Visual Informatics 4 (2020) 41–50

Contents lists available at ScienceDirect

Visual Informatics

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

Analyzing gaze behavior for text-embellished narrative visualizations under different task scenarios

Chris Bryan [a,*], Aditi Mishra [a], Hidekazu Shidara [b], Kwan-Liu Ma [b]

[a] School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, United States
[b] Department of Computer Science, University of California, Davis, United States

ARTICLE INFO

Article history:
Received 25 June 2020
Received in revised form 3 August 2020
Accepted 6 August 2020
Available online 24 August 2020

Keywords:
Narrative visualization
Eye tracking
Perception
User study

ABSTRACT

We conduct an eye tracking study to investigate perception text-embellished narrative visualizations under different task conditions. Study stimuli are data visualizations embellished with text-based elements: annotations, captions, labels, and descriptive text. We consider three common viewing tasks that occur when these types of graphics are viewed: (1) simple observation, (2) active search to answer a query, and (3) information memorization for later recall. The overarching goal is to understand, at a perceptual level, if and how task affects how these visualizations are interacted with. By analyzing collected gaze data and conducting advanced semantic scanpath analysis, we find, at a high level, diverse patterns of gaze behavior: simple observation and information memorization lead to similar optical viewing strategies, while active search significantly diverges, both in regards to which areas of the visualization are focused upon and how often embellishments are interacted with. We discuss study outcomes in the context of embellishing visualizations with text for various usage scenarios.

1. Introduction

Narrative visualization is a form of data storytelling where statistical graphics are used to complement (or even replace parts of) a story (Segel and Heer, 2010). Narrative visualizations can be categorized – based on their afforded interactivity – into *static* and *interactive* charts. For static charts and infographics, *narrative rhetoric* can be attained not only by framing the chart's design choices (Hullman and Diakopoulos, 2011), but also by embellishing the base visualization with additional cues: glyphs, pictures, labels, annotations, descriptive captions, highlights, etc. Static narrative visualizations are commonly found in written articles – e.g., in a newspaper – or as a part of media presentations – e.g., on PowerPoint slides – to provide contextualizing insights that support the author's narrative. For interactive charts and dashboards, the user manipulates the visualization(s) to explore the underlying dataset, via the use of tooltips, filters, linked views, animations, transitions, pop-ups, etc. In today's digital world, narrative visualizations have widespread adoption in computational journalism and online media (Cohen et al., 2011).

In this paper, we study a specific subset of narrative visualizations: *static charts embellished with text-based narrative cues*. There is good motivation for this. Narrative visualizations are commonly employed for public consumption, such as in newspapers, media sites, government reports, and websites (Cohen et al., 2011), and should be designed to ensure effective and desired comprehension. The information that such visualizations provide is in part likely a combination of the visual features (or properties) of the chart (Matzen et al., 2018), the task (or expectations) of the user (Healey and Enns, 2012), and the type of information provided by the visualization (see Shah and Hoeffner, 2002 for a review).

To help better understand the visual importance of features present in static narrative visualizations when a person is performing various tasks, we conduct a controlled eye tracking user study. Eye tracking is one way to understand visual saliency of a scene, as gaze is typically thought to be closely linked to attention (Rayner, 2009a). In particular, we collect and analyze the eye movement data of participants ($n = 16$) to evaluate how gaze behavior and visual scanning strategies differ based on the participant's current task. We formalize this exploration around the following two-part research question:

Research Question: (1) When viewing static narrative visualizations embellished with text elements, are gaze behavior and visual scanning strategies affected by the viewer's current task? (2) If so, how?

We consider three tasks in our study, each reflecting common reasons that a person might look at a visualization:

* Corresponding author.
E-mail address: cbryan16@asu.edu (C. Bryan).
URL: https://chrisbryan.github.io (C. Bryan).

https://doi.org/10.1016/j.visinf.2020.08.001

Observation: The chart is viewed simply because it is present, perhaps as part of a news story shown on TV, as a supplementary figure in an article, or as part of a slide in a presentation.

Search: The chart is explicitly looked at to learn specific information, perhaps as a response to an information-seeking query.

Recall: The chart is observed for a period of time. At a later point in time, when the chart is no longer present, information shown on the chart needs to be recalled from memory.

Our assumption about the first part of the research question is a trivial, "*yes, task affects visual perception*". In part, we base this on prior research into how humans translate sight into memories. *Guided search* (Wolfe, 1994) is the process of spatially focusing attention to perform complex cognitive operations – e.g. looking up and validating a target as part of a search task – using input from earlier attentional processes. In simple observation scenarios that lack any specified target, guided search will either not occur or it will be self-driven by the viewer. In contrast, recall tasks where the stimulus is no longer available require longer information storage using either short- or long-term memory (Atkinson and Shiffrin, 1968).

The second (and more interesting) part of our research question looks at *how* task affects gaze and scanning behavior. For example, we assume that when no specific target is provided (as in the observation and recall tasks) there will be a disproportionate focus on the text elements in the stimuli.

To conduct our study, we use an eye tracker that records the eye movements of participants while they look at displayed stimuli while performing the three tasks. Since attention is generally thought to be linked to gaze, eye tracking can provide a window into cognitive processes as a person takes in a scene (Holmqvist et al., 2011). Prior work has shown that gaze can be affected both by the stimuli (e.g. Netzel et al., 2017a; Goldberg and Helfman, 2011) and the task that is being performed (Yarbus, 1967). To make the study grounded, we use real-world published visualizations as stimuli, which are carefully selected and balanced to minimize the potential for confounding variables (Borkin et al., 2013, 2016).

We analyze the collected study data at various levels of semantic complexity. First, we assess traditional "point-based" eye tracking metrics: fixation durations and saccade distances. For advanced analysis, we conduct an extensive manual area of interest (AOI) tagging that accounts for subject viewing semantics. For 16 subjects, this results in 960 scanpaths composed of over 47,000 AOI tags. We analyze both AOI transitions and aggregate scanpaths to discern variations in viewing strategies between the tasks.

At a high level, the study results validate the first part of our research question: yes, that task does impact gaze behavior of text-embellished narrative visualizations. For the second part, the study results indicate task-based gaze differences exist both at the perceptual level (point-based gaze metrics) and for higher-level viewing strategies (AOI- and scanpath-based data). Notably, we see a large divergence in gaze strategies when comparing the concrete task (search) to the two open-ended tasks (observation and recall).

We conclude by discussing the contributions of this work: (1) a robust anonymized eye tracking dataset with both point-based and semantic/annotated gaze data (over 47,000 manually tagged AOIs),[1] (2) an empirical validation that task affects gaze behavior, which reinforces several findings from previous and related studies, (3) new insights how the importance of text as a "focus first" feature on gaze behavior for narrative visualizations, which can inform in creating task- and temporally-aware design guidelines, quality metrics, and saliency maps.

[1] Supplemental materials are hosted at https://github.com/chrisbryan/StudyData_AnalyzingGazeBehaviorByTask.

