Readability as a multi-measure construct in data visualization

Anne-Flore Cabouat
Université Paris-Saclay, Inria, LISN, Télécom Paris, Institut Polytechnique de Paris, i3, CNRS
France
given-name.family_name@inria.fr

Samuel Huron
Télécom Paris, Institut Polytechnique de Paris, I3 CNRS
France
firstname.family_name@telecom-paris.fr

Tobias Isenberg, Petra Isenberg
Université Paris-Saclay, Inria, LISN, CNRS
France
firstname.family_name@inria.fr

Abstract

In this paper, we argue that readability cannot be meaningfully discussed without considering multiple complementary measures, and that relying on a single measure constitutes an epistemological choice that constrains the conclusions that can be drawn.

CCS Concepts

• **Human-centered computing → Visualization theory, concepts and paradigms**; **Hypertext / hypermedia**; **Empirical studies in HCI**.

Keywords

Readability, Evaluation, Data visualization

1 Introduction

While reading is often associated with written text, readers routinely engage with non-textual elements such as graphs, diagrams, tables, and other visual elements accompanying text, which they also describe as being read [7]. Reading technologies aim to support readers in extracting information from both textual and visual representations, making readability a central concern in related research areas, such as text display or data visualization. Across domains, researchers study readability through indirect measures that can only capture some of the reading process facets [5], using, for example, metrics based on the representation's structural or formal properties, task performance measures, or subjective experience reports. As a result, what it means for an artifact to be readable often depends on how reading itself is framed and measured.

Data visualization research offers a clear illustration of this issue. Visualization researchers approach readability in distinct ways: as an intrinsic property of the visual artifact, as reflected in task-dependent user performance and behavioral metrics, or through subjective accounts of readers' experiences. Each of these approaches foregrounds different aspects of reading while obscuring others. As a result, conclusions about whether a visualization is readable for certain tasks and readers depend strongly on how readability is defined and measured [9].

This point raises a broader methodological question: *What do different readability measures actually allow researchers to claim about the ease of reading?* We argue that readability cannot be meaningfully discussed without considering multiple complementary measures—relying on a single measure constitutes an epistemological choice that constrains the conclusions that can be drawn. We examine how readability is currently specified in research on visualization reading and discuss why acknowledging the measurement choices is needed for interpreting empirical results.

2 Background: How is readability evaluated in data visualization research?

Early text readability research treated reading difficulty as an intrinsic property of the text, assessed through formulas relying on surface linguistic features such as sentence length, word frequency, and syntactic features. While such formulas offer simple and scalable indicators of cognitive processing constraints imposed by a text's form, they have also been criticized for their weak relationship to actual reading comprehension processes and for neglecting reader and task effects [4]. In response to these limitations, readability research diversified its evaluation methods, incorporating performance-based measures, comprehension assessments, behavioral indicators, and subjective ratings. This diversification resulted in a heterogeneous landscape of non-equivalent measures [5]. In this context, the very concept of readability resists a unified definition because its meaning shifts depending on how it is measured.

We observe a similar epistemological tension in data visualization research, where readability has likewise been assessed through a variety of measurement types. The most commonly reported considerations are participant layout metrics, task-based evaluations, and subjective feedback. In the next paragraphs, we briefly review each of those measures and reflect on the divergent epistemological stances they represent.

2.1 Layout metrics

Stemming from efforts to optimize node-link diagram layout, readability in data visualization was early-on associated to metrics approximating representations' desired visual properties [12]. In node-link visualizations, these are called aesthetics metrics, and they include edge crossings, node overlapping, neighborhood preservation, or global symmetry [6, 22]. Beyond node-link representations, Giovannangeli *et al.* [16] proposed a drawing algorithm to improve the readability of scatterplots, which they frame as a "visibility" issue. Goffin *et al.* [17] computed metrics based on objects bounding boxes to quantify how different placements of word-scale visualizations affected the readability of documents. Similar to formulas of text readability, such approaches reflect an epistemological stance that readability is intrinsic to the artifact, and do not consider individual factors influencing readers' visual perception and understanding, such as their cultural background or visual abilities.

2.2 Task-based evaluations

Task-based evaluations aim to infer readability from participants' behavior while they perform predefined reading tasks. These evaluations rely on two broad families of measures: process-oriented measures and outcome-oriented measures.

