What Makes a Visualization Visually Complex?

Kylie Lin
klin368@gatech.edu
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA

Sean Sheng-Tse Ru
College of Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
sru3@gatech.edu

David N. Rapp
Department of Psychology
Northwestern University
Evanston, Illinois
rapp@northwestern.edu

Hui Guan
College of Information and Computer Science
University of Massachussetts, Amherst
Amherst, Massachussetts, USA
huiguan@cs.umass.edu

Cindy Xiong Bearfield
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
cxiong@gatech.edu

Abstract

The world is filled with overly complex visualizations that obscure key insights and overwhelm audiences. While prior work in visualization has identified general best practices, it has not fully explored how specific design features contribute to perceptions of visual complexity. To address this, we augmented the MASSVIS dataset—one of the largest static visualization datasets—with new metadata on design features (text, color, data, and layout) and perceived complexity ratings. Analyzing feature distributions and their effects, we trained machine learning models to predict perceived complexity and identified feature importance. Our findings show that the number of charts in a visualization most strongly influences perceived complexity, with more visual elements generally leading to higher complexity. Additionally, we present case studies demonstrating how the augmented dataset and models can support research on human cognition and visualization design, enabling designers to refine visualizations for optimal complexity. All supplemental materials are available at https://osf.io/k4uta/?view_only=ab57e05d70324f0b9e26255c77646c9a

CCS Concepts

• **Human-centered computing** → **Empirical studies in visualization**.

Keywords

Visual complexity, Perception, Static visualizations, Predictive modeling.

ACM Reference Format:
Kylie Lin, Sean Sheng-Tse Ru, David N. Rapp, Hui Guan, and Cindy Xiong Bearfield. 2025. What Makes a Visualization Visually Complex?. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3706599.3719983

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/2025/04
https://doi.org/10.1145/3706599.3719983

1 Introduction

Well-designed visualizations enhance critical thinking by aligning with our perceptual and cognitive strengths, aiding information comprehension and decision-making [10, 14]. Conversely, poor design can obscure patterns and lead to miscommunication, with even well-meaning designers falling prey to the curse of expertise, producing overly complex visuals that overwhelm audiences [45].

To mitigate this, the visualization research community has developed best practices to improve the readability, interpretability, and memorability of a visualization [3, 5, 6]. However, understanding how combinations of design choices influence the visual complexity of static visualizations remains a challenge. Visual complexity, often defined as the "amount of detail or intricacy" in an image [36], can render visualizations perplexing when excessive. Attempts to quantify complexity, such as approximate entropy for line charts [29] or edge counts in network graphs [40], have focused on specific chart types rather than static visualizations broadly [1, 20].

So, what makes a visualization visually complex? Building on existing work on visualization affordances [4, 12, 26], we examine the relationship between perceived complexity and design features in **Text**, **Color**, **Data**, and **Design** categories for 5,800 static visualizations. We extended the MASSVIS dataset [4, 5] with crowdsourced visual complexity ratings and labels of key design features. This enriched dataset enabled (1) statistical analyses and (2) the training of regression models to quantify how design elements influence perceived complexity. We find that visualizations with more elements are generally perceived as more complex. Feature importance techniques [11] revealed the weighted impact of these elements. We propose that future research can leverage these insights to help designers optimize visual complexity for clarity and interoperability.

Contributions: We contribute insights on the relationship between visualization design features and perceived visual complexity as well as a set of design feature labels and perceived visual complexity ratings for the MASSVIS dataset. We use the augmented dataset to train machine learning models that predict human perceptions of visual complexity. We also offer two case studies, demonstrating how (1) our models can help designers create visualizations with consideration for visual complexity and (2) the augmented dataset can be used in cognitive studies to shed light on how visualizations can be designed to elicit trust and encourage critical thinking.

