Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study with Eye Movement Analysis

Arzu Çöltekin, Simone Garlandini, Benedikt Heil, Sara Irina Fabrikant
Dept. of Geography, University of Zurich, Winterthurerstr. 190, 8057 Zurich, Switzerland
{arzu.coltekin, simone.garlandini, benedikt.heil, sara.fabrikant}@geo.uzh.ch

ABSTRACT: This paper proposes combining traditional usability methods with the analysis of eye movement recordings to evaluate interactive map interfaces, and presents a case study in support of this approach. This case study evaluates of two informationally equivalent but differently-designed online interactive map interfaces presented to users. In a mixed factorial experiment thirty participants were asked to solve three typical map use tasks using one of the two interfaces, while we measured user *Satisfaction*, *Efficiency* (completion time) and *Effectiveness* (accuracy) with standard *SEE* usability metrics. While traditional (bottom line) usability metrics can reveal a range of usability problems, they may be enhanced by additional procedural measures such as eye movement recordings. Eye movements have been shown to reveal the amount of cognitive processing a display requires and where these cognitive resources are required. Therefore, we can establish how a display may or may not facilitate task completion by analyzing eye movement recordings. User satisfaction information related to stimuli (i.e., collected through standardized questionnaires) can also be linked to eye tracking data for further analysis. We hope that the presented methodology and case study will help cartographers and map interface designers to better identify design issues in their products, and that these insights will eventually lead to more effective and efficient online map interfaces.

KEYWORDS: geographic visualization, human computer interaction, eye movement analysis, interactive maps, interface evaluation

Introduction

About ten years ago, Howard and MacEachren (1996) predicted that digital "softcopy" maps were becoming the norm replacing paper maps, and as a result, design of interface tools would become as fundamental to cartography as the design of the maps themselves. The spread of high-bandwidth Internet, and access to it through the increasing use of location-based services with mobile devices (i.e., in-car navigation systems, personal digital assistants, cell phones, etc.) seems to have validated this prediction. In accordance with this development, new complex representation forms and interactive methods for visualizing geospatial data are available to large audiences with different levels of experience in handling them (Koua et al., 2006; Fabrikant et al., 2008). The need to assess the impact, usefulness and usability of these tools is increasing at the same rate as their rising availability and spreading versatility (Fuhrman et al., 2005; Koua et al., 2006; Nivala et al., 2008; Haklay and Zafiri, 2008). However, as identified by ***MacEachren and Kraak (2001), already at the turn of the century "the lack of established paradigms for conducting cognitive or usability studies with highly interactive visual map interfaces" remains as one of the challenges in geovisualization. We tackle this issue by bringing modern usability engineering techniques together with eye movement analysis.

Usability Engineering Usability engineering refers to a set of techniques and concepts for assessing a product or system's ease of use based on controlled experiments, system inspection and inquiry methods (Good et al. 1986; Nielsen, 1993). Typically users are provided with a specific set of tasks with a particular usage scenario, and in a specific context. Usability performance metrics such as satisfaction, efficiency and effectiveness (SEE) are employed to assess how easy the product or system is to use. Satisfaction refers to a user's attitude or preferences about the system, efficiency refers to how quickly the tasks are completed, and effectiveness refers to whether or not a task is successfully completed.

Iterative evaluation sessions allow usability researchers to identify most usability problems (Nielsen, 1993). In typical usability studies, human-system interactions are evaluated with direct observation, pen and pencil questionnaires, video analysis, key stroke and mouse click recordings. These evaluation procedures are often supported with other standard empirical methods such as think aloud protocols and interviews. When the procedure involves self-reports or interviews, various psychological and social factors can influence human behavior (and performance) and create bias in the results, such as short-term memory problems, anxiety or desire to "succeed" in a test situation. Therefore, with self-reports or interviews, what people say they do, is not always what they actually do. This can be particularly relevant in highly interactive systems used to solve complex problems when people might not be able to fully verbalize their own complex inference making. A viewer's *cognitive load* might become so high during task completion that verbal reports, or think-aloud protocols interfere with the quality of inference making. Eye movement recordings, on the other hand, can offer additional unobtrusive evidence of overt user behavior. Eye movement recordings are frequently viewed as a window into internal cognitive processes (Bojko, 2006; Goldberg et al., 2002). By studying them, we may be able to compensate for the excessive cognitive load that prevents the participant from remembering processes when self-reporting. Eye movement recordings are also very useful for identifying *where* problem areas are in system use and *how* the information is processed (Dix et al., 2004).

Within the cartography and geovisualization domains, traditional (static) map display evaluation methods also have been based on standard usability approaches (*i.e.* testing, inspection and inquiry methods). Recent research papers frequently point at the need for better, more suitable methods for evaluating maps, stating that the standard methods "may no longer be suitable for the growing range of map users, usage scenarios, and digital map devices" (Nivala, 2008), in particular with new interactive visualizations (Koua and Kraak, 2004). In this study, to explore the potential of eye tracking in this domain, inspired by Fabrikant et al. (2008), we propose a combination of traditional usability engineering methods with eye movement analysis for the empirical evaluation of interactive map interfaces.