Process-oriented measures focus on how reading activities unfold during task execution. Completion time is commonly used as a proxy for processing efficiency, under the assumption that more readable visualizations require less effort to decode, and that a lower amount of effort translates into faster task completion. Eye-tracking recordings extend this approach by providing spatio-temporal gaze and physiological data that characterize readers' visual behavior while reading a visualization [20].

Outcome-oriented measures, on the other hand, focus on success for specific reading tasks, such as retrieving the value of a data point, detecting a trend, or comparing visual elements in a representation. Such measures typically include accuracy (i. e., error rates), and sometimes numerical precision.

Although a few studies evaluated readability based solely on error rates or numerical precision of answers (e. g., [26]), more often researchers collect and analyze both time and accuracy. For example, Bu *et al.* [8] assessed the readability of stacked area graphs using accuracy measures for tasks designed by Thudt *et al.* [27], including reading the thickness of an individual layer, the variation of a layer's thickness, and the overall thickness of aggregated layers.

Within this framing, outcome-oriented measures capture the extent to which readers correctly extract relevant information, while process-oriented measures characterize the efficiency and dynamics of the underlying reading activity. Task-based evaluations thus adopt a behavioral stance, in which readability is not measured directly but inferred from observable performance. Researchers adopting this stance (e. g., [15, 27, 28]) define readability as the extent to which a visual representation supports readers in completing a task successfully and efficiently.

Task-based measures, however, not only reflect the efficiency of interaction and visual design choices but also other factors such as the reader's prior knowledge, beliefs, and task strategies as well as task complexity. As a result, assessing readability through task performance alone generally does not allow researchers to draw clear conclusions about the efficiency of a visual design or an interaction technique, except under highly controlled experimental conditions (e. g., large sample sizes, constrained tasks, or controlled population parameters).

2.3 Subjective experience reports

Finally, visualization researchers commonly use subjective assessments in experiments such as rating scales [10], sketching observations [21], interviews [30], think-aloud protocols [25], and open-ended comments in surveys [2]. Such methods allow researchers to collect insights into participants' reading strategies and to capture how readable a visualization *feels* to a reader, rather than how efficiently or accurately information can be extracted.

Subjective reports are often motivated by the observation that readability cannot be reduced to performance alone, echoing broader discussions that distinguish usability from user experience [1, 29]. Accounts of subjective reader experiences allow researchers to access facets of the reading experience that remain conflated in behavioral measures, such as perceived effort, confusion, or confidence in answers. As such, they contribute to a more holistic understanding of how people engage with data visualizations [24].

At the same time, subjective assessments reflect readers' interpretations of their own experience and behavior rather than directly observable processes. Reported readability may be biased by a range of individual factors, including expertise with the represented data, familiarity with a type of visualization, aesthetic preferences, cultural background, and meta-cognitive abilities, or by contextual and experimental factors such as social desirability and post-rationalization effects. Like task-based evaluations, subjective measures adopt a reader-centered stance. They reflect, however, a different epistemological viewpoint: rather than inferring readability from task efficiency or performance outcomes, subjective measures frame readability as a property of the experience lived and articulated by the reader.

Taken together, these approaches illustrate that readability is conceptualized in fundamentally different ways across evaluation methods: layout metrics treat it as an intrinsic property of the visual artifact; task-based evaluations infer it from behavioral observations; and subjective reports frame it as a lived experience. No single method is sufficient to characterize readability, as each captures distinct and non-interchangeable aspects of reading.

3 Readability as a multi-measure construct

As highlighted in the previous section, readability in data visualization can be evaluated through a collection of measures that each capture different facets of reading. Layout metrics, task-based evaluations, and subjective experience reports do not provide equivalent evidence on readability, nor do they answer the same research questions. This diversity points to conceptual differences: readability is not a directly observable property, but a construct whose meaning depends on how reading itself is framed and examined. Yet, such underlying assumptions are rarely made explicit in visualization readability research. Treating readability as a single measurable quantity further obscures these conceptual differences and limits results interpretation. We therefore argue that readability should be approached through multiple complementary measures.

Our perspective challenges approaches that infer readability solely from properties of the representation and generic assumptions about perceptual decoding. Reading is not merely seeing: for both textual and graphical representations, it involves interpreting established representational systems, such as writing systems for language or coordinate systems for quantitative relationships. Readers have to mobilize their prior knowledge [14, 19], expectations [31], and attention [18] in order to retrieve meaningful information. From this perspective, **readability does not reside solely in the visual artifact itself, but in how a reader can leverage this artifact for their reading goal.** This view echoes critiques in text readability research, which argue that readability resides in the use of a text rather than in the text alone [4]. Similarly, for visualizations, readability resides in the visualization activity, not in isolation in the visual artifact.