2. Related work

2.1. Graphical perception and visual attention

Graphical perception considers how people perceive and interpret the mark and channel encodings in charts, graphs, and other visualization techniques (Cleveland and McGill, 1984). Healey and Enns provide an excellent overview for how attention and memory affect perception with regards to data visualization and graphics (Healey and Enns, 2012). Visual attention is a complex process that combines low-level sensory and perceptual processes with cognitive considerations. These directly affect what we actually "see" when looking at a scene, based not only on where we look but also on what is in our minds. For example, being able to interpret the salient features or regions within a chart is important for overall comprehension (Hoffman and Singh, 1997).

It has long been known that perception can be affected by task. Neisser's classic study – where a woman holding an umbrella walks through a basketball game – demonstrates that inattentional blindness can occur due to concentration on a separate target (Neisser, 1979). When observers were asked to count basketball passes, only a small minority even noticed the woman in the scene. When the task was simple observation, 100% of participants noticed the woman.

2.2. Narrative visualization

In addition to Segel and Heer's formal categorization of narrative visualizations (Segel and Heer, 2010), the importance of visualization to data storytelling had been recognized across many application fields, including InfoVis (Gershon and Page, 2001), SciVis (Ma et al., 2012), news and media (Cohen et al., 2011), and business and industry (Knaflic, 2015). Moere and Leuven argue that visualizations used for communication should have a focus on *aesthetics* in addition to soundness and utility (Moere and Purchase, 2011). In practice, this commonly means nicely designed charts and graphics that are styled with rhetorical cues and flourishes to promote a desired interpretation (Hullman and Diakopoulos, 2011).

Graphical embellishments such as annotations and other design cues (sometimes derogatorily referenced as "chart junk") are considered an effective way to "point" to regions or features of interest on visualizations (Segel and Heer, 2010). In cognitive psychology and educational domains, annotated figures have been shown to improve active engagement, germane processing, and creativity, while leading to superior learning outcomes (Sedig and Parsons, 2013; Mayer et al., 1995, 2005). Recent research in InfoVis has demonstrated several benefits of embellished static visualizations, including enhanced recall, memorability, and comprehension (Bateman et al., 2010; Borgo et al., 2012; Borkin et al., 2013, 2016). However, a careful design balance must be struck, as too much "junk" hinders not only memorability but other tasks like visual search (Borgo et al., 2012). That said, "*a memorable visualization is often an effective one*" for later recognition (Borkin et al., 2016) and appending interesting visual cues a common approach for creating memorability (Borkin et al., 2013).

2.3. Eye tracking

To quantify foveal focus for a given scene, eye trackers record how a person's eyes move and fixate on objects (Holmqvist et al., 2011). While eye tracking is not a perfect proxy for perception and attention, they are generally considered closely linked (Rayner, 2009b). There are two primary types of point-based gaze data: *Fixations* are locations where the eye briefly remains approximately still. *Saccades* are optical "jumps" between successive fixation points. The trajectory of alternating fixations and saccades on a stimulus is called a *scanpath*. Meaningful regions in a stimulus can be organized into *areas of interest* (AOIs). Visits to and transitions between AOIs provide insight into higher-level viewing behavior. For example, a scanpath can be interpreted as a sequence of categorical AOI tags (each fixation pertaining to one AOI), enabling analysis using event sequence and language processing techniques like n-gram analysis (we use this technique in Section 6).

Using eye tracking, different types of visualizations have been shown to elicit different gaze behavior. Goldberg and Helfman compared radial and linear versions of several chart types, showing different fixation patterns for the same lookup task (Goldberg and Helfman, 2011). They found subjects performed a three-stage cognitive process based on the order in which they looked at chart elements. More broadly, it has been shown (going as far back as Yarbus' work in the 1960s) that varying task can affect how people observe a stimulus (Yarbus, 1967).

Even for the same task, visual scanning strategies can vary. Studying maps, Netzel et al. found that different map variants promoted different viewing strategies for a search task (Netzel et al., 2017a). In a separate study on metro charts, Netzel et al. (2017b) clustered scanning strategies for a path-following task—finding different gaze behavior depending on how subjects solved a line-tracing problem.

Eye tracking also sheds light on how narrative visualizations are perceived. Acartürk (2012) investigated fixation distributions for a search task on annotated line charts, finding that fixations on annotations come at the expense of focus on other parts of the chart. In Bateman et al.'s "chart junk" study (Bateman et al., 2010), the overall percentage of fixations for various chart AOIs were provided to demonstrate aggregate attention to various types of chart elements. Using the MASSVIS dataset, Borkin et al. conducted several studies on the memorability and recall of narrative visualizations (Borkin et al., 2013, 2016). (We use their dataset as a candidate pool for our study stimuli, see Section 4.) Matzen et al. performed temporal analysis of fixation percentages on embellished visualizations (Matzen et al., 2017), similar to parts of our analysis in Section 6. However, their study design only looks at aggregate and temporal fixation distributions (similar to our hypotheses H4–H6), whereas we investigate gaze behavior over multiple semantic levels of complexity.

3. Hypotheses

To investigate the research question posed in Section 1, we list a set of hypotheses. As previously described, we consider three tasks (observation, search, recall). In the full study design (see Section 5), we test over four types of visualizations (bar chart, line chart, map, point-based chart), meaning there are two independent variables in the study. Specific data points collected (i.e., the dependent variables) include task performance (the search and recall tasks ask subjects to answer a question for each trial), point-based gaze data, AOI visits, and scanpaths.

The four hypotheses consider each of these dependent variables individually and are written to determine *if* task has an effect on gaze behavior for that data point.

H1 Task performance will be better for the search task compared to the recall task.
H2 Point-based gaze data will primarily vary based on the task.
H3 Chart features, specifically text-based narrative embellishments, will be focused upon at different frequencies and at different times for each task.
H4 Aggregate viewing behaviors will vary based on task.

By analyzing the data about each hypothesis in detail, we can explore *how* gaze behavior changes based on the task. Below, we briefly outline assumptions for how we believe task will affect gaze behavior for each hypothesis.

3.1. Subject performance by task [H1]

As the focus of the study is gaze behavior, H1 is given primarily as a sanity check to ensure that in tasks that difficulty is appropriately reflected in the search and recall tasks. When the stimulus is present (search task), the answer can simply be looked up, so performance should be higher compared to the recall task which uses memory to answer the question. If performance is similar between these tasks, it could indicate a study confound due to unbalanced questions.

3.2. Point-based gaze data [H2]

Average fixation durations and average saccade lengths are the two most common metrics for point-based gaze data. Fixation durations are sometimes considered as a proxy for cognitive processing (Holmqvist et al., 2011). Higher values can indicate that a subject is spending more time dwelling on a stimulus feature, possibly due to visual complexity or scene novelty, while lower values can result from stress and/or frenetic scanning behavior. Longer saccade lengths can indicate exploratory or searching behavior; shorter ones either a longer focus within small regions or short jumps across the stimuli (Holmqvist et al., 2011).

Our assumption is study participants will peruse the charts at different speeds (varying fixation durations) and with different jump patterns (varying saccade lengths). Reading text generally results in short saccades (Rayner, 1998), which we believe will be a disproportionate focus during the observation task.

3.3. Focusing on chart features [H3]

To assess H3, we first consider that chart elements (labeled as AOIs) can be categorized either as "base" chart features (required elements of the visualization, such as axes and marks) or as embellishments. (Table 1 shows the specific categorization of chart elements.) Our assumption is that optical focus on these items will vary based on task. For example, since observation is an undirected task, participants will overly focus on the embellishments (such as reading a chart's caption, which is a cognitively easier process than interpreting abstract marks and channels). For the search task, subjects will likely skim over the embellishments if they are irrelevant, leading to less focus on these features.

