Similarity Framework for Visualization Retrieval

Huyen N. Nguyen[*]
Harvard Medical School

Nils Gehlenborg[†]
Harvard Medical School

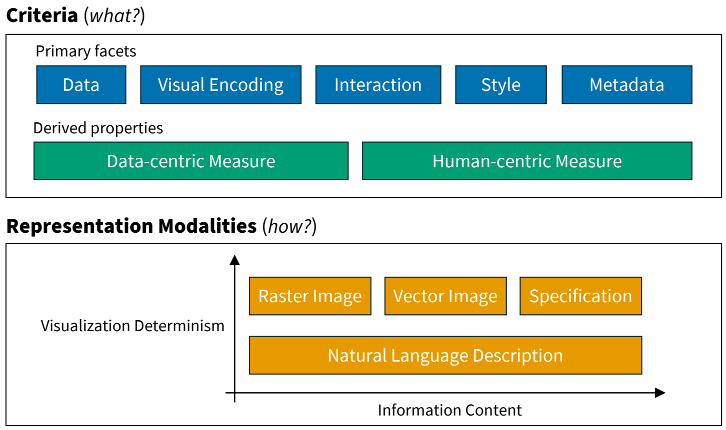

Figure 1: Our proposed similarity framework for visualization retrieval establishes clear criteria and representation modalities. The framework characterizes comparison criteria determining *what* aspects of visualizations should be compared, while representation modalities define *how* these visualizations are represented during comparison process, with regard to information content and visualization determinism–the degree to which a representation format guarantees a single, consistent visual rendering.

Abstract

Effective visualization retrieval necessitates a clear definition of similarity. Despite the increasing body of work in specialized visualization retrieval systems, a systematic approach to understanding visualization similarity remains absent. We introduce the Similarity Framework for Visualization Retrieval (Safire), a conceptual model that frames visualization similarity along two dimensions: comparison criteria and representation modalities. Comparison criteria identify what aspects make visualizations similar: data, visual encoding, interaction, style, and metadata, while considering derived properties such as data-centric and human-centric measures. Safire connects what to compare with how comparisons are executed through representation modalities. We categorize existing representation approaches into four groups based on abstraction level: raster image, vector image, specification, and natural language description, guiding what is computable and comparable. We analyze several visualization retrieval approaches with Safire to demonstrate its practical value in clarifying similarity considerations. The findings reveal how specific similarity and representation aspects align in different use cases. One significant insight is that the choice of representation modality is not only specific to implementation but an important decision that shapes retrieval capabilities and constraints. Based on our analysis, we provide recommendations and discuss broader implications for multimodal learning, AI applications, and visualization reproducibility.

Index Terms: Visualization retrieval, visualization similarity, comparison, representation modality.

[*]e-mail: huyen_nguyen@hms.harvard.edu
[†]e-mail: nils@hms.harvard.edu

1 Introduction

Designing effective visualization retrieval systems involves unique challenges due to the complex nature of visualizations. While the overarching goal aligns with general information retrieval: finding relevant examples for a query, designers must first address what relevance means for visualizations. A fundamental question arises: *What constitutes similarity between two visualizations?* This leads to a series of exploratory considerations: What criteria should guide comparison? Should we compare underlying data, visual encoding choices, interactive features, or aesthetic styles? Additionally, which representation format best captures a visualization's essence? These similarity modeling questions are critical in specialized visualization retrieval systems [8, 19, 21] and broader search platforms [3, 31, 32]. Despite their recurrence across various scenarios, a systematic approach to clarifying essential dimensions of visualization similarity is currently lacking.

To address this gap, we propose a Similarity Framework for Visualization Retrieval (Safire). Safire (pronounced similarly to sapphire) provides a structured framework for understanding visualization similarity along two key dimensions, as shown in Figure 1. The comparison **criteria** determine *what* aspects of visualizations should be compared, while **representation modalities** define *how* visualizations are represented for comparison. We ground Safire in visualization theory and contextualize it with practical applications to ensure its applicability in real-world systems.