Task-based performance measures and subjective experience reports move the assessment of readability beyond the visual artifact itself by anchoring it in readers' actions and visualization experiences. We can find evidence that readability is a multi-dimensional construct within each of these categories. For instance, task-based

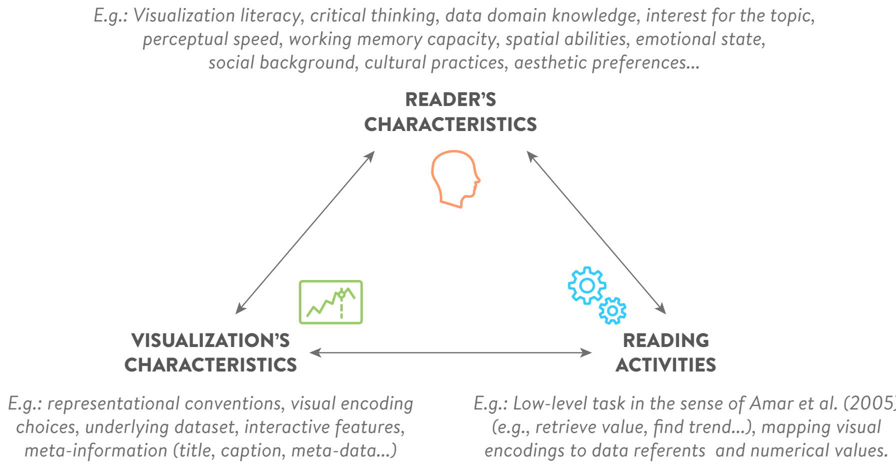

Figure 1: A triadic reciprocal representation of factors influencing readability in data visualization. We adapt this figure from Cabouat *et al.*'s work [9], under the rules of the CC BY 4.0 license.

evaluations often distinguish between efficiency-related indicators (e. g., reading time, gaze fixations, and saccades) and outcome-related indicators (e. g., answer correctness or precision). Such measurements are not conflated but presented as separate evidence. Prior work on perceived readability further shows that subjective judgments are not unidimensional: the PREVis questionnaire [10] exhibits a four-factor structure, including judgments about the artifact's visual clarity, task-related subjective experience (e. g., ease of retrieving data values and perceiving data patterns), and perceived ease of interpreting and making sense of the visual encodings.

These observations suggest that readability cannot be attributed to a single source, nor captured through a single type of measure. We thus frame readability as an emergent property of the interaction between three components: the visual artifact, the reader, and the reading activity (see Fig. 1). The visual artifact provides an underlying dataset and visual encoding choices, emerging from sociotechnical processes that include institutional data practices, representational norms, and the designer's situated decisions. The reader brings perceptual abilities, prior knowledge, skills, a sociocultural background, and other individual characteristics (e. g., personality traits, motivation). Finally, the reading activity itself defines tasks and contextual constraints (e. g., collaborative settings or time-constrained activities). Readability does not reside in any one of these components alone, but emerges from their reciprocal influence during a visualization reading experience.

This triadic perspective helps us explain why different readability measures foreground different aspects of reading and why no single measure can fully characterize readability across readers and contexts. While visualization user studies frequently collect heterogeneous forms of evidence, these measures are often reported side by side rather than interpreted together as evidence about readability. Typical analyses enumerate differences in speed, accuracy, or preference across conditions, but rarely articulate what these differences jointly imply about how readable a visualization is, for whom, and under which conditions. As a result, rather than addressing the complexity of readability, current practices often leave it implicit. This gap calls for a more systematic examination of how different readability measures relate to one another, what aspects of reading they foreground or obscure, and how they can be interpreted together.

4 A research program on readability evaluation in data visualization

If readability cannot be reduced to a single source or a single measure, then evaluating readability requires making explicit what aspects of reading are being captured, for whom, and under which conditions. From this perspective, readability can be understood as a lens through which heterogeneous measures contribute partial views of a reading situation. In this section, we outline a research program for readability evaluation in data visualization that builds on this perspective. First, we should examine how different readability measures foreground or obscure distinct aspects of reading. We then need to consider how individual differences interact with these measures and shape their interpretation. Finally, we should investigate how to interpret heterogeneous readability measures together through principled triangulation.