Eye Movement Studies for Interface Evaluation and Usability Eye tracking has a history of nearly one hundred years in psychology, but early technology was cumbersome, difficult to use, and prohibitively expensive (Dix et al., 2004; Duchowski, 2007). Technological developments in recent decades have made eye tracking systems more accessible and available. Modern eye tracking systems allow fairly accurate recordings of pupil diameter, number of fixations (a fixation is when eyes are stationary during a given threshold of ~50 to ~500 milliseconds, in Irwin, 2004; Henderson and Ferreira, 2005 and Bojko, 2006), fixation durations and saccades (a saccade is the rapid eye movement that occurs between fixations) multiple times per second during a session. With these developments, eye movement analysis has been increasingly and successfully employed in various fields such as software, design and interactive web interface evaluation research and practice (e.g. Goldberg and Kotval, 1999; Byrne et al., 1999).

When utilizing eye tracking analysis to evaluate an interface, some common assumptions are such that *more fixations may indicate a less efficient search strategy, longer fixations may indicate difficulty with the display* and *plotting scan paths and fixations will allow documenting what people look at, how often and how long* (Goldbeg and Kotval, 1999; Dix et al., 2004; Bojko, 2006). When users are searching to find the correct link, button or another control on an online interface, typically two types of processes occur: a perceptual one (where user should locate/notice the target) and a cognitive one (where user cognitively computes the visual input and understands the function of the target). Eye movement analysis provides valuable quantitative and qualitative information on both stages of visual search and thereby complement SEE metrics (Goldberg and Kotval, 1999; Jakob and Karn, 2003). These observations, in accordance with earlier predictions (e.g. by Dix et al., 2004), have led some recent academic and industrial interface evaluation studies to combine eye movement analysis with other usability methods (e.g. Pretorius et al., 2005; Bojko, 2006).

Evaluating Interactive Map Interfaces with Usability Engineering Methods and Eye Movement Analysis Utilizing information that can be gathered by recording eye movements to understand the relationship between map reading and map design was reported as early as the 1970s (Steinke, 1987) The cartographic community showed interest in eye tracking until the 1980s, but after this decade, the interest seems to have nearly disappeared (Steinke, 1987; Brodersen et al., 2001; Fabrikant et al., 2008). This trend can be a result of a suboptimal cost-benefit relationship; eye movement analysis was financially costly to start and effort-intensive to finish. Today, eye tracking hardware is affordable and even though analyzing eye movement data still is a time-consuming and complex process, digital processing can arguably make it easier to process very large datasets in comparison to the analog methods used in 1970s and 1980s.

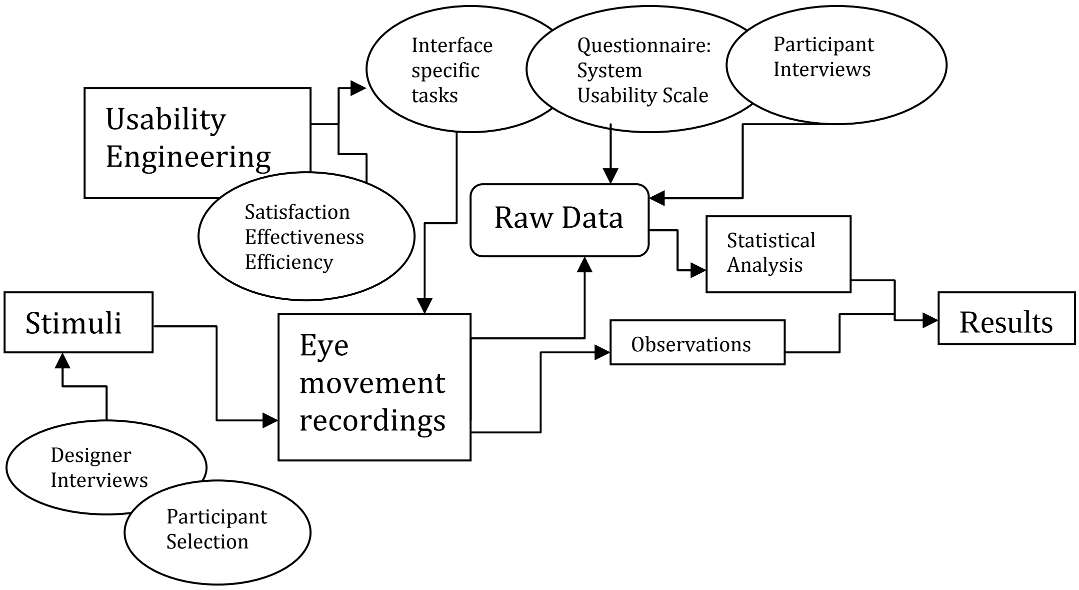
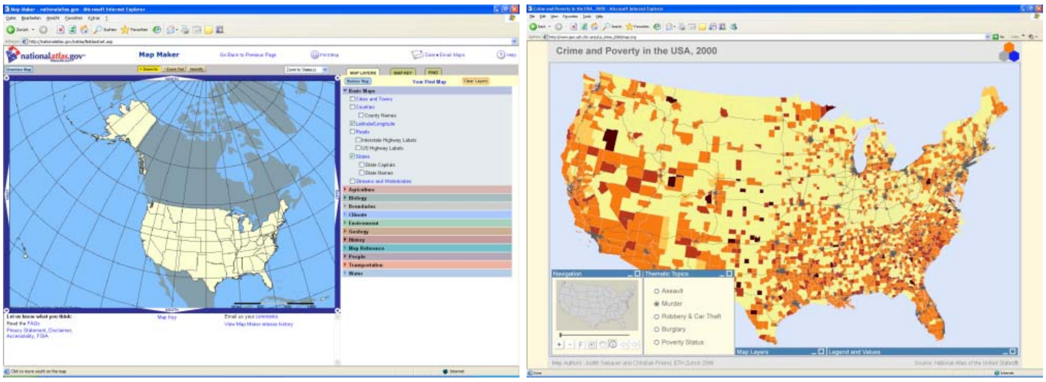


Figure 1. An overview of the process flow for proposed methodology.

