DOI: 10.1111/cgf.70099

EUROGRAPHICS 2025 / W. Aigner, N. Andrienko, and B. Wang
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 44 (2025), Number 3

Modeling and Measuring the Chart Communication Recall Process

A. Arunkumar[1], L. Padilla[1], and C. Bryan[2]

[1]Northeastern University, Boston, United States, [2]Arizona State University, Tempe, United States

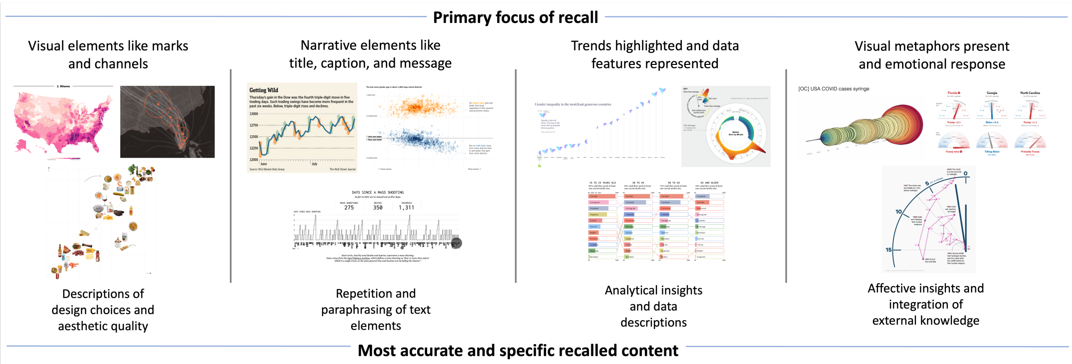

Figure 1: *Examples of visualization stimuli that trigger recall with different primary focii, as well as varying levels of accuracy and specificity.*

Abstract

Understanding memory in the context of data visualizations is paramount for effective design. While immediate clarity in a visualization is crucial, retention of its information determines its long-term impact. While extensive research has underscored the elements enhancing visualization memorability, a limited body of work has delved into modeling the recall process. This study investigates the temporal dynamics of visualization recall, focusing on factors influencing recollection, shifts in recall veracity, and the role of participant demographics. Using data from an empirical study (n = 104), we propose a novel approach combining temporal clustering and handcrafted features to model recall over time. A long short-term memory (LSTM) model with attention mechanisms predicts recall patterns, revealing alignment with informativeness scores and participant characteristics. Our findings show that perceived informativeness dictates recall focus, with more informative visualizations eliciting narrative-driven insights and less informative ones prompting aesthetic-driven responses. Recall accuracy diminishes over time, particularly for unfamiliar visualizations, with age and education significantly shaping recall emphases. These insights advance our understanding of visualization recall, offering practical guidance for designing visualizations that enhance retention and comprehension. All data and materials are available at: `https://osf.io/ghe2j/`.

CCS Concepts

• ***Human-centered computing*** → ***Empirical studies in visualization; Visualization theory, concepts and paradigms; Visualization techniques;***

1. Introduction

Studying memory in the context of data visualizations is crucial because the efficacy of a visualization is determined not just by its immediate clarity but also by how effectively the viewer retains its information [BARM*12, BVB*13]. Memory plays a pivotal role in our ability to make informed decisions, draw conclusions, and take actions based on the data presented. By understanding how different visual representations impact memory retention and recall, designers can create more effective visualizations that present data comprehensibly, ensuring that the information is remembered and can be acted upon in the future.

Prior research on visualization memorability has focused on static recall– evaluating what aspects of a visualization remain memorable after a single timepoint [BARM*12,BVB*13,KRH17]. However, recall is a dynamic process rather than a fixed outcome; it unfolds over time as elements are retained, reinforced, or forgotten. Additionally, individual differences such as age and education play a significant role in recall patterns but remain underexplored in this context [APBB24, PAEE19].

In this work, we address these gaps by modeling recall as a temporal process rather than a static outcome. We conducted an empirical study (n = 104) collecting verbal recall for 100 real-world visualizations. Using structural equation modeling (SEM) with temporally clustered and manually tagged recall data, we analyzed sequential recall patterns. Additionally, we employed a long short-term memory (LSTM) network– a recurrent neural network designed to model sequential data– enhanced with attention mechanisms to focus on important elements within the recall sequence [YYD*16], benchmarking against simpler baselines. Incorporating demographics, perceived informativeness, and visualization metadata into the LSTM provides insights into how individual differences shape recall sequences. This framework combines descriptive analysis with dynamic modeling, offering a comprehensive view of how visualizations are processed and remembered.

Our findings reveal that text elements (titles, captions, annotations) strongly influence recall focus, corroborating prior research on text-visual integration [SSC*22] (see Figure 1). Recall accuracy declines over time across all visualizations, but the rate of decline varies with informativeness and familiarity. Highly informative or familiar visualizations elicit narrative-driven insights, while less informative or unfamiliar ones rely on aesthetic or emotional responses. Demographics like age and education further shape these patterns, affecting the depth, specificity, and accuracy of recall. Our predictive modeling shows that incorporating temporal dynamics and demographics significantly improves recall predictions, offering actionable insights for enhancing visualization design.

We contribute: (i) A coded dataset of verbal recall for 100 diverse, real-world visualizations. (ii) A framework modeling recall as a temporal process, integrating participant demographics, visualization metadata, and perceived informativeness (Sec.3). (iii) Quantitative insights into recall patterns based on visualization elements and participant impressions of familiarity and informativeness (Sec.4). (iv) Design recommendations for optimizing memory retention in visualizations (Sec. 5). To ensure reproducibility, all materials, recall tags, metadata, participant demographics, analysis, and results will be publicly available at: https://osf.io/ghe2j/.

2. Background and Related Work

2.1. Communicative Visualizations and Memory

Communicative visualizations aim to present insights, tell stories, and educate rather than enable exploratory analysis [AL21, CMS99]. They balance cognitive efficiency with emotional engagement to create impactful, data-driven narratives [BSH21]. Prior research has focused on selecting visualization methods aligned with data characteristics [Mac86] and enhancing audience engagement through emotional appeal [LRA22, KRH17]. Memory also plays a crucial role in a visualization's long-term impact, ensuring that insights persist beyond initial exposure [BARM*12, BVB*13]. Most research on visualization memorability assesses recall at a single timepoint [BBK*16, KRH17]. For instance, Borkin et al. [BBK*16] measured recall accuracy in written descriptions, while Kim et al. [KRH17] examined how textual and visual anchors guide attention.

However, these studies treat recall as a static outcome, overlooking its evolution over time; memory retrieval involves recalling, reinforcing, or forgetting elements in sequence. Consequently, our study models recall as a dynamic process, leveraging verbal recall data and temporal modeling to examine how design, narrative structures, and semantic content shape what aspects of a visualization are retained or lost over time.

Semantic Level	Description
L1	Consists of elemental or encoded aspects of the chart, such as the overall topic or a description of the content of an axis.
L2	Consists of statistical or relational components, such as a comparison between two points or identification of extrema.
L3	Describes perceptual or cognitive aspects, such as an overall pattern or changes in trend.
L4	Provides external context to the chart, such as past events which affect the topic depicted.

Table 1: *Conceptual model of semantic content of chart annotations by Lundgard et al. [LS21].*

2.2. Role of Text in Recall

Text plays a crucial role in visualizations by providing titles, annotations, and captions that guide interpretation [KM13]. Designers use text to convey messages, highlight trends, and clarify complexities [KLK18, SH10]. Titles and annotations are often the most memorable elements, anchoring attention and shaping understanding [BBK*16, PAEE19]. Self-explaining visual information further enhances comprehension by fostering accurate internal representations [KRH17].

Semantic content—from low-level descriptions (L1) to high-level insights (L4)—significantly impacts how viewers process and recall visualizations [LS21] (see Table 1). Integrating text and visuals improves comprehension by reinforcing internal representations [KRH17, SSC*22]. However, the role of semantic levels in recall dynamics remains underexplored. We address this by analyzing recall across visualizations with varying annotation levels, scoring recall dimensions to assess how semantic content influences immediate retention. Standardized recall durations (30 seconds) ensure consistency in evaluating memorable aspects.

2.3. Informativeness, Visual Saliency, and Recall Dynamics

Prior research has examined how visualization design impacts memorability and comprehension across chart types [HAS11, Few20, Tuf85, IPTO11]. Studies on memorability typically use written descriptions or recognition tasks to assess recall after a delay [SSC*22, SNL*21, KSL*16, XSB*22, STD19, SSL*23]. While these studies identify memorable visual elements, few explore how informativeness—the effectiveness of a visualization in conveying insights [APBB24]—interacts with temporal recall. Some exceptions address recall accuracy and content quality [BBK*16, APB25]. Additionally, verbal descriptions enhance comprehension by converting internal representations into external formats [KRH17, SAE03].

