Patterns of Attention: How Data Visualizations are Read

Laura E. Matzen, Michael J. Haass, Kristin M. Divis, Mallory C. Stites

Sandia National Laboratories, Albuquerque, USA

lematze@sandia.gov, mjhaass@sandia.gov,
kmdivis@sandia.gov, mcstite@sandia.gov

Abstract. Data visualizations are used to communicate information to people in a wide variety of contexts, but few tools are available to help visualization designers evaluate the effectiveness of their designs. Visual saliency maps that predict which regions of an image are likely to draw the viewer's attention could be a useful evaluation tool, but existing models of visual saliency often make poor predictions for abstract data visualizations. These models do not take into account the importance of features like text in visualizations, which may lead to inaccurate saliency maps. In this paper we use data from two eye tracking experiments to investigate attention to text in data visualizations. The data sets were collected under two different task conditions: a memory task and a free viewing task. Across both tasks, the text elements in the visualizations consistently drew attention, especially during early stages of viewing. These findings highlight the need to incorporate additional features into saliency models that will be applied to visualizations.

Keywords: Data visualizations, text, eye tracking

1   Introduction

Data visualizations are widely used to convey information, yet it is difficult to evaluate whether or not they are effective. Previous work on graph comprehension has suggested that the effectiveness of a graph depends on the relationships between the visual properties of the graph, the experience and expectations of the user, and the type of information to be extracted from the graph (reviewed in [27]). As such, the recommendations for the "best" way to present as dataset may differ for every new visualization created.

Eye tracking can provide insight into how people comprehend data visualizations. It is a useful measure of where visual attention is being directed, as attention is typically closely linked with gaze location (see [25] for review). Eye tracking measures are divided into fixations (periods of relative stability) and saccades (ballistic movements, during which effectively no new visual information is processed). In general, people tend to spend more time looking at, and make more fixations on, areas of a display that are difficult to process or important to their current task goals [25]. Graph comprehension researchers have devised various metrics to evaluate ease of processing information from graphs. For example, the time to the first fixation in a region is taken as

an indicator of how easy the region was to find. The time from landing in a region to making a decision about a graph is taken as an indicator of how easy the information was to process after it was found (see [5] and [12] for discussions of other useful metrics). In this way, eye movement patterns can provide a window into the ongoing cognitive processes taking place as people comprehend data visualizations.

Although eye tracking metrics have the potential to be useful in evaluating the effectiveness of a data visualization in conveying information to a viewer, they must be evaluated within the context of many different factors that affect viewers' eye movement patterns. One factor is the viewer's task, which has a large impact on his or her eye movements. For example, Goldberg and Helfman [12] found more fixations to a graph when viewers subtracted or added data than when they were tasked with simply extracting values. Similarly, Strobel et al. [28] found more fixations to line graphs than bar graphs when users were performing trend analyses. The type of visualization technique used also impacts how users take in the same information, with, for example, more fixations for unfamiliar or difficult visualizations [10,11]. Characteristics of the viewer also influence eye movement behaviors. More experienced users can extract information in less time and may pay attention to different aspects of a visualization than less experienced viewers [21].

To address the diversity of factors that can influence what aspects of a data visualization draw the viewer's attention, it is useful to distinguish between top-down and bottom-up visual attention. Top-down, or goal-oriented, visual attention is driven by the viewer's goals and expectations. Meanwhile, bottom-up visual attention is driven by the physical characteristics of the image, such as color and contrast [9,23]. There are existing models of bottom-up visual attention that use the visual properties of an image to predict which parts of the image will draw a viewer's attention (cf. [16]). These models take an input image and generate a map of visual saliency, where the salient regions are those that are more likely to attract bottom-up visual attention. To assess the ability of the models to predict where people will look, the saliency maps are compared to eye tracking data collected under free viewing conditions (i.e. the participants view the images for a fixed amount of time with no specific task to complete; [2]).

In prior work, we developed evaluation approaches for data visualizations that incorporate eye tracking data, saliency maps, and sensor phenomenology [21]. We demonstrated that comparing saliency maps to eye tracking data collected from experienced and inexperienced viewers can highlight the differences between features that are highly salient and features that are highly task-relevant. Using saliency maps and eye tracking data in combination was informative for teasing apart which aspects of the data drew viewers' attention from both the bottom-up and top-down perspectives. This information can then be applied to improving the visual representation of the data and to assessing feature detection algorithms.

In subsequent work, we have attempted to extend this general approach from the realm of sensor data into the domain of abstract data visualizations. Predicting what parts of a visualization will draw the user's attention would be a useful first pass at evaluation [26]. However, our work has found that existing saliency maps do not work well for predicting where viewers will look in abstract data visualizations. In Haass et

al. [13], we evaluated the ability of multiple models of visual saliency to explain viewing behaviors in natural scenes as well as data visualizations. The models performed well for natural scenes, but they were poor predictors of viewing patterns for abstract data visualizations. Based on comparisons of the saliency maps and fixations, a large part of the discrepancy seems to be due to people attending to text in the data visualizations. The text elements received a high proportion of the viewers' fixations, but were generally not identified as salient in the saliency maps. The visual properties of text are quite different from those of features in natural scenes, so models designed to predict eye movement in scene viewing do not account for the text's influence on the viewer's patterns of attention.

