Beyond Memorability: Visualization Recognition and Recall

Michelle A. Borkin*, *Member, IEEE*, Zoya Bylinskii*, Nam Wook Kim, Constance May Bainbridge,
Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, *Senior Member, IEEE*, and Aude Oliva
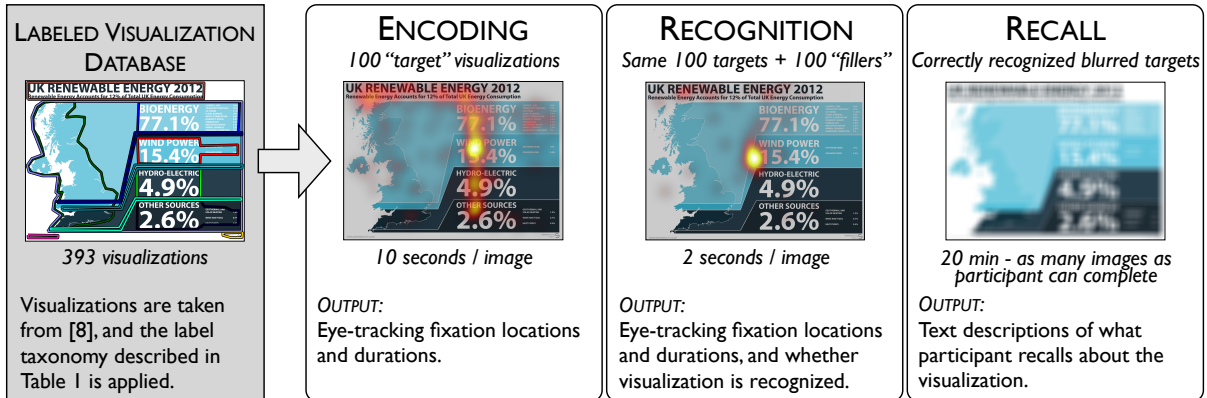
Fig. 1. Illustrative diagram of the experiment design. From left to right: the elements of the visualizations are labeled and categorized, eye-tracking fixations are gathered for 10 seconds of "encoding", eye-tracking fixations are gathered while visualization recognizability is measured, and finally participants provide text descriptions of the visualizations based on blurred representations to gauge recall.

Abstract— In this paper we move beyond memorability and investigate how visualizations are recognized and recalled. For this study we labeled a dataset of 393 visualizations and analyzed the eye movements of 33 participants as well as thousands of participant-generated text descriptions of the visualizations. This allowed us to determine what components of a visualization attract people's attention, and what information is encoded into memory. Our findings quantitatively support many conventional qualitative design guidelines, including that (1) titles and supporting text should convey the message of a visualization, (2) if used appropriately, pictograms do not interfere with understanding and can improve recognition, and (3) redundancy helps effectively communicate the message. Importantly, we show that visualizations memorable "at-a-glance" are also capable of effectively conveying the message of the visualization. Thus, a memorable visualization is often also an effective one.

Index Terms—Information visualization, memorability, recognition, recall, eye-tracking study

1 INTRODUCTION

Understanding the perceptual and cognitive processing of a visualization is essential for effective data presentation as well as communication to the viewer. Memorability, a basic cognitive concept, has important implications for both the design of visualizations that will be remembered but also lays the groundwork for understanding higher cognitive functions such as comprehension. In our previous study [8], the memorability scores for hundreds of real-world visualizations were collected on Amazon's Mechanical Turk (AMT). The results of this research demonstrate that visualizations have inherent memorability, consistent across different groups of observers. We also found that the most memorable visualization types are those that are visually distinct

(e.g., diagrams, tree and network diagrams, etc.), and that elements such as color, visual complexity, and recognizable objects increase a visualization's memorability. However a few questions remain: What visual elements do people actually pay attention to when examining a visualization? What are the differences in memorability when given more time to view a visualization? What information do people use to recognize a visualization? What exactly do people recall about a visualization?

To answer these questions, in this paper we present the results of a three-phase experiment (see experimental design in Fig. 1). In contrast to the previous study, where each visualization was presented for 1 second ("at-a-glance"), we presented each visualization for 10 seconds ("prolonged exposure") in the *encoding phase* of our experiment to allow participants more time to explore visualizations and to facilitate the collection of eye movements. During the *recognition* phase, participants viewed the *target* visualizations, mixed in with previously-unseen (*filler*) visualizations for 2 seconds each and responded to indicate which visualizations they recognized. During both phases, we collected eye fixations to examine which elements a person focuses on when visually encoding and retrieving a visualization from memory. Finally, in the *recall* phase, participants were presented with the target visualizations correctly identified during the recognition phase in a randomized order and asked to "Describe the visualization in as much detail as possible." This last experiment phase provides insight into what visualization elements, types, and concepts were easily recalled from memory. Together, the three phases of the experiment, along with a new detailed labeling of the 393 target visualizations from [8], allow us to analyze and quantify what visualization elements people encode, retrieve, and recall from memory.

- *Michelle A. Borkin is with the University of British Columbia, and Harvard University. E-mail: borkin@cs.ubc.ca.*
- *Zoya Bylinskii, Constance Bainbridge, and Aude Oliva are with the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT), E-mail: {zoya,oliva}@mit.edu*
- *Nam Wook Kim, Hanspeter Pfister, and Chelsea S. Yeh are with the School of Engineering & Applied Sciences, Harvard University, E-mail: {namwkim,pfister}@seas.harvard.edu*
- *Daniel Borkin is with the University of Michigan. E-mail: drborkin@umich.edu*
- **Equal contribution.*

Contributions: This work represents the first study incorporating eye-tracking as well as cognitive experimental techniques to investigate which elements of visualizations facilitate subsequent recognition and recall. In addition, we present an analysis of the new labeled visualizations in our database[1] in order to characterize visualization design characteristics, including data and message redundancy strategies, across different publication venues. Based on the results of our experiment, we are able to offer quantitative evidence in direct support of a number of existing conventional qualitative visualization design guidelines, including that: (1) titles and supporting text should convey the message of a visualization, (2) if used appropriately, pictograms do not interfere with understanding and can improve recognition, and (3) redundancy helps effectively communicate the message.

2 RELATED WORK

Perception and Memorability of Visualizations: Many important works in the visualization community have studied how different visualization types are perceived, and the effect of different data types and tasks [13, 29, 38]. The effect of "visual embellishments" on the memorability and comprehensibility of visualizations is also an active area of research [4, 5, 7, 8, 15, 22, 31, 35, 43, 46]. The effect and role of specific visual elements have also been investigated within the context of specific visualization types, e.g., attributes of node link diagrams [1, 16], specific visual elements such as pictographs [18], visual distortions [37], and more broadly [19]. It has been demonstrated that interactions with visualizations affect memory and recall (e.g., [25, 32]). As described in Sec.1, we evaluated the memorability of visualizations using thousands of un-edited real-world visualizations from the web [8]. We found that some visualization types are more memorable than others, and particular visual elements (e.g., human recognizable objects, color, etc.) seem to increase a visualization's memorability. In this study we build on all of this previous work and move beyond basic memorability by studying what visualization types and features contribute to visualization recognition and recall.

Visual Cognition and Memory: Work on image memorability has reported high consistency between participants and experimental settings with regards to which images are memorable and forgettable. Studies have shown that the memorability rank of images is stable over time [23] and over experimental contexts [12]. Memorability has been demonstrated to be an intrinsic property of scene images [12, 23], faces [3], and graphs [8]. Studies in visual cognition and memory have demonstrated high fidelity of memories over time and massive storage capacity in long-term memory [9, 12, 27, 47]. People remember objects that have defined meanings and use cases [10], people remember a lot of details from visual objects and scenes [27, 47], and people can easily distinguish between different exemplars of the same category [10, 12, 27]. Also, unique or distinct visual stimuli have been found to be easier to remember in massive memory studies [27, 12, 8]. Brady et al. have shown that when an object is retrieved from memory, many details are retrieved along with it [10]. In this paper we go beyond the memorability of visualizations [8] and explore what details of visualizations are remembered along with it, and what elements of a visualization lead to better recognition and recall.