3.4. Aggregate viewing behaviors [H4]

The last hypothesis deals with overall viewing behavior when looking at a chart. By analyzing aggregate scanpaths of participants when viewing a stimulus, we can investigate how gaze behavior for the overall scene varies.

Table 1
Classified chart elements (AOIs) and action tags. Every fixation during the study is classified into an AOI under one of the four AOI groups: general scanpath AOIs, SCEs, NTEs, NPEs, or a Q&A elements. Some AOIs can also have action tags prepends to them, see Section 5.5. Note that Z and E AOIs are omitted from analysis of H4–H6 in Section 6.3.

AOI	Description
General scanpath AOIs	
Z	Start — The starting fixation on the scanpath trajectory.
E	End — The ending fixation on the scanpath trajectory.
N	Nothing — An area of the page that could not be associated with any AOIs.
Standard Chart Elements (SCEs)	
D	Data Mark — A point, line, or area mark.
B	Background — The chart's background area.
K	Key — The key or legend.
X	X-Axis — The x-axis.
XL	X-Axis Label — The text label denote the x-axis' values.
Y	Y-Axis — The y-axis.
YL	Y-Axis Label — The text label denoting the y-axis' values.
Narrative Text Elements (NTEs)	
T	Title — The chart's title.
C	Caption — The text caption or subtitle.
S	Source — A text label denoting the data source or publishing information.
DL	Data Label — A text label referencing a data mark's value or timestep.
A	Annotation — An overlaid text annotation, text box, or descriptive sentence that contextualizes the chart.
Narrative Pictorial Elements (NPEs)	
H	Highlight — A shape or glyph that points to or highlights a set of data points or an area on the chart.
P	Pictures — A non-text embellishment (picture, drawing, etc.); can also act as a background element.
Q&A elements (Search Task Only)	
Q	Question — The question text that is asked to the subject.
QA	Answers — The answers to the question (a set of input boxes).
Action tags	
J	Jump — Directing foveal attention to a new area, starting a new, unconnected sequence of fixations.
R	Reference — Directing attention to a chart element explicitly based on the prior chart element.

4. Stimuli overview

Before running the study, we must first select an appropriate set of visualizations to use as stimuli. For this, we turn to the MASSVIS Dataset.[2] A total of 66 visualizations were selected from this dataset: 20 charts for each task plus 6 charts for training.

Selected charts are only used for one task, and no two charts are exactly the same. This can lead to potential study confounds as no two charts are exactly the same: there might be slight variations in styling, data density, and composition, text content, etc. To account for this, we carefully select stimuli in triplets: three similar visualizations were chosen together – based on several design and stylistic constraints – and one each assigned to the observation, search, and recall tasks.

4.1. Selecting stimuli from the MASSVIS dataset

The MASSVIS Dataset is a large collection of static visualizations scraped from several public online sources: government reports, infographic blogs, design sites, news and media websites, and scientific journals (Borkin et al., 2013, 2016). It was originally published as a resource for researchers to gain deeper insight into how the elements of visualization affect memorability, recall, and comprehension. With over 5000 total image files, it represents an extensive cross-sampling of current "in the wild" design trends and patterns. A broad range of visualization types, styles, and complexities is represented.

4.1.1. General stimuli constraints

We use the "targets393" subset of the MASSVIS Dataset as a candidate pool for stimuli. This subset contains 393 visualizations that meet specific design constraints: they are single-panel, stand-alone charts that have been rigorously labeled according to a number of encoding and metadata properties.[3] Significantly, the targets393 subset has been used in prior eye tracking studies and verified to contain appropriate aspect ratios conducive to eye tracking (Borkin et al., 2016).

In selecting candidate charts from the targets393 subset, the following constraints were initially applied:

Visualization Type. A dozen visualization types are contained in the targets393 subset. Of these, we restrict potential stimuli to four of the most common types: bar chart, line chart, map, and point-based charts. These charts make up approximately 60% of the targets393 subset (233/393 total charts). Each of our three study tasks has five visualizations of each type (for 20 total charts per task).

Text-based narrative elements. Only visualizations containing text-based embellishments were considered as potential stimuli (see NTEs in Table 1). Generally, these types of embellishments are manually appended by the chart designer to provide interpretation and framing (Hullman and Diakopoulos, 2011).

Legibility. We discarded potential stimuli that the targets393 metadata notes that some stimuli were labeled as hard-to-read in prior studies, primarily because the image files were pixelated and contained blurry text or data marks.

4.1.2. Selecting triplets of stimuli for each task

In addition to the above constraints, when selecting stimuli triplets we also evaluated similarity by balancing the following factors:

Visual Density. This is the density of the visualized data marks with respect to the overall image. We did not want stimuli to be too dense as this normally corresponded to very complex charts.

Data-Ink Ratio. The amount of data to non-data elements in the chart. Like visual density, we balanced between overly dense charts and those containing only a few sparse elements.

Distinct Colors. Black-and-white visualizations were discarded. For color visualizations, the number of distinct colors was usually between 2–6.

Number of text-based embellishments. The count and make-up of text-based embellishments for each triplet was balanced.

[2] http://massvis.mit.edu/.

[3] Metadata and labeling information for the targets393 subset can be found at the following URL: https://github.com/massvis/dataset.

Overall, charts had an average of 15.35 narrative elements, though it is important to note that variance was sometimes high between different triplet sets. For example, while all charts had at most one title element, maps usually contained several labels identifying individual countries, states, cities, geographic features, etc. Bar charts regularly label each bar mark (sometimes with the numerical values redundantly encoded). Such variance highlights the importance of selecting stimuli as triplets, as it evenly distributes individual variances across charts among each task.

Total Word Count. We summed all words in the visualization, including from base chart elements (keys and axes, see SCEs in Table 1). Overall, charts had an average of 85.85 words. Again, though variance between individual charts was sometimes high, chart triplets were balanced to have a similar number of words.

Visual Complexity. Visual complexity is loosely defined as the amount of detail in the chart (Borgo et al., 2012). This is partly a function of aforementioned factors (density, data-ink ratio, word count), but also depends on the chart's perceived degree of structure, data variety, and organization. Visual search can be highly dependent on the visual complexity of a chart (Reppa et al., 2008). To mitigate this, charts with similar visual complexity were selected as triplets.

Pictorial Embellishments. Narrative visualizations can also contain non-text embellishments (see NPEs in Table 1). These include arrows, circles, lines, and other glyph shapes, which are used to highlight data—for example, connecting a text label to a data point. Embellishments can also include human-recognizable objects, such as drawings or pictures overlaid on the chart or set as a background.

Rendering Style. Since the MASSVIS Dataset aggregates from several sources, there is a wide range of design stylings and aspect ratios in its charts. For example, government charts tend to be more minimalist and traditional in design, with clean borders and serif-text font. Infographic charts from blogs generally have artistic flourishes, such as utilizing multiple font styles and adding pictorial embellishments.

