Computers & Graphics 72 (2018) 26–38

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

Exploring visual attention and saliency modeling for task-based visual analysis

Patrik Polatsek[a], Manuela Waldner[b], Ivan Viola[b], Peter Kapec[a], Wanda Benesova[a,*]

[a] Faculty of Informatics and Information Technologies, Slovak University Of Technology in Bratislava, Slovak Republic
[b] Faculty of Informatics, TU Wien, Austria

ARTICLE INFO

Article history:
Received 3 October 2017
Revised 26 January 2018
Accepted 26 January 2018
Available online 7 February 2018

Keywords:
Information visualization
Eye-tracking experiment
Saliency
Visual attention
Low-level analytical tasks

ABSTRACT

Memory, visual attention and perception play a critical role in the design of visualizations. The way users observe a visualization is affected by salient stimuli in a scene as well as by domain knowledge, interest, and the task. While recent saliency models manage to predict the users' visual attention in visualizations during exploratory analysis, there is little evidence how much influence bottom-up saliency has on task-based visual analysis. Therefore, we performed an eye-tracking study with 47 users to determine the users' path of attention when solving three low-level analytical tasks using 30 different charts from the MASSVIS database [1]. We also compared our task-based eye tracking data to the data from the original memorability experiment by Borkin et al. [2]. We found that solving a task leads to more consistent viewing patterns compared to exploratory visual analysis. However, bottom-up saliency of a visualization has negligible influence on users' fixations and task efficiency when performing a low-level analytical task. Also, the efficiency of visual search for an extreme target data point is barely influenced by the target's bottom-up saliency. Therefore, we conclude that bottom-up saliency models tailored towards information visualization are not suitable for predicting visual attention when performing task-based visual analysis. We discuss potential reasons and suggest extensions to visual attention models to better account for task-based visual analysis.

1. Introduction

Visualization designers use a large variety of visual channels to effectively encode data. Due to the limited computational capacity of the brain, parsing and interpreting these visual channels cannot be performed rapidly. Instead, visual attention is serially directed to different regions in the visualization, and the information is gradually decoded. Visual attention is a set of cognitive processes that selects relevant information and filters out irrelevant information from the environment [3]. Attention is driven by both bottom-up and top-down factors. The aim of exogenous and very rapid *bottom-up* attention is to warn of impending danger. It is guided by low-level salient visual features which stand out from their neighborhood (the so-called "pop-out effect"), such as intensity or color contrasts, texture and motion. Visual channels used in information visualizations are perceived either with specialized receptors of the human visual system (e.g. red-green opponency [4], orientation or spatial frequency [5]) or by multiple receptors in the case of complex channels, such as shape. In contrast to this stimulus-driven attention, endogenous and much slower *top-down* attention is biased by cognitive factors. It involves prior knowledge, expectations, tasks and goals that enhance bottom-up attention. Real scene *perception*, referring to the organization and interpretation of sensory information, lies in the interaction between bottom-up and top-down processing of attention [6].

When users interpret visualizations, top-down factors of attention are incorporated in scene perception. Visual search is an important component of the process of interpreting visualizations. It is the process of finding a specific target object in a scene among non-targets. Visual attention thereby guides the user's gaze and the visual search, respectively. Understanding visual attention is therefore essential for selecting appropriate visual channels and designing effective visualizations.

Computational saliency models have been developed to predict users' visual attention (see Section 2). These models are quite accurate for simple stimuli and natural images [7–9]. While saliency models have also been used as a quality measure in the information visualization domain [10,11], it has been shown that these models' accuracy for predicting visual attention in visualizations

* Corresponding author.
E-mail addresses: patrik.polatsek@stuba.sk (P. Polatsek), waldner@cg.tuwien.ac.at (M. Waldner), viola@cg.tuwien.ac.at (I. Viola), peter.kapec@stuba.sk (P. Kapec), vanda_benesova@stuba.sk (W. Benesova).

https://doi.org/10.1016/j.cag.2018.01.010
0097-8493/

is significantly poorer than for natural images [12]. Matzen et al. [13] therefore recently introduced a saliency model tailored towards information visualization, and showed that fixations during exploratory visual analysis could be predicted more accurately.

While it has been shown that bottom-up factors captured by this new model have a strong influence on users' visual attention during exploratory visual analysis, it is still unknown how strong top-down guidance influences attention during task-based visual analysis. This work investigates human gaze behavior and saliency prediction when performing typical low-level analytical tasks with visualizations. To this end, we performed an eye-tracking study with charts from the MASSVIS database [1]. During this study, users solved three different low-level analytical tasks. We compared the data of this study with eye tracking data of the memorability experiment [2] with conditions closer to natural image viewing. We could show that fixations are more coherent between users, but correlate less with the visualization-tailored saliency map when performing a low-level analytical task. We discuss some potential extensions of saliency models to incorporate these added top-down factors.

2. Visual attention models

In recent decades, various attention models have been proposed that differ in how they predict human visual attention. As pioneers, Itti et al. [4] defined a computational bottom-up saliency model using local center-surround differences of intensity, color and orientation features at multiple spatial scales. This approach of feature extraction has been adopted in many attention models. Harel et al. [14] also followed this approach. Their model computes saliency using graph-based dissimilarity measures. Hou et al. [15] introduced a model analyzing the frequency domain instead of the spatial domain to predict saliency. The bottom-up model presented by Zhang and Sclaroff [16] is based on the principle of figure-background segregation. The model identifies figures in a set of Boolean maps generated by thresholding feature maps. A work presented by Bruce and Tsotsos [17] defines saliency as the self-information of visual features of the image. Zhang et al. [18] proposed a Bayesian framework that incorporates top-down information dependent on the target's features with bottom-up saliency that is represented as the self-information derived from the statistics of natural images. Goferman et al. [19] proposed saliency detection based on patches with unique low-level features, visual organization rules according to which regions close to salient pixels are also salient and high-level factors, such as human faces. Vig et al. [20] and Cornia et al. [21] introduced saliency models that employ neural networks to predict fixations.

Visual saliency predicted by computational models can be applied in many areas of computer science including image processing [22–24], computer graphics [25], robotics [26,27], surveillance systems [28,29] and human-computer interaction [30,31]. Saliency models have been widely evaluated against different datasets that usually contain natural scenes and fixations from free viewing [7,32–35]. The benchmarks [36–38] show for some image databases a small difference between the state-of-the-art models and human inter-observer (IO) that outputs a fixation map integrated from other subjects than the one under test. The map serves as an upper bound for prediction accuracy. Generally, the prediction accuracy of the models is higher for simple images with few salient objects. However, predicting human fixations in complex images with multiple objects is a challenging task [39,40].

The models are commonly used to predict where the observer's attention is directed in natural images. However, they have also been used in visualization research to predict attention and to derive quality measures, respectively. For instance, Jänicke and Chen [11] describe a saliency-based quality metric for visualizations. It compares a saliency map using the cognitive model by Itti et al. [4] and importance of visualized data, generated automatically from data or manually by visualization designers. The metric is then computed as the difference of these maps. Attention in dynamic geovisualizations was studied by Gagarlandini and Fabrikant [10]. Saliency of dynamic visualizations was predicted by the spatiotemporal model by Itti et al. [41]. The highest saliency value was predicted in regions of the change. Saliency was also applied in volume visualizations to guide attention to selected regions [42]. Saliency was first determined for each voxel, and was then enhanced by center-surround operations between voxels inspired by the standard cognitive saliency model [4].

These works are based on the assumption that saliency models, originally developed for natural image viewing, are equally accurate for predicting the attention when interpreting visualizations. However, there are some notable differences between natural images and classic charts used in information visualization. Graphical marks, such as dots or lines, are usually abstract and simple compared to complex objects in natural images. Also, the background is mostly uniform and the visualizations contain a lot of textual information, such as labels and legends. Graphical marks and visual channels are chosen by a visualization designer according to design guidelines and visualization domain knowledge with the goal to expressively and effectively represent the underlying data. Thereby, visualization designers choose their visual channels to maximize the amount of information displayed [43]. Matzen et al. [13] also note that most saliency models tend to omit fine-grained visual features, like thin lines, which are highly relevant for information visualization.

Therefore, special variations of saliency models have been developed for information visualization. Lee et al. [44], for instance, introduced a saliency model for categorical map visualizations. They define point saliency as color difference of each point against its neighborhood. The class visibility quantifies the point saliency values that correspond to a given category. Most relevant for our work, Matzen et al. [13] proposed a novel saliency model that combines the model of Itti et al. with text saliency to predict saliency in data visualization with higher precision. The presented work evaluated saliency models on the MASSVIS database [1]. The results highlight the importance of text in visual attention since the model that relies only on text saliency outperforms classic saliency models in most evaluation metrics.

In our work, we will compute all above mentioned saliency models for the visualizations used in our experiment and compute the correlations to the obtained fixations from our eye tracking data.

3. Related work

To explore the applicability of saliency models beyond natural images, Borji et al. [7] compared the performance of four saliency models to eye tracking data obtained during free viewing of 20 different image categories, including abstract patterns and line drawings. In their study, saliency models predicted fixations surprisingly well for sketches. Matzen et al. [45] compared fixation maps of novices and professional analysts when analyzing synthetic aperture radar imagery to Itti et al.'s [4] saliency model. They showed that fixation maps of novices were more correlated with the saliency maps, compared to those of the professionals. They concluded that novices are much more likely to be directed by bottom-up salient features than experienced users.