In the psychology literature, there is a dual process model of memory called *process dissociation* [24, 34, 45, 30]. The two memory processes include a fast, automatic "familiarity-driven" process, and a slower, conscious "recollection-driven" process. The former is thought to be the result of the perceptual system's quicker processing of stimuli [24]. Thus, the experiments presented in this paper may be thought of as accessing two memory processes: single-response recognition and the longer, detail-retrieving recall. However, even with our 2-second recognition phase, the processes at play are not just perceptual, as participants have time to go back to some of the textual elements before making their response. We are thus likely capturing both memory processes during both experimental phases, and both play an important role in how and what people remember in visualizations.

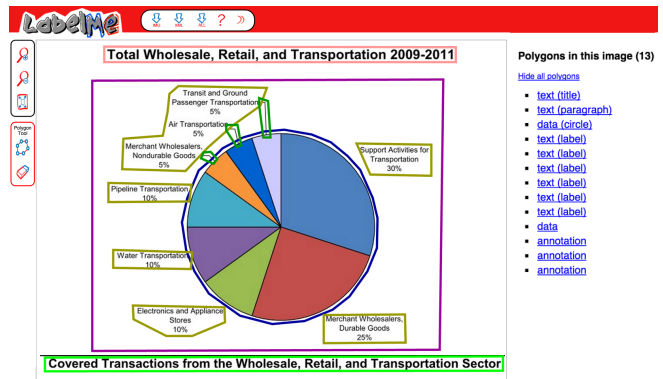[1] Massachusetts (Massive) Visualization Dataset (**MASSVIS**) available at http://massvis.mit.edu.

Fig. 2. Example labeled visualization in the LabelMe system [41]. This pie chart from a government publication has 13 labeled elements including where the data, annotations, and title are located.

Eye-tracking Evaluations in Visualization: Eye-tracking evaluations can be an effective tool for understanding how a person views and visually explores a visualization [6]. They have been used in the visualization community for evaluating specific visualization types such as graphs [20, 21, 28, 39], tree diagrams [11], and parallel coordinates [42], for comparing multiple types of visualizations [17], and for evaluating cognitive processes in visualization interactions [25]. There has also been research in the area of understanding different types of tasks and visual search strategies for visualizations through the analysis of eye-tracking fixation patterns as well as insights into cognitive processes [39, 40]. The work presented in this paper does not focus on specific tasks, nor on a specific type of visualization, but rather uses eye-tracking for fixation location and duration analysis on hundreds of labeled and categorized real-world visualizations from the web with dozens of study participants. Within the context of our experimental design, we are able to more deeply understand the specific cognitive processes involved in recognition and recall of visualizations.

3 DATA COLLECTION AND ANNOTATION

We used the database of visualizations from our previous memorability study [8]. The database was generated by scraping multiple sources of real-world visualization publication venues online covering government reports, infographic blogs, news media, and scientific journals. The diversity and distribution of these visualizations represent a broad set of data visualizations "in the wild". For our present study, we used the same 393 target visualizations utilized in the previous study [8] along with 393 visualizations selected from the remaining "single" (i.e., stand alone single-panel) visualizations in the collection as filler images (see Sec. 5.3).

To gain deeper insight into what elements of a visualization affect its recognition and recall, three experts manually labeled the polygons of each of the visual elements in the 393 target visualizations using the LabelMe system [41]. All labels were reviewed for accuracy and consistency by three visualization experts. Examples of labeled visualizations are shown in the leftmost panel of Fig. 1 and in Fig. 2.

The labeling taxonomy was based on the visualization taxonomy from [8]. As described in Table 1, the labels classify the visualization elements to be either data encoding, data-related components (e.g., axes, annotations, legends, etc.), textual elements (e.g., title, axis labels, paragraphs, etc.), human recognizable objects (HRO), or graphical elements with no data encoding function. Labels could overlap in that a single region could have a number of labels (e.g., an annotation on a graph has an annotation label and a graph label). Additionally, the title of each visualization was manually coded for further analysis.

We also documented whether each visualization exhibited one of two types of redundancies: *data* and *message* redundancy. A visualization exhibits data redundancy if the data being presented is visually encoded in more than one way. This can include the addition of quantitative values as labels (e.g., numbers on top of bars in a bar chart, as illustrated in Fig. 3, or on sectors in a pie chart), or the use of channels

Table 1. The visualization labeling taxonomy used to annotate our target visualizations. The data subtypes taxonomy is taken from [8].

| LABEL	[OPTIONAL SUBTYPES] DESCRIPTION
Annotation	[Arrow] Outline of any visual elements annotating the data. A specific subtype of "arrow" was included to denote whether the annotation was in the form of an arrow.
Axis	[Time] Outline of where an axis is located including any tick marks and numeric values along the axis. A specific subtype of "time" was included to denote an axis involving time.
Data	Outline of where the full data plot area is located (e.g., the area between the x-axis and y-axis in a 2D plot).
Data (type)	[Area, Bar, Circle, Diagram, Distribution, Grid & Matrix, Line, Map, Point, Table, Text, Trees & Networks] Outline of where the actual data values are visually encoded (e.g., bars in a bar graph, points in a scatterplot, etc.).
Graphical Element	Outline of any visual elements that are not related to the visual representation or description of the data.
Legend	Outline of any legends or keys that explain the data's visual encoding (e.g., color scales, symbol keys, map legends, etc.).
Object	[Photograph, Pictogram] Outline of any human recognizable objects (HRO) in the image. Objects are either realistic in representation (photograph) or abstract drawings (pictogram). Descriptions of each object were also recorded.
Text	[Axis Label, Header Row, Label, Paragraph, Source, Title] Outline of any text in the image. Subtypes cover all the common representations from prose to labels.
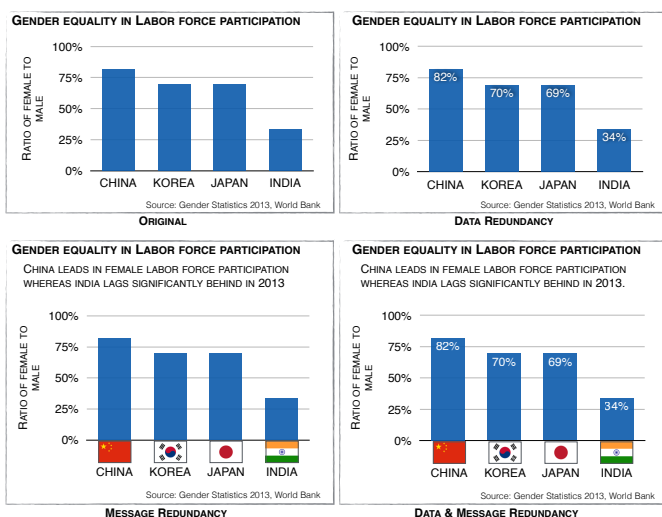
Fig. 3. Illustrative examples of data redundancy (i.e., additional quantitative encodings of the data) and message redundancy (i.e., additional qualitative encodings of the main trend or message of the data). More examples are provided in the Supplemental Material.
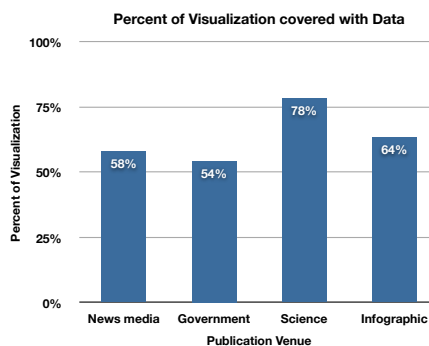

Fig. 4. Percentage of visualization average pixel area covered by the *data* label. Scientific visualizations on average had the highest percentage of image area devoted to data presentation.

such as color, size, or opacity to represent a value already exhibited in a visualization such as the x- or y-axis values. In contrast, a visualization exhibits message redundancy if the main conclusion or message of the visualization is explicitly presented to the viewer in multiple ways: the addition of explanatory annotations, labels, text, and pictures. A visualization can exhibit *both* data and message redundancy (e.g., lower right panel of Fig. 3).