We develop criteria for what makes visualizations similar, distinguishing between primary facets used in visualization construction and derived properties observed afterward. The framework identifies five key primary facets: data, visual encoding, interaction, style, and metadata, drawing from both visualization theory and practical system needs. Derived properties cover both data-centric computational metrics and human-centric perceptual aspects.

The framework connects what to compare with how comparisons are operated through representation modalities. Appropriate representation forms the basis of effective retrieval, and this principle applies to visualization retrieval as well. The chosen representation format (e.g., declarative specification, raster image) dictates which aspects are captured and which similarity criteria are accessible for comparison. Based on the information content and visualization determinism, we categorize the existing representation modalities into four groups: raster image, vector image, specification, and natural language description.

We analyze several visualization retrieval work with Safire to demonstrate its practical value in clarifying similarity. We found that the choice of representation is not only an implementation detail but a decision that shapes the possibilities and limitations of the retrieval process. Drawing from these observations, we provide recommendations and discuss implications in the bigger context of retrieval, multimodal learning, AI application and reproducibility. Our contributions are two-fold:

- A similarity framework for visualization retrieval, Safire, outlining both comparison criteria and representation modalities. This conceptual model serves as a practical tool for visualization retrieval designers to clarify design choices and align similarity dimensions with specific use cases.

- Application of Safire to analyze existing visualization retrieval systems, highlighting how different aspects of similarity and representation are prioritized in different use cases.

2 Related Work

This section presents related work to the formulation of the framework as a whole. Specific related work to the finer details will be introduced in the corresponding places. The formulation of Safire as a framework was greatly inspired by how the Nested Model [16] frames different facets of visualization design, along with its extension [15] for inter- and intra-level blocks. FaEvR [29] provides an exemplar model that gathers insights from real-world visualizations to build a framework, then applies this framework back to analyze these visualizations from a different angle.

In terms of similarity, GraphScape [10] provides an example of visualization similarity based on transition cost between graph nodes: the distance between mark types indicates the estimated transition cost between them. ChartSeer [35] improved the idea of using embeddings for comparison and employs deep learning techniques to map charts to semantic vectors to measure chart similarity and generate charts. Cavallo and Demiralp [2] introduced a tool for guided visual clustering analysis, based on the premise that clustering facilitates grouping data points by similarity. We also build upon the idea of a visualization workflow using D3 [1]: from imperative programming, to vector graphics (SVG) with interactions [4, 17, 18], to vector/raster image export [14]. The foundation of our work is developed and informed by insights from prior visualization retrieval systems [3, 6, 8, 11, 21, 25, 31, 32, 33].

3 Safire: Similarity Framework for Visualization Retrieval

3.1 Criteria for Comparison

In our framework, the criteria answer the question of 'what' aspects should be compared for understanding similarity between visualizations. As presented in the top panel of Figure 1, we distinguish primary facets that directly contribute to constructing a visualization, from derived properties that are extracted after a visualization is built. This distinction acknowledges the fundamental difference between the contributing parameters that define how a visualization is created and the emergent characteristics that can only be observed in the final visual output. The following sections elaborate on how we developed these criteria.

3.1.1 Primary Facets

Our framework integrates criteria from theoretical visualization models with empirical retrieval systems, resulting in the five primary comparison facets: data, visual encoding, interaction, style, and metadata. We ground our approach in fundamental models of visualization design, particularly the nested model by Munzner [16] and its subsequent extensions to inter-level and intra-level blocks [15] by Meyer et al., which systematically deconstruct visualization design into core elements constructing a visualization. These models conceptualize visualization creation as a cascade of design decisions transforming domain problems into data-task abstractions, visual encodings, and implementations.

By analyzing these distinct design layers, we identify the first two fundamental comparison groups: (1) underlying **data** and (2) **visual encoding** that maps data attributes to visual features. Given the increasing importance of interactivity in visualization workflows [5, 7, 20], we deem it only appropriate to include (3) **interaction** as a separate dimension focused on user-centric exploration. Observations from practical visualization retrieval systems and broader design considerations [8, 31, 25, 19, 23] emphasize the importance of including (4) **visual styles** and (5) **contextual metadata** as additional criteria. The criteria are defined as follows:

Data Covering data-related properties, including transformation methods, parsing, data types, and aggregation parameters (e.g., binning size). This criterion facilitates searching for visualization examples handling specific data types or wrangling approaches.