The findings of Haass and colleagues [13] highlight the point that abstract data visualizations are very different from natural scenes – each element was chosen by a designer and is there for a reason. In this way, data visualizations share some commonalities with print ads, which are also comprised of a combination of images and text to convey a message. Eye-tracking techniques have been applied to the print ad literature (see review in [14]), and their findings have largely echoed the graph comprehension literature in showing that the viewer's goals have a huge influence over eye movement guidance. One robust finding is that when viewers are asked to learn about a product or decide on a product to purchase, they tend to look at the text of an ad earlier and for more time—roughly 70% of viewing time—than when they are evaluating an ad for its likeability or effectiveness (in which case viewers show a preference for fixating the images). Readers are also more likely to fixate, and spend more time viewing, ads with large text relative to small text, although the same is not true for photo size. Importantly, the characteristics of eye movements also change when people look at different elements of ads: readers make longer fixation durations and saccades on graphical elements compared to text.

It is worth noting that the graphical elements in ads and data visualizations serve different purposes (display a product versus convey numeric information, respectively), and so different mechanisms might influence viewing patterns for these two visualization types. However, gaining an understanding of the features that drive eye movements in a range of visualizations is an important first step in understanding how viewers allocate their attention between text and graphics during successful comprehension. Uncovering these basic features will help inform models of visual saliency. Our previous work has already shown that simple saliency maps are not sufficient to explain viewing patterns in visualizations [13]. Updating these models to incorporate insights regarding how users allocate their attention between text and graphics might help visualization designers to assess their designs more accurately than models that treat text similarly to graphics.

In the present study, we take a closer look at viewers' attention to text in data visualizations. First, we analyzed eye tracking data collected by Borkin and colleagues [3] in the context of a memory study. While their study included a wide range of visualizations, we selected and analyzed a subset of the data that included frequently-used graph types, such as bar charts and line graphs. We then assessed how much attention participants devoted to different regions of the visualizations, paying particular attention to how attention was allocated to regions that contained text compared to those that

did not. The data collected by Borkin et al. [3], henceforth referred to as the MASSVIS data, was collected during a memory study. The parameters of this task are somewhat different from those used in the eye tracking datasets that are commonly used to evaluate visual saliency models. To address this, we collected eye tracking data from a new group of participants who completed a free viewing task for the same subset of the MASSVIS images and an additional set of newly created data visualizations.

2   Viewing Data Visualizations in a Memory Task

To study how viewers divide their attention between text and graphics in data visualizations, we began with an analysis of a subset of the MASSVIS dataset (http://massvis.mit.edu/). These data were collected during a memory study in which participants viewed images for 10 seconds and were later tested on their memory for the visualizations via recognition and recall tests [3].

For the present analysis, we selected a subset of 35 images from the MASSVIS study. These images represented a variety of commonly used types of data visualizations, all of which contained some combination of text and graphical representations of data. The subset included four area plots, four bar charts, one bubble plot, four column charts (including two double Y-axis plots in which a line graph was overlaid on the column charts), three correlation plots, three line graphs, two map-based visualizations, three network diagrams, three pie charts, and five scatter plots. In addition to these 32 images, we included the three visualizations that had the best match between the eye tracking data and the saliency maps in our prior evaluation of saliency models [13]. These included two infographics and one line graph.

Regions of interest (ROIs) were defined for the stimulus set, dividing the visualizations into the following regions: Title, Data, Data Area, X-Axis, X-Axis Label, Y-Axis, Y-Axis Label, Legend, Data Labels, and Text. For each visualization, the ROIs were marked using GIMP software (www.gimp.org). The ROIs were tightly drawn to the edges of each region.

Scan paths, representing the sequence of fixations across the ROIs for each participant and each visualization where constructed using MATLAB [20]. Fixations were counted as falling within an ROI if their center, defined as the geometric median of all points in the fixation, fell within a 1 degree viewing angle of the ROI, approximating the participants' useful field of view. If the same fixation could be assigned to multiple ROIs, multiple variants of the scan path were generated. However, for the purpose of this analysis, only the first variant was used. A total of 562 scan paths were analyzed, with an average of 16 scan paths from different participants for each visualization. There were an average of 36 fixations per scan path (range 6-51).

Analyses

For each visualization, the number of participants who fixated within each ROI in the visualization at least once was calculated. The average proportion of participants who fixated on an ROI (when present) across all of the visualizations is shown in Table

1. Unsurprisingly, participants nearly always fixated on the data in the visualizations. They were also highly likely to fixate on the title, legend, and data labels, when those ROIs were present.

To determine where the participants allocated their attention in the visualizations, we calculated the proportion of each participant's fixations that fell within each ROI for each visualization. The average proportion of fixations in each ROI is also shown in Table 1. The Data ROI received the highest average proportion of fixations, but this proportion was relatively low. On average, only 27% of the participants' fixations were in the Data ROI, while the Title and Data Labels ROIs received similar proportions of fixations (25% and 26%, respectively).

Table 1. Attention to each ROI in the analysis of the MASSVIS data, including average proportions and (standard deviations).