Annotating redundancy for each target visualization allows us to evaluate whether redundancy enables more effective recognition and recall. Each of the two types of redundancies were recorded as present or absent, by three visualization experts. The visualizations were reviewed and discussed until unanimous consensus was found.

4 ANALYSIS OF LABELED VISUALIZATIONS

The labeled visualizations enable us to gain insight into and study the distribution and type of visual elements employed across publication venues and visualization types. These insights also help us understand and put into context the observed trends and results of our study (Sec. 6). Examining the proportion of image area covered by the *data* label, as shown in Fig. 4, we see that it is highest for the scientific journal visualizations. This is probably due to the publishing context of scientific journal figures in which the visualization occurs as part of a paper narrative. Additionally, authors commonly have enforced page limits and figure limit constraints, so maximizing information per unit area is a constraint.

Breaking down this measure of image area for data display by vi-

sualization type, we see that diagrams, maps, and tables cover a larger percentage of the image area than other visualization types. These types of visualizations tend to have annotations and text labels incorporated into the data representation itself, thus requiring less of the image area around the data plot for additional explanations.

Another observation is the difference in the average total number of elements in a visualization across sources. Visualizations from government sources have on average 11.9 labeled elements per visualization, significantly fewer compared to other visualization sources ($p < 0.001, t(177) = 4.79$ when compared[2] to the numerically closest visualization source, science). In contrast, visualizations from infographic sources have nearly twice as many elements ($M = 38.7$) as compared news media ($M = 19.7$, $p < 0.001$, $t(212) = 5.73$) and scientific visualizations ($M = 18.4$, $p < 0.001$, $t(169) = 5.23$). The additional elements in the infographic visualizations are mostly in the form of more text, objects, and graphical elements around the data.

Finally, there is a difference between publication venues in the percentage of the visualization's area covered by human recognizable objects (HRO). There are no such objects in the government visualizations, and the percentages are generally less for scientific journal visualizations ($M = 14\%$) as compared to news media ($M = 25\%$) and infographic ($M = 29\%$) visualizations. The human recognizable objects are primarily in the form of company logos (McDonalds, Twitter, Apple, etc.), international flags commonly used in the news media visualizations, and pictograms or photographs of human representations and computer/technology depictions (see also the word cloud visualization in the Supplemental Material).

In addition to the labels, each visualization was examined to determine if it exhibited message redundancy, data redundancy, both, or neither. As shown in Fig. 5, all publication venues included visual-

---

[2]For all the p-values reported in this paper, t-tests were corrected for multiple comparisons.

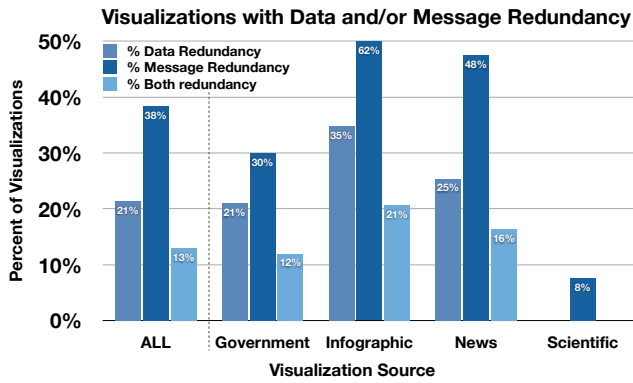**Visualizations with Data and/or Message Redundancy**

Fig. 5. The percent of visualizations that exhibit data and message redundancies by publication source. The largest percentage of visualizations with message redundancy are from the Infographic and News media venues. Overall and across all visualization sources, there is more frequent use of message redundancy as compared to data redundancy.

izations with message redundancy, with the highest rates in the infographic and news media visualization. This is probably due to both venues prioritizing clear communication of the visualization message. In contrast, the scientific visualizations in our sample had the least message redundancy and no data redundancy. When examining redundancy across visualization types, the message redundancy rates are comparable across all visualization types but highest for circle (53% of circle visualizations contain message redundancy), table (51%), line (44%), and bar chart (39%) visualizations.

5 EXPERIMENT OVERVIEW

5.1 Experiment Set-up & Participants

As discussed in Sec. 3, we labeled all the visual elements in the 393 target visualizations for which we have memorability scores [8]. We also carefully selected 393 visualizations from the database of single visualizations [8] to use as fillers during the recognition portion of the experiment. These visualizations match the exact distribution of visualization types and original sources as the target visualizations. All target and filler visualizations were resized, while preserving aspect ratios, so that their maximum dimension was 1000 pixels.

For the eye-tracking portions of our experiment, we used an SR Research EyeLink1000 desktop eye-tracking system with a chin-rest mount 22 inches from a 19 inch CRT monitor (resolution: 1280 x 1024 pixels). At the beginning of an experimental session, participants performed a randomized 9-point calibration and validation procedure. At regular intervals, a drift check was performed and, if necessary, recalibration took place. Optional breaks were offered to participants.

A total of 33 participants (17 females, 16 males) participated in the experiment. All of the participants were recruited from the local communities of Cambridge and Boston with an average age of 22.9 years ($SD = 4.2$). All participants had normal color vision. In a single experiment lasting 1 hour, a participant would see about 100 randomly-selected (out of a total of 393) target visualizations. A single session of the experiment lasted about 1 hour. Each participant was monetarily compensated $25 for their time.

5.2 Encoding Phase

The first portion of the experiment, the encoding phase, lasted 20 minutes. As illustrated in Fig. 1, participants were each shown about 100 visualizations, randomly selected from the 393 labeled target visualizations. For this phase of the experiment, participants examined each visualization for 10 seconds while being eye tracked. Visualizations were separated by a 0.5 second fixation cross to clear their field of view. A 10 second 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 was short enough to avoid too much redundancy in refixations as well as explorative strategies. The mea-

sures collected during this phase were the (x,y) fixation locations[3] (in pixel coordinates on each visualization) and durations of each fixation (measured in ms).

5.3 Recognition Phase

The recognition phase of the experiment, directly following the encoding phase, lasted 10 minutes. Participants were shown the same 100 visualizations they saw in the encoding phase as well as 100 filler visualizations. These 200 visualizations were presented in a random permutation for 2 seconds each with a 0.5 second fixation cross between visualizations. Participants pressed the spacebar to indicate recognition of a visualization from the previous experimental phase. A feedback message was presented (i.e., correct, incorrect) after each visualization. The 2 second duration time was chosen to be brief enough to quickly collect responses, but long enough to capture meaningful eye-tracking data. The measures collected during this phase were the (x,y) fixation locations and durations for each fixation (in ms), and the number of correctly-recognized visualizations (*HITs*).

5.4 Recall Phase

The final portion of the experiment, the recall phase, lasted 20 minutes. Participants' gazes were not recorded. In this phase, all the visualizations the participant correctly recognized in the previous phase were presented in a randomized sequence. Each visualization was presented at 60% of its original size and blurred by a 40-pixel wide Gaussian filter to make the text unreadable. The purpose of blurring visualizations was to allow the visualization to be recognizable, but not contain enough visual detail to enable the extraction of any new information.

Next to each blurred visualization was an empty text box with the instruction: "Describe the visualization in as much detail as possible." The goal was to elicit from the participant as much information about the visualization as they could recall from memory. Participants were given 20 minutes in total to write as many descriptions as possible. There was no limit to how much time or text length was spent on each visualization, nor any requirement to complete a certain number of visualizations. The measures collected during this phase were the participant-generated text descriptions of what they could recall of a given visualization.