Maps have also changed since 1980s. Interactive digital maps (as opposed to static paper maps) have become more complex to analyze due to added dynamic features. Digital interactive map interfaces typically come with two display elements: a cartographic data display area, where map itself is presented, and a set of graphical user interface (GUI) elements which allow for the interaction with the presented map data. The usability of such maps relies heavily on interface design (You et al., 2007). By employing the eye movement data collection method for complex interactive map interfaces, we can monitor a user's inference making process while interacting both with the map and the interface elements at the same time (Fabrikant et al., 2008). The questions *where*, *when*, *how long*, *how often* and *in which order* a display element was attended to during a task may allow us to more effectively interpret why task completion or inference making might be facilitated (or hindered) with a particular interface design, and whether the map interface is indeed utilized as intended by the designers. Procedural (eye movement) data combined with baseline effectiveness and efficiency data (i.e., accuracy and speed of response) provides added value to the process of systematically evaluating interactive map interfaces. An overview of our proposed combination of these methods can be seen as a process flow diagram in Figure 1.

Experiment

The proposed evaluation methodology has been applied to a controlled experiment comparing two interactive online map interfaces: the *Map Maker* service of the *National Atlas of the U.S.A.* (Natlas, 2008), and an interactive thematic map published on the *carto.net* website (Carto.net, 2008). Participants were asked to perform a set of map use tasks while their eye and mouse movements were being recorded. While the maps include the same statistical data, they differ significantly in the approach they take to map interface design (Figure 2). In other words, while they are informationally equivalent we contend that they are not computationally equivalent (Larkin and Simon, 1987).



(a) (b)

Figure 2. Screenshots of two interactive maps used in the study, a) National Atlas of the United States, b) Carto.net's SVG map showing crime and poverty data taken from National Atlas.

Based on Larkin and Simon's computational equivalence concept we hypothesize that map users will perform less efficiently with Natlas as it requires comprehension of a more complex interface.

Designer Interviews To better understand the design process and usage contexts, we first studied the stimuli by documenting both the technology that was involved in making them, and by interviewing the individuals who were involved in decision-making roles of the development and design process. Display designers responded to a ten-question online form that was inspired by the system usability scale (SUS) questionnaire, a standard usability measurement tool typically used to measure a user's attitudes and system preferences. Designer's answers to the re-engineered SUS questionnaire were later employed as a baseline to compare actual user responses to their SUS questionnaire answers (see section: System Usability Scale and Participant Interviews).

The interview revealed that *Natlas* was developed using ESRI's Map Objects Internet Map Services to render the maps and ArcIMS to manage the communication between the Web server and five spatial services. The graphical user interface (GUI) has evolved through several development environments, all of which are still present. They include HTML, Javascript, Cold Fusion and Active Server pages. There are also tables for zip codes and geographic names which are managed by an Oracle database on a Sun server (Unix). The system does not require a special plug-in, and it runs in all standard web browsers.

Carto.net designers, on the other hand, developed the interface based on Scalable Vector Graphics (SVG) to render the graphics, and ECMAScript (ECMA is for *European Computer Manufacturers Association*) to handle map interaction. Only Internet-Explorer users need a specific Scalable Vector Graphics (SVG) plug-in (i.e., the Adobe SVG viewer). All other browsers can display the map without additional installations.

The intended audience for both maps is “the average internet user”, that is, non-domain experts, without any specific additional technical expertise. For both maps, the design team included at least one cartographer. *Natlas* required a two-year design and development period, with 12 people involved at different stages over the entire development time. Natlas was designed with frequent usage in mind. The system was thoroughly tested before public launch, including a classical usability study that led designers to “*deliberately remove[d] functionality that was too complex for our average users*”. Currently, it is maintained by one employee and receives more than 150,000 unique visits per month.

Carto.net’s implementation, on the other hand, grew out of a student’s class project. Four people have been involved in the development and design process; two in supervising roles, and two actually implementing the system. While developers made different technical, cartographic, and interface design choices, the map data were taken from Natlas, and thus are identical. The *Carto.net* designer that we interviewed stated that “*the application is not too flexible and customizable*” and that it could be improved in these areas. The map interface was not designed with frequent usage in mind, and no usage statistics are available for this map. It was not subjected to any pilot testing or human subject testing, before it was launched. However, the designers’ intention was that it “*should be easy to use for everyone.*”

Experimental Design Our experiment was designed to provide a balance between experimental control and ecological validity. In a between-subject design, we monitored user responses (N = 30) to three different typical map use tasks (independent variables). We used tasks that have different levels of complexities: two are close-ended questions and require an inference related to an attribute or a location (Qestions 1 and 2) and one is an open ended question that requires the participant to compare two spatial distributions (Question 3). Test questions were as follows:

> **1)** What is the number of assaults in Washington County (Maine) in the year 2000?
> **2)** Which county in the state of Oregon has the highest murder rate in the year 2000?
> **3)** Looking at the map of the U.S.A, overall, do you see a relationship (if any) between poverty rates and burglaries in the year 2000?

