Grand Unified Grammar of Graphics (GUGOG)

Cynthia A. Huang[1]
LMU Munich, Munich Center for Machine Learning (MCML)

Matthew Kay[2]
Northwestern University

Susan Vanderplas[3]
University of Nebraska-Lincoln

Heike Hofmann[4]
University of Nebraska-Lincoln

Joyce Robbins[5]
Columbia University

Evangeline Reynolds[6]
University of Denver

Abstract

Following Wilkinson's seminal "Grammar of Graphics" (2005), visualization communities in both statistics and computer science have developed various grammar-based approaches to visualization problems, workflows and usage scenarios. While this diversity reflects the richness of visualization challenges, it also reveals fundamental questions: Why do these grammars differ? What core principles unite them? What opportunities exist for synthesis? Which properties make a visualization system a 'graphical grammar'? Despite scattered attempts to survey and understand diversity in grammar based systems, we lack systematic frameworks for understanding how these grammars relate, where they succeed or struggle, and what a more unified theoretical foundation might look like. The first workshop for a grand unified grammar of graphics (GUGOG) aims to facilitate interdisciplinary discussion and exploration of these open questions. We invite reflections on past work and recent developments in visualization grammars, synthesis of parallel and overlapping contributions across statistical graphics and information visualization communities, and visions for the future of grammar-based visualization research.

Index Terms: Grammar of Graphics, visualisation grammars, graphical grammars

1 Motivation

Declarative, modular and compositional approaches to the specification of visualizations, otherwise known as graphical or visualization grammars, often claim Wilkinson's Grammar of Graphics [35] as motivation, inspiration or foundation. Despite the shared intellectual lineage, visualization grammars encompass a broad array of different abstractions, primitives, and compositional strategies, along with a diversity of target users and usage scenarios. There are specialized grammars for specific chart types [21, 31], declaring or managing graphical components [6, 22, 26, 37], particular types of data and/or semantic domains (e.g. uncertainty visualization [13, 18, 24], time-orientated data [30, 36] or databases [38]). There are also a number of general-purpose graphical grammars covering static [14, 20, 34], interactive [3, 28], linked and animated graphics [12, 29].

The design goals and implementation sophistication of graphical grammars also varies greatly. Even tools built upon existing frameworks, such as extensions of ggplot2 [32], vary greatly in scope and complexity, ranging from convenience wrappers for common chart types, to new grammar systems and graphical elements tied to a particular domain and data semantics. The adoption and user base of graphical grammars also varies greatly across various academic and non-academic contexts, suggesting there is not a single best approach to 'grammar of graphics' implementations. The popularity of open-source visualization grammars such as ggplot2, Vega-Lite and Observable Plot amongst different user bases, coupled with differences in their approaches to operationalizing and implementing Wilkinson's Grammar of Graphics give rise to questions about what makes a grammar system successful and how evaluation methods can reflect these various multi-faceted success criteria.

Our current conceptualization and evaluation of grammar-based approaches lack comprehensive integration of these differences and nuances. Instead, grammars are often compared qualitatively [23, 27], or along rather narrowly defined dimensions, such as coverage and generative power over specific visualization design spaces, or cognitive aspects of the notation and syntax [17, 19]. Unfortunately, existing commentary and reflections about the design, implementation and evaluation of graphical grammars are difficult to review because such perspectives are can usually only be included as secondary or tertiary contributions alongside more traditional research contributions (e.g. grammar-based systems and accompanying user studies). Furthermore, these discussions are also fragmented across multiple research and practitioner communities and venues. Relevant discussions can occur anywhere from research literature across visualization and statistical graphics, to the documentation and development repositories for popular grammar systems and beyond.

2 Workshop Goals

This workshop aims to seek a unified view of the diverse application contexts, functional goals and design philosophies embedded in the proliferation of grammar-based visualization tools, by identifying and addressing gaps in our conceptualization and understanding of visualization grammars. While IEEE VIS regularly features papers on individual visualization systems and grammars, it rarely provides structured opportunities for cross-system comparison, meta-theoretical reflection, community consensus building, and practitioner perspectives. This workshop creates a dedicated forum for systematic reflection on the current state of graphical grammars and charting paths toward more principled, unified approaches.