Displaying blurred visualizations at this phase of the experiment has its limitations, including the potential to bias participants to more easily recall visual elements. However, we chose to use this modality due to the following advantages: (1) the study can be performed at-scale, showing participants many visualizations and collecting text responses in a later phase, (2) no assumptions are made about what participants can remember and at what level they extract the content, and thus free-form responses can be collected, and (3) participants can be queried on specific visualizations they have stored in long-term memory via visual hints (i.e., the blurred visualizations).

5.5 Performance Metrics

We computed fixation measures by intersecting fixation locations with labeled visualization elements (Table 1) to determine when fixations landed within each element. We analyze the total duration of a viewer's fixations landing on a given visual element throughout the entire viewing period. We also measure **refixations** - the total number of times a viewer returns to an element during the entire viewing period (including the first time the element is fixated). Consecutive fixations on the same element are not counted. Note that a single fixation can land on several elements at once (e.g., an annotation on a graph). In this case, we count the fixation as belonging to all of those elements. We collect all of a viewer's fixations during a particular viewing period (i.e., 10 seconds for encoding, 2 seconds for recognition), and we discard as noise fixations lasting less than 150 milliseconds. All fixation measures are averaged across viewers and different sets of visualizations, and compared using Bonferroni-corrected t-tests.

---

[3]The ordering of fixations is also available in the data, but was not used for the present study.
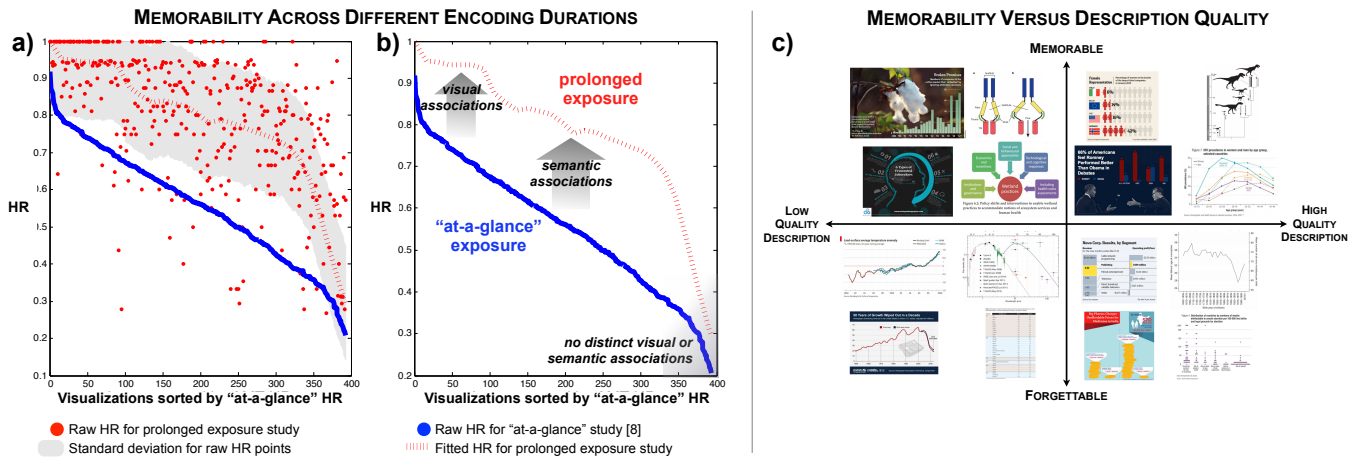
Fig. 6. **a)** Plot comparing the raw HR (i.e., memorability) scores of target visualizations from the recognition phase of this experiment (after 10 seconds of encoding) to the raw HR scores of the same visualizations from the previous experiment [8] (with 1 second of encoding). **b)** Summarizing the same data by box filtering the raw HR scores makes the main trends clearer. The most memorable "at-a-glance" visualizations are still the most memorable after prolonged exposure, likely due to visual associations. The increase in memorability across experiments for some visualizations can be explained by additional semantic associations (e.g., title, text, etc.) being invoked. **c)** The top and bottom ranked visualizations across description quality and memorability, for all four publication source categories. The y-axis represents recognition HR, and the x-axis represents average text description quality at recall. Description quality is also correlated with memorability (Sec. 6.1.4).

We also computed the recognition hit rate (*HR*) for each visualization. This is the fraction of participants who correctly recognized a visualization when they saw it during the recognition phase of the experiment. This value ranges from 0 to 1. See the Supplemental Material for a discussion of the measurement of memorability in this experiment compared to [8].

To quantify the qualitative aspects of the text descriptions collected during the recall phase of the experiment, three visualization experts went through all 2,449 descriptions. The **description quality** is a measure of how well a participant recalled the content and message of the visualization. Quality was rated from 0 to 3 where 0 was an incorrect or incoherent description, and 3 was a text description that touched upon the visualization topic, what data or information is presented in the visualization, the main message of the visualization, and one additional specific detail about the visualization. The quality ratings were assigned based on unanimous consensus. Each description was also reviewed for the visual components explicitly discussed or referenced from the visualization including the title and other textual elements. Finally, descriptions were flagged if they were not perfectly accurate (i.e., contained a factual error).

6 EXPERIMENTAL RESULTS AND DISCUSSION

In our previous study [8], we showed that people are highly consistent in which visualizations they recognize and which they forget. Visualizations were only shown for 1 second at a time, so we were able to measure how memorable visualizations are "at-a-glance". In this study we want to test not just recognizability, but we also want to discover which visualization elements people encode and are consecutively able to recall (top right quadrant of Fig. 6c). We also want to test what aspects of a visualization impact how well the main message of the visualization is understood. For this purpose we extended encoding time from 1 to 10 seconds and analyzed eye movements at encoding, responses at recognition, and textual descriptions at recall.

6.1 "At-a-glance" vs. "prolonged exposure" memorability

6.1.1 Does "at-a-glance" memorability generalize?

During the recognition phase of the current study, as discussed in Sec. 5, hit rates (HR) were generated for each visualization (i.e., what percentage of participants correctly identified the visualization as occurring during the encoding phase). For the current memorability study discussed in this paper (with 10 seconds of encoding), the mean HR is 79.70% ($SD = 17.98\%$) as compared to a mean HR of 55.61%

($SD = 15.30\%$) from the previous study [8] (with 1 second of encoding). We compare the memorability scores (HR) of both studies for all the target visualizations in Fig. 6a-b. When time to encode a visualization is increased from 1 to 10 seconds, it is natural for the absolute memorability scores to increase. However, what we are more interested in is the stability of the relative scores (i.e., ranks) of the visualizations across studies. We find a Spearman rank correlation of 0.62 ($p < 0.001$) between the memorability scores of the two studies. Note that the inter-participant consistency in the initial study was not perfectly correlated either, with a rank correlation of 0.83 [8]. Thus, the relatively high correlation between the two studies (despite the difference in experimental set-up, participant population, and encoding time) points to the stability of memorability ranks of the visualizations. We also see that this trend holds if we look separately within each source category. The Spearman rank correlations between the HR scores of the two studies are: 0.44 for infographics, 0.38 for news, 0.56 for government, and 0.59 for science ($p < 0.001$ for all). **Thus, visualizations that were memorable "at-a-glance" (after only 1 second of encoding) are often the same ones that are memorable after "prolonged exposure" (10 seconds of encoding)**. The same holds for the forgettable visualizations. Thus our initial study's findings generalize to a longer encoding time when people can process more of the visual and textual input. However, now that encoding time has increased, we also notice that 8% of the visualizations that were not previously in the top third most memorable visualizations move up in rank as compared to the previous study. In the next section we discuss why some visualizations become more memorable after 10 seconds of encoding than after 1 second, and why others remain forgettable in both cases.

6.1.2 What are the differences between the most and least recognizable visualizations?

We use the eye movements of a participant at recognition time, just before a response is made, as indicators of which parts of the visualization trigger the response (i.e., help retrieve the memory of the visualization). We can compare the differences in eye movements for the visualizations that are "at-a-glance" the most and least recognizable [8]. By considering the eye movements made on these visualizations during 10 seconds of encoding and at recognition time, we can see what parts of a visualization are encoded and what parts are required for recognition.