Visual Encoding Representing the mapping of data to visual attributes, such as mark types, layout structures, and visual channels to encode value (e.g., bar height, circle radius). This criterion enables identification of visually similar representations, such as bar charts using bar length to indicate magnitude of value.

Interaction Capturing user interactivity with visual elements, including brushing and linking, click, details-on-demand features. This criterion supports exploration of interactive techniques, e.g., linking overview with detailed views following user selection.

Style Corresponding to non-data-encoding visual attributes [8] that contribute to aesthetic and perceptual aspects, including typography, background colors, and decorative elements. This criterion facilitates discovery of visual language applications, e.g., similar color palette usage across different contexts.

Metadata Comprising information that describes and contextualizes the visualization, including titles, subtitles, legends, and annotations. This criterion supports identification of effective approaches for enhancing visualization comprehensibility through supplementary elements.

It is important to note that these five primary facets are mutually non-exclusive. Depending on the specific domain problem and task, an attribute can belong to multiple categories. For example, stroke width can be visual encoding when it corresponds to value magnitude, or style when its purpose is to enhance legibility.

3.1.2 Derived Properties

Having established primary facets that define visualization construction, our framework now addresses derived properties: features extracted or computed from the resulting visualization. This characterization aligns with the role of visualization in visual analytics (VA) workflows: providing the means of communicating about data and information, where humans and machines cooperate [9]. Inspired by the systematic considerations in VA by Sun et al. [28], we divide derived properties into two categories:

Data-centric Measure Referring to computational properties derived from data, designed for analytical interpretation. Examples include distribution, outliers, and cluster-related measures [33]. This criterion enables finding visualizations with specific computational targets, topologies, or statistical measures.

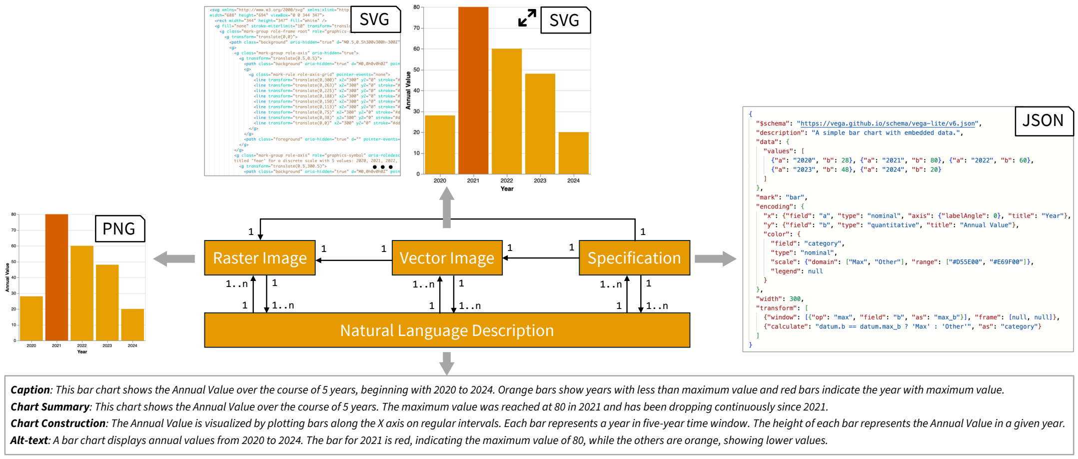

Figure 2: Visualization representation across four modalities: a Vega-Lite JSON specification (right) rendered as SVG vector–with accompanying SVG markup, and PNG raster images, along with multiple natural language descriptions. Specification, vector, and raster formats maintain 1:1 mapping relationships (directed arrows), while natural language enables one-to-many interpretations (multiple text examples).

Human-centric Measure Characterizing how users perceive information, involving human cognitive processing of visual information. Examples include metrics for perceptual similarity [22], reflecting how observers group plots based on concepts like orientation, edges, or density. This criterion supports identifying visualizations grouped based on human perceptual judgments.