- **Intellectual Integration:** While individual grammar systems are well-documented, the field lacks comparative frameworks for understanding their similarities, differences, and complementarity. One notable exception is [17], which focuses on assessing ease of traversal between visualization designs. However, depending on the use case and design goals of a graphical grammar, other features, such as correctness and consistency are also relevant. Such functional requirements and features of graphical grammars, would also benefit from systematic examination, and can help extend evaluation criteria beyond established notions of usability (e.g. cognitive dimensions of notation).

[1] cynthia.huang\@lmu.de
[2] matthew.kay\@u.northwestern.edu
[3] susan.vanderplas\@unl.edu
[4] hhofmann4\@unl.edu
[5] jtr13\@columbia.edu
[6] evangeline.mae\@gmail.com

- **Bridging Theory and Practice:** Many practitioners struggle to choose between grammar systems or adapt them for novel visualization challenges. By examining current state-of-the-art and identifying similarities, differences and opportunities for improvement or complementary usage, we hope to develop some practical guidance for both tool builders and users. We will also consider issues of learnability and adoption, including challenges and opportunities associated with the steeper learning curve of grammar-based systems. This could include developing natural language interfaces for graphical notations to minimize the initial burden of learning specialized syntax for new users.
- **Community Building:** Bringing together researchers working on different grammar systems—from ggplot2 extensions to Vega-Lite variants to novel formalisms—will foster cross-pollination and potentially collaborative efforts toward more unified approaches. Although, there have been various attempts at cross-disciplinary dialogue between Statistics and InfoVis [4, 5, 16, 33], the interdisciplinary dialogue in this workshop is supported by parallel foundations in mathematical and formal logic (i.e. type theory) across visualization grammar research communities in CS and statistics. In particular, we hope to uncover complementary and possibly duplicated insights across fields, such as algebraic approaches to visualization validity in InfoVis [15] and statistical graphics [2], or formal guarantees in programming languages and preservation of statistical semantics in Wilkinson's grammar of graphics and closely related work.

The intended outcome of this workshop is the development of a cohesive interdisciplinary research agenda for grammar-based approaches in visualization research. We hope to coalesce a new subfield of visualization research that incorporates complementary perspectives and insights from the various researcher and practitioner communities developing, using and critically analysing graphical grammars. We expect to see follow-up initiatives such as an associated event at JSM 2027, funding applications for extended projects or collaborative meetings based on topics discussed at the workshop, and will plan to publish summary notes of workshop discussions on the workshop website.

3 Related Scientific Initiatives

This workshop relates to past and forthcoming initiatives in both statistics and computer science. Multiple members of the organizing committee have been involved in one or more of these initiatives. Within VIS itself, grammar focused sessions are a recurring part of the main program, reflecting the proliferation of grammar-inspired research. Unfortunately, the presentation constraints of the full and short paper tracks leave little room for discussion of generalizable and transferable design and implementation lessons. The 'Special Interest Group on Visualization Grammars' at CHI21 facilitated discussion around what makes a grammar "good", including design choices, usability considerations, evaluation criteria and methods. This event predominantly focused on engaging visualization researchers from CS venues (i.e. CHI, VIS, UIST, CSCW, and KDD). By contrast, we will incorporate research developers and engage with researchers from beyond the CS community – namely from statistics and statistical graphics. The Joint Statistical Meeting (JSM) program has also increasingly featured grammar-related research, including the paper session 'Who are the ggplot2 extenders, and how can you become one?' in 2025, and the forthcoming invited panel 'The Past, Present, and Future of ggplot2 Extension Packages'. The past two John M. Chamber Statistical Software Awards presented at JSM have also been awarded to grammar-based visualization tools [1, 18].

4 Workshop Activities

We propose a half-day hybrid workshop with expert panels curated from submissions and invited speakers, and structured breakout sessions. The panels will serve to establish foundational understanding of the shared landscape, and facilitate smoother cross-disciplinary discussions in the breakout sessions.

