Eye Fixation Metrics for Large Scale Evaluation and Comparison of Information Visualizations

Zoya Bylinskii, Michelle A. Borkin, Nam Wook Kim,
Hanspeter Pfister, and Aude Oliva

Abstract An observer's eye movements are often informative about how the observer interacts with and processes a visual stimulus. Here, we are specifically interested in what eye movements reveal about how the content of information visualizations is processed. Conversely, by pooling over many observers' worth of eye movements, what can we learn about the general effectiveness of different visualizations and the underlying design principles employed? The contribution of this manuscript is to consider these questions at a large data scale, with thousands of eye fixations on hundreds of diverse information visualizations. We survey existing methods and metrics for collective eye movement analysis, and consider what each can tell us about the overall effectiveness of different information visualizations and designs at this large data scale.

1 Introduction

Eye movements can provide us with clues about the elements of a visual display that people pay attention to, what they spend most time on, and how they redirect their attention between elements. The eyes can also be used as indicators of higher-level cognitive processing like memory, comprehension, and problem solving [22, 24, 33, 40, 41, 55].

Zoya Bylinskii and Aude Oliva
Computer Science and Artificial Intelligence Lab at the Massachusetts Institute of Technology, 32 Vassar St., Boston, MA. e-mail: {zoya,oliva}@mit.edu

Michelle A. Borkin
College of Computer and Information Science at Northeastern University, 360 Huntington Ave., Boston, MA. e-mail: m.borkin@neu.edu

Nam Wook Kim and Hanspeter Pfister
School of Engineering & Applied Sciences at Harvard University, 33 Oxford Street, Boston, MA. e-mail: {namwkim,pfister}@seas.harvard.edu

Eye movement analyses have been used to study the perception of natural scenes, simple artificial stimuli, webpages, user interfaces, and increasingly, information visualizations. In human-computer interaction (HCI), eye tracking has often been used for evaluating the usability of systems and studying the related question of interface design [14, 20, 30, 48]. Duchowski provides a survey of different eye tracking applications in domains ranging from industrial engineering to marketing [14].

In the visualization community, eye tracking analyses have been used to independently evaluate different visualizations (e.g., graphs [26, 27, 28, 40, 50], node-link diagrams [1], tree diagrams [9], parallel coordinates [63]) and to directly compare visualization types [7, 12, 18]. Eye-tracking has also been used to understand how a person visually perceives, explores, searches, and remembers a visualization, providing a window into the cognitive processes involved when interacting with visualizations [1, 3, 7, 12, 27, 38, 50, 51, 54].

Information visualizations are specifically designed to be parsed and understood by human observers. Visualizations can be created to help convey a specific message to a general audience, or to help data analysts extract trends and meaning from the data. As visualizations are amenable to specific tasks, observer performance on those tasks can be directly measured (e.g., ability to find a specific piece of information, to solve an analysis task, to remember the content for later retrieval, etc.). Eye movement analyses can then be used to provide possible explanations of task performance (e.g., why a task was completed quicker with one visualization design as compared to another), as complementary performance measurements that take into account human perception. Eye movements can provide a window into the cognitive processing taking place when an observer examines an information visualization.

Although different eye movement metrics have been previously reviewed within the context of different tasks [1, 18, 30, 52], in this manuscript we focus specifically on eye fixation metrics that can be used for *collective analysis* (the aggregation of data across a population of observers and visualizations) of information visualization designs. We provide a review of metrics that can be used for the *quantitative comparison* of different visualization designs in a large data setting. Unlike many previous studies, our analyses are broad, spanning a large diversity of visualization types and sources. We discuss and visualize ways in which different metrics can be used to evaluate the effectiveness of different visualization designs, and we use the MASSVIS dataset [7] to provide some specific examples. The review provided in this manuscript is intended to motivate further research into large-scale eye movement analysis for the broad comparison and evaluation of visualization designs.

2 Methods

2.1 Visualization data

We used the MASSVIS dataset of 393 labeled target visualizations[1], spanning four different **source categories**: government and world organizations, news media, infographics, and scientific publications [7]. These visualizations were manually labeled using the LabelMe system [60] and Borkin et al.'s visualization taxonomy [8] (Fig. 1a). Labels classify **visualization elements** as: data encoding, data-related components (e.g., axes, annotations, legends), textual elements (e.g., title, axis labels, paragraphs), pictograms or human recognizable objects, or graphical elements with no data encoding function. Labels can overlap in that a single region can have a number of labels (e.g., an annotation on a graph has an annotation label and a graph label). Labels are available for analyses as segmented polygons.

2.2 Eyetracking experiments

We used eye movements collected during the *encoding* experimental phase from the study by Borkin et al. [7]. During this phase, each visualization was shown to participants for 10 seconds, producing an average of 37.4 (SD: 3.2) eye fixations per visualization, or an average 623 (SD: 93) total fixations per visualization. This duration proved to be of sufficient length for a participant to read the visualization's title, axes, annotations, etc., as well as explore the data encoding, and short enough to avoid too much redundancy in fixation patterns and explorative strategies. Participants were told to remember as many details of each visualization as possible for subsequent experimental phases. During the *recognition* and *recall* phases, respectively, participants completed a memory task and were asked to write descriptions of the visualizations they remembered. We do not directly use this additional data in the present manuscript, but refer to the conclusions made from the eye movement analyses in the context of memory performances.

Eye movements of 33 participants were recorded on 393 target visualizations, with an average of 16.7 viewers (SD: 1.98) per visualization. Equipment included an SR Research EyeLink1000 desktop eye-tracker [64] with a chin-rest mount 22 inches from a 19 inch CRT monitor (1280 x 1024 pixels). For each eye fixation, available for analysis are its spatial location in pixel coordinates, duration in milliseconds, and ordering within the entire viewing episode (scanpath).

[1] Dataset available at http://massvis.mit.edu

2.3 Metrics and visualizations

Depending on the analysis being performed, different aspects of eye movement behavior can be measured including fixation locations, fixation durations, and saccades[2]. **Fixations** are discrete samples of where an eye was looking on a visual display obtained from continuous eye movement data[3] (Fig. 1b). By segmenting the visual stimulus into elements or Areas of Interest, denoted **AOI**, fixations falling on different AOIs can be separately analyzed (Fig. 1a). Consecutive fixations on a specific region or AOI can be further clustered into **gazes** (Fig. 1c).