Haass et al. [12] compared the performance of three different saliency models between the cat2000 dataset [7] and the MASSVIS dataset from Borkin et al.'s memorability experiment [2] using eight different comparison metrics. They found that saliency models performed worse for information visualizations than for the natural images. One possible explanation by the authors is that text labels in visualizations attract the users' attention to a higher extent than indicated by the saliency models. Indeed, Matzen et al. [46] and Bylinskii et al. [31] showed that most fixations in visualizations can be accounted to regions containing text. In the DVS model [13], Matzen et al. linearly combined a variation of Itti et al.'s model [4] with text saliency, and could thereby increase the performance of the saliency model significantly. Our work extends this knowledge by comparing the effect of bottom-up visual saliency on visual attention between exploratory and task-based visual analysis. In combination with analysis of fixation patterns, we can characterize the top-down guidance imposed by low-level analytical tasks.

Eye tracking is becoming a popular alternative for evaluating visualizations, compared to classic completion time and accuracy evaluations [47]. Mostly, it is employed to understand how users read a particular encoding, such as parallel coordinates [48], or to compare visual exploration of different encodings, such as different tree layouts [49,50], graph layouts [51], linear and radial charts [52], or user strategies for sorting in tabular visualizations [53]. Task-dependent areas of interest can be used to assess in which order and frequency users fixate crucial chart elements when decoding the visualization [52,54,55]. In contrast to prior task-based eye tracking experiments, we analyze the influence of visual saliency on users' attention when solving different tasks.

In the memorability experiment by Borkin et al. [2], eye tracking data was gathered in two separate treatments using various charts of the MASSVIS dataset [1]: In a 10-second encoding phase, users examined a visualization without a pre-defined goal. In the subsequent 2-second recognition phase, they were asked to indicate whether they had seen the particular visualization before. Finally, in the 20-minute recall phase without eye tracking, they were presented with small blurred versions of recognized visualizations, and were asked to write down everything they remember being shown. The outcomes of this experiment give indications which visualization elements attract the users' attention, and which elements make a visualization memorable.

Based on the data obtained during the memorability experiment [2], Bylinskii et al. [56] explored different eye fixation metrics to assess the MASSVIS dataset. The presented metrics are based on overall fixations, fixations within areas of interest, gaze coverage and human IO. These metrics can reveal interesting differences between visualization types, such as the observation that fewer visualization elements lead to more consistent viewing patterns between participants. To the best of our knowledge, our work is a first attempt to assess the influence of bottom-up saliency on visual attention when performing task-based analysis of a visualization given different low-level tasks, compared to task-free exploration in the memorability experiment [2].

4. Task-based visual analysis experiment

Data analysis using visualizations is commonly divided into three categories [57]:

1. *exploratory analysis*: to formulate a new hypothesis about the data,
2. *confirmatory analysis*: to confirm or reject given hypotheses about the data,
3. *presentation*: to communicate facts efficiently and effectively.

We can assume that the impact of top-down factors on the user's visual attention varies for these activities. Exploratory analysis is less driven by top-down factors than confirmatory analysis to answer a specific question. For presenting known facts, often highlighting is used to effectively draw the attention to specific regions, but the task-imposed guidance is also low.

In our experiment, we compare confirmatory (or task-based) visual analysis and exploratory visual analysis. The purpose is to test if task-based visual analysis is indeed strongly driven by top-down factors, so that bottom-up saliency has negligible influence on the user's attention during the task.

4.1. Hypotheses

Our experiment is based on two major hypotheses, which we further sub-divided:

H1: *Overall, top-down factors, such as a particular task, play such an important role in guiding visual attention that bottom-up factors have a negligible effect on the recorded fixation patterns.* We reason that fixations of users will be strongly guided by the task during task-based visual analysis. To solve a task, users have to look at pre-defined areas of interest within the visualization, which will require most of their attention. On the other hand, we expect that during exploratory analysis, users will be more strongly guided by bottom-up factors. We therefore expect the following results:

- H1.1: Fixations between users solving the same low-level analytical task will be more coherent than when exploring a visualization without a specific task.
- H1.2: When solving a low-level analytical task, users fixate on a sequence of specific chart areas in a task-dependent order.
- H1.3: The similarity between the recorded fixation maps and bottom-up saliency maps will be higher when users explore a visualization without a specific task than when performing a low-level analytical task.

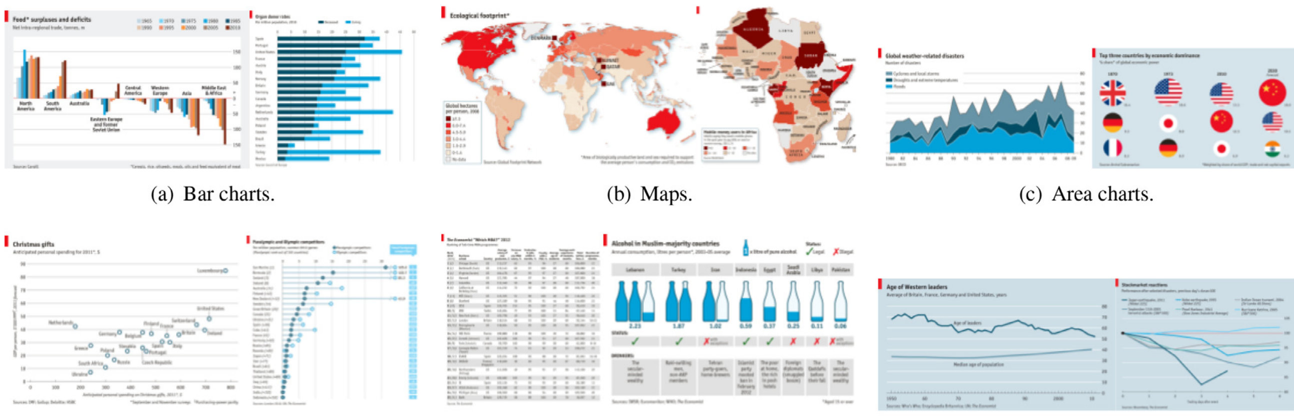
H2: *Bottom-up factors have an influence on visual attention when performing a visual search for a target.* While we assume that bottom-up saliency does not have a strong influence on users' fixations (see H1.3), we do believe that it has an influence on visual search efficiency for target areas when solving a low-level analytical task. In particular, extreme values should stand out in their associated visual channel, for instance as the longest bar in a bar chart or the darkest region in a choropleth map. We therefore assume that extreme values should also show up as salient regions in the saliency maps, and that salient target data points are therefore fixated more quickly than non-salient data points. As a consequence, we assume that users can find extreme values more efficiently than retrieving values of specific items or items associated with specific values. Our specific hypotheses are the following:

- H2.1: Efficiency of visual search for a target area depends on the area's visual saliency.
- H2.2: Extreme data points show up as highly salient graphical marks in saliency maps.
- H2.3: Extreme values can be found most efficiently.

4.2. Image data

Since our hypotheses are not targeted towards a specific visualization type, we chose the MASSVIS database [1] as source for our image data. This database contains around 5000 static visualizations obtained from different online sources, such as news media or blogs. The contained visualizations are targeted towards a broad audience and are therefore a popular choice to evaluate how non-experts read visualizations [1,2,12,13,56]. We selected a subset of 30 visualizations from the dataset with the goal to cover a large variety of visualization types, such as bar charts, maps, area charts, tables, point charts and line charts (Fig. 1) from the "news media" and "infographics" categories. Thereby, we only chose visualizations with associated eye tracking data.

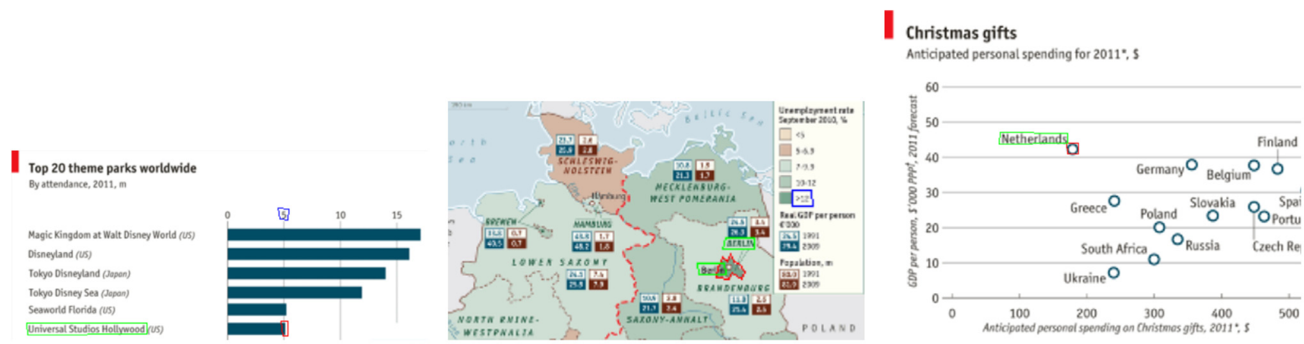
16 charts contain human recognizable objects (HROs) such as pictograms and real objects (e.g., bottles as in Fig. 1)(e). We selected visualizations with a rather low average data-ink ratio (ratio



(a) Bar charts. (b) Maps. (c) Area charts.

(d) Point charts. (e) Tables. (f) Line charts.

Fig. 1. Categories of visualizations used in the experiment.



(a) What is the attendance of Universal Studios Hollywood? (b) Which German states have an unemployment rate of more than 12%? (c) In which country do people anticipate to spend the least money for personal Christmas gifts?

Fig. 2. Target-dependent AOIs of sub-parts of visualizations for the RV-task (a), the F-task (b), and the FE-task (c), respectively. Red, green and blue outlines define the target data points, their item labels and value labels respectively. To complete the RV-task, a target item label has to be searched. Then, the value label of the target data point (i.e., bar) is read. For the F-task, participants search the value label that satisfies the given condition. Then, they search data points (i.e. states) with the color that corresponds to this value and read their names. The scatterplot has data points sorted by the anticipated personal spending on Christmas gifts. Thus, participants only have to find the left-most dot without reading the actual value label. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Original labels of the MASSVIS dataset [1].