As shown in Fig. 7, the heat maps overlaid on the visualizations represent the average of all of the participants' encoding fixations on the visualization. The fixation patterns in the encoding phase demon-
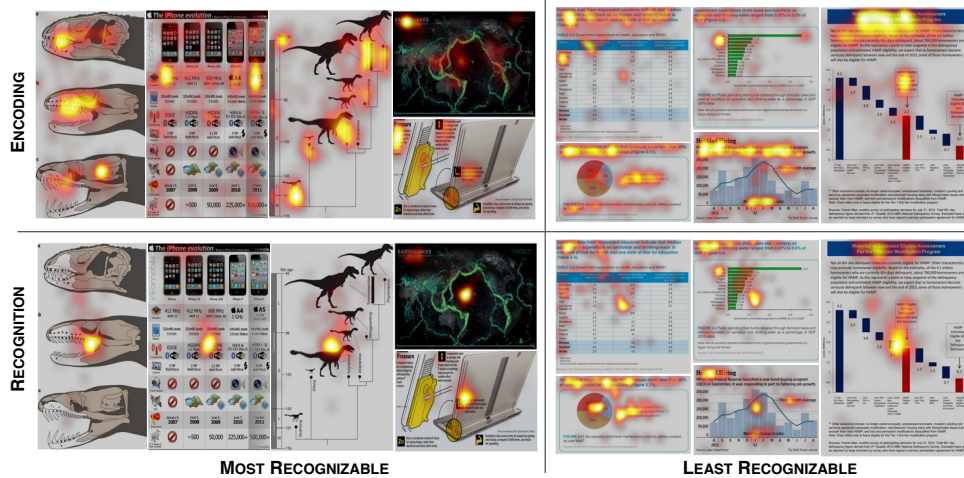
Fig. 7. Examples of the most and least recognizable visualizations from [8]. **TOP**: Eye-tracking fixation heat maps (i.e., average of all participants' fixation locations) from the *encoding* phase of the experiment in which each visualization was presented for 10 seconds. The fixation patterns demonstrate visual exploration of the visualization. **BOTTOM:** Eye-tracking fixation heat maps from the *recognition* phase of the experiment in which each visualization was presented for 2 seconds or until response. The most recognizable visualizations all have a single focus in the center indicating quick recognition of the visualization, whereas the least recognizable visualizations have fixation patterns similar to the encoding fixations indicative of visual exploration (e.g., title, text, etc.) for recognition.

strate patterns of visual exploration (e.g., view graph, read title, look at legend, etc.), corresponding to the trends described in Sec. 6.2. These visual exploration patterns are seen in both the most and least recognizable visualizations.

The heat maps generated for recognition fixations are obtained by taking only the fixations until a positive response is made, i.e., only the fixations that lead up to a successful recognition, so that we can determine which parts of a visualization participants look at to recognize a visualization. In Fig. 7 we see a distinct difference between the fixation heat maps of the most and least recognizable visualizations in the *recognition phase*, where the most recognizable visualizations have a fixation bias towards the center of the visualization. This indicates that a fixation near the center, with the accompanying peripheral context, provides sufficient information to recognize the visualization without requiring further eye movements. In contrast, the least recognizable visualizations have fixation heat maps that look more like the fixation patterns of visual exploration in the encoding phase. These visualizations are not recognizable "at-a-glance" and participants visually search the visualization for an *association*, i.e., a component in the visualization that will help with recognition. The fixations along the top of the heat map for the least recognizable visualizations generally correspond to the location of the title and paragraph text describing the visualization in further detail. We conducted statistical tests to verify that the fixation patterns are significantly different between the most and least recognizable visualizations, and we describe these analyses in the Supplemental Material.

We see that **the recognition fixations are significantly different between the most and least recognizable visualizations in that the least recognizable visualizations require more exploration before they can be retrieved from memory**. There is more eye movement during recognition for the least recognizable visualizations indicating that people need to look at more parts of a visualization before they can recognize it. The visual appearance is no longer enough and people start exploring the text. In the following section we will discuss some possible explanations for this difference and what visualization elements people may be using for recognition.

6.1.3 Which visualization elements help recognition?

We have demonstrated that there is a distinct difference between the fixation patterns of the most and least recognizable visualizations. Which visual elements in the visualization contribute to this difference in fixation patterns, and which elements help explain the overall HR increase for recognition? We believe that people are utilizing two

types of associations to help with visualization recognition: *visual associations* and *semantic associations*. The recognizable visualizations tend to contain more human recognizable objects (objects are present in 74% of the top third most recognizable visualizations) compared to the least recognizable visualizations (objects are present in 8% of the bottom third least recognizable visualizations). These human recognizable objects are examples of visual associations.

The visualizations that are not visually distinct must be recognizable due to semantic associations, such as the visualization's title or similar text. In fact, the elements with the highest total fixation time during recognition are the titles ($M = 274ms$) followed by human recognizable objects ($M = 246ms$, $p < 0.01$). The titles serve as a semantic association, and the objects as a visual association for recognition. Note that in Fig. 7 the fixations outside of the central focus in the least recognizable visualizations during recognition correspond to the visualizations' title and other textual elements near the top. Thus these types of semantic associations, i.e., textual components, are used for recognition if the visualization is not visually distinct or does not contain sufficient visual associations.

The least recognizable visualizations likely do not have a strong visual association nor a strong semantic association to assist with recognition (Fig. 6b). The least recognizable visualizations after 10 seconds have significant overlap with the least memorable visualizations from [8] (after 1 second). Also, of the 1/3 least recognizable visualizations 54% are from government sources (compare that to only 3% of government sources in the top 1/3 most recognizable visualizations), while of the 1/3 most recognizable visualizations, 49% are from infovis sources (and only 2% of the bottom 1/3 least recognizable visualizations are from infovis). Government visualizations tend to use the same templates and similar aesthetics, contributing to confusion between visualizations during recognition. The government visualizations are also heavily composed of the least memorable visualization types including bar and line graphs. **The least memorable visualizations "at-a-glance" are also the least recognizable even after prolonged exposure**.

6.1.4 What do people recall after 10 seconds of encoding?

Across 374 target visualizations[4], the study participants generated 2,733 text descriptions (see Supplemental Material for examples of participant-generated descriptions). The mean length of a descrip-

---

[4]We removed visualizations for which participants complained about the text being too small to read.

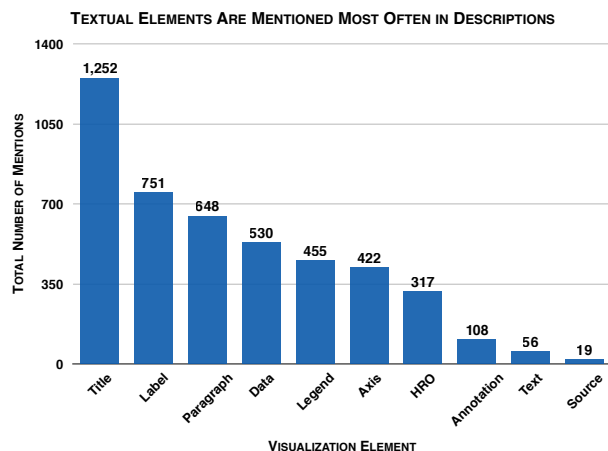**TEXTUAL ELEMENTS ARE MENTIONED MOST OFTEN IN DESCRIPTIONS**

Fig. 8. The total number of mentions of specific visualization elements in the participant-generated recall text descriptions. Textual elements received the most mentions overall, and specially the title received the most mentions across all visualizations.

Table 3. Most frequently mentioned visualization elements by publication source (percent of time element was mentioned in the text description out of total number of times it was present). Titles, labels, and paragraphs are mentioned most often.