ROI Name	Number of visualizations containing ROI	Average proportion of participants viewing ROI	Average proportion of fixations to ROI
Title	26	0.94 (0.10)	0.25 (0.10)
Data	35	0.98 (0.05)	0.27 (0.17)
Data Area	21	0.55 (0.26)	0.04 (0.03)
X-Axis	24	0.64 (0.20)	0.05 (0.03)
X-Axis Label	11	0.67 (0.14)	0.06 (0.05)
Y-Axis	24	0.70 (0.22)	0.12 (0.17)
Y-Axis Label	15	0.73 (0.25)	0.10 (0.08)
Legend	23	0.89 (0.15)	0.20 (0.11)
Data Label	15	0.88 (0.22)	0.26 (0.16)
Text	24	0.56 (0.28)	0.07 (0.10)

To test our hypothesis that participants disproportionately pay attention to text in data visualizations, the ROIs were categorized based on whether or not they contained text for each stimulus. For example, the X-Axis ROIs contained text in some visualizations but not in others. For each visualization, we then calculated the proportion of fixations that fell in ROIs containing text, the proportion of fixations to the data and data area, and the proportion of fixations that fell in other ROIs that did not contain text (including graphics, symbols, numbers, etc.). On average across all of the visualizations, 59.9% (SD = 16.1%) of the participants' fixations fell into ROIs containing text relative to 30.0% (SD = 15.6%) of fixations in the data ROIs and 10.1% (SD = 6.6%) of fixations in the other non-text ROIs.

As another measure of how participants weighted the relative importance of each ROI, we assessed how often each ROI was one of the first three ROIs visited by a participant. This was calculated as the proportion of scan paths in which the ROI was one of the first three fixated (for visualizations where that ROI was present). Note that this does not necessarily mean that one of the first three *fixations* in the trial fell in that ROI. For example, if a participant began a trial by fixating four times on the title, then

fixating three times on the data, and then fixating once on the legend, then the title, data, and legend would be counted as the first three ROIs visited on that trial. In other words, we assessed the order in which the ROIs were viewed irrespective of the number of fixations in the sequence.

The Title ROI was the most likely to be one of the first three ROIs visited. When the Title ROI was present in a visualization, it was one of the first three visited in 87.8% of the scan paths. The Data ROI was a close second at 83.5%. The proportions were much lower for the other ROIs (51.1% for Data Labels; 39.8% for Legend; 34.7% for the combination of Y-Axis and Y-Axis Labels; 17.0% for the combination of X-Axis and X-Axis Labels; 14.8% for Text). Some of the X- and Y-Axis ROIs contained words (e.g. the names of countries or months) while others were numerical (e.g. years or values). The axis ROIs were subdivided into those that contained text (other than the axis labels) and those that did not. When the X-Axis ROI contained text, it was one of the first three ROIs visited in 48.5% of the scan paths.[1] When the X-Axis ROI did not contain text, it was one of the first three ROIs visited in 12.4% of the scan paths. The difference was even more dramatic for the Y-Axis ROI, which was in the first three ROIs visited in 80.9% of the scan paths when the ROI included text, but only 13.0% of the scan paths when it did not.

To explore the data further, we looked at correlations between the number of words in an ROI and the proportion of fixations in the ROI. If a participant is spending time reading the text in a particular ROI, we would expect to see a high correlation between the number of words and the proportion of fixations. The correlations were significant for the Title ($R^2 = 0.73$, $p < 0.001$), Text ($R^2 = 0.82$, $p < 0.001$), X-Axis Label ($R^2 = 0.69$, $p < 0.02$), and Y-Axis Label ($R^2 = 0.83$, $p < 0.001$) ROIs. For the Legend and Data Label ROIs, which received relatively high proportions of fixations on average, there was not a significant correlation between the number of words and the proportion of fixations (Legend: $R^2 = 0.39$, $p = 0.07$; Data Labels: $R^2 = 0.41$, $p = 0.15$).

The axes themselves provide an interesting opportunity for investigating the effect of text on where viewers spend their time when studying a visualization. As mentioned above, some of the X- and Y-Axis ROIs contained words and others contained only numbers. When the axes contained words, there was a significant correlation between the number of words and the proportion of fixations to the axis (X-Axis: $R^2 = 0.48$, $p < 0.02$; Y-Axis: $R^2 = 0.90$, $p < 0.001$). In contrast, when the X-Axis contained only numerical values, there was no correlation between the number of numerical values and the proportion of fixations ($R^2 = 0.09$, $p = 0.68$). When the Y-Axis contained only numerical values, there was a significant *negative* correlation ($R^2 = -0.46$, $p < 0.03$).

Discussion

The results of our analyses indicate that participants disproportionately viewed regions of the visualizations that contained text in the MASSVIS study. Although the participants did spend time looking at the visualized data, the majority of their fixations

---

[1] However, there were only two visualizations in this category, with a total of 33 scan paths. The other groupings contained much higher numbers of visualizations and scan paths.

were devoted to regions containing text. For some of those regions, including the Title, Text and Axis Label ROIs, significant correlations between the number of fixations and the number of words in the ROIs indicate that participants were spending time reading the text. For other regions, namely the Legend and Data Label ROIs, there was not a significant correlation between the number of fixations and the number of words. These ROIs received relatively high proportions of fixations overall, so the absence of a correlation between the number of words and the proportion of fixations in these regions likely indicates that the participants read the text in those regions but also referred back to them more than once as they studied the visualizations.

Interestingly, the axes of graphs seemed to attract participants' attention when they contained text but not when they contained numbers. Axes containing text were much more likely to be one of the first three ROIs viewed than axes containing only numbers, and for the Y-Axis ROI there was a significant negative correlation between the number of fixations and the number of numerical values along the axis. There are several possible explanations for this pattern, but it seems plausible that numerical axes can be comprehended at a glance, making repeated fixations and revisits unnecessary.