Apart from summarizing the number and duration of fixations on a visual design or its constituent elements, the spatial and sequential aspects of a viewing episode can be used to compute additional measurements of eye movement behavior for visual design analysis. For instance, the spatial distribution of fixations can be captured by the moments of the distribution or the **coverage** (proportion of visual design fixated at a particular threshold value, Sec. 3.3). The temporal ordering (sequence) of fixations is often referred to as the **scanpath** [46] and is common for analyzing search tasks (Fig. 1d). For instance, one can consider the sequence of AOIs observers fixate while searching for a target or a specific piece of information.

Quantitative eye movement measurements used by previous visualization studies are summarized in Table 1. A review of the most common eye measurements across usability studies more generally is provided by Jacob and Karn [30]. The 5 most common metrics reported across 24 usability studies also appear in Table 1. Different metrics emphasize different aspects of eye movement behavior, which are in turn linked to different underlying cognitive processes. The number or density of fixations allocated to a visual area has been linked to its importance [30, 53]; fixation duration in a visual area has been linked to the area's information content or complexity [33]; and the transitions between fixations have been found to be related to the search behavior and expectations of the viewer [16, 45, 55]. Patterns in the fixation data of a group of observers can also be used to highlight design features or diagnose potential problems. For instance, the order of fixations has been found to be indicative of the efficiency of the arrangement of visual elements [16]. A visualization designer might be interested in ensuring that the important elements are more likely to be fixated early.

The use of different types of visualizations for highlighting properties of eye movement data have also been useful for complementing and facilitating analysis over groups of observers [1, 19, 42, 58, 65, 67, 69, 70]. A number of previous visualization studies relied mostly on such qualitative analyses (Table 1). Blascheck et al. provide a review of visualizations and visual analytics tools for eye movement data [3]. While visualizations can facilitate data exploration, inferences made from eye movement data are more meaningful when supported by quantitative metrics.

---

[2] Saccades are intervals between fixations: the motion of the eyes from one fixation point to the next. The analysis of saccades is beyond the scope of the present manuscript, for which additional metrics would be necessary [41, 52].

[3] The eye has to be recorded as "still" according to prespecified parameters [25, 61]. We use the standard thresholds set by the EyeLink Eyetracker [64].

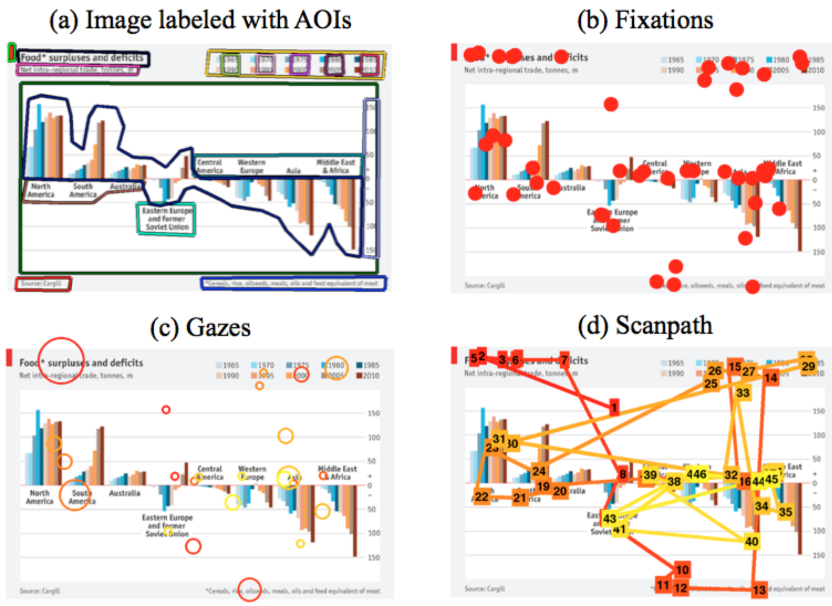

Fig. 1 We plot the fixations of a single observer for demonstration purposes, to visually depict a few key terms used throughout this manuscript. (a) The images we use are labeled with AOIs (Areas of Interest), which are elements like the title, axes, and legend. (b) Fixations are the discrete locations that an observer's eyes have landed on at some point during the viewing episode. (c) Multiple consecutive fixations that land on the same AOIs of an image can be further clustered into gazes. The size of the gaze marker is proportional to the number of fixations making up the gaze, with the marker centered at the mean of those fixation locations. (d) A scanpath is the sequence of fixations made. Here, to denote the temporal ordering, fixations are connected by lines, numerically labeled, and colored such that the earliest are in red and the latest in yellow.

For the explorative analysis of the MASSVIS eye movement data, we utilize **fixation heatmaps** due to their versatility, scalability, and interpretability. Fixation heatmaps are constructed by aggregating a set of fixations and placing a Gaussian[4] at each fixation location. The result is a continuous distribution that can be plotted on top of the image to highlight elements receiving the most attention. This simple visualization is particularly amenable to collective analysis, allowing us to visualize the fixations of any number of observers on a single image. To highlight different trends in the eye movements, we aggregate over different subsets of the data: distinct fixation durations (Fig. 2), time points during the viewing episode (Fig. 3), and observers (Fig. 5). Our coverage plots are also just thresholded fixation heatmaps (Fig. 4).

[4] Typically, the sigma of the Gaussian is chosen to be equal to 1 or 2 degrees of visual angle, to model the uncertainty in viewing location.

We note that eye movement analyses are most informative in the context of an objective task that an observer performs. In such cases, eye movements are more likely to be related to task completion itself. Furthermore, eye movement analyses can be used to complement, and provide possible explanations for, other objective performance measurements (e.g., speed or accuracy of task completion). Considered in isolation, eye movement measurements can be open to interpretation, and thus they should complement, not replace, other measurements. For example, the eye movements from the MASSVIS dataset were collected in the context of memory and recall tasks. Participants' fixations were recorded as they examined visualizations, knowing they would have to retrieve the details from memory later. In this manuscript, our focus is on the eye movement metrics themselves and how they can be used for the evaluation and comparison of information visualizations more broadly. We use the MASSVIS dataset for demonstrative examples.