Rank:	1st	2nd	3rd
Overall	Title (46%)	Label (27%)	Paragraph (24%)
Infographics	Title (72%)	Label (28%)	Paragraph (24%)
News	Paragraph (45%)	Title (43%)	Label (33%)
Government	Title (55%)	Legend (26%)	Data (21%)
Science	Label (27%)	Axis (14%)	Legend (13%)

tion is 13.1 words ($SD = 9.6$), with an average of 7.3 descriptions ($SD = 4.0$) per visualization. Table 2 contains a breakdown of the description statistics per source category. We can see that infographics sources generated the most number of descriptions in total (877), with the largest average number of descriptions per visualization ($M = 11.2$, $SD = 3.0$) and the longest descriptions in terms of word count ($M = 14.3$, $SD = 10.0$). More importantly, they also correspond to the highest-quality descriptions ($M = 2.09$, $SD = 0.79$), statistically higher than the descriptions for the news media visualizations ($M = 1.99$, $SD = 0.88, p < 0.001$). As discussed in Sec. 5, the description quality is a measure of how well participants recalled the content and message of the visualization. A higher description quality for infographics implies that participants recalled more details about the source that had the most recognizable visualizations. For reporting statistics related to description quality, we only consider visualizations with at least 3 participant-generated descriptions (so that computations of average description quality are more meaningful). The source with least recognizable visualizations, government, also had the fewest total descriptions (465), the fewest descriptions per visualization ($M = 4.7$, $SD = 2.9$), and the shortest descriptions ($M = 10.8$, $SD = 8.4$). The average description quality was 1.37 ($SD = 0.93$), similar to science visualizations. Thus participants were able to recall fewer things about the least recognizable visualizations.

Overall, the Spearman rank correlation between "at-a-glance" HR [8] and description quality (after prolonged exposure) is 0.34 ($p < 0.001$), and within just infographics, the source with the highest-quality descriptions, the correlation is 0.24 ($p < 0.01$). The positive correlation is also present (though not significant) for the other 3 sources, possibly due to insufficient data (fewer visualizations and fewer descriptions per visualization than in the infographics category). Fig. 6c contains example visualizations categorized by how memorable they were and the quality of descriptions generated for them (i.e., whether what was memorable was the main content and message). Sample visualizations with descriptions, sorted by source, are

included in the Supplemental Material.

At this point, although we cannot make any causality assumptions, the findings above demonstrate that visualization recognition is related to description quality. **In other words, visualizations that were most recognizable after only 1 second of viewing also tend to be better described (after 10 second exposure)**. It might very well be that a third factor (e.g., relevance of topic to viewer) contributes both to recognizability and description quality. However, even in that case, knowing something about recognizability can tell us something about description quality. Thus, visualizations that were memorable for their visual content after 1 second of viewing are memorable after 10 seconds of viewing, and more importantly, their semantic content (the message of the visualizations) is correctly recalled by experimental participants (hence the higher description quality).

6.2 The effects of visualization elements on memorability and comprehension

In this section we investigate which visualization elements attract attention during encoding through the analysis of eye movements, and which elements are recalled when a visualization is described from memory through the analysis of textual descriptions. We use the eye movements on a visualization as a proxy for what information is being encoded into memory. From the content of the recall text descriptions, we can infer from which elements of a visualization a participant extracted information. As discussed in Sec. 5, the description quality score corresponds to the extent to which a participant recalled details and the main message of the visualization. These two metrics (eye movements and textual descriptions) together are evidence for which elements of a visualization contribute to a participant's memory and understanding of a visualization.

6.2.1 Titles

When participants were shown visualizations with titles during encoding, the titles were fixated 59% of the time. Correspondingly, during the recall phase of the experiment, titles were the element most likely to be described if present (see Fig. 8). When presented visualizations with titles, 46% of the time the participants described the contents, or rewording, of the title. For the infographics sources, titles were fixated and described most often (80% and 72% of the time, respectively). A full breakdown by source category of elements described most often is presented in Table 3.

What is the contribution of visualization titles to the quality of the textual descriptions? Across the 330 visualizations with at least 3 descriptions, the average description quality of visualizations with titles is 1.90 as compared to 1.30 for visualizations without titles (the difference in quality is statistically significant at $p < 0.001$). This trend is partially driven by the absence of titles for the scientific visualizations. We believe this might be one explanatory factor as to why the scientific visualizations have the lowest quality descriptions.

Titles were also more likely to be fixated (76% of the time) when present at the top of the visualization than when present at the bottom (fixated 63% of the time). This trend holds across the source categories. Note that government visualizations, the least memorable category with poorer descriptions, contained the most occurrences of titles at the bottom of the visualization (73% of all visualizations that had a title at the bottom are from a government source). Interestingly, for government visualizations, table header rows were fixated more frequently and longer than titles (see Supplemental Material). For government sources, titles that were found at the top of the visualization were fixated 85% of the time and described 59% of the time, compared to titles that were found at the bottom of the visualization, which were fixated only 62% of the time and described 46% of the time (see Table 2 in the Supplemental Material for the complete breakdown). Thus, titles are more likely to be paid attention to and later recalled when present at the top of a visualization.

Across all textual elements, the title is among the most important. A good or a bad title can sometimes make all of the difference between a visualization that is recalled correctly from one that is not. We

Table 2. Summary of statistics related to visualization descriptions by source (** = $p < 0.001$). Mean description quality is computed only on visualizations with at least 3 participant-generated descriptions. Infographics had the most, longest, and highest-quality descriptions.

Source	Total # of vis.	Total # of desc.	Mean # of desc. per vis.	Mean desc. length (# words)	Vis. with no desc.	Vis. with at least 3 desc.	Correlation (HR, # of desc.)	Mean desc. quality (0-3 scale)	Inaccurate desc.
Overall	374	2733	7.3 (SD:4.0)	13.1 (SD: 9.6)	6%	84%	0.57**	1.77 (SD: 0.94)	13.1%
Infographics	78 (20.9%)	877 (32.1%)	11.2 (SD:3.0)	14.3 (SD: 10.0)	15%	85%	0.51**	2.09 (SD: 0.79)	14.9%
News	120 (32.0%)	848 (31.0%)	7.1 (SD:3.8)	14.1 (SD: 10.0)	3%	88%	0.43**	1.99 (SD:0.88)	15.7%
Government	99 (26.5%)	465 (17.0%)	4.7 (SD:2.9)	10.8 (SD: 8.4)	3%	74%	0.36**	1.37 (SD: 0.93)	8.6%
Science	77 (20.6%)	543 (19.9%)	7.1 (SD:3.4)	11.5 (SD: 8.7)	3%	90%	0.49**	1.24 (SD: 0.93)	10.1%

observed that when the title included the main message of the visualization (e.g., "66% of Americans feel Romney Performed Better Than Obama in Debates"), more participant-generated descriptions recalled the visualization's message compared to instances when the title was more generic or less clear (e.g., "Cities"). Select participant-generated descriptions for exemplar visualizations with good and bad titles are presented in the Supplemental Material. Note that prior psychology studies have demonstrated that concrete, imageable words (those that can be easily imagined or visualized) are easier to remember than abstract ones [26, 36, 48, 33]. Recent work by Mahowald and colleagues [33] has begun investigating the memorability of individual words, opening up future extensions to designing more memorable titles.

6.2.2 Pictograms

Pictograms, i.e., human recognizable objects (HRO), did not seem to distract participants during encoding. In fact, averaged over all visualization elements, the total fixation time spent on pictograms was less than all the other visualization elements (see Supplemental Material). Infographics is the only source where pictograms were refixated the most and fixated for the longest time in total, but these are also the visualizations with the highest recognizability and quality descriptions.

