

Image or Information? Examining the Nature and Impact of Visualization Perceptual Classification

Anjana Arunkumar, Lace Padilla, Gi-Yeul Bae, and Chris Bryan

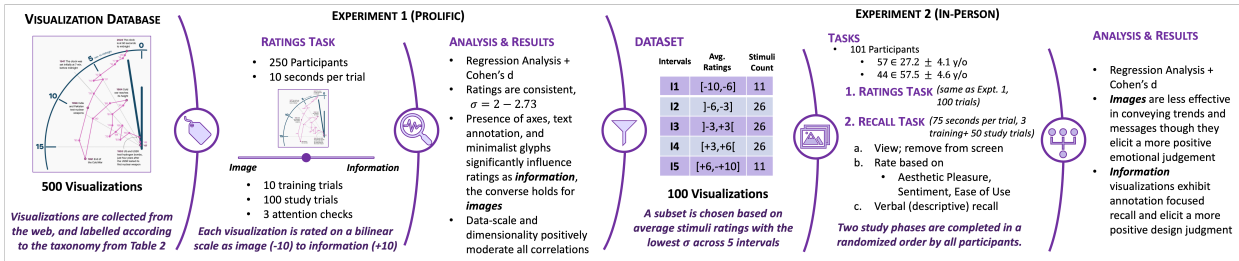


Fig. 1: Illustrative diagram of the experiments: 500 visualizations are labeled and categorized. In Experiment 1, *image/information* ratings are gathered, and regression analysis and Cohen's d are used to determine the impact of different design elements of the charts on the ratings. In Experiment 2, a representative subset of 100 visualizations is curated, and two tasks are performed: *image/information* ratings and free recall; analysis compares the impact of ratings on the nature of the recall.

Abstract— How do people internalize visualizations: as *images* or *information*? In this study, we investigate the nature of internalization for visualizations (i.e., how the mind encodes visualizations in memory) and how memory encoding affects its retrieval. This exploratory work examines the influence of various design elements on a user's perception of a chart. Specifically, which design elements lead to perceptions of visualization as an *image* (aims to provide visual references, evoke emotions, express creativity, and inspire philosophic thought) or as *information* (aims to present complex data, information, or ideas concisely and promote analytical thinking)? Understanding how design elements contribute to viewers perceiving a visualization more as an image or information will help designers decide which elements to include to achieve their communication goals. For this study, we annotated 500 visualizations and analyzed the responses of 250 online participants, who rated the visualizations on a bilinear scale as 'image' or 'information.' We then conducted an in-person study ($n = 101$) using a free recall task to examine how the image/information ratings and design elements impacted memory. The results revealed several interesting findings: Image-rated visualizations were perceived as more aesthetically 'appealing,' 'enjoyable,' and 'pleasing.' Information-rated visualizations were perceived as less 'difficult to understand' and more aesthetically 'likable' and 'nice,' though participants expressed higher 'positive' sentiment when viewing image-rated visualizations and felt less 'guided to a conclusion.' The presence of axes and text annotations heavily influenced the likelihood of participants rating the visualization as 'information.' We also found different patterns among participants that were older. Importantly, we show that visualizations internalized as 'images' are less effective in conveying trends and messages, though they elicit a more positive emotional judgment, while 'informative' visualizations exhibit annotation focused recall and elicit a more positive design judgment. We discuss the implications of this dissociation between aesthetic pleasure and perceived ease of use in visualization design.

Index Terms—Information Visualization; Human-Centered Computing; Perception & Cognition; Takeaways.

1 INTRODUCTION

Communicative visualizations are rhetorical devices that inform and convince an audience of an idea by amplifying cognition [20]. These represent the bulk of exposure any individual has to visualizations [1]. Scholars commonly define the *purpose* of a communicative chart in terms of the *intents* of a visualization designer [6]; these can be modeled as learning objectives, which can be *cognitive* (analytical task performed) or *affective* (emotional reaction elicited) [57] and dictate design choices.

Prior work has extensively examined the design elements that make some visualizations intrinsically more *memorable* than others

across charts in the wild [9, 10, 32]. Concurrently, there have been attempts to define and evaluate higher-level concepts like *engagement* [11, 35, 61, 64] and *effectiveness* [4, 17, 83] across a plethora of (often synthetic and minimalist) chart types. However, the difference in the nature of charts studied through these different lenses poses challenges in reconciling good empirical results and rules of thumb with good exemplars of real-world visualizations. To disentangle these confounding factors, we set out to answer a fundamental question: *how do people actually internalize visualizations: as image or information?* We define *images* as seeking to provide visual references, evoke emotions, express creativity, and inspire philosophic thought [21, 47, 54, 55, 63]. *Information* visualization, on the other hand, focuses on presenting complex data, information, or ideas in a clear and concise manner, and promotes analytical thinking [10, 16, 30, 53, 76, 81] (discussion of definition in Section 2.2). While both use visual elements to convey meaning, they differ in their purpose and design.

In this paper, we present the results of two experiments (see the experimental design in Fig.1). We first annotated a dataset of 500 visualizations used "in the wild" across social and scientific domains based on a taxonomy of static visualizations derived from literature [10, 14] (see Table 2). We then presented each visualization for 10 seconds to be rated on a bilinear scale as 'image' or 'information' (see Fig.2 for definitions) by 250 online subjects. Based on response analysis, we took

• Anjana Arunkumar, Gi-Yeul Bae, and Chris Bryan are with Arizona State University. Email: aarunku5@asu.edu, giyeulbae@asu.edu, cbryan16@asu.edu.

• Lace Padilla is with Northeastern University, Email: l.padilla@northeastern.edu.

a representative sample of 100 visualizations from our database. We investigated the agreement between externalized ratings and internalized memory through a free recall task in a within-subjects experiment for 101 in-person subjects in two different age groups (20-35 y/o and 50+y/o). At a high level, the study results indicate consistency across age groups in rating patterns for the first experiment. For the second experiment, we observe similar descriptive patterns in the framing and focus of free recall verbiage for similar externalized rating values. Notably, we see that the presence of axes and text annotations heavily influence the nature of internalization.

Contributions: This work represents the first study that systematically investigates which design elements of visualizations dictate the nature of internalization. To promote reproducibility, all study materials, visualization labeling and metadata, demographic details of study participants, study analysis and results, are publicly available at: <https://github.com/aarunku5/Image-or-Information-Vis-2023> (Note: anonymous repository is being used for review round). Based on the results of our experiment, we can offer quantitative evidence in direct support of several existing conventional qualitative visualization design guidelines promoting ease of use, including: (1) use of appropriate text annotation, (2) minimalist design style, and (3) presence of axes. We further identify a disconnect between aesthetic pleasure and perceived ease of use in visualization design, where *image-like* visualizations may be less effective in conveying trends and messages, even though they tend to elicit a more positive emotional response. In contrast, *informative* visualizations include more text, exhibit annotation-focused recall, and receive a more favorable design evaluation.

2 BACKGROUND

2.1 Perception and Memorability of Visualizations

Broadly, the topic of graphical perception is concerned with how visualizations are perceived and interpreted [23]. Perception and readability depend on many factors, including the visual encodings being used (i.e., the marks and channels), the number of data dimensions being encoded [45, 65, 67], the amount of data shown [49], how the chart is styled [52], what rhetorical elements are present [44], and even the current cognitive focus of the person viewing the chart [38].

Previous work has focused on understanding the role of such visualization design choices on memorability and comprehensibility across a wide variety of charts [31, 43, 46, 75]. Additionally, several studies aiming to evaluate the impact of embellishments on visualization memorability and comprehension [5, 8] have been linked to a “slow analytics” movement that encourages ownership and retention of analytical tasks rather than precision [12, 59]. Studies have demonstrated that some visualization types and visual elements are more memorable than others; for example, human-recognizable objects and color can increase the memorability of a visualization [9, 10].

In this study, we aim to move beyond memorability and investigate the nature of internalization for visualizations, i.e., how is a visualization encoded in memory (image/information rating), and how does the nature of memory encoding affect its retrieval (free recall verbiage patterns)? By identifying the most dominant features for the nature of rating and recall, designers can better understand how to reconcile empirical design guidelines with stylistic embellishments to improve the communicative efficacy of visualizations.

2.2 Communicative Visualizations

Scholars often frame communicative visualizations in the context of “learning about data” [2, 13, 29], embedded in an analytical workflow of identifying and presenting an insight [40, 78]. Research and practices focused on cognitive efficiency recommend the best visualization for accurate message decodings, such as generalized systems like APT and Tableau, which prioritize design based on specifications of the data of interest or broad analytical targets [51, 60, 68]. However, in addition to visualizing data to support a claim accurately, design choices can appeal to emotion to persuade audiences to believe the data [1, 15, 25]. For instance, researchers have utilized visualizations to bring attention

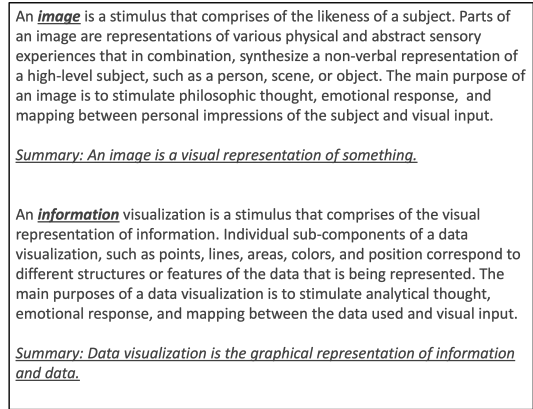


Fig. 2: Constructed definitions of *image* and *information*.

to and provoke action for many issues [56, 70]. Designers can intentionally (or unintentionally) use conventions, such as minimalist aesthetics, the inclusion of data sources, or the increased use of text, to create an illusion of increased objectivity and transparency in such data representations for the audience [28, 50]. Notably, visual literacy plays a crucial role in determining the effectiveness of a communicative visualization in achieving such learning objectives for its target audience [57].