Table 1 Eye movement metrics commonly reported in usability studies [30] and for evaluation and comparison of information visualizations. Different perception studies have used these metrics to make conclusions about the importance and noticeability of different visual elements, and to reason about the difficulty of the perception task and the complexity of the visual design [52]. AOI refers to an *Area of Interest*, which can be a component of a graph like the title, axis, or legend.

Quantitative measurements	**Visualization studies**	**Possible interpretations**
Summary measurements		
Total number of fixations*	[18, 40]	Efficiency of searching or engagement [13, 20, 33]
Total number of gazes	[12]	Complexity of inferential process [12]
Mean fixation duration*		Complexity or engagement [33]
AOI measurements		
Fixations on AOIs* (proportion or number)	[9, 38, 63]	Element importance or noticeability [53]
Gazes on AOIs* (proportion or number)	[12]	Element importance or noticeability [30]
Viewing time on AOIs* (proportion or total)	[12, 38, 63]	Information content, complexity, or engagement [33]
Time to first fixation on an AOI	[18, 40, 63]	Attention-getting properties [11]
Mostly qualitative analysis	[26, 27, 28, 50, 54]	Relative complexity or efficiency of different designs

* The marked metrics are the 5 most commonly-reported across a total of 24 usability studies surveyed by Jacob and Karn [30].

3 Analyses

In this section we demonstrate how the metrics listed in Table 1 can be used for collective eye movement analysis over a large dataset of visualizations and observers. We use the MASSVIS dataset for our examples. Summary fixation measurements (Sec. 3.1) can be used for a very coarse analysis of the fixation data to compare groups of visualizations, for instance by source type. Having areas of interest labeled on individual visualizations allows us to perform a finer-grained analysis (Sec. 3.2) to investigate which elements capture observer attention the earliest, the most number of times, and for the longest interval of time. The advantage of a fixed set of labels is that statistics can be aggregated over many different visualizations to discover general trends. Aside from using the common metrics from Table 1, we also show the utility of coverage (Sec. 3.3) and inter-observer consistency (Sec. 3.4) analyses to derive additional diagnostics about visualization designs.

3.1 Summary fixation measurements

To summarize fixation behavior across images and observers for a given task, eye tracking studies often consider the average number and duration of fixations required for task completion. The advantage of these coarse measurements are that they are easy to compute, independent of image content, and can be aggregated over any number of data points. These measurements are particularly meaningful when there is an objective task for an observer to complete, such as searching for a particular piece of information in a visualization. Studies can investigate whether fewer fixations are required to solve a task using one visualization design compared to another. These measurements can also be used to make inferences about observer engagement, with the caveat that there may be confounding factors such as the amount of information on a visualization, and relying on these metrics alone may not be sufficient. All the results reported below correspond to numerical values computed on 393 MASSVIS visualizations, and reported in Table 2 in the Appendix.

Total number of fixations: Aggregating over target visualizations from different source categories, the news media visualizations contained the most fixations on average, significantly more than the other visualization sources.

Total number of gazes: By aggregating fixations into gazes, we can avoid double counting fixations with different pixel coordinates on the image, but still falling within the same set of AOIs. For instance, for an observer reading a piece of text, all *consecutive* fixations falling on the text are considered part of a single gaze. Analyzing gazes, we find that the same patterns hold as with fixation counts, with the news media visualizations containing the largest number of gazes on average. This shows that the eyes moved around most between elements on the news media visualizations than on any of the other visualization sources. Is there more to look at on the news media visualizations? The number of visualization elements is actually highest for the infographics. We can use these metrics to hypothesize that observers were more engaged by the news media visualizations, but additional user studies would be needed for validation.

Mean fixation duration: The duration of individual fixations has significance in the psychology literature. For instance, shorter-duration fixations, less than about 200-250 ms, are sometimes considered involuntary (the eyes move there without a conscious decision) [21]. Fixations longer than about 300 ms are thought to be encoded in memory. Across the MASSVIS target visualizations, the mean fixation duration is longer for infographics and scientific visualizations. These visualizations contain many diagrams and other visually-engaging elements, and have been found to be the most memorable [7].

By plotting heatmaps of fixations at various durations in Fig. 2, we can see which elements of a visualization are explored for shorter or longer periods of time, and thus potentially differentially processed. Durations of fixations have been found to be related to the complexity and difficulty of the visual content and task being performed [16, 49, 55]. Thus, considering locations in a visualization receiving fixations of increased duration could be used to discover elements of the visualization that are engaging the cognitive resources of the observer.

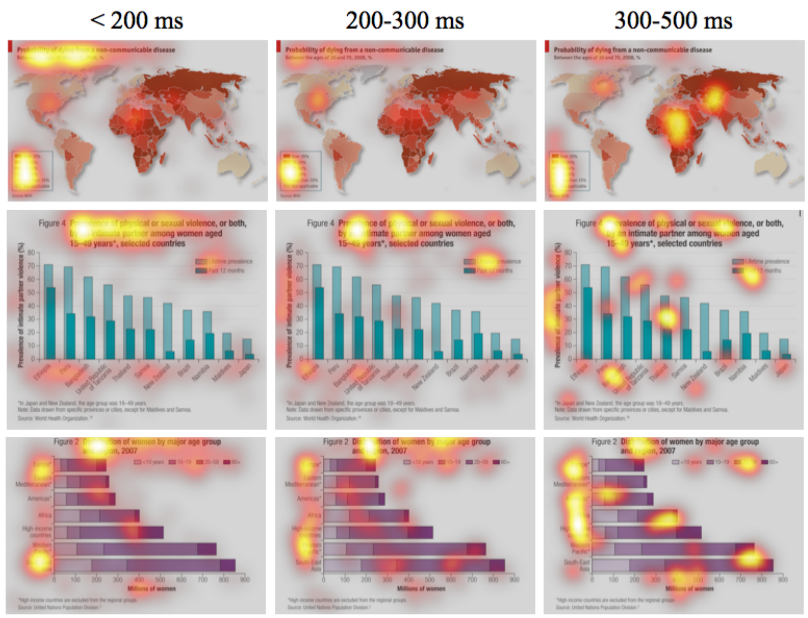

Fig. 2 Heatmaps created by selectively aggregating fixations of different durations, across all observers. Here we see that longer-duration fixations (300-500 ms) are used to explore more of the data elements. Fixation durations have been linked to the complexity and informativeness of a visual area [16, 49, 55].