An important point to note is that the MASSVIS eye tracking dataset was collected in the context of a memory study, which may have had a substantial influence on how participants allocated their attention. For example, they may have devoted a lot of attention to the titles of the graphs, thinking that the titles would be easier to remember than the details of the visualized data. To explore the impact of the task on patterns of attention to the visualizations, we conducted a study in which participants viewed data visualizations in a free viewing task.

3   Viewing Data Visualizations in a Free Viewing Task

When eye tracking datasets are used to assess saliency maps, the participants in the eye tracking studies are typically given a free viewing task. For example, in the widely used MIT Saliency Benchmark eye tracking datasets (http://saliency.mit.edu), participants completed a free viewing task in which they viewed each image for 5 seconds [2, 6, 17]. In this study, we used the same task and presentation duration to examine eye movement patterns on a larger set of data visualizations and a larger group of participants. Participants viewed the same subset of MASSVIS stimuli that were used in the analysis described above and an additional 27 data visualizations in the context of a larger free viewing experiment.

Method

Participants.
Thirty participants were recruited from students, faculty, and staff in the University of Illinois community (10 males; mean age = 30.53 years, SD = 13.06) and compensated $20 for their time. All participants were tested for color vision deficiencies (24 plate Ishihara Test [15]) and near vision acuity prior to completing the study. Data from an additional five participants was discarded because: they failed the colorblindness

and/or acuity tests prior to beginning the experiment (2 participants); the eye tracker failed to successfully capture their eye movements for a significant portion of the experiment (1 participant); they fell asleep for any portion of the experiment (1 participant); or there was a problem with the experimental apparatus (1 participant).

Materials.

Four blocks of images were used in this study, consisting of a total of 108 images. Each image was centered and gray padded to fill the dimensions of the screen.

Two of the blocks consisted of line drawings (30 images) and fractals (16 images) drawn from the MIT Saliency Benchmark CAT2000 dataset [2]. Those blocks are not analyzed in the present study. One block contained thirty-five data visualizations pulled from the MASSVIS dataset [3, 4]. These were the same visualizations as those analyzed in section 2. The final block contained twenty-seven data visualizations that were created specifically for this experiment (3 bar charts, 3 boxplots, 3 bubble graphs, 3 column charts, 3 line plots, 3 parallel coordinates plots, 3 pie charts, 3 scatterplots, and 3 violin plots[2]). These stimuli were selected to represent a variety of common types of data visualizations. To mirror the visualizations in the MASSVIS set, not all of the visualizations contained all of the possible ROIs and the placement of specific ROIs (such as the Legend) varied across visualizations. The newly generated visualizations also differed from the MASSVIS set because they did not contain infographics or additional text, such as text indicating the source of the data.

The order in which the four blocks of images were presented was counterbalanced across participants. Within each block, the stimuli were shown in a random order.

Procedure.

The experiment was completed in a dark room at a nominal viewing distance of 0.8 meters. Stimuli were presented on a large monitor (0.932 x 0.523 meters; 1920 x 1080 pixels) while eye movements were recorded with two Smart Eye Pro cameras. Participants first underwent the standard Smart Eye camera setup procedure and 9-point calibration.

Participants were instructed to view each image as it was presented. Each trial began with a 2-second fixation cross in the center of the screen. The fixation cross was followed by the presentation of an individual image, which was displayed on the screen for 5 seconds.

Analysis.

In the resulting dataset, fixations were defined as samples for which the velocity over the preceding 200 milliseconds (ms) was less than 15 degrees per second. The first

---

[2]Due to a programming error, 11 of these images were dropped (leaving a total of 97 images in this experiment). Because they were still of interest, the dropped images were included in a subsequent data collection. The participants in that data collection were recruited in the same manner as the initial group of participants. The group consisted of thirty participants (7 males; mean age = 29.57, stdev = 13.79). Two participants completed both data collection sessions.

fixation in each trial and any fixations with a duration less than 100 ms were dropped from the analysis. For all of the analyses described below, the visualizations pulled from the MASSVIS set and the visualizations created specifically for this experiment are pooled together. A total of 1834 scan paths were included in the analysis. There were an average of 11 fixations per scan path (range 1-19).

As in our earlier analysis, the number of participants who fixated within each ROI at least once was calculated for each visualization. The average proportion of participants who fixated on an ROI (when present) across all of the visualizations is shown in Table 2. In addition, we calculated the proportion of each participant's total fixations that fell within each ROI for each visualization. The average proportion of fixations in each ROI is also shown in Table 2. As before, the three ROIs receiving the highest proportion of fixations were the Data (37%), Title (22%) and Data Label (19%) ROIs.

Table 2. Attention to each ROI for the visualizations in the second analysis, including average proportions and (standard deviations).

ROI Name	Number of visualizations containing ROI	Average proportion of participants viewing ROI	Average proportion of fixations to ROI
Title	43	0.71 (0.21)	0.22 (0.14)
Data	62	0.91 (0.12)	0.37 (0.18)
Data Area	43	0.53 (0.23)	0.10 (0.06)
X-Axis	46	0.43 (0.18)	0.07 (0.04)
X-Axis Label	23	0.17 (0.11)	0.02 (0.02)
Y-Axis	47	0.52 (0.22)	0.10 (0.10)
Y-Axis Label	33	0.39 (0.23)	0.07 (0.07)
Legend	42	0.68 (0.21)	0.14 (0.08)
Data Label	17	0.70 (0.30)	0.19 (0.13)
Text	24	0.24 (0.29)	0.05 (0.08)

The ROIs were categorized based on whether or not they contained text for each stimulus. For each visualization, we then calculated the proportion of fixations that fell in ROIs containing text, the proportion of fixations to the data and data area, and the proportion of fixations that fell in other ROIs that did not contain text (including graphics, symbols, numbers, etc.). On average across all of the visualizations, 40.8% (SD = 19.5%) of the participants' fixations fell into ROIs containing text relative to 44.4% (SD = 18.3%) of fixations in the data ROIs and 14.8% (SD = 0.07%) of fixations in the other non-text ROIs.