Category	Description
Annotation	Visual elements annotating the data
Axis	Axes including tick marks and values
Data	Area where data are plotted
Data (type)	Visual representation of the actual data
Graphical element	Elements not related To data
Legend	Legends or keys
Object	HROs
Text	Textual elements

of data and non-data elements) of 1.5 (measured on a scale from 1=low to 3=high). The dataset is accompanied by manually defined labels summarized in Table 1.

4.3. Tasks

Visualization research has produced various task taxonomies that formalize activities with visualizations. We adopted the taxonomy proposed by Amar et al. [58] that specifies low-level analytic tasks in the field of information visualization. According to Amar et al., these are simple, easily solvable tasks, where users have to analyze the visualized data points. In contrast to high-level tasks, users do not require higher-level domain knowledge. Amar et al.'s collection of simple tasks is one of the most frequently cited task taxonomies in the visualization community [59]. We picked the three simplest tasks out of this taxonomy, which are easily solvable across a wide range of visualization types – in particular the visualization types in the MASSVIS database.

For each visualization, we formulated a question according to the task:

- *retrieve value* of a specific data element (RV-task) (Fig. 2(a)),
- *filter* data elements based on specific criteria (F-task) (Fig. 2(b)),
- *find* an *extremum* attribute value within a dataset (FE-task) (Fig. 2(c)).

For comparing task-based visual analysis to more exploratory analysis, we additionally analyzed the eye tracking data of the memorability experiment [2] (Mem-task). The aim of the memorability experiment was to encode displayed visualizations and later recall as many details as possible. Participants were shown about 100 visualizations, each for 10 s. For the evaluation, we used fixations from the first phase (encoding phase) of the memorability experiment, when participants were shown visualizations for the first time.

Table 2
Task-dependent AOIs.

Category	Description
Value label	Textual value label of a target attribute
Value annotation	Textual elements annotating values of an attribute
Value legend	Legend or keys of attribute values of data points
Data point	Target data point
Item label	Textual identification of a target
Item legend	Legend of item encodings

Table 3
Optimal viewing order of task-dependent AOIs.

Task	Step 1	Step 2	Step 3
RV	Search item label	Map to the item	Read the value label
F	Search value label(s)	Map to the item(s)	Read the item label(s)
FE	Search value label(s) search item(s)	Map to the item(s)	Read the item label(s)

4.4. Areas of interest

To be able to more specifically analyze eye tracking data with respect to the given task, we additionally defined task-dependent areas of interest (AOIs) for each visualization and low-level task (RV, F, FE), respectively, listed in Table 2. They comprise all elements of the visualization that need to be attended to correctly answer the question. It is important to note that not all visualizations contain all AOIs, such as legends. Fig. 2 shows exemplary visualizations with task-dependent AOIs.

Depending on the task, there is an optimal viewing sequence in which these task-dependent AOIs should be examined in order to answer the question. For their eye tracking experiments, Goldberg and Helfman [54] defined AOIs for three sequential steps required to retrieve values in linear or radial graphs: "find dimension", "find associated datapoint", "get datapoint value". We adopted these three steps for the three low-level tasks in our experiment (Table 3). Step 1 thereby is always a visual search for an item label, a value label, or a target data point, respectively. Item or value labels can be axis labels, listed in legends, or directly associated with a data point. In the second step, this label has to be mapped to the actual data point (i.e., the graphical mark). In Step 3, the associated value or item label has to be read.

The F-task and the FE-task share the same goal – namely, to read one or multiple item labels of targets whose attribute values fulfill the given criteria. When solving the FE-task, users may directly search for the data point without reading value labels if data entities are sorted according to the attribute's value in question (e.g., as shown in Fig. 2(c)). Mind that our experiment included 27 F-tasks and four FE-tasks with multiple targets.

The optimal viewing order of the RV-task is reversed. A target label is given, and the goal is to find the value of its required attribute, as shown, for instance, in Fig. 2(a). For the RV-task, users always had to find only a single value.

4.5. Experimental design and procedure

Using a within-subjects design, participants were shown the same subset of 30 visualizations without any repetitions. The order of appearance was counterbalanced with a Latin square across participants. We formulated one RV-task, one F-task and one FE-task for each visualization, but participants solved only one task type per visualization. The order of the types was randomized with the equal distribution of each type and balanced across participants.

Participants had to correctly solve the task as quickly as possible. The procedure of task completion consisted of three steps that were repeated for each visualization:

1. *Task description*: First, participants were shown a question. After they understood and remembered the question, they pressed the spacebar.
2. *Visualization*: Next, they saw a visualization which they should analyze to answer the question. We did not show a central fixation cross before displaying the visualization. In order to keep the same viewing conditions as in the original memorability experiment [2], the task description that would affect participants' scanning sequence, was not displayed in this step of the experiment. As soon as they found the answer, they pressed the spacebar again.
3. *Answer form*: Finally, participants were shown a form where they entered their answer.

The experiment started with three example tasks to familiarize with each task type. The whole experiment took 29 minutes per participant, on average. Prior to this experiment, a pilot test was performed with three participants to ensure that task descriptions are easy to understand and remember.

4.6. Measures and analysis

For each user and visualization, we recorded eye tracking data, the task completion time, and whether the given response was correct. For each visualization, we additionally created a saliency map. From this raw data, we used the following measures in our analysis:

Correctness refers to the ratio of correctly answered questions for a given task per user. An answer was considered as correct when it contained all target labels or their values. Task correctness was checked manually after the experiment. We used the correctness to test if the complexity of the tasks was similar, and only included measures of correctly answered samples for further analysis.

From the recorded eye tracking data, we computed several fixation and AOI fixation measures:

To measure *fixation similarity* within and across tasks, we built a binary fixation map for each participant with ones at exact fixation locations and blurred the maps (Gaussian filter: size = 200, $\sigma = 32$). The *inter-participant fixation similarity* corresponds to the average value of *correlation coefficients (CC)* between each participant pair's fixation map solving the same task for the same visualization. This measure reveals the coherence of the fixations between users solving the same task (H1.1). The *inter-task fixation similarity* is the average of CC between each task pair's fixation map for the same visualization.

For the AOI fixation measures, we set the maximum distance between a fixation and an AOI to 50 px. This corresponds approximately to 1.3° of visual angle. The *first fixation time (FF)* – or time to first fixation – describes how much time passed from stimulus onset until the first fixation was registered within an AOI. The FF is used to compare the fixation sequence of task-dependent AOIs between tasks (H1.2).

To evaluate the prediction ability of saliency models and to measure the impact of saliency on attention (the *fixation-saliency similarity*), we generated saliency maps from 12 saliency algorithms described in Section 2, denoted **Itti** [4] (implementation by Harel [14]), **AIM** [17], **GBVS** [14], **SUN** [18], **CAS** [19], **Sign**[15], **BMS** [16], **eDN** [20], **SAMv** and **SAMr** [21] (feature maps extracted by the convolutional neural model based on VGG-16 [60] and ResNet-50 [61], respectively), **DVS** [13] (with the optimal weight of text saliency for MASSVIS database) and **TextS** [13] (text saliency of the DVS model separately).

We used two evaluation scores – *Area under the Receiver Operating Characteristic Curve* and *Normalized Scanpath Saliency*. The *Receiver Operating Characteristic* (ROC) *curve* represents the trade-off between the true positive rate and the false positive rate. A saliency map is treated as a binary classifier. Saliency pixels at fixations and the same number of random pixels are extracted. Fixations with saliency above a threshold that is gradually increasing and random pixels above the threshold are considered as true positives and false positives, respectively. Then, the ROC is plotted and the area under the curve (AUC) is computed. An AUC value of 1 corresponds to a perfect fit between fixation map and saliency map, while 0.5 corresponds to chance level. The *Normalized Scanpath Saliency* (NSS) first normalizes saliency to have a zero mean and a unit standard deviation. The NSS score is then the average of saliency pixels at fixation locations. For NSS, a value of 0 corresponds to chance level, and the higher the NSS score, the better the fit. We chose these two metrics, because AUC is the most widely used metric, and NSS has been shown to be the fairest comparison metric in a formal evaluation [62]. We also report the score of *human IO* that generates output maps from fixations of all participants except one under the test. The score can be considered as upper bound to the fixation-saliency similarity evaluation scores. Fixation-similarity measures are compared between the three low-level analytical tasks and the Mem-task to verify hypothesis H1.3.

The *AOI saliency* is computed as the average saliency value in an AOI. We computed the correlation between the AOI saliency and its FF to test its visual search efficiency depending on its saliency value (H2.1). Also, we compared the AOI saliencies of target data points between the tasks to test if extreme data points are more salient (H2.2).

Finally, the *task completion time* (TCT) is measured after understanding of a task, from the initial display of a visualization to the press of the spacebar. We use TCT to test if the FE-task can be solved more efficiently (H2.3).

Eye-tracking data of our experiment are publicly available at http://vgg.fiit.stuba.sk/2018-02/taskvis/. Eye-tracking data of the original memorability experiment [2] can be found at http://massvis.mit.edu/.

4.7. Apparatus

We recorded eye-tracking data using Tobii X2-60 eye-trackers at 60 Hz. All stimuli were displayed on 24.1-inch monitors with a resolution of 1920 × 1080 pixels at a viewing distance of approximately 60 cm. The gaze data were recorded with Tobii Studio and processed by Tobii I-VT fixation filter. Users' heads were not fixed, but they were instructed to avoid unnecessary head movements. The experiment was conducted in normal indoor daylight lighting conditions.

4.8. Participants

We recorded eye-tracking data, response times, and the textual responses of 47 students participating in a data visualization course and a computer vision course. Students were aged 20 to 25 years; 44 were male, three female. Participants who normally wore glasses or contact lenses for distant viewing were asked to wear them during the experiment. None of the participants had any color vision problems. All participants gave their informed consent to the study and received an explanation of the experiment. The study was performed at the end of the course and participation was compulsory to gain all credits for the course. However, once having started the study, students were free to stop the experiment at any time without having their data recorded and losing any course credits.