4.1 Scope of Topics

Workshop topics include, but are not limited to:

- Novel grammar proposals
- Comparative analyses
- Implementation experiences
- User perspectives
- Theoretical frameworks
- Vision statements

5 Publication Plan

The workshop will accept two types of non-archival submissions:

- Short position papers of less than 4 pages. We will not accept system descriptions unless they are illustrative prototypes closely tied to a specific conceptual or philosophical argument.
- Discussion abstracts of up to 500 words. We invite provocations, opinions and observations on all aspects of grammar-based visualization research.

We will host accepted submissions on the workshop's website, and also recommend authors share their publications using open-access services. The accepted submissions will not be considered archival, and we will encourage the reuse of content in follow-up publications. Submissions will not be anonymous; authors must include their full names, emails, and affiliations. Lastly, at least one author per accepted submission must register for the conference.

5.1 Tentative Schedule (3.5 hours)

Given the interdisciplinary scope of the workshop, we plan to organize a fully hybrid workshop that is open to remote attendance. This workshop is planned as a half-day event, with the following tentative schedule:

- **Session 1:** Introduction to the workshop goals and activities (15 min), followed by lightning talks and brief introductions by participants (30 min), a panel discussion (30 min).
- **Break:** Coffee break, and in parallel, collection of discussion topic suggestions via sticky notes or similar participatory tool.
- **Session 2:** Announcement of discussion topics (5 min), breakout discussions (40 min), reports by the groups with a short summary discussion (10 min). Topics will be based on submissions and Session 1 discussion.
- **Summary and Call to Action:** To conclude the workshop, we will summarize the insights from the panel and breakout discussions, emphasizing potential follow-up initiatives and future research directions (10 min).
- **Networking:** After the workshop session, a voluntary workshop lunch will be planned to encourage community building and networking among working groups.

Back-up policy: If we receive insufficient submissions, we plan to invite a keynote speaker or expert panellists to provide inspiration for the discussion groups planned for the second part of the workshop. Potential speakers and panellists include core developers of open-source grammars, past presenters at the ggplot2 extenders

club, and authors of highly cited or novel work on grammar-based approaches.

5.2 Call for Participation and Timeline

- Call opens: May 1, 2026
- Submission deadline: Aug 1, 2026
- Notifications: Sep 1, 2026
- Camera-ready: Oct 7, 2026

We intend to advertise across CS and statistics communities, including advertising on IEEE and ACM mailing lists, at UseR!, DSSV, JSM and social media (X, LinkedIn, BlueSky, etc.). The discussion abstract deadline is scheduled for a month after the JSM panel session to allow time for attendees of that session to submit.

6 Organizers

Our organizing committee spans multiple research and practitioner communities, allowing us to reach a broader range of individuals working on visualization grammars and grammar-based approaches.

Cynthia Huang is a postdoctoral researcher at the Social Data Science and AI chair at the Ludwig Maximilian University of Munich, Germany where she works on data preparation and visualization tools that incorporate statistical, computational and usability considerations. Previously, she was a PhD candidate, then research fellow, in the department of Econometrics and Business Statistics at Monash University, Australia. During this time, she also undertook an extended visit to the InfoVis group at the University of British Columbia, Vancouver, Canada. Her related work includes a novel graph-based representation and workflow grammar for validating and implementing ex-post data harmonization workflows [9], a grammar-based system for extracting canonical tables from spreadsheets [39], and time-aware grammar of graphics geometries, scales and coordinate systems [10]. She was co-chair of WOMBAT2025, where she chaired a panel discussion on the design of statistical software, and will be a panellist at the forthcoming ggplot2 session at JSM 2026. Website: cynthiahqy.com