3.2 Representation Modalities

Representation modality defines how visualization information is represented. Before creating vector embeddings, it is essential to characterize raw modalities that capture different aspects of the information as illustrated in Figure 2. Common modalities include raster images (PNG, JPG), vector graphics (SVG), and declarative JSON specifications (Vega-Lite [24], Gosling [13]).

We categorize the raw modalities along two dimensions: information content and visualization determinism, as shown in the lower half of Figure 1. Higher information content enables user to recreate the visualization more accurately and extract more meaningful information. Visualization determinism refers to the degree to which a representation format guarantees a singular, consistent visual outcome without requiring additional interpretation. These two dimensions are essential in retrieval context due to immediate association with how much information is captured and how consistently that information translates to a specific visual form. We define the representation as follows:

Raster Image Rendering a visualization as a fixed grid of pixels (e.g., PNG, JPG). Each pixel stores only color information without preserving data relationships or visual mark semantics. As a raw modality for visualization retrieval, raster images require visual feature extraction via a pre-defined taxonomy or deep learning models to interpret chart types [32, 31]. While suitable for image-based retrieval or search-by-sketch scenarios, they lack structural relationship to the underlying data.

Vector Image Preserving visualization geometry through scalable paths, shapes, and text elements that can be scaled without loss of quality. Examples include a SVG file of a scatter plot represents each point as a circle with properties like position, radius, and color. SVG uses HTML-tag markup that, along with its visual rendering, can enable structure-aware retrieval [11].

Specification Defining the visualization's structure, data bindings, encoding rules and potentially interaction, at a high level with predefined schema. Specifications offer machine-readable access to high-level semantics. They are ideal for precise matching and retrieval based on structural similarity or query-by-example, including searching for *interaction*. Examples include retrieval systems for Chart2Vec JSON [3], Gosling JSON [19], and recommender system with Tableau Workbook XML [21].

Natural Language (NL) Description Capturing a semantic content of a visualization using NL to convey and contextualize insights. Examples include alt-text, which is the most abstracted, human-readable interpretation of the visualization [26, 27]. Other examples are captions (general interpretation), chart summaries (richer descriptions of patterns, insights, and context, but may lack encoding information), and chart construction (procedural instructions for building charts–similar to grammar-based specification but in NL). NL descriptions inherently contain ambiguity: visualizations can have multiple captions for different audiences, and different charts of the same data may deliver the same message in their summaries.

Figure 2 demonstrates the interconnections between these modalities. A Vega-Lite JSON specification defines the visualization structure, rendered as an SVG vector image (along with its markup snippet) and captured as a PNG raster image. The figure also shows four types of NL descriptions: caption, chart summary, chart construction, and alt-text. While specification, vector, and raster representations maintain a 1:1 mapping (along directed arrows), NL descriptions exhibit one-to-many relationships, as shown by the four different textual representation types.

4 Application Examples

In this section, we analyze several previous visualization retrieval systems in terms of our framework, to provide examples of framing the retrieval problem with regard to visualization similarity in Safire.

4.1 Searching D3 Visualizations

Hoque and Agrawala present a system for searching D3 visualizations by visual style and structure [8], as shown in Figure 3. Their retrieval system deconstructs and indexes visualizations based on data, visual encoding, style, and metadata criteria. The system generates a representation similar to a Vega-Lite [24] specification from each visualization, which also serves as the query input format. NL text and metadata are indexed separately alongside the deconstructed specification. This work demonstrates flexibility of specification in encoding chart semantics. By extracting both data- and non-data-encoding attributes, this approach enables comprehensive searches across visual and structural dimensions, even with partial specifications.

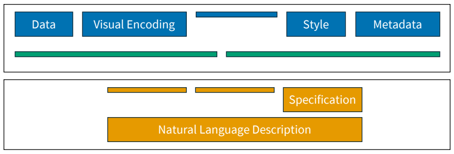

Figure 3: Searching D3 visualizations [8]

4.2 Multimodal Retrieval of Genomics Visualizations

Nguyen et al. [19] present a multimodal retrieval system for genomics data visualizations, covering all five comparison criteria: data, visual encoding, interaction, style, and metadata. Their system uses three modalities: raster images, Gosling [13] grammar specifications, and NL descriptions (both alt-text and LLM-enriched versions).