We construct definitions to clarify the nuance between ‘image’ and ‘information’ type visualizations (see Fig. 2) in the context of this communicative framework. Although both use visual components to communicate a message, their purpose and style vary. Our definitions are curated from a synthesis of literature [3, 10, 16, 21, 30, 47, 53–55, 63, 76, 81], and consider: (i) *design*: images often prioritize aesthetics, composition, and visual impact, while information is typically presented in a structured, organized manner to facilitate understanding and retention, (ii) *purpose*: images visually communicate an idea or evoke emotional engagement, while information educates, informs, or instructs the viewer, and (iii) *response*: images seek to elicit visceral, emotional responses and rely on intuitive processing, while information facilitates the viewer’s analysis/understanding of complex topics, and promotes long-term recall.

2.3 Infographics and Relationships with Data Visualization

Infographics are a form of data representation primarily intended for effective communication. In comparison, data visualizations can be used not just to communicate but also for exploration, discovery, and analysis, among other tasks [18, 19]. Outside the visualization community, infographics are used to promote audience engagement [42] and comprehension of information [27], particularly for individuals with low-to-medium graph literacy [18, 33].

Past research has demonstrated increased memorability of infographics compared to unembellished charts [5, 9, 10], with a more accurate short-term recall of chart message/data points [32, 36] for colorful, non-complex representations. Some studies find infographics are less effective than other narrative visualization techniques like data comics [66, 79] and primarily serve as a curated exploration method rather than an analytical tool. Additionally, studies have shown that the internal memory of a visualization is influenced by a viewer’s impression of a canonical representation of that visualization type [32, 77].

The visualizations used in our study are within the purview of static infographics (with varying levels of text annotation). We analyze which design elements in such infographics detract the most from perceived ease of use, i.e., when the visualization is rated as highly image-like, for study stimuli and their relationship with aesthetic appreciation.

2.4 Visualization Aesthetics

Prior work across domains like advertising and market design has demonstrated that high ratings of aesthetic quality are associated with greater visual exploration and perceived usability [62, 74]. However, these works use stimuli that cannot be categorized as infographics or

Source	Websites	Topics	Total (500/100)
Government	Department of Energy, Los Angeles Planning Commission, Nasa, The Struggle for Five Years in Four , Washington Census, White House Office of Management and Budget, US Treasury Dept., World Health Organization (WHO)	Demographics, Economics, Environment, Health, Power, Traffic, Space	28/5
News	BBC, Bloomberg, Chicago Tribune, Economist, Guardian, Huff Post, Microsoft, NBC, National Post, New York Times, Quartz, Sports Reference, Visual Capitalist, Wall Street Journal, Washington Post	Demographics, Economics, Food, Geography, Geology, Health, Language, Politics, Pop Culture, Power, Resources, Space, Sports, Technology, Traffic, Weather	112/22
Media	3CS.CH, Beautiful Evidence, Behance, Bike MS, DataViz Catalogue, Envisioning Information, Exploring Data, Facebook, FiftyThirtyEight, Flickr, Forbes, Fortune, Fundamentals of Data Visualization, Golreland, Hive Systems, Mapbox, NFL, NPM-JS, Neon, Observable, Reuters, Selfiecity, Tech in Review, The Pudding, Velir, Vengage, Verilogue, Visual Capitalist	Demographics, Economics, Environment, Food, Geography, Geology, Health, History, Language, Lifestyle, Miscellaneous, Politics, Pop Culture, Power, Resources, Science, Sports, Technology, Traffic, Weather	52/15
Blogs	Adventures in Mapping, Flowing Data, Georgia Lupi, Matt Mckoon, Midwest Uncertainty Collective, Multiple Views, Nicolas Rapp, Nightingale, R-Graph Gallery, Towards Data Science, Vis Guides, Visual Cinnamon	Demographics, Economics, Environment, Food, Geography, Geology, History, Health, Language, Lifestyle, Military, Politics, Pop Culture, Power, Resources, Science, Space, Sports, Technology, Traffic, Weather	86/21
Social Media	Flickr, Pinterest, Reddit, Twitter	Demographics, Economics, Environment, Food, Geography, Geology, Health, History, Language, Lifestyle, Miscellaneous, Politics, Pop Culture, Power, Resources, Sports, Technology, Traffic, Weather	56/13
Infographics	Beautiful News, Visual.ly, Information is Beautiful, Tableau Public	Culture, Demographics, Economics, Environment, Food, Geography, Geology, Health, History, Language, Lifestyle, Miscellaneous, Politics, Pop Culture, Power, Resources, Science, Sports, Technology, Traffic, Weather	137/22
Scientific Publications	IEEE, Industrial Ecology, Nature, OpenCellID, Oxford, Perspecta, R-chie, University of Santa Clara, Vassar	Demographics, Economics, Geography, Geology, History, Military, Resources, Science, Technology, Traffic, Weather	29/2

Table 1: List of visualization sources, their topics, and the respective number of visualizations from the full corpus (500) / subset (100).

visualizations, though they may comprise part of such elements. In the visualization context, *aesthetics* refers to a quality or characteristic of a visual representation distinct from how clear, informative, or memorable it is. An alternate definition refers to the visual appeal or beauty of the representation [26, 37]. We seek to systematically evaluate the applicability of the premise that beauty and functionality are intrinsically intertwined [5, 10, 22, 38, 39, 41] for visualization. Is what is useful often beautiful? Is what is beautiful often useful? In our free recall task, we ask participants to explicitly rate aspects of aesthetic pleasure based on emotional response and design quality considerations using the Beauvis scale [37] (see Fig. 4b). We compare these ratings with the image/information scores for stimuli, as well as ratings on perceived ease of use (i.e., participants’ ‘experiencing difficulty in understanding’ and ‘feeling guided to a conclusion’ by stimuli).

2.5 Visualization Engagement

Visualization researchers have offered varying definitions of *user engagement*. Some define user engagement as the willingness of users to invest effort in exploring a visualization to gain more information [11, 35], while others define it as the perceived effectiveness of visualizations in terms of participation or use [64]. Different aesthetic styles of visualization, such as ‘sketchy’ [82], ‘magazine’ [64], ‘pictographic’ [35], and ‘analytical’ [64], and interaction measures such as hover [11], clicks [35], and willingness to annotate [82] have been analyzed for user engagement. However, factors such as users’ goals, interest, and familiarity of the data, as well as accessibility and display parameters, can affect their willingness to explore the visualizations and also eventually determine the insights they gain [84].

Mahyar et al. [61] propose a five-level taxonomy for engagement (Expose–Involve–Analyze–Synthesize–Decide) based on prior literature [7] which posits that the degree of engagement increases as a user performs higher-level cognitive tasks such as synthesizing information, and making final decisions. To further ground this work in current visualization efforts, we measure the agreement between participants’ internalization ratings (i.e., *image/information?*) and the elements of visualizations they focus on during free recall (i.e., what is mentioned and what is emphasized) in the context of this taxonomy.

3 STIMULI OVERVIEW

To obtain many real-world examples encompassing a breadth of visualization types, design aesthetics, and domains, we initially collected 500 data visualizations (for Experiment 1) from multiple sources, as

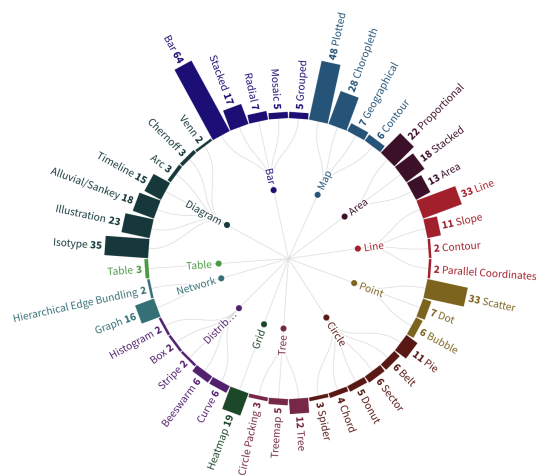


Fig. 3: Count for different visualization types in the database.

shown in Table 1 and Fig. 3. These particular sources were chosen in line with data collection efforts like MASSVIS.¹ In total, we identified 460 *single* visualizations, i.e., stand-alone visualizations with one panel, and 40 *multi-panel* visualizations, that contain multiple related visualizations as part of a single narrative. We note that some *single* visualizations may use other visualizations as glyphs (for instance, maps can be overlaid with pies/bars/networks across various geographic regions of interest).

A higher proportion of the visualizations collected are from infographic sources or social media. These sites are more likely to be accessed and encountered by non-expert users; such sources have more ‘diagrams.’ Skeuomorphism is also more commonly seen in these sources, where the chart’s overall structure or individual elements mimic real-world counterparts relevant to its topic/data domain. ‘Maps’ are highly represented across all sources except for scientific publications, which, along with government sources, use basic visual encoding techniques, such as line graphs, bar charts, and point plots. Overall, infographic and scientific sources have a relatively larger proportion of *multi-panel* charts compared to other sources.

¹<http://massvis.mit.edu/>.