Matthew Kay is an Associate Professor jointly appointed in Computer Science and Communications Studies at Northwestern University. He works in human-computer interaction and information visualization, with a particular focus on uncertainty visualization, visualization literacy, and the design of human-centered tools for data analysis. His research has been funded by multiple NSF awards, including an award to develop grammars for uncertainty visualizations, which lead to an honorable-mention paper at CHI [24], a paper at VIS [13], and the development of the ggdist R package, a widely-used ggplot2 extension for uncertainty visualization. He has also studied how data analysts use visualization grammars in practice [23] and organized a SIG on visualization grammars at CHI 2021 [25]. He has received multiple best paper awards across human-computer interaction and information visualization venues, including ACM CHI and IEEE VIS, and he received the 2023 IEEE VGTC Visualization Significant New Researcher Award. He was co-chair of the CHI Visualization Papers Subcommittee in 2022 and 2023 and VIS Area Paper Chair for Area 1 (Theoretical and Empirical) at VIS 2024 and 2025. He co-directs the Midwest Uncertainty Collective (https://mucollective.northwestern.edu/).

Susan Vanderplas is an associate professor in Statistics at the University of Nebraska-Lincoln. She received her PhD in statistics from Iowa State University in 2015. She researches data visualization, reproducible computing, and machine learning algorithms for forensic pattern matching. Susan previously worked on the R package animint [7] to extend the ggplot2 grammar of graphics to include interactivity, and has contributed to the 'ggpcp' R package to use the grammar of graphics for parallel coordinate plots [31]. Website: srvanderplas.github.io

Heike Hofmann Heike Hofmann is the Professor of Methodology in Observational Data and Exploratory Data Analysis in the Department of Statistics at the University of Nebraska-Lincoln. Before joining UNL in Aug 2024, Prof. Hofmann was a faculty member in the Department of Statistics at Iowa State since 2002. She is interested in developing tools and methodology for visualizing large, multivariate, and complex data, and in applications across a range of areas, such as forensic sciences and bioinformatics. She co-founded the Graphics Group, a network of almost 100 members interested in statistical graphics, established in 2002 with weekly meetings around topics on data visualization and statistical computing during the academic year. Her prior work includes numerous ggplot2 and grammar of graphics extensions [8, 11, 31].

Joyce Robbins is a Senior Lecturer in the Department of Statistics at Columbia University. Her main interests are data visualization and statistics education, with current work focused on mapping the ggplot2 extension package ecosystem. Since 2017, she has been the sole instructor of Exploratory Data Analysis and Visualization, a core requirement in Columbia's Data Science Institute Masters Program. She is an active member of the Statistical Graphics Section of the American Statistical Association, having served as publications officer and later program chair, and has organized sessions on interactive graphics, high dimensional visualization, and ggplot2 extension packages at recent JSM conferences. Through RForwards, she has co-taught free virtual R package writing workshops as part of the organization's mission to increase the participation of underrepresented groups in the R community.

Evangeline Reynolds is a data scientist and educator whose work focuses on tools that make data analytics more fluid and intuitive, with particular attention to reducing cognitive load in both the interpretation of data visualizations and the writing of code to build them. She has taught quantitative methodology at TU Dresden, the University of Denver, and West Point, and also worked as a Data Scientist at West Point. Her software contributions in the ggplot2 extension space include ggcirclepack, ggrgions, and ggdims, among others. More recently, she works in developer relations for the ggplot2 extender community, supported by Posit PBC. She co-founded and organizes the ggplot2 extenders club and has developed associated educational resources including 'easy geom recipes' and 'express' methodologies for ad hoc extension and experimental work. She co-founded R Ladies Denver in 2019 and is active in the American Statistical Association at both the national level and through the Colorado-Wyoming chapter. Website: github.com/EvaMaeRey

References

[1] A. Bartonicek, "Plotscaper: Explore Your Data with Interactive Figures." 2025.

[2] A. Bartonicek, S. Urbanek, and P. Murrell, "No More, No Less than Sum of Its Parts: Groups, Monoids, and the Algebra of Graphics, Statistics, and Interaction," *Journal of Computational and Graphical Statistics*, vol. 34, no. 3, pp. 1063–1074, July 2025, doi: 10.1080/10618600.2024.2429708.