3.2 AOI fixation measurements

Having labeled (pre-segmented) visualization elements allows statistics to be aggregated over observers and visualizations, to relate eye movements back to these elements, and get a finer-grained picture of observer attention. In the eye tracking literature, segmented image regions for quantifying eye movement behavior are often called Areas of Interest (AOIs) or Regions of Interest (ROIs). Note that in some cases, as in ours, the AOIs are meaningful parts of the visual content and are pre-segmented for analysis. In other cases AOIs may be defined by clustering the eye movements during post-processing[5]. All of the results reported below correspond to the plots included in Fig. 6-7 of the Appendix.

Fixations on AOIs: Fixation statistics across AOIs can be aggregated over all visualizations to make conclusions about general design principles. For example, over all 393 target visualizations of the MASSVIS dataset, the legend, table header row (i.e., top label row of a table), paragraph, and title elements receive on average the largest number of fixations. However, when aggregating over multiple instances in a visualization of each element type, we find that observers make more fixations on the paragraph and label element types, although any individual label in a visualization would receive fewer fixations than the legend.

Gazes on AOIs: Within a single gaze, paragraphs receive the most fixations. But by aggregating fixations into gazes, the header row and legend receive the most gazes. Observers return to header rows and legends most frequently, which is why they end up with the most fixations overall. These specific elements allow the information in a visualization to be clarified and integrated.

Viewing time on AOIs: The viewing time (in ms) can be a measure of the importance or information content of a visualization element [33]. We find that of the total number of time spent fixating visualizations, legends, header rows, paragraphs, and titles were fixated the longest. This corresponds to the fact that these elements received the most fixations overall, another measure of importance.

Time to first fixation on an AOI: An analysis of scanpaths can indicate which elements are fixated first and which elements are fixated multiple times during the entire viewing episode. Over all observers and visualizations, we can find the average fixation number on which each element was first fixated. Across the MASSVIS target visualizations, the elements fixated earliest are titles, objects, paragraphs, and header rows. These are textual elements from which an observer can expect to learn the most about what the visualization is conveying (important elements) and visual depictions that attract attention (noticeable elements). A complementary visualization can depict these trends. We selectively aggregated over fixations at different time points in the viewing episode, splitting the viewing time into 3 segments of 3 seconds each, and computed fixation heatmaps. As depicted in Fig. 3, titles consistently receive attention in the first 3 seconds of viewing time. Then fixations move to the paragraphs, other explanatory text, and data elements.

Overall, observers tend not to dwell on pictograms and purely-visual elements, and instead spend most of the time reading text. This supports previous findings that viewers start by visiting, and spend more time on, textual elements than pictorial elements [56]. This does not mean that observers do not look at pictograms. However, fixations on these elements do not last long: observers look at these elements, and move on. Considering a number of different fixation metrics concurrently paints a clearer picture of observer eye movement behavior.

[5] Goldberg and Helfman [18] discuss implementation choices and issues arising when working with AOIs and fixations.

Fig. 3 Viewing behavior unfolding over time is visualized by aggregating fixations during specific intervals of time. Titles are consistently fixated earliest, followed by explanatory paragraphs. The data itself is explored after much of the text.

Of all the textual elements, titles are often first to be examined, and in general, receive a lot of attention during the viewing episode. Our eye movement analyses point to the importance of these elements, while additional quantitative analyses reported in Borkin et al. confirm that titles are highly memorable elements that are often recalled by participants, and can aid or hinder comprehension of a visualization [7]. In such a way, eye movement measurements can complement additional task-specific analyses.

3.3 Coverage

Coverage, related to spatial density metrics [13, 20], is computed by aggregating the fixations of all observers, thresholding the resulting fixation heatmap at some fixed value, and measuring the amount of image area covered by fixations [69]. Coverage can be visualized (as in Fig. 4) by masking out image regions with fixation values below the threshold. Image regions that survive high thresholds are those that receive the most fixations. Applying the same threshold to different information visualizations can facilitate comparison across designs. A lower coverage value indicates that observers tend to look at a smaller portion of the visualization.

Analyzing coverage can help diagnose potential design issues. If a large part of the visualization is covered in data but fixation coverage is low, observers may have missed important components of the visualization or crucial parts of the message (Fig. 4). Across the MASSVIS target visualizations, infographic visualizations have on average more coverage than any of the other visualization sources. Although these differences are not statistically significant, a trend surfaces across 3 different threshold values. Another way to look at this trend is that among the 50 visualizations with highest coverage (at a 20% threshold), 38% are infographics, while of the 50 visualizations with lowest coverage, 38% are news media. Does this contradict the coverage finding? Both infographics and news media visualizations receive a high number of fixations, indicating high observer engagement, but the news media visualizations in the MASSVIS dataset tend to be simpler and have fewer elements. As a result, fixations on the news media visualizations are more clustered around a few components, leading to lower coverage. By considering multiple fixation metrics, a fuller story unfolds.

3.4 Inter-Observer Consistency

Inter-observer consistency (**IOC**) is used in saliency research[6] to quantify the similarity of observer fixations on an image. IOC for an image is a measure of how well the fixation heatmap of N-1 observers predicts the fixation heatmap of the remaining observer, averaging over all N observers, under some similarity metric[7]. We propose that IOC analysis can be used to determine how the design of an information visualization guides observers. High IOC implies that observers tend to have similar fixation patterns, while a low IOC corresponds to different observers examining a visualization in different ways. In the latter case, it is worth measuring if the different possible fixation patterns will lead observers to derive similar conclusions from the visualization. Will the message of the visualization be clear no matter

[6] This has also been called inter-subject consistency [68], the inter-observer (IO) model [4], and inter-observer congruency (IOC) [43].

[7] Area under receiver operating characteristic curve (AUROC or AUC) is the most commonly used similarity metric [15]. Note that IOC analysis can be extended to the ordering, instead of just the distribution, of fixations [19, 32, 43, 46].