Properties		Possible Values
Dimension	the number of variables being represented	low (<3) / medium (3-4) / high (>4)
Multiplicity	is the visualization stand-alone or somehow grouped with other visualizations	single/multi-panel
Warping	is the visualization radial in nature or have curvilinearly styled axes	y/n
Temporal	is the visualization representing time-series data	y/n
Pictorial	encoding is a pictogram; individual pictograms represent units of data such as in Isotype visualizations	y/n
Attributes		Possible Values
Black and White	is the visualization only in gray-scale	y/n
Number of Distinct Colors	how many distinct color hues are present	2-7
Data Ink Ratio	the ratio of data to non-data elements	low/medium/high
Visual Density	overall density of visual elements in the image without distinguishing between data and non-data elements	low/medium/high
Skeuomorphism	includes some element or has an overall appearance which resembles a real-world object or aligns with the broad topic of the chart, though its rendering may not be photo-realistic	y/n
Human Recognizable Objects (HRO)	includes skeuomorphic glyphs/annotations, logos, symbols	y/n
Human Depiction (HD)	presence of a photo-realistic or skeuomorphic representation of a human	y/n
Glyph/Overlay	presence of a glyph/overlay element in the visualization	circles/squares/planes/food/etc.
Ordering	how are visualization elements ordered, if applicable	increasing/decreasing/temporal/random/nan
3D	is the visualization depicted as a 3-dimensional figure	y/n
Data Volume	what is the scale of data being visualized	low (<10) / medium (<100) / high (>100)
Background Color	what kind of background color hue is present	white/dark/light
Text Volume	what is the volume of text present in the visualization (based on word count)	low (<10) / medium (<25) / high (>25)
Data Source	is the data source specified within the visualization	y/n
Photo-Realism	includes some element which depicts a real-world object with the exactness of a photograph	y/n
Sentiment	what is the emotional response evoked on viewing	positive/negative/neutral
Purpose	what is the level of engagement expected from a viewer	expose/involve/analyze
Presence/Absence: Axes, Title, Key, Caption, Aggregate Data, Data Redundancy, Message Redundancy, Grid Lines, Annotations of Interest (AOI) - Text, Lines, Arrows, Highlights		

Table 2: Summary of classified chart elements, properties, attributes.

We manually labeled visualizations using 5 ‘expert’ annotators (graduate students with 3+ years of research experience in data visualization) based on the visual taxonomy developed by Borkin et al. [10], which classifies visualizations based on underlying data structures, visual encodings used, and the perceptual tasks enabled by these encodings. All coders annotated the full corpus, and majority coder agreement was used for cases of disagreement Metadata and labeling information for the dataset can be found in the anonymized repository that hosts supplemental material. Additionally, our experts annotated the stimuli based on a set of constraints that may apply to any of the visualization types (see Table 2), drawing on prior work [10, 14].

4 EXPERIMENT OVERVIEW

4.1 Set-up and Participants

Experiment 1: As discussed in Section 3, we annotated 500 visualizations. We resized these while preserving aspect ratios to ensure their maximum dimension was 1000 pixels. For each trial, participants were shown a visualization for 10 seconds. They were asked to perform a rating task where they positioned a slider on a bilinear scale ranging from -10 (Image) to +10 (Information), as shown in Fig. 4a. Each participant completed ten training trials, followed by 100 session trials. The 100 stimuli were randomly selected from the dataset of 500 visualizations, such that no chart appears twice during a session. An attention check appeared after the 25th, 50th, and 75th questions. Before running the main study, we conducted a pilot study with three participants to validate the design. We recruited 265 participants on Prolific² using the following user filters: (i) self-reported normal or corrected-to-normal vision, (ii) a first language of English, (iii) located in the U.S./U.K. with at least a high-school level of education, (iv) at least 100 previously completed Prolific surveys, and (v) an approval rate of 90% or more.³ Participants were paid \$3.00 for participation. Study duration averaged roughly 15:55 ($\pm 2:38$) minutes, resulting in an average hourly pay of \$12.73 ($\pm \2.11). SurveyMonkey⁴ was used to display the study and store participant responses. We excluded 15 participants as their performance indicated they did not understand the task (i.e., more than 70% of ratings had the values of 0, -10 or +10) during/after training, or they took more than 20 minutes to complete the study ($>2\sigma$). Each visualization was seen at least 50 times.

Experiment 2: We selected a subset of 100 stimuli from the dataset as exemplar visualizations. Of the exemplars, 11 are extreme examples

of ‘image’ ($\overline{score} \in [-10, -6]$), 11 are extreme examples of ‘information’ ($\overline{score} \in [+6, +10]$), and the other 78 are in-between on the image-information spectrum, sampled equally across score intervals of $]-6, -3[$, $]-3, 3[$, $[3, 6[$. Choosing them this way allows us to measure the effects of different visual elements while minimizing bias. The exemplar visualizations chosen exhibited the highest levels of inter-participant agreement within their respective score-interval, i.e., for each interval, charts with the lowest σ for their average scores were picked (see Fig. 6). Thus, the exemplar population comprises the best-fitting visualizations for each interval and is representative of a real-world perceptual distribution. Participants perform two sets of tasks in succession:⁵ (i) *Rating*: This task set-up is identical to that used in Experiment 1. (ii) *Free Recall*: For each trial, participants viewed a stimulus for 30 seconds. Then, the screen was cleared, and participants were given 15 seconds to explicitly report aesthetic pleasure, perceived ease of use, and sentiment about the visualization. Finally, participants were asked to orally describe what they remember and feel about the visualization they viewed in the trial for 30 seconds. Each participant completed 50 trials, where each trial randomly selected one chart. Three training trials were performed before beginning the recall phase of the experiment. Similar to Experiment 1, a pilot study was first conducted to validate the design with three users. We recruited 101 participants from local communities belonging to two distinct groups. College-aged participants were recruited from a local university, and older participants were recruited from residential complexes in the locality via the leasing offices.⁶ Our data was collected over a period of 4 months.³ We recruited 57 college-aged (27.2 ± 4.1 y/o) participants educated in STEM and 44 older participants (57.5 ± 4.6 y/o) who are high-school (minimum) educated. All participants had normal color vision. Participants completed both phases of the experiment over a session lasting about 75 minutes. Participants volunteered their time in the study. The *rating* and *recall* phases were randomly ordered across participants in order to minimize effects of order bias. Each of the 100 visualizations was viewed at least once in the ratings task; 50 stimuli were seen for a second time in the recall task.

4.2 Data Analysis

Ratings: We conducted a set of analyses for both experiments, comparing the rating scores to various visualization attributes (elicited based on prior research and the taxonomy (Table 2) used to categorize the visualizations). The visualizations of the data analysis were constructed by

²<https://www.prolific.co>

³See supplemental material for further demographic details.

⁴<https://www.surveymonkey.com>

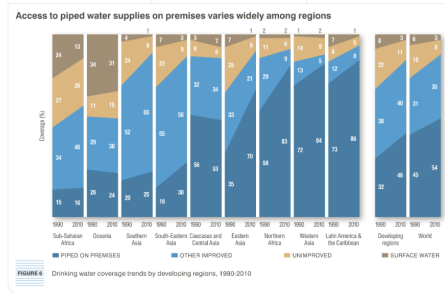
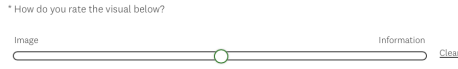
⁵All 100 stimuli are rated first; on completion, participants then perform the recall task on 50 randomly selected stimuli.

⁶Anonymized for the review stage

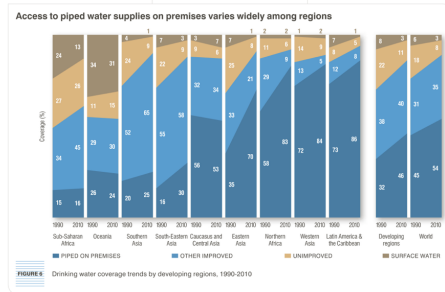
summarizing across all visualizations. Our initial hypothesis was that different visualizations produce distinct internalizations. We further hypothesized that specific visual elements lead to diverging internalizations. However, as this work was exploratory, we did not have explicit predictions about which visual elements would produce various internalizations. We performed regression analyses [34, 48] and calculated Cohen’s d to determine the effect strength of each factor (i.e., visual features) [24] to preserve statistical power during analyses.⁷

power of visualizations had a high volume of results focused on surveying college-aged participants. One of the rare exceptions to this is Peck et al. [69], who explicitly recruited a subset of participants that were older in age (they also considered these age brackets in their analyses). We follow a similar approach here, recruiting both college-aged students and also an older group of participants, to determine if age was a significant factor in rating patterns.

Recall: The explicit ratings for aesthetic pleasure, perceived ease of use, and sentiment were compared to the internalization scores using regression analyses. The free recall was analyzed using the engagement taxonomy proposed by Mahyar et al. [61]. Higher levels of engagement include the formulation of a hypothesis based on the patterns and trends observed in the data. This hypothesis can be an initial assumption or idea about what the data may represent or indicate, which can then be further explored or tested. Subsequently, decision-making can be performed after validating the hypothesis. The formation of a hypothesis when viewing a visualization is often influenced by the viewer’s prior knowledge and experience with the subject matter, as well as their cognitive biases and assumptions. Since participants viewed each stimulus for up to 30 seconds (recall) and also do not necessarily have domain expertise, we do not consider formulation and evaluation of hypotheses as contextual elements, as the quality of hypotheses has the potential to significantly vary between individuals.



(a) Ratings Task



To what extent do you agree that this visual representation is:

	1	2	3	4	5
Enjoyable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Likeable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pleasing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Appealing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

This visual representation is:

Negative Neutral Positive

To what extent do you agree that this visual representation

	1	2	3	4	5
Guided you to a conclusion	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Was difficult to understand	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(b) Recall Task (+ Verbal Recall)

Fig. 4: Description of tasks performed in Experiments 1 and 2.