We assessed how often each ROI was one of the first three ROIs fixated by a participant using the same procedure defined above. In this experiment, the Data ROI was most often one of the first three ROIs fixated. It was one of the first three ROIs fixated for 80.5% of the scan paths. The Title ROI was second at 67.5%. Once again, the proportions were lower for the other ROIs (50.8% for Data Labels; 40.5% for Legend; 40.3% for the combination of Y-Axis and Y-Axis Labels; 18.7% for the combination

of X-Axis and X-Axis Labels; 13.8% for Text). The axis ROIs were subdivided into those that contained text (other than the axis labels) and those that did not. When the X-Axis ROI contained text, it was one of the first three ROIs viewed in 22.2% of the scan paths. When the X-Axis ROI did not contain text, it was one of the first three ROIs viewed in 14.4% of the scan paths. The Y-Axis ROI was one of the first three ROIs viewed in 56.4% of the scan paths when the ROI included text and 22.0% of the scan paths when it did not.

As before, we also assessed the correlations between the number of words in an ROI and the proportion of fixations in the ROI. The correlations were significant for the Title ($R^2 = 0.90$, $p < 0.001$), Text ($R^2 = 0.81$, $p < 0.001$), X-Axis Label ($R^2 = 0.57$, $p < 0.01$), Y-Axis Label ($R^2 = 0.64$, $p < 0.001$), Legend ($R^2 = 0.39$, $p < 0.02$) and Data Label ($R^2 = 0.60$, $p < 0.02$) ROIs.

As in the first analysis, some of the X- and Y-Axis ROIs contained words and others contained only numbers. For the X-Axis, there was not a significant correlation between the number of items and the proportion of fixations for axes consisting of words ($R^2 = 0.27$, $p = 0.07$) or numbers ($R^2 = 0.03$, $p = 0.86$). For the Y-Axis, there was a significant correlation between the proportion of fixations and the number of words ($R^2 = 0.89$, $p < 0.001$), and, as in the first analysis, a significant negative correlation for numbers ($R^2 = -0.41$, $p < 0.01$).

For a more detailed assessment of how participants allocated their attention to the ROIs, plots were created to show the time course of attention to various parts of the visualizations. Every trial was divided into 313 consecutive 16 ms time windows, from trial onset until the five second trial cutoff time. For each time window, we calculated whether a fixation was made, and if so, which ROI the fixation fell into. An ROI was given a value of 1 for the time window if it received a fixation, and a 0 if it did not. Time windows of 16 ms were chosen to coincide with the sampling rate of the eye-tracker. Fixations were counted as occurring within a time bin if any part of the fixation fell in the window (i.e., even if the fixation ended or started during the time window). Only one fixation was allowed to occur in a single 16 ms time window; if multiple fixations occurred during a time window, only the first ROI visited was counted, and the fixation to the second ROI was assigned as starting in the next time window. However, given that it takes roughly 30-50ms to make a saccade, it is highly unlikely that two separate fixations would have been possible in the small time window. The first fixation of the trial was excluded, as it began with the disappearance of the fixation cross and did not represent a volitional look to any ROI.

The data plotted in Figure 1 shows the viewing patterns collapsing across all visualizations. The x-axis represents time from trial start, the y-axis represents the probability of fixating an ROI, and each line represents a different ROI. Note that the probabilities do not necessarily sum to 1 at every time point, because not every participant made a fixation during every time point (e.g., due to saccades or track loss). Overall, participants tended to look at the Title ROI early in the trial, with Title fixations peaking between 750-1000 ms after trial onset and then quickly declining. Fixations to the Data ROI surpassed looks to the Title beginning ~1500 ms after trial onset, and continued to increase throughout the duration of the trial until peaking at ~4500 ms. The next most-