4.1 What aspects of reading are highlighted or hidden by different readability measures?

A first line of inquiry concerns the relationship between readability measures and the specific aspects of reading they make visible or leave unobserved. As we highlighted in Sect. 2, layout metrics foreground structural properties of the visual artifact, task-based evaluations emphasize efficiency and correctness in information extraction, eye-tracking data can highlight patterns of visual attention, and subjective reports capture readers' strategies and experiences.

In order to shift the question from whether a visualization is readable to which aspects of reading are being assessed, we should study situations in which different measures diverge. For example, we can characterize situations in which fast task completion co-occurs with

[1]In Fig. 1, we mainly refer to low-level tasks as they were defined in Amar *et al.*'s taxonomy [3].

low subjective confidence, or in which high visualization literacy scores coexist with high error rates. Such divergences will expose which aspects of reading a given measure implicitly prioritizes.

Relatedly, readability measures are also often sensitive to task framing. For example, completion time may indicate decoding efficiency in tightly constrained lookup tasks, but reflect other processes in open-ended reading contexts, such as constructing an initial understanding of the artifact or forming questions about the data. Examining how the same time-based measure behaves in free exploration versus directed lookup can therefore clarify what the measure captures in each reading context.

4.2 How do individual differences interact with readability measures?

A second line of inquiry concerns how readability measures interact with individual differences among readers. While readability evaluations often aim to characterize properties of a visualization artifact, reader-based measures and reports are necessarily mediated by readers' perceptual abilities, prior knowledge, familiarity with visualization encodings, and other individual characteristics. Yet, how these factors modulate different readability measures remains poorly understood. For instance, domain expertise or familiarity with a visualization type may differentially affect task performance, perceived ease, and confidence in answers, leading to situations where measures diverge not because of the visualization itself, but because readers engage with it differently.

Similarly, personality traits such as conscientiousness can affect the way in which study participant engage in tasks [32], and language fluency can affect their ability to process information or report their experience [23]. Social factors, such as educational background, graphical conventions, and norms around expressing uncertainty or confidence, may further shape how readability is experienced and reported.

It is therefore necessary to examine how individual characteristics shape what a given measure is able to capture, and how the evidential value of that measure varies across populations. Doing so is essential to distinguish between variability in readability arising from the visual artifact, the reader, or the task, and distortions introduced by the measurement method itself. An example would be a case in which experts confidently rate a visualization on a questionnaire, whereas novices provide more hesitant or inconsistent responses—not necessarily because they find it less readable, but because they are less at ease with reporting their opinion. Studying how various measures amplify or mask such discrepancies can clarify whether observed effects should be interpreted as limitations of a design, of a measure, or of its applicability to a given population. More broadly, this line of work invites a shift from treating individual variability as noise to be controlled toward examining it as a key factor shaping the interpretation of readability measures.

4.3 How to interpret different readability measurements together?

A final line of inquiry concerns how different readability measures should be interpreted together. In mixed-methods research, this problem is commonly addressed through triangulation, understood as the deliberate combination of different data sources or methods to examine a phenomenon from multiple perspectives [11, 13]. In this view, divergence between measures is not necessarily a problem to be resolved, but can be seen as a source of insight into the complexity of the underlying phenomenon. An important research question is thus how different configurations of measures should be interpreted in relation to the reading situation under study. When measures agree, which aspects of reading are rendered interpretable, and under which assumptions? Conversely, when they diverge, what does this reveal about the limits of each measure and the reading situation itself? Echoing our argument in Sect. 4.1, this shifts the focus from whether measures agree to why they agree or conflict in particular reading contexts.

Similarly, rather than assuming a fixed sequence in which one type of measure explains another (e.g., performance first, subjective reports second), we should consider them as distinct lenses on the same reading activity, each shaped by its own assumptions and sensitivities. Measures can inform one another without being hierarchically or temporally ordered, for instance by using qualitative material to reinterpret quantitative patterns, or observational data to challenge subjective reports. A productive research direction is therefore to develop ways of articulating and comparing the epistemic status of readability measures within a study: what they are taken to stand for, what they are not intended to capture, and how the researchers choose to articulate them.