The results from the online study (Experiment 1) confirmed the existence of a distinct separation in the nature of visualization internalization and showed that there seemed to be some replicable variation in rating patterns across different age groups. However, since our online participants had high variation in (i) age, (ii) education, and (iii) each participant only rated a randomized subset of 100 stimuli from the full set of 500, we could not determine what proportion of noted patterns could be accounted for by individual differences due to small sample sizes. While formulating the online study, we noted that prior exploratory work that examined different aspects of the communicative

⁷Detailed statistics are fully reported in the supplementary material

Descriptive Elements	
Text Annotation Names	Verbatim repetition of Title, Caption, Message Arrows/Highlights/Text Labels Using Text Labels/Legend to refer to specific Marks
Structure Data Marks	Chart type, Skeuomorphism Data type/source/domain Mentioning specific Data Points without reference to Name elements
Channels Trends	Features like Color, Shape, Size, etc. Mentioning patterns observed in data without reference to Text or Name elements
Opinion	Likes/Dislikes of design choices (suitability of Structure, Mark, Channel elements)
Sentiment	Emotional Response on viewing (e.g.: happy, sad, angry, positive, negative, etc.)

Table 3: Descriptive elements to parse in free recall verbiage.

We therefore use the first three levels of Mahyar’s taxonomy to determine the focus of free recall (see Table 3): (i) Expose: the user can superficially understand what the visualization represents as a whole in terms of the source, domain, or topic of data. (ii) Involve (Interacting): the user can read and understand individual data points and visual encodings. (iii) Analyze (Finding Trends): the user can analyze the data to find trends, outliers, etc. We identified various descriptive elements to flag and parse during free recall, as shown in Table 3. The recall was manually transcribed during the study sessions; post study, 3 coders reviewed the transcripts and utterances were judged to identify which descriptive elements they could be mapped to. The presence or absence of these elements during recall, as well as the emphasis or order in which these elements are recalled, were noted.

5 EXPERIMENTAL RESULTS AND DISCUSSION

5.1 Internalization Comparisons

Our experiments primarily present baseline results to identify the design elements that cause visualizations to be internalized in memory as ‘images’ instead of ‘information’. We further examine how the nature of internalization affects (i) the process of accessing and verbally recalling the visualization in the short term, (ii) aesthetic pleasure experienced on viewing the visualization, and (iii) the perceived ease of use for a visualization at-a-glance. We do not evaluate in the context of cognitive objectives that require longer viewing time, domain expertise, interactive charts, or long-term memorability. We measure the consistency of our internalization ratings by randomly splitting participants into two independent groups for comparison and averaging over 25 such half-splits to obtain a high Spearman’s correlation of 0.82 (see Fig. 5).

We represent 30 visualizations with the lowest σ (i.e., highest rating agreement) across various score-intervals (I1–I5) in Fig. 6. Overall,

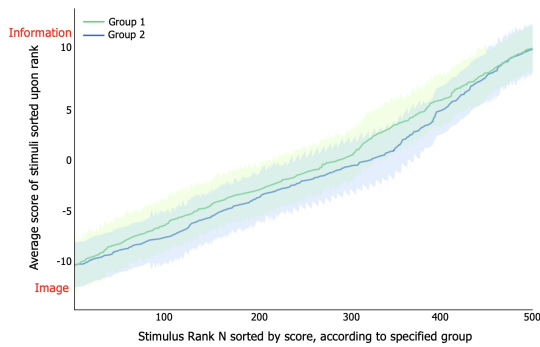


Fig. 5: Participants were split into two independent sets, Group 1 and Group 2. Visualizations were ranked by internalization scores from participants in Group 1 (green line) or Group 2 (blue line) and plotted against the overall average internalization scores given by participants from both groups. Plots are averaged across 25 such random splits. Contours depict standard deviation averaged across ratings.

there is a strong agreement in chart ratings: $\sigma \in [2 - -2.73]$. Additionally, in Experiment 2, we find that older participants exhibit higher inter-rater agreement (mean $\sigma = 2.25$) across all score intervals for our subset of 100 visualizations, in comparison to younger participants (mean $\sigma = 2.68$). We also compare the ratings between Experiments 1 and 2 for the selected stimuli subset and find that in Experiment 2, the stimuli fall in the same scoring intervals and have ratings within 0–2 points of that in Experiment 1.

5.2 The Effects of Visualization Elements on the Nature of Internalization

We consider the individual impact of each design element⁸ in our discussion below and represent average internalization scores in Figs. 7 and 8. A common trend observed in our analysis is that older participants display larger effect sizes (d values are 0.09 ± 0.03 greater on average) for individual factors, that lead them to rate visualizations in a more consistent manner when compared to younger participants.

Chart-type, Skeuomorphism, Human Recognizable Objects, Human Depiction (♥) : The internalization scores for different chart types are summarized in Fig. 8. Of our 500 target visualizations, we observe that 283 contained either skeuomorphism, photographs, or pictograms of human recognizable objects (including human depiction). Visualizations containing pictograms are on average considered more *image-like* (internalization score -6.85 ± 2.08 ⁹) than those without pictograms (internalization score 3.22 ± 3.01). This supports the observation that isotype charts and chernoff diagrams are in the more image-like categories (**I1** and **I2** in Fig. 8) ($R^2=0.84$, $d=0.92$). Charts with photo-realistic glyphs (such as the usage of real-world objects or human faces) also follow this trend ($R^2=0.82$, $d=0.83$). Area charts that rely on proportional comparison without axes also follow this trend and are found in **I1** and **I2** ($R^2=0.79$, $d=0.80$); the individual data elements can be regarded as variably sized pictograms in these visualizations.

Surprisingly, we observe that cartographic representations also feature heavily in **I1** and **I2** ($R^2=0.76$, $d=0.77$), particularly for maps that do not have overlaid text annotation ($R^2=0.80$, $d=0.81$). A potential explanation for this might be that geographical landmasses are skeuomorphic in nature, and can be treated akin to pictograms. Indeed, on further analysis, we note that skeuomorphic charts are perceived as more image-like ($R^2=0.71$, $d=0.72$). Radial warping (such as in chord diagrams, donut charts, timelines, radial bar charts, etc.) also leads to ratings between **I1** and **I2**, though with lower effect size ($R^2=0.72$, $d=0.65$). Pie charts, though radial in appearance, do not exhibit such an effect. We posit that the presence of curvature may influence this internalization of relatively unfamiliar, visually complex representations,

⁸The symbols ▲/▼ denote increase/decrease in informational value of stimuli based on element presence.

⁹All results discussed are statistically significant, with $p < 0.001$

though further investigation is needed to determine the exact cause.

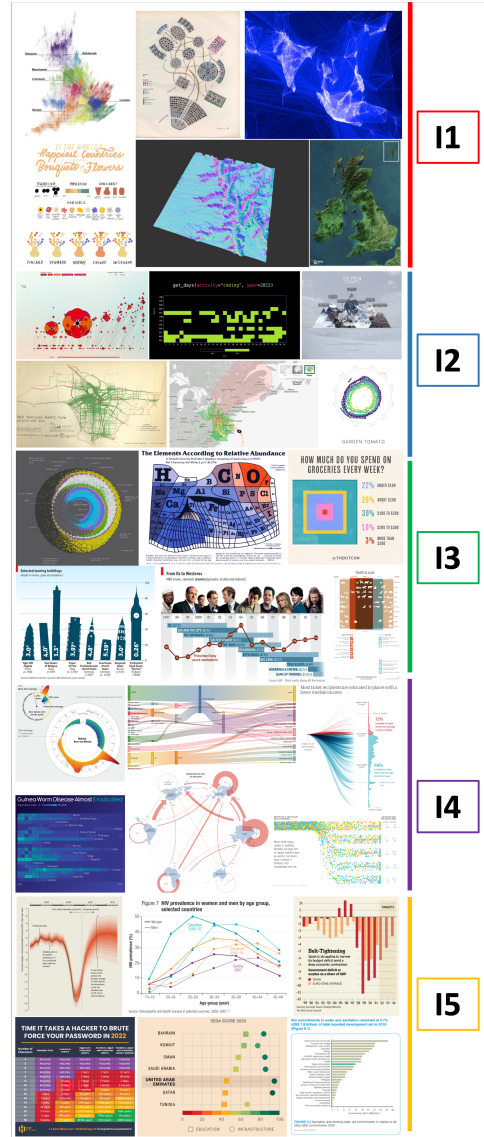


Fig. 6: The top six visualizations with the highest inter-rater agreement for different score intervals. We organized the visualizations into groups (I1–I5) with I1 being those rated as the most image-like and I5 as the most information-like.

In terms of *information* ratings, on average we find that $\text{bar} > \text{point} > \text{line} > \text{pie} > \text{area}$ charts; these visualizations are also perceived as more informative than all other visualization types in the study. However, scores exhibit wide, seemingly random variation between **I3–I5** when considering chart-type as the only determining factor. We attribute increased *information* ratings to participant familiarity with such charts, with other design elements’ presence/absence dictating the sub-categorization of these charts amongst score intervals.

Presence of Axes (▲) : There is an observable trend of visualizations with conventional axes representation (i.e., the presence of linear x and y-axes) being rated in **I3–I5** ($R^2=0.87$, $d=0.89$). The presence of tick marks and axes labels increases the scores to **I4–I5** ($R^2=0.86$, $d=0.91$) across different chart types, data domains, data scales, visual density, and data-ink ratio. Given that participants have a high familiarity with and exposure to visualizations like point (scatter/bubble-plots), line, bar, and area charts in the real world, we believe that the presence of axes intrinsically boosts the perception of stimuli as more ‘informative.’