Our study builds on this to examine how perceived informativeness and visualization familiarity influence recall dynamics. Instead of relying on delayed written descriptions, we analyze immediate verbal recall, capturing first impressions and retention patterns. Using a dataset of diverse communicative visualizations, we tag and cluster recall temporally across dimensions such as aesthetics, narrative, design, and user experience, extending beyond accuracy to uncover deeper recall patterns.

2.4. Demographics and Modeling

Demographic factors, such as age and education, shape how individuals engage with and recall visualizations [PAEE19, APBB24]. Older viewers, with reduced working memory and a preference for structured information, rely on narrative elements as semantic anchors, aiding comprehension [MC05, GNB23, BSH21]. Younger viewers, in contrast, tend to explore advanced encodings and integrate external knowledge for deeper insights [BDF15].

Education further influences graph literacy and familiarity with complex visualizations [KRH17]. Higher education enhances confidence in interpreting abstract or minimally annotated visuals, while those with less formal education benefit from clear textual explanations or human-recognizable objects to reduce cognitive load [PAEE19, APBB24].

Our study integrates these factors into predictive models, analyzing temporal recall dynamics using verbal recall data. By leveraging LSTM models with attention mechanisms, we identify nuanced recall patterns influenced by age and education, contributing to personalized visualization design.

3. Study

3.1. Research Questions

We explore three primary research questions outlined below, focusing on what visualization characteristics determine the ***progression*** (RQ1) and ***veracity*** (RQ2) of recall, as well as the role of ***individual differences*** (RQ3). Our analysis, detailed in Sec. 4, is structured around these questions and their corresponding hypotheses.

RQ1: How does the process of visualization recall unfold, and what factors influence recollection over time? This question investigates the temporal dynamics of recall, focusing on how visualization elements are prioritized and remembered across time steps. Beyond retention, recall involves capturing attention early and guiding viewers to retain critical insights [BBK*16]. Traditional static evaluations overlook these dynamics [APBB24], while a dynamic approach reveals which features effectively capture focus and remain memorable. We hypothesize that highly informative visualizations will prioritize narrative-driven elements, with higher-level semantic content (L3, L4) guiding early comprehension, whereas aesthetic-driven features will dominate recall for less informative visualizations.

RQ2: How does the veracity of recall shift over time? This question examines how accurately and specifically participants recall visualization content over time, as memory accuracy often declines with time [HKF15]. Understanding this decline can reveal which visualization elements are resilient to memory decay and how they influence recall fidelity [QLTL20]. By tagging recall data for accuracy and specificity across time steps, we identify patterns of degradation across visualization types. We hypothesize that the rate and nature of veracity decline are influenced by perceived informativeness and familiarity with the visualization type.

RQ3: How do participant demographics influence the temporal dynamics of recall, and can their individual recall be predicted? This question investigates how demographics, such as age and education, shape the focus and progression of recall. Individual differences influence how participants process and engage with visualizations based on their experiences, cognitive styles, and familiarity with visualization formats [PAEE19, APBB24, GNB23]. We hypothesize that age and education impact preferences for narrative and affective recall. Additionally, we explore whether LSTM models augmented with demographic data can predict individual recall patterns, enabling tailored visualization design.

3.2. Stimuli

Communicative visualizations [CMS99], such as infographics, engage audiences and improve understanding, particularly for individuals with lower graph literacy [BXF*22] (low familiarity scores in our study). This study uses a dataset of 100 static visualizations curated by Arunkumar et al. [APBB24], encompassing diverse annotation styles and design elements. Building on prior large-scale efforts [BVB*13, BBK*16], this dataset systematically categorizes real-world visualizations based on key attributes such as human-recognizable objects, data-ink ratios, and encoding types. Prior research also demonstrated that visualization sources follow distinct design conventions [BVB*13]; for example, government and news sources often use text-heavy bar, pie, and line charts, while scientific and digital media favor richly annotated tree/network charts. Highly pictorial/isotype charts are more common in blogs and infographic platforms. Similarly, the dataset used [APBB24] includes visualizations from governmental, media, infographic, and academic sources, allowing us to examine how annotation and encoding variations influence recall. Further details on dataset composition are provided in the supplementary material.

3.3. Set-Up and Participants

Participants completed the study in two phases (Fig. 2), followed by demographic data collection (age, education, gender). In Phase 1, participants completed 50 trials with randomly selected visualizations, assessing *familiarity* by identifying the base chart type, data types, and creation source (e.g., government, news media) after a 3-second stimulus exposure. In Phase 2, participants viewed the same 50 stimuli in random order, rated their *informativeness* after a 30-second viewing, and then verbally described what they remembered and felt about the stimulus for 30 seconds following a 15-second preparation period. Participants completed three training trials pre-study and two attention checks per phase.

We recruited 104 participants from two groups: 57 college-aged participants (23.6 ± 5.2 years) and 47 older participants (57.4 ± 6.9 years). Among the college-aged group, 31 participants had received a high school diploma and were currently enrolled in their first semester of college, while the remaining 26 participants had at least an undergraduate STEM degree. The older group had 23 participants with only a high school diploma and 24 participants who had completed a college degree (either undergraduate or graduate level in STEM). College-aged participants were recruited from Arizona State University, while older participants were recruited from nearby residential complexes on a volunteer basis. All participants had normal color vision. Data was collected over one month, with

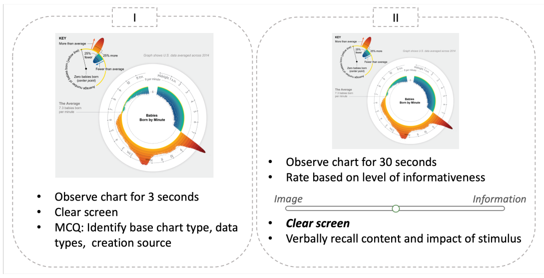

Figure 2: *Study procedure: (I) Phase 1: 50 trials, tests familiarity with stimulus. (II) Phase 2: 50 trials, rate for perceived informativeness and collect verbal recall.*

each participant completing a 90-minute session. Phase 1 lasted an average of 10:09 (±2:18) minutes, and Phase 2 averaged 75:20 (±6:38) minutes, with up to two 10-minute breaks allowed to prevent fatigue. Each visualization was viewed at least 50 times.

3.4. Data Coding

Visualization Features: Taxonomy data from the dataset was mapped to binary (0/1) or ternary (0/1/2) scales.

Familiarity: Familiarity: Participant responses in Phase 1 were scored as correct (1), partially correct (0.5), or incorrect (0) based on visualization metadata (chart type, data types, and source of creation), following Arunkumar et al. [APB25]. Chart type reports were evaluated at a broad category level rather than specific subtypes (e.g., both bar chart and column chart fall under Bar) [BVB*13, APBB24, APB25]. Responses were mapped to the closest taxonomy category to account for synonymous terms. This scoring captured both accuracy and depth of understanding, with aggregated scores forming a composite familiarity metric, where higher scores indicate greater recognition and interpretation ability.

Perceived Informativeness: Following Arunkumar et al. [APBB24, APB25], informativeness was measured on a bilinear scale from -10 (Image) to +10 (Information), assessing how well a visualization conveys structured, interpretable data. Highly informative visualizations featured clear narratives and contextual captions (L3/L4), while less informative ones had sparse annotations or cluttered layouts, often resembling mere 'images' rather than structured data representations. To minimize subjectivity, participants were trained with reference examples spanning the informativeness scale. Study instructions emphasized evaluating informativeness based on data clarity, not topic familiarity. Overall, visualizations were categorized as highly (36%), moderately (27%), or less informative (37%). Ratings were aggregated to identify trends, with a 1.95 standard deviation from prior studies [APBB24], ensuring consistency.

Recall Coding: To analyze recall data, we implemented a multi-step methodology combining unsupervised clustering techniques and handcrafted feature-based tagging to ensure both quantitative rigor and contextual fidelity. Initially, we applied topic modelling as a supervised approach [US23], leveraging pre-defined thematic categories based on the visualization taxonomy [BVB*13, APBB24]. However, this produced poor results, with incoherent clusters that failed to align meaningfully with the semantic structure of the recall data. This issue likely stemmed from the variability in participant responses, which spanned diverse perspectives and lacked the consistency required to fit into rigid, predefined categories. Furthermore, participant descriptions often included overlapping or mixed themes, making it challenging for supervised methods to assign clear-cut labels. This variability underscored the need for more flexible, data-driven approaches.

We therefore adopted an unsupervised clustering approach to allow the data to define its own structure. Token embeddings were extracted using pre-trained language models (BERT [KT19]) to capture semantic relationships in recall data. Initial attempts with k-means and hierarchical agglomerative clustering (HAC) on static data produced unclear, overly generalized clusters with low silhouette scores (<0.5). As silhouette scores quantify how well a data point fits within its assigned cluster relative to its proximity to other clusters [Rou87], the lower scores indicated poor separation between the clusters and highlighted the need for a dynamic approach to account for the temporal and contextual nature of recall.