An example stimuli triplet is shown in Fig. 1. It is important to note that these factors are not considered control variables, since it is extremely difficult (if not impossible) to rigorously mitigate so many variables in a single study. Despite this, each factor was reviewed and balanced when selecting and dividing charts among the three tasks.

The supplemental materials contain a list of statistics for countable variable; these were used as quantitative reference points when balancing charts. By maximizing the similarities between stimuli triplets, we minimize potential confounding effects due to variation in chart presentation. Such a balancing approach is similar to prior studies (Borgo et al., 2012; Matzen et al., 2017) which also use real-world, narrative-based datasets for stimuli and must comprehensively balance several subtle-but-important factors.

4.2. Recall and search questions

To make the search and recall tasks tractable, we needed a mechanism that motivated subjects to earnestly perform each task. To do this, we formatted the search and recall tasks as "answer the question" exercises. Subjects were given a multiple choice question about data shown on a chart with four possible answers. Each stimulus was assigned one question, meaning 40 total questions were generated. Questions for each stimulus ask the subject to either (1) identify a displayed data value or extremum (2) understand a trend, pattern, or theme shown in the chart, both of which are common visualization tasks. The supplemental materials show an example of how questions in the search and recall tasks were rendered to participants during the study.

For each set of stimuli in the search and recall tasks, 10 visualizations were randomly assigned "data value" questions and 10 were assigned "trend" questions. Questions were written such that they could be answered by looking at the data visualization only. That is, we did not ask questions that required reading any embellishments, though in a small number of instances – 2 times in search, 7 times in recall – a text embellishment redundantly answered or provided a hint to the answer, though this did not affect the gaze behavior of participants in the study.

Questions for each chart also varied in relative difficulty, so as not to promote uniform review and search strategies. Questions were worded so that, as best possible, a subject could not use his or her prior knowledge to select the answer. Fortunately, many of the charts selected as stimuli visualized niche data (as shown in Fig. 1), making it highly unlikely that participants could use prior knowledge. In the study interface, questions were displayed using four clickable answer buttons.

5. User study

As mentioned previously, the study has two independent variables: task and visualization type. The design is within-subject; an outline is shown in Fig. 2. Task order is randomized and the same stimuli are used for each task, but trial order within each task is randomized (mitigating potential learning effects). With 16 users and 20 stimuli for each of the three tasks, the study contains $16 \times 20 \times 3 = 960$ total trials.

5.1. Protocol

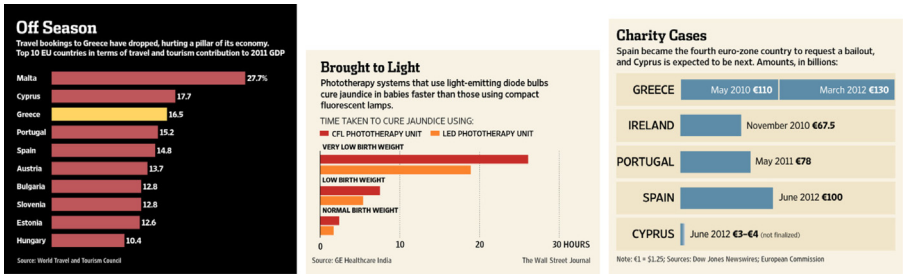
A participant begins the study by entering demographic information. Submitting these advances to an intro page which gives instructions for the first task. For example, the observational task reads as follows: *Each visualization will be shown for 10 s. Your task is simply to look at the visualization. Nothing else is required!* Progressing to the next page begins a set of 22 repetitions, as outlined in Fig. 2. Specific procedures for each task are as follows:

Observation. The stimulus is displayed for 10 s and then a break screen is shown. Pressing the space bar proceeds to the next stimulus.

Recall. The stimulus is shown for 10 s, then a gray screen is shown for 5 s, and then a multiple choice question about the just-seen stimulus is shown with four answers. Selecting an answer proceeds to a break screen, whereupon pressing the space bar proceeds to the next stimulus.

Search. The question and answers are shown for 10 s with a blank space present where the stimulus would normally be. After 10 s, the stimulus appears in the blank space. At this point, selecting an answer proceeds to a break screen, where pressing the spacebar to the next stimulus.

The first two repetitions for each task are considered training and not included as results. During training, the subject is allowed to question the session proctor if they are confused. The break screen allows subjects to rest, relax, and adjust position between trials. A subject can remain on this screen for as long as desired. When the 20 trials for a task are completed, the study redirects to an intro page for the next task. Upon completing the three tasks, a finish page states the study is finished and reports the number of questions the subject has correctly answered during search and recall tasks.



(a) Observation stimulus (b) Recall stimulus (c) Search stimulus

Fig. 1. An example chart triplet used in the study. Three charts are selected based on similarity according to several factors described in Section 4.1.2, including design styling, aspect ratio, text embellishments, and amount of information content.

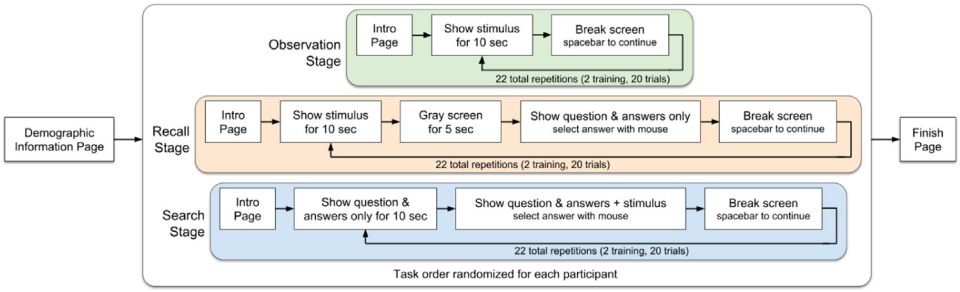


Fig. 2. The study design. After performing an eye tracker calibration, each subject begins by filling out demographic information and then completes the three task stages in random order. For each task, 22 total repetitions are performed: 2 training, then 20 trials. Stimuli order within each task stage is randomized.

5.2. Pilot study

Before beginning the main study, we ran a short pilot study with three participants. Each completed the primary study as designed, but was instructed to verbalize any issues or confusion that occurred during the process. This allowed us to verify several aspects of the design: tasks were clearly explained, task stages flowed as intended, the study duration did not cause undue fatigue by being overly long or tedious, stimuli were appropriate, etc. Based on feedback, the study's overall length and design were maintained. No stimuli were deemed inappropriate and replaced, but the phrasing of three questions was changed to be more explicit.

5.3. Testing setup and hardware

The study was conducted in a campus research laboratory, a quiet, office-like setting with normal artificial lighting conditions. Subjects sat at a desk in front of a 19-inch Dell monitor with a screen resolution of 1980 × 1080 pixels, running Microsoft Internet Explorer 11 in full-screen mode. Stimuli larger than 1100 × 900 pixels were downsized to fit these constraints (while maintaining aspect ratio).

A Tobii X2-60 eye tracker was mounted at the base of the monitor. Subject eye height was approximately one-third from the top of the monitor (though this depended on subject height) at a distance of approximately 60–65 cm from the eye tracker (the recommended sweet spot according to Tobii). The X2-60 eye tracker records gaze positions for each eye with a sampling rate of 60 Hz. Tobii Pro Studio software processed the recorded eye tracking data. Before beginning the study, each participant performed a 9-point eye calibration. To convert raw gaze data into fixations and saccades, the I-VT filter in Tobii Studio was used.