fixated ROI was the Legend region, which had a numerically higher probability of fixation than the rest of the ROIs from ~750 ms after trial onset until the end of the trial. However, the low probability of fixating the other ROIs could be due the fact that not all ROIs were present in all visualizations, meaning that many ROIs had zeros for several visualizations. This plot highlights that although users made more fixations to the data ROI *overall*, this pattern was only true in the later part of the viewing period. Upon first viewing a new visualization, users tended to look at the Title first, after which they shifted their attention to other areas of the visualization.
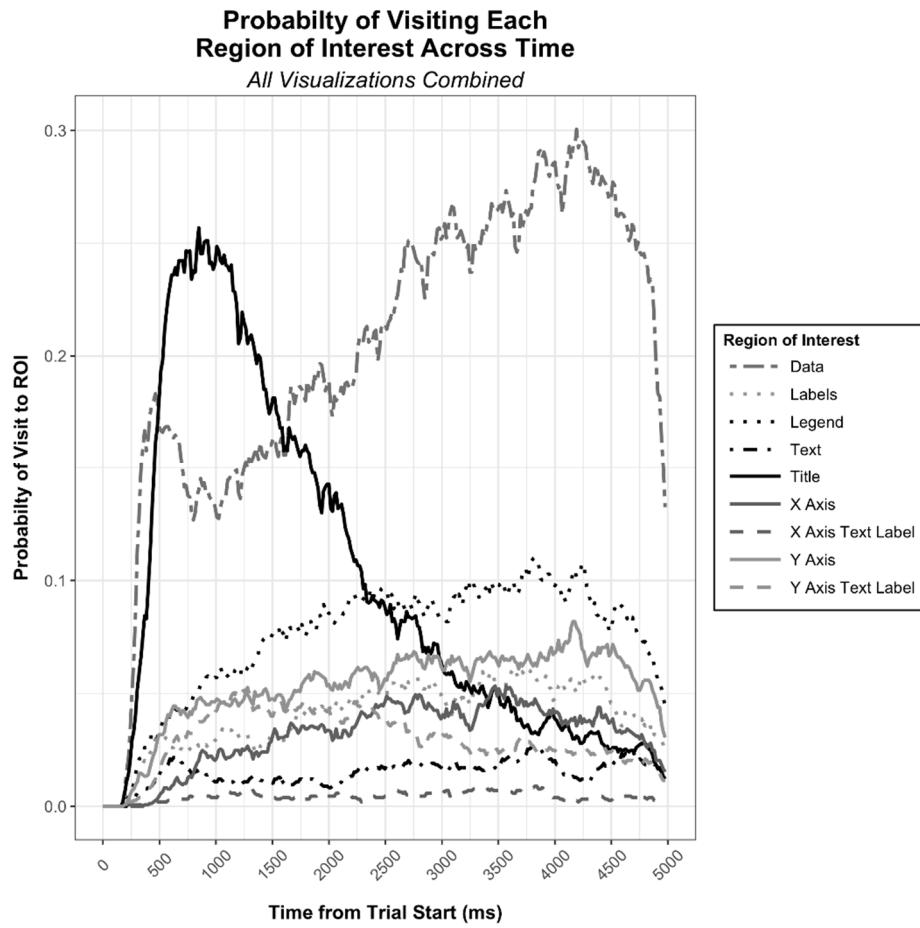
Fig. 1. Probability of fixating each ROI across time, collapsing across all visualizations.

The data plotted in Figure 2 shows viewing patterns to visualizations without text in the y-axis (top panel) versus with text in the y-axis (bottom panel). In both cases, Title fixations peaked early in the trial (~500 ms in vis without y-axis text and ~1000 ms in vis with y-axis text).
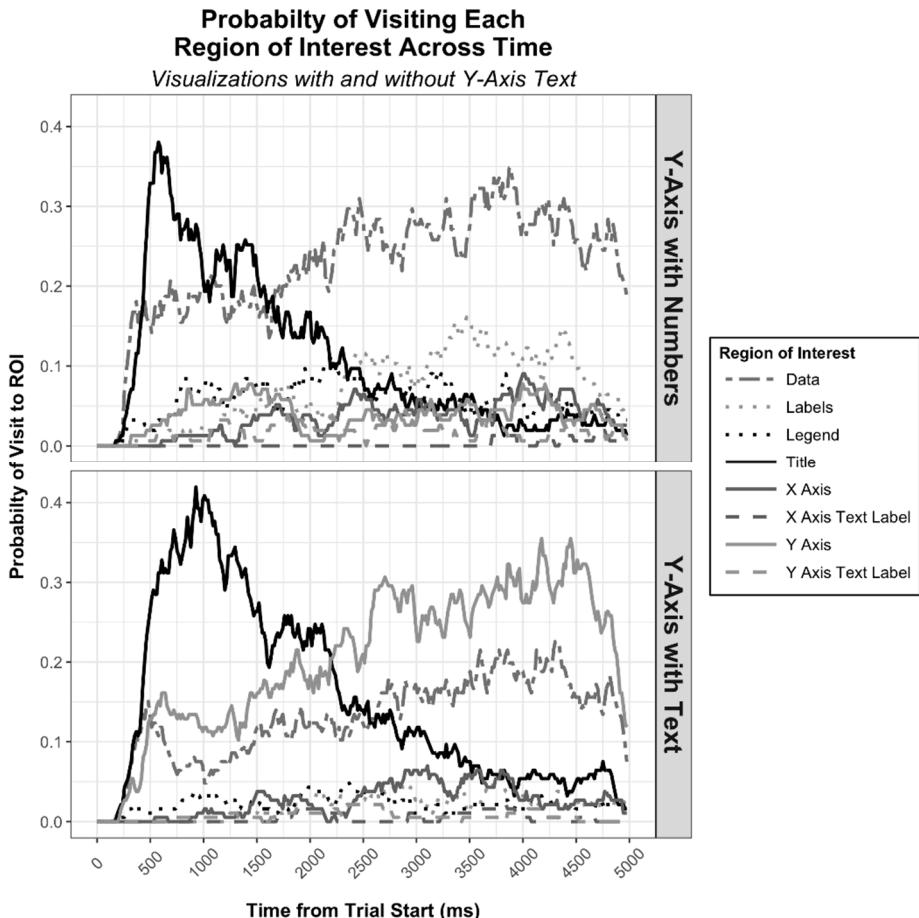
Fig. 2. Probability of fixating each ROI across time, plotted separately for visualizations without y-axis text (top panel) and with y-axis text (bottom panel).