To address these limitations, we segmented recall data at the sentence level instead of using fixed time intervals. Equal-duration windows (e.g., every 5 seconds) resulted in uneven comparisons, as participants varied in recall speed and detail. Instead, we applied dynamic time warping (DTW) [SC78], a sequence alignment technique that dynamically adjusts segment boundaries to match equivalent recall structures across participants. We define these DTW-aligned segments as **recall timesteps**, representing structurally equivalent recall moments rather than absolute speech timing. This approach preserves the natural recall progression, allowing for a content-driven comparison of recall evolution rather than a rigid time-based division. After DTW segmentation, we applied k-means clustering to sentence-level embeddings extracted using Sentence-BERT [Rei19] to identify recurring thematic structures in recall. This clustering process revealed how recall content evolves across timesteps, capturing shifts in focus over the course of recall.

Despite insights from clustering, some contextual nuances were lost, particularly for mixed or nuanced responses. To address this, we developed a comprehensive set of handcrafted features (Table 2) and manually tagged recall data at the phrase level. Handcrafted features were systematically derived from existing visualization literature on memorability [BBK*16, APBB24, SNL*21] through an iterative process of literature review, author discussion, and pilot annotations. Specifically, we incorporated elements coded in prior works on visualization recall (e.g., [BVB*13, BBK*16, APBB24, APB25]) and extended them to include dimensions unique to verbal recall. Tags encompassed data-driven elements (e.g., topic, chart type, encoding) and user-experience-driven aspects (e.g., opinions on effectiveness). Three annotators, each with expertise in data visualization and/or recall research, independently coded responses. To ensure reliability, they first engaged in discussions with one of the authors to familiarize themselves with recurring themes in a subset of recall responses from prior work [BVB*13, APBB24]. Following this training phase, annotators independently coded the full dataset. Inter-annotator agreement was assessed using Fleiss' κ, which yielded a score of 0.77, indicating substantial agreement [Fle71]. Discrepancies were resolved through discussion among the authors and annotators to refine the

Recall Tag	Constituent Elements	Definition	Examples	References
Data	Data Source, Data Domain, Conjecturing what the Data is if not mentioned	Context for raw data used to populate stimulus	*I see it's WHO [data]*; *medical records from 2014*; *maybe it's something to do with broadband networks*	[BVB*13, PAEE19, TARB23]
Scale	Data Volume, Axes and Axes Labels	Reference to relative magnitude and quantity of stimulus data	*so many points, but the clustering separates them out [a bit]*; *monetary commitments in 2010 by the US vs departments*; *it's a timeline but made radial*	[IS11, Tuf85]
Trends	Identification of data patterns	Analysis of stimulus information that may/may not be explicitly stated in text	*the curve on the right is very bright*; *everything leans irrespective of when it was built*; *lots of tiny flowers in recent years, so could be less death per war but cumulatively more deaths*	[XSB*22, SCA*23]
Metaphor	Skeuomorphism, Photorealistic Elements, Pictorial Units, Human-Recognizable Objects, Human Depiction	Comparison of stimulus to real-world concepts, that may/may not align with the underlying information represented	*looks like the earth at night*; *mountains as triangular bars*; *thought it was like the US flag at first*	[XSG*22, BMG*10]
Author	Title, Message, Caption, Bias, Redundancy	Contribution to narrative in stimulus and explicit reflection of the designer's high-level intent	*most white people stay rich if they are raised rich, but not black people*; *only pro-Trump states are shown*; *it says total internet users*; *size, color, and labels show owned property values*	[SSC*22, PAEE19, KLK18, LRA22]
Appearance	Marks, Channels, Grid Lines, Background Color	Objective description of stimulus appearance	*it's connecting yellow lines on a darker globe*; *curved lines between shapes with more little shapes inside*; *split into two curves made of circles decreasing in size*	[Sza17, SS19]
Schema	Chart Type, Dimensionality, Multiplicity, Ordering	Objective reference to basic structural elements and type of stimulus	*looks like a network*; *it's a [fancy] bar chart essentially*; *lots of features are shown, about ten*	[BVB*13, BBK*16, APBB24]
Implicit Emphasis	Channel-based Highlights, Directionality, Grouping, Gridlines, Background Shading	Elements conveying additional information, context, or insights without directly labeling or annotating	*Rwanda stands out immediately at the top*; *the center band is the most opaque*	[BMSM20, LBLR23]
Explicit Emphasis	Text Labels, Text Volume, Arrows, Reference Lines, Legends	Clearly defined and labeled elements that provide direct information, context, or insights	*World War I and II are pointed out*; *lots of text explaining what's happening and shading the timeline for years of interest*	[BMSM20, SSC*22, SNL*21]
Uncertainty	Conjecture and Confusion Statements	Statement of confusion, frustration, or conjecture about the stimulus information and impact	*not sure what I'm supposed to be seeing*; *why are there no yellow bars for the last few years*; *it's pretty confusing going back and forth with the table*	[BVB*13, BBK*16]
Sentiment	Affective Goals, Emotional Response	Reference to intended emotional impact of stimulus or viewer's affective response to stimulus impact	*it's shocking to see the sheer costs of a literal arm and leg*; *I'm like wow and also meh [you know?]*; *you've certainly turned my worldview on its head [dear], I guess I'll stop checking my grandson's candy*	[LRA22, KRH17]
Opinion	Likes, Dislikes, Suggestions	Qualitative statements about aesthetic appeal and suitability of stimulus	*I love the blue tint, very artsy looking*; *looks like clumsy photoshop though I get the big picture*; *a legend would go a long way to making head or tail of this*	[APBB24, BMG*10, Few20, BBK*16]
Purpose	Cognitive Goals, Stating Intent	Reference to intended informational value of stimulus or viewer's cognitive response to stimulus impact	*it gets the message across but I'd like to drill down into more factors than race*; *definitely seeing more wars if not more death*; *now I know 8AM is a bad time to book appointments with my OBGYN*	[HKF15, BDF15]
Quality	Visual Density, Data Ink Ratio, Complexity	Impact on ease or clarity in reading and understanding stimulus	*it's a very busy graph for so few actual data points*; *I'm more focused on how cool it is that the gridlines are fishnets than on the labels*; *more difficult because of all the petals I keep thinking width means duration instead of following the lines*; *well-designed though I know nothing about stocks*	[HAS11, Few20, Tuf85, IPTO11]
Extrapolation	Integrating external knowledge	Engaging in self-explanation of stimulus	*Covid might have changed things at hospitals*; *typical that the government gets the most funding*; *Nordic countries have amazing social infrastructure so it makes sense*	[KRH17, FPS*21, PAEE19]

Table 2: *Summary of recall tags used for annotation. Whenever one of the constituent elements was referenced, the recall was tagged.*

coding framework and ensure consistency. Each phrase was annotated with feature tags, ranked by relative proportions at different time steps. Annotators also assessed *veracity* by evaluating whether recall accurately reflected stimulus content and referred to specific features. This manual tagging preserved thematic depth and contextual details, mitigating the limitations of unsupervised clustering.

Factor analysis of handcrafted features identified five latent constructs, while unsupervised clustering revealed four thematic clusters in recall data. Although these constructs and clusters overlap conceptually, they capture distinct recall aspects: handcrafted features focus on finer-grained phrase-level descriptions, whereas thematic clusters reflect sentence-level patterns capturing overarching trends and temporal dynamics. We elaborate on this distinction in Section 4. Hence, clustering and handcrafted features serve complementary roles in our analysis, demonstrating the value of combining data-driven and theory-driven approaches.

4. Results & Discussion

In this section, we present results for research questions ***RQ1–RQ3***, based on the hypotheses outlined in Section 3.1. We conducted multivariate SEM (Structural Equation Modeling) [Jör73] using R (4.3.3) with Robust Maximum Likelihood Estimate, maintaining a statistical power of 0.8 [HRS11]. Results are reported for models with the best fit, determined by established fit indices and statistical significance ($p < 0.05$). For further analysis details, please refer to the supplementary material.