[3] M. Bostock, V. Ogievetsky, and J. Heer, "D^3 Data-Driven Documents", *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2301–2309, Dec. 2011, doi: 10.1109/TVCG.2011.185.

[4] A. Gelman and A. Unwin, "Infovis and Statistical Graphics: Different Goals, Different Looks," *Journal of Computational and Graphical Statistics*, vol. 22, no. 1, pp. 2–28, Jan. 2013, doi: 10.1080/10618600.2012.761137.

[5] A. Gelman and A. Unwin, "Tradeoffs in Information Graphics," *Journal of Computational and Graphical Statistics*, vol. 22, no. 1, pp. 45–49, Jan. 2013, doi: 10.1080/10618600.2012.761141.

[6] G. Grolemund and H. Wickham, "Visualizing Complex Data With Embedded Plots," *Journal of Computational and Graphical Statistics*, vol. 24, no. 1, pp. 26–43, Jan. 2015, doi: 10.1080/10618600.2014.896808.

[7] T. Hocking *et al.*, "Animint2: Animated Interactive Grammar of Graphics." 2024.

[8] H. Hofmann, H. Wickham, and K. Kafadar, "Letter-Value Plots: Boxplots for Large Data," *Journal of Computational and Graphical Statistics*, vol. 26, no. 3, pp. 469–477, July 2017, doi: 10.1080/10618600.2017.1305277.

[9] C. A. Huang, "A Unified Statistical And Computational Framework For Ex-Post Harmonisation Of Aggregate Statistics," no. arXiv:2406.14163. arXiv, June 2024, doi: 10.48550/arXiv.2406.14163.

[10] C. A. Huang, M. O'Hara-Wild, R. J. Hyndman, and M. Kay, "Ggtime: A Grammar of Temporal Graphics," no. arXiv:2510.25656. arXiv, Oct. 2025, doi: 10.48550/arXiv.2510.25656.

[11] H. Jeppson and H. Hofmann, "Generalized Mosaic Plots in the Ggplot2 Framework," *The R Journal*, vol. 14, no. 4, pp. 50–78, Feb. 2023, doi: 10.32614/RJ-2023-013.

[12] Jonathan Zong, Josh Pollock, Dylan Wootton, ArvindJonathan SatyanarayanZong, and Josh Pollock, "Animated Vega-Lite: Unifying Animation with a Grammar of Interactive Graphics.," *IEEE Transactions on Visualization and Computer Graphics*, Oct. 2022, doi: 10.1109/tvcg.2022.3209369.

[13] M. Kay, "Ggdist: Visualizations of Distributions and Uncertainty in the Grammar of Graphics," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2023, doi: 10.1109/TVCG.2023.3327195.

[14] H. Kibirige *et al.*, "Has2k1/Plotnine: V0.15.3." Jan. 2026, doi: 10.5281/ZENODO.1325308.

[15] G. Kindlmann and C. Scheidegger, "An Algebraic Process for Visualization Design," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2181–2190, Dec. 2014, doi: 10.1109/TVCG.2014.2346325.

[16] R. Kosara, "InfoVis Is So Much More: A Comment on Gelman and Unwin and an Invitation to Consider the Opportunities," *Journal of Computational and Graphical Statistics*, vol. 22, no. 1, pp. 29–32, Jan. 2013, doi: 10.1080/10618600.2012.755465.

[17] N. Kruchten, A. M. McNutt, and M. J. McGuffin, "Metrics-Based Evaluation and Comparison of Visualization Notations," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2023, doi: 10.1109/TVCG.2023.3326907.

[18] H. Mason, D. Cook, S. Goodwin, and S. VanderPlas, "Ggdibbler: Add Uncertainty to Data Visualisations." 2026.

[19] A. M. McNutt, "No Grammar to Rule Them All: A Survey of JSON-style DSLs for Visualization," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2022, doi: 10.1109/TVCG.2022.3209460.

[20] Observable, Inc., "Observable Plot." 2025.

[21] D. Park, S. M. Drucker, R. Fernandez, and N. Elmqvist, "\textsc{Atom}: A Grammar for Unit Visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 12, pp. 3032–3043, Dec. 2018, doi: 10.1109/TVCG.2017.2785807.