This would support more transparent reasoning about readability without assuming that multiple data sources should converge on a single answer. Such an approach also emphasizes the situated nature of readability and the context-dependent validity of empirical findings rather than aiming to derive universally applicable design prescriptions or guidelines.

5 Conclusion

In this position paper, we argue that different measures illuminate different aspects of reading, each grounded in distinct assumptions about how meaning is extracted from visual representations. Treating these measures as complementary is therefore essential to making sense of empirical results. Only by making their underlying assumptions explicit can readability findings be meaningfully interpreted and compared across studies. Building on this view, we outline a research program that examines how different measures foreground distinct facets of reading, how individual differences shape their interpretation, and how heterogeneous measures can be meaningfully interpreted together. Making such relationships explicit enables a more coherent and cumulative understanding of readability in visualization, in which empirical heterogeneity becomes a resource for theory building.

References

[1] Anshu Agarwal and Andrew Meyer. 2009. Beyond usability: Evaluating emotional response as an integral part of the user experience. In *CHI Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, 2919–2930. doi:10/frgsks

[2] Kiran Ajani, Elsie Lee, Cindy Xiong, Cole Nussbaumer Knaflic, William Kemper, and Steven Franconeri. 2022. Declutter and focus: Empirically evaluating design guidelines for effective data communication. *IEEE Transactions on Visualization and Computer Graphics* 28, 10 (2022), 3351–3364. doi:10/gn984r

[3] Robert Amar, James Eagan, and John Stasko. 2005. Low-level components of analytic activity in information visualization. In *Proceedings of the IEEE Symposium on Information Visualization*. IEEE Computer Society, Los Alamitos, 111–117. doi:10/bwrm27

[4] Alan Bailin and Ann Grafstein. 2001. The linguistic assumptions underlying readability formulae: A critique. *Language & Communication* 21, 3 (2001), 285–301. doi:10/d6jngm

[5] Rebekah George Benjamin. 2012. Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review* 24, 1 (2012), 63–88. doi:10/bdjfkd

[6] Chris Bennett, Jody Ryall, Leo Spalteholz, and Amy Gooch. 2007. The aesthetics of graph visualization. In *Proceedings of the Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*. Eurographics Association, Goslar, 57–64. doi:10/gn9kwr

[7] Jacques Bertin. 1983. *Semiology of Graphics: Diagrams, Networks, Maps.* Esri Press, Redlands, California. https://www.esri.com/en-us/esri-press/browse/semiology-of-graphics-diagrams-networks-maps

[8] Chuan Bu, Quanjie Zhang, Qianwen Wang, Jian Zhang, Michael Sedlmair, Oliver Deussen, and Yunhai Wang. 2021. SineStream: Improving the readability of streamgraphs by minimizing sine illusion effects. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1634–1643. doi:10/ghv58n

[9] Anne-Flore Cabouat, Tingying He, Florent Cabric, Tobias Isenberg, and Petra Isenberg. 2024. Position paper: A case to study the relationship between data visualization readability and visualization literacy. In *Proceedings of the CHI Workshop "Toward a More Comprehensive Understanding of Visualization Literacy"*. Article 5, 10 pages. url:hal.science/hal-04523790.

[10] Anne-Flore Cabouat, Tingying He, Petra Isenberg, and Tobias Isenberg. 2025. PREVis: Perceived readability evaluation for visualizations. *IEEE Transactions on Visualization and Computer Graphics* 31, 1 (2025), 1083–1093. doi:10/njnz

[11] Nancy Carter, Denise Bryant-Lukosius, Alba DiCenso, Jennifer Blythe, and Alan J. Neville. 2014. The use of triangulation in qualitative research. *Oncology Nursing Forum* 41, 5 (2014), 545–547. doi:10/dmt5

[12] Peter Eades. 1984. A heuristic for graph drawing. *Congressus Numerantium* 42 (1984), 149–160.