Table 3 presents two distinct analytical layers: the top half lists the five latent constructs derived from factor analysis of handcrafted features coded at the phrase level, while the bottom half details the four thematic clusters obtained via unsupervised clustering of sentence-level recall data across recall timesteps (as described in

Latent Variable Construct	Constituent Variables	Description
Data-Centric	Data, Scale, Trends	Focuses on the characteristics, nature, and visual aspects of data representation
Design-Centric	Metaphor, Appearance, Schema	Examines the visual aspects and creative elements of the visualization, along with the overall design structure
Narrative	Author, Implicit Emphasis, Explicit Emphasis	Deals with how the visualization communicates an insight, story, or message
User Experience	Sentiment, Opinion, Purpose	Comprises indicators of how viewers perceive, interpret, and respond to the visualization
Effectiveness	Uncertainty, Quality, Extrapolation	Quantifies clarity, quality, and alignment with intended purpose
Clusters	**Constituent Latent Variables**	**Description**
Factual	Data-Centric, Effectiveness, Narrative	Focuses on repetition and paraphrasing of text content directly present in the visualization
Cognitive	Data-Centric, Narrative, User Experience	Comprises insights about trends that may or may not be explicitly referenced in the stimulus
Affective	User Experience, Effectiveness	Refers to how viewers respond to and perceive the visualization
Aesthetic	Design-Centric, Effectiveness	Deals with descriptions and judgement of the visual features present in the chart

Table 3: *Latent Dependent Variable constructs used in SEM, computed over both the proportions and veracity of clustered and hand coded recall data collected in Phase 2 of the study.*

Section 3.4). To enhance interpretability, we contextualize the thematic clusters using the latent constructs, highlighting their complementary roles in structuring recall. For instance, the cognitive cluster aligned with data-centric features and insights inferred beyond explicit visualization content, while the affective cluster was characterized by user experience and sentiment-related phrases. These names were assigned based on feature co-occurrence and validated through author discussions to ensure alignment with recall patterns. Across recall timesteps, an average of 72% of phrase-level annotations corresponded to the dominant thematic cluster for that segment, indicating substantial overlap in structure.

Figure 3 illustrates the structure of our SEM approaches, which connect all possible combinations of < independent ◯–moderator ◯–dependent □ > paths. Hence, this model allows us to investigate how the proportion and veracity of both the latent handcrafted feature constructs as well as the thematic clusters are influenced by visualization properties, participant characteristics, and recall timesteps. Below, we present the normalized regression results; references to recall timesteps (T1-T5) correspond to DTW-derived recall segments, as defined in Section 3.4.

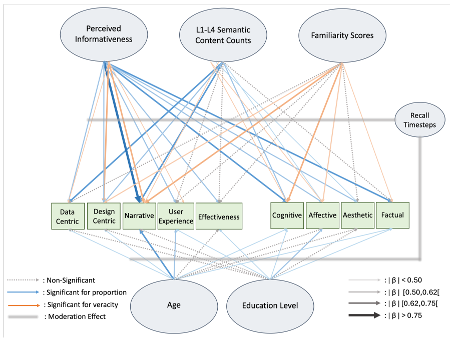

Figure 3: *SEM analysis is conducted separately for each latent variable derived from handcrafted features. Both the proportion (counts) and veracity (accuracy) of these variables are treated as observed measures, with recall timesteps as moderators. Independent variables include informativeness, familiarity, counts of semantic content (L1-L4), and participant demographics. Link color and thickness indicate interaction coefficients (β) for the observed proportion and veracity measures.*

4.1. RQ1: Recall focus and progression

RQ1 asks: how does recall unfold and what factors influence its focus?

Summary of Findings: Both perceived informativeness and the semantic level of text content significantly influence the focus (frequency) and progression (order) of recall. Text volume correlates strongly with perceived informativeness, shaping recall priorities. **Highly informative** visualizations prioritize *narrative* elements, such as titles, messages, and text labels, while **moderately informative** ones also highlight the chart's *purpose* or *quality*. **Less informative** stimuli, often with minimal text, focus on *visual features*, particularly human-recognizable objects, followed by *aesthetic* and *sentiment* statements. Additionally for **text**, L2/L3 semantic content appears early in recall, aiding trend insights, while L1/L4 content surfaces later, often aligning with affective responses. Overall, recall progresses from lower to higher granularity, driven by text volume and the visualization's informativeness.

Details of Analysis: Independent variables include perceived informativeness ratings and L1–L4 semantic content counts, while recall time steps serve as moderators to assess their influence on the dynamic variation in the *proportions* of recall elements.

H1a: ***Overall perceived informativeness dictates the focus and course of recall.*** A strong positive association was found between informativeness and *narrative* features, such as titles, messages, and highlight/arrow annotations ($\beta = 0.82$, $f^2 = 0.51$), as defined in Table 3. Text is often the first element mentioned during recall (83%; see Figure 4 for examples of how recall content and structure differ across informativeness levels), with higher text volume correlating with increased informativeness [APBB24]. Moderate associations were observed for *data-centric* features (e.g., data domain, trends; $\beta = 0.59$, $f^2 = 0.31$) and *user-experience* features (e.g., uncertainty, purpose, goals; $\beta = 0.56$, $f^2 = 0.28$). In contrast, *effectiveness* ($\beta = -0.59$, $f^2 = 0.31$) and *design-centric* features ($\beta = -0.56$, $f^2 = 0.28$), like human-recognizable objects and data-ink-ratio, were negatively associated with informativeness. Design-centric features (human-recognizable objects, data-ink ratio, and visual density) were rarely mentioned in highly informative visualizations (13%) but frequently appeared in less informative ones, often early in recall (74% of cases).

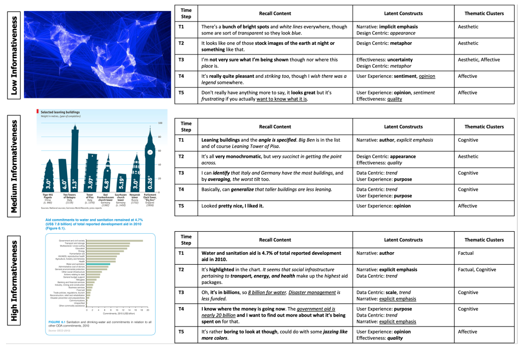

Figure 4: *Coded recall process for stimuli is shown at different levels of informativeness, across hand-crafted features and clusters. T_i represents each recall timestep. The order of appearance and relative proportion in which tags occur are highlighted.*

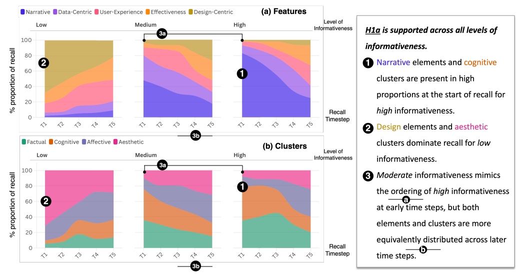

Figure 5: *Small multiple stacked area charts show variation in the relative proportions of different hand crafted features (represented as latent variables from Table 3) and clusters across recall timesteps (x-axis) for different levels of informativeness. Color indicates the feature/cluster examined. Flow size indicates the relative proportions at each time step.*

In general, extreme levels of informativeness (very low or very high) led to the omission or delayed recall of certain features. As shown in Figure 5, for visualizations perceived as moderately informative, the relative proportions of individual feature categories were more balanced, though the progression of recall still closely mirrored that observed for highly informative visualizations. Clustering analysis reinforced these findings: *cognitive* ($\beta = 0.74$, $f^2 =$

0.46) and *factual* ($\beta = 0.63$, $f^2 = 0.38$) clusters, reflecting *narrative* (e.g., title, caption, labels) and *data-centric* (e.g., data domains, axes information) elements, were positively associated with informativeness. Conversely, *aesthetic* ($\beta = -0.43$, $f^2 = 0.25$; e.g., icons, chart type, color, visual density) and *affective* ($\beta = -0.54$, $f^2 = 0.27$; e.g., emotional response, complexity, uncertainty) clusters, tied to *design*, *effectiveness*, and *user experience*, were negatively associated.

While informativeness is a perceptual measure, its strong correlation with structured recall patterns suggests that it captures meaningful differences in how well visualizations support comprehension. Highly informative visualizations often include textual elements, such as annotations and captions, which anchor attention on narrative and cognitive insights, reducing ambiguity and streamlining recall by highlighting key patterns or trends. These features lower cognitive effort and focus viewers on the intended message.

While human-recognizable objects offer some initial engagement, their aesthetic appeal or emotional resonance often outweighs their utility for delivering actionable insights or supporting deeper engagement. Additionally, the strong correlation between text volume and informativeness underscores how narrative-driven designs promote recall sequences that prioritize meaning-making over aesthetic impressions or subjective responses. Specifically, increased text density provides explicit semantic cues, allowing viewers to form connections between different elements of the visualization, leading to more coherent and structured recall. These findings highlight perceived informativeness as a key factor shaping recall progression, particularly when synthesizing complex visual and textual information. **Hence hypothesis H1a is supported.**

H1b: ***L1–L4 semantic content in visualization text primes the recall sequence.*** Different levels of semantic content (L1–L4) act as cognitive anchors in visualizations, shaping focus and understanding [SSC*22, APB25]. By analyzing recall at these levels, we ensure that textual elements are not merely restatements of data but play distinct roles in meaning-making and user engagement (as shown in Figure 4). This effect is particularly strong in visualizations with moderate to high text volumes, where recall is dominated by narrative (Table 3) text elements ($\beta = 0.69$, $f^2 = 0.35$). L2 and L3 content are most frequently mentioned at the first timestep (71% of instances), with L2 often preceding L3 when highlighting maxima, minima, or in the absence of human-recognizable objects. This prioritization of salient data points aligns with the prominence of *cognitive* clusters (64%) and *narrative* elements emphasizing actionable insights. In contrast, L4 content is mentioned less frequently and typically appears during the middle timesteps ($\beta = 0.44$, $f^2 = 0.23$), often in titles, captions, or keys, and is associated with *affective* clusters tied to user experience and emotional responses (77% of mentions). L1 content is most often mentioned last, serving as a peripheral or supplementary anchor in memory.

