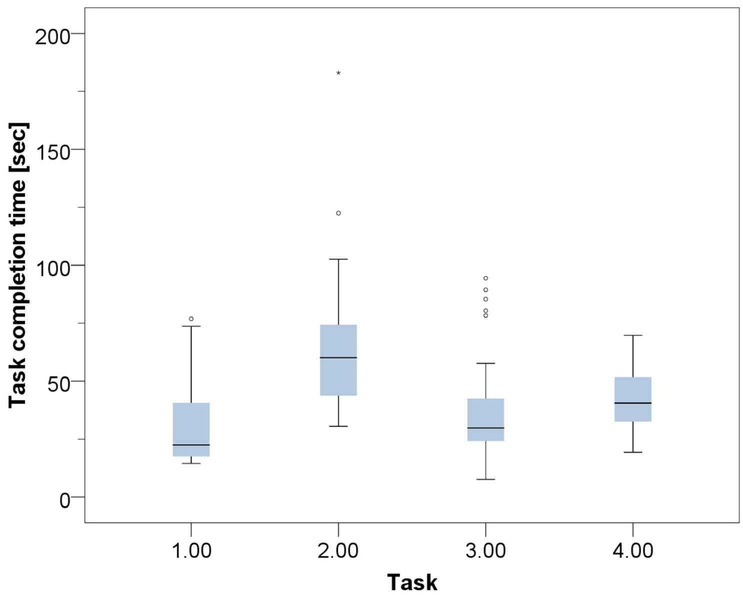
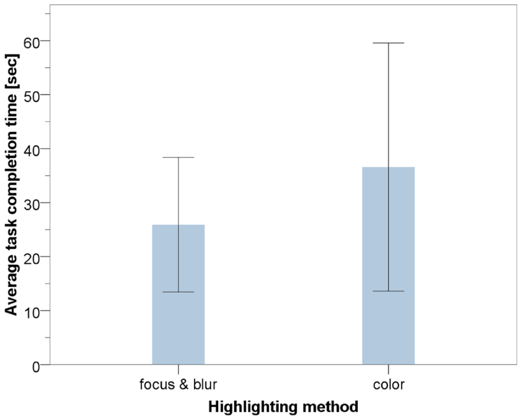


Figure 13: Box-plot of the four tasks completion times.

The lower whiskers in Task 1 are a lot nearer to the 25th percentile than the others. A grand majority completed the task relatively fast, while only a few of the participants required more time. The outliers for all tasks are solely on the upper part of the graph, which suggests that there have been a few participants that needed considerably more time to complete the tasks than the majority.

4.2.3 Task 1

Efficiency



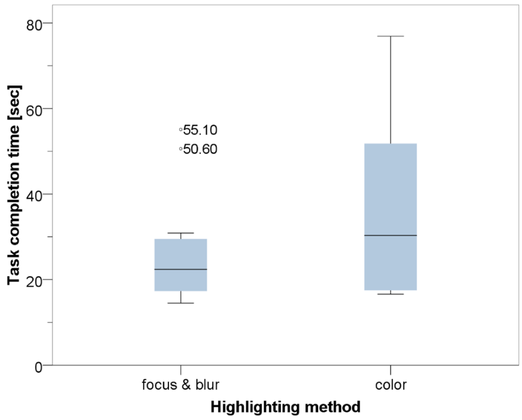


Figure 14: Task 1. Average task completion times for focus & blur and color. (Error bars = σ)
Figure 15: Task 1. Box plot of response times for focus & blur and color.

The mean task completion time in Task 1 for the focus & blur condition lies with 25.91 seconds (N = 14) quite a bit lower than for the color condition which shows a mean value of 36.59 seconds (N = 8). The data values from the focus & blur group are not normally distributed (Shapiro-Wilk, $p_{blur} = 0.003 < .05$), even if the outliers are accounted for. On the other hand, the color values have a higher variance than focus & blur.

Comparing the two highlighting methods in Task 1 with a Mann-Whitney's U test comparing focus & blur and color reveals no significant effect of group ($U = 46.0$, $Z = -0.683$, $p = 0.495 > 0.05$). Accordingly, the two highlighting methods do not differ significantly in efficiency for Test 1. It should be noted that the number of values for color (N = 8) is relatively low, due to the many mistakes that have been made in this task. Removing the two outliers from the focus & blur group does not result in a significant effect of group.

Effectiveness

In this task the color highlighting group scored an error rate of 46.6% which is unquestionably high. Considering that the response times for the participants who made an error are not significantly higher than the completion times from participants who gave a correct answer, it seems that the participants valued speed higher than correctness. The focus & blur group scored a lower error rate of 12.5%.

Confidence

The mean confidence ratings for Task 1 are:

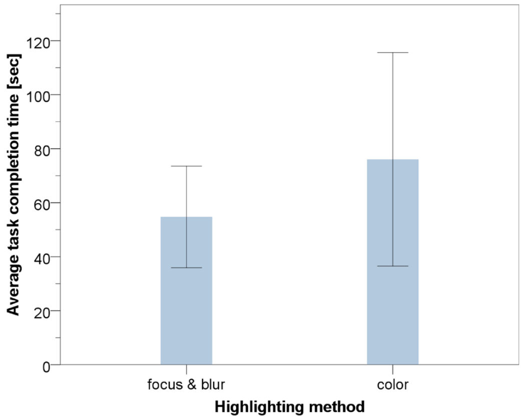
	focus & blur	color
correct answers	N = 14 4.9	N = 8 4.6
incorrect answers	N = 2 4.0	N = 7 3.1

Figure 16: Confidence Ratings for Task 1. (Likert scale, 1 to 5)

The confidence ratings are not too different from one another. Focus & blur scored slightly better and reflects the trend that this highlighting method scored slightly better in terms of efficiency (not significant) and effectiveness. Both mean confidence ratings are high, expecially the focus & blur group reported a very high mean confidence rating which is just one tick below the maximum of 5, which suggests that the highlighting method supported them well in their task. Participants who gave a wrong answer did also indicate their uncertainty with a lower mean confidence rating of 4.0 and 3.1 respectively.

4.2.4 Task 2

Efficiency



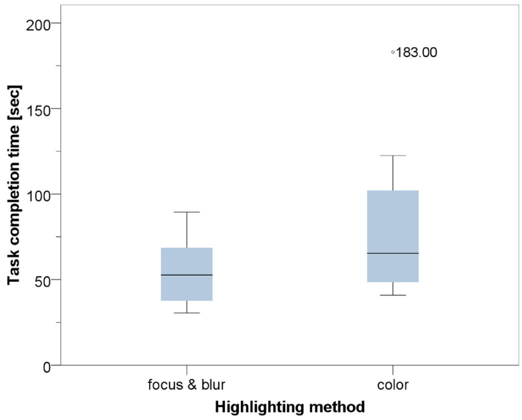


Figure 17: Task 2. Average task completion times for focus & blur and color. (Error bars = σ)
Figure 18: Task 2. Box-plot of response times for focus & blur and color.

Figure 17 and Figure 18 indicate that focus & blur had a lower mean (54.75 seconds, N = 15) response time than color (76.06 seconds, N = 14). Furthermore, the variance is greater for color than for focus & blur and sports a hefty outlier with a value of 183.0 seconds. The data values from the color group are not normally distributed, even when considering the extreme value (Shapiro-Wilk, $p_{color} = 0.005 < 0.05$).

The Mann-Whitney's U test comparing focus & blur (mean = 54.75 seconds) and color (mean = 76.06 seconds) attests no significant effect of group ($U = 68.0$, $Z = -1.615$, $p = 0.112 > 0.05$). The two highlighting methods therefore do not differ significantly in efficiency for Test 2.

Effectiveness

There has been one wrong answer for both highlighting methods (error rate = $6.25\%_{blur}$, $6.66\%_{color}$). The task was more complex than Task 1 and did take more time to complete. This could have led participants to be more careful before giving an answer to question.

Confidence

The average confidence ratings are for Task 2 are:

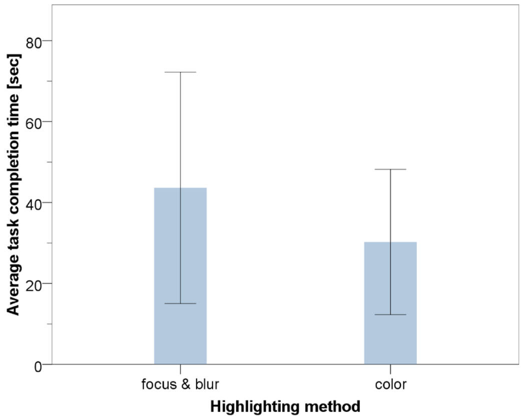
	focus & blur	color
correct answers	N = 15 4.5	N = 14 4.6
incorrect answers	N = 1 3.0	N = 1 4.0

Figure 19: Confidence Ratings for Task 2. (Likert scale, 1 to 5)

Participants reported a very similar and high mean confidence rating for both highlighting methods, which goes along well with the efficiency and effectiveness ratings in this task. Obviously, both highlighting approaches helped the participant complete the task with sufficient confidence.

4.2.5 Task 3

Efficiency



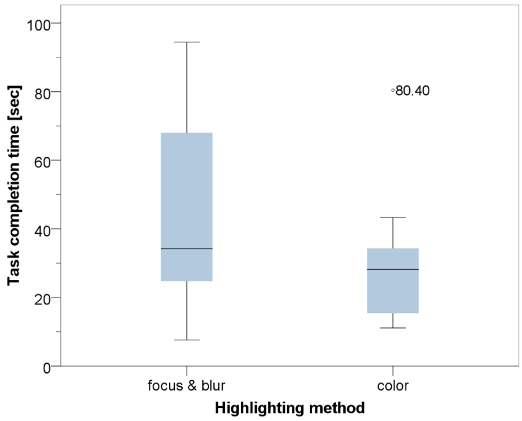


Figure 20: Task 3. Average task completion time for focus & blur and color. (Error bars = σ)
Figure 21: Task 3. Box-plot of response times for focus & blur and color.

Participants in this task who were shown the color highlighting performed faster (mean = 30.25 sec, N = 13), than the other group using focus & blur highlighting (43.63 seconds, N = 16). In addition, the variance for focus & blur is higher and the 75th percentile hovers far above the one from color (Figure 21). Both data sets are not normally distributed (Shapiro-Wilk, $p_{blur} = 0.036$, $p_{color} = 0.011 < 0.05$), but when taking the extreme value for color into account, the distribution for color is normal.

The Mann Whitney's U test, comparing focus & blur with color, confirmes no significant difference ($U = 75.0$, $Z = -1.272$, $p = 0.211 > 0.05$). Even when considering the extreme value for color highlighting, no significant difference can be observed. The two highlighting methods, therefore, do not have a significantly different efficiency.

Effectiveness

The focus & blur group scored a 0% error rate, while the color group shows an error rate of 13.33% with two wrong answers.

Confidence

The average confidence ratings for Task 3 are:

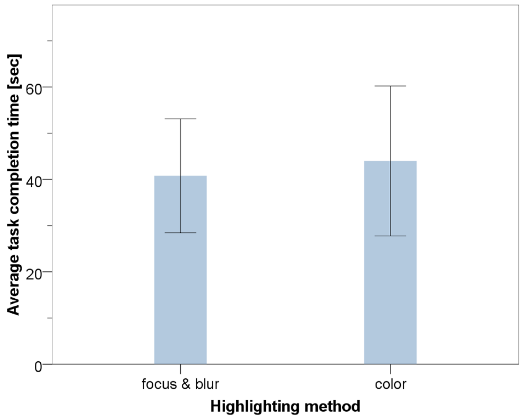
	focus & blur	color
correct answers	N = 16 4.6	N = 13 4.8
incorrect answers		N = 2 4.5

Figure 22: Confidence Ratings for Task 3. (Likert scale, 1 to 5)

Participants from both highlighting methods reported high mean confidence rating. The color highlighting did produce a slightly higher confidence rating than for focus & blur, which reflects the trend that the color highlighting condition was slightly more efficient (not significant) in this task, than the focus & blur condition.

4.2.6 Task 4

Efficiency



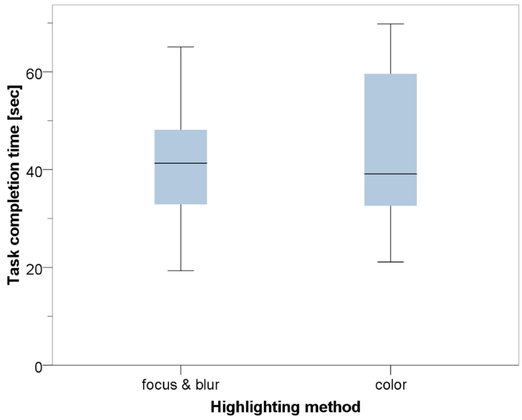


Figure 23: Task 4. Average task completion times for focus & blur and color. (Error bars = σ)
Figure 24: Task 4. Box-plot of response times for focus & blur and color.

The bar chart in Figure 23 reveals that the mean values for focus & blur (40.80 seconds, N = 12) and color (44.00 seconds, N = 9) are almost the same. The box-plot in Figure 24 suggests that the variance for color highlighting is slightly higher than for focus highlighting. The median value for the color condition is considerably lower than its mean value. Both data sets are normally distributed.

An independent samples t-test reveals that there is no significant difference between the two data sets, $t(19) = -0.515$, $p = 0.613 > 0.05$). The two highlighting methods are thus not significantly different in efficiency for this task.

Effectiveness

The error rate in this task is considerably high for both groups. Participants from the focus & blur group achieved an error rate of 25% and the participants from the color group an error rate of 40%. However, since this is not a boolean question, the deviation from the correct answers are of interest. Participants from the focus and blur group had a mean absolute value estimation error of 1.74 units, while on the other hand the color condition had 1.30 units.