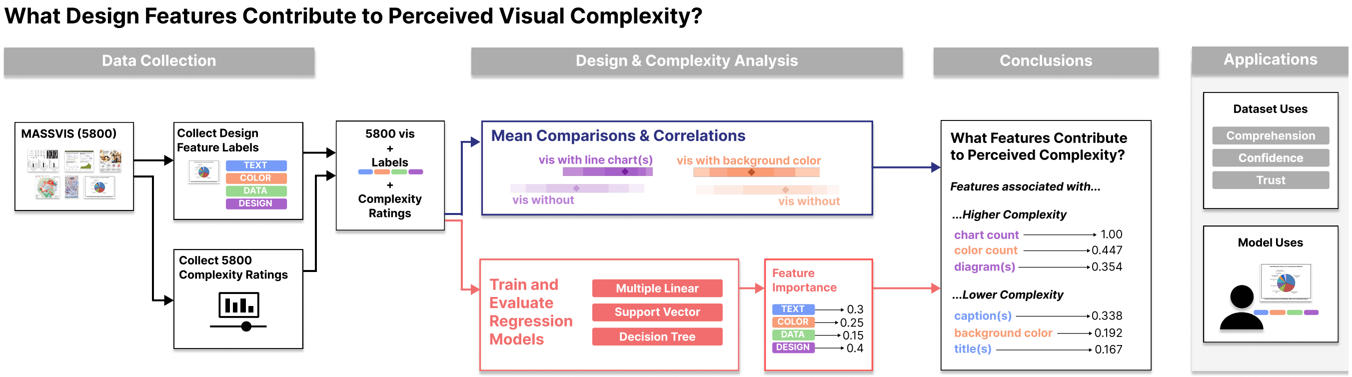

Figure 1: The flow of our investigation spans four phases. We conduct *Data Collection* using the MASSVIS dataset, collecting ratings of complexity and novel design feature labels for 5,800 visualizations. With the combined data, we then conduct *Design & Complexity Analysis*, including 1) means comparison analysis of perceived complexity across design features and 2) training and evaluation of simple machine learning models to derive design feature importance. We end up with *Conclusions* about what design features contribute to visual complexity. Finally, we detail *Applications* for the augmented dataset and our trained models.

2 Related Work

Visual complexity has been defined variably across fields of study. Perceptual psychology often describes it as "the amount of detail or intricacy" in an image [36], while vision science uses measures like feature congestion, subband entropy, and edge density to approximate visual complexity [28]. Algorithmic measures, such as Kolmogorov complexity, evaluate complexity by the shortest program needed to reproduce an image [8], while human-computer interaction considers the number of visual elements in webpages [27] or the complexity of icons based on structural features [7, 13].

These definitions, however, fall short for data visualizations, which function more like paragraphs than static images, involving multilayered interpretation [15, 32]. Visual saliency models trained on natural images often fail to predict attention in visualizations [22], highlighting the need for unique measures of visual complexity specific to this medium. Some propose a human-centric approach, emphasizing perceived complexity to capture subjective interpretations [21]. However, perceived complexity does not always align with inherent complexity, and the roles of elements like text, color, and chart type remain underexplored. Research on the complexity of visualizations often references principles like Tufte's data-ink ratio and the avoidance of chartjunk for simplicity [38]. While some support minimalism for cognitive efficiency [20], others argue that decorative elements can enhance reflection and engagement [2, 16]. Recent visualization literature has called for the community to consider both visual and non-visual complexity as 'design material' that is useful for informing effective data visualization design [43]. Still, no consensus exists on how design characteristics influence perceived complexity [23, 33], and definitions remain inconsistent across studies.

Machine learning offers promising tools to bridge these gaps. Models have been applied to improve visualization layouts [41], refine saliency predictions [22, 34], and generate descriptive text for patterns [35]. By predicting feature importance, machine learning can help identify how design elements contribute to perceived complexity, advancing our understanding of effective visualization design.

3 MASSVIS Dataset Labeling

We generated a labeling scheme based on the original MASSVIS taxonomy and related visualization literature, capturing design features related to text (**Text**), color (**Color**), underlying data (**Data**), and chart layout/design (**Design**). We crowdsourced labels for the 5,800 MASSVIS visualizations using our labeling scheme, recruiting 160 participants ($M_{age} = 38.19$, $SD_{age} = 13.62$) via Prolific [24]. We designed one survey per labeling category, where each survey contained combinations of multiple-selection, multiple-choice, and text-response survey questions. Labels collected were either categorical (C), binary (0/1), or numeric (#). Since crowdsourcing can provide low-quality responses, we manually checked all crowdsourced labels to ensure data quality.