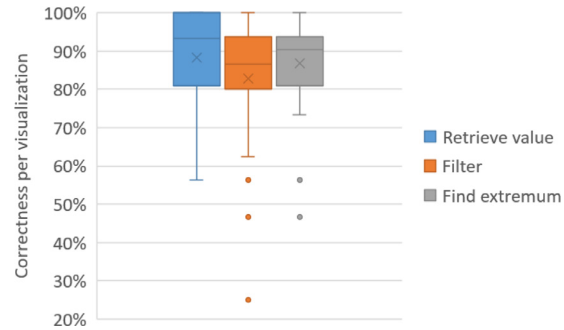


Fig. 3. Correctness of answers per visualization.

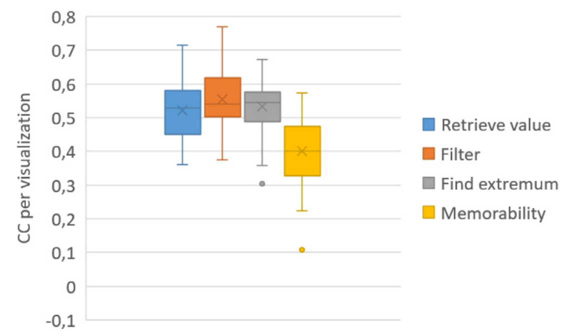


Fig. 4. Similarity between fixations of the same type of activity.

5. Results

Each of the 47 users answered 30 questions in total. This corresponds to 10 answers per task for each participant, resulting in a total of 1410 gathered responses. From these responses, 199 (14.1%) were incorrectly answered and excluded from further analysis, leaving 1211 responses. While, in total, the highest number of incorrect answers was given for the F-task (81), there is no significant difference in *correctness* between the three tasks (Friedman test: $\chi^2(2) = 5.081$, $p = .079$; Fig. 3). The difficulty of the three tasks therefore seems to be comparable.

5.1. Fixation similarities

To test if fixation patterns are more coherent between users solving the same low-level task than when trying to memorize the visualized information, we compared the *inter-participant fixation similarity* between the three low-level analytical tasks, as well as the Mem-task. An ANOVA with Bonferroni-adjusted post-hoc comparisons showed that fixation similarity between participants is indeed significantly higher for the three analytical tasks of our experiment than for the Mem-task (as visualized in Fig. 4; $F(3, 87) = 20.274$; $p < .001$; $\eta^2 = .411$). *This means that users solving the same low-level task indeed have more coherent fixations, thereby confirming hypothesis H1.1.*

We illustrate this finding with a map visualization example in Fig. 5. The matrix in Fig. 6 visualizes the similarity between fixation maps of individual participants for this particular visualization. Fixations obtained for the same task are usually more similar, indicated by blocks of red cells along the matrix diagonal. However, it can also be seen that fixations between the FE-task and the Mem-task are quite similar (top right quarter of the matrix).

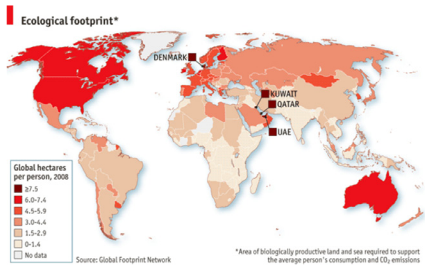


Fig. 5. A map used in our experiment.

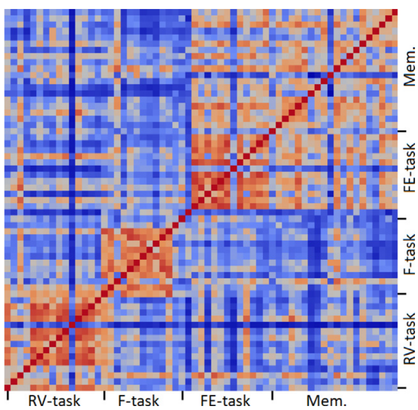


Fig. 6. Fixation similarity matrix for the map in Fig. 5: The matrix visualizes CCs between all participant pairs. Participants are ordered by the activity they performed. The more red, the more similar are the fixation maps. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

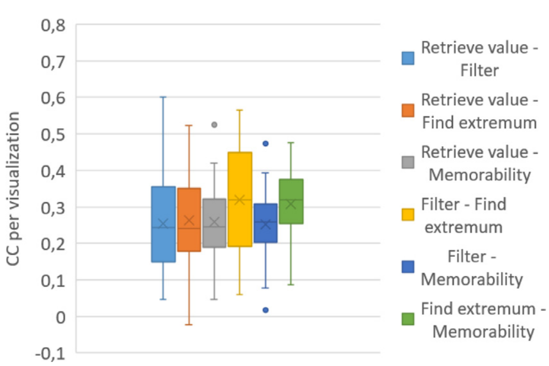


Fig. 7. Similarity between fixations of different types of activity.

Mind that the targets of the FE-task – the countries with the highest ecological footprint – are clearly popping out in Fig. 5.

To further explore this similarity between the fixations of the FE-task and the Mem-task, we compared the *inter-task fixation similarities* between all four tasks using an ANOVA with Bonferroni-corrected post-hoc comparisons. We found that the similarity between the FE-task and the Mem-task (rightmost bar in Fig. 7) is significantly higher than between the F-task and the Mem-task, as well as the RV-task and the Mem-task ($F(5, 145) = 3.136$; $p = .010$; $\eta^2 = .098$). We did not find any statistical differences between inter-task fixation similarities of any other task pairs. In other words, the gaze patterns obtained during the Mem-task are indeed most similar to those of the FE-task, while the other two low-level analytical tasks lead to significantly less similar fixation patterns to the Mem-task.

5.2. Task-dependent fixation sequence

We then tested if the high similarity between the fixation maps of users solving the same low-level analytical task can be explained by the sequence of fixations in the pre-defined task-dependent AOIs. We therefore compared the first fixation times (FF) of the three task-dependent AOIs defined in Table 3 for each task type to test whether users follow the optimal viewing sequence (Fig. 9). We conducted Friedman tests to compare the FFs on the target item label, data point, and value label, respectively, for the RV-task and F-task. For the FE-task, we conducted a Wilcoxon-Signed Rank test to compare the FFs on the target data point and its associated item label. We only found a significant difference in FFs for the RV-task ($\chi^2(2) = 111.972$, $p < .001$). Wilcoxon-Signed Rank pairwise post-hoc comparisons showed that all FFs were significantly different from each other, with the lowest FF for the item label, and the highest for the value label, as predicted in Table 3. The lowest median FF was recorded for data point in the FE-task, and interestingly also in the F-task. However, these differences are not statistically significant. *This only partially confirms hypothesis H1.2: while the task-dependent sequence of AOIs could predict the sequence of first fixations for the RV-task, this sequence could not be observed in the scanpaths recorded for the F-task, and is not pronounced for the FE-task.*

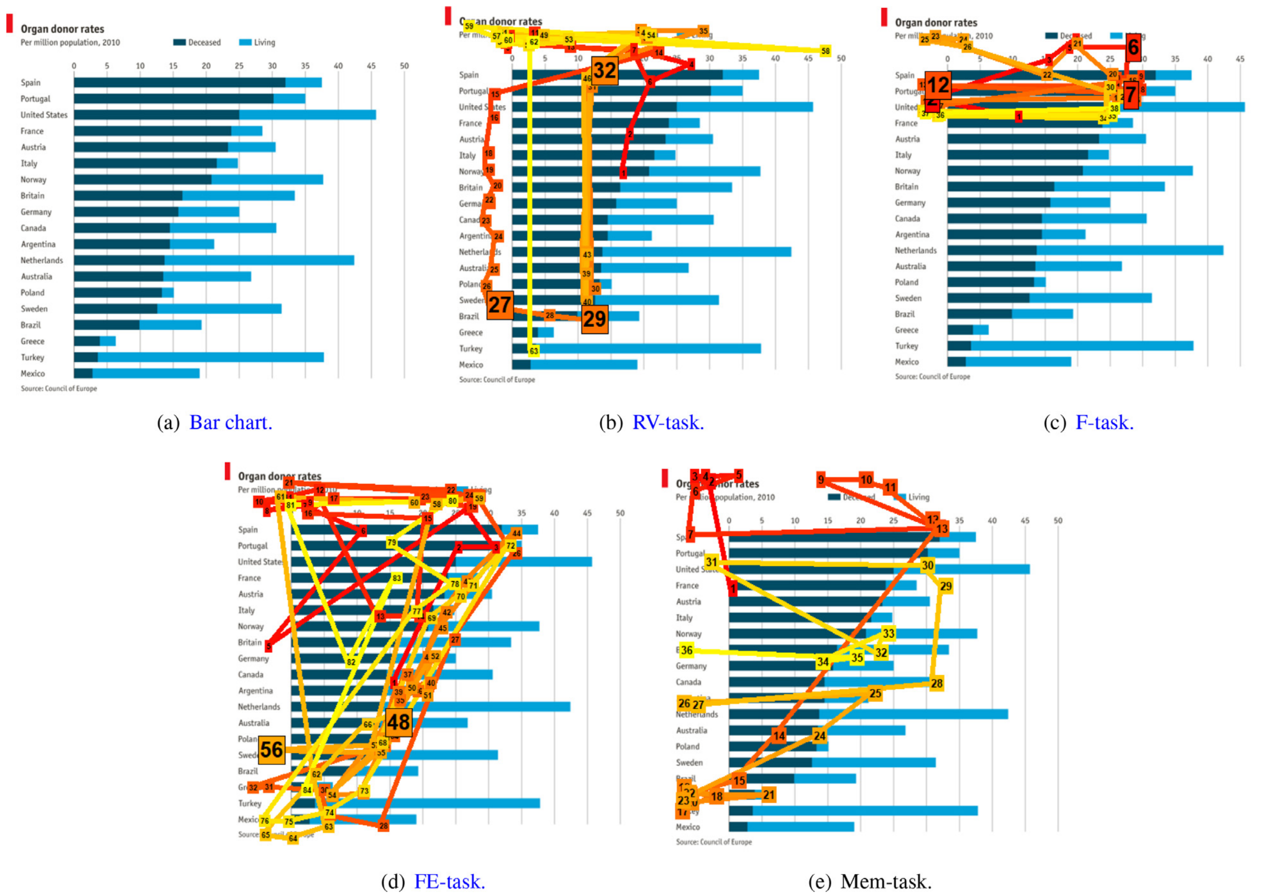
Fig. 8 illustrates the fixation sequences using scanpaths of a selected visualization and selected users performing one of the four tasks each. It is interesting to note how the user of the RV-task sequentially searched for the correct item label first (fixations 15 to 27) before finding the associated data point (fixation 29) and finally its value (fixation 32). In the example scanpaths for the F- and FE-task, users scanned the target data point and its associated labels more often. In all three tasks, the top area containing the legends was visited repeatedly during the task. In contrast, observe how the user of the Mem-task parsed the header and the value legend first, before switching the attention to selected items.