Confidence

The average confidence ratings for Task 4 are:

	focus & blur	color
correct answers	N = 12 3.8	N = 9 4.2
incorrect answers	N = 4 4.0	N = 6 4.0

Figure 25: Confidence Ratings for Task 4. (Likert scale, 1 to 5)

The confidence ratings are considerably lower in this task than in the others. However, this was the only task where participants needed to calculate a difference between two values and not just simply report an observed relative value position. The fact that they had to do a mathematical computation, as benign as it was, might have affected peoples' confidence. Interestingly, the mean confidence rating in the focus & blur group is higher for the wrong answers than for the right answers.

4.3 SUS Results

The two groups in this user study from the different highlighting methods both filled out the 10 questions of the adapted SUS at the end of the session. The odd numbers represent positively formulated statements, while the even numbers represent negatively formulated statements.

Nr	Statement
1	I think that I would like to use this highlighting method frequently
2	I found the highlighting method unnecessarily complex
3	I found the highlighting method visually pleasing
4	I think that I could not see the highlighting very well
5	I found that the highlighting method helped me perform the tasks
6	I found the performance of the highlighting method not satisfyingly fast enough
7	I would imagine that most people would learn to use this highlighting method very quickly
8	I found the highlighting method distracted me from performing the tasks
9	I felt very confident using this highlighting method
10	I need to practice a lot more before I could get going with this highlighting method

Table 4: The 10 adapted SUS statements, based on the original SUS from Brooke (1996).

The overall SUS scores are 77.5 for focus & blur and 81.17 for color. Bangor et al. (2008) tested more than 2300 SUS questionnaires and came to the conclusion that SUS scores that are over 73 represent good to excellent results (Figure 26). Both highlighting approaches therefore scored good results and are acceptable from a subjective usability point of view.

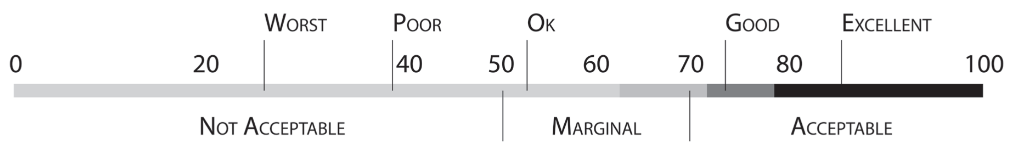


Figure 26: Quality rating for final SUS score. Source: Adapted from Bangor et al. (2008)

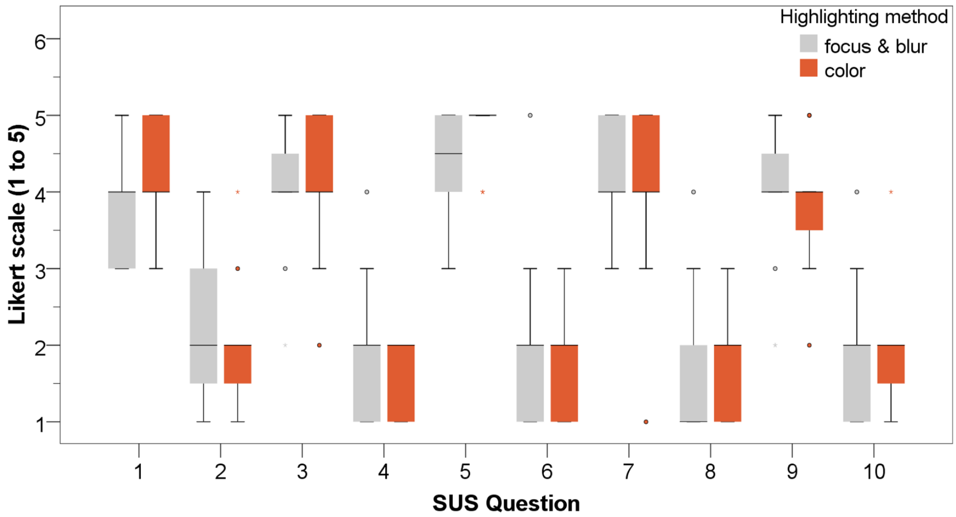


Figure 27: Results from the SUS questionnaire. Likert Scale (1 to 5).

Figure 27 displays the different results for both highlighting methods for each SUS statement. The boxplot suggests that there are no relevant differences between the two highlighting approaches. Mann Whitney's U tests revealed, subsequently, no significant differences in any of the 10 SUS statements for the two highlighting approaches. Both highlighting methods were thus satisfyingly enough.

5 Eye-Tracking Results

This section presents the eye-tracking results. The metrics of the eye-tracking device give insight into how long participants gazed at an area, how long it took them to look at a specific area for the first time and ultimately in what sequence they looked at these areas. Areas of interests (AOIs) were defined for each task and placed at the locations where the participants most likely needed to look at to complete the tasks (top-down AOI). After qualitatively assessing the eye-tracking data, no additional areas of interest needed to be added (bottom-up AOI).

First, the results from Test 1 are discussed. The results from the second test, with a focus on transition times and fixation lengths on specific AOIs are then presented. Finally, a sequence analysis of the eye movement data of both highlighting methods in Task 2 are discussed.

5.1 Test 1

5.1.1 Qualitative

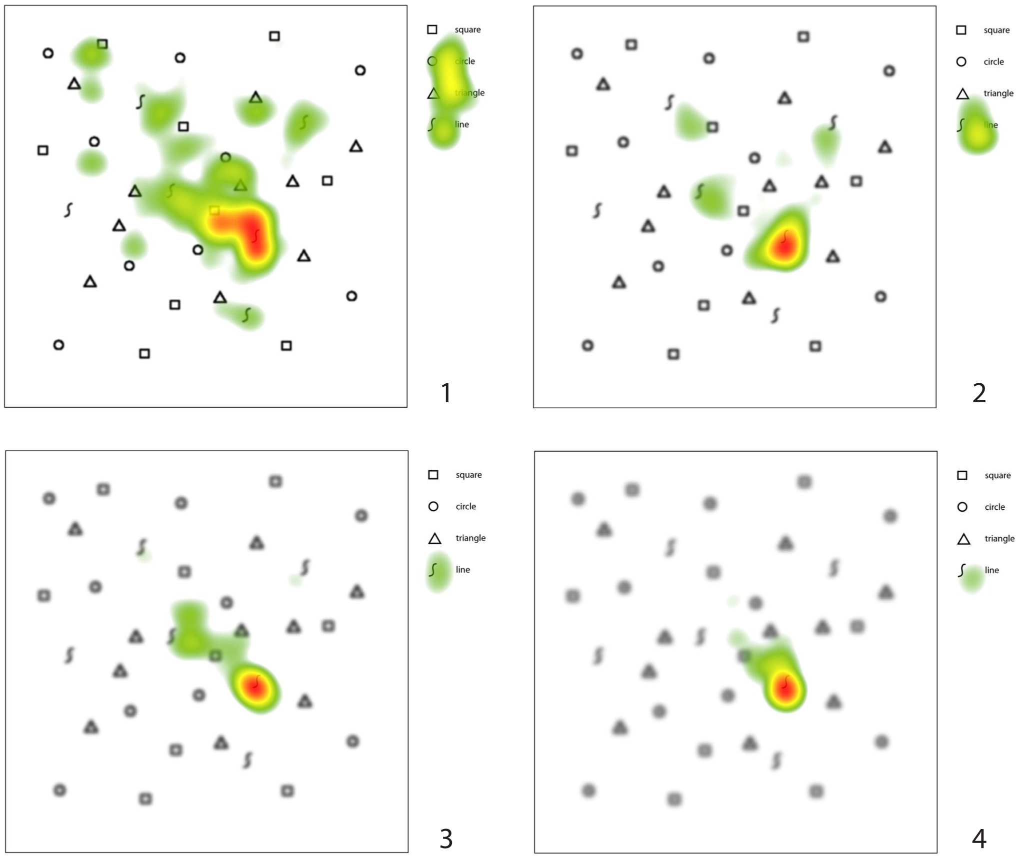


Figure 28: Mean fixation lengths for the 4 levels of blur. (relative gaze duration)

The mean fixation lengths images in Figure 28 clearly show how the participants looked at different regions across the four levels of blur. Participants who looked at blur level 1 stimuli, had to look around to find the highlighted symbol, while participants from level 4 had no problem whatsoever to find the highlighted symbol. The images from level 2 and 3 suggest that participants needed less time to look around and found the highlighted target relatively fast. In level 3, however, participants required less time to look at the remaining symbols within the frame, which can be seen by the smaller heatmap smudges.

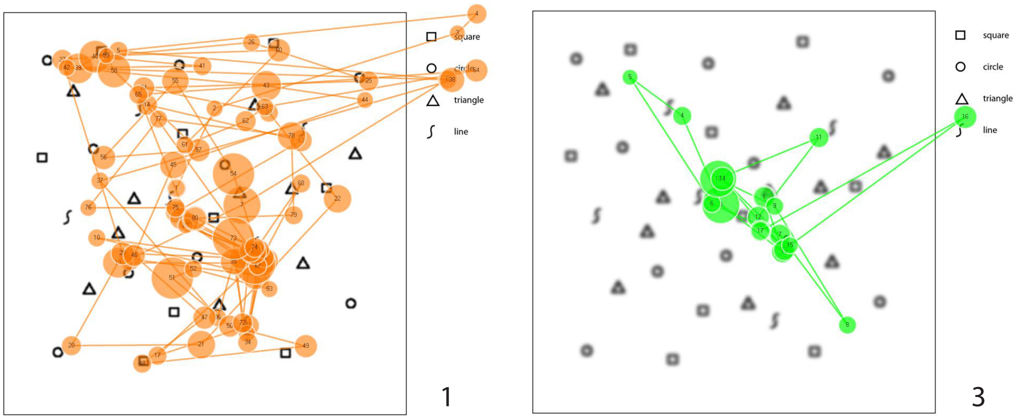


Figure 29: Typical gazeplots of level 1 blur and level 3 blur stimuli.

The gazeplots in Figure 29 make it clear how different the search was performed in the two blur levels. Generally, participants who were shown level 1 blur stimuli, performed a serial search, while participants who were shown blur levels 2 and 3 instantly found the highlighted target and had no problems finding the five other symbols within the frame.

5.1.2 Quantitative

The eye-tracking parameter that can be best used to assess the level of blur is the time needed to fixate the highlighted object for the first time. Participant from the blur level group 1 required in average 6.49 seconds, while the level 2 blur group needed in average 1.37 seconds, the level 3 blur group 0.47 seconds and the level 4 group 0.75 seconds. The boxplot in Figure 30 shows the value distribution for the four different blur levels. The blur level 3 seemed to have had the shortest first fixation times, while it took participants from level 1 clearly the longest for a first fixation. An ANOVA test revealed that at least one blur level was significantly different from the others ($F(3,47.238) = 15.517$, $p = 0.000 < 0.05$). A subsequent post hoc test with Tukey's HSD confirmes that the blur level 1 group fixated the highlighted symbol significantly later than the others. Another interesting result is the 0 seconds value in blur level 4, which is probably because the participant fixated that area before the stimuli was shown and therefore occurred by chance.

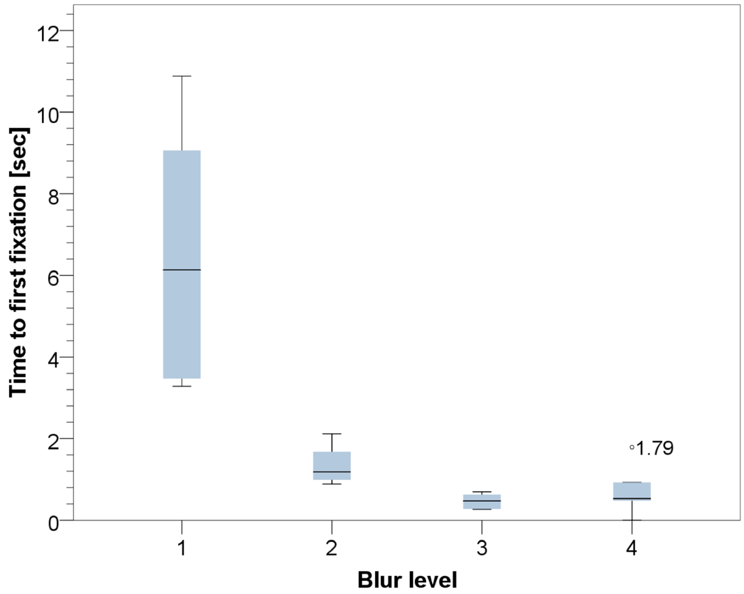


Figure 30: Test 1. Time needed to fixate the highlighted symbol for the first time.

5.2 Test 2

In this section the four tasks are analyzed with the help from the eye-tracking data. For each task, a qualitative overview and specific occurrences are given first, The quantitative analysis is presented afterwards.

5.2.1 Qualitative Analysis

Task 1

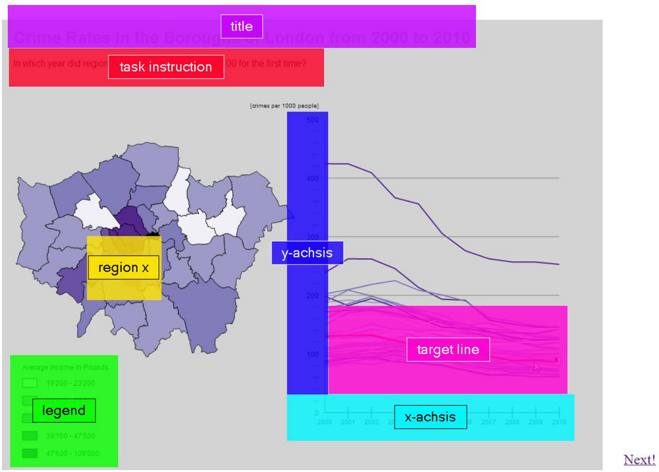
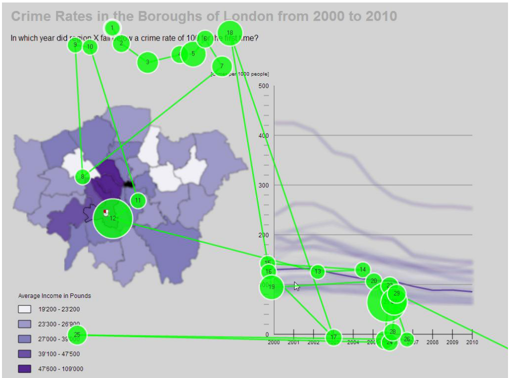


Figure 31: AOI for Task 1.