These patterns indicate that the hierarchical nature of semantic content governs recall progression. L2 and L3 content, offering mid-to-high-level interpretative depth, effectively guide initial comprehension and attention allocation. This aligns with cognitive processing theories [Heg11], which emphasize the salience of intermediate abstraction levels (e.g., trends, patterns, comparative values) for forming mental models. In contrast, L4 content, providing highly contextual or abstract insights, requires more cognitive effort and is therefore potentially typically recalled later during emotional or reflective processing. L1, being basic and descriptive, is relegated to the final recall stages, possibly because it contributes less to overarching insights or emotional framing. **Hence this hypothesis is partially supported.**

4.2. RQ2: Recall veracity

RQ2 asks: how does recall veracity shift over time?

Summary of Findings: Recall veracity decreases over time, with degradation influenced by perceived informativeness and familiarity. For **highly informative** visualizations, cognitive insights and narrative text are recalled with high accuracy early on and are either repeated verbatim or moderately accurate if newly mentioned later. In contrast, design and aesthetic elements are rarely recalled, appearing only in later stages with lower accuracy. **Less informative** visualizations prioritize design, aesthetic, and affective statements early, but these degrade significantly over time, especially visual channels not tied to legends or annotations. Late affective statements also show reduced accuracy. **Moderately informative** visualizations exhibit less overall degradation, with cognitive and data-centric elements maintaining higher veracity across time. Design-centric recall retains accuracy but loses specificity over time. Familiarity with visualization types further reduces degradation at later stages, though its impact is weaker in early recall.

Details of Analysis: We used the same analysis procedure described in Section 4.1, but added on familiarity score as an independent variable, moderated by recall time steps. We also assessed influence on the *accuracy* of recall elements; *specificity* was analyzed qualitatively.

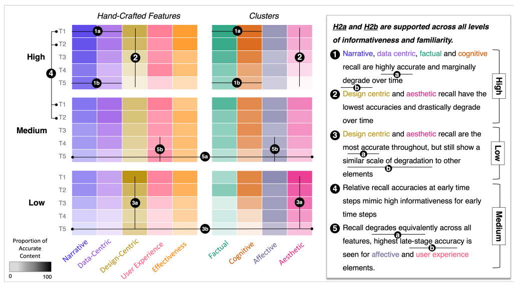

Figure 6: *Heatmaps show the degradation of recall accuracy over time. We averaged over both perceived informativeness and familiarity levels for easy of interpretation as they grouped nicely to high, medium and low levels.*

H2a: ***For less informative charts, the veracity of recall will significantly decrease over the course of recall, compared to stimuli that participants perceive as more informative.*** Recall veracity diminishes over time across all visualizations, consistent with memory decay principles [HNN13]. The rate and nature of this decline, however, are strongly influenced by perceived informativeness. Highly informative visualizations maintain higher recall accuracy for *narrative* (Table 3) features ($\beta = 0.72$, $f^2 = 0.41$), with accuracy dropping only slightly over time ($-\Delta$ 11%) and often being repeated verbatim, preserving specificity. In contrast, visual *design* and *aesthetic* elements degrade more significantly ($-\Delta$ 45%), though human-recognizable objects show slightly better retention ($-\Delta$ 38%). Less informative visualizations prioritize *design-centric* features ($\beta = -0.68$, $f^2 = 0.39$) early in recall, but these degrade sharply over time ($-\Delta$ 28%), especially when visual channels lack legends or annotations ($-\Delta$ 42%). Human-recognizable objects, however, remain more consistent ($-\Delta$ 15%). Moderately informative visualizations display more balanced recall patterns, with *cognitive* and *data-centric* elements retaining both accuracy ($-\Delta$ 14%) and specificity across time. *Design-centric* and *affective* recall remain fairly accurate ($-\Delta$ 19%), though specificity declines in later stages, particularly for features unrelated to human-recognizable objects. Figures 6, 7 illustrate how moderately informative visualizations better mitigate veracity degradation compared to their less or highly informative counterparts.

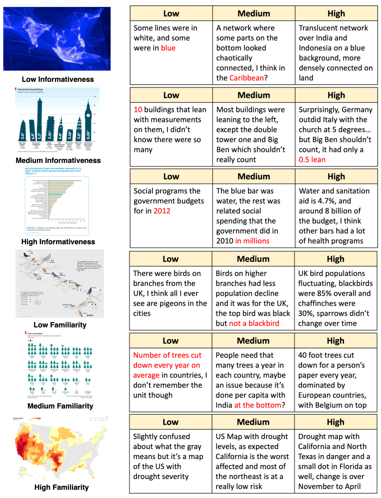

Figure 7: *Variation of the specificity of recall at different levels of perceived informativeness and familiarity. Recall is taken from timesteps: T1/T2– high, T2/T3– moderate, and T4/T5– low specificity. Red text denotes inaccurate aspects of recall.*

These findings build on patterns observed in **H1a**, demonstrating how perceived informativeness shapes cognitive resource allocation and recall priorities over time, reinforcing its role in shaping long-term retention. Highly informative visualizations anchor recall early, maintaining high accuracy even at later stages due to reduced cognitive effort [APB25, APBB24]. Legends and annotations further mitigate recall degradation by providing clear reference points for interpreting visual channels. In contrast, less informative visualizations rely on design-centric and affective elements, like human-recognizable objects, which capture early attention but offer limited cognitive scaffolding, leading to faster accuracy decline. Degradation in specificity for design and affective features at later stages likely results from lack of explicit informational anchors, hindering recall reconstruction. Moderately informative visualizations, which balance narrative and design elements, mitigate these effects by fostering both cognitive engagement and aesthetic appeal, minimizing veracity degradation. These findings highlight how perceived informativeness shapes recall durability and progression, particularly in contexts requiring synthesis of diverse visual and textual elements. **Hence hypothesis H2a is supported.**

H2b: ***For unfamiliar charts, the veracity of recall will significantly decrease over the course of recall compared to chart types that participants are more familiar with.*** Recall veracity is strongly associated with familiarity for widely recognized chart types, such as bar charts or line graphs ($\beta = 0.65$, $f^2 = 0.36$). Familiar visualizations exhibit less degradation over time, particularly for *cognitive* and *narrative* elements (Table 3), which retain high accuracy ($-\Delta$ 7%) and specificity across timesteps. Familiarity allows participants to draw on prior knowledge, reducing cognitive load and enhancing accurate recall of insights and details. In contrast, less familiar chart types, such as chord diagrams and radial plots, experience a steeper decline in veracity ($-\Delta$ 31%). Basic geometric shapes, as opposed to human-recognizable objects, show lower retention ($-\Delta$ 23%). Early recall of these visualizations is dominated by affective and design-centric features ($\beta = 0.54$, $f^2 = 0.29$), with participants often focusing on initial impressions or aesthetic reactions. However, the absence of interpretive schemas results in lower accuracy and specificity at later timesteps, leading to conjectures or omissions. Figures 6, 7 illustrate how familiarity moderates recall trajectories for both familiar and less familiar chart types.

These findings align with schema-driven memory recall theories, which suggest that familiarity with a visualization type provides cognitive scaffolding, enabling participants to encode, organize, and retrieve elements more effectively [HNN13, GNB23]. Familiar visualizations seem to leverage pre-existing mental models to identify patterns and integrate them into coherent understanding, preserving recall veracity over time. Features such as standard axis labels, consistent color schemes, and simple data representations potentially further support accuracy by aligning with viewers' mental frameworks. In contrast, less familiar visualizations lack these interpretive aids, which may prompt reliance on surface-level features like human-recognizable objects, which capture initial attention but fail to sustain accurate recall. Scaffolding elements, such as legends and stepwise annotations, could potentially mitigate these challenges by reducing cognitive load and enhancing recall veracity. These results highlight the importance of aligning visualization design with users' mental models, particularly for complex or unfamiliar chart types. **Hence hypothesis H2b is supported.**

4.3. RQ3: Individual Differences

RQ3 asks: how do participant demographics influence recall, and can this be modeled?