Titles and Keys (▲) : When participants were shown stimuli with titles during the rating task, they found those visualizations to be rela-

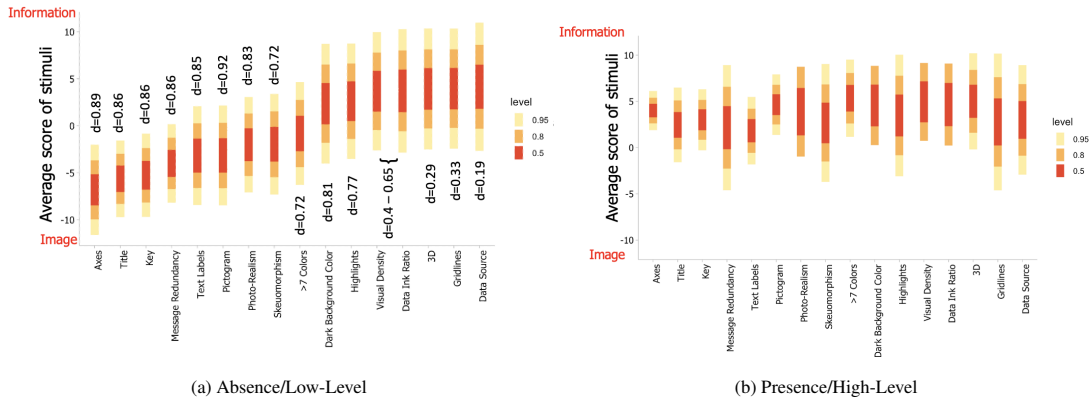


Fig. 7: Internalization scores averaged for visualizations with different visual elements. Each element is considered in isolation for its reported score, with intervals depicting the score distribution and effect size denoted using cohen’s d values.

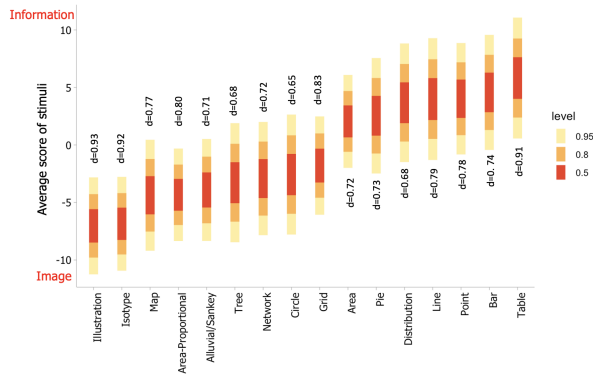


Fig. 8: Internalization scores averaged for visualizations based on visualization type. Note that a category is taken to comprise of all the chart sub-types other than the ones that have been explicitly separated with intervals depicting the score distribution and effect size denoted using cohen’s d values.

tively more informative (both within and across all five score intervals), than those lacking these titles ($R^2=0.83$, $d=0.86$). However, for charts with glyphs representing more than one dimension, the combined presence of a title and a key exhibit higher *information* effect size ($R^2=0.78$, $d=0.73$). In the case of charts with only a key present, the *information* rating is boosted, but shows a smaller effect size than titled charts ($R^2=0.77$, $d=0.69$). These observations align with prior literature that highlights ‘titles’ as high-priority design elements that draw visual attention and increase the perceived quality of visualizations [10, 80].

Annotation (▲) : We observe that visualizations with high levels of text annotation are scored as highly *informative* \in **15** ($R^2=0.92$, $d=0.85$). The presence of message redundancy ($R^2=0.79$, $d=0.86$) and mark labels ($R^2=0.74$, $d=0.83$) also significantly bias *information* internalization. On the other hand, the use of highlighting and overlay of glyphs (like arrows) for annotation show relatively smaller effect sizes ($R^2=0.77$, $d=0.68$ and $R^2=0.72$, $d=0.65$, respectively). We attribute this to an increased preference for charts annotated with more text than sparser counterparts, in line with previous work [72]. Additionally, the text explicitly describes statistical or relational components of a chart (or even summarizes chart takeaways through message redundancy) in comparison to highlights or arrows, which tend to focus on elemental or encoded components. We therefore speculate that participants perceive charts that explicitly state the output of higher-level cognitive tasks such as summarization or decision making as ‘easy to use’ and therefore internalize them as *information*.

Usage of Color and 3D (▼) : More colorful visualizations exhibit *image* internalization: visualizations with seven or more colors display the largest effect size ($R^2=0.73$, $d=0.72$), followed by those with 2–6 colors ($R^2=0.71$, $d=0.70$), and finally visualizations with one

color or black-and-white gradient ($R^2=0.65$, $d=0.68$). We also analyze visualization scores based on a background color; stimuli with white backgrounds are more likely to be rated as *information* ($R^2=0.83$, $d=0.79$), while non-white backgrounds promote *image* ratings. The effect of the background color is more pronounced for visualizations with darker hues ($R^2=0.86$, $d=0.81$). Additionally, the presence of gridlines and usage of 3D glyphs show a relatively low impact on internalization ratings (with $d=0.33$ and $d=0.29$, respectively). Hence, we do not further analyze internalization scores on these bases.

Data Ink Ratio, Visual Density (▲) : Considering all the visualizations together, the respective influences of data-ink ratio and visual density on internalization have moderate effect sizes for the above results ($d \in 0.4–0.65$). Low data-ink ratio and high visual density bias ratings towards **11–13**. Additionally, moderate–high data-ink ratio and low–moderate visual density bias ratings towards **13–15**. The broader range of variation in rating scores indicates that these factors, while significant, are not of high priority during internalization.

Data Source and Domain (▲) : Some visualizations display explicit information about the ‘source’ and ‘domain’ of data via text annotation, typically located bottom-left/right of the stimuli. These elements display negligible effect sizes ($d \leq 0.19$) and therefore do not significantly influence chart internalization.

Data Scale and Dimensionality (▼) : We observe that the scale and dimensionality of data serve as moderators for all the attributes discussed above while displaying negligible effect sizes ($d \leq 0.15$) when considered in isolation. Moderate–high dimensionality and moderate–high data scale lead to the relatively increased internalization of visualizations as *images* overall attributes; lower values of the two do not significantly impact ratings. We can attribute this phenomenon to an increase in the perceived complexity of charts with larger data volumes, which can overwhelm viewers, decreasing their efficiency in distilling relevant information from the chart.

Synthesizing these Results: We note that the presence of elements does not show a clear opposite trend to absence of elements in Fig. 7; we attribute this to these elements being considered in isolation in our analysis. For instance, a chart in **15** might comprise axes, title, key, and message, and these factors might in combination produce an average rating >5 . We accordingly synthesize a rudimentary ranking of the influence various design elements have on the internalization of a visualization: **axes** $>$ **titles/keys** $>$ **message redundancy** **annotation** $>$ **text label annotation** $>$ **pictogram/photo-realistic glyphs** $>$ **more than 7 colors/dark background-color** $>$ **highlight or glyph-based annotation** $>$ **visual density** $>$ **data-ink ratio** $>$ **3D/gridlines** $>$ **data source/domain**. Here, attributes are ordered such that their **presence/high-level** or **absence/low-level** increases the likelihood that the visualization is internalized as an *image*. Data scale and dimensionality levels increase the effect sizes for this ranking. By reversing the **presence/absence** of these design elements in the ranking, we can expect an increased likelihood that the visualization considered will be internalized as *information*.

5.3 “At-a-glance” vs. “Prolonged Exposure” Internalization

We restrict our analysis of data collected during free recall by juxtaposing only against **I1–I5** for rated stimuli (see Fig. 9 for rating patterns and high-level insights). This constraint follows from the non-standardization of dimensionality and data volume across visualizations, which significantly impact the nature of element-wise design choices made during visualization creation.

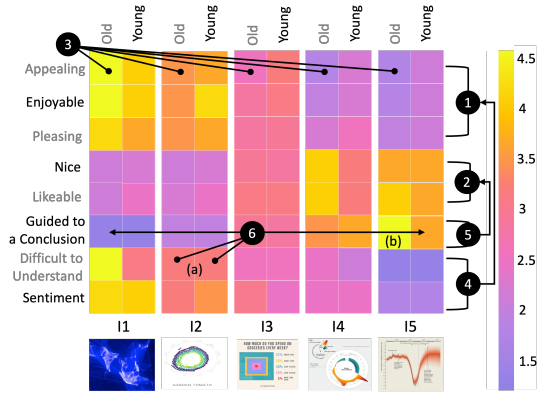


Fig. 9: Internalization score intervals for visualizations juxtaposed against aesthetic pleasure, perceived ease of use, and viewer sentiment across different user groups. We consider sentiment ratings of positive/neutral/negative to correspond to scores of 5/3/1. (1) *Emotion-Aesthetics* scores are higher when stimuli are perceived as more image-like. (2) *Data-Aesthetics* scores are higher when stimuli are perceived as more informational. (3) Older participants show greater variation in *Emotion-Aesthetics* ratings between I1–I5. (4) *Sentiment and Perceived Ease of Use* are proportional to *Emotion-Aesthetics* scores. (5) *Feeling Guided* is proportional to *Data-Aesthetics* scores. *Feeling Guided* ratings: (6a) display the most equivalence between the two participant groups. (6b) show the highest range of variation between I1–I5.