Each of these AOIs, defined in Figure 31, depict an important area for the successful completion of this task. Most importantly are the following: region x, target line, task instruction and the two axes. The legend to the classification is not needed in this task. Figure 32 portrays a fast performer of the focus & blur group and it clearly shows that the participant did not gaze at unnecessary areas and quickly focused from region x to the target line.



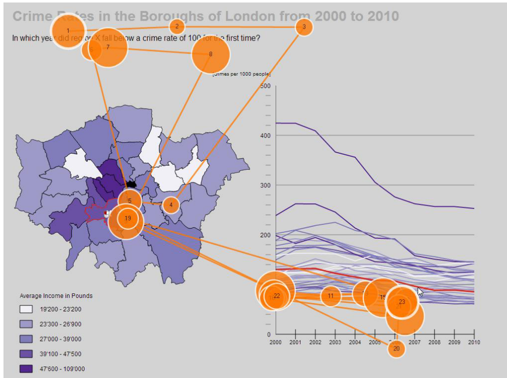
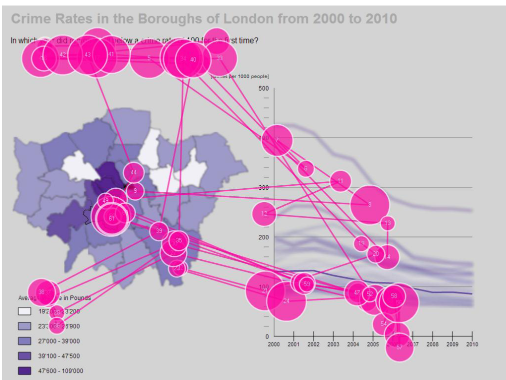


Figure 32: Gazeplot of Task 1 with focus & blur highlighting. Typical fast performer.
Figure 33: Gazeplot of Task 1 with color highlighting. Typical fast performer.

Overall, participants who were less efficient seemed to look around more often and double checked the task instructions (Figure 34), while more efficient users went for a more straight and simple path pattern. Figure 35 shows a participant who gave a wrong answer. The relative gaze time shows that he fixated the task instructions and specifically the term *below*, which suggests that he did confuse *below* with *above*. All in all, there is no visible difference in the gazeplots between the two highlighting methods for fast and slow participants.



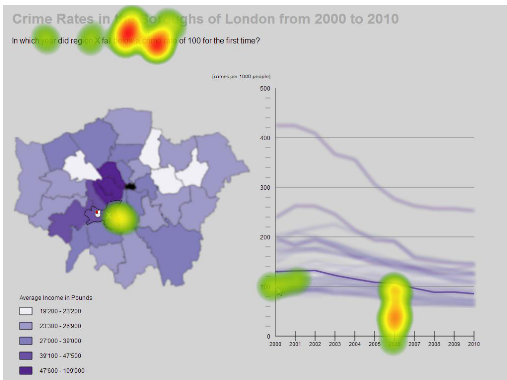


Figure 34: Heatmap of Task 1 with focus & blur highlighting (relative gaze time). Slow performer.
Figure 35: Heatmap of Task 1 with color highlighting (relative gaze time). Wrong answer.

Task 2

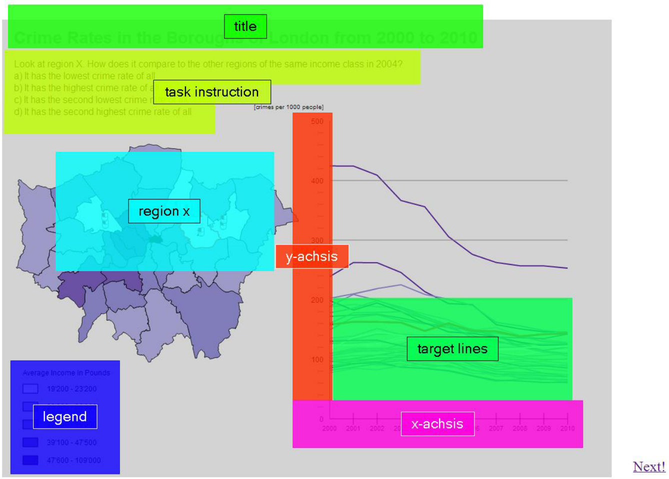
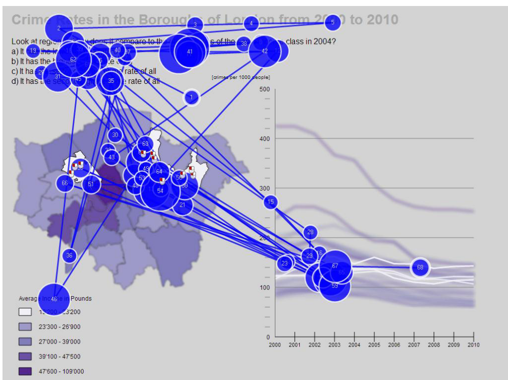


Figure 36: AOI for Task 2.

The AOIs for Task 2 are not too different from the Task 1 AOIs. However, because there are three separate region x areas, a single big AOI for region x was created. The size of the regions and the distance between them are to small for the accuracy of the eye-tracker to have them separately labeled.



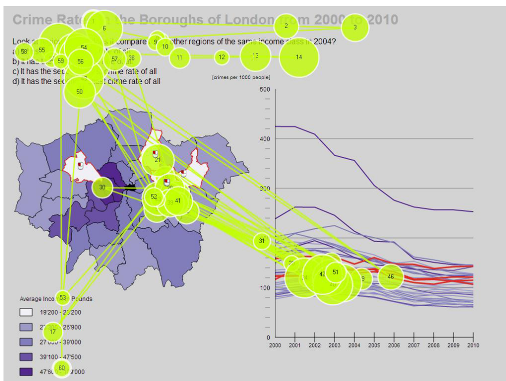
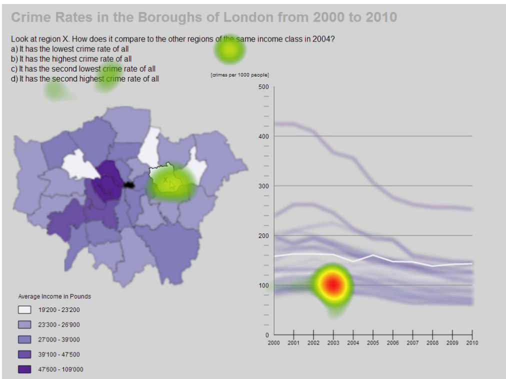


Figure 37: Gazeplot of Task 2 with focus & blur highlighting. Fast performer.
Figure 38: Gazeplot of Task 2 with color highlighting. Fast performer.

The gazeplots from the two different highlighting methods are strikingly similar. Both represent an efficient performer and both made use of the click and freeze function to allow for multiple objects to be highlighted at the same time (Figure 37 and Figure 38). Moreover, the legend, which this time was needed or at least supportive in the task, was fixated by all the fast performers at least once.



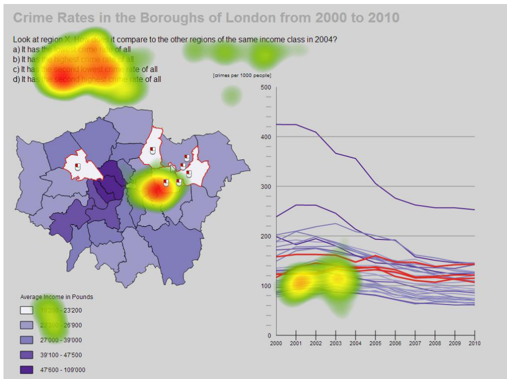


Figure 39: Heatmap of Task 2 with focus & blur highlighting (relative gaze time). Slow performer.
Figure 40: Heatmap of Task 2 with color highlighting (relative gaze time). Slow performer.

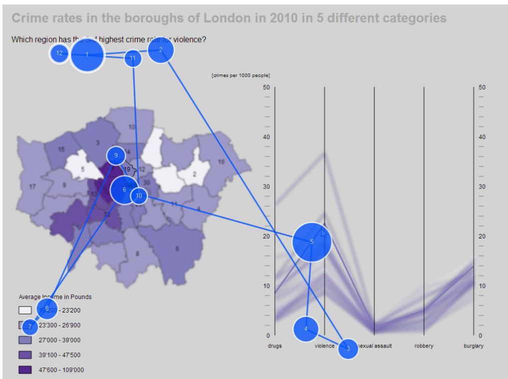
Participants who were less efficient to solve this task did not have a very different gazeplot pattern, but the distributions of gaze-lengths are different. The slower performer with focus & blur condition in Figure 39 did not make use of the click and freeze function, but tried to see where the other data lines of the same income category are placed within the graph. The participant did manage to solve the task correctly but needed more time. On the other hand, the slower performer with the color condition focused his time evidently on the task instructions. This suggests that the participant had trouble remembering or understanding the task instructions, which lead to the relative high gaze-time on that region.

Overall, the gazeplots are not very different between the two highlighting methods, but reveal clear differences between the faster and slower performers, regardless of the applied highlighting method.

Task 3

Figure 41: AOI for Task 3.

In this task, participants had to start from the graph (third highest crime rate for violence) and find the corresponding region. Other than that, the AOIs are the same as for the other tasks.



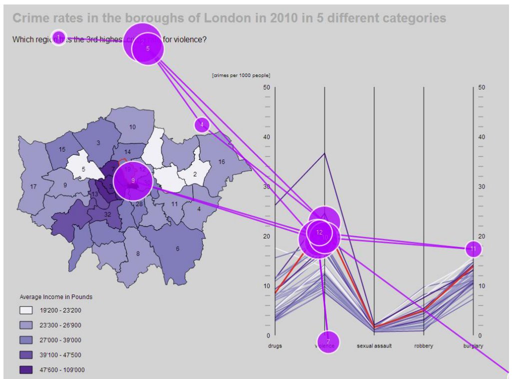
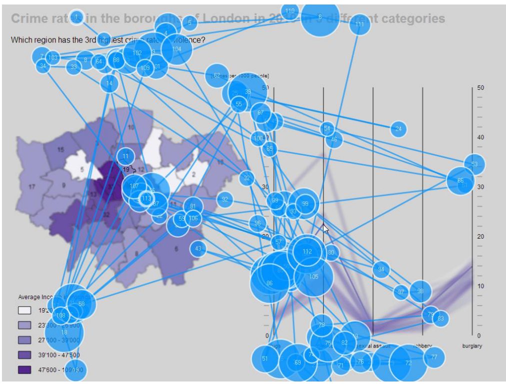


Figure 42: Gazeplot of Task 3 with focus & blur highlighting. Quickest performer.
Figure 43: Gazeplot of Task 3 with color highlighting. Fast performer.

While the focus & blur participant in Figure 42 was the fastest performer, the participant in Figure 43 was the second fastest performer. Both gazeplots are very similar and show little fixations on irrelevant areas of the stimuli. The focus & blur user fixated shortly the legend in the lower left corner, while the color user fixated the right y-achsis element shortly. Other than that, they went straight from the task instruction to the target line and then to the corresponding area in the map.



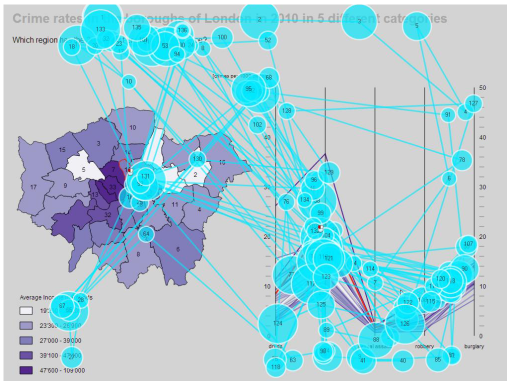


Figure 44: Gazeplot of Task 3 with focus & blur highlighting. Typical slow performer.
Figure 45: Gazeplot of Task 3 with color highlighting. Typical slow performer.

Slow performers in this task fixated a lot of areas on the stimuli and even put considerate time on the legend, which was of no help for this task. The scan-paths show clearly that they went from task instruction to target line to region x back and forth.

Overall, no substantial differences in the gazeplots can be detected between the two highlighting methods, but clear differences between slow and fast performers, regardless of highlighting approach, are observable.

Task 4

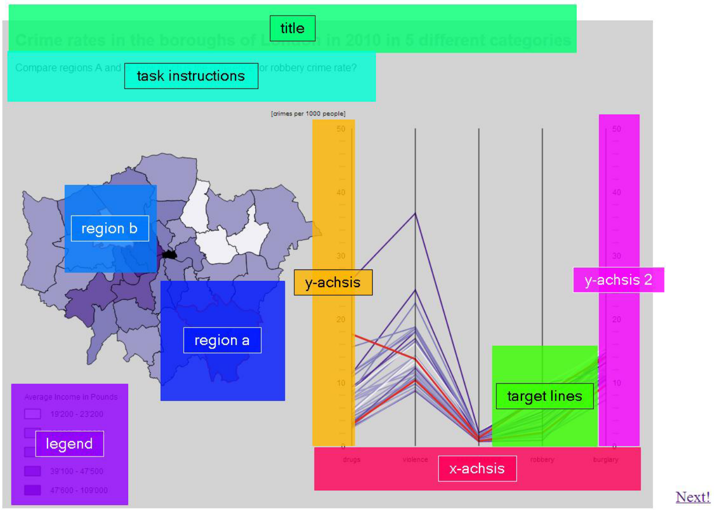
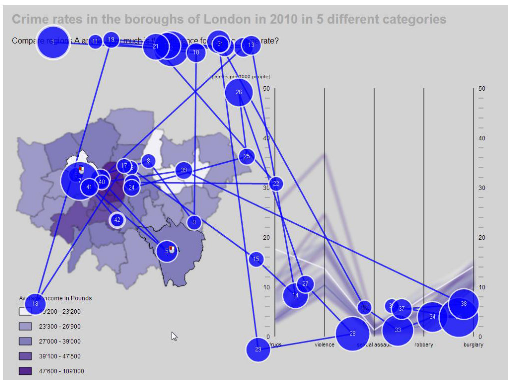


Figure 46: AOI for Task 4.