The map use tasks (within-subject) were shown in a systematic rotation to counter-balance for a potential learning effect. The dependent variables include the traditional usability measures such as absolute and relative response time (efficiency measures) and accuracy of response (quality measure). These performance measures are complemented by self-reports collected on a standardized system usability scale (SUS). Additionally, eye movement recordings including gaze plots and fixations patterns in selected areas of interest in the interface (i.e., as a result of a cognitive walkthrough session before the experiment and where participants reported having trouble after the experiment) allow us to link traditional usability (success) measures with users’ interface interaction processes. Finally, participants also provided qualitative interface preference feedback. Even though the experimental design included two professional groups (geography-educated and others), in the scope of this paper we focus on identifying the usability problems based on map interface designs rather than on differences between the two groups.

Participants Thirty participants (11 females, 19 males) participated in this study. The average age is 28 years. Fifteen participants have college-level training in geography (5 point Likert scale participants reported an average of 4, where 1 meant no training and 5 meant proficient) and fifteen participants have a non-geographic educational or professional background. All participants are non-native, but fluent English speakers (the interfaces of both maps are in English). They have reported a high level of experience with the relevant operating system (4.7 of 5), the Internet (4.8 of 5) and the relevant browser (3.30 of 5). It is also important to note that the participants were, on average, fairly experienced in using graphical (4.2 of 5) and spatial (3 of 5) data. They were offered no compensation for their participation.

Materials Two interactive online map interfaces (see Figure 2) were selected as stimuli for the study (*Natlas* and *Carto.net* as introduced earlier). Both maps display a dataset that represents thematic information related to "Crime and Poverty in the USA, 2000" and provide several interactive features to display and query this dataset. The two maps have different interface designs: *Natlas* devides the screen into three main parts, where the top part of the screen has the title of the map and several interactive buttons and links, the map is displayed in 43% of the screen, and 57% is reserved for interacting via buttons, links and pull-down menus which are distributed in three distinct tabs (Map Layers, Map Key and Find). When a query is committed, the answer is returned on a pop-up window. *Carto.net* uses a larger area of the screen for displaying the map (72%) and 4 small windows overlaid on the map's area. These windows can be minimized and/or moved. The queries are made via radio buttons, and the query answer is returned in the top bar of the legend window as the mouse moves to the relevant area. It is noteworthy that while *Carto.net* provides data and queries only for Crime and Poverty in the USA in 2000, Natlas has a very large selection of other themes as well as this dataset. To make sure that what was immediately visible on one interface was also immediately visible on the other, in other words to ensure that the two maps are informationally equivalent, appropriate tab and pull-down menus were left open on both *Natlas* and *Carto.net* before the participants viewed the interfaces.

Setup The experiment was performed on a Windows workstation, running Tobii Studio software for automatic stimuli display and eye movement recordings. The SUS survey was delivered digitally via Morae usability software. Interface stimuli were displayed on a 24 inch flat screen at a 1600x1200 screen resolution. Eye movements were recorded with a Tobii X120 eye tracker, at a 60Hz sampling resolution.

Procedure After welcoming the participants, we requested them to sign a consent form that provided general information about the experiment. This was followed by an anonymous participant background questionnaire. Participants were then trained to locate the States of Oregon and Maine on a digital map of the coterminous U.S., where all other state names were removed, and the locations of these two states were highlighted. The purpose of the training was to make sure that participants knew were these states were located, as the goal of the experiment was not related to finding the two States on the map. In making sure that they know where the states are located, we can be confident that their response time is related to solving the planned map use tasks instead of trying to locate the states. Before recording began, participants were instructed to assume a comfortable position and not move too much to maximize the eye movement recording's accuracy. Then a calibration with eye tracker followed. At this stage, participants were ready to solve the tasks with the map interfaces and recording began. The experiment leader provided verbal instructions for carrying out the tasks during the experiment and participants also responded verbally. In order to limit the duration of the experiment, the experiment leader provided participants with help after 5 minutes (and the task was considered incomplete). After completing the map use tasks, participants filled out a closed-ended feedback questionnaire and responded to three additional qualitative preference questions. After completing the questionnaire, participants were debriefed, and thanked for their participation.

Results

All participants completed 3 tasks using the two interfaces (Carto.net min: 180.5 s, max: 1305.0 s, Natlas min: 337.4 s, max: 1120.9 s, excluding technical problems and task delivery times). One outlier was detected and removed from the statistical analysis. Removing an outlier is a common practice when it is multiple standard deviations away from the others (e.g. Hegarty and Waller, 2004).

Mean response times and accuracies are shown in Figure 3. To confirm their statistical significance, response time and accuracy scores were subjected to a one-way analysis of variance with map interface types as between subject independent variables. Overall, confirming our hypothesis, participants are significantly more efficient (faster) using *Carto.net*'s interface, F=7.359, p=.011<.05, but significantly more effective (accurate) using *Natlas*, F=5.095 and p=.032<.05.