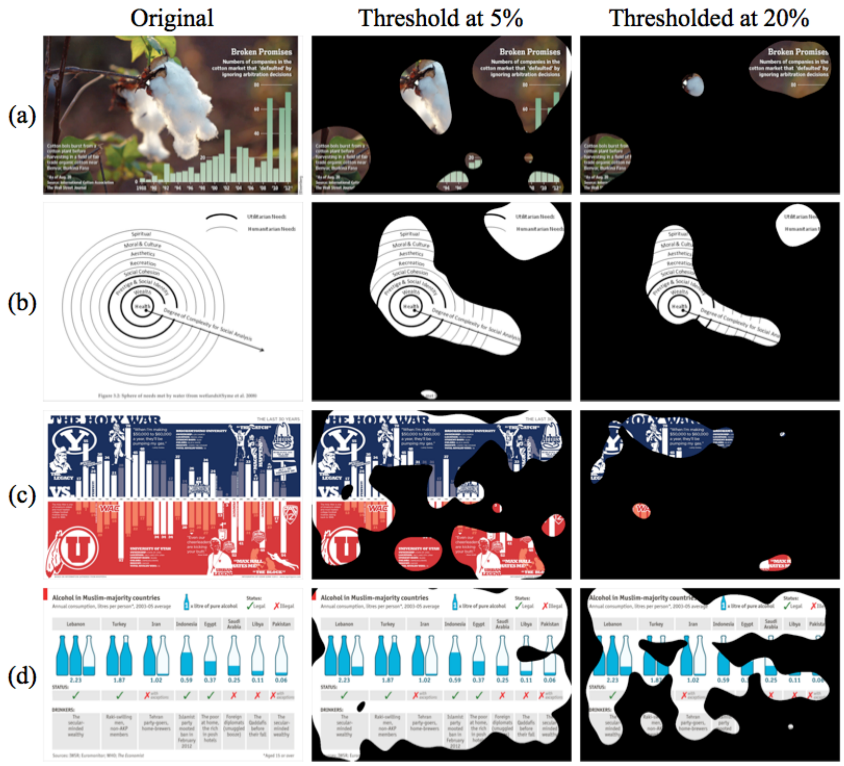

Fig. 4 Analyzing fixation coverage can help diagnose potential design issues. (a) The photographic element may have distracted observers, who paid no attention to the bar graph; (b) The title at the bottom, explaining the visualization, was missed; (c) Less crucial quotes captured more attention than explanatory text; (d) A visualization with many components and high coverage - observers were engaged, and examined the majority of the visualization. Different thresholds for plotting coverage can be used to visualize regions of an image fixated by different proportions of observers. We plot the thresholds at 5% and 20% of the maximum heatmap value.

how the visualization is examined? Did the designer of the visualization intend the visualization to be viewed in a particular way? Fig. 5 contains example fixation heatmaps for a visualization with high IOC and one with low IOC. In general, dense and crowded visualizations with a lot of information have low IOC; there is a lot to look at, and different observers choose to look at different things. Simple, well-structured visualizations (e.g., with a standard layout) direct observer attention, so different observers look at these visualizations in similar ways. For example, across the MASSVIS target visualizations, infographic visualizations have lower IOC than any of the other source categories, and news media visualizations have the highest IOC. This goes along with the coverage storage: with fewer elements to look at in a visualization, observers are more consistent about where they look.

Fig. 5 **Top row:** A visualization with high IOC. All observers have a similar fixation pattern on this visualization. This visualization tends to consistently guide the observer's attention. **Bottom row:** a visualization with low inter-observer consistency (IOC). Different observers examine the visualization in different ways but will they get the same information out of it? For ease of comparing the fixation patterns of different observers, the underlying visualizations have been gray-scaled.

4 Conclusion

In this manuscript we reviewed a number of existing eye movement metrics and considered their utility for the collective analysis of large, diverse datasets of visualizations. By aggregating statistics over observers and visualizations, these metrics can be used to quantitatively evaluate different types and designs of visualizations. We also discussed techniques for visualizing properties of fixation behavior that these metrics aim to capture[8]. Whereas we focused mostly on the distribution of eye fixations, a more thorough investigation of other properties of eye movement behavior like scanpaths and saccades are likely to provide additional insights. This manuscript contributed a discussion of broader, more large-scale comparison methods than prior visualization studies.

The need will only increase for metrics and analyses that can scale to processing data of potentially hundreds of observers on thousands of images. New methodologies are opening up opportunities of collecting user attention patterns, to approximate or replace costly eye tracker recordings, at larger scales than previously possible [31, 37, 59].

[8] Labeled visualizations, eye movement data, and code for the visualizations in this manuscript are available at http://massvis.mit.edu.

Moreover, some of the design evaluations discussed might be possible without collecting any user data at all. Many computational models have been developed over the past couple of decades to predict eye movements, specifically fixations and attention patterns on natural images[9]. In recent years, computational predictions have begun to come very close to ground truth human eye movements on photographs [10]. Models for predicting eye movements on graphic designs, webpages, and visual interfaces are also beginning to show promise [47, 62]. As computational models continue to evolve, opportunities will open up to evaluate visual designs, including information visualizations, in a fully automatic manner. For instance, O'Donovan et al. computationally predict the importance of different visual elements in graphic designs [47], Berg et al. predict the importance of elements and objects in natural images [2], and Khosla et al. predict the memorability of different image regions, automatically generating a kind of importance map per image [36]. Le Meur et al. directly predict inter-observer congruency (IOC) for images without user data [44]. Automatic predictions of image interestingness [23], style [34], aesthetics [57], and memorability [29, 35] are already possible. Such computational predictions have the potential of making their way into designer tools, to provide real-time feedback on visual designs and visualizations. Importantly, these computational predictions are all informed by studies and measurements of human perception and cognition. The results of eye movement analyses thus have the potential to make simultaneous contributions to the understanding of human cognitive and perceptual processes, visual content design principles, and better automatic design predictions in the future.

Acknowledgements This work was partly funded by awards from Google and Xerox to A.O., NSERC Postgraduate Doctoral Scholarship (PGS-D) to Z.B., NSF Graduate Research Fellowship Program and NSERC Discovery grant to M.B., and a Kwanjeong Educational Foundation grant to N.K.