[22] J. Pollock and A. Satyanarayan, "GoFish: A Grammar of More Graphics!," *IEEE Transactions on Visualization and Computer Graphics*, vol. 32, no. 1, pp. 549–559, Jan. 2026, doi: 10.1109/TVCG.2025.3634250.

[23] X. Pu and M. Kay, "How Data Analysts Use a Visualization Grammar in Practice," 2023, doi: 10.1145/3544548.3580837.

[24] X. Pu and M. Kay, "A Probabilistic Grammar of Graphics," *International Conference on Human Factors in Computing Systems*, pp. 1–13, Apr. 2020, doi: 10.1145/3313831.3376466.

[25] X. Pu, M. Kay, S. M. Drucker, J. Heer, D. Moritz, and A. Satyanarayan, "Special Interest Group on Visualization Grammars," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, May 2021, pp. 1–3, doi: 10.1145/3411763.3450406.

[26] D. Rahman, R.-u.- Zaman, A. McNutt, and P. Rosen, "AnnoGram: An Annotative Grammar of Graphics Extension."

[27] A. Satyanarayan *et al.*, "Critical Reflections on Visualization Authoring Systems," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 461–471, 2020, doi: 10.1109/TVCG.2019.2934281.

[28] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, "Vega-Lite: A Grammar of Interactive Graphics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 341–350, Jan. 2017, doi: 10.1109/tvcg.2016.2599030.

[29] C. Sievert, S. VanderPlas, J. Cai, K. Ferris, F. U. F. Khan, and T. D. Hocking, "Extending Ggplot2 for Linked and Animated Web Graphics," *Journal of Computational and Graphical Statistics*, vol. 28, no. 2, pp. 299–308, Apr. 2019, doi: 10.1080/10618600.2018.1513367.

[30] V. Stoiber, N. Gehlenborg, W. Aigner, and M. Streit, "Time-i-Gram: A Grammar for Interactive Visualization of Time-based Data." Open Science Framework, May 2024, doi: 10.31219/osf.io/m9ubg.

[31] S. VanderPlas, Y. Ge, A. Unwin, and H. Hofmann, "Penguins Go Parallel: A Grammar of Graphics Framework for Generalized Parallel Coordinate Plots," *Journal of Computational and Graphical Statistics*, vol. 32, no. 4, pp. 1572–1587, Oct. 2023, doi: 10.1080/10618600.2023.2195462.

[32] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

[33] H. Wickham, "Graphical Criticism: Some Historical Notes," *Journal of Computational and Graphical Statistics*, vol. 22, no. 1, pp. 38–44, Jan. 2013, doi: 10.1080/10618600.2012.761140.

[34] H. Wickham, "A Layered Grammar of Graphics," *Journal of Computational and Graphical Statistics*, vol. 19, no. 1, pp. 3–28, Jan. 2010, doi: 10.1198/jcgs.2009.07098.

[35] L. Wilkinson, *The Grammar of Graphics*, 2nd ed. New York: Springer, 2005.

[36] G. Wills, *Visualizing Time: Designing Graphical Representations for Statistical Data*. New York, NY: Springer New York, 2012.

[37] K. Wongsuphasawat, "Encodable: Configurable Grammar for Visualization Components," in *2020 IEEE Visualization Conference (VIS)*, Oct. 2020, pp. 131–135, doi: 10.1109/VIS47514.2020.00033.

[38] E. Wu, X. Y. Tuang, A. Li, and V. Bainwala, "A Formalism and Library for Database Visualization," no. arXiv:2504.08979. arXiv, Apr. 2025, doi: 10.48550/arXiv.2504.08979.

[39] K. Xiong, C. A. Huang, M. Wybrow, and Y. Wu, "TableCanoniser: Interactive Grammar-Powered Transformation of Messy, Non-Relational Tables to Canonical Tables," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Apr. 2025, pp. 1–20, doi: 10.1145/3706598.3714321.