Summary of Findings: Participant demographics, particularly age and education, significantly influence recall focus and progression. Older participants prioritize *narrative* elements **across all levels of perceived informativeness**. In the absence of narrative elements, they often provide statements on communicative effectiveness, avoiding detailed descriptions of visual features except human-recognizable objects. Younger participants, however, tend to recall abstract *visual features* and data-centric elements more frequently at later recall stages. Education has a nuanced impact. While it does not significantly affect the frequency of narrative recall, higher education (college degree or higher) correlates with earlier and more frequent mentions of *cognitive* insights, including those absent from the visualization text. College-educated participants also integrate external knowledge and produce deeper affective statements tied to broader contexts, unlike less educated participants (high school diploma), whose responses often remain surface-level. Notably, age and education do not significantly affect veracity but shape the sequence and emphasis of recall content.

Details of Analysis: We used the same analysis procedure described in Section 4.1, but used age (older vs. younger) and education level (high school vs. college-educated) as independent variables following prior work [APBB24, PAEE19], moderated by recall time steps. We assessed their influence on the dynamic variation in both the *proportion* and *accuracy* of recall elements.

H3a: Participants who are older (or) who have less formal education will mention narrative elements earlier and more often than younger or more educated participants. Older participants exhibited higher proportions of *user-experience* (Table 3) recall tags ($\beta = 0.57$, $f^2 = 0.22$) when recalling visualizations with low to moderate levels of text and annotation. For highly annotated visualizations, recall was dominated by *narrative* tags ($\beta = 0.65$, $f^2 = 0.26$). In contrast, younger participants favored a *design-centric* recall strategy for low-text visualizations ($\beta = 0.48$, $f^2 = 0.21$), emphasizing visual elements over narrative structures (Figure 8). Education level did not significantly influence the narrative recall trajectory ($p = 0.07$). Similarly, age and education had no direct impact on *veracity* of narrative recall, although older participants experienced faster veracity degradation over time ($-\Delta 26\%$). These findings align with prior research indicating that older participants prioritize structured, narrative-driven elements as semantic anchors, simplifying cognitive processing and aiding interpretation of complex or minimally annotated visualizations [GNB23, GC23]. In low-text settings, their focus on user-experience tags and human recognizable objects reflects a compensatory strategy to enhance communicative effectiveness [BSH21]. Conversely, younger participants emphasize more abstract design-centric elements, such as geometric marks, likely due to greater familiarity with modern visualization tools and trends that highlight aesthetic exploration [PAEE19, RM15]. Their inclination toward abstract features over narrative cues may stem from a cognitive preference for engaging with complex encodings [BDF15, APB25]. The lack of a significant education effect on narrative recall likely reflects the intuitive accessibility of elements like titles and captions, which require no specialized skills [BSH21]. Finally, faster degradation of narrative recall veracity among older participants may result from age-related memory decline affecting long-term accuracy [MC05]. **Hence hypothesis H3a is partially supported.**

H3b: Participants who are older (or) who have less formal education will make affective statements earlier and more often than younger or more educated participants. Higher levels of education were positively associated with the frequency and depth of *affective* statements (Table 3) in recall ($\beta = 0.58$, $f^2 = 0.24$). College-educated participants often integrated affective responses with *external knowledge* or *cognitive* insights not explicitly present in the visualization text ($\beta = 0.42$, $f^2 = 0.18$), framing their emotional reactions within societal or domain-specific contexts (Figure 8). In contrast, high-school-educated participants focused on surface-level affective statements, for e.g., the immediate emotional impact of human recognizable objects. Age did not significantly affect the frequency or timing of affective recall ($p = 0.08$), contrary to expectations. For both groups, affective recall was more frequent in stimuli with less narrative content, suggesting a compensatory mechanism where emotional engagement replaces text cues. These findings indicate that affective recall is more strongly influenced by education than age, emphasizing the role of formal education in fostering nuanced emotional engagement. Educated participants are potentially better equipped to contextualize emotional responses within broader interpretative frameworks, aligning with theories of cognitive-affective integration [PAEE19]. In contrast, less-educated participants seem to focus on surface-level affective and aesthetic impressions, consistent with research linking lower graph literacy to reliance on immediate visual or emotional cues [BXF*22]. The absence of a significant age effect challenges the "positivity effect" hypothesis [MC05], suggesting that emotional engagement in recall is more driven by visualization content/design than age-related affective goals. **Hence hypothesis H3b is partially supported.**

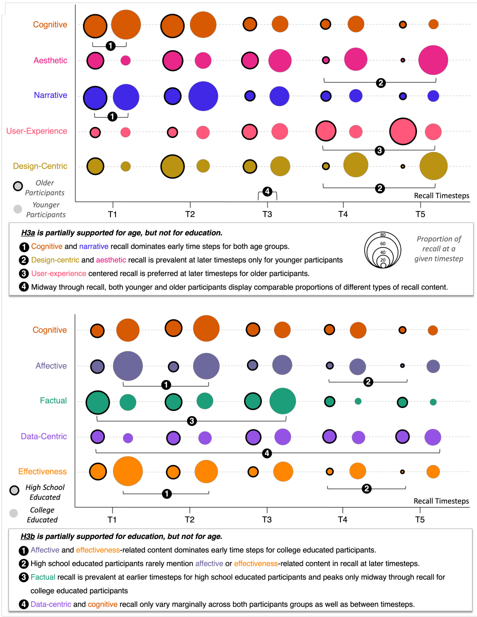

Figure 8: *Bubble charts based on recall timesteps (x-axis) across demographic groups (top: age, bottom: education) depicting variation in the proportions (bubble-size) of recall elements. Recall content categories are depicted on the y-axis and are color coded. The border thickness of bubbles differentiates demographic groups.*

Model	Overall Performance		Low Informativeness		Moderate Informativeness		High Informativeness	
Accuracy	Acc_P	Acc_V	Acc_P	Acc_V	Acc_P	Acc_V	Acc_P	Acc_V
LSTM With Demographics	89.4	87.3	86.5	81.9	89.3	88.3	92.5	91.7
LSTM Base Model	82.6	78.4	76.2	69.7	84.7	80.4	89.3	86.1
Logistic Regression	65.4	61.8	52.5	43.7	60.1	63.6	65.4	69.5
Feedforward Neural Network	70.2	67.4	64.3	55.7	70.8	67.9	71.2	75.1

Table 4: *Model performance results (%) for LSTM models, as well as for baseline models applied to each timestep. Both LSTM models outperform the baselines models across all stimuli in predicting both the veracity and proportion of recall content across the course of recall. Demographics help marginally improve performance in comparison the the initial LSTM Base Model.*

Modeling Temporal Dynamics. To analyze the influence of visualization impressions (informativeness and familiarity) and participant demographics on recall progression, we trained an LSTM model with attention mechanisms [YYD*16] on recall data from 79 visualizations, reserving 21 (n = 7 per informativeness level) for validation and testing in a 5-fold cross-validation setup. Recall sequences were segmented into five sentence-based time steps, with DTW aligning participant responses, resulting in a total of 395 recall data points per participant and 39,500 data points per visualization for the training process. Input features included handcrafted features and clustering proportions and accuracies, perceived informativeness, and visualization metadata. The model consists of multidimensional input feature vectors, followed by two LSTM layers with 128 hidden units, a dropout layer (rate 0.2) to reduce overfitting, and an attention mechanism to prioritize important features at each timestep. The output layer uses softmax activation to predict categorical recall features, optimized with categorical cross-entropy loss and the Adam optimizer [KB14]. Training was performed for 100 epochs (batch size = 32). Performance was evaluated using two metrics: Acc_P– proportion of time steps correctly predicting dominant clusters or handcrafted features, and Acc_V– proportion accurately predicting veracity. Results, stratified by informativeness levels (Table 4), were compared to baseline models (logistic regression and feed-forward neural networks trained separately for each recall time step, using the same data as LSTM), emphasizing the advantages of temporal modeling.

The LSTM model outperformed baselines in predicting recall dynamics, achieving 82.6% overall accuracy for Acc_P and 78.4% for Acc_V. Stratified performance was highest for highly informative visualizations (Acc_P: 89.3%, Acc_V: 86.1%), moderate for moderately informative visualizations (84.7%, 80.4%), and lowest for less informative ones (76.2%, 69.7%). Structured narrative elements improved predictability, while less informative visualizations introduced variability. Baseline models (logistic regression, FFN) lagged behind, with accuracies of 65.4% and 70.2% for Acc_P, and 61.8% and 67.4% for Acc_V, highlighting the LSTM's ability to capture temporal dependencies. Attention analysis showed that narrative features were the strongest predictors, while design-centric and aesthetic features influenced early recall but had limited long-term impact. Adding participant demographics (age, education) improved Acc_P and Acc_V by 6.8% and 8.9%, with the largest gains for less informative visualizations (10.3%, 12.2%), emphasizing the importance of individual differences in recall modeling. The LSTM model offers a dynamic understanding of recall, enabling prediction of the effectiveness of specific elements, particularly narrative and cognitive features, to optimize visualizations for better long-term recall. Additionally, the model's ability to predict recall across time steps opens the door for interactive visualizations that adapt based on user recall patterns, leading to more personalized and effective user experiences.