The AOIs for this task have two important regions (a and b). The target line AOI is concentrated around the robbery crime data axis. Additionally, a supplementary AOI is introduced for the second y-axis.



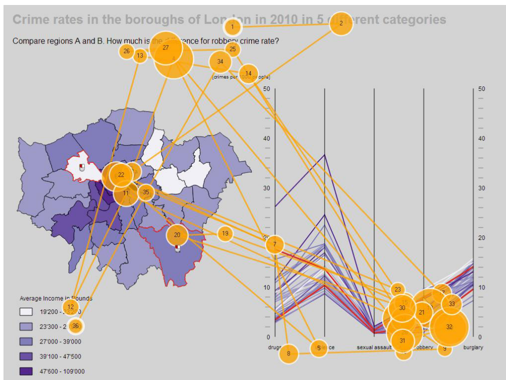
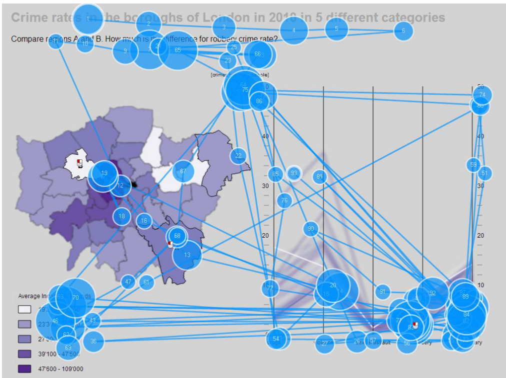


Figure 47: Gazeplot of Task 4 with focus & blur highlighting. Typical fast performer.
Figure 48: Gazeplot of Task 4 with color highlighting. Typical fast performer.

The typical efficient performer in this task did not fixate many irrelevant areas on the stimuli. Since this task required participants to calculate a difference, some of the fixations may be credited to participants calculating and not directing their gaze at will or just staring at the middle of the screen.



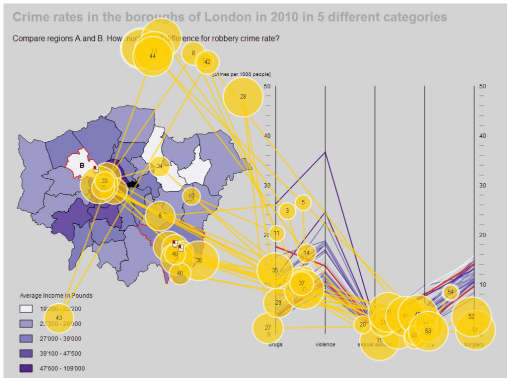


Figure 49: Gazeplot of Task 4 with focus & blur highlighting. Specific slow performer.
Figure 50: Gazeplot of Task 4 with color highlighting. Typical slow performer.

In Figure 49, the specific slow performer fixated the legend quiet a lot and for a considerate amount of time. This participant also went back and forth from target line area to task instruction and legend. Quite possibly, the participant mixed up the income classification legend with the crime rate data and got confused. The slow performer in Figure 50 has a similar gaze-path pattern than the faster performers, but took considerately more time at each area, which can be observed by the bigger fixation circles.

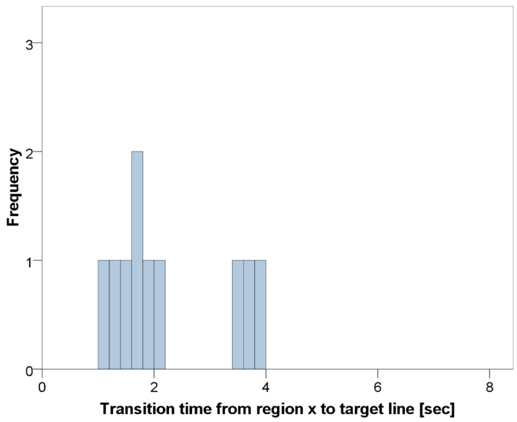
Overall, there are no obvious qualitative differences in the gazeplots between the two highlighting approaches. However, since the task required participants to calculate a simple value difference, some seemed to have had problems to do so under the self inflicted time pressure. The gazeplot from Figure 49 is also found in similar form in the color highlighting group.

5.2.2 Quantitative Analysis

Time to first fixation

An interesting metric of eye-tracking systems is the time to first fixation on a specific AOI. Because participants had to read the instructions on the stimuli itself, a simple time to first fixation is not a meaningful variable to determine the efficiency of a highlighting method. Some people might have taken more time to read the task instruction or looked at them in a later stage of their task performance. The target line AOI is of most importance for successfully solving the task and should be fixated right after region x has been detected and highlighted. Therefore, the transition time from region x to target line is further analyzed.

Task 1



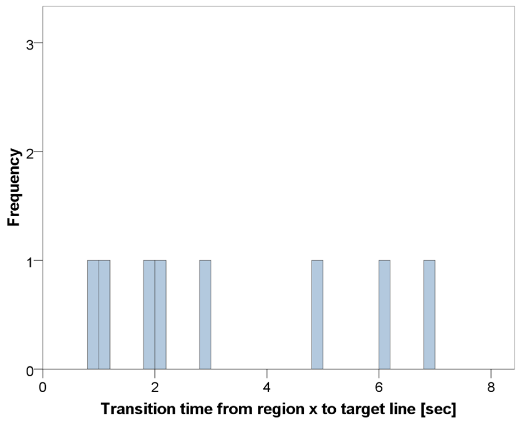
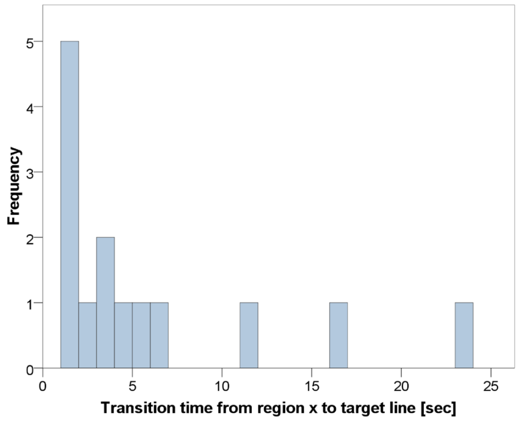


Figure 51: Transition times for Task 1, focus & blur highlighting. (interval = 0.2 sec)
Figure 52: Transition times for Task 1, color highlighting. (interval = 0.2 sec)

In task 1 the transition times from region x to target line are slightly faster in the focus and blur highlighting group (mean = 2.216 seconds), than for the color highlighting group (mean = 3.370 seconds). However, an independent samples t-test revealed no significant difference, $t(16) = -1.412$, $p = 0.177 > 0.05$. It must be noted, though, that the quantity of values is relatively low ($N = 10$ for focus & blur, $N = 8$ for color) and it is therefore only possible to comment on a trend.

Task 2



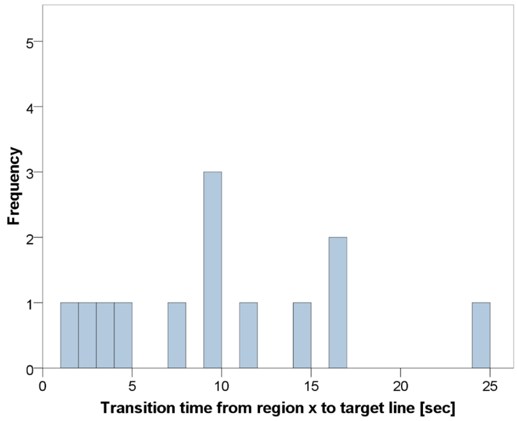


Figure 53: Transition time from region x to target line for Task 2, focus & blur highlighting. (interval = 1 sec)
Figure 54: Transition time from region x to target line for Task 2, color highlighting. (interval = 1 sec)

Figure 53 and Figure 54 show the transition times for Task 2. The frequency graphs suggest that the transition times for the focus & blur group are slightly faster (mean = 6.047 seconds), than for the color group (mean = 10.143 seconds). Because the data set of the focus & blur group is not normally distributed a Mann Whitney's U test has been used and showed a significant difference for transition times between the two highlighting methods ($U = 50.5$, $Z = -1.966$, $p = 0.049 < 0.05$). Figure 55 clarifies the histograms and shows the two extreme values for focus & blur more clearly.

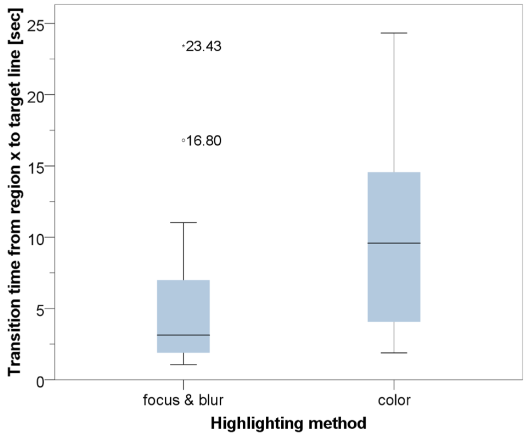
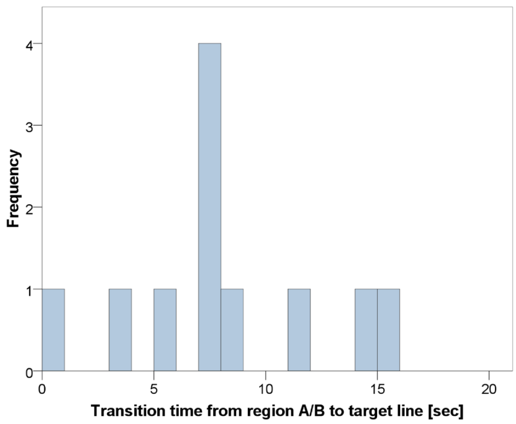


Figure 55: Boxplot of the transition times for Task 2.

Task 3

The transition times for Task 3 are unfortunately impossible to calculate, because the search task went from graph to map, almost all participants looked at the map first, which would distort the result (negative values).

Task 4



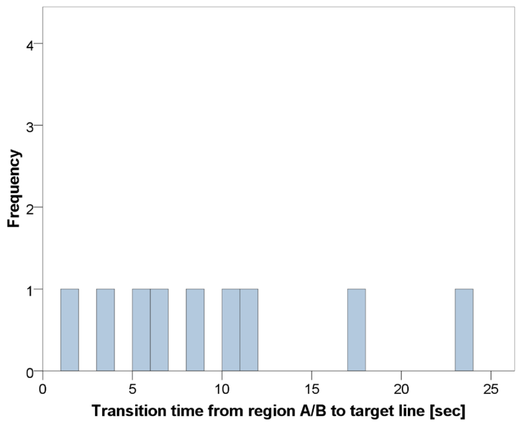


Figure 56: Transition times for Task 4, focus & blur highlighting. (interval = 1 sec)
Figure 57: Transition times for Task 4, color highlighting. (interval = 1 sec)

The transition times in Task 4 are from the look of the two histograms in Figure 56 and

Figure 57 quite similar. An independent samples t-test confirmed that there is no significant difference between the two highlighting methods, $t(18) = -0.650$, $p = 0.524 > 0.05$.

Fixation length per AOI

The fixation lengths per AOI shows how long participants gazed at any of the areas of interest. The fixation lengths are shown in a percentage of the total task completion time.

Task 1

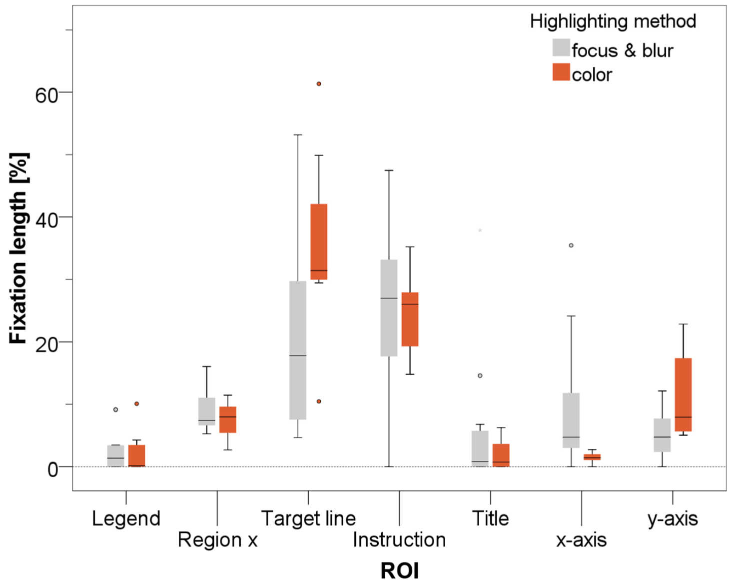


Figure 58: Fixation times for the different AOI in Task 1.

Figure 58 shows the boxplots for each of the AOIs in Task 1. There seems to be no difference in gaze-length for the legend and title. Region X is also fixated in average for the same amount of time for both highlighting methods. However, the boxplot suggests that the mean fixation lengths for the target line is quite different for the two highlighting methods. The task instructions are fixated for approximately the same mean time (24.48 and 25.55 seconds), but the focus & blur group has a higher variance. The x-axis is in average less looked at by the color group, than by the focus & blur group. On the other hand, the y-axis is fixated longer by the color group than by the focus & blur group.

The comparisons between the two highlighting approaches reveal that AOI target line is significantly shorter looked at by the focus & blur participants (independent samples t-test, t(20) = -2.399, p = 0.026 < 0.05). Furthermore, a Mann Whitney's U test discloses that the x-axis is significantly fixated for less time by the group with the color condition (U= 17.0, Z = -2.665, p = 0.008 < 0.05). The y-axis is, moreover, in average fixated for less time by the participants with the focus & blur condition (U = 27.0, Z = -1.979, p = 0.048 < 0.05).