5.4. Participants

The 16 study participants were recruited from the University of California, Davis (13 males, 3 females, age $M = 26.5$, $SD = 5.3$). 13 participants had a background in computer science, and one each a background in economics, design, and biochemistry. Based on a 5-point Likert scale, subjects reported moderate prior familiarity with data storytelling ($M = 2.4$, $SD = 0.9$) and high proficiency in reading English text ($M = 4.6$, $SD = 0.8$). Participants took an average of 29.7 min ($SD = 4.4$) to complete the study, timed from loading the demographics page to reaching the finish page.

All subjects self-reported good vision with no forms of color-blindness. Five wore eye glasses, two wore contacts, one previously had corrective eye surgery, and eight required no vision correction. Eye tracker accuracy can vary from subject to subject, being negatively affected by the presence of glasses, the shape of a subject's face, distance and angle to the eye tracker, etc. Therefore, gaze sample quality was recorded during the study and reviewed afterward to assess its validity. All subjects recorded above a collection threshold of 75% of gaze samples (45+ samples per second). Therefore, no subjects needed to be discarded.

5.5. Creating AOIs and scanpaths

To evaluate H3 and H4, individual fixations must be labeled into AOIs according to the chart elements that subjects were looking at during each fixation. It is important to note that eye tracking has inherent limitations: accuracy and precision are never 100% and peripheral vision is not considered. While it is highly likely that foveal vision is the primary information-gathering region of vision (reading relies on the ability to fixate on words directly with the fovea Drieghe, 2011), peripheral vision plays an important role in helping us determine where to look next.

Despite this, turning fixations into AOIs and chaining them together into scanpaths is a popular and effective technique for understanding high-level gaze behavior for a scene. We create

a scanpath for each study trail, using an extensive manual AOI-tagging method based on a 2016 study by Netzel et al. (2016). In contrast to polygon-based approaches that automatically assign fixations to AOIs based on shapes drawn on the stimulus (the approach used in Tobii Studio), manual AOI-tagging overcomes the following challenges: (1) Text elements on a chart are normally quite small and may be placed on top of other AOIs (e.g. city/state labels on a map), leading to ambiguity. (2) The sub-scanpath around the current fixation can provide additional semantic information about the subject's gaze. For example, if a subject looks at a data mark and then looks at the text label annotating that mark, it is reasonable to infer the text AOI is being viewed in reference to the data mark. (3) Since fixation accuracy and precision may be slightly off, automatic labeling can result in mislabeling. If a fixation is recorded as just outside an AOI's border, it will be incorrectly classified in an automatic scheme.

The manual annotation of AOIs resulted in the labeling of 47,778 fixations from the 960 study trials. Table 1 lists all AOIs to which a fixation can be assigned, as well as their aggregate "AOI groups" (NTE, SCE, NPE, Q&A). Like Netzel et al. (2016), we use the previous and subsequent fixations in the scanpath to label each fixation according to (1) the chart element being viewed, and (2) if the user is reasonably performing one of two transitional actions: a *jump* (J) or *reference* (R). These two action tags denote specific eye actions that can happen when looking at a stimulus. When a transitional action occurs, a fixation receives two tags, one denoting the action and one denoting the AOI; e.g. "J DL" indicates a user is *jumping* to a *data label*. As Netzel et al. (2017a) note, since this approach indirectly considers nearby saccadic information, it "*mimics to a certain degree the peripheral information that led to an action*".

We include this fixation-to-AOI labeling dataset in supplemental materials.

6. Results

We organize the results according to hypotheses H1–H4. Where appropriate, we use a threshold of $\alpha = 0.05$ to assess significance difference between conditions. In these cases, a Shapiro–Wilk test was first applied to verify that data values were normally distributed.

6.1. Analyzing subject performance by task [H1]

For the search and recall tasks, participant performance was measured as the number of correctly answered questions. Within each task, a paired sample t-test indicates question type (data value vs. trend) does not have an effect on performance ($p > 0.05$). As expected, between tasks performance was higher in the search task ($M = 16.25/20$, $SD = 1.53$) compared to the recall task ($M = 8.5/20$, $SD = 1.86$). A paired sample t-test indicates that task has a significant effect on subject performance: $t(15) = 16.812$, $p < 0.01$. Therefore, **H1 is supported.**

6.2. Analyzing point-based gaze data [H2]

To analyze point-based gaze data, we consider fixation durations and saccade lengths. Fig. 3 plots these values by task and visualization type. We conduct to analyses on this point-based gaze data:

(1) As an initial assessment, we perform a pair of two-way repeated measures ANOVAs using average saccade length and average fixation duration as the dependent variables and task as the independent variable. The first ANOVA indicates that task does not have a statistically significant effect on average fixation duration: ($F_{2,30} = 1.034$, $p = 0.368$, $\eta^2 = 0.064$). The second ANOVA indicates that task has a significant effect on average saccade length: ($F_{2,30} = 63.94$, $p < 0.01$, $\eta^2 = 0.81$). Bonferroni post hoc tests reveal a statistically significant difference between saccade lengths among all three tasks ($p < 0.5$).

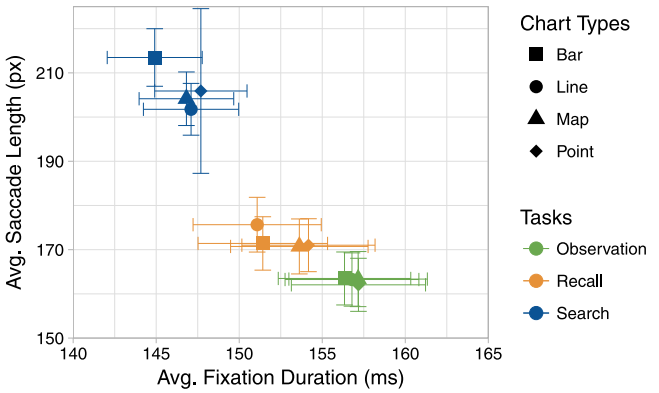


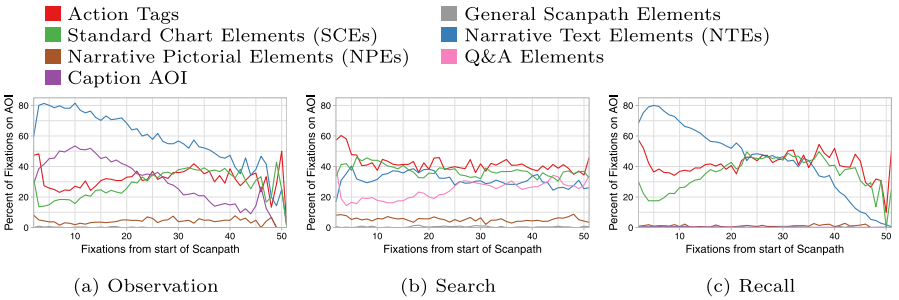
Fig. 3. Average fixation durations and averaged saccade lengths plotted by chart type and task. Error bars display 95% confidence intervals.