5. Conclusions and Future Work

Our study examines how perceived informativeness, familiarity, and participant demographics shape recall dynamics in visualizations. Through temporal modeling, we highlight the role of text, icons, and annotations in recall, likely due to their visual saliency. We validated predictive models using visualization metadata and participant differences, demonstrating their potential to forecast recall scores and explain demographic variations. By focusing on immediate recall, we assess information retained in short-term memory [Cow08], revealing how attentional filtering [Bro82] and memory decay [ZL09] impact recall. Our findings provide a holistic measure of visualization effectiveness, integrating cognitive, emotional, and design-based factors. Below, we summarize key insights and their practical implications:

(D1) Enhancing recall through narrative elements. Highly informative visualizations prioritize narrative elements, such as titles and text labels, which reduce cognitive load and direct attention toward actionable insights. *Design Guideline:* Use structured narrative components, including clear titles and concise messages, to enhance clarity and recall, particularly for complex data.

(D2) Mitigating recall degradation for less informative and unfamiliar visualization types. In less informative or unfamiliar visualizations, human-recognizable objects act as intuitive anchors, engaging users and supporting recall accuracy. However, they are less effective than narrative elements for deeper cognitive engagement. *Design Guideline:* Use human-recognizable objects as entry points in less informative or unfamiliar visualizations, paired with intuitive legends or labels to bridge understanding gaps.

(D3) Guiding recall with semantic text content. Mid-level semantic content (L2/L3) drives early recall by offering interpretative depth, while abstract content (L4) supports later reflective processing. *Design Guideline:* Highlight mid-level semantic content prominently to aid initial comprehension, with abstract insights positioned for deeper exploration.

(D4) Tailoring visualizations to audience demographics. Older participants favor narrative elements, while younger or more educated audiences engage with abstract features and design-centric elements. *Design Guideline:* For older audiences, prioritize clear titles, descriptive annotations, and straightforward layouts. For younger or more educated viewers, emphasize intricate, layered visual encodings and less textual guidance to boost engagement.

While our findings provide actionable design insights, several limitations warrant consideration. First, our focus on static visualizations excludes interactive contexts common in real-world use. Future work should examine interaction metrics' (e.g., clicks, zooms) influence on recall. Second, reliance on self-reported recall and manual tagging introduces potential biases; automating these processes with NLP and multimodal data (e.g., gaze tracking) could improve reproducibility. Third, in our participant pool, although younger and older groups included both high school and college-educated participants, all younger high-school-educated participants were in their first semester of college, limiting the isolation of high school education effects. Since demographic variables often exhibit partial dependence [TF19], our models treated age and education separately. Future work could recruit a more stratified sample and broaden cultural and linguistic representation for generalizability. Lastly, the LSTM model's reliance on handcrafted features and clustering may oversimplify recall behavior. Incorporating richer embeddings from advanced models and multimodal data could enhance interpretability and predictive accuracy.

Overall, this study underscores the importance of understanding recall dynamics in visualization design. By linking design elements to cognitive processes and audience needs, we contribute to creating visualizations that are not only memorable but also meaningful across diverse contexts. These insights pave the way for future exploration of innovative design strategies that enhance clarity, engagement, and inclusivity.

References

[AL21] ADAR E., LEE E.: Communicative visualizations as a learning problem. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2021), 946–956. doi:10.1109/TVCG.2020.3030375. 2

[APB25] ARUNKUMAR A., PADILLA L., BRYAN C.: Mind drifts, data shifts: Utilizing mind wandering to track the evolution of user experience with data visualizations. *IEEE Transactions on Visualization and Computer Graphics 31*, 1 (2025), 1169–1179. doi:10.1109/TVCG.2024.3456344. 2, 4, 7, 8, 9

[APBB24] ARUNKUMAR A., PADILLA L., BAE G.-Y., BRYAN C.: Image or information? examining the nature and impact of visualization perceptual classification. *IEEE Transactions on Visualization and Computer Graphics 30*, 1 (2024), 1030–1040. doi:10.1109/TVCG.2023.3326919. 2, 3, 4, 5, 6, 8, 9

[BARM*12] BORGO R., ABDUL-RAHMAN A., MOHAMED F., GRANT P. W., REPPA I., FLORIDI L., CHEN M.: An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2759–2768. doi:10.1109/TVCG.2012.197. 1, 2

[BBK*16] BORKIN M. A., BYLINSKII Z., KIM N. W., BAINBRIDGE C. M., YEH C. S., BORKIN D., PFISTER H., OLIVA A.: Beyond memorability: Visualization recognition and recall. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 519–528. doi:10.1109/TVCG.2015.2467732. 2, 3, 4, 5

[BDF15] BOY J., DETIENNE F., FEKETE J.-D.: 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* (New York, NY, USA, 2015), CHI '15, Association for Computing Machinery, p. 1449–1458. doi:10.1145/2702123.2702452. 3, 5, 9

[BMG*10] BATEMAN S., MANDRYK R. L., GUTWIN C., GENEST A., MCDINE D., BROOKS C.: 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* (New York, NY, USA, 2010), CHI '10, Association for Computing Machinery, p. 2573–2582. doi:10.1145/1753326.1753716. 5

[BMSM20] BRYAN C., MISHRA A., SHIDARA H., MA K.-L.: Analyzing gaze behavior for text-embellished narrative visualizations under different task scenarios. *Visual Informatics 4*, 3 (2020), 41–50. doi:10.1016/j.visinf.2020.08.001. 5

[Bro82] BROADBENT D. E.: Task combination and selective intake of information. *Acta Psychologica 50*, 3 (1982), 253–290. doi:10.1016/0001-6918(82)90043-9. 10

[BSH21] BULLOCK O. M., SHULMAN H. C., HUSKEY R.: Narratives are persuasive because they are easier to understand: Examining processing fluency as a mechanism of narrative persuasion. *Frontiers in Communication 6* (2021). doi:10.3389/fcomm.2021.719615. 2, 3, 9

[BVB*13] BORKIN M. A., VO A. A., BYLINSKII Z., ISOLA P., SUNKAVALLI S., OLIVA A., PFISTER H.: What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2306–2315. doi:10.1109/TVCG.2013.234. 1, 2, 3, 4, 5

[BXF*22] BURNS A., XIONG C., FRANCONERI S., CAIRO A., MAHYAR N.: Designing with pictographs: Envision topics without sacrificing understanding. *IEEE Transactions on Visualization and Computer Graphics 28*, 12 (2022), 4515–4530. doi:10.1109/TVCG.2021.3092680. 3, 9

[CMS99] CARD S. K., MACKINLAY J. D., SHNEIDERMAN B.: Using vision to think. *Readings in information visualization: using vision to think 579* (1999). 2, 3

[Cow08] COWAN N.: Chapter 20 what are the differences between long-term, short-term, and working memory? In *Essence of Memory*, Sossin W. S., Lacaille J.-C., Castellucci V. F., Belleville S., (Eds.), vol. 169 of *Progress in Brain Research*. Elsevier, 2008, pp. 323–338. doi:10.1016/S0079-6123(07)00020-9. 10

[Few20] FEW S.: The chartjunk debate–a close examination of recent findings. 2011, 2020. 2, 5

[Fle71] FLEISS J. L.: Measuring nominal scale agreement among many raters. *Psychological bulletin 76*, 5 (1971), 378. doi:10.1037/h0031619. 4

[FPS*21] FRANCONERI S. L., PADILLA L. M., SHAH P., ZACKS J. M., HULLMAN J.: The science of visual data communication: What works. *Psychological Science in the public interest 22*, 3 (2021), 110–161. doi:10.1177/15291006211051956. 5

[GC23] GUSKJOLEN A., CEMBROWSKI M. S.: Engram neurons: Encoding, consolidation, retrieval, and forgetting of memory. *Molecular psychiatry 28*, 8 (2023), 3207–3219. doi:10.1038/s41380-023-02137-5. 9

[GNB23] GREENE N. R., NAVEH-BENJAMIN M.: Adult age-related changes in the specificity of episodic memory representations: A review and theoretical framework. *Psychology and Aging 38*, 2 (2023), 67. doi:10.1037/pag0000724. 3, 8, 9

[HAS11] HULLMAN J., ADAR E., SHAH P.: Benefitting infovis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2213–2222. doi:10.1109/TVCG.2011.175. 2, 5

[Heg11] HEGARTY M.: The cognitive science of visual-spatial displays: Implications for design. *Topics in cognitive science 3*, 3 (2011), 446–474. doi:10.1111/j.1756-8765.2011.01150.x. 7

[HKF15] HAROZ S., KOSARA R., FRANCONERI S. L.: Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (2015), pp. 1191–1200. doi:10.1145/2702123.2702275. 3, 5