Aesthetic Pleasure: The Beauvis scale [37] assesses a visualization’s aesthetic pleasure (or beauty). The authors use the terms ‘appealing,’ ‘enjoyable,’ ‘likable,’ ‘nice,’ and ‘pleasing’ to target *emotion-oriented* or affective judgment of a visualization; this scale does not target *data-aesthetic* judgment, which hinges on the combination of data and design. We find that *image* visualizations \in **I1–I2** elicit higher ratings for being ‘appealing’ ($R^2=0.82$, $d=0.73$), ‘enjoyable’ ($R^2=0.85$, $d=0.74$), and ‘pleasing’ ($R^2=0.77$, $d=0.73$). However, such visualizations receive lower scores for ‘likable’ ($R^2=0.74$, $d=0.71$) and ‘nice’ ($R^2=0.68$, $d=0.70$). We conjecture that ‘nice’ and ‘likable’ are partially conflated as a data-aesthetic judgment and are not grouped with the other emotion-oriented terms, as observed in Figure 9(2). On the other hand, visualizations rated as *information* \in **I3–I5** elicit higher ratings for being more aesthetically ‘likable’ ($R^2=0.73$, $d=0.65$) and ‘nice’ ($R^2=0.69$, $d=0.67$), and low–moderate ratings for being ‘appealing’ ($R^2=0.76$, $d=0.71$), ‘enjoyable’ ($R^2=0.76$, $d=0.68$) and ‘pleasing’ ($R^2=0.74$, $d=0.69$). This could be attributed to the manner of defining ‘images’ within the scope of our study: we associated images with invoking a sensory reaction which did not necessarily evoke analytical thought. Based on this definition, participants seem to have associated ‘appealing,’ ‘enjoyable,’ and ‘pleasing’ with such a sensory experience. Older participants show clearer patterns of high emotion-oriented aesthetic enjoyment of *image*-rated visualizations (d values are 0.07 ± 0.04 greater on average) compared to younger subjects; in the case of data-aesthetic judgments, the effect size is smaller across both groups of subjects, though it remains significant, as shown in Figure 9(3).

Perceived ease of use and sentiment: Participants feel less ‘guided to a conclusion’ by *image* visualizations \in **I1–I2** ($R^2=0.78$, $d=0.68$) and express higher *positive* sentiment ($R^2=0.69$, $d=0.70$) when viewing these stimuli. However, they simultaneously rate such visualizations

as comparatively ‘difficult to understand’ ($R^2=0.73$, $d=0.71$). Feeling ‘guided’ is strongly associated with data-aesthetics measures (Figure 9(5)); it is also the measure for which both groups of participants have the most equivalent rating (Figure 9(6a)) across I1–I5, which can be potentially attributed to the lower volumes of text present in image-like stimuli. The absence of text places a higher internal cognitive demand on a viewer to process the information shown and form a useful takeaway. The presence of elements like title/annotation/message/etc. in the visualizations causes the cognitive demand on participants to decrease rapidly, in turn reflected in the ratings for ‘feeling guided’ to fluctuate to a greater extent (Figure 9(6b)) compared to other measures.

Information visualizations \in **I4–I5** are considered less complex and provide explicit guidance but are associated with more neutral/negative sentiments. Sentiment hence is associated by participants with emotion-oriented aesthetic judgements (Figure 9(4)), and ease of use follows a similar trend. This confound indicates a dissociation between the *pleasure derived* when viewing a visualization versus the *perceived communicative utility* of the visualization.

Designer Intent: As mentioned in Table 2, visualizations were annotated by experts based on their *sentiment* (positive/negative/neutral emotional response) and *purpose* (expose/involve/analyze). The experts annotated the visualizations by considering them from a designer’s perspective; they estimated how a viewer would respond emotionally to these visuals and how engaging they would find them. We found that experts coded **I4–I5** visualizations as *positive*, compared to other intervals ($R^2=0.82$, $d=0.78$), as shown in Fig. 10. Additionally, these intervals were associated with viewer intent to *analyze* ($R^2=0.85$, $d=0.80$). **I3** visualizations are tagged as *neutral*, and **I1–I2** as *neutral/negative*. Experts, however, annotated very few visualizations from **I1–I3** (<5% with the *expose* level of engagement; it is only used when there is no text present. Instead, *involve* is prevalent for low text volume. Experts, therefore, seem to associate more *positive* sentiment with visualizations seen as more *informative*. This however, contradicts participant sentiment ratings as shown in Fig. 9, wherein more informative stimuli were rated as having neutral/negative sentiment.

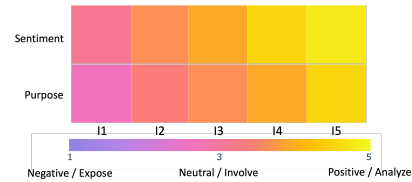


Fig. 10: Internalization score intervals compared to expert annotations for sentiment and purpose, rated as 5/3/1, with 5 corresponding to more positive and higher engagement levels, respectively.

Verbal Recall Patterns: Our analysis of verbal recall patterns can be summarized as follows: *Image*-rated visualizations focus (in priority order) on the recall of (i) individual data points (often those that are either annotated with a label/highlight, or have max-values), (ii) description of visual encoding appearance (typically in order of type of axes, human-recognizable glyphs, color, shape), (iii) positive or neutral opinion on the aesthetic appeal of encoding choices, and (iv) stating the data domain and reporting moderate–high difficulty of understanding data nuances. Such charts, therefore, seek to *expose* and, in some cases, *involve* the user in performing low-level tasks related to data interpretation. *Information*-rated visualizations focus (in priority order) on the recall of (i) the title/message of the visualization, (ii) text-label annotations on marks and axes labels, (iii) reporting high ease of use, observed trend, and re-stating the title/message (paraphrased and in context of an observed trend or annotation) of the visualization, (iv) stating the presence of highlight or overlay annotations like arrows and chart-type (if known). Such charts, therefore, seek to *involve* the user in performing tasks to *analyze* data trends and generate chart takeaways.

Notably, recall patterns align with the expert view of a visualization’s purpose (as discussed above); experts rank more informative stimuli as serving analytical purposes (‘analyze’ level), while image-like stim-

uli are attributed to being designed for creating awareness ('expose' level) on the topic of the visualization to viewers. Accordingly, recall patterns for more informative stimuli (I4–I5) focus on stating a trend or takeaway. Stimuli in I2–I3 tended to have a sufficient amount of annotation for a viewer to be involved in recalling particular data points of interest, in accordance with the intermediate 'involve' level of Mahyar's taxonomy. Image-like stimuli (I1–I2) on the other hand, are primarily described in terms of their appearance, with statements about the domain (or topic) of the visualization taking low priority primarily via referencing.

To summarize, visualizations perceived as 'images' may be less efficacious for communicating trends and messages, although they tend to evoke a more positive emotional response. On the other hand, 'informative' visualizations focus on annotation-based recall and receive more positive design evaluations.

6 CONCLUSIONS AND FUTURE WORK

Our research has shed light on how people internalize and recall visualizations, and provides strong quantitative evidence supporting traditional guidelines for designing and presenting visualizations. We have found that certain design elements can affect how viewers pay attention to and remember visualizations. These elements can serve as effective cues for recalling and guiding viewers through the visualization. Particularly, we find that elements such as axes, title, key, high text volume, etc., minimalist representations, and more familiar chart types (such as bar, pie, line, etc.) are viewed as contributing to stimuli having high perceived informational value. We speculate this result arises from the overall appearance of such stimuli being closer to the abstract canonical impression of particular visualization types. Furthermore, we observe a distinct trend of older participants exhibiting more uniform patterns of rating and recall, with larger effect sizes and lower standard deviation. A potentially stronger dependency (arising from accumulated knowledge and exposure) on 'canonical' representations, combined with potential declines in memory due to aging [58, 73] might partially account for this phenomenon, though it merits deeper inquiry.

We note that we frame our constructed definitions of 'image' vs. 'information' such that study participants consider an image to evoke a primarily superficial sensory reaction that afforded the development of little to no analytical insights, although a viewer might form strong aesthetic-based judgements. We chose a bilinear scale in accordance with this view, as a visualization can have high aesthetic impact, while leading to varying levels and quality of analytical insight [9]. This primarily arose from a lack of vocabulary to define something with purely sensory impact and no informational value in terms of visuals. Additionally, our definitions consider design, purpose, and response, though purpose and design are often studied as a cohesive unit [71]. However, while a visualization's design might be suitable for a viewer to elicit an accurate takeaway, the takeaway generated might be of low value; for instance slow analytics [12] prioritizes retention and understanding as the higher purpose of a visualization, while 'precision' based takeaways are of lower value. As this work is exploratory, future studies can better study the competing and nuanced considerations of image and information.

The following conclusions summarize our high-level insights:

Information elicits positive design judgments and negative emotional judgments. Visualizations with high communicative utility likely require less cognitive load in discerning visualization takeaways. Importantly, when these visualizations are retrieved from memory, many annotation-related details of the visualization are also retrieved. Thus, participant-generated descriptions tend to be of higher quality for these visualizations. However, such visualizations simultaneously evoke a low-level, negative-leaning emotional response, indicating a confound between designing for affective vs. cognitive objectives.

Images elicit positive emotional judgment and are regarded as more visually complex. Visualizations regarded as aesthetically pleasing are perceived to have low analytical utility. This is observed even in the case of pictograms, which have been previously shown to serve as effective visual hooks into memory [9], allowing visualizations to be better recognized and described. The nature of description, however,

supports this view of increased 'complexity,' as descriptions tend to center around the appearance of visual encodings or individual data points. Users simultaneously exhibit low analytical engagement and high emotional engagement with a given visualization.