(2) Next, we analyze the relationship between task and chart type to see if chart type has a large effect on point-based gaze data, regardless of task. We perform a pair of two-way repeated measures ANOVAs using average fixation duration and average saccade length as the dependent variables and task and visualization type as the independent variables. Each ANOVA indicates there is a significant task × visualization type interaction, both for fixation durations ($F_{6,90} = 2.862$, $p < 0.05$, $\eta^2 = 0.16$) and saccade lengths ($F_{6,90} = 7.562$, $p < 0.01$, $\eta^2 = 0.3$). However, the effect sizes for fixation durations ($\eta^2 = 0.16$) are very small, indicating that, though an effect exists, it is unsubstantial and can therefore be ignored (Cohen, 1992). For saccade lengths, the effect size ($\eta^2 = 0.3$) is considered medium (Cohen, 1992). For post hoc analysis, we run a within-subjects contrasts test on saccade lengths. Contrasts on this interaction term indicate that when the difference in saccade lengths between the observation and recall tasks was compared to maps and points chart types, there was a significant difference ($p < 0.05$). However, no other interaction effects were observed. This can be verified by examining Fig. 3, which shows there is no clear pattern or ordering of saccade length values within each task. (For example, bar chart has the highest saccade length in the search task, line chart is the highest for recall).

In sum, the latter analysis (2) on task × visualization type indicates the interaction for fixation durations is insignificant, and for saccade lengths does not show a clear ordering, suggesting that visualization type is not a driving factor in differences for point-based gaze data. Paired with the initial "task-only" analysis, which shows average fixation durations are not affected by task, but average saccade lengths are, **H2 is partially supported**.

6.3. Analyzing focusing on chart features [H3]

To understand the chart features that participants focus on, we analyze labeled AOIs (created via the manual process in Section 5.5). Specifically, Fig. 3 lists three AOI groups: NTE, SCE, and Q&QA (search task only). We first compare the overall normalized distribution of AOI group visits against each other for each task using paired sample t-tests. For observation, AOIs in the NTE group were viewed at a higher rate ($M = 0.45, SD = 0.02$) than SCE AOIs ($0.21, SD = 0.08$): $t(15) = 8.7, p < 0.01$. For search, AOIs in the SCE group were viewed at a higher rate ($M = 0.27, SD = 0.03$) than both NTE ($M = 0.21, SD = 0.03$) and Q&QA



(a) Observation (b) Search (c) Recall

Fig. 4. For each task, the percentages of fixations for each AOI group and the percentage of fixations that contain an action tag for the first 50 fixations of the scanpath. In (a), the caption AOI is additionally rendered to demonstrate that this tag makes up the majority of NTE fixations during the observation task.

AOIs ($M = 0.18, SD = 0.03$): respectively, $t(15) = 4.5, p < 0.01$ and $t(15) = 6.625, p < 0.01$. For recall, AOIs in the NTE group were viewed at a higher rate ($M = 0.38, SD = 0.05$) than SCE AOIs ($M = 0.24, SD = 0.008$): $t(15) = 7.38, p < 0.01$). In other words, each task shows significant differences in the frequencies at which chart features are focused upon.

To better understand the dynamical nature of how participants focused on charts, we next analyze how participants fixated upon (i.e., looked at) AOIs over time. Fig. 4 plots the fixation percentages for AOI groups over time. Interestingly, temporal AOI fixations show a similar trend to the point-based gaze data shown in Fig. 3, in that the observation and recall tasks are similar as compared to the recall task.

To get a quantifiable sense of how AOI distributions change over time, for each task we bin fixations in groups of 5 along the scanpath (i.e. fixations 0–5, 6–10, 11–15, etc.). Within each bin, we aggregate fixations into AOI groups and run paired sample t-tests between AOI groups (ignoring the NPE group, since it has minimal focus). For observation, the tests indicate that until the "25-30" fixation bin, the NTE group has a significantly higher distribution ($p < 0.05$, $p > 0.05$ for subsequent bins). For search, while some bins show NTE has a higher distribution than SCE, this is not always the case (for example, our results show that SCE > NTE in bins 5–10, 20–25, 30–35, and 40–45), indicating no consistent pattern between focus on NTE and SCE AOIs. For recall, the t-tests indicate that until the "20-25" fixation bin, the NTE group has a significantly higher distribution ($p < 0.05$, $p > 0.05$ for subsequent bins).

As the observation and recall tasks displayed strikingly similar behavior up to the "20–25" bin, we compared these two tasks over this portion of the scanpath. For each fixation bin 0–5, 6–10, 11–15, 16–20, and 20–25, we perform a repeated measures one way ANOVA, where task is the independent variable and the normalized number of NTE tags of each subject in each task is the dependent variable. The ANOVAs indicate that task has a statistically significant effect on the NTE distribution across tasks for each bin ($p < 0.05$) up until the 20–25 bin. Post hoc Bonferroni tests indicate that the NTEs in observation have a higher distribution of AOI visits during this span compared to the search and recall tasks.

To sum, the observation and recall tasks display a similar pattern for how subjects focused on chart features: NTE AOIs (i.e., text-based embellishments) dominate the early part of the viewing experience, though this is more pronounced in observation. In contrast, the search task had no clear pattern for viewing NTE and SCE AOIs. Therefore, **H3 is supported**.

6.4. Aggregate viewing behaviors [H4]

Finally, we analyze whether aggregate viewing behavior will vary based on task. We do this by considering the entire scanpath for each trial in our study.

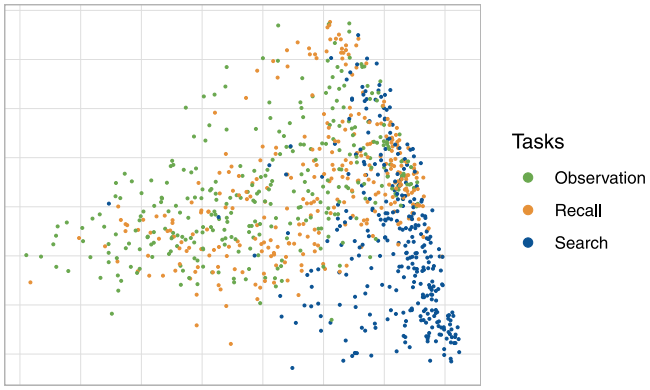


Fig. 5. Scanpaths for the 960 study trials are colored by task and visualized using multidimensional scaling based on similarity. Search scanpaths cluster at lower right, while observation and recall scanpaths are more widely distributed.

To begin, we first plot the 960 trials by their scanpath pairwise distances. We calculate the distances between each pair of scanpaths using the Needleman–Wunsch algorithm (Needleman and Wunsch, 1970). This algorithm originated in bioinformatics as a way to assign cost in aligning protein sequences, but works for any categorical sequence (i.e. string) dataset. (Note that, to allow fair comparison, we removed the Q and QA AOIs for scanpaths from the search task.) The algorithm calculates the distance between two sequences by penalizing the cost to match items at each position along the sequences. We use the following penalty values when calculating the matching cost: {match: 0, insertion/delete: 2, substitution: 2}.

Using the calculated distance matrix, the 960 scanpaths are plotted in Fig. 5 using multidimensional scaling (Borg and Groenen, 2003). Each scanpath is rendered as a single circle and colored by task. Search scanpaths (blue circles) cluster towards the lower right of the plot, while observation and recall show a wider distribution over the rest of the figure.