References

1. G. Andrienko, N. Andrienko, M. Burch, and D. Weiskopf. Visual analytics methodology for eye movement studies. *IEEE TVCG*, 18(12):2889–2898, 2012.
2. A. C. Berg, T. L. Berg, H. Daume III, J. Dodge, A. Goyal, X. Han, A. Mensch, M. Mitchell, A. Sood, K. Stratos, et al. Understanding and predicting importance in images. In *Computer Vision and Pattern Recognition*, pages 3562–3569. IEEE, 2012.
3. T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. State-of-the-art of visualization for eye tracking data. In *Proceedings of EuroVis*, volume 2014, 2014.
4. A. Borji, D. Sihite, and L. Itti. Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE Transactions on Image Processing*, 22(1):55–69, 2012.
5. A. Borji, D. N. Sihite, and L. Itti. Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE Transactions on Image Processing*, 22(1):55–69, 2013.

[9] We suggest the following surveys: [5, 6, 17, 39, 66]

6. A. Borji, H. R. Tavakoli, D. N. Sihite, and L. Itti. Analysis of scores, datasets, and models in visual saliency prediction. In *IEEE International Conference on Computer Vision*, 2013.
7. M. Borkin*, Z. Bylinskii*, N. Kim, B. C.M., C. Yeh, D. Borkin, H. Pfister, and A. Oliva. Beyond memorability: Visualization recognition and recall. *IEEE TVCG*, 22(1):519–528, 2016.
8. M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE TVCG*, 19(12):2306–2315, 2013.
9. M. Burch, N. Konevtsova, J. Heinrich, M. Hoeferlin, and D. Weiskopf. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *IEEE TVCG*, 17(12):2440–2448, 2011.
10. Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba. MIT Saliency Benchmark. http://saliency.mit.edu/.
11. M. D. Byrne, J. R. Anderson, S. Douglass, and M. Matessa. Eye tracking the visual search of click-down menus. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 402–409. ACM, 1999.
12. P. A. Carpenter and P. Shah. A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2):75, 1998.
13. L. Cowen, L. J. Ball, and J. Delin. An eye movement analysis of web page usability. In *People and Computers XVI*, pages 317–335. Springer, 2002.
14. A. T. Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4):455–470, 2002.
15. T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
16. P. M. Fitts, R. E. Jones, and J. L. Milton. Eye movements of aircraft pilots during instrument-landing approaches. *Ergonomics: Psychological mechanisms and models in ergonomics*, 3:56, 2005.
17. S. Frintrop, E. Rome, and H. I. Christensen. Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception (TAP)*, 2010.
18. J. H. Goldberg and J. I. Helfman. Comparing information graphics: A critical look at eye tracking. In *BELIV'10*, pages 71–78. ACM, 2010.
19. J. H. Goldberg and J. I. Helfman. Scanpath clustering and aggregation. In *Proceedings of the 2010 symposium on eye-tracking research & applications*, pages 227–234. ACM, 2010.
20. J. H. Goldberg and X. P. Kotval. Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24(6):631–645, 1999.
21. W. Graf and H. Krueger. Ergonomic evaluation of user-interfaces by means of eye-movement data. In *Proceedings of the third international conference on human-computer interaction*, pages 659–665. Elsevier Science Inc., 1989.
22. E. R. Grant and M. J. Spivey. Eye movements and problem solving guiding attention guides thought. *Psychological Science*, 14(5):462–466, 2003.
23. M. Gygli, H. Grabner, H. Riemenschneider, F. Nater, and L. Gool. The interestingness of images. In *International Conference on Computer Vision*, pages 1633–1640, 2013.
24. M. Hayhoe. Advances in relating eye movements and cognition. *Infancy*, 6(2):267–274, 2004.
25. K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press, 2011.
26. W. Huang. Using eye tracking to investigate graph layout effects. In *APVIS '07*, pages 97–100, Feb 2007.
27. W. Huang and P. Eades. How people read graphs. In *Proceedings of the 2005 Asia-Pacific symposium on Information Visualisation*, volume 45, pages 51–58, 2005.
28. W. Huang, P. Eades, and S.-H. Hong. A graph reading behavior: Geodesic-path tendency. In *PacificVis '09*, pages 137–144, April 2009.
29. P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 145–152. IEEE, 2011.
30. R. Jacob and K. S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, 2(3):4, 2003.

31. M. Jiang, S. Huang, J. Duan, and Q. Zhao. Salicon: Saliency in context. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
32. S. Josephson and M. E. Holmes. Visual attention to repeated internet images: testing the scanpath theory on the world wide web. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 43–49. ACM, 2002.
33. M. A. Just and P. A. Carpenter. Eye fixations and cognitive processes. *Cognitive psychology*, 8(4):441–480, 1976.
34. S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller. Recognizing image style. *arXiv preprint arXiv:1311.3715*, 2013.
35. A. Khosla, A. S. Raju, A. Torralba, and A. Oliva. Understanding and predicting image memorability at a large scale. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2390–2398, 2015.
36. A. Khosla, J. Xiao, A. Torralba, and A. Oliva. Memorability of image regions. In *NIPS*, pages 305–313, 2012.
37. N. W. Kim, Z. Bylinskii, M. A. Borkin, A. Oliva, K. Z. Gajos, and H. Pfister. A crowdsourced alternative to eye-tracking for visualization understanding. In *CHI'15 Extended Abstracts*, pages 1349–1354. ACM, 2015.
38. S.-H. Kim, Z. Dong, H. Xian, B. Upatising, and J. S. Yi. Does an eye tracker tell the truth about visualizations?: Findings while investigating visualizations for decision making. *IEEE TVCG*, 18(12):2421–2430, 2012.
39. A. Kimura, R. Yonetani, and T. Hirayama. Computational models of human visual attention and their implementations: A Survey. *IEICE TRANS INF. and SYST.*, 2013.
40. C. Körner. Eye movements reveal distinct search and reasoning processes in comprehension of complex graphs. *Applied Cognitive Psychology*, 25(6):893–905, 2011.
41. E. Kowler. The role of visual and cognitive processes in the control of eye movement. *Reviews of oculomotor research*, 4:1–70, 1989.
42. C. Lankford. Gazetracker: software designed to facilitate eye movement analysis. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 51–55. ACM, 2000.
43. O. Le Meur and T. Baccino. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavioral Research Methods*, 45(1):251–266, 2013.
44. O. Le Meur, T. Baccino, and A. Roumy. Prediction of the inter-observer visual congruency (iovc) and application to image ranking. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 373–382. ACM, 2011.
45. G. R. Loftus and N. H. Mackworth. Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human perception and performance*, 4(4):565, 1978.
46. D. Noton and L. Stark. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision research*, 11(9):929, 1971.
47. P. O'Donovan, A. Agarwala, and A. Hertzmann. Learning Layouts for Single-Page Graphic Designs. *IEEE TVCG*, 20(8):1200–1213, 2014.
48. B. Pan, H. A. Hembrooke, G. K. Gay, L. A. Granka, M. K. Feusner, and J. K. Newman. The determinants of web page viewing behavior: an eye-tracking study. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 147–154. ACM, 2004.
49. J. B. Pelz, R. Canosa, and J. Babcock. Extended tasks elicit complex eye movement patterns. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 37–43. ACM, 2000.
50. M. Pohl, M. Schmitt, and S. Diehl. Comparing the readability of graph layouts using eyetracking and task-oriented analysis. In *Computational Aesthetics in Graphics, Visualization and Imaging*, pages 49–56, 2009.
51. M. Pomplun, H. Ritter, and B. Velichkovsky. Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception*, 25:931–948, 1996.
52. A. Poole and L. J. Ball. Eye tracking in HCI and usability research. *Encyclopedia of human computer interaction*, 1:211–219, 2006.