Text *Text Types (C)*: A visualization can be composed solely of text, have no text, or include multiple types of text in the form of axes label(s), axes text, title(s), short annotation(s), captions(s), and legend-related text such as title or content text.

Color *Black and White (0/1)*: This label was coded as 'yes' if the visualization was in black and white and 'no' if it was in color. *Number of Colors (#)*: This label captures crowdsourced participants' best approximations of the number of distinct colors present in the visualization. *Background Color (0/1)*: We measured background color, which was coded as 'yes' if the the visualization had a non-white background color and 'no' if it did not.

Data *Number of Quantitative Variables (#)*: This label captures approximations of the number of quantitative variables in a visualization. *Number of Categorical Variables (#)*: This label captures approximations of the number of categorical variables that underlie a given visualization.

Design *Number of Charts (#)*: This label represents the amount of charts in a visualization. *Chart Types (C)*: This label captures the presence of different chart types, which are the 12 visualization types in the MASSVIS static visualization taxonomy.

The types of charts are: Area, Bar, Circle, Diagram, Distribution, Grid & Matrix, Line, Map, Point, Table, Text, and Trees & Networks. *Number of Chart Types (#)*: This label is derived directly from the *Chart Types* label and notes the number of unique chart types included in a given visualization.

4 Experiment: Perceptions of Visual Complexity

We recruited 400 participants (M_{Age} = 37.01, SD_{age} = 12.84) from Prolific to provide visual complexity ratings for the 5800 MASSVIS visualizations. Participants were randomly assigned to one of 40 visualization group conditions, each containing a unique group of 145 random MASSVIS visualizations so that all 5800 visualizations were viewed 10 times. They provided complexity ratings using a slider task, resulting in 10 unique ratings for each visualization.

For binary labels, we ran comparisons of the visual complexity ratings for visualizations with a given label versus those without, such as comparing visualizations that are black and white to visualizations that are *not* black and white. For numeric labels, such as the number of colors, we conducted a Pearson's correlation test to determine the feature's relationship with complexity ratings. We do not claim that these analyses reveal causal relationships between the presence of a given design feature and perceived complexity, but rather use the comparisons to provide a summary overview of how design and complexity ratings were distributed in the MASSVIS space (2). We describe the results below:

Text Complexity ratings for visualizations with no text were not significantly different from those with text ($p = 0.075$). Visualizations that contained captions ($p < 0.001$), titles ($p < 0.001$), or only text ($p < 0.001$) were rated as significantly less complex as opposed to visualizations without those features. In contrast, visualizations that contained annotations ($p < 0.001$), axes labels ($p < 0.001$), axes text ($p < 0.001$), legend titles ($p < 0.001$), or legend text ($p < 0.001$) had significantly higher complexity scores than visualizations without those features.

Color Black-and-white visualizations were rated as significantly more complex than visualizations in color ($p < 0.001$). Visualizations with a non-white background ($p < 0.001$) were rated significantly less complex than visualizations with a white background. We found a weak, positive relationship between the number of colors in a visualization and perceived complexity ($p < 0.001$).

Data We found a weak, positive relationship between perceived complexity and both the number of quantitative variables ($p < 0.001$) and the number of categorical variables ($p < 0.001$).

Design Visualizations that contained Area ($p = 0.012$), Diagram ($p < 0.001$), Distribution ($p < 0.001$), Grid & Matrix ($p < 0.001$), Line ($p < 0.001$), Point ($p < 0.001$), Table ($p < 0.001$), or Tree & Network charts ($p < 0.001$) were perceived as significantly more complex than visualizations without. Visualizations with Maps ($p < 0.001$) had significantly lower complexity ratings as opposed to visualizations without maps. Text ($p = 0.386$), Bar ($p = 0.524$), or Circle ($p = 0.283$) charts did not significantly impact complexity ratings. However, we did find a moderate, positive relationship between the number of charts in a visualization and perceived complexity ($p < 0.001$) as well as between the number of chart types and perceived complexity ($p < 0.001$).