[13] Nigel G. Fielding. 2012. Triangulation and mixed methods designs: Data integration with new research technologies. *Journal of Mixed Methods Research* 6, 2 (2012), 124–136. doi:10/gdtkbw

[14] Eric G. Freedman and Priti Shah. 2002. Toward a model of knowledge-based graph comprehension. In *Diagrammatic Representation and Inference. Proceedings of Diagrams*. Springer, Berlin, 18–30. doi:10/fr56t9

[15] Mohammad Ghoniem, Jean-Daniel Fekete, and Philippe Castagliola. 2005. On the readability of graphs using node-link and matrix-based representations: A controlled experiment and statistical analysis. *Information Visualization* 4, 2 (2005), 114–135. doi:10/bncqv6

[16] Loann Giovannangeli, Frederic Lalanne, Romain Giot, and Romain Bourqui. 2024. Guaranteed visibility in scatterplots with tolerance. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 792–802. doi:10/gtgz8h

[17] Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg. 2014. Exploring the placement and design of word-scale visualizations. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2291–2300. doi:10/f6qjwg

[18] Christopher Healey and James Enns. 2012. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics* 18, 7 (2012), 1170–1188. doi:10/ch7sc2

[19] Walter Kintsch. 1988. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review* 95, 2 (1988), 163–182. doi:10/cxhhsw

[20] Maurice Koch, Kuno Kurzhals, Michael Burch, and Daniel Weiskopf. 2023. Visualization psychology for eye tracking evaluation. In *Visualization Psychology*, Danielle Albers Szafir, Rita Borgo, Min Chen, Darren J. Edwards, Brian Fisher, and Lace Padilla (Eds.). Springer, Cham, Chapter 10, 243–260. doi:10/gtgz8m

[21] Carolina Nobre, Kehang Zhu, Eric Mörth, Hanspeter Pfister, and Johanna Beyer. 2024. Reading between the pixels: Investigating the barriers to visualization literacy. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, Article 197, 17 pages. doi:10/n98g

[22] Helen C. Purchase. 2002. Metrics for graph drawing aesthetics. *Journal of Visual Languages & Computing* 13, 5 (2002), 501–516. doi:10/fdqd3w

[23] Noëlle Rakotondravony, Priya Dhawka, and Melanie Bancilhon. 2023. Beyond English: Centering multilingualism in data visualization. In *Proceedings of the IEEE VIS Workshop on Visualization for Social Good.* Article 3, 5 pages. arXiv:2309.06659 doi:10/qq69

[24] Bahador Saket, Alex Endert, and John Stasko. 2016. Beyond usability and performance: A review of user experience-focused evaluations in visualization. In *Proceedings of the Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization.* ACM, New York, 133–142. doi:10/gfw4vt

[25] Michael Sedlmair, Annika Frank, Tamara Munzner, and Andreas Butz. 2012. RelEx: Visualization for actively changing overlay network specifications. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2729–2738. doi:10/f4ft3t

[26] Drew Skau and Robert Kosara. 2017. Readability and precision in pictorial bar charts. In *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers.* Eurographics Association, Goslar, 91–95. doi:10/gtgz8f

[27] Alice Thudt, Jagoda Walny, Charles Perin, Fateme Rajabiyazdi, Lindsay MacDonald, Riane Vardeleon, Saul Greenberg, and Sheelagh Carpendale. 2016. Assessing the readability of stacked graphs. In *Proceedings of the Graphics Interface Conference.* Canadian Human-Computer Communications Society, Waterloo, 167–174. doi:10/gtgz8d

[28] Ying Tu and Han-Wei Shen. 2007. Visualizing changes of hierarchical data using treemaps. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1286–1293. doi:10/fq2wwj

[29] Eliane Z. Victorelli, Anne-Flore Cabouat, Emanuele Santos, Florent Cabric, and Petra Isenberg. 2025. We should change how we measure user experience in visual analytics systems. In *Proceedings of the International EuroVis Workshop on Visual Analytics.* Eurographics Association, Goslar, Article eurova.20251095, 6 pages. doi:10/g9wq2t

[30] Zhen Wen, Yihan Liu, Siwei Tan, Jieyi Chen, Minfeng Zhu, Dongming Han, Jianwei Yin, Mingliang Xu, and Wei Chen. 2024. Quantivine: A visualization approach for large-scale quantum circuit representation and analysis. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 573–583. doi:10/gt2v32

[31] Cindy Xiong, Chase Stokes, Yea-Seul Kim, and Steven Franconeri. 2023. Seeing what you believe or believing what you see? Belief biases correlation estimation. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 493–503. doi:10/gtg7mh

[32] Caroline Ziemkiewicz and Robert Kosara. 2009. Preconceptions and individual differences in understanding visual metaphors. *Computer Graphics Forum* 28, 3 (2009), 911–918. doi:10/bdghfc

Received 12 February 2026; accepted 26 February 2026