Overall (as in Table 4), the average description quality for visualizations with pictograms (2.01) is statistically higher than for visualizations without pictograms (1.50, $p < 0.001$). Within each source category, the mean description quality for visualizations with pictograms is always higher than for visualizations without pictograms. This is not always a significant difference, but what this demonstrates is that even if pictograms do not always help, they do not seem to hinder description quality. In fact, if we consider the top third best described visualizations, 20% of them have pictograms, compared to only 2% among the visualizations in the bottom third (Supplemental Material). Across all source categories, the visualizations with the best descriptions are more likely to contain pictograms than the visualizations with the lowest-quality descriptions. **Thus, what participants recall about visualizations is not hindered by pictograms**. In other words, with pictograms the message of the visualization is correctly recalled (and in fact, often better). Participants may use pictograms in these cases as visual associations (see Sec. 6.1.3). Pictograms might also provide a form of message redundancy (see Sec. 6.3).

Note, however, that some pictograms *can* hurt the understanding of a visualization. Any time people spend on a pictogram is time not spent on the data or the message (unless the pictogram contains part of the data or message), and if the link between the data/message and the pictogram is not clear, people may misunderstand and/or misrecall what the visualization was about. In the Supplemental Material, we qualitatively review the lowest quality text descriptions and see that pictograms that do not relate to the visualization's data or message can produce confusion and misdirect attention.

6.2.3 Other Elements

Across all visualizations, the elements that were refixated the most were the legend, table header row (often acting as a title), and title.

Table 4. Across all sources, visualizations with pictograms have similar or better quality descriptions than visualizations without pictograms (** = $p < 0.001$ when comparing visualizations with and without pictograms).

Source	Mean description quality (0-3 scale)	
	With pictogram	**Without pictogram**
Overall	2.01**	1.50**
Infographics	2.09	2.07
News	2.10**	1.84**
Government	1.46	1.36
Science	1.52**	1.12**

The elements that were fixated the longest were the paragraph, legend, and header row (a full breakdown by source is provided in the Supplemental Material.) We find that the elements that most commonly contribute to the textual descriptions, across all visualizations, include title, label, and paragraph (see Table 3). For news media sources, the paragraph is used more often than the title and label in the recall descriptions. This may be due to the commonly short, and not necessarily informative, titles provided. For example, a viewer will gain more from the explanatory paragraph "Numbers of companies in the cotton market that 'defaulted' by ignoring arbitration decisions" instead of its title "Broken Promises". Recall that the science visualizations we have do not contain titles, a common convention in scientific journals, and so participants refer instead to the labels, axis, and legend in their descriptions (since those are often the only elements that contain any explanatory content). **Thus, when the title is not sufficient, people turn to the other text in a visualization, including the explanatory paragraph and annotations (labels).**

6.3 The Importance of Redundancy

The average description quality is higher for visualizations that contain message redundancy ($M = 1.99$) than for visualizations that do not ($M = 1.59$, $p < 0.001$). Similarly, visualizations that contain data redundancy also have better quality descriptions ($M = 2.01$) than those that do not ($M = 1.70$, $p < 0.001$). We can also see this trend by examining redundancy in the visualizations of the top third and bottom third of average description quality rankings. Of the visualizations in the top third of description quality ranks, 57% contain message redundancy and 34% contain data redundancy. In contrast, of the visualizations in the bottom third of description quality ranks, only 22% of visualizations contain message redundancy and 12% contain data redundancy. These trends hold within each of the source categories (see Table 5) in that the mean description quality tends to be better for visualizations with message/data redundancy than without, and similarly, the visualizations with the best descriptions (the top third) are more likely to contain both forms of redundancy.

Overall, if we compare the source categories to each other (see Table 5), infographic visualizations contain the most message redundancy, followed by news media, government, and science. The exact same trend holds for data redundancy. Thus, although no causal

Table 5. The effect of redundancy on the description quality of visualizations: across all sources, the presence of message or data redundancy is linked to higher description quality on average (* = $p < 0.01$ when comparing visualizations with and without message or data redundancy).

Source	Mean description quality (0-3 scale)					% with msg redundancy			% with data redundancy		
	All images	Message redundancy		Data redundancy		All images	Top 1/3 (good)	Bottom 1/3 (bad)	All images	Top 1/3 (good)	Bottom 1/3 (bad)
		With	Without	With	Without						
Overall	1.77	1.99*	1.59*	2.01*	1.70*	39%	57%	22%	22%	34%	12%
Infographics	2.09	2.11	2.06	2.16	2.05	63%	69%	58%	37%	38%	27%
News	1.99	2.05	1.90	2.02	1.97	50%	51%	43%	27%	26%	26%
Government	1.37	1.41	1.36	1.40	1.37	28%	25%	21%	20%	21%	8%
Science	1.24	1.60*	1.19*	NA	1.24	7%	13%	0%	0%	0%	0%

conclusions can be made at this point, we can see that redundancy is related to how well a visualization is recognized and recalled. The importance of data and message redundancy has also been observed for animated and video visualizations [2]. **Data and message redundancy in visualizations help people grasp on to the main trends and messages, and improves visualization recognition and recall.**

7 CONCLUSIONS

Based on the results presented in the preceding sections, we summarize below the key observations from our study. In addition to our experimental results shedding light on the fundamental theory of how visualizations are encoded, recognized, and recalled from memory, our results also provide direct quantitative empirical evidence in support of many conventional established qualitative visualization design and presentation guidelines. The conclusions listed below, with related established design principles where appropriate, relate to having a good and clear presentation, making effective use of text and annotations, drawing a viewer's attention to the important details, providing effective visual hooks for recall, and guiding the viewer through a visualization using effective composition and visual narrative.

Visualizations that are memorable "at-a-glance" have memorable content. Visualizations that are most memorable "at-a-glance" are those that can be quickly retrieved from memory (i.e., require less eye movements to recognize the visualization). Importantly, when these visualizations are retrieved from memory, many details of the visualization are retrieved as well. Thus, participant-generated descriptions tend to be higher quality for these visualizations (Sec. 6.1).

Titles and text are key elements in a visualization and help recall the message. Titles and text attract people's attention, are dwelled upon during encoding, and correspondingly contribute to recognition and recall. People spend the most amount of time looking at the text in a visualization, and more specifically, the title. If a title is not present, or is in an unexpected location (i.e., not at the top of the visualization), other textual elements receive attention. As exhibited by these results, the content of a title has a significant impact on what a person will take away from, and later recall, about a visualization (Sec. 6.2).

"Words on graphics are data-ink. It is nearly always helpful to write little messages on the plotting field to explain the data, to label outliers and interesting data points." (Edward Tufte [44])

Pictograms do not hinder the memory or understanding of a visualization. Visualizations that contain pictograms tend to be better recognized and described. Pictograms can often serve as visual hooks into memory, allowing a visualization to be retrieved from memory more effectively. If designed well, pictograms can help convey the message of the visualization, as an alternative, and addition to text (Sec. 6.2).

"The same ink should often serve more than one graphical purpose. A graphical element may carry data information and also perform a design function usually left to non-data-ink. Or it might show several different pieces of data. Such multi-functioning graphical elements, if designed with care and subtlety, can effectively display complex, multivariate data." (Edward Tufte [44])

Redundancy helps with visualization recall and understanding. When redundancy is present, to communicate quantitative values (data redundancy) or the main trends or concepts of a visualization (message redundancy), the data is presented more clearly as measured through better-quality descriptions and a better understanding of the message of the visualization at recall (Sec. 6.3).