5.3. Fixation-saliency similarities

To test whether the *fixation-saliency similarity* is higher for the Mem-task than for the three low-level analytical tasks, we created saliency maps of all the 30 visualizations using 12 different algorithms and computed the fixation-saliency similarities for all four conditions. In Table 4 and 5, we report average AUC and NSS scores of these saliency models and the *human IO*.

Comparing the performance of Itti et al.'s saliency model [4] to the human IO score in Tables 4 and 5, there is a remarkable gap between the saliency prediction and human visual attention for all four tasks. These scores are similar to the AUC and NSS scores of data visualization eye tracking data compared to Itti et al.'s [4] saliency model, reported by Haass et al. [12] (0.68 and 0.64, respectively). For reference, the average AUC and NSS scores they computed for natural images were 0.77 and 1.06, respectively.

We statistically compared the performance of two selected saliency models between the four tasks: the widely used saliency model by Itti et al. [4], as well as the state-of-the-art for modeling visual attention for visualizations [13] (DVS) using Kruskal–Wallis H tests. For both, AUC and NSS scores, we did not find any statistically significant differences between the tasks using Itti et al.'s saliency model (AUC: $\chi^2(3) = .017$; $p = .999$, NSS: $\chi^2(3) = .117$; $p = .990$). However, we found significant differences for DVS (AUC: $\chi^2(3) = 10.666$; $p = .014$, NSS: $\chi^2(3) = 16.972$; $p = .001$). Bonferroni-corrected Mann–Whitney U post-hoc comparisons showed a significantly higher AUC-score for the Mem-task than for the F-test and a significantly higher NSS-score for the Mem-task than all three low-level analytical tasks.



(a) Bar chart. (b) RV-task. (c) F-task.

(d) FE-task. (e) Mem-task.

Fig. 8. Example scanpaths during the RV-task (How many deceased organ donors per million population does Brazil have?), the F-task (Which country has 25 of deceased organ donors per million population?), the FE-task (Which country has the lowest number of living organ donors per million population?) and the Mem-task, respectively. Fixations are colored according to their order in the scanpath, from red to yellow color. The scanpaths recorded during low-level tasks all contain the optimal viewing sequence of target-dependent AOIs defined in Table 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
The average AUC for each task (saliency models sorted by publication year).

Task	Itti	AIM	GBVS	SUN	CAS	Sign	BMS	eDN	SAMv	SAMr	TextS	DVS	IO
RV	.684	.646	.608	.593	.595	.576	.621	.596	.630	.632	.647	.702	.812
F	.690	.645	.642	.593	.604	.622	.651	.595	.618	.631	.624	.692	.819
FE	.679	.654	.599	.602	.601	.600	.638	.568	.637	.647	.651	.705	.809
Mem	.686	.675	.553	.622	.637	.589	.652	.554	.653	.664	.696	.738	.781

Table 5
The average NSS for each task (saliency models sorted by publication year).

Task	Itti	AIM	GBVS	SUN	CAS	Sign	BMS	eDN	SAMv	SAMr	TextS	DVS	IO
RV	0.66	0.53	0.42	0.39	0.43	0.27	0.39	0.34	0.70	0.65	0.69	0.80	2.00
F	0.66	0.52	0.55	0.41	0.44	0.45	0.51	0.33	0.68	0.65	0.56	0.71	2.00
FE	0.64	0.55	0.41	0.45	0.46	0.35	0.48	0.24	0.80	0.78	0.72	0.81	1.93
Mem	0.67	0.62	0.24	0.53	0.63	0.34	0.54	0.19	0.83	0.88	1.03	1.06	1.50

For the DVS model [13], we can therefore confirm our hypothesis H1.3 that bottom-up saliency strongly influences fixations of users when freely exploring the visualization, but has a significantly lower effect on visual attention when performing a low-level analytical task.

The major difference between the saliency model by Itti et al. [4] and DVS by Matzen et al. [13] is that the latter explicitly encodes text regions within visualizations as highly salient. A potential explanation for the significantly worse performance of DVS for the low-level analytical tasks compared to the Mem-task could be that users direct their attention more towards the data areas than the text areas when performing low-level analytical tasks, than when trying to memorize the visualization. Therefore, we explored the *AOI fixation ratios* in task-independent AOIs defined in Table 1. Indeed, a Kruskal–Wallis H test with Mann-Whitney U post-hoc comparisons revealed that the data areas of visualizations were fixated more frequently during task-based analysis than during the memorability experiment ($\chi^2(3) = 41.435$; $p < .001$; see Fig. 10). The reason for this could be that users were seeking more explicitly for a particular data point and spent less time reading annotations, legends, and titles to memorize textual information, which was irrelevant for the present task. For both, text elements and legends, the fixation ratio was significantly lower for the F-task

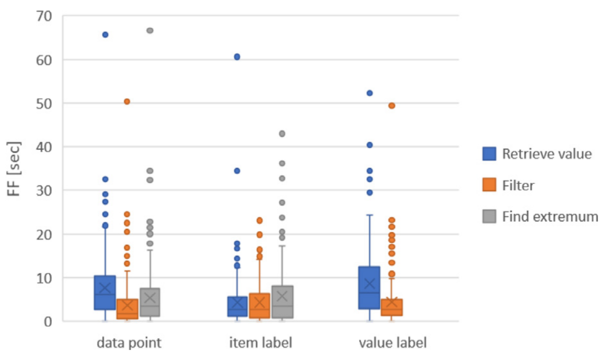


Fig. 9. First fixation times of target-dependent AOIs defined in Table 3.

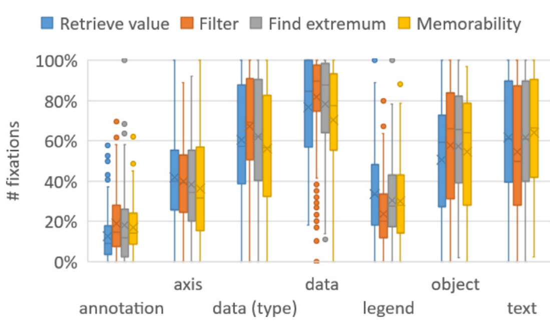
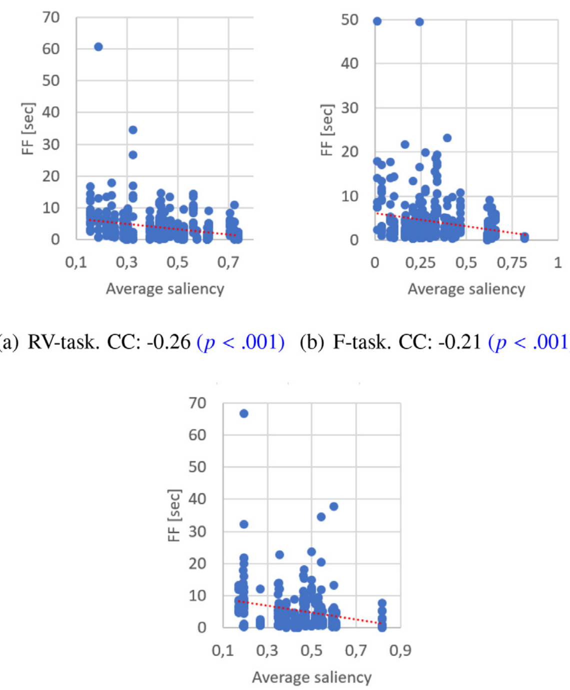


Fig. 10. Fixations in task-independent AOIs.



(a) RV-task. CC: -0.26 ($p < .001$) (b) F-task. CC: -0.21 ($p < .001$)

(c) FE-task. CC: -0.26 ($p < .001$)

Fig. 11. Relationship between the average saliency value of the target (Itti et al.'s algorithm [4]) and the first fixation time (the item label for RV-task, the value label for F-task and the data point itself for FE-task). The red dashed lines are lines of best fit. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

than for all other tasks (Kruskal–Wallis H test: $\chi^2(3) = 23.701$; $p < .001$ and $\chi^2(3) = 37.121$; $p < .001$). However, there is no significant difference between the Mem-task and the other two low-level analytical tasks.