However, striking differences are apparent in the pattern of looks to the y-axis. In visualizations *with* y-axis text, users showed clear preference for fixating the y-axis over the data area after ~500 ms into the trial, and fixations to the y-axis exceeded Title fixations after ~2250 ms. Conversely, in visualizations without y-axis text, participants made very few looks to the y-axis, and instead focused most of their fixations on the Title early in the trial, and to the Data ROI later in the trial (after ~1500 ms). There was a small preference for fixating the Labels ROI, relative to the non-Data ROIs, from ~3000-4500 ms, suggesting the need to seek out text to understand the plots when it was not present in the y-axis. This pattern clearly shows that users' viewing patterns to the y-axis were strongly influenced by the presence of text. Users made many more y-axis fixations when text was present compared to when it was not, and even made more fixations to the y-axis than to the Data when text was present, highlighting the emphasis that users place on text during visualization comprehension.

General Discussion

Overall, the results of these analyses suggest that viewers devote a great deal of attention to the text in data visualizations. For the eye tracking data collected as part of the MASSVIS study, the majority of the participants' fixations were devoted to ROIs that contained text. In the second eye tracking dataset, collected using a larger set of data visualizations and a larger group of participants along with a free view rather than memory task, the proportion of fixations devoted to text was comparable to the proportion of fixations devoted to the data.

For both datasets, it was instructive to examine the participants' attention to the axes, which contained text in some visualizations and numbers in others. The axes were one of the first three ROIs fixated more often when they contained text than when they did not. Interestingly, for the Y-Axis ROI in both datasets, there was a significant correlation between the proportion of fixations and the number of words in the ROI, and a significant negative correlation between the proportion of fixations and the number of numerical values. An analysis of the time course of fixations for the second dataset indicated that when the Y-Axis ROI contained text, it had a high probability of being visited throughout the trials, and was the most likely ROI to be viewed in the second half of the trials, after participants had turned their attention away from the title of the visualization. When the Y-Axis ROI did not contain text, it had a low probability of visits throughout the trial, with participants devoting more attention to the Data and Legend ROIs.

It is important to note that the two datasets are different in several ways. The MASSVIS data was collected in the context of a memory study where the visualizations were displayed for 10 seconds each. It consisted of visualizations that were found "in the wild." Although we selected a subset of the visualizations that represented common types of data visualizations, these images often contained descriptive titles, annotations, and text noting the source of the data. In other words, the data itself was contextualized by the text in the visualizations. In the second study, we added an additional set of visualizations that were generated in the lab rather than being found in the wild. These visualizations tended to be simpler and had less contextual information. In addition, to mirror the experimental parameters that have been used for assessing visual saliency maps, participants were given a free viewing task[3] with only 5 seconds for examining the visualizations. The simpler text and shorter viewing times in the second dataset may have driven the difference in the overall proportions of fixations to the text versus the data. However, even in the second dataset, the ROIs containing text were viewed almost as often as the data ROIs, indicating that the text still draws viewers' attention even when they have little time and the text provides relatively little information.

---

[3] It is worth noting that a free viewing task may be more representative of how people interact with visualizations in the wild than a memory task. When a person encounters a data visualization in *The Economist*, for example, they are essentially doing a free viewing task.

Our finding that viewers focused on the text elements in data visualizations is consistent with prior research. Some studies have found that users spend as much as 60-70% of viewing time reading the title, data labels and axes of simple graphs [1, 8, 18]. Users are also more likely to re-fixate text-based areas, such as the legend [3, 22, 29]. In our current analysis, we investigated a wider variety of visualization types and complexities, but the overall tendency to devote a large amount of viewing time to text-based regions remained the same.

The analyses presented here have several limitations. First, the relatively small size of the text in visualizations may necessitate more direct fixations due to the limits of visual acuity [24]. This may have an impact on overall viewing time. Second, the participants in these studies had no particular expertise with interpreting data visualizations, and their tasks did not require them to find specific information in the visualizations, or even to understand the gist of the data presented. While this approach may be realistic for understanding how people process visualizations that they encounter in daily life, such as an infographic presented in a magazine, patterns of attention are likely to be quite different in cases where a viewer is using a visualization to obtain specific information in the context of a larger task. Domain experience also plays an important role in how people attend to data visualizations. Our own prior work found large differences between professional imagery analysts and novice viewers looking at radar imagery [21], and other researchers have found that even brief instructions on how to interpret a plot can change how people allocate their attention [7]. Individual differences in information processing also play an important role. For example, dyslexic individuals spend disproportionately more time on text than typical readers [18]. None of these factors operate in isolation, and taking their combination into account can result in complex interactions between such factors as chart type, task difficulty, and the user's perceptual speed [29].

Despite these limitations, the general finding that text in data visualizations draws the viewer's attention has important implications for the development of visual saliency models that apply to visualizations. As discussed above, the ability to make predictions about where viewers will look in data visualizations could be a useful evaluation tool. To make accurate predictions, these models must take attention to text into account. In our future work, we plan to develop a new saliency model that incorporates text as a visual feature. We will test how to weight this feature relative to the other visual features that are commonly used in saliency models (color, contrast, and orientation). If successful, this approach will provide an improved tool that will allow visualization designers to evaluate their designs from the perspective of human visual processing.