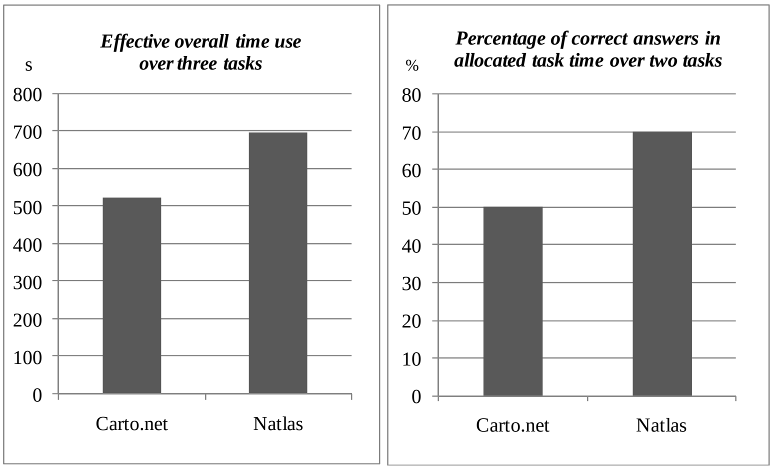
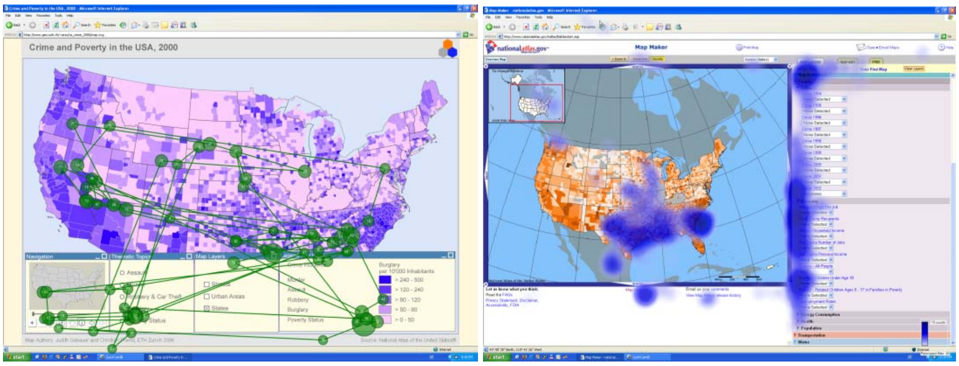


Figure 3. Overall mean response times and accuracies for two map types.

Metrics regarding the response time and completion rate (accuracy) reveal an interesting problem for evaluating the interfaces: one of the designs allow users to perform faster while the other gives more accurate results. This information tells us both designs have elements that make the viewer perform better or worse, but which elements are these? This is where eye movement analysis offers help by allowing us to study micro-level behaviors linked to people's visual attention and internal cognitive processes.



(a) (b)

Figure 4. Screenshot of evaluated interfaces including overlaid eye movement recordings. First image (a) is an example of a *gaze plot* and second image (b) is an example of a *density map*.

The left image in Figure 4 shows a user's *gaze plot*, that is, eye fixations (graduated circles) and saccades (connecting lines) after solving a map use task with the Carto.net interface. On the right of Figure 4, a user's eye movement behavior is depicted with an interpolated fixation density surface (to clear possible confusion, this is referred to as *heat map* in Tobii Studio as well as some of the eye tracking literature) overlaid onto the Natlas map interface. While gaze plots show a discrete object view of individual fixation locations, including their sequence and fixation durations, the density surface maps provide a continuous field perspective of the space-time data, that is, a snapshot view of fixation concentrations during one trial.

For this study, the fixation filter was set to a radius 50 of pixels, minimum fixation duration to 100 ms. Overall fixation counts for each map are Natlas (sum): 19554, Carto.net (sum):

17902 indicate that users had a less efficient search strategy with Natlas in comparison with Carto.net. Overall mean fixation durations for Natlas (M= 7715.1 s, SD: 139.8 s) and Carto.net (M= 6642.1 s, SD: 111.3 s) also indicates that in overall, Natlas has a more complex interface. To analyze the sources of issues regarding search efficiency and difficulties we conduct a deeper level study of certain areas of interest.

Areas of Interest (AOIs) analysis and identified usability issues: The areas of interest were defined before data collection began, based on which interface elements would have to be used to successfully solve the tasks. For this purpose the authors performed a cognitive walkthrough session before running any sessions, to determine which interface elements participants would likely use, and in which sequence. The identified interface elements and sequence were also cross-validated in a pilot experiment with a small set of test participants. Participant interviews later also confirmed the anticipated problem areas. Taking the above into consideration, along with the *whole screen* and *map area* on both maps, the identified areas of interest with potential usability issues were determined as *Identify* and *Redraw Map* buttons on Natlas and the *mouse roll-over behavior* that reflects on the *legend-bar* in Carto.net.

For Natlas, once the Identify and Redraw map buttons were discovered (time to first fixation: Identify M=116.2 s, SD=102.1; Redraw Map M=50.6 s, SD=30.6) and their functions were understood (time to first mouse click: Identify M=186 s, Redraw Map M=54 s), tasks were successfully completed. The difference between the mouse clicks and the first fixations tell us that the labeling of the Redraw Map was more quickly understood by the users in comparison with the Identify. Considering the average task completion time for the first task was 190 seconds for Natlas, participants spent 61% of their time to locate Identify and 27% for redraw map. The fact that there are 423 fixations before Identify and 302 fixations before Redraw Map buttons were located tells us participants were searching them in other parts of the screen. Looking at the scan paths, we observe that majority of the people (75%) spent time on the Menu area (22% of the screen with tabs, located on the right side of the screen and very rich in information) looking for a tool that would help them. Redraw map button has a salient behavior (it flashes and pops a small window telling the user to press the button), even then 30% of the users have not used this button but has sought alternative ways to fulfill its function. Both of these buttons are probably too small (Identify 0.08% of the screen, Redraw Map 0.13% of the screen). It is also noteworthy that 100% of the participants who needed assistance (5 of 15 participants needed assistance to continue) within the five-minute task duration limit had trouble with the Identify button, confirming that the identified usability problem regarding this feature.