Redundancy improves quality of both design and emotional response. When a visualization contains redundant information, such as repeating quantitative values (data redundancy), the audience tends to make more emotional/affective judgments. On the other hand, if the redundancy covers the main trends or concepts of the visualization (message redundancy), the audience tends to make more design-related judgments. It is essential to strike a balance between the levels of data and message redundancy to provide clarity and organization to complex information, for instance, by manipulating the relative positions of different text annotations with respect to visual encodings used. This balance can also facilitate comparisons across different parts of the data, resulting in a visually appealing and balanced design.

Embellishing familiar charts can preserve the perceived communicative utility and increase emotional response. A stylistic embellishment of common visualization types, such as bar charts or bubble charts, can partially bridge the gap between emotion and design-based judgments. Participant familiarity with such visualization types may be able to offset the detrimental impact of increased visual density from embellishments on perceived design quality; further examination of the relationship between perceived vs. encoded information is necessary to develop more concrete guidelines on what is "useful" communicative aesthetic. A similar observation holds for cartographic visualizations minimally overlaid with text annotations. Additionally, the dichotomy of our rating task between image and information necessitates that users associate 'images' with evoking a primarily superficial sensory reaction as opposed to cognitive processing. Detailed examination of stimuli in I3 might provide further insights on balancing functional (*information*) and aesthetic design (*image*) considerations.

Simply understanding how the brain processes visualizations is the initial stage towards comprehending how to effectively communicate the crucial and pertinent aspects of the data or trend that the designer intends to convey in a way that ensures the viewer retains them. To gain further insights into visualization internalization, we plan to expand our current database, particularly considering that only a best-fit subset of our current dataset is studied in the context of recall.

Additionally, while participants rate the aesthetics, sentiment, and ease of use (Fig. 10) prior to verbal recall, we did not analyze the recall against each of these individual ratings. We do observe broad general trends between recall and ratings; for instance, for a stimuli in I1, the ratings are high for 'aesthetics' and 'sentiment', low for 'data-aesthetics' and 'ease of use', and verbal recall primarily mentions individual annotated data points or describes visual encoding appearance. It's possible that rating prior to recall helps participants organize their thoughts for recall, enabling uniform qualitative analysis by coding for particular aspects of recall. In the future, we intend to code and analyze recall data as well as use more fine-grained definitions and measures of abstract concepts like visual density and data scale to better understand specific subtleties in various sub-types of visualizations. We also plan to investigate the impact of time on the quality of visualization recall by conducting experiments based on eye-tracking, change-detection, short-term and long-term memorability, and determining which visual elements stick in viewers' minds longer. Furthermore, we aim to directly demonstrate in future work how the nature of internalization significantly affects visualization comprehension and how this may lead to mind-wandering during visualization viewing and recollection. Lastly, we plan to design and evaluate interaction techniques that can help intrinsically resolve user uncertainty when exploring image-like visualizations to boost analytic engagement while maintaining emotional engagement.

By gaining a better understanding of how visualizations are internalized in memory, future studies can ask more high-stakes questions about what makes a visualization effective or engaging. Understanding the impact of low-level visual elements on internalization can help control for them in future experiments, ultimately leading to guidelines to creatively engineer high-impact holistic visualizations.

SUPPLEMENTAL MATERIALS

Refer to the following repository: <https://github.com/aarunku5/Image-or-Information-Vis-2023> for visualization metadata, demographic details of study participants, study analysis and results.

FIGURE CREDITS

A. Fig. 1 image credit: Amanda Shendruk / Quartz, January 2023

B. Fig. 4 image credit: WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation / Progress on Drinking Water and Sanitation 2012-Update

C. Fig. 6 image credits (top-left to bottom-right for each score interval, separated by ;): **I1**: MIT Senseable City Lab/ December 2010 ; Teaching and Learning / Fortune Magazine, April 1961 ; Paul Butler / Facebook, December 2010 ; Sam Shannon / World Visualization Data Prize, Information is Beautiful February 2019 ; Abhishek Tripathi / Mapbox, September 2019 ; Steve Parker / GISP, Visual Wall Maps, January 2023. **I2**: Valentina D’Efilippo / Behance, 2014 ; u/snickerdoodle_codes / Reddit, r/dataisbeautiful, January 2023 ; Audree Lapierre / Behance, December 2012 ; The Regional Planning Commission, County of Los Angeles / 1946; Nathan Yau / Flowing Data, based on New York Times Hurricane Tracker, October 2012 ; Moritz Stefaner / Truth & Beauty in association with Google, Rhythm of Food, November 2016. **I3**: Nathan Yau / Flowing Data, August 2021 ; Wm. F. Sheehan / University of Santa Clara, Ref. Chemistry, Vol. 49, No.3, p 17-18, 1976 ; Kitchn Editors / Kitchn, October 2020 ; Graphic Detail / Economist, October 2011 ; Graphic Detail / Economist, August 2011 ; David McCandless / Information is Beautiful, January 2019. **I4**: Nadieh Bremer / Visual Cinnamon for Scientific American, Graphic Science, July 2017 ; Mike Bostock / Observable, October 2021 ; Nadieh Bremer / Visual Cinnamon for The Guardian, December 2017 ; Beautiful News / Information is Beautiful, September 2020 ; Eduardo Porter and Karl Russel / New York Times, June 2018 ; Emily Badger, Claire Cain Miller, Adam Pearce, and Kevin Quealy / Upshot, New York Times, March 2018. **I5**: Jen Christiansen / Inflation Report, Bank of England, February 2010 ; World Health Organization / Women’s Health Report 2009 ; Wall Street Journal / December 2011 ; Hive Systems / March 2022 ; Information is Beautiful / December 2019 ; World Health Organization / Water and Sanitation, GLAAS Report, 2012.

ACKNOWLEDGMENTS

This research was supported in part by the U.S. National Science Foundation through grants DUE-2216452 and IIS-2238175.

REFERENCES

- [1] E. Adar and E. Lee. Communicative visualizations as a learning problem. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):946–956, 2021. doi: 10.1109/TVCG.2020.3030375 1, 2
- [2] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 111–117. IEEE, 2005. 2
- [3] R. Arnheim. *Art and visual perception: A psychology of the creative eye*. Univ of California Press, 1969. 2
- [4] A. Arunkumar, S. Ginpalli, and C. Bryan. Bayesian modelling of alluvial diagram complexity. In *2021 IEEE Visualization Conference (VIS)*, pp. 51–55. IEEE, 2021. 1
- [5] 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 *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 2573–2582, 2010. 2, 3
- [6] E. Bertini, M. Correll, and S. Franconeri. Why shouldn’t all charts be scatter plots? beyond precision-driven visualizations. In *2020 IEEE Visualization Conference (VIS)*, pp. 206–210. IEEE, 2020. 1
- [7] B. S. Bloom and D. R. Krathwohl. *Taxonomy of educational objectives: The classification of educational goals. Book 1, Cognitive domain*. longman, 2020. 3
- [8] R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Rappa, L. Floridi, and M. Chen. An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2759–2768, 2012. 2
- [9] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. Beyond memorability: Visualization recognition and recall. *IEEE transactions on visualization and computer graphics*, 22(1):519–528, 2015. 1, 2, 9
- [10] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE transactions on visualization and computer graphics*, 19(12):2306–2315, 2013. 1, 2, 3, 4, 7
- [11] J. Boy, F. Detienne, and J.-D. Fekete. Storytelling in information visualizations: Does it engage users to explore data? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1449–1458, 2015. 1, 3
- [12] A. J. Bradley, V. Sawal, and C. Collins. Approaching humanities questions using slow visual search interfaces. 2021. 2, 9
- [13] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics*, 19(12):2376–2385, 2013. 2
- [14] C. Bryan, A. Mishra, H. Shidara, and K.-L. Ma. Analyzing gaze behavior for text-embellished narrative visualizations under different task scenarios. *Visual Informatics*, 4(3):41–50, 2020. 1, 4
- [15] R. Buchanan. Wicked problems in design thinking. *Design issues*, 8(2):5–21, 1992. 2
- [16] R. Burkhard. Learning from architects: the difference between knowledge visualization and information visualization. In *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004.*, pp. 519–524, 2004. doi: 10.1109/IV.2004.1320194 1, 2
- [17] D. Burlinson and D. Szafir. Shape size judgments are influenced by fill and contour closure. *Journal of Vision*, 20(11):1640–1640, 2020. 1
- [18] A. Burns, C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar. Designing with pictographs: Envision topics without sacrificing understanding. *IEEE transactions on visualization and computer graphics*, 28(12):4515–4530, 2021. 2
- [19] A. Cairo. *The truthful art: Data, charts, and maps for communication*. New Riders, 2016. 2
- [20] S. K. Card, J. D. Mackinlay, and B. Shneiderman. Using vision to think. *Readings in information visualization: using vision to think*, 579, 1999. 1
- [21] C.-S. Chan. Mental image and internal representation. *Journal of Architectural and Planning Research*, 14(1):52–77, 1997. 1, 2
- [22] C. Chen. Top 10 unsolved information visualization problems. *IEEE computer graphics and applications*, 25(4):12–16, 2005. 3
- [23] W. S. Cleveland and R. McGill. Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society: Series A (General)*, 150(3):192–210, 1987. 2
- [24] J. Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013. 5
- [25] J. Conklin. A taxonomy for learning, teaching, and assessing: A revision of bloom’s taxonomy of educational objectives complete edition, 2005. 2
- [26] S. Di Bartolomeo, V. Schetinger, J. L. Adams, A. M. McNutt, M. El-Assady, and M. Miller. Doom or deliciousness: Challenges and opportunities for visualization in the age of generative models. 2023. 3
- [27] M. Dick. Interactive infographics and news values. *Digital Journalism*, 2(4):490–506, 2014. 2
- [28] J. Emerson, M. L. Satterthwaite, and A. V. Pandey. The challenging power of data visualization for human rights advocacy. *Practice*, 171:97, 2018. 2
- [29] J.-D. Fekete, J. J. Van Wijk, J. T. Stasko, and C. North. The value of information visualization. *Information Visualization: Human-Centered Issues and Perspectives*, pp. 1–18, 2008. 2
- [30] S. Few. *Now you see it: simple visualization techniques for quantitative analysis*. Analytics Press, 2009. 1, 2
- [31] S. Few. The chartjunk debate—a close examination of recent findings. 2011, 2020. 2
- [32] S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman. The science of visual data communication: What works. *Psychological Science in the public interest*, 22(3):110–161, 2021. 1, 2
- [33] R. Garcia-Retamero and E. T. Cokely. Communicating health risks with visual aids. *Current Directions in Psychological Science*, 22(5):392–399, 2013. 2
- [34] J. F. Hair. *Multivariate data analysis*. 2009. 5
- [35] S. Haroz, R. Kosara, and S. L. Franconeri. Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd annual ACM conference on human factors in computing*