Task 2

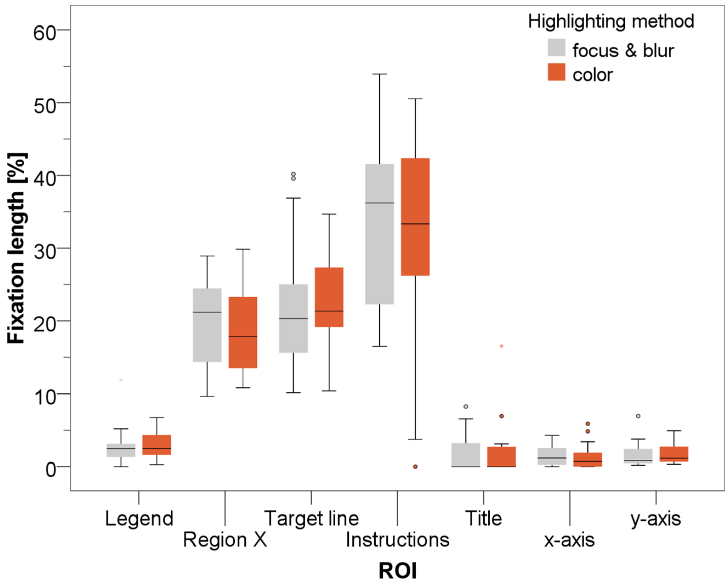


Figure 59: Fixation times for the different AOI in Task 2.

The fixation lengths for the different AOIs in Task 2 seem to be very similar for both highlighting methods. Figure 59 shows the boxplots for each AOIs and suggests no significant differences. Interestingly, the AOI instructions has the highest variance among all others for both highlighting approaches. The instructions were the longest in this task and it is therefore of no surprise that participants needed to look at them for quite some time and in varying lengths. The task could be solved without looking at any of the axes, which is exactly what the boxplots suggest. Overall, none of the AOIs provide significantly different mean fixation lengths for either of the highlighting approaches.

Task 3

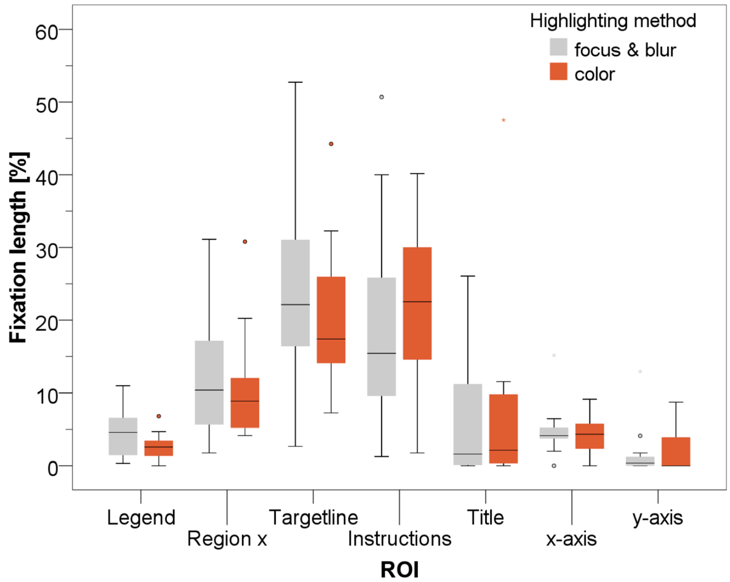


Figure 60: Fixation times for the different AOI in Task 3.

The boxplot from Figure 60 portrays the different fixation lengths for all AOIs. This task required participants to go from graph to map. The x-axis was used to find the violence data in the PCP graph, thus people fixated that AOI longer than the y-axis. Region x was the answer to the question and it seams that the participants from the color group needed slightly less time looking at it. This goes well with the findings from section 5.2.5 which suggested that the color group needed slightly less time to complete the task (not significantly). Furthermore, participants from the color group needed slightly less time looking at the targetline than the focus & blur participants. Interestingly, focus & blur users looked slightly longer at the legend than the color users, which might be explained by the fact that the color highlighting overrides the data lines initial color, while focus & blur lets the user still see the classification color. This might have distracted some of the focus & blur participants.

Task 4

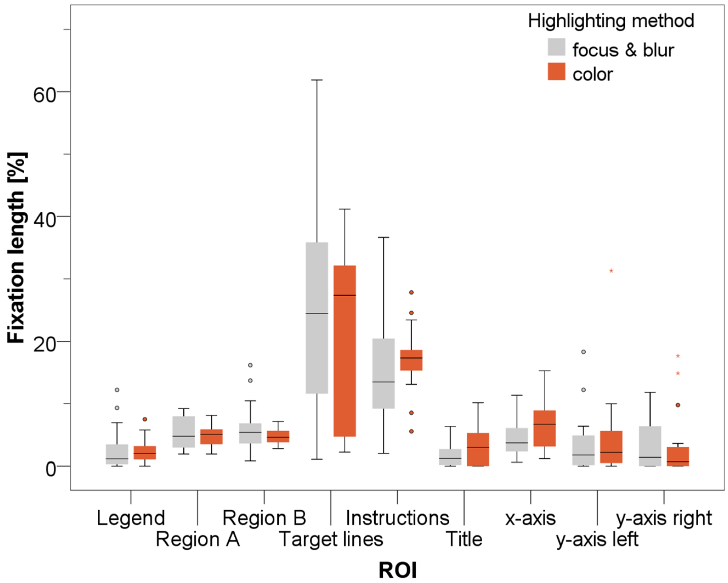


Figure 61: Fixation times for the different AOI in Task 4.

Task 4 was completed by both highlighting groups in similar efficiency (see section 5.2.6). The boxplots from Figure 61 reflect these findings. There are no significant differences in mean fixation lengths for any of the AOIs. All AOIs, except for the target lines and instructions, were fixated relatively short over the period of solving the task. The target lines and instructions took up most of the time and these are the AOIs where participants varied the most. However, the variance is equally for both highlighting approaches.

5.2.3 Task 2 Sequence Analysis

Task 2 was the most complex task in this user study. Participants had to compare region x with three other regions from the same income class. The task is therefore ideally suited to analyze sequences in the eye movement data.

Similarity in gaze sequence

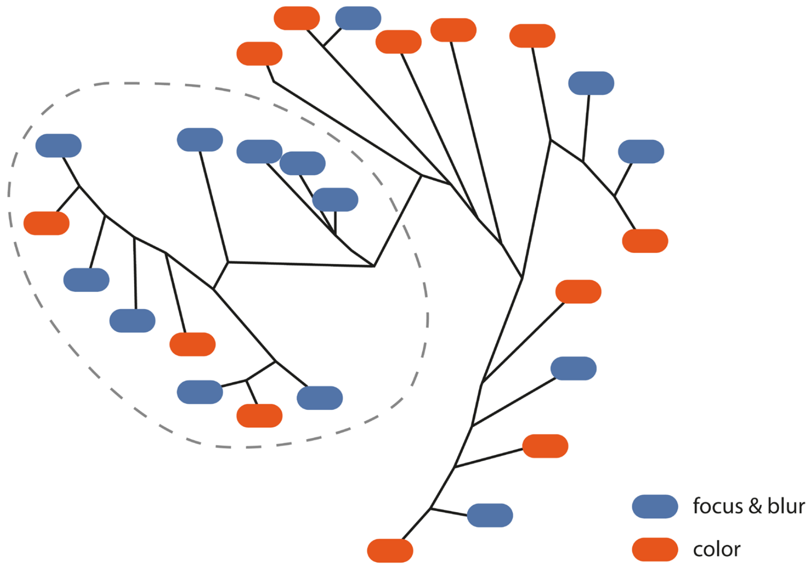


Figure 62: Clustering based on Levenshtein (1966) distances and average linkage clustering for Task 2

Figure 62 portrays the clustered sequences from the eye movement data. As stated in section 4, focus & blur was in average slightly faster than color highlighting, although not significantly. The sequence analysis with Levenshtein (1966) distances and average linkage clustering shows that focus & blur has many similar sequences and they tend to be the fast and medium values in terms of efficiency (Figure 63). It should be noted that the similarity is not based on distance in the dendrogram but on the count of branches between two values (West et al. 2006).

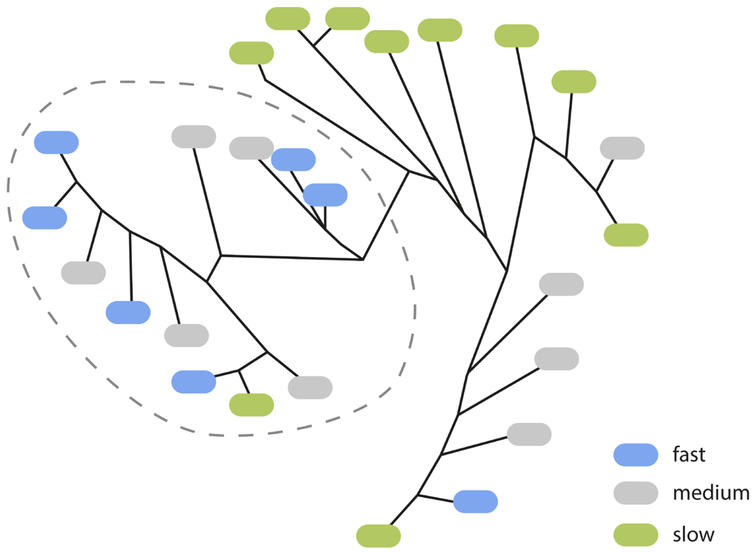


Figure 63: Clustering based on Levenshtein (1966) distances and average linkage clustering for Task 2.

In conclusion, the two dendrograms suggest that the faster performers have more similar sequences of which, evidently, the majority are from the focus & blur highlighting group. The leading factor for sequence similarities, thus, is most likely performance efficiency and not the different highlighting methods in general. Nonetheless, in order to validate the findings, another clustering algorithm is presented in Figure 64. The Needleman and Wunsch (1970) sequence alignment algorithm validates the findings from the Levenshtein (1966) measure by also showing a clustering among the medium and fast performing sequences.

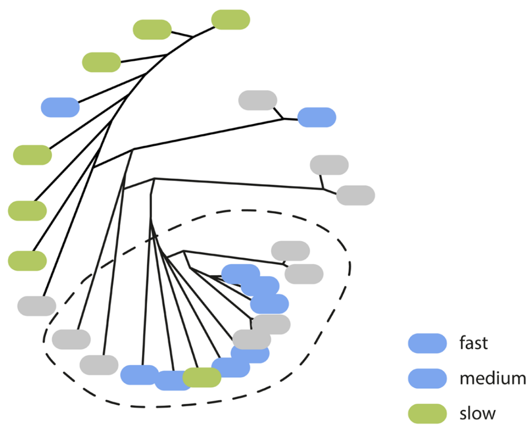


Figure 64: Clustering based on Needleman & Wunsch (1970) sequence alignment and average linkage clustering for Task 2.

In another look at the sequence similarities, the top 5 performers were analyzed separately. The dendrogram in Figure 65 shows the difference between the two top 5 performers for each of the highlighting methods. The focus & blur sequences tend to cluster more clearly, with the exception of the slowest of the five values.

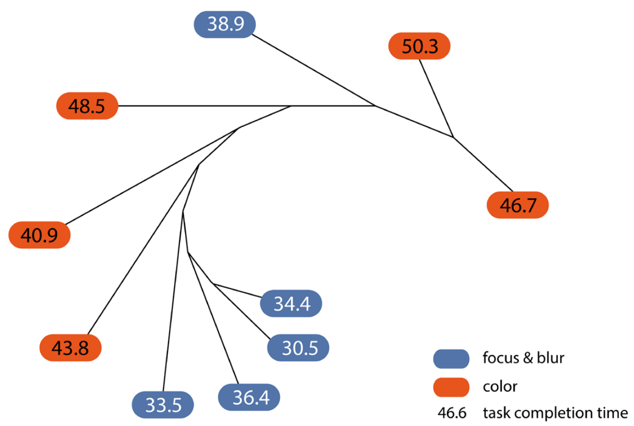


Figure 65: Clustering based on Needleman & Wunsch (1970) sequence alignment and average linkage clustering for the top 5 performers of both highlighting methods.

The mean values for the top 5 performers are indeed quite different (34.74 seconds for focus & blur and 46.04 seconds for color). An independent t-test has confirmed the significant difference between the two top 5 performers ($t(8) = -5.164$, $p = 0.001 < 0.05$). The top five performers' sequences, therefore, clustered due to efficiency and not primarily due to the used highlighting method.

Microlevel analysis

For subsequent understanding, the coding for the AOIs are: title = T, task instruction = J, region x = R, target lines = G, x-axis = X, y-axis = Y and legend = L

The eyePatterns software from West et al. (2006) can also analyze patterns within the sequences. A theory driven pattern approach (top-down) would assume a standard pattern would be a sequence like this: JRGRLG. This pattern has, however, only occurred in one sequence from a focus & blur participant and the performance of this user was not the best (74.7 seconds), which suggests that faster participants did not need to look at the legend. Another possibility would be that the participants use a similar sequence but with the legend right after region x (JRLGRG). Again, this has been found only once by an average performer from the focus & blur highlighting group. So what is the most used pattern for fast performers?

If the starting AOI has to be the task instruction, then the most used pattern (bottom-up) for fast performers is the following: JRGRG, which is used by 44% of the fast performers, but also by 50% of the slow performers. The most used pattern overall, however, is the back and forth between region x and target lines, which can go from three letters up to ten letters (i.e. RGr, RGRG, RGRGR,...,RGRGRGRGRG). The three letter pattern is used by 96.15% of all sequences, while the ten letter pattern is still used by 69.23%. These findings have been assumed to occur, since the task was based on comparing four different regions, which thus made participants look back and forth from map to graph. The next most common pattern has the task instruction AOI in it, though not at the beginning (i.e. RGRGJ), again in any combination of length. These findings advocate that the most efficient performers concentrated on R and G and J with little else. In order to find out more, the following Table 5 illustrates a transition probability matrix (West et al. 2006), which may give more insight into the different sequences of the two highlighting methods.