"Redundancy, upon occasion, has its uses; giving a context and order to complexity, facilitating comparisons over various parts of the data, perhaps creating an aesthetic balance." (Edward Tufte [44])

"Telling things once is often not enough: redundancy helps restore messages damaged by noise." (Jean-Luc Doumont [14])

All of these findings come down to the following well-phrased communication advice:

"Effective communication is getting messages across. Thus it implies someone else: it is about an audience, and it suggests that we get this audience to understand something. To ensure that they understand it, we must first get them to pay attention. In turn, getting them to understand is usually nothing but a means to an end: we may want them to remember the material communicated, be convinced of it, or ultimately, act or at least be able to act on the basis of it." (Jean-Luc Doumont [14])

The previous paper on visualization memorability [8] presented findings on which visualizations are most memorable and which are most forgettable, when participants are only given 1 second to encode each visualization. We were able to show that this "at-a-glance" memorability is very consistent across participants, and thus likely a property of the visualizations themselves. In this paper, we extended encoding to give participants enough time to process the content (textual components, messages, and trends) of a visualization. We measured memorability and analyzed the information that participants were able to recall about the visualizations. Participants remembered and forgot the same visualizations whether given 1 or 10 seconds for encoding, and more importantly, participants better recalled the main message/content of the visualizations that were more memorable "at-a-glance". Thus, to get people to understand a visualization, *"we must first get them to pay attention"* - memorability is one way to get there. Then, with care, one can design and introduce as necessary text, pictograms, and redundancy into the visualization so as to get the messages across.

REFERENCES

[1] B. E. Alper, N. Henry Riche, and T. Hollerer. Structuring the space: a study on enriching node-link diagrams with visual references. In *CHI '14*, pages 1825–1834. ACM, 2014.

[2] F. Amini, N. H. Riche, B. Lee, C. Hurter, and P. Irani. Understanding data videos: Looking at narrative visualization through the cinematography lens. In *CHI '15*, pages 1459–1468. ACM, 2015.

[3] W. A. Bainbridge, P. Isola, and A. Oliva. The intrinsic memorability of face photographs. *J. of Exp. Psychology: General*, 142(4):1323, 2013.

[4] S. Bateman, R. L. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *CHI '10*, pages 2573–2582, 2010.

[5] A. F. Blackwell and T. Green. Does metaphor increase visual language usability? In *Visual Languages*, pages 246–253. IEEE, 1999.

[6] 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.

[7] R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Reppa, L. Floridi, and M. Chen. An empirical study on using visual embellishments in visualization. *IEEE TVCG*, 18(12):2759–2768, 2012.

[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] T. F. Brady, T. Konkle, and G. A. Alvarez. A review of visual memory capacity: Beyond individual items and toward structured representations. *J. of Vision*, 11(5):1–34, 2011.

[10] T. F. Brady, T. Konkle, G. A. Alvarez, and A. Oliva. Visual long-term memory has a massive storage capacity for object details. *PNAS*, 105(38):14325–14329, 2008.

[11] 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, Dec 2011.

[12] Z. Bylinskii, P. Isola, C. Bainbridge, A. Torralba, and A. Oliva. Intrinsic and extrinsic effects on image memorability. *Vision Research*, 2015. (In Press).

[13] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. of the Am. Stat. Assoc.*, 79(387):531–554, 1984.

[14] J. Doumont. *Trees, Maps, and Theorems: Effective Communication for Rational Minds*. Principiae, 2009.

[15] S. Few. The chartjunk debate: A close examination of recent findings. Visual Business Intelligence Newsletter, 2011.

[16] S. Ghani and N. Elmqvist. Improving revisitation in graphs through static spatial features. In *Proceedings of Graphics Interface 2011*, pages 175–182. Canadian Human-Computer Communications Society, 2011.

[17] J. H. Goldberg and J. I. Helfman. Comparing information graphics: A critical look at eye tracking. In *BELIV'10*, pages 71–78. ACM, 2010.

[18] S. Haroz, R. Kosara, and S. L. Franconeri. Isotype visualization–working memory, performance, and engagement with pictographs. In *CHI '15*, pages 1191–1200. ACM, 2015.

[19] S. Haroz and D. Whitney. How capacity limits of attention influence information visualization effectiveness. *IEEE TVCG*, 18(12):2402–2410, 2012.

[20] W. Huang. Using eye tracking to investigate graph layout effects. In *APVIS '07*, pages 97–100, Feb 2007.

[21] W. Huang, P. Eades, and S.-H. Hong. A graph reading behavior: Geodesic-path tendency. In *PacificVis '09*, pages 137–144, April 2009.

[22] J. Hullman, E. Adar, and P. Shah. Benefitting infovis with visual difficulties. *IEEE TVCG*, 17(12):2213–2222, 2011.

[23] P. Isola, J. Xiao, D. Parikh, A. Torralba, and A. Oliva. What Makes a Photograph Memorable? *PAMI*, 36(7):1469–1482, 2014.

[24] L. L. Jacoby. A process dissociation framework: Separating automatic from intentional uses of memory. *J. of Memory and Language*, 30:513–541, 1991.

[25] 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, Dec 2012.

[26] P. Klaver, J. Fell, T. Dietl, S. Schür, C. Schaller, C. Elger, and G. Fernández. Word imageability affects the hippocampus in recognition memory. *Hippocampus*, 15(6):704–712, 2005.

[27] T. Konkle, T. F. Brady, G. A. Alvarez, and A. Oliva. Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *J. of Exp. Psychology: General*, 139(3):558–578, 2010.

[28] C. Körner. Eye movements reveal distinct search and reasoning processes in comprehension of complex graphs. *Applied Cognitive Psychology*, 25(6):893–905, 2011.

[29] S. M. Kosslyn. Understanding charts and graphs. *Applied Cognitive Psychology*, 3(3):185–225, 1989.

[30] J. P. Leboe and B. W. A. Whittlesea. The inferential basis of familiarity and recall: Evidence for a common underlying process. *J. of Memory and Language*, 46:804–829, 2002.

[31] H. Li and N. Moacdieh. Is chart junk useful? an extended examination of visual embellishment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 58, pages 1516–1520. SAGE Publications, 2014.

[32] H. R. Lipford, F. Stukes, W. Dou, M. E. Hawkins, and R. Chang. Helping users recall their reasoning process. In *2010 IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 187–194. IEEE, 2010.

[33] K. Mahowald, P. Isola, E. Fedorenko, A. Oliva, and E. Gibson. Memorable words are monogamous: An information-theoretic account of word memorability. 2015. (Under Review).

[34] G. Mandler. Recognizing: The judgment of previous occurrence. *Psychological Review*, 87:252–271, 1980.

[35] K. Marriott, H. Purchase, M. Wybrow, and C. Goncu. Memorability of visual features in network diagrams. *IEEE TVCG*, 18(12):2477–2485, 2012.

[36] A. Paivio. Mental imagery in associative learning and memory. *Psychological Review*, 76(3), 1969.

[37] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini. How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques. In *CHI '15*, pages 15–03, 2015.

[38] S. Pinker. A theory of graph comprehension. *Artificial intelligence and the future of testing*, pages 73–126, 1990.

[39] 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.

[40] M. Raschke, T. Blascheck, M. Richter, T. Agapkin, and T. Ertl. Visual analysis of perceptual and cognitive processes. In *IJCV*, 2014.

[41] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: a database and web-based tool for image annotation. *International J. of Computer Vision*, 77(1-3):157–173, 2008.

[42] H. Siirtola, T. Laivo, T. Heimonen, and K.-J. Raiha. Visual perception of parallel coordinate visualizations. In *Intern. Conf. on Information Visualisation*, pages 3–9, July 2009.

[43] D. Skau, L. Harrison, and R. Kosara. An evaluation of the impact of visual embellishments in bar charts. In *Proceedings of EuroVis*, volume 2015, 2015.

[44] E. Tufte. *The Visual Display of Quantitative Information*. Cheshire (Conn.), 2001.

[45] E. Tulving. Memory and consciousness. *Canadian Psychology*, 26:1–12, 1985.

[46] A. Vande Moere, M. Tomitsch, C. Wimmer, B. Christoph, and T. Grechenig. Evaluating the effect of style in information visualization. *IEEE TVCG*, 18(12):2739–2748, 2012.

[47] S. Vogt and S. Magnussen. Long-term memory for 400 pictures on a common theme. *Experimental Psychology*, 54(4):298–303, 2007.

[48] I. Walker and C. Hulme. Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *J. of Exp. Psychology: Learning, Memory, and Cognition*, 25(5), 1999.