The reason why Carto.net is faster than Natlas might be partly because of the size and locations of these two critical buttons but we speculate that it has also to do with the map size: the map is 43% of the screen for Natlas where it is 72% of the screen for Carto.net. This leads to lesser use of zoom buttons, and the fact that users did not need to use zoom in and out buttons as much as they did on Natlas supports that map size is indeed a factor in faster results: the difference between fixation counts on Zoom-in and Zoom-out buttons between two maps are 0.7 and 1.6 respectively (Natlas fixated more times). The difference between the fixation lengths (Zoom-in difference M=0.3s, SD=0.4; Zoom-out difference M=0.7, SD=1.0) also support this claim as this indicates either the complexity of a feature or its importance. In this case, Zoom-in and Zoom-out buttons were both common designs and there was no reason for the participant to have difficulty to comprehend.

While all indicators regarding efficiency (speed) point Carto.net to be favorable, one usability issue regarding this interface manifests in the accuracy scores (task completion success). Carto.net presents the queried data as people roll their mouse over to the relevant geographic area on the bar above the legend. This proves counter intuitive for most users: all participants (80%) have clicked on the map at least once and expected to see a result before they found out about this function, and some (40%) used the right mouse click to explore what other options may be ‘hidden’. Also a majority (80%) of the participants who needed assistance (5 of 15 participants needed assistance, 4 of 5 needed help with the use of the legend) at the end of the five minute limit was with this feature. The legend is designed to present information in two columns which appears to mislead the participants: overall they spent more time looking at the first half of the legend (48% of the participants, observation length M=24.5 s, SD=23.5, fixation length M=20.7 s SD=19.9) than the second half where the information is more relevant (observation length M=15.7 s, SD=19.6, fixation length M=13.3 SD=17.0). On the contrary, accuracy is high with Natlas interface because the responses are isolated within pop-up windows, hence difficult to miss. Isolating the query results therefore reveals itself as a favorable design choice.

An additional observation regarding Carto.net’s design is that it is possible to move the information windows, yet only 4 users of out of 15 (26.6%) have discovered that these windows can be moved. While this is not task relevant, it is interesting to note. It is a useful feature, but maybe participants need to be made aware of it in an explicit manner. For Natlas an additional usability issue may be about the help button, as only 7% of the participants ever fixated on it and only one participant used it.

System Usability Scale (SUS) and Participant Interviews There are a number of standardized usability surveys to measure participant’s attitudes or preferences, such as the Questionnaire for User Interface Satisfaction (QUIS), the Computer System Usability Scale (CSQU), the Software Usability Measurement Inventory (SUMI) and the System Usability Scale (SUS). In a systematic comparison carried out by Tullis and Stetson (2004), the SUS, one of the simplest questionnaires studied, provided among the most reliable results across sample sizes (references for these tests can be found in Tullis and Stetson (2004)). For this reason, at the end of the eye movement recording session, participants filled in an interactive System Usability Scale (SUS) questionnaire, including 10 Likert-style preference questions (Brooke, 1996). As mentioned earlier, designer interviews before the experiment included questions that were rephrased for them instead of a user. In doing so, participants’ responses can be compared to system designer intentions.

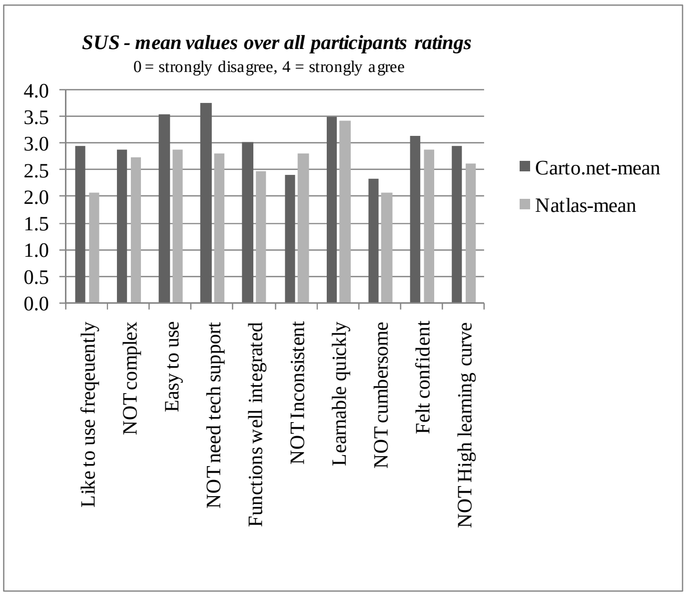


Figure 5. SUS results based on a 0-4 Likert scale. Participants seem to prefer Carto.net. However, only in two of the questions (Questions 1 and 4) preferences were *significantly* different.