	T	R	L	X	Y	J	G	Total	focus & blur
T	0	4	0	0	0	5	1	10	
	0.0%	40.0%	0.0%	0.0%	0.0%	**50.0%**	10.0%	100%	
	0.0%	2.1%	0.0%	0.0%	0.0%	5.82%	0.54%		
R	1	0	11	2	4	41	131	190	
	0.53%	0.0%	5.79%	1.06%	2.11%	21.58%	**68.95%**	100%	
	20.0%	0.0%	**47.83%**	6.67%	13.34%	**47.68%**	**69.69%**		
L	0	13	0	1	1	6	2	23	
	0.0%	**56.53%**	0.0%	4.35%	4.35%	26.09%	8.7%	100%	
	0.0%	6.81%	0.0%	3.34%	3.34%	6.98%	1.07%		
X	0	5	0	0	0	1	20	26	
	0.0%	19.24%	0.0%	0.0%	0.0%	3.85%	**76.93%**	100%	
	0.0%	2.62%	0.0%	0.0%	0.0%	1.17%	10.64%		
Y	0	8	1	2	0	6	13	30	
	0.0%	26.67%	3.34%	6.67%	0.0%	20.0%	**43.34%**	100%	
	0.0%	4.19%	4.35%	6.67%	0.0%	6.98%	6.92%		
J	4	42	10	6	8	0	21	91	
	4.4%	**46.16%**	10.99%	6.6%	8.8%	0.0%	23.08%	100%	
	80.0%	21.99%	43.48%	20.0%	26.67%	0.0%	11.18%		
G	0	119	1	19	17	27	0	183	
	0.0%	**65.03%**	0.55%	10.39%	9.29%	14.76%	0.0%	100%	
	0.0%	**62.31%**	4.35%	**63.34%**	**56.67%**	31.4%	0.0%		
Total	5	191	23	30	30	86	188	553	

Table 5: Transition probability Matrix for both highlighting methods. (focus & blur on upper table, color on lower table).

	T	R	L	X	Y	J	G	Total	Color
T	0	1	0	0	0	8	2	11	
	0.0%	9.1%	0.0%	0.0%	0.0%	**72.73%**	18.19%	100%	
	0.0%	0.42%	0.0%	0.0%	0.0%	7.7%	0.91%		
R	2	0	20	1	6	49	158	236	
	0.85%	0.0%	8.48%	0.43%	2.55%	20.77%	**66.95%**	100%	
	28.58%	0.0%	**42.56%**	3.34%	16.67%	**47.12%**	**71.18%**		
L	0	27	0	2	1	8	6	44	
	0.0%	**61.37%**	0.0%	4.55%	2.28%	18.19%	13.64%	100%	
	0.0%	11.3%	0.0%	6.67%	2.78%	7.7%	2.71%		
X	1	3	4	0	3	1	18	30	
	3.34%	10.0%	13.34%	0.0%	10.0%	3.34%	**60.0%**	100%	
	14.29%	1.26%	8.52%	0.0%	8.34%	0.97%	8.11%		
Y	1	10	3	2	0	5	14	35	
	2.86%	28.58%	8.58%	5.72%	0.0%	14.29%	**40.0%**	100%	
	14.29%	4.19%	6.39%	6.67%	0.0%	4.81%	6.31%		
J	3	56	15	6	4	0	24	108	
	2.78%	**51.86%**	13.89%	5.56%	3.71%	0.0%	22.23%	100%	
	42.86%	23.44%	31.92%	20.0%	11.12%	0.0%	10.82%		
G	0	142	5	19	22	33	0	221	
	0.0%	**64.26%**	2.27%	8.6%	9.96%	14.94%	0.0%	100%	
	0.0%	**59.42%**	10.64%	**63.34%**	**61.12%**	31.74%	0.0%		
Total	7	239	47	30	36	104	222	685	

The transition probability matrix contains three values, The first number is the count, the second is the row probability and the third is the column probability. The row probability indicates which AOI from the row leads most likely to the AOI in the column. The column probability indicates which AOI from the column comes from the AOI in the row.

J$_{blur}$ to **R** 46.16%	**R**$_{blur}$ to **G** 68.95%	**G**$_{blur}$ to **R** 65.03%	**L**$_{blur}$ to **R** 56.53%
J$_{color}$ to **R** 51.86%	**R**$_{color}$ to **G** 66.95%	**G**$_{color}$ to **R** 64.26%	**L**$_{color}$ to **R** 61.37%

Table 6: Row transition probabilities for both highlighting methods (highest value only).

The transition probabilities for both highlighting methods do not reveal significant differences. The highest probabilities in Table 5 for both highlighting methods are marked with bold letters. The structure is very much identical and the transitions differ only in the level of probability. There are some notable differences, such as the higher probability for color highlighting participants to go from J to R or from L to R. The probability that the legend was fixated after the target lines, for example, is doubled for the color group (10.64% versus 6.35%). Nonetheless, the differences are subtle, which was expected since the two highlighting methods did not perform significantly different, even though focus & blur was slightly more efficient.

6 Discussion

This section discusses the user study results in light of the proposed research questions and hypothesis. Furthermore, the findings of this thesis are put into the context of the newest developments in research.

6.1 Research Questions

6.1.1 RQ 1

What is the ideal level of blur for focus & blur highlighting with regard to efficiency and effectiveness of object identification and context awareness?

The findings of Test 1 revealed that the ideal level of blur lies somewhere around 150% of the highlighted object's outline thickness. The 150% blur level was the most effective (100% correct answers) and together with the 100% blur level the most efficient. The 50% blur level was the least efficient, and the eye-tracking results moreover showed that the average time to first fixation was significantly the longest. The 50% blur level was also the only parameter where participants perceived other symbols as the highlighted object, which suggests that the blur level was not salient enough. The 200% blur level yielded both slightly longer average task completion times and two wrong answers where a false amount of remaining symbols was reported. This indicates that participants were not able to find the other blurred symbols very quickly. It should be noted, however, that the task completion times were not significant for any of the blur levels. Nonetheless, the user study shows a tendency towards the 150% level blur. Now, how useful is this finding? The relative blur value can only be applied when using a normal Gaussian blur filter. Also, the document's pixel density should be considered. In normal desktop viewing conditions, the level of blur is calculated by multiplying the object's outline thickness in pixel with 0.020833[20] multiplied with the ppi of the document. For example, if the document has a pixel density of 300ppi and a graph containing lines with thickness of 5 pixels, which should be highlighted, the blur level for these lines should be set at 31.25 Gaussian pixel blur radius. This level of blur is by no means the final answer, yet it should give potential users a good starting point and a basis to work with. The same amount of blur has been used for Test 2 to evaluate the highlighting approach in visual analytics tasks.

In conclusion, the results are comparable to the findings of Kosara et al. (2002b), where they noted that a pixel diameter of around 11 was sufficient, while 7 pixels were not enough (similar to the 50% level blur in this study). Kosara et al. (2002b) calculate the 11 pixels as the diameter of a circle over which information of 1 pixel is spread. Adapted to the blur level in this user study, the spread of the 150% blur level is 9 pixel and the 200% level is 11 pixel.

6.1.2 RQ 2

Is there a significant difference in efficiency and effectiveness between focus & blur and color highlighting to support visual analytics tasks in coordinated displays?

Efficiency

The overall results showed that there is no significant difference in efficiency between focus & blur and color highlighting in any of the tasks. There is even a slight trend towards focus

20 For this user study, the documents ppi was set to 72 (2 pixel lines * 0.020833 * 72 = 3 pixel, which is 150% of the pixels line thickness)

& blur being slightly more efficient than color. In Tasks 1, 2 and 4, the participants with the focus & blur condition were slightly more efficient, while in Task 3, the color highlighting was slightly more efficient. The eye-tracking results showed that from a qualitative point of view, the gaze plots do not differ significantly between the two highlighting methods, but show clear differences between slower and faster performers, regardless of the highlighting method. The quantitative eye-tracking data, moreover, did not reveal significant differences across the board. However, some aspects of the four tasks should be discussed in more detail in order to fully understand the results.

In Task 1, where participants with the focus & blur method were moderately more efficient, the eye-tracking results showed that there was a significant difference in the average fixation time of the AOI target lines. This suggests that the focus & blur group could see the position of the highlighted timeline better than the color group. The fact that, in the focus & blur highlighting method, the context loses contrast through the blurring could have led to a better relative positioning of the highlighted timeline with regard to the y-axis. For the color condition group, however, the highlighted timeline's context did not lose contrast and could therefore have had an impact on the relative positioning of the timeline with the 100 crime rate mark. This assumption is further confirmed by the fact that the color group looked significantly longer at the y-axis than the focus & blur group.

The eye-tracking results for Task 2 revealed that the mean transition times from region x to the target lines were significantly shorter for the focus & blur condition. This, however, must not be mistaken for a better guidance of attention, for the task required to compare the region x with other regions of the same income class, which led to the fact that some participants, after having fixated region x, looked at the legend first and only afterwards at the target lines. The transition probability matrix reveals that the probability for a participant to go from region x to the legend is 2.69% higher for the color highlighting group. This blends well with the findings of the sequence analysis which suggested that the focus & blur condition had slightly more similar sequences, especially for their top performers. Another interesting issue is that this task could have been solved by the color highlighting group by highlighting only region x. The timelines were in fact classified with the same color as the boroughs of London to represent the different income classes. Now, a witty participant could have seen the three other timelines lying below region x's timeline for 2004 and, as a matter of fact, for the whole timespan of 2000 to 2010. However, every single participant from the color highlighting group either highlighted all boroughs from the same income class sequentially or simultaneously. Thus, none made use of the better context awareness that the color highlighting provides. The same would have been possible with focus & blur highlighting too, but one would have needed very sharp eyes to see the other timelines while they were blurred.

Task 3 was in average solved slightly faster by the color group. The eye-tracking data show no significant difference in mean transition times from target line to region x. When region x was found, participants were asked to verbally state the borough's number. It could be that the surrounding boroughs and their numbers, which were all blurred, confused participants.

Effectiveness

The results of this user study have shown some differences in effectiveness between the two highlighting methods. Overall, participants from the color highlighting group have in average answered 26.6% of the questions wrong, while the focus & blur participants answered only 10.9% of the answers wrong. Most errors on the color condition side occurred in Task 1, where the participants obviously had more problems with the positioning of the timeline with respect to the y-axis and the 100 crime rate mark. This suggests that the design of the visual analytics environment in this user study should be improved in the future, so that the color highlighting participants can see the y-axis helpline equally good as the focus & blur participants. Several participants from the color highlighting group moreover noted at the end of the user study that the y-axis could not be perceived very well.

Task 4 was a further task that raised problems for both highlighting methods. The differences between the two highlighting approaches in this task are not substantial, though. The effectiveness results should therefore be put into perspective and should not be overinterpreted in favor of focus & blur highlighting.

6.1.3 RQ 3

Is there a significant difference in satisfaction between focus & blur and color highlighting when used in visual analytics tasks?

The SUS questionnaire that participants had to fill out after the testing was done provided no significant differences between the two highlighting methods. Color highlighting achieved a slightly higher final SUS score (81.7 versus 77.5), but both values represent good results (Bangor et al. 2008). Participants found both highlighting methods visually pleasing and helpful in the performance of the tasks. They also felt confident about using both highlighting approaches. In addition to the SUS questionnaire, participants could give overall statements regarding the highlighting methods at the end of the user study. Most focus & blur participants stated that they found the highlighting method good looking, but two indicated that they felt slightly dizzy after having looked at the blurred stimuli. One participant even mentioned that for him/her everything looked blurry and not just the objects that indeed were blurred by the highlighting. These statements suggest that the blur level should be personally adjustable by every user. On the other hand, one participant noted that s/he did not recognize actively that some objects were highlighted when s/he moved their mouse over an area or data line, but still managed to solve the tasks just fine. As expected, the participants from the color highlighting group did not make any subjective statements. Color highlighting is indeed widely known and people are used to it.

6.1.4 Hypothesis

Based on the findings in this thesis and the discussion of research questions, the hypothesis proposed in the introduction can be retained: there is no significant difference in efficiency, effectiveness and satisfaction between focus & blur highlighting and color highlighting.

6.1.5 Other Findings

Participants needed to indicate the frequency of their video games consumption, specifically that of 3D action games. The reason for this were recent findings which suggest that action video game players (VGPs) have increased stimulus-response mappings in visual search tasks (Castel et al. 2005), better selective attention management (Green & Bavelier 2003), and even higher spatial resolution of vision (measured by crowding, or target-distractor distance, (Green & Bavelier 2007). DOF is a method commonly used in 3D games; therefore, it was assumed that the VGPs would perform slightly more efficient than the NVGPs. Yet, the user study in this thesis showed mixed results. Some of the avid VGPs performed far above average, while others who indicated to be VGPs were below average; there was no significant effect of group. The number of regular VGPs in this study was most likely too small (N = 9) to come to any conclusion.

6.1.6 Research Context

Ware (2012), Wolfe and Horowitz (2004) and Kosara (2001, 2002a,b) have already stated that blur is a good candidate for the guidance of attention. The findings in this thesis are congruent with these statements. All participants from the focus & blur condition found the highlighted areas just as good or sometimes even slightly faster than the participants from the color condition. Kosara et al (2002b) found in their user study that SDOF was not significantly different for visual search tasks than color highlighting. Neither does this thesis' user study show significant differences in efficiency and effectiveness between focus & blur and color for visual analytics tasks. They are thus comparable to other empirical findings and theoretical assumptions (i.e. Robinson 2011 or Liang and Huang 2010).

6.1.7 Limitations and Outlook

Do the findings in this thesis suggest that focus & blur highlighting should be preferred over color highlighting from now on? No, of course not. Firstly, color highlighting is very much capable and has been used and proven for many years. Focus & blur highlighting should rather be viewed as an alternative to color highlighting for when color is not the ideal solution (for example, when the information must be color coded to represent another attribute). Moreover, the findings of this thesis solely show a tendency that focus & blur highlighting can be used just as successfully as color for visual analytics tasks in a tow-view coordinated display and for the use of approximately 20 minutes continuously. Some participants of the focus & blur group have indicated that they felt slightly dizzy after the testing session. Hence, further studies should be performed to investigate the effects of using blur over longer periods of time as a highlighting method. Furthermore, as Nakayama and Silverman (1986) have shown, blur can be used in conjunction with other visual variables and still be attentive in early vision. Combinations with blur should thus be examined more closely in future studies.