- systems, pp. 1191–1200, 2015. 1, 3
- [36] L. Harrison, K. Reinecke, and R. Chang. Infographic aesthetics: Designing for the first impression. In *Proceedings of the 33rd Annual ACM conference on human factors in computing systems*, pp. 1187–1190, 2015. 2
- [37] T. He, P. Isenberg, R. Dachsel, and T. Isenberg. Beauvis: A validated scale for measuring the aesthetic pleasure of visual representations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):363–373, 2022. 3, 8
- [38] C. Healey and J. Enns. Attention and visual memory in visualization and computer graphics. *IEEE transactions on visualization and computer graphics*, 18(7):1170–1188, 2012. 2, 3
- [39] C. G. Healey, L. Tateosian, J. T. Enns, and M. Remple. Perceptually based brush strokes for nonphotorealistic visualization. *ACM Transactions on Graphics (TOG)*, 23(1):64–96, 2004. 3
- [40] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 1029–1038, 2007. 2
- [41] W. Hogarth. *The analysis of beauty: Written with a view of fixing the fluctuating ideas of taste*. Georg Olms Verlag, 1753. 3
- [42] P. S. Houts, C. C. Doak, L. G. Doak, and M. J. Loscalzo. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient education and counseling*, 61(2):173–190, 2006. 2
- [43] J. Hullman, E. Adar, and P. Shah. Benefitting infovis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2213–2222, 2011. 2
- [44] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2231–2240, 2011. 2
- [45] N. Iliinsky and J. Steele. *Designing data visualizations: Representing informational Relationships*. " O'Reilly Media, Inc.", 2011. 2
- [46] P. Isola, D. Parikh, A. Torralba, and A. Oliva. Understanding the intrinsic memorability of images. *Advances in neural information processing systems*, 24, 2011. 2
- [47] V. John-Steiner. *Notebooks of the mind: Explorations of thinking*. Oxford University Press, 1997. 1, 2
- [48] H. Kefi. Using a systems thinking perspective to construct and apply an evaluation approach of technology-based information systems. In *Emerging Systems Approaches in Information Technologies: Concepts, Theories, and Applications*, pp. 294–309. IGI Global, 2010. 5
- [49] D. Keim, H. Qu, and K.-L. Ma. Big-data visualization. *IEEE Computer Graphics and Applications*, 33(4):20–21, 2013. 2
- [50] H. Kennedy, R. L. Hill, G. Aiello, and W. Allen. The work that visualisation conventions do. *Information, Communication & Society*, 19(6):715–735, 2016. 2
- [51] S. Kerpedjiev, G. Carenini, N. Green, J. Moore, and S. Roth. Saying it in graphics: from intentions to visualizations. In *Proceedings IEEE Symposium on Information Visualization (Cat. No. 98TB100258)*, pp. 97–101. IEEE, 1998. 2
- [52] N. W. Kim, E. Schweickart, Z. Liu, M. Dontcheva, W. Li, J. Popovic, and H. Pfister. Data-driven guides: Supporting expressive design for information graphics. *IEEE transactions on visualization and computer graphics*, 23(1):491–500, 2016. 2
- [53] A. Kirk. *Data Visualization: a successful design process*. Packt publishing LTD, 2012. 1, 2
- [54] S. M. Kosslyn. Can imagery be distinguished from other forms of internal representation? evidence from studies of information retrieval times. *Memory & Cognition*, 4:291–297, 1976. 1, 2
- [55] S. M. Kosslyn and J. R. Pomerantz. Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, 9(1):52–76, 1977. doi: 10.1016/0010-0285(77)90004-4 1, 2
- [56] S. Lambert. And what do i do now? using data visualization for social change. *Center for Artistic Activism*, 2016. 2
- [57] E. Lee-Robbins and E. Adar. Affective learning objectives for communicative visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1–11, 2023. doi: 10.1109/TVCG.2022.3209500 1, 2
- [58] U. Leonards, V. Ibanez, and P. Giannakopoulos. The role of stimulus type in age-related changes of visual working memory. *Experimental Brain Research*, 146(2):172–183, 2002. 9
- [59] G. Lupi. Data humanism: the revolutionary future of data visualization. *Print Magazine*, 30(3), 2017. 2
- [60] J. Mackinlay. Automating the design of graphical presentations of relational information. *Ac Transactions On Graphics (Tog)*, 5(2):110–141, 1986. 2
- [61] N. Mahyar, S.-H. Kim, and B. C. Kwon. Towards a taxonomy for evaluating user engagement in information visualization. 1, 3, 5
- [62] L. Maughan, S. Gutnikov, and R. Stevens. Like more, look more. look more, like more: The evidence from eye-tracking. *Journal of Brand management*, 14(4):335–342, 2007. 2
- [63] W. J. T. Mitchell. What is an image? *New Literary History*, 15(3):503–537, 1984. 1, 2
- [64] A. V. Moore, M. Tomitsch, C. Wimmer, B. Christoph, and T. Grechenig. Evaluating the effect of style in information visualization. *IEEE transactions on visualization and computer graphics*, 18(12):2739–2748, 2012. 1, 3
- [65] T. Munzner. *Visualization analysis and design*. CRC press, 2014. 2
- [66] A. Ovans. What makes the best infographics so convincing. *Harvard Business Review*, 2014. 2
- [67] L. M. Padilla, S. H. Creem-Regehr, and W. Thompson. The powerful influence of marks: Visual and knowledge-driven processing in hurricane track displays. *Journal of experimental psychology: applied*, 26(1):1, 2020. 2
- [68] A. Pandey, A. Srinivasan, and V. Setlur. Medley: Intent-based recommendations to support dashboard composition. *IEEE Transactions on Visualization and Computer Graphics*, 1912. 2
- [69] E. M. Peck, S. E. Ayuso, and O. El-Etr. Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300474 5
- [70] J. C. Roberts, C. Headleand, D. Perkins, and P. D. Ritsos. Personal visualisation for learning. In *Personal Visualization: Exploring Data in Everyday Life Workshop, IEEE Conference on Visualization (VIS), Chicago, IL, USA: Vis4Me Personal15*, 2015. 2
- [71] L. Shen, E. Shen, Z. Tai, Y. Song, and J. Wang. TaskVis: Task-oriented Visualization Recommendation. In M. Agus, C. Garth, and A. Kerren, eds., *EuroVis 2021 - Short Papers*. The Eurographics Association, 2021. doi: 10.2312/evs.20211061 9
- [72] C. Stokes, V. Setlur, B. Cogley, A. Satyanarayan, and M. A. Hearst. Striking a balance: Reader takeaways and preferences when integrating text and charts. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1233–1243, 2023. doi: 10.1109/TVCG.2022.3209383 7
- [73] A. C. Tas, M. C. Costello, and A. T. Buss. Age-related decline in visual working memory: The effect of nontarget objects during a delayed estimation task. *Psychology and aging*, 35(4):565, 2020. 9
- [74] N. Tractinsky, A. S. Katz, and D. Ikar. What is beautiful is usable. *Interacting with computers*, 13(2):127–145, 2000. 2
- [75] E. R. Tufte. The visual display of quantitative information. *The Journal for Healthcare Quality (JHQ)*, 7(3):15, 1985. 2
- [76] E. R. Tufte. Envisioning information. *Optometry and Vision Science*, 68(4):322–324, 1991. 1, 2
- [77] B. Tversky and D. J. Schiano. Perceptual and conceptual factors in distortions in memory for graphs and maps. *Journal of Experimental Psychology: General*, 118(4):387, 1989. 2
- [78] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *IEEE transactions on visualization and computer graphics*, 13(6):1121–1128, 2007. 2
- [79] Z. Wang, S. Wang, M. Farinella, D. Murray-Rust, N. Henry Riche, and B. Bach. Comparing effectiveness and engagement of data comics and infographics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019. 2
- [80] D. L. Wanzer, T. Azzam, N. D. Jones, and D. Skousen. The role of titles in enhancing data visualization. *Evaluation and Program Planning*, 84:101896, 2021. 7
- [81] C. Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2019. 1, 2
- [82] J. Wood, P. Isenberg, T. Isenberg, J. Dykes, N. Boukhelifa, and A. Slingsby. Sketchy rendering for information visualization. *IEEE transactions on visualization and computer graphics*, 18(12):2749–2758, 2012. 3
- [83] C. Xiong, V. Setlur, B. Bach, E. Koh, K. Lin, and S. Franconeri. Visual arrangements of bar charts influence comparisons in viewer takeaways. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):955–965, 2021. 1
- [84] J. S. Yi, Y.-a. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: how do people gain insights using information

visualization? In *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaluation methods for Information Visualization*, pp. 1–6, 2008. 3