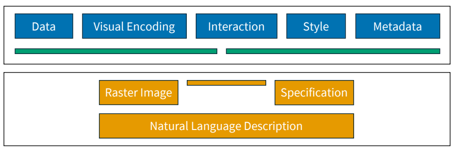

Figure 4: Multimodal Retrieval of Genomics Data Visualizations [19]

The multimodal representations approach enables the system to capture both the semantic structure and visual characteristics of genomics visualizations, supporting flexible querying by example images, text queries, or specification-based queries.

4.3 WYTIWYR: User Intent-Aware Framework

Xiao et al. present WYTIWYR [31], a retrieval tool comparing charts based on visual attributes and style cues. To better understand user intents, the authors first conducted a preliminary study to formulate chart attributes, along three dimensions: colormap, data trends and view layout.

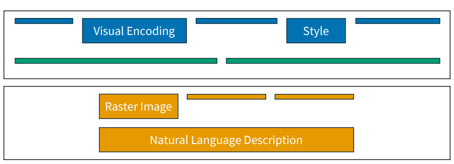

Figure 5: WYTIWYR: User Intent-Aware Framework [31]

The system processes raster images as visualization inputs with optional text prompts expressing user intent, combining them via a CLIP-based multi-modal encoder.

4.4 VAID: Indexing View Designs in VA system

Ying et al. present VAID [33], an index structure for complex and composit visualizations. VAID compares both primary facets (data-related, visual encoding, and style) and derived data-centric measures: graph-related metrics (e.g., clusters, topology) and tabular structures (e.g., correlation, distribution, outliers).

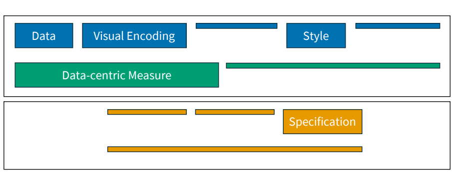

Figure 6: VAID: Indexing View Designs in VA system [33]

Though VAID provides multiple criteria for comparison, it indexes views solely through specifications, using an extended Vega-Lite grammar, demonstrating the comprehensiveness of specification-based representation.

5 Discussion

NL Description Is Highly Nondeterministic. As opposed to specifications and vector images that map one-to-one to a single rendering, NL descriptions introduce ambiguity, creating one-to-many relationships between descriptions and visualizations. The same chart can be described in many different ways: some descriptions detail data bindings and encodings, while others focus on broader patterns or insights. Specifically, *chart construction* involves procedural instructions that essentially function as specifications written in NL rather than formal grammar. On the other hand, a *chart summary* can convey insights and go beyond the visual channel, e.g., mark type. Visual features in this case are merely the medium to extract the message. These observations complement the four-level model of semantic content [12] by adding practical nuance from the ambiguous nature of descriptions, depending on the communicative intent and context. Natural language descriptions associated with visualizations presents a rich area for further investigation.w

Representation Modality Shapes Retrieval Capabilities and Reproducibility. We find that data- and interaction-related criteria are only comparable when specification is involved. In fact, specification is one of the most versatile modalities, encompassing all five primary facets (Section 4.2) and multiple data-centric measures (Section 4.4). Vector images bridge between raster images and specifications but often feature complex, highly nested markup. Meanwhile, NL descriptions can capture high-level insights and context missing in other modalities, yet their inherent ambiguity challenges precise matching and retrieval. Recognizing these trade-offs, multimodal retrieval presents a promising approach that integrates complementary strengths from each modality to create a more comprehensive understanding. In terms of reproducibility and information content, specifications rank highest, followed by vector images and raster images. This aspect is essential for visualization authoring [30] where retrieved examples serve both as inspiration and templates for adaptation. Specifications enable efficient programmatic modifications, while raster images serve as great visual references but with limited editability.

LLM Guidance in the Context of AI. The five primary facets of visualization can help guide large language models (LLMs) to focus on key elements and steer their interpretation of charts toward clearer, more accurate understanding. By structuring prompts around these facets: data, visual encoding, interaction, style, and metadata, we can direct the LLMs' attention on where they might otherwise overlook. Furthermore, these same facets create a systematic way for evaluating LLM performance in visualization comprehension tasks, revealing which aspects remain challenging and may require additional prompt engineering or model training for improvement.