The focus & blur highlighting method has the property of making the context of the highlighted object harder to read and identify (Robinson 2011). Task 2 in this user study also tried find out whether contextualization is still possible with this method. However, participants mostly switched back and forth between the different regions and highlighted them sequentially or simultaneously with the use of the click and freeze option. Tasks 1 and 4, however, showed that the reduction of readability of context even helped participants to solve these tasks. The experience from this user study showed that relative positioning within the blurred context is easily possible; yet the detection of other objects' colors, especially in line graphs, proved to be problematic. The questions that need to be further investigated are: what tasks in visual analytics require which level of contextualization and in which situations is the lack of contextualization even helpful or more efficient? Additionally, implementations where the blur level can be reduced momentarily to improve context awareness is something that also needs additonal research.

7 Conclusion

This thesis evaluated the usability of the alternative highlighting method focus & blur. Based on literature findings (Robinson 2011, Koara 2001, Kosara et al. 2002a,b) which suggested that blur is a preattentive feature and a probable variable for guidance of attention, a user study was conducted. 31 participants were tested to find out whether this highlighting approach was significantly different in efficiency, effectiveness and satisfaction from standard color highlighting for visual analytics tasks. In a short pretest, participants were shown four stimuli with different levels of blur and were asked to perform a visual search task. The results suggested that too low levels of blur proved to be ineffective and less efficient, while too high levels of blur reduced the ability to perceive the context of the highlighted object[21]. The second test required participants to solve four visual analytics tasks in a two-view coordinated display. The testing was recorded using a near infrared eye-tracking device to provide further insight into how well the highlighting methods worked as a guidance of attention. The results showed that there is no significant difference in efficiency, effectiveness and satisfaction between focus & blur highlighting and color highlighting. The eye-movement data showed that there are no significant differences overall between the two highlighting methods, but a clear discrepancy between fast performers and slow performers. The subjective usability metric satisfaction was assessed with the use of an adapted SUS questionnaire. The final SUS scores of both highlighting methods ranged well above the *good* mark proposed by Bangor et al.

21 The most efficient and effective blur ratio is presented in section 6.

(2008). However, some participants from the focus & blur group stated that they felt slightly dizzy after the testing session.

The results of this thesis align with the conclusions of Kosara et al. (2002b) who noted that SDOF is not significantly different in efficiency and effectiveness from color for visual search tasks. In conclusion, this user study suggests that focus & blur can be used as a capable alternative to color highlighting; yet further research is needed to find out how focus & blur performs over extended periods of time, how it can be conjuncted with other visual variables for visual highlighting, and, finally, how it affects context awareness for better or worse in visual analytics tasks.

8 References

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction, 24(6). 574–594.

Barkowsky, T., Freksa, C., Hegarty, M., & Lowe, R. (2005). Reasoning with mental and external diagrams: Computational modeling and spatial assistance. In Proceedings. 21–23

Baudisch, P., Good, N., & Stewart, P. (2001). Focus plus context screens: combining display technology with visualization techniques. Proceedings of UIST'01, 31-34, Orlando.

Becker, R., & Cleveland, W. (1987). Brushing scatterplots. Technometrics, 29(2), 127–142.

Bertin, J. (1983 / 2010). Semiology of Graphics. Redlands, USA. *Esri Press*.

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), Usability evaluation in industry. 189–194. London, UK. Taylor & Francis.

Carpendale, M. S. T. (2003). Considering visual variables as a basis for information visualisation. Computer Science TR# 2001-693(16).

Castel, A. D., Pratt, J., & Drummond, E. (2005). The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. Acta psychologica, 119(2). 217–30.

Collins, C., & Carpendale, S. (2007). VisLink: revealing relationships amongst visualizations. IEEE transactions on visualization and computer graphics, 13(6). 1192–9.

Çöltekin, a., Fabrikant, S. I., & Lacayo, M. (2010). Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. International Journal of Geographical Information Science, 24(10). 1559–1575.

Çöltekin, A. (2006). Foveation for 3D visualization and stereo imaging. PhD Thesis. Institute of Photogrammetry and Remote Sensing, Helsinki University of Technology.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. The Behavioral and brain sciences, 24(1), 87–114

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. Annual review of neuroscience, 193–222.

Di Lollo, Vincent Di, Enns, J., & Rensink, R. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. Journal of Experimental Psychology: General, 129(4), 481–507.

Di Lollo, V., Kawahara, J., Zuvic, S. M., & Visser, T. A. W. (2001). The preattentive emperor has no clothes: a dynamic redressing. Journal of Experimental Psychology: General, 130(3), 479.

Duchowsky, A. (2007). Eye Tracking Methodology. London: Springer.

Egeth, H. E., & Yantis, S. (1997). Visual attention: control, representation, and time course. Annual review of psychology, 48, 269–97.

Giller, V., Tscheligi, M., & Schrammel, J. (2001). Experimental evaluation of semantic depth of field, a preattentive method for focus+ context visualization. Technical Paper TR-VRVis-2001-021, VRVis Research Center, http://www.vrvis.at/, Austria.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. Nature, 423(6939), 534–7.

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. Psychological science, 18(1), 88–94.

Griffin, A., & Robinson, A. (2010). Comparing color and leader line approaches for highlighting in geovisualization. GIScience 2010, Zurich, Switzerland.

Häberling, C. (2003). Topographische 3D-Karten: Thesen für kartographische Gestaltungsgrundsätze. PhD thesis, ETHZ, Zurich, Switzerland

Hardisty, F., & Robinson, A. C. (2011). The GeoViz Toolkit: Using component-oriented coordination methods for geographic visualization and analysis. International journal of geographical information science : IJGIS, 25(2), 191–210.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. Nature, 394(6693), 575–7.

Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (2010). Mastering the information age solving problems with visual analytics. Goslar, Germany. Eurographics Association.

Kerren, A., Ebert, A., & Meyer, J. (2007). Human-centered visualization environments.

Kosara, Robert. (2001). Semantic Depth of Field-Using Blur for Focus+ Context Visualization. PhD Thesis. Institute of Computer Graphics and Algorithms, Vienna University of Technology.

Kosara, R., Miksch, S., & Hauser, H. (2001). Semantic depth of field. IEEE Symposium on Information Visualization. INFOVIS 2001, 97–104.

Kosara, R., Miksch, S., & Hauser, H. (2002a). Focus+context taken literally. IEEE Computer Graphics and Applications, 22(1), 22–29.

Kosara, Robert, Miksch, S., & Hauser, H. (2002b). Useful properties of semantic depth of field for better f+ c visualization. Proceedings of the symposium on Data Visualization. Eurographics Association, 2002.

Kruskal, J. B. & Wish, M. (1978). Multidimensional Scaling. Sage Publications, Beverly Hills, California.

Lee, D., & Chun, M. M. (2001). What are the units of visual short-term memory, objects or spatial locations? Perception & psychophysics, 63(2). 253–7.

Lewis, J., & Sauro, J. (2009). The factor structure of the system usability scale. Human Centered Design, 1, 94–103.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. Doklady Physics, 10, 707- 710.

Liang, J., & Huang, M. L. (2010). Highlighting in Information Visualization: A Survey. 2010 14th International Conference Information Visualisation, 79–85.

Liebovitch, L. S. (2005). Why the Eye Is Round. Advances in Organ Biology, 10, 1–19.

Liston, K., Fischer, M., & Kunz, J. (2000). Designing and Evaluating Visualization Techniques for Construction Planning. Computing in Civil and Building Engineering. 1293–1300.

MacEachren, A., Hardisty, F., Dai, X., & Pickle, L. (2003). Supporting visual analysis of federal geospatial statistics. Communications of the ACM, 46(1), 63–64.

MacEachren, A. M. (2004). How maps work: Representation, Visualization, and Design. Paperback edition. New York, USA. Guilford Press.

Nakayama, K., & Silverman, G. (1986). Serial and parallel processing of visual feature conjunctions. Nature, 320(6059). 264-265.

Needleman, S. B. & Wunsch C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins, Journal of Molecular Biology, 48, 3. 443-453.

Nielsen, J. (1993). What is Usability. In: Usability Engineering. California, USA. Morgan Kaufmann.

Nunnally, J. C., Bernstein, I. H., & Berge, J. M. F. (1967). Psychometric theory (Vol. 2). McGraw-Hill New York.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. Perception & psychophysics, 64(7), 1055–67.

Philipsen, G. (1994). Effects of six different highlighting modes on visual search performance in menu options. International Journal of Human Computer Interaction, 6(3), 319-335.

Roberts, J. C., & Wright, M. a. E. (2006). Towards Ubiquitous Brushing for Information Visualization. Tenth International Conference on Information Visualisation (IV'06), 151–156.

Robinson, AC. (2006). Highlighting techniques to support geovisualization. Proceedings of the ICA Workshop on Geovisualization and Visual Analytics.

Robinson, AC. (2011). Highlighting in Geovisualization. Cartography and Geographic Information Science, 38(4). 373-383.

Russ, J. C. (2006). The Image Processing Handbook. (Springer, Ed.) Image Processing, 56. CRC Press, 275–86.

Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? Proceedings of the 2011 annual conference on Human factors in computing systems, ACM, 2215 – 2224.

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

Sperling, G. (1960). The information available in brief visual presentations. Psychological monographs: General and applied, 74(11).

Steinberger, M., Waldner, M., Streit, M., Lex, A., & Schmalstieg, D. (2011). Context-preserving visual links. IEEE transactions on visualization and computer *graphics*, 17(12), 2249–58.

Swienty, O. (2008). Attention-Guiding Geovisualisation. PhD Thesis. Technischen Universität München.

Thomas, J. J., & Cook, K. A. (2005). Illuminating the Path. The Research and Development Agenda for Visual Analytics. New York, USA, IEEE CS Press.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. Nature, 381(6582), 520-522.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. Nature, 428(6984), 751–4.

Treisman, AM, & Gelade, G. (1980). A feature-integration theory of attention. Cognitive psychology, 136, 97–136.

Treisman, Anne. (1985). Preattentive Processing in Vision. Computer vision, graphics, and image processing, 177, 156–177.

Treisman, Anne, & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. Psychological review, 95(1), 15–48.

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. Journal of cognitive neuroscience, 13(4). 454–61.

Waldner, M., Puff, W., & Lex, A. (2010). Visual links across applications. Proceedings of Graphics Interface 2010. Canadian Information Processing Society, 129–136.

Ware, C. (2012). Information visualization Perception for Design. USA, Morgan Kaufmann.

Ware, C., & Bobrow, R. (2004). Motion to support rapid interactive queries on node--link diagrams. ACM Transactions on Applied Perception, 1(1), 3–18.

Ware, C., & Bobrow, R. (2005). Supporting visual queries on medium-sized node–link diagrams. Information Visualization, 4(1), 49–58.

West, J.M., Haake, A. R., Rozanski, E. P., & Karn, K. S. (2006). EyePatterns: software for identifying patterns and similarities across fixation sequences. In: Proceedings of the ETRA 2006, 27–29 March 2006, San Diego, CA. New York: ACM Press, 149–154.

Wickens, C. D., Alexander, A. L., Ambinder, M. S., & Martens, M. (2004). The role of highlighting in visual search through maps. Spatial vision, 17(4-5), 373–88.

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. Psychonomic Bulletin & Review, 1(2), 202–238.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? Nature reviews. Neuroscience, 5(6), 495–501.

Yantis, S. (2005). How visual salience wins the battle for awareness. Nature neuroscience, 8(8), 975–7.

Zhai, S., Wright, J., Selker, T., & Kelin, S.-A. (1997). Graphical Means of Directing User's Attention in the Visual Interface. Human-computer interaction: Interact'97: IFIP TC13 International Conference on Human-Computer Interaction. Sydney, Australia, 59–66.

Ziemkiewicz, C., & Kosara, R. (2010). Laws of attraction: From perceptual forces to conceptual similarity. Visualization and Computer, 16(6), 1009–1016.