53. A. Poole, L. J. Ball, and P. Phillips. In search of salience: A response-time and eye-movement analysis of bookmark recognition. In *People and Computers XVIII*, pages 363–378. Springer, 2004.
54. M. Raschke, T. Blascheck, M. Richter, T. Agapkin, and T. Ertl. Visual analysis of perceptual and cognitive processes. In *International Journal of Computer Vision*, 2014.
55. K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372, 1998.
56. K. Rayner, C. M. Rotello, A. J. Stewart, J. Keir, and S. A. Duffy. Integrating text and pictorial information: eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, 7(3):219, 2001.
57. K. Reinecke, T. Yeh, L. Miratrix, R. Mardiko, Y. Zhao, J. Liu, and K. Z. Gajos. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2049–2058. ACM, 2013.
58. G. Ristovski, M. Hunter, B. Olk, and L. Linsen. EyeC: Coordinated views for interactive visual exploration of eye-tracking data. In *17th International Conference on Information Visualisation*, pages 239–248, 2013.
59. D. Rudoy, D. B. Goldman, E. Shechtman, and L. Zelnik-Manor. Crowdsourcing gaze data collection. *arXiv preprint arXiv:1204.3367*, 2012.
60. B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1-3):157–173, 2008.
61. D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.
62. C. Shen and Q. Zhao. Webpage saliency. In *European Conference on Computer Vision*, pages 33–46. Springer, 2014.
63. H. Siirtola, T. Laivo, T. Heimonen, and K.-J. Raiha. Visual perception of parallel coordinate visualizations. In *International Conference on Information Visualisation*, pages 3–9, July 2009.
64. SR Research Ltd. *EyeLink Data Viewer User's Manual, Version 1.8.402 (2008)*.
65. H. Y. Tsang, M. Tory, and C. Swindells. eSeeTrack: Visualizing Sequential Fixation Patterns. *IEEE TVCG*, 16(6):953–962, 2010.
66. J. K. Tsotsos and A. Rothenstein. Computational models of visual attention. *Scholarpedia*, 6, 2011.
67. J. M. West, A. R. Haake, E. P. Rozanski, and K. S. Karn. eyePatterns: software for identifying patterns and similarities across fixation sequences. In *Proceedings of the 2006 symposium on Eye tracking research & applications*, pages 149–154. ACM, 2006.
68. N. Wilming, T. Betz, T. C. Kietzmann, and P. König. Measures and limits of models of fixation selection. *PLoS ONE*, 6(9), 2011.
69. D. S. Wooding. Eye movements of large populations: Deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods, Instruments, & Computers*, 34(4):518–528, 2002.
70. M. M. A. Wu and T. Munzner. SEQIT: Visualizing Sequences of Interest in Eye Tracking Data. In *IEEE TVCG*, 2015.

Appendix

Table 2 Eye fixations on a total of 393 MASSVIS visualizations are analyzed and discussed in Sec. 3.1. Measurements are first aggregated across all observers per visualization, to obtain an average value for each visualization. Then statistics are computed over all the visualizations per source category for a comparison across the categories: infographic, news media, scientific, and government. The t-statistic is reported for each pairwise t-test in the final column (Bonferonni-corrected for multiple comparisons). Colored markers indicate which pairwise comparison each t-statistic corresponds to. Tests with $p < 0.05$ are marked with (*) and those corresponding to $p < 0.01$ are marked with (**). Note, for clarity, not every pairwise comparison is reported. The highest value for each measurement is highlighted in gray.

Summary measurements	**Infographics** (92 vis)	**News** (122 vis)	**Science** (79 vis)	**Government** (100 vis)	**Pairwise comparisons**
Number of elements	M = 38.7 🔵🟢 (SD = 32.9)	M = 19.7 🔵 (SD = 14.0)	M = 18.4 🟠🟢 (SD = 10.8)	M = 11.9 🟠 (SD = 7.4)	🔵 t(212)=5.73** 🟠 t(177)=4.79** 🟢 t(169)=5.23**
Total number of fixations	M = 37.3 🔵🟢 (SD = 3.1)	M = 39.0 🔵🔴 (SD = 2.6)	M = 34.6 🟠🟢 (SD = 3.1)	M = 37.7 🟠🔴 (SD = 2.5)	🔵 t(212)=4.39** 🟠 t(177)=7.56** 🟢 t(169)=5.77** 🔴 t(220)=3.80**
Total number of gazes	M = 33.7 🟢 (SD = 3.7)	M = 33.9 🔵🔴 (SD = 3.5)	M = 32.3 🔵 (SD = 3.7)	M = 31.9 🟢🔴 (SD = 3.3)	🔵 t(199)=3.20* 🟢 t(190)=3.65* 🔴 t(220)=4.56**
Mean fixation duration	M = 238.6 🔵🟢 (SD = 26.5)	M = 218.6 🔵 (SD = 16.1)	M = 245.3 🟠 (SD = 26.9)	M = 221.3 🟠🟢 (SD = 15.9)	🔵 t(212)=6.82** 🟠 t(177)=7.44** 🟢 t(190)=5.55**
Coverage (5%)	M = 0.59 (SD = 0.15)	M = 0.55 (SD = 0.12)	M = 0.57 (SD = 0.14)	M = 0.57 (SD = 0.12)	
Coverage (10%)	M = 0.43 (SD = 0.15)	M = 0.39 (SD = 0.12)	M = 0.41 (SD = 0.13)	M = 0.42 (SD = 0.12)	
Coverage (20%)	M = 0.26 (SD = 0.12)	M = 0.23 (SD = 0.09)	M = 0.23 (SD = 0.09)	M = 0.25 (SD = 0.09)	
IOC (20%)	M = 0.81 (SD = 0.05)	M = 0.83 (SD = 0.03)	M = 0.82 (SD = 0.04)	M = 0.82 (SD = 0.03)	