5.4. Correlation between target point saliency and first fixation time

To test the influence of a target area's visual saliency on visual search performance, we computed the correlation between the task-dependent *AOI saliency* (using the model by Itti et al. [4]) and its *FF* for each of the three low-level analytical tasks. We only analyzed the first task-dependent AOI to be fixated according to the task-dependent AOI sequence shown as Step 1 in Table 3. Since each task has a different optimal solution process, we set the target item label as visual search target for the RV-task, the target value label for the F-task, and the target data point for the FE-task. According to our hypothesis H2.1, there should be a negative correlation between the visual search target's AOI saliency and its FF – in other words: the more salient the target, the faster it should be fixated by the user. As visualized in Fig. 11, there is a negative correlation, but this correlation is weak. *In other words, the visual search efficiency for a target in the course of a low-level analytical task does not strongly correlate with its target saliency. Therefore, we have to reject hypothesis H2.1.*

5.5. Saliency of extreme data points

Our assumption is that data points with extreme values usually stand out visually, i.e., have a higher saliency than target data points for the RV-task or the F-task. However, a Friedman test on the *AOI saliency* values computed using Itti et al.'s saliency model showed that there is, in fact, no difference in target data point saliency ($\chi^2(2) = 2.381$; $p = .304$). *Therefore, we have to reject hypothesis H2.2: target data points of the FE-task do not show up as more salient in saliency maps than target data points of the other two tasks.*

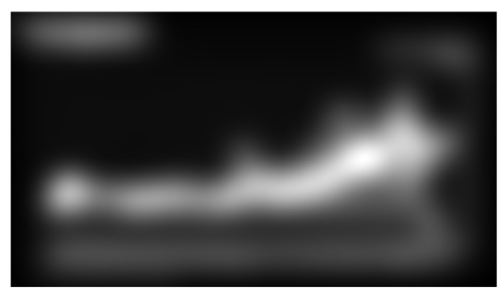


Fig. 12. Saliency map (Itti et al. [4]) of the visualization shown in Fig. 14(b).

For illustration of this result, consider Fig. 14(b): The peak in the line chart intuitively stands out. However, neither is this peak the most salient region in the saliency map (see Fig. 12), nor is it the extremum that was requested in the task question, which is the highest blue bar.

5.6. Find-extremum task efficiency

Finally, we tested the hypothesis that extrema can be found more efficiently than values associated with given items or items associated with given value ranges. We therefore compared the task completion time between the three low-level analytical tasks. A Friedman test showed that there is a significant difference between the tasks: $\chi^2(2) = 16.128$, $p < .001$. Post-hoc Wilcoxon Signed-Rank tests revealed that the F-task takes significantly longer to be solved than the RV-task ($Z = -3.757$, $p < .001$) and the FE-task ($Z = -4.011$, $p < .001$), but there is no difference in task completion time between the RV-task and the FE-task ($Z = -.328$, $p = .743$). *We therefore also have to reject hypothesis 2.3: the FE-task was not more efficient to solve than the RV-task.*

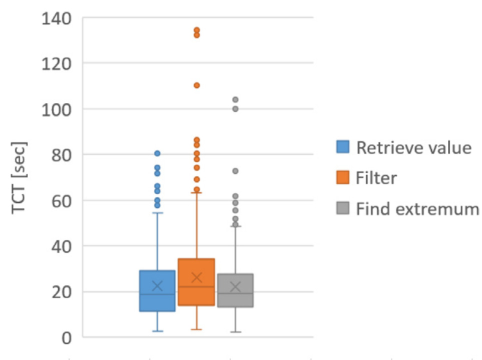
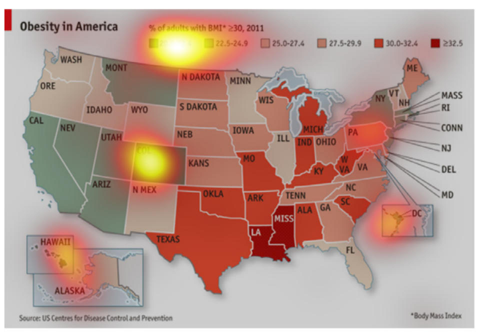
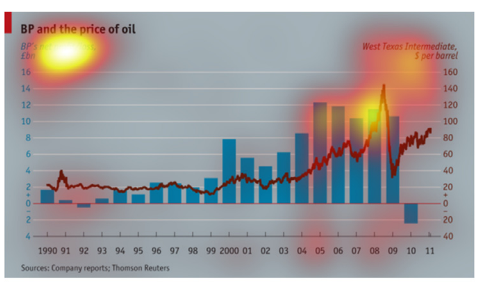


Fig. 13. Task completion time per task.



(a) Which American states have the lowest obesity rates?



(b) When did BP achieve the highest net profit?

Fig. 14. Fixation heat maps of visualizations with the lowest number of correct answers (a) and highest task completion time (b) for the FE-task.

To further explore the difference of task efficiency between the FE-task and the other two low-level analytical tasks, we compared how quickly participants fixated the target data points for the first time (FF of target data points, see first block in Fig. 9). Search for the target data point took significantly longer for the RV-task than for the other two task, and was accomplished fastest during the F-task (Kruskal–Wallis H: $\chi^2(2) = 56.512$; $p < .001$). In other words, visual search for an extreme data point was more efficient than finding a data point associated with a given item label. However, visual search for any target data point within a given value range was even more efficient. Mind how this finding contrasts the differences in task completion time. This illustrates that efficient visual search for a target data point does not automatically lead to an efficient overall task performance.

For illustration purposes, we show the visualizations leading to the lowest correctness score and highest task completion time (i.e., lowest overall performance), respectively, for the FE-task in Fig. 14. In Fig. 14(a), the low correctness was caused by the fact that the second target item – Hawaii – has a salient color, but is very small and therefore easy to miss. In Fig. 14(b) (i.e. the example leading to the highest task completion time in the FE-task) users were confronted with charts encoding two attributes in a one-dimensional chart and had to figure out the visual mapping of the target attribute first. This can explain why most fixations were captured around the legend on the top left in Fig. 14(b). This can also explain why extreme targets do not necessarily show up as salient regions in the visualization, since they can be confounded by other data attributes, such as the size and shape given by the geography of the state in Fig. 14(a) or a second dependent variable encoded in Fig. 14(b).

6. Discussion

For our discussion, we will relate our results to our hypotheses and finally discuss the implications.

6.1. Influence of bottom-up saliency during task-based visual analysis

Fixation patterns of users solving different low-level analytical tasks showed that fixations between users solving the same task highly correlate, while the fixation map correlation between users trying to memorize the visualization is significantly lower. This result was expected (**H1.1**). However only in the RV-task could we show that users clearly fixated the areas of interest in the optimal sequence for solving the task. The given low-level analytical task therefore seems to have a measurable top-down guidance for the users *where* to look, but not necessarily in which order (H1.2).

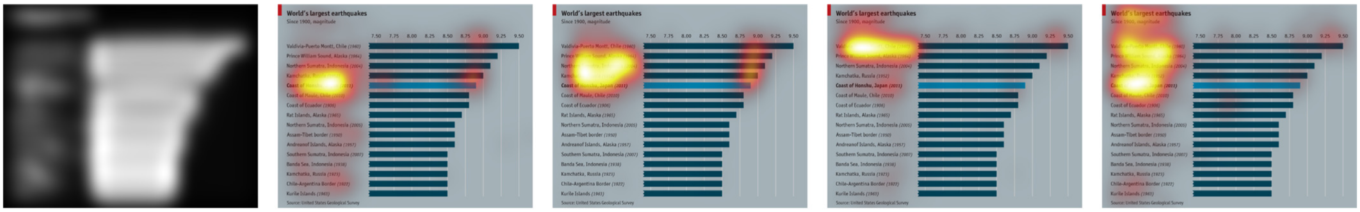
An unexpected finding during our experiment was that fixation maps of the memorability experiment much closer resembled those of the find-extremum task than those of the retrieve-value or filter task. There are two possible explanations for this observation. A naive assumption could be that because extreme values are highly salient, the users' attention is guided there during exploratory analysis. This might be true in some cases (e.g., Fig. 5), but is often not the case (e.g., Fig. 8 and Fig. 12). In fact, we found that target data points in our FE-tasks did not have a higher bottom-up saliency than those of the RV-task and the F-task (H2.2).

An alternative explanation is that users were intentionally seeking for extrema as representative values to memorize the content of the visualization. This tendency is reflected in the selected descriptions of visualization content of users in Borkin et al.'s [2] memorability experiment (supplemental material). Most of the listed user descriptions contain a short summary of what is visualized together with one or more extreme items. This would mean that a memorability task would lead to similar top-down guidance as a find-extremum task.

Despite the higher diversity of the fixations during the memorability experiment, the fixations are more likely to co-incide with highly salient regions than during low-level analytical tasks. This is true for the DVS model recently presented by Matzen et al. [13] (**H1.3**). However, for the seminal saliency model by Itti et al. [4], the fixation-saliency similarities are equally low for the low-level analytical tasks as for the memorability task. The difference between these two saliency models is that DVS explicitly detects regions containing text and marks these regions as highly salient. Since it has been found that users attend textual elements for a long time when trying to memorize or free-viewing a visualization [46], DVS can better predict fixation patterns while performing exploratory visual analysis. During task-based visual analysis, however, users' attention is more strongly directed towards the data areas of the visualization, while, presumably, text areas are targeted only selectively.

6.2. Influence of bottom-up saliency during visual search

Our second hypothesis was that bottom-up visual saliency may be a useful tool to predict the efficiency of visual search that needs



(a) Saliency map. (b) RV-task. (c) F-task. (d) FE-task. (e) Memorability [2].

Fig. 15. Saliency map [4] and fixations during the RV-task (What was the earthquake magnitude in Kamchatka in 1952?), the F-task (Which earthquakes reached a magnitude of 9 and higher?), the FE-task (What was the largest earthquake since 1900?), and the Mem-task for a bar chart.

to be conducted in the course of task-based visual analysis. Depending on the task, this may be a visual search for an item label, a value label, or an extreme data point. Especially in the latter case, we expected to see that extreme data points stand out in the saliency maps, and that therefore find-extremum tasks are more efficient to solve overall.