To analyze how close the clusters from each other, we determine both the between- and within-cluster distances. The between-cluster distances between the observation and recall cluster centroids are quite low: $d = 0.182$, while the distance to the search cluster's centroid is 0.65 and 0.53, respectively. In other words, the search cluster is much farther from the observation and recall clusters than they are from each other.

The within-cluster distances for observation, recall, and search are 0.41, 0.375 and 0.285, respectively. This indicates that scanpaths for the observation task have the most divergence, while scanpaths for the search task are most similar to one another (i.e. the search cluster is the most homogeneous). This reflects what is shown in Fig. 5; the observation and recall tasks show relatively similar behavior (i.e. similar distributions), but search

is substantively different. This behavior is also similar to the analysis of the point-based gaze data and AOI visits: observation and recall are similar, while search is the outlier. Therefore, **H4 is supported.**

7. Discussion

Based on the study results – H1, H3, and H4 supported, H2 partially supported – the answer the first part of the overarching research question posed in Section 1 is a straightforward, *yes, task significantly affects gaze behavior for text-embellished narrative visualizations*. This result is not surprising, as it replicates aspects of previous studies (both for visualization and not) that demonstrate divergent gaze behavior under different task scenarios (see Section 2.3). The second part of the question – *how* does task affect the gaze behavior between the observation, search, and recall tasks? – is more interesting. We discuss takeaways here.

7.1. The search task is the outlier at all levels of gaze behavior

One particularly interesting result from our study is that observation and recall displayed strikingly similar gaze behavior at all three levels of analysis: point-based gaze data (fixations and saccades), focusing on chart features (AOI analysis), and for aggregate viewing behavior (scanpath-based analysis). In contrast, the search task led subjects to demonstrate significantly divergent gaze behavior. This is likely because only the search task employs guided search (Wolfe, 1994). Subjects are given a definite target by which to complete the task. Since observation and recall are "open-ended", with no specified target given while the stimulus is present, these two tasks promote a similar "breadth-first" gaze behavior that encourages subjects to peruse the entire scene.

Paradoxically, prior research has demonstrated that scanpaths both can differ even for the same task on a visualization stimulus (Netzel et al., 2017b), and also be used to infer low-level perception tasks (Steichen et al., 2013). (Note that this latter paper only considers guided search tasks such as looking up and comparing variables.) Scanpath variation has also been shown to not negatively affect scene memory (Rayner, 1998). Despite this, our results show a high-level distinction, where observation and recall gaze behavior can roughly be placed in one bin, and search placed in the other.

7.2. Understanding focus on text embellishments

Even though gaze behavior differed between tasks, participants devoted a large amount of gaze towards narrative text elements regardless of the task. For example, Fig. 4(b) shows that, even for the search task, NTEs were looked at almost as much as SCEs during the first 50 fixations of the scanpath. This helps demonstrate that, even with a guided task, text embellishments still draw a large amount of the viewer's focus, even though they are not required. This aligns with prior work (e.g., Borkin et al., 2016; Ottley et al., 2019) which shows that text is an essential feature of the communicative aspects of visualization.

For the other two tasks, NTEs dominated the early fixation (Fig. 4a,c). Looking more closely at the data for the observation task in Fig. 4a, we even see that the primary NTE that was visited was the Caption AOI. This is likely a validation of Hegarty's theoretical modeling of how mental models are constructed (Hegarty and Just, 1993).

7.3. Future directions

Though this study focuses on task as the primary driver for gaze behavior, when evaluating H2 we saw interaction effects between task and visualization type for average saccade length and fixation duration (though for fixation durations, the effect size was very small). We want to be careful about broadly extrapolating these results since each task only contained five examples of each visualization type, but future studies can more deeply investigate the specific influences that chart type and NTE styling can have on gaze behavior when performing specific tasks. In addition, future studies can explore factors such as assessing how NTEs affect gaze behavior (and information retrieval) when they directly provide hints or solutions to the task (such as providing the answer to a question). We contribute study materials including our detailed AOI labeling dataset as supplemental materials to help with such studies.

The same premise also applies to the text content contained in the visualizations. As described in Section 4, when selecting MASSVIS charts as stimuli triplets, we carefully several balanced chart design features such chart styling and text content (though the amount of text itself was not controlled as an independent variable). As a sanity check, we tested whether the amount of text on a chart affected point-based gaze data. For each task, we found the correlation coefficient between the total number of words in a chart and the average saccade lengths and fixation durations.

$$cor(x, y) = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

Correlation coefficients can range between -1 and 1, to indicate negative and positive correlations. For the study, coefficient values were near 0 for all tasks, indicating little to no correlation between text amount and point-based data for our stimuli. In other words, the amount of text content was not the primary driver for saccade length and fixation duration values, but was instead the subject's current task.

However, we note here that rigorously considering text features – not just the amount of text, but other stylistic options such as font size, weight, and color – as factors in future studies will provide better insight into the relationships that chart type and text have with task performance, gaze behavior, and attention. As an example, recent work has shown that naively pairing visualization and text does not necessarily lead to improved reasoning (e.g., Ottley et al., 2019).

This demonstrates the need for more research into understanding how narrative visualizations that use text embellishments can impact the viewer's ability to extract information in the context of different task scenarios. For example, in observational and memory (recall) text content is likely to be referenced early and therefore drive initial information extraction. Similar to other empirical study papers about task-based visualization perception, this work can help inform the design of advanced task- and/or temporally-aware visualization design guidelines, quality metrics, and saliency models. In general, such work does not consider these sorts of semantics (e.g Shen and Zhao, 2014; Bylinskii et al., 2017; Matzen et al., 2018 only consider aggregate fixations). Even when task is taken into account, such as the recent work by Polatsek et al. (2018), temporally-aware metrics remain an underexplored research area.

8. Conclusion

We present an eye tracking study to investigate the effects of task on gaze behavior for text-embellished narrative visualizations. By analyzing a carefully curated set of real-world narrative visualizations, we find similar viewing strategies during

observation and recall tasks, where subjects primarily focused on narrative text elements early in viewing stages. In contrast, for a search task that required a subject to find an answer on the chart, very different behavior was observed for a number of eye tracking data points, including point-based metrics, AOIs, and scanpaths.

We hypothesize this is due to the search task motivating guided search for a specific target. Despite this, text elements were still viewed at a high rate for all three tasks, which is perhaps explained by Hegarty's theoretical modeling of how cognitive models are constructed. These results how text-based chart elements attract viewer gaze – even for tasks where they are not necessary – at all levels of perception (point-based, feature, level, and aggregate viewing behavior). We discuss potential future work to investigate possible interactions between task, chart type, and text content, and discuss how our results can be applied to task-based design guidelines and task- and temporally-aware visual saliency maps.

CRediT authorship contribution statement

Chris Bryan: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project management, Supervision. **Aditi Mishra:** Data curation, Formal analysis, Writing - original draft. **Hidekazu Shidara:** Investigation, Data curation, Formal analysis. **Kwan-Liu Ma:** Funding acquisition, Supervision, Project management.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research has been sponsored in part by the National Science Foundation, United States through grant IIS-1528203.