SUS results reveal that Carto.net users have a more positive attitude towards the interface than the Natlas users (Figure 5). This difference becomes statistically significant in two accounts: more people (question 1, F=4.95, p=0.034<.05) declared that they would use the Carto.net interface frequently and more people (question 4: F=7.22, p=.012<.05) declared that they felt they would need tech support when using Natlas interface. These results further confirm our hypothesis that the two maps are informationally equivalent however not computationally equivalent. The results of this questionnaire were cross-checked against the designer responses. On the two accounts that were significant, for example, Natlas expects very high frequency of visitors (4 on the Likert scale of 1-4), but the participants' average is only 2.1. The next question that yielded a significant difference between the two interfaces was regarding the expected technical skills of a user. The designers report that they hope the users with minimum or no technical skills can comfortably use the interface (Natlas=3, Carto.net=4). On the other hand, the participants' average reveals that a majority of them (Natlas=2.8 Carto.net=3.7) feel that they would need technical support to be able to successfully complete the tasks.

Qualitative questions and observations To complement the SUS, we added three additional qualitative questions:

1) Would you use this interactive digital map instead of a traditional Atlas? Please explain why.
2) Would you recommend this interactive digital map to a friend? Please explain why.
3) Which of the following features do you believe need major improvement?
   Server and process speed / Map coloring scheme / Size of the buttons / Placement of the buttons / Wording of the buttons / All of the above / None of the above / Other (please tell us which)

Participants were asked to check one or more of the offered categories, and/or add their own categories. Summarizing participants' open-ended comments, it seems that even though they have complaints, they find both interactive maps superior to paper maps, 36% of them explicitly list the maps' interactivity and responsiveness as main advantages. One of the Natlas users mentions that s/he would skip the service if it weren't for the experimental situation, and look for the answer to the question elsewhere. Indeed, it was very informative to see how several participants actually tried to do this when they got frustrated with the map. We observed 5 Natlas viewers out of 15 (33 %) and 2 Carto.net viewers out of 15 (13%), after getting frustrated with the map interfaces, tried to leave the map web page and use other web search tools to find the answer to the test questions (e.g. Google, US census bureau link, SVG source code).

One participant, responding to the Natlas interface, also offered a comment about interfaces in general: *"Although it should not necessarily be the case always, I have a feeling that a poorly designed GUI (graphical user interface) is also an indication of a poorly implemented system"*.The following comment from another participant responding to the Carto.net interface user suggests: *"They [legend information] did NOT always mean what I assumed. I think people don't like reading too much, and it would be best if the legends would match what most people assume by default."* This comment is particularly noteworthy as only one (7%) of our observed map users ever perused the Help button when they got stuck! A large number of (40%) of Carto.net viewers complained about the legends (legend design and legend description). This even included one participant claiming that the legend was wrong: *"The legend and value reporting is wrongly designed."*

Conclusions

We presented a methodology that integrates eye movement analysis and traditional usability performance and satisfaction metrics for interactive map interfaces. Identifying usability problems in interactive map interfaces can be a tedious, iterative process.

With a case study, we demonstrated that the information gathered from eye movement analysis can enhance usability studies both quantitatively and qualitatively. Overall results of SEE metrics and eye movement analysis for the case study confirm our hypothesis and indicate that Carto.net has a faster interface. The study also reveals that the users, at least on two accounts, *prefer* Carto.net, more participants declaring that they would like to use it frequently. However, Natlas is more accurate which indicates both designs have usability problems, and the interfaces could be improved. Usability researchers are well aware of the fact that professional deformation from years of training leaves the designer with little clue as to what is difficult for a non-trained person and what is easy-to-use.

The amount of data generated by modern eye tracking devices is very high albeit the rapid progress in recent hardware and software technologies, simpler and more elegant solutions for processing these large volumes of data are still desirable. However, eye movement analysis provides us with information on visual attention which is commonly accepted as a proxy for mental attention (Webb and Renshaw, 2008) and it is valuable for understanding how users process the interfaces. In the demonstrated case study, eye movement data revealed micro-level usability issues regarding Identify and Redraw Map buttons on Natlas as well as the mouse roll-over behavior on Carto.net, that the usability test alone would not be sufficient to identify.

Future Work What we report in this paper is only part of the analysis, several items that are left for a follow-up publication include (but not limited to): individual task analysis, scanpath analysis and professional bias. A future direction may also be a comparative analysis of both interfaces after modifying them with the lessons learned from this study.

Acknowledgements We would like to offer our sincere thanks to our participants and the designers of Carto.net and Natlas interfaces for their time, patience and kind cooperation.

REFERENCES

Bojko, A. (2006) Using Eye Tracking to Compare Web Page Designs: A Case Study. ***Journal of Usability Studies***, 3, 1, pp. 112-120.

Brodersen, L. Andersen, H.H.K. and Weber, S. (2001) Applying Eye-Movement Tracking for the Study of Map Perception and Map Design. ***National Survey and Cadastre, DK***, 4, 9.

Brooke, J. (1996) SUS: A "Quick and Dirty" Usability Scale. ***Usability Evaluation in Industry***. 1ed. London: Taylor and Francis.

Byrne, M.D. Anderson, J.R. Douglas, S. and Matessa, M. (1999) Eye Tracking the Visual Search of Clickdown Menus. *Proceedings of CHI'99,* NY: ACM Press. pp. 402-409.

Dix, A. Finlay, J. Abowd, G. and Beale, R. (2004). ***Human-Computer Interaction.*** 3rd ed. Prentice Hall. ISBN 0-13-239864-8.