However, we could neither find a strong correlation between the target point's saliency and its first fixation time (H2.1), nor a significantly higher saliency of extreme data points compared to data point of the other the two tasks (H2.2), nor an increased task efficiency for the find-extremum task (H2.3). In other words: extreme data points are not necessarily more salient in classic visualizations, and more salient data points are not necessarily faster to detect during task-based visual analysis.

We showed several examples why extreme data points in these in-the-wild visualizations are often not more salient than other data points. In many examples, there is more than one attribute encoded in the visualization, so that two visual encodings compete for the user's attention (see, for instance, Fig. 8 or Fig. 14).

Another example for a mismatch between apparent pop-out effect and bottom-up visual saliency, as modeled by Itti et al.'s saliency model, is illustrated in Fig. 15. Here, the highlighted bar is associated with a low saliency value, because the contrast of the bright bar to the bright background is lower than of the remaining bars.

6.3. Alternative visual attention models for information visualization

As illustrated in Fig. 15, classic saliency models may contradict the intuitive impression of which data points stand out from a visualization. In information visualization, attention is affected not only by pixels that differ from their neighborhood. Instead, attention may rather be attracted by graphical marks whose visual channels differ notably from the remaining marks. Note that graphical marks and visual channels are generated by a visualization designer. Saliency models targeted towards information visualization therefore do not need to be computed from the resulting image. Instead, the saliency of a graphical mark could be computed when the visualization is constructed from its relative visual prominence with respect to all remaining graphical marks.

In the field of information visualization, a few specialized saliency models quantifying the visual prominence of graphical marks have been proposed. For instance, Lee et al. [44] introduced *point and class saliency* measures to quantify the color saliency of a single data point or a class of data points in a categorical map visualization. Waldner et al. [63] derived a visual prominence measure of data points in scatterplots that use luminance and blur for highlighting. In the future, it will therefore be of interest to compare the prediction accuracy of these measures to classic saliency models, or to find ways how to combine them.

However, while we could observe that more attention is targeted towards the actual data points when performing a low-level analytical task as compared to the memorability task (see Fig. 10), a considerable amount of attention is also attributed to textual elements, like labels and legends. By taking text into account, the DVS model [13] can therefore achieve considerably higher accuracy than the classic saliency models for the memorability task. When performing low-level analytical tasks, users also direct a lot of attention towards text, but it is more selective. For instance, in the RV-task, users have to search for the matching item label. For the example shown in Fig. 15, this would correspond to a visual search for a text label on the vertical axis of the chart. In the F-task, users have to find a matching value range, which corresponds to a label on the horizontal axis.

For real-world scenes, Wang and Pomplun [64] showed that users are likely to direct their attention to text content, if they assume that it is at an informative location. Low-level features, like color, are not the main attractors. In the case of visual analysis, whether or not users assume text to be at an informative location is task-dependent. For the RV-task in Fig. 15, for instance, the vertical location and length of the textual item label might be a more reliable prediction for fixations than its visual saliency. For the FE-task, on the other hand, a saliency model targeted towards graphical marks and their visual channels, could be more applicable.

6.4. Study limitations

Our study highlighted some unknown aspects about visual attention during task-based visual analysis. However, there are some limitations to our study.

First, the set of visualizations and the task questions were quite heterogeneous. The disadvantage is that there are, therefore, many factors potentially confounding the results, such as different visualization types, varying numbers of dependent variables encoded in the visualizations, or the usage of human recognizable objects. We suspect that some hypotheses would require a more controlled setup to be fully verified. One advantage of this variety, however, is that the visualizations of the MASSVIS database are considered as representative samples of in-the-wild visualizations that non-expert users are regularly confronted with. Another advantage is that we were able to collect a variety of eye tracking samples and could informally explore various factors that may have an influence on visual attention during goal-driven analysis.

Second, our user group was composed of data visualization students who have gained some experience in analyzing visualization compared to novices. As also shown in a prior study [45], it can be expected that novices are less strongly guided by top-down factors when performing confirmatory analysis than more experienced users. The participants of Borkin et al.'s [2] experiment were "recruited from the local communities of Cambridge and Boston", but no information about their visualization literacy is provided. If these participants were novices, an alternative explanation to the lower human IO scores could be that users of the memorability experiment analyzed the visualization in a less structured way than the users of our task-based visual analysis experiment.

Third, we only tested three very simple analytical tasks in our experiment. However, in the task taxonomy by Amar et al. [58], there are more low-level tasks, like sort, determine range, cluster, or characterize distribution. Many of these tasks are not straightforward to solve on some of the visualization types in the MASSVIS database or require very elaborate responses. We therefore did not include them in this experiment. The eye movement characteristics during these tasks are yet to be investigated in future work.

7. Conclusions

Our results show that despite having improved bottom-up saliency models for information visualization, like DVS [13], the influence of bottom-up visual saliency is drastically reduced during task-based visual analysis. We showed that users focus more on data areas of the visualization during task-based visual analysis than when trying to memorize a visualization. Therefore, the added text saliency in the DVS model did not increase the accuracy for task-based visual analysis in the same extent as for exploratory visual analysis. However, despite the increased attention in the data area, we did not find a strong correlation between a task-dependent area's saliency and visual search efficiency. This means that visual attention is only slightly affected by early features when performing task-based visual analysis using information visualization – in contrast to observation of natural images or during exploratory visual analysis. Yet, fixations between users are more similar than during the memorability experiment. This means that task-based visual analysis is strongly guided by top-down factors imposed by the task.

To improve existing saliency models and tailor them more towards task-based visual analysis, we therefore recommended to merge classic image-based saliency models with object-based saliency models. When quantifying how much individual graphical marks stand out from their surrounding marks, the model should localize and identify the marks, compare their features at object level (e.g. color, orientation, size and shape) and estimate their relationships. In addition, other element types in a visualization, such as text areas, legends or axes, should be also incorporated in the model. For instance, saliency of text labels should vary with the task, so that text at informative locations for a given task receives higher saliency.

Our experiment was conducted using rather simple in-the-wild visualizations and low-level analytical tasks. These tasks could be easily solved by attending to only a few areas of interest in the visualization. More complex exploratory or task-based visual analysis requires more complex visual encodings, multiple (coordinated) views, and user actions to explore the data, like dynamic queries or brushing and linking. Ultimately, saliency models tailored to information visualization have to be extended to model the user's attention in these more complex displays that dynamically change as the user is interacting.

In the future, it will be important to perform more systematic comparisons of eye tracking patterns between low-level analytical tasks by using carefully selected simple charts, such as bar charts. Also, further low-level tasks, such as assessing correlations in scatterplots, need to be explored. Finally, our user group had a quite high level of visualization literacy. Comparing groups of visualization experts with non-experts could reveal whether top-down guidance of low-level analytical tasks is similarly strong for non-expert users. Since neural network models successfully reduced the gap between human fixations and saliency prediction for natural scenes, a similar approach could be applied in data visualizations, too. By training a neural network on viewers' fixation data acquired during a specific task, saliency maps could be tailored for a particular viewer and task, respectively.

Acknowledgments

The authors would like to thank for financial contribution from the Slovak Scientific Grant Agency, VEGA 1/0874/17, and from the Research and Development Operational Programme for the project "University Science Park of STU Bratislava", ITMS 26240220084, co-funded by the European Regional Development Fund. This research has also been financed through the Austrian Science Fund FWF (project T 752-N30) and the Vienna Science and Technology Fund WWTF (project VRG11-010).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.cag.2018.01.010

References

[1] Borkin MA, Vo AA, Bylinskii Z, Isola P, Sunkavalli S, Oliva A, et al. What makes a visualization memorable? IEEE Trans Vis Comput Graph 2013;19(12):2306–15.
[2] Borkin MA, Bylinskii Z, Kim NW, Bainbridge CM, Yeh CS, Borkin D, et al. Beyond memorability: visualization recognition and recall. IEEE Trans Vis Comput Graph 2016;22(1):519–28.
[3] Binder MD, Hirokawa N, Windhorst U. Encyclopedia of Neuroscience, 3166. Berlin Heidelberg: Springer; 2009.
[4] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 1998;20(11):1254–9.
[5] Treisman AM, Gelade G. A feature-integration theory of attention. Cogn Psychol 1980;12(1):97–136.
[6] Goldstein E. Encyclopedia of perception. SAGE Publications; 2010. ISBN 978-1-4129-4081-8.
[7] Borji A, Itti L. CAT2000: a large scale fixation dataset for boosting saliency research. CoRR 2015. abs/1505.03581
[8] Parkhurst D, Law K, Niebur E. Modeling the role of salience in the allocation of overt visual attention. Vis Res 2002;42(1):107–23.
[9] Peters RJ, Iyer A, Itti L, Koch C. Components of bottom-up gaze allocation in natural images. Vis Res 2005;45(18):2397–416.
[10] Garlandini S, Fabrikant SI. Evaluating the effectiveness and efficiency of visual variables for geographic information visualization. In: Proceedings of international conference on spatial information theory. Springer; 2009. p. 195–211.
[11] Jänicke H, Chen M. A salience-based quality metric for visualization. In: Computer graphics forum, 29. Wiley Online Library; 2010. p. 1183–92.
[12] Haass MJ, Wilson AT, Matzen LE, Divis KM. Modeling human comprehension of data visualizations. Cham: Springer International Publishing; 2016. p. 125–34. ISBN 978-3-319-39907-2.
[13] Matzen LE, Haass MJ, Divis KM, Wang Z, Wilson AT. Data visualization saliency model: a tool for evaluating abstract data visualizations. IEEE Trans Vis Comput Graph 2018;24(1):563–73.
[14] Harel J, Koch C, Perona P. Graph-based visual saliency. In: Proceedings of the 19th international conference on neural information processing systems, NIPS'06. Cambridge, MA, USA: MIT Press; 2006. p. 545–52.
[15] Hou X, Harel J, Koch C. Image signature: highlighting sparse salient regions. IEEE Trans Pattern Anal Mach Intell 2012;34(1):194–201.
[16] Zhang J, Sclaroff S. Saliency detection: a Boolean map approach. In: Proceedings of 2013 IEEE international conference on computer vision; 2013. p. 153–60.
[17] Bruce NDB, Tsotsos JK. Saliency based on information maximization. In: Proceedings of the 18th international conference on neural information processing systems, NIPS'05. Cambridge, MA, USA: MIT Press; 2005. p. 155–62.
[18] Zhang L, Tong MH, Marks TK, Shan H, Cottrell GW. Sun: a bayesian framework for saliency using natural statistics. J Vis 2008;8(7):32.
[19] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection. IEEE Trans Pattern Anal Mach Intell 2012;34(10):1915–26.
[20] Vig E, Dorr M, Cox D. Large-scale optimization of hierarchical features for saliency prediction in natural images. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2014. p. 2798–805.
[21] Cornia M, Baraldi L, Serra G, Cucchiara R. Predicting human eye fixations via an LSTM-based saliency attentive model. CoRR 2016. abs/1611.09571
[22] Achanta R, Süsstrunk S. Saliency detection for content-aware image resizing. In: 2009 16th IEEE international conference on image processing (ICIP); 2009. p. 1005–8.
[23] Marchesotti L, Cifarelli C, Csurka G. A framework for visual saliency detection with applications to image thumbnailing. In: Proceedings of the 2009 IEEE 12th international conference on computer vision; 2009. p. 2232–9.
[24] Ardizzone E, Bruno A, Mazzola G. Saliency based image cropping. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 773–82. ISBN 978-3-642-41181-6.
[25] Dong L, Lin W, Zhu C, Seah HS. Selective rendering with graphical saliency model. In: Proceedings of the 10th IVMSP workshop: perception and visual signal analysis; 2011. p. 159–64.