References

Acartürk, C., 2012. Points, lines and arrows in statistical graphs. In: Diagrammatic Representation and Inference. Springer, pp. 95–101.
Atkinson, R.C., Shiffrin, R.M., 1968. Human memory: A proposed system and its control processes1. In: Psychology of learning and motivation, Vol. 2. Elsevier, pp. 89–195.
Bateman, S., Mandryk, R.L., Gutwin, C., Genest, A., McDine, D., Brooks, C., 2010. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 2573–2582.
Borg, I., Groenen, P., 2003. Modern multidimensional scaling: theory and applications. J. Educ. Meas. 40 (3), 277–280.
Borgo, R., Abdul-Rahman, A., Mohamed, F., Grant, P.W., Reppa, I., Floridi, L., Chen, M., 2012. An empirical study on using visual embellishments in visualization. IEEE Trans. Vis. Comput. Graphics 18 (12), 2759–2768.
Borkin, M.A., Bylinskii, Z., Kim, N.W., Bainbridge, C.M., Yeh, C.S., Borkin, D., Pfister, H., Oliva, A., 2016. Beyond memorability: Visualization recognition and recall. IEEE Trans. Vis. Comput. Graph. 22 (1), 519–528.
Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., Pfister, H., 2013. What makes a visualization memorable?. IEEE Trans. Vis. Comput. Graphics 19 (12), 2306–2315.
Bylinskii, Z., Kim, N.W., O'Donovan, P., Alsheikh, S., Madan, S., Pfister, H., Durand, F., Russell, B., Hertzmann, A., 2017. Learning visual importance for graphic designs and data visualizations. In: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology. ACM, pp. 57–69.
Cleveland, W.S., McGill, R., 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. J. Amer. Statist. Assoc. 79 (387), 531–554.
Cohen, J., 1992. A power primer.. Psychol. Bull. 112 (1), 155.
Cohen, S., Hamilton, J., Turner, F., 2011. Computational journalism. Commun. ACM 54 (10), 66–71.
Drieghe, D., 2011. Parafoveal-on-Foveal Effects on Eye Movements During Reading. Oxford University Press.
Gershon, N., Page, W., 2001. What storytelling can do for information visualization. Commun. ACM 44 (8), 31–37.
Goldberg, J., Helfman, J., 2011. Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. Inf. Vis. 10 (3), 182–195.
Healey, C., Enns, J., 2012. Attention and visual memory in visualization and computer graphics. IEEE Trans. Vis. Comput. Graph. 18 (7), 1170–1188.
Hegarty, M., Just, M.-A., 1993. Constructing mental models of machines from text and diagrams. J. Mem. Lang. 32 (6), 717–742.
Hoffman, D., Singh, M., 1997. Salience of visual parts. Cognition 63 (1), 29–78.
Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2011. Eye Tracking: A Comprehensive Guide to Methods and Measures. OUP Oxford.
Hullman, J., Diakopoulos, N., 2011. Visualization rhetoric: Framing effects in narrative visualization. IEEE Trans. Vis. Comput. Graph. 17 (12), 2231–2240.
Knaflic, C.N., 2015. Storytelling with Data: A Data Visualization Guide for Business Professionals. John Wiley & Sons.
Ma, K.-L., Liao, I., Frazier, J., Hauser, H., Kostis, H.-N., 2012. Scientific storytelling using visualization. IEEE Comput. Graph. Appl. 32 (1), 12–19.
Matzen, L.E., Haass, M.J., Divis, K.M., Stites, M.C., 2017. Patterns of attention: How data visualizations are read. In: International Conference on Augmented Cognition. Springer, pp. 176–191.
Matzen, L.E., Haass, M.J., Divis, K.M., Wang, Z., Wilson, A.T., 2018. Data visualization saliency model: A tool for evaluating abstract data visualizations. IEEE Trans. Vis. Comput. Graph. 24 (1), 563–573.
Mayer, R.E., Hegarty, M., Mayer, S., Campbell, J., 2005. When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. J. Exp. Psychol.: Appl. 11 (4), 256.
Mayer, R.E., Steinhoff, K., Bower, G., Mars, R., 1995. A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. Educ. Technol. Res. Dev. 43 (1), 31–41.
Moere, A., Purchase, H., 2011. On the role of design in information visualization. Inf. Vis. 10 (4), 356–371.
Needleman, S.B., Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48 (3), 443–453.
Neisser, U., 1979. The control of information pickup in selective looking. Percept. Dev.: Tribut. Eleanor J. Gibson 201–219.
Netzel, R., Burch, M., Weiskopf, D., 2016. Interactive scanpath-oriented annotation of fixations. In: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications. ACM, pp. 183–187.
Netzel, R., Hlawatsch, M., Burch, M., Balakrishnan, S., Schmauder, H., Weiskopf, D., 2017a. An evaluation of visual search support in maps. IEEE Trans. Vis. Comput. Graphics 23 (1), 421–430.
Netzel, R., Ohlhausen, B., Kurzhals, K., Woods, R., Burch, M., Weiskopf, D., 2017b. User performance and reading strategies for metro maps: An eye tracking study. Spat. Cogn. Comput. 17 (1–2), 39–64.
Ottley, A., Kaszowska, A., Crouser, R.J., Peck, E.M., 2019. The curious case of combining text and visualization.. In: EuroVis (Short Papers). pp. 121–125.
Polatsek, P., Waldner, M., Viola, I., Kapec, P., Benesova, W., 2018. Exploring visual attention and saliency modeling for task-based visual analysis. Comput. Graph. 72, 26–38.
Rayner, K., 1998. Eye movements in reading and information processing: 20 years of research. Psychol. Bull. 124 (3), 372.
Rayner, K., 2009a. The 35th sir frederick bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. Quart. J. Exp. Psychol. 62 (8), 1457–1506.
Rayner, K., 2009b. Eye movements and attention in reading, scene perception, and visual search. Quart. J. Exp. Psychol. 62 (8), 1457–1506.
Reppa, I., Playfoot, D., McDougall, S.J., 2008. Visual aesthetic appeal speeds processing of complex but not simple icons. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 52. SAGE Publications Sage CA, Los Angeles, CA, pp. 1155–1159.
Sedig, K., Parsons, P., 2013. Interaction design for complex cognitive activities with visual representations: A pattern-based approach. AIS Trans. Hum.-Comput. Interact. 5 (2), 84–133.
Segel, E., Heer, J., 2010. Narrative visualization: Telling stories with data. IEEE Trans. Vis. Comput. Graph. 16 (6), 1139–1148.
Shah, P., Hoeffner, J., 2002. Review of graph comprehension research: Implications for instruction. Educ. Psychol. Rev. 14 (1), 47–69.
Shen, C., Zhao, Q., 2014. Webpage saliency. In: European Conference on Computer Vision. Springer, pp. 33–46.
Steichen, B., Carenini, G., Conati, C., 2013. User-adaptive information visualization: using eye gaze data to infer visualization tasks and user cognitive abilities. In: Proceedings of the 2013 International Conference on Intelligent User Interfaces. ACM, pp. 317–328.
Wolfe, J.M., 1994. Guided search 2.0 a revised model of visual search. Psychon. Bull. Rev. 1 (2), 202–238.
Yarbus, A.L., 1967. Eye Movements and Vision. Plenum, New York.