6 Conclusion

In this work, we introduced Safire, a framework for modeling visualization similarity that connects comparison criteria with representation modalities. By distinguishing five primary facets: data, visual encoding, interaction, style, and metadata, we provide a structured way to define what aspects make visualizations similar. We then linked these criteria to four common representation formats: raster images, vector graphics, specifications, and natural language, each with different implications for retrieval, comprehension, and reproducibility. Applying Safire to existing visualization retrieval systems demonstrated its practical value in guiding design choices in visualization retrieval and aligning similarity dimensions with intended use cases. While comparing algorithms is beyond our current scope, prior work on formal algorithm evaluation [34] suggests promising directions for future research.

Acknowledgments

This work was supported in part by the National Institutes of Health (R01HG011773) and the Advanced Research Projects Agency for Health (AY2AX000028).

References

[1] M. Bostock, V. Ogievetsky, and J. Heer. D³ data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011. 2

[2] M. Cavallo and Ç. Demiralp. Clustrophile 2: Guided visual clustering analysis. *IEEE transactions on visualization and computer graphics*, 25(1):267–276, 2018. 2

[3] Q. Chen, Y. Chen, R. Zou, W. Shuai, Y. Guo, J. Wang, and N. Cao. Chart2vec: A universal embedding of context-aware visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 1, 2, 3

[4] T. Dang, H. N. Nguyen, and V. Pham. WordStream: Interactive Visualization for Topic Evolution. In J. Johansson, F. Sadlo, and G. E. Marai, eds., *EuroVis 2019 - Short Papers*. The Eurographics Association, 2019. 2

[5] E. Dimara and C. Perin. What is interaction for data visualization? *IEEE transactions on visualization and computer graphics*, 26(1):119–129, 2019. 2

[6] L. Ferres, M. Dumontier, and N. Villanueva-Rosales. Semantic query answering with time-series graphs. In *2007 Eleventh International IEEE EDOC Conference Workshop*, pp. 117–124. IEEE, 2007. 2

[7] 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. 2

[8] E. Hoque and M. Agrawala. Searching the visual style and structure of d3 visualizations. *IEEE transactions on visualization and computer graphics*, 26(1):1236–1245, 2019. 1, 2, 3, 4

[9] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In *Information visualization: Human-centered issues and perspectives*, pp. 154–175. Springer, 2008. 2

[10] Y. Kim, K. Wongsuphasawat, J. Hullman, and J. Heer. Graphscape: A model for automated reasoning about visualization similarity and sequencing. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 2628–2638, 2017. 2

[11] H. Li, Y. Wang, A. Wu, H. Wei, and H. Qu. Structure-aware visualization retrieval. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2022. 2, 3

[12] A. Lundgard and A. Satyanarayan. Accessible visualization via natural language descriptions: A four-level model of semantic content. *IEEE transactions on visualization and computer graphics*, 28(1):1073–1083, 2021. 4

[13] S. LYi, Q. Wang, F. Lekschas, and N. Gehlenborg. Gosling: A grammar-based toolkit for scalable and interactive genomics data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):140–150, 2021. 3, 4

[14] M. Mauri, T. Elli, G. Caviglia, G. Uboldi, and M. Azzi. Rawgraphs: A visualisation platform to create open outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*, CHItaly '17. Association for Computing Machinery, New York, NY, USA, 2017. 2

[15] M. Meyer, M. Sedlmair, and T. Munzner. The four-level nested model revisited: blocks and guidelines. In *Proceedings of the 2012 BELIV workshop: Beyond time and errors-novel evaluation methods for visualization*, pp. 1–6, 2012. 2

[16] T. Munzner. A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics*, 15(6):921–928, 2009. 2

[17] H. N. Nguyen, F. Abri, V. Pham, M. Chatterjee, A. S. Namin, and T. Dang. Malview: Interactive visual analytics for comprehending malware behavior. *IEEE Access*, 10:99909–99930, 2022. 2