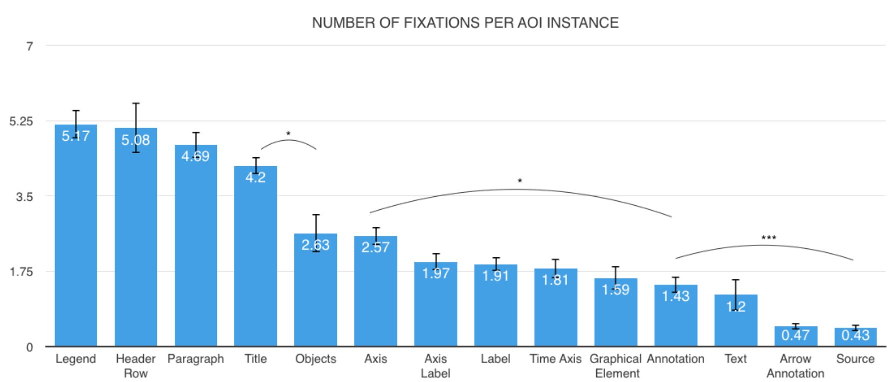

Fig. 6 These plots correspond to the results reported in Sec. 3.2. Note that Bonferonni-corrected pairwise t-tests with $p < 0.05$ are marked with (*), $p < 0.01$ with (**), and $p < 0.001$ with (***). For clarity, not all pairwise comparisons are plotted.

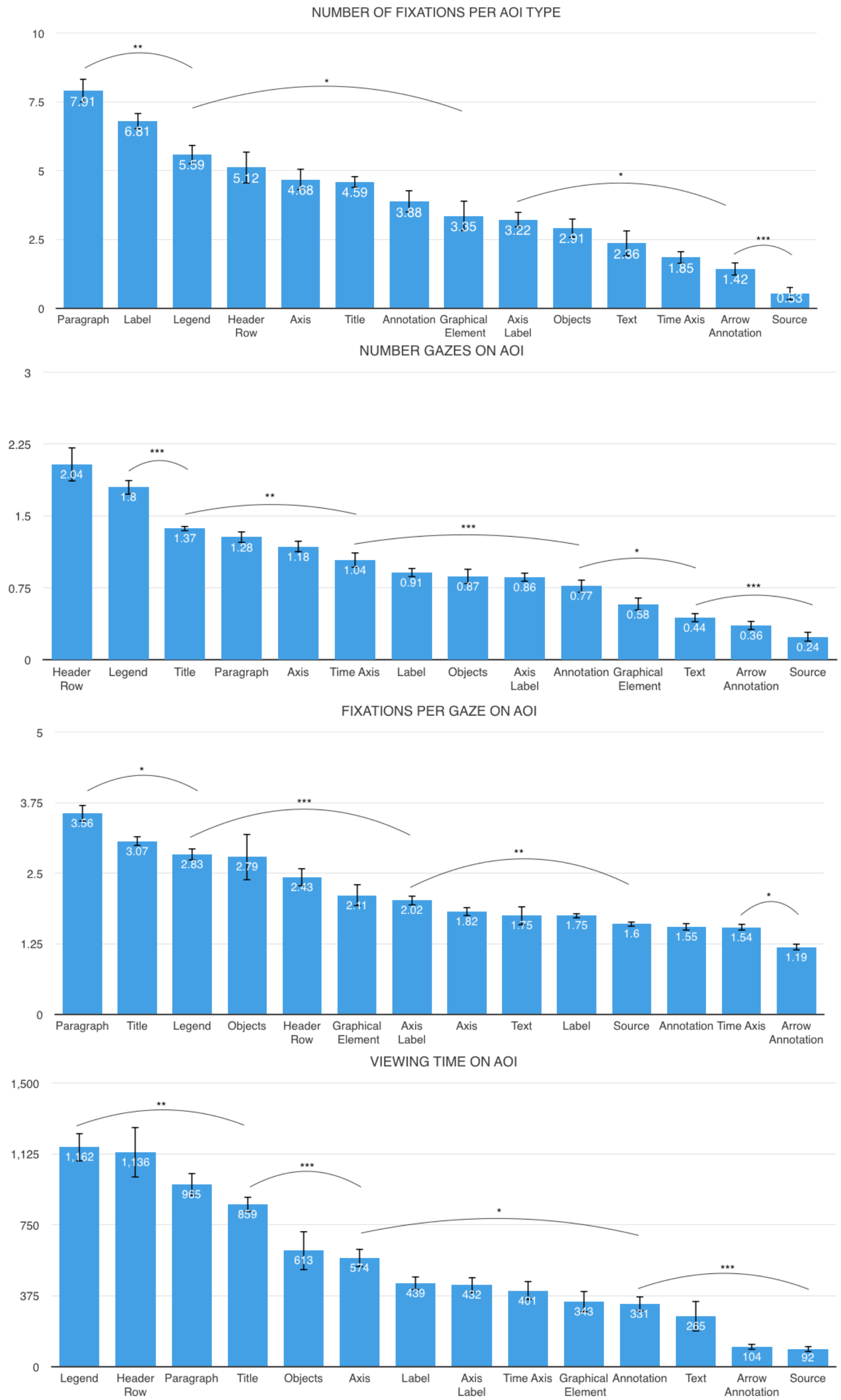

Fig. 7 These plots correspond to the results reported in Sec. 3.2. Note that Bonferonni-corrected pairwise t-tests with $p < 0.05$ are marked with (*), $p < 0.01$ with (**), and $p < 0.001$ with (***). For clarity, not all pairwise comparisons are plotted.