9 Appendix

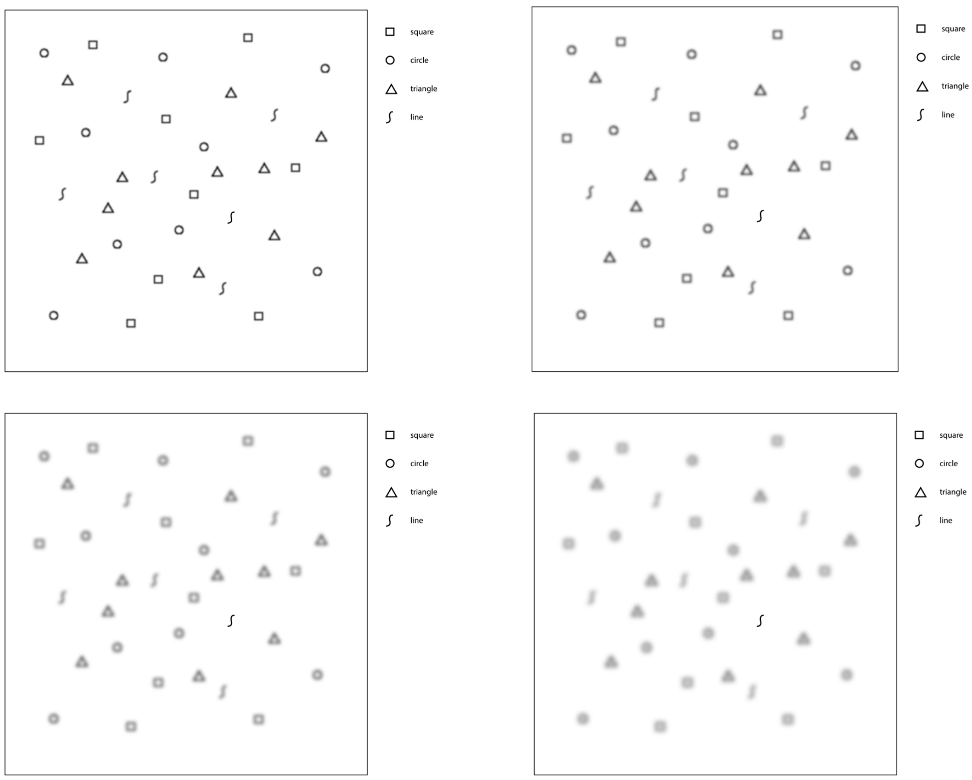


Figure 66: Test 1 Stimuli. Blur level 1 to 4

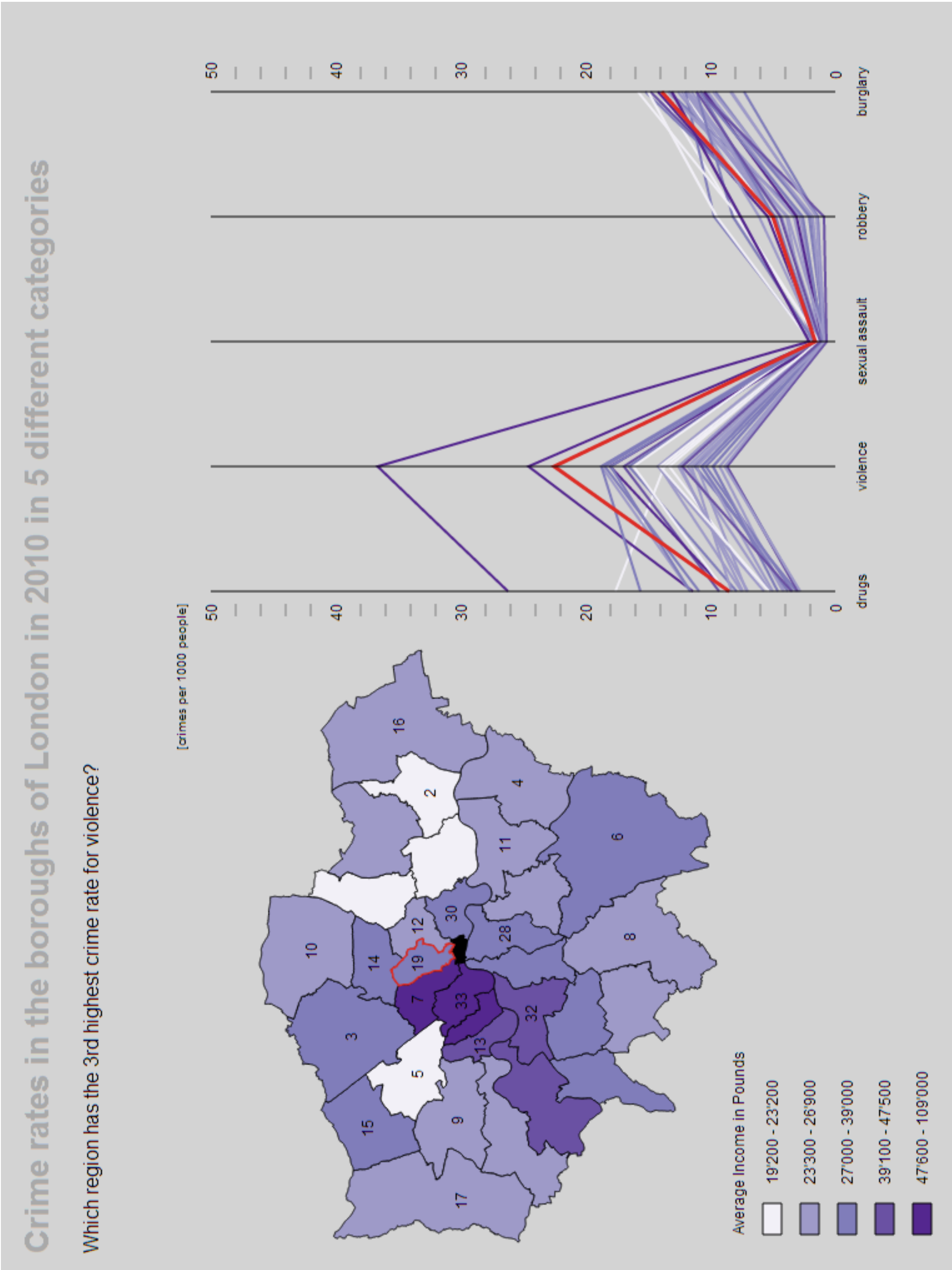


Figure 67: Test 2. Color stimuli, PCP

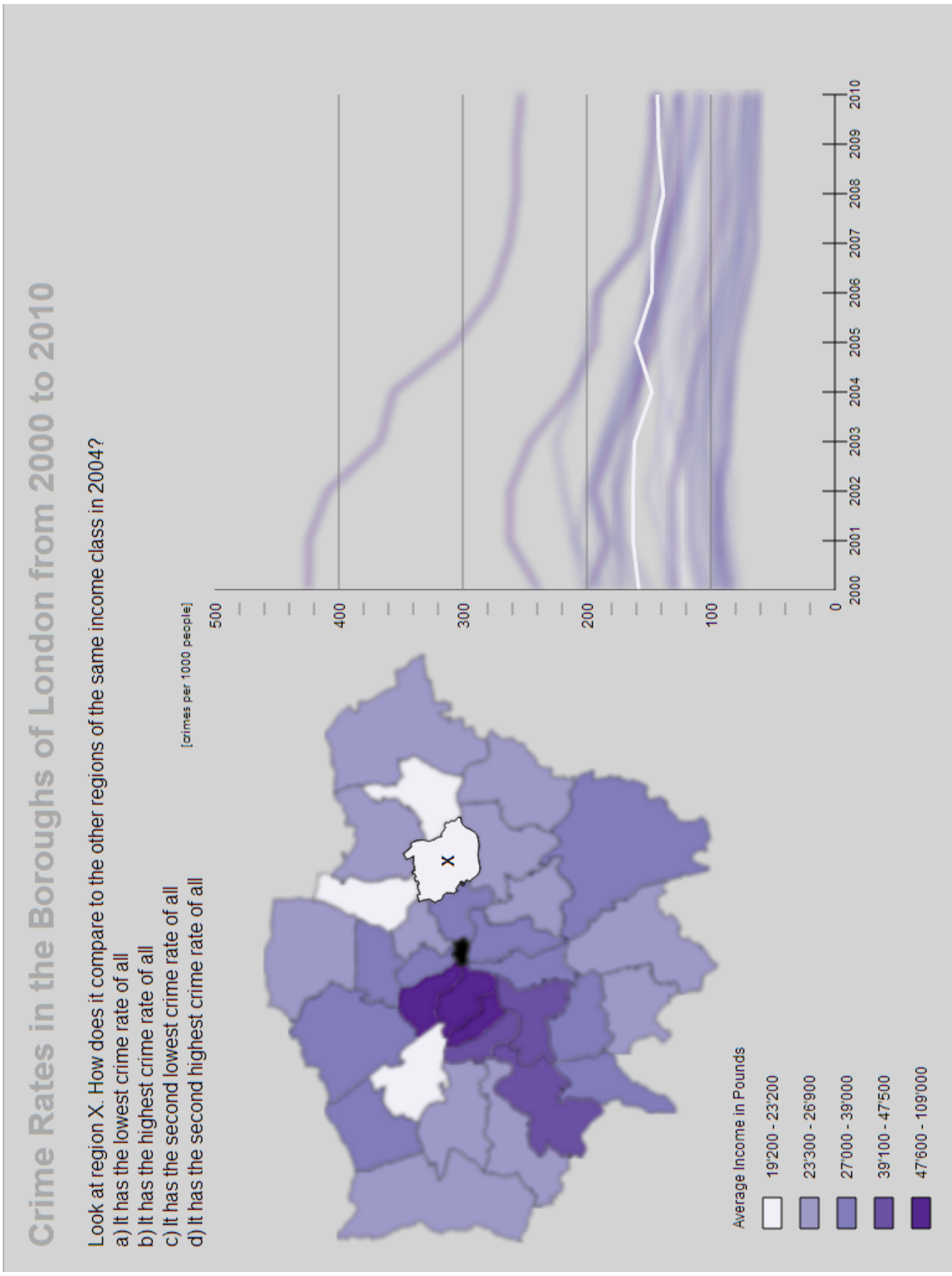


Figure 68: Test 2. Focus & blur stimuli, timeline.

```
/////////////////////
//Function to highlight the mouseover target [boroughs]
/////////////////////

function highlight(evt){

nplace = document.getElementById('newPlace');

	if(evt.type=="mouseover"){
		var highlight = evt.target;
		var point_pool = document.getElementById('london').childNodes.length;
		point_length = (point_pool-1)/2;
		var point_id = highlight.getAttributeNS(null,'id')

	for (j = 1; j <= point_length; j++){

		if( point_id == j){
			highlight2 = highlight.cloneNode(true);
			nplace.appendChild(highlight2);
			highlight2.setAttributeNS(null,"id","newBoro")
			highlight2.setAttributeNS(null,"stroke","black");
			highlight2.setAttributeNS(null,"stroke-width","120");
			highlight2.setAttributeNS(null,"pointer-events","none");
			highlight2.setAttributeNS(null,"filter","none");
			}

		else{
			var highlight_obj = document.getElementById(j);
			highlight_obj.setAttributeNS(null,"filter","url(#f1)");
		}
	}
	}

	else if(evt.type=="mouseout"){
		var targ = evt.target.getAttributeNS(null,'id');

			if(friz == 1){
				for (j = 1; j <= point_length; j++){
				var highlight_obj = document.getElementById(j);
				highlight_obj.setAttributeNS(null,"filter","none");
				}
			}
	nplace.removeChild(highlight2);
	}
}
```

Figure 69: Sample code for the highlighting function in Javascript.

The University of Zurich - **Participant Information Statement and Consent Form** **A usability evaluation of focus highlighting**
October 23-31 and November 1-3, 2012
Participant No:

Purpose of study

You are invited to participate in a study regarding an evaluation of highlighting methods. We hope to learn more about the design of blur as a highlighting method.

Description of study and risks

If you decide to participate, we will ask you to begin by filling out a short background questionnaire including demographic information. This will be followed by a session at the computer where you will be asked to use a thematic map and diagram. During this process we will record your interactions with the computer using a webcam, audio recorder and eye tracking. The eye tracking device is non-contact, uses near infrared light and should not cause any discomfort. After the experiment we will ask you to fill out a second questionnaire.

The whole procedure should take approximately 30 minutes and there are no particular risks or benefits to you from participating in this experiment.

Confidentiality and disclosure of information

Any information that can be identified with you in connection with this study will remain confidential and will be disclosed only with your permission. If you give us permission by signing this document, we plan to publish the results of this research in scientific publications. In any publication, information will be provided in such a way that you cannot be identified.

Compensation

We do not provide any compensation for your participation in this experiment, nor are there any costs for you for your participation.

Your consent

Your decision whether or not to participate will not prejudice your future relations with University of Zurich. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without prejudice.

If you have any questions, please feel free to ask me. If you have any additional questions later, I (kaspar.fischer@geo.uzh.ch) will be happy to answer them.

You will be given a copy of this form to keep.

Page 1 of 2

The University of Zurich - **Participant Information Statement and Consent Form (continued)** A usability evaluation of focus highlighting
October 23-31 and November 1-3, 2012
Participant No:

You are making a decision whether or not to participate. Your signature indicates that, having read the information provided above, you have decided to participate.

……………………………………. ………………………………

Signature of Research Participant Signature of Experimenter

……………………………………. ………………………………

Please PRINT name Please PRINT name

…………………………………….

Date and Place

The University of Zurich - **Participant Information Statement and Consent Form (continued)** Evaluating blur as a highlighting method for visual analytics tasks
October 23-31 and November 1-3, 2012
Participant No:

REVOCATION OF CONSENT

Evaluating blur as a highlighting method for visual analytics tasks

I hereby wish to WITHDRAW my consent to participate in the research proposal described above and understand that such withdrawal WILL NOT jeopardize any treatment or my relationship with The University of Zurich.

……………………………………. ………………………………

Signature Date

…………………………………….

Please PRINT name

This section of Revocation of Consent should be forwarded to Kaspar Fischer, Geographic Information Visualization and Analysis, Dept. of Geography, University of Zurich, CH-8057, Zurich.

Pre Test Questions

A.

Name: ....

Age: ....

Sex: ....

Semester: ....

Major / Minor: ....

Diopter: ....

B.

1. On a scale 1 to 5 how often do you use thematic maps in your daily life?

☐	☐	☐	☐	☐
1 (least)	2	3	4	5 (most)

2. On a scale 1 to 5 how familiar are you with GIS software?

☐	☐	☐	☐	☐
1 (least)	2	3	4	5 (most)

3. On a scale 1 to 5 how proficient are you with the English language?

☐	☐	☐	☐	☐
1 (least)	2	3	4	5 (most)

4. How often do you play 3D video games?

☐	☐	☐	☐	☐
never	once a month	once a week	2-4 times a week	every day

2. SUS QUESTIONNAIRE

Statement	Strongly disagree				Strongly Agree
1. I think that I would like to use this highlighting method frequently	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
2. I found the highlighting method unnecessarily complex	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
3. I found the highlighting method visually pleasing	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
4. I think that I could not see the highlighting very well	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
5. I found that the highlighting method helped me perform the tasks	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
6. I found the performance of the highlighting method not satisfyingly fast enough	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
7. I Would imagine that most people would learn to use this highlighting method very quickly	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

8. I found the highlighting method distracted me from performing the tasks	Strongly disagree				Strongly Agree
	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

9. I felt very confident using this highlighting method	Strongly disagree				Strongly Agree
	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

10. I need to practice a lot more before I could get going with this highlighting method	Strongly disagree				Strongly Agree
	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

3. Comments, remarks?

Personal declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Zurich, 31 January 2013

Kaspar Fischer