[18] H. N. Nguyen, T. Dang, and K. A. Bowe. Wordstream maker: A lightweight end-to-end visualization platform for qualitative time-series data, 2022. 2

[19] H. N. Nguyen, S. L'Yi, T. C. Smits, S. Gao, M. Zitnik, and N. Gehlenborg. Multimodal retrieval of genomics data visualizations, Apr 2025. doi: 10.31219/osf.io/zatw9_v1 1, 2, 3, 4

[20] H. N. Nguyen, C. M. Trujillo, K. Wee, and K. A. Bowe. Interactive qualitative data visualization for educational assessment. In *Proceedings of the 12th International Conference on Advances in Information Technology*, pp. 1–9, 2021. 2

[21] M. Oppermann, R. Kincaid, and T. Munzner. Vizcommender: Computing text-based similarity in visualization repositories for content-based recommendations. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):495–505, 2020. 1, 2, 3

[22] A. V. Pandey, J. Krause, C. Felix, J. Boy, and E. Bertini. Towards understanding human similarity perception in the analysis of large sets of scatter plots. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 3659–3669, 2016. 3

[23] B. Saleh, M. Dontcheva, A. Hertzmann, and Z. Liu. Learning style similarity for searching infographics. In *Proceedings of the 41st Graphics Interface Conference*, GI '15, p. 59–64. Canadian Information Processing Society, CAN, 2015. 2

[24] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer. Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics*, 23(1):341–350, 2016. 3

[25] V. Setlur, A. Kanyuka, and A. Srinivasan. Olio: A semantic search interface for data repositories. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–16, 2023. 2

[26] T. C. Smits, S. L'Yi, A. P. Mar, and N. Gehlenborg. Altgosling: automatic generation of text descriptions for accessible genomics data visualization. *Bioinformatics*, 40(12):btae670, 2024. 3

[27] T. C. Smits, S. L'Yi, H. N. Nguyen, A. P. Mar, and N. Gehlenborg. Explaining unfamiliar genomics data visualizations to a blind individual through transitions. In *2024 1st Workshop on Accessible Data Visualization (AccessViz)*, pp. 24–28. IEEE, 2024. 3

[28] M. Sun, Y. Ma, Y. Wang, T. Li, J. Zhao, Y. Liu, and P.-S. Zhong. Toward systematic considerations of missingness in visual analytics. In *2022 IEEE Visualization and Visual Analytics (VIS)*, pp. 110–114. IEEE, 2022. 2

[29] S. Vaidya and A. Dasgupta. Knowing what to look for: A fact-evidence reasoning framework for decoding communicative visualization. In *2020 IEEE Visualization Conference (VIS)*, pp. 231–235. IEEE, 2020. 2

[30] A. van den Brandt, S. L'Yi, H. N. Nguyen, A. Vilanova, and N. Gehlenborg. Understanding visualization authoring techniques for genomics data in the context of personas and tasks. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 4

[31] S. Xiao, Y. Hou, C. Jin, and W. Zeng. Wytiwyr: A user intent-aware framework with multi-modal inputs for visualization retrieval. In *Computer Graphics Forum*, vol. 42, pp. 311–322. Wiley Online Library, 2023. 1, 2, 3, 4

[32] Y. Ye, R. Huang, and W. Zeng. Visatlas: An image-based exploration and query system for large visualization collections via neural image embedding. *IEEE Transactions on Visualization and Computer Graphics*, 2022. 1, 2, 3

[33] L. Ying, A. Wu, H. Li, Z. Deng, J. Lan, J. Wu, Y. Wang, H. Qu, D. Deng, and Y. Wu. Vaid: Indexing view designs in visual analytics system. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2024. 2, 4

[34] Z. Zeng, P. Moh, F. Du, J. Hoffswell, T. Y. Lee, S. Malik, E. Koh, and L. Battle. An evaluation-focused framework for visualization recommendation algorithms. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):346–356, 2021. 4

[35] J. Zhao, M. Fan, and M. Feng. Chartseer: Interactive steering exploratory visual analysis with machine intelligence. *IEEE Transactions on Visualization and Computer Graphics*, 28(3):1500–1513, 2020. 2