Duchowski, A.T. (2007) *Eye Tracking Methodology, Theory and Practice.* London: Springer.

Fabrikant**,** S.I. Rebich-Hespanha, S. Andrienko, N. Andrienko, G. and Montello, D.R. (2008) Novel Method to Measure Inference Affordance in Static Small Multiple Displays Representing Dynamic Processes. ***The Cartographic Journal****, 3, 45* (in press).

Fuhrmann, S. Ahonen-Rainio, P. Edsall, R. M. Fabrikant, S.I. Koua, E.L. Tobon, C. Ware, C. and Wilson, S. (2005) Making Useful and Useable Geovisualisation: Design, and Evaluation Issues. In ***Exploring Geovisualisation***, Elsevier Ltd., pp. 553-566.

Goldberg, J.H.. and Kotval, X.P. (1999) Computer Interface Evaluation Using Eye Movements: Methods and Constructs. ***International Journal of Industrial Ergonomics***, 24, pp. 631-645.

Goldberg, J.H. Stimson, M.J. Lewenstein, M. Scott, N. and Wichansky, A.M. (2002) Eye Tracking in Web Search Tasks: Design Implications. In ***Proceedings of theEye Tracking Research and Applications Symposium 2002,*** NY: ACM Press. pp. 51-58.

Good, M. Spine, T.M. Whiteside, J. and George, P. (1986) User-derived Impact Analysis as a Tool for Usability Engineering. In ***Proceedings of Human Factors in Computer Systems**: CHI '86* NY: ACM. pp., 241-46.
Haklay, M. and Zafiri, A. (2008) Usability Engineering for GIS: Learning from a Screenshot. Special Issue on Use and User Issues of the ***Cartographic Journal**, 2, 45*, pp. 87-97(11).
Hegarty, M. and Waller, D. (2004) A Dissociation Between Mental Rotation and Perspective Taking Spatial Abilities. ***Intelligence***. 32, 2, pp. 175-191.
Henderson, J. M. and Ferreira, F. (2004) Scene Perception for Psycholinguistics. In ***Introduction to the interface of vision, language, and action: Eye Movements and the Visual World***. NY: Psychology Press. pp. 1-58.
Howard, D.L. and MacEachren, A.M. (2006) Interface Design for Geographic Visualization: Tools for Representing Reliability. ***Cartography and Geographic Information Science***, 2, 23, pp. 59-77.
Irwin, E. (2004) Fixation Location and Fixation Duration as Indices of Cognitive Processing. ***The Interface of Language, Vision and Action: Eye Movements and the Visual World***, NY: Psychology Press, pp. 105-134.
Jacob, R. and Karn, K. (2003) Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research.* Elsevier, Amsterdam, pp. 573-605.
Koua, E.L. MacEachren A.M. and Kraak, M.-J. (2006) Evaluating the Usability of Visualisation Methods in an Exploratory Geovisualization Environment, ***International Journal of Geographical Information Science***, 20, pp. 425-448.
Koua, E.L. and Kraak, M.-J. (2004) A Usability Framework for the Design and Evaluation of an Exploratory Geovisualisation Environment, ***Proceedings of the 8th International Conference on Information Visualisation***, IV'04, IEEE Computer Society Press.
MacEachren, A.M. and Kraak, M-J. (2001) Research Challenges in Geovisualization. ***Cartography and Geographic Information Science***. Special Issue, 1, 28, pp. 3-12.
Nielsen, J. (1993) ***Usability Engineering.***, Boston, MA.
Nivala, A.-M. Brewster, S.A. and Sarjakoski, L.T. (2008) Usability Evaluation of Web Mapping Sites. Special Issue on Use and User Issues of the ***Cartographic Journal**, 2, 45,* pp. 130-140.
Larkin, J.H. and Simon, H.A. (1987) Why a Diagram is (Sometimes) Worth Ten Thousand Words. ***Cognitive Science***, 1, 11, pp. 65-100.
Pretorius, M.C. Calitz, A.P. and van Greunen, D. (2005) The Added Value of Eye Tracking in the Usability Evaluation of a Network Management Tool. In ***Proceedings of the 2005 Annual Research Conference of the South African institute of Computer Scientists and information Technologists on IT Research in Developing Countries,*** ACM International Conference Proceeding Series, 150., pp. 1-10.
Steinke, T.R. (1987). Eye Movement Studies in Cartography and Related Fields. *Studies in Cartography, Monograph 37,* ***Cartographica***, *24* (2), pp. 40-73.
Tullis, T.S. and Stetson, J.N. (2004) A comparison of Questionnaires for Assessing Website Usability. In ***Proceedings of Usability Professionals Association Conference, UPA 2004***.
Webb, N. and Renshaw, T. (2008) Eyetracking in HCI. *Research Methods for Human Computer Interaction,* Cambridge University Press, pp. 35-69.
You, M. Chen, C.-W. Liu, H. and Lin, H. (2007) A Usability Evaluation of Web Map Zoom and Pan Functions. ***International Journal of Design***, 1, 1, pp. 15-25.

Tested web maps

Carto.net (2008) An Online Interactive Map Interface showing the Crime and Poverty Rates in the U.S.A. in 2000. http://www.carto.net/papers/svg/us_crime_2000/ Last visited 16/06/2008.
Natlas (2008). Map maker, an Online Interactive Map Interface of National Atlas of the United States. http://nationalatlas.gov/natlas/Natlasstart.asp Last visited 16/06/2008.