[HNN13] HARDT O., NADER K., NADEL L.: Decay happens: the role of active forgetting in memory. *Trends in cognitive sciences 17*, 3 (2013), 111–120. doi:doi:10.1016/j.tics.2013.01.001. 7, 8

[HRS11] HAIR J. F., RINGLE C. M., SARSTEDT M.: Pls-sem: Indeed a silver bullet. *Journal of Marketing theory and Practice 19*, 2 (2011), 139–152. doi:10.2753/MTP1069-6679190202. 5

[IPTO11] ISOLA P., PARIKH D., TORRALBA A., OLIVA A.: Understanding the intrinsic memorability of images. *Advances in neural information processing systems 24* (2011). URL: https://proceedings.neurips.cc/paper_files/paper/2011/file/286674e3082feb7e5afb92777e48821f-Paper.pdf. 2, 5

[IS11] ILIINSKY N., STEELE J.: *Designing data visualizations: Representing informational Relationships.* " O'Reilly Media, Inc.", 2011. 5

[Jör73] JÖRESKOG K. G.: Analysis of covariance structures. In *Multivariate analysis–III*. Elsevier, 1973, pp. 263–285. doi:10.1016/B978-0-12-426653-7.50024-7. 5

[KB14] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014). doi:10.48550/arXiv.1412.6980. 10

[KLK18] KONG H.-K., LIU Z., KARAHALIOS K.: Frames and slants in titles of visualizations on controversial topics. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (2018), pp. 1–12. doi:10.1145/3173574.3174012. 2, 5

[KM13] KOSARA R., MACKINLAY J.: Storytelling: The next step for visualization. *Computer 46*, 5 (2013), 44–50. doi:10.1109/MC.2013.36. 2

[KRH17] KIM Y.-S., REINECKE K., HULLMAN J.: Explaining the gap: Visualizing one's predictions improves recall and comprehension of data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), pp. 1375–1386. doi:10.1145/3025453.3025592. 1, 2, 3, 5

[KSL*16] Kim N. W., Schweickart E., Liu Z., Dontcheva M., Li W., Popovic J., Pfister H.: Data-driven guides: Supporting expressive design for information graphics. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2016), 491–500. doi:10.1109/TVCG.2016.2598620. 2

[KT19] Kenton J. D. M.-W. C., Toutanova L. K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT* (2019), vol. 1, Minneapolis, Minnesota, p. 2. doi:10.18653/v1/N19-1423. 4

[LBLR23] Luisa Barrera-Leon F. C., Russis L. D.: How the preattentive process is exploited in practical information visualization design: A review. *International Journal of Human–Computer Interaction 39*, 4 (2023), 707–720. doi:10.1080/10447318.2022.2049137. 5

[LRA22] Lee-Robbins E., Adar E.: Affective learning objectives for communicative visualizations. *IEEE Transactions on Visualization and Computer Graphics 29*, 1 (2022), 1–11. doi:10.1109/TVCG.2022.3209500. 2, 5

[LS21] Lundgard A., Satyanarayan A.: Accessible visualization via natural language descriptions: A four-level model of semantic content. *IEEE Transactions on Visualization and Computer Graphics 28*, 1 (2021), 1073–1083. doi:10.1109/TVCG.2021.3114770. 2

[Mac86] Mackinlay J.: Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog) 5*, 2 (1986), 110–141. doi:10.1145/22949.22950. 2

[MC05] Mather M., Carstensen L. L.: Aging and motivated cognition: The positivity effect in attention and memory. *Trends in cognitive sciences 9*, 10 (2005), 496–502. doi:10.1016/j.tics.2005.08.005. 3, 9

[PAEE19] Peck E. M., Ayuso S. E., El-Etr O.: 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* (New York, NY, USA, 2019), CHI '19, Association for Computing Machinery, p. 1–12. doi:10.1145/3290605.3300474. 2, 3, 5, 9

[QLTL20] Qin X., Luo Y., Tang N., Li G.: Making data visualization more efficient and effective: a survey. *The VLDB Journal 29*, 1 (2020), 93–117. doi:10.1007/s00778-019-00588-3. 3

[Rei19] Reimers N.: Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019). doi:10.48550/arXiv.1908.10084. 4

[RM15] Ruchikachorn P., Mueller K.: Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE Transactions on Visualization and Computer Graphics 21*, 9 (2015), 1028–1044. doi:10.1109/TVCG.2015.2413786. 9

[Rou87] Rousseeuw P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics 20* (1987), 53–65. doi:10.1016/0377-0427(87)90125-7. 4

[SAE03] Stern E., Aprea C., Ebner H. G.: Improving cross-content transfer in text processing by means of active graphical representation. *Learning and Instruction 13*, 2 (2003), 191–203. doi:10.1016/S0959-4752(02)00020-8. 2

[SC78] Sakoe H., Chiba S.: Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing 26*, 1 (1978), 43–49. doi:10.1109/TASSP.1978.1163055. 4

[SCA*23] Shrestha H., Cachel K., Alkhathlan M., Rundensteiner E., Harrison L.: Help or hinder? evaluating the impact of fairness metrics and algorithms in visualizations for consensus ranking. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (New York, NY, USA, 2023), FAccT '23, Association for Computing Machinery, p. 1685–1698. doi:10.1145/3593013.3594108. 5

[SH10] Segel E., Heer J.: Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics 16*, 6 (2010), 1139–1148. doi:10.1109/TVCG.2010.179. 2

[SNL*21] Srinivasan A., Nyapathy N., Lee B., Drucker S. M., Stasko J.: Collecting and characterizing natural language utterances for specifying data visualizations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2021), CHI '21, Association for Computing Machinery. doi:10.1145/3411764.3445400. 2, 4, 5

[SS19] Smart S., Szafir D. A.: Measuring the separability of shape, size, and color in scatterplots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–14. doi:10.1145/3290605.3300899. 5

[SSC*22] Stokes C., Setlur V., Cogley B., Satyanarayan A., Hearst M. A.: Striking a balance: reader takeaways and preferences when integrating text and charts. *IEEE Transactions on Visualization and Computer Graphics 29*, 1 (2022), 1233–1243. doi:10.1109/TVCG.2022.3209383. 2, 5, 7

[SSL*23] Shen L., Shen E., Luo Y., Yang X., Hu X., Zhang X., Tai Z., Wang J.: Towards natural language interfaces for data visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics 29*, 6 (2023), 3121–3144. doi:10.1109/TVCG.2022.3148007. 2

[STD19] Setlur V., Tory M., Djalali A.: Inferencing underspecified natural language utterances in visual analysis. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (New York, NY, USA, 2019), IUI '19, Association for Computing Machinery, p. 40–51. doi:10.1145/3301275.3302270. 2

[Sza17] Szafir D. A.: Modeling color difference for visualization design. *IEEE transactions on visualization and computer graphics 24*, 1 (2017), 392–401. doi:10.1109/TVCG.2017.2744359. 5

[TARB23] Tandon S., Abdul-Rahman A., Borgo R.: Measuring effects of spatial visualization and domain on visualization task performance: A comparative study. *IEEE Transactions on Visualization and Computer Graphics 29*, 1 (2023), 668–678. doi:10.1109/TVCG.2022.3209491. 5

[TF19] Tabachnick B. G., Fidell L. S.: *Using Multivariate Statistics*, 7th ed. Pearson, Boston, MA, 2019. 10

[Tuf85] Tufte E. R.: The visual display of quantitative information. *The Journal for Healthcare Quality (JHQ) 7*, 3 (1985), 15. 2, 5

[US23] Uthirapathy S. E., Sandanam D.: Topic modelling and opinion analysis on climate change twitter data using lda and bert model. *Procedia Computer Science 218* (2023), 908–917. doi:10.1016/j.procs.2023.01.071. 4

[XSB*22] Xiong C., Setlur V., Bach B., Koh E., Lin K., Franconeri S.: Visual arrangements of bar charts influence comparisons in viewer takeaways. *IEEE Transactions on Visualization and Computer Graphics 28*, 1 (2022), 955–965. doi:10.1109/TVCG.2021.3114823. 2, 5

[XSG*22] Xiong C., Sarvghad A., Goldstein D. G., Hofman J. M., Demiralp Ç.: Investigating perceptual biases in icon arrays. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022), pp. 1–12. doi:10.1145/3491102.3501874. 5

[YYD*16] Yang Z., Yang D., Dyer C., He X., Smola A., Hovy E.: Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (San Diego, California, June 2016), Knight K., Nenkova A., Rambow O., (Eds.), Association for Computational Linguistics, pp. 1480–1489. doi:10.18653/v1/N16-1174. 2, 10

[ZL09] Zhang W., Luck S. J.: Sudden death and gradual decay in visual working memory. *Psychological science 20*, 4 (2009), 423–428. doi:10.1111/j.1467-9280.2009.02322.x. 10