4   References

1. Acarturk, C., Habel, C., Cagiltay, K., & Alacam, O.: Multi-media comprehension of language and graphics. J Eye Mov Res **1**(3):2, 1-15. (2008). doi: 10.16910/jemr.1.3.2
2. Borji, A., & Itti, L.: Cat2000: A large scale fixation dataset for boosting saliency research. CVPR 2015 workshop on "Future of Datasets" (2015), arXiv preprint arXiv:1505.03581
3. Borkin, M., Bylinskii, Z., Kim, N., C.M., B., Yeh, C., Borkin, D., Pfister, H., & Oliva, A.: Beyond memorability: Visualization recognition and recall. IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis) (2015). doi: 10.1109/TVCG.2015.2467732
4. Borkin, M., Vo, A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. What makes a visualization memorable? In IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis) (2013). doi: 10.1109/TVCG.2013.234
5. Bylinskii, Z., & Borkin, M. A.: Eye fixation metrics for large scale analysis of information visualizations. In Proceedings of ETVIS 2015, First Workshop on Eyetracking and Visualizations (2015).
6. Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., & Torralba, A.: MIT saliency benchmark. http://saliency.mit.edu/
7. Canham, M., & Hegarty, M.: Effects of knowledge and display design on comprehension of complex graphics. Learn Instr **20**, 155-166. (2010). doi: 10.1016/j.learninstruc.2009.02.014
8. Carpenter, P. A., & Shah, P.: A model of the perceptual and conceptual processes in graph comprehension. J Exp Psychol Appl **4**(2), 75–100. (1998). doi: 10.1037//1076-898x.4.2.75
9. Connor, C.E., Egeth, H.E., & Yantis, S.: Visual attention: Bottom-up versus top- down. Curr Biol **14**(19), R850–R852. (2004). doi: 10.1016/j.cub.2004.09.041
10. Fu, B., Noy, N. F., & Storey, M. A.: Eye tracking the user experience–An evaluation of ontology visualization techniques. Semant Web **8**(1) 23-41. (2017). doi: 10.3233/SW-140163
11. Goldberg, J. H., & Helfman, J. I.: 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, pp. 71-78. (2010). doi: 10.1145/2110192.2110203
12. Goldberg. J. H., & Helfman, J. I.: Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. Inf Vis **10**(3), 182-195. (2011). doi: 10.1177/1473871611406623
13. Haass, M. J., Wilson, A. T., Matzen, L. E., & Divis, K. M.: Modeling Human Comprehension of Data Visualizations. In International Conference on Virtual, Augmented and Mixed Reality, pp. 125-134. (2016). doi: 10.1007/978-3-319-39907-2_12
14. Higgins, E., Leigenger, M., & Rayner, K.: Eye movements when viewing advertisements. Front Psychol, **5**, 210. (2014). doi: doi.org/10.3389/fpsyg.2014.00210
15. Ishihara, S. (1972). Tests for Colour-Blindness; 24 Plates Edition. Tokyo: Kanehara Shuppan Co., Ltd.
16. Itti, L., & Koch, C.: Computational modelling of visual attention. Nat Rev Neurosci **2**, 194-203. (2001). doi: 10.1038/35058500

17. Judd, T., Durand, F., & Torralba, A. (2012). A benchmark of computational models of saliency to predict human fixations. [https://dspace.mit.edu/handle/1721.1/68590]
18. Kim. S., & Lombardino, L. J.: Comparing graphs and text: Effects of complexity and task. J Eye Mov Res **8**(3):2, 1-17. (2015). doi: 10.16910/jemr.8.3.2
19. Kim. S., Lombardino, L. J., Cowles, W., & Altmann, L. J.: Investigating graph comprehension in students with dyslexia: An eye-tracking study. Res Dev Disabil, **35**, 1609-1622. (2015). doi: 10.1016/j.ridd.2014.03.043
20. MATLAB Release 2015b, The MathWorks, Inc., Natick, Massachusetts, United States
21. Matzen, L. E., Haass, M. J., Tran, J. & McNamara, L. A.: Using eye tracking metrics and visual saliency maps to assess image utility. Electronic Imaging **16**, 1-8. (2016). doi: doi.org/10.2352/ISSN.2470-1173.2016.16.HVEI-127
22. Peebles, D., & Cheng. P. C.-H.: Modeling the effect of task and graphical representation on response latency in a graph reading task. Hum Factors **45**(1), 28-46. (2003). doi: 10.1518/hfes.45.1.28.27225
23. Pinto, Y., van der Leij, A., Sligte, I. G., Lamme, V. A. F., & Scholte, H. S.: Bottom-up and top-down attention are independent. J Vis **13**, 1-14 (2013). doi: 10.1167/13.3.16
24. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychol Bull **124**(3), 372-422. (1998). doi: 10.1037/0033-2909.124.3.372
25. Rayner, K.: Eye movements and attention in reading, scene perception, and visual search. Q J Exp Psychol **62**(8), 1457-1506. (2009). doi: 10.1080/17470210902816461
26. Rosenholtz, R., Dorai, A., & Freeman, R.: Do predictions of visual perception aid design? In ACM Transactions on Applied Perception (TAP), **8**(2), 12. (2011). doi: 10.1145/1870076.1870080
27. Shah, P., & Hoeffner, J.: Review of graph comprehension research: implications for instruction. Educ Psychol Rev **14**(1), 47-69. (2002). doi: 10.1023/A:1013180410169
28. Strobel, B., Sass, S., Lindner, M. A., & Köller, O.: Do graph readers prefer the graph type most suited to a given task? Insights from eye tracking. J Eye Mov Res **9**(4):4, 1-15. (2016). doi: 10.16910/jemr.9.4.4
29. Toker, D., Conati, C., Steichen, B., & Carenini, G.: Individual user characteristics and information visualization: connecting the dots through eye tracking. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 295-304. (2013). doi: 10.1145/2470654.2470696