[26] Butko NJ, Zhang L, Cottrell GW, Movellan JR. Visual saliency model for robot cameras. In: Proceedings of 2008 IEEE international conference on robotics and automation; 2008. p. 2398–403.

[27] Siagian C, Itti L. Biologically inspired mobile robot vision localization. IEEE Trans Robot 2009;25(4):861–73.

[28] Endres D, Vintila F, Bruce N, Bouecke J, Kornprobst P, Neumann H, et al. Hooligan detection: the effects of saliency and expert knowledge. In: Proceedings of perception ECVP, 39. Lausanne, Switzerland; 2010. p. 193.

[29] Lee W, Lee G, Ban S-W, Jung I, Lee M. Intelligent video surveillance system using dynamic saliency map and boosted Gaussian mixture model. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 557–64. ISBN 978-3-642-24965-5.

[30] Shen C, Zhao Q. Webpage saliency. Cham: Springer International Publishing; 2014. p. 33–46. ISBN 978-3-319-10584-0.

[31] Bylinskii Z, Kim NW, O'Donovan P, Alsheikh S, Madan S, Pfister H, et al. Learning visual importance for graphic designs and data visualizations. In: Proceedings of the 30th Annual ACM symposium on user interface software and technology; 2017a. p. 57–69.

[32] Bruce NDB, Tsotsos JK. Saliency, attention, and visual search: an information theoretic approach. J Vis 2009;9(3):5.

[33] Judd T, Ehinger K, Durand F, Torralba A. Learning to predict where humans look. In: Proceedings of 2009 IEEE 12th international conference on computer vision; 2009. p. 2106–13.

[34] Kootstra G, Nederveen A, De Boer B. Paying attention to symmetry. In: Proceedings of british machine vision conference (BMVC2008). The British Machine Vision Association and Society for Pattern Recognition; 2008. p. 1115–25.

[35] Ramanathan S, Katti H, Sebe N, Kankanhalli M, Chua T-S. An eye fixation database for saliency detection in images. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 30–43. ISBN 978-3-642-15561-1.

[36] Borji A, Tavakoli HR, Sihite DN, Itti L. Analysis of scores, datasets, and models in visual saliency prediction. In: Proceedings of 2013 IEEE international conference on computer vision; 2013. p. 921–8.

[37] Huang X, Shen C, Boix X, Zhao Q. Salicon: reducing the semantic gap in saliency prediction by adapting deep neural networks. In: Proceedings of 2015 IEEE international conference on computer vision (ICCV); 2015. p. 262–70.

[38] Zhang J, Sclaroff S. Exploiting surroundedness for saliency detection: a Boolean map approach. IEEE Trans Pattern Anal Mach Intell 2016;38(5):889–902.

[39] Liang M, Hu X. Predicting eye fixations with higher-level visual features. IEEE Trans Image Process 2015;24(3):1178–89.

[40] Xu L, Zeng L, Duan H, Sowah NL. Saliency detection in complex scenes. EURASIP J Image Video Process 2014;2014(1):31.

[41] Itti L, Dhavale N, Pighin F. Realistic avatar eye and head animation using a neurobiological model of visual attention. In: Proceedings of SPIE 48th annual international symposium on optical science and technology; 2003. p. 64–78.

[42] Kim Y, Varshney A. Saliency-guided enhancement for volume visualization. IEEE Trans Vis Comput Graph 2006;12(5):925–32.

[43] Haroz S, Whitney D. How capacity limits of attention influence information visualization effectiveness. IEEE Trans Vis Comput Graph 2012;18(12):2402–10.

[44] Lee S, Sips M, Seidel HP. Perceptually driven visibility optimization for categorical data visualization. IEEE Trans Vis Comput Graph 2013;19(10):1746–57.

[45] Matzen LE, Haass MJ, Tran J, McNamara LA. Using eye tracking metrics and visual saliency maps to assess image utility. Electron Imag 2016;2016(16):1–8.

[46] Matzen LE, Haass MJ, Divis KM, Stites MC. Patterns of attention: how data visualizations are read. Cham: Springer International Publishing; 2017. p. 176–91. ISBN 978-3-319-58628-1.

[47] Kurzhals K, Fisher B, Burch M, Weiskopf D. Evaluating visual analytics with eye tracking. In: Proceedings of the fifth workshop on beyond time and errors: novel evaluation methods for visualization, BELIV '14. New York, NY, USA: ACM; 2014. p. 61–9. ISBN 978-1-4503-3209-5.

[48] Siirtola H, Laivo T, Heimonen T, Räihä KJ. Visual perception of parallel coordinate visualizations. In: Proceedings of the 2009 13th international conference information visualisation; 2009. p. 3–9.

[49] Burch M, Andrienko G, Andrienko N, Höferlin M, Raschke M, Weiskopf D. Visual task solution strategies in tree diagrams. In: Proceedings of 2013 IEEE Pacific visualization symposium (PacificVis); 2013. p. 169–76.

[50] Burch M, Konevtsova N, Heinrich J, Hoeferlin M, Weiskopf D. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. IEEE Trans Vis Comput Graph 2011;17(12):2440–8.

[51] Pohl M, Schmitt M, Diehl S. Comparing the readability of graph layouts using eyetracking and task-oriented analysis. In: Proceedings of the Fifth Eurographics conference on computational aesthetics in graphics, visualization and imaging, computational Aesthetics'09. Aire-la-Ville, Switzerland: Eurographics Association; 2009. p. 49–56. ISBN 978-3-905674-17-0.

[52] Goldberg JH, Helfman JI. Comparing information graphics: a critical look at eye tracking. In: Proceedings of the 3rd BELIV'10 workshop: beyond time and errors: novel evaluation methods for information visualization, BELIV '10. New York, NY, USA: ACM; 2010. p. 71–8. ISBN 978-1-4503-0007-0.

[53] Kim SH, Dong Z, Xian H, Upatising B, Yi JS. Does an eye tracker tell the truth about visualizations?: Findings while investigating visualizations for decision making. IEEE Trans Vis Comput Graph 2012;18(12):2421–30.

[54] Goldberg J, Helfman J. Eye tracking for visualization evaluation: reading values on linear versus radial graphs. Inf Vis 2011;10(3):182–95.

[55] Raschke M, Blascheck T, Richter M, Agapkin T, Ertl T. Visual analysis of perceptual and cognitive processes. In: Proceedings of 2014 international conference on information visualization theory and applications (IVAPP); 2014. p. 284–91.

[56] Bylinskii Z, Borkin MA, Kim NW, Pfister H, Oliva A. Eye fixation metrics for large scale evaluation and comparison of information visualizations. Cham: Springer International Publishing. p. 235–55. ISBN 978-3-319-47024-5.

[57] Mazza R. Introduction to information visualization. Springer; 2009. ISBN 978-1-84800-218-0.

[58] Amar R, Eagan J, Stasko J. Low-level components of analytic activity in information visualization. In: Proceedings of IEEE symposium on information visualization, 2005. INFOVIS 2005.; 2005. p. 111–17.

[59] Brehmer M, Munzner T. A multi-level typology of abstract visualization tasks. IEEE Trans Vis Comput Graph 2013;19(12):2376–85.

[60] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. CoRR 2014. abs/1409.1556

[61] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of 2016 IEEE conference on computer vision and pattern recognition (CVPR); 2016. p. 770–8.

[62] Bylinskii Z, Judd T, Oliva A, Torralba A, Durand F. What do different evaluation metrics tell us about saliency models? CoRR 2016. abs/1604.03605

[63] Waldner M, Karimov A, Gröller E. Exploring visual prominence of multi-channel highlighting in visualizations. In: Proceedings of spring conference on computer graphics; 2017.

[64] Wang H-C, Pomplun M. The attraction of visual attention to texts in real-world scenes. J Vis 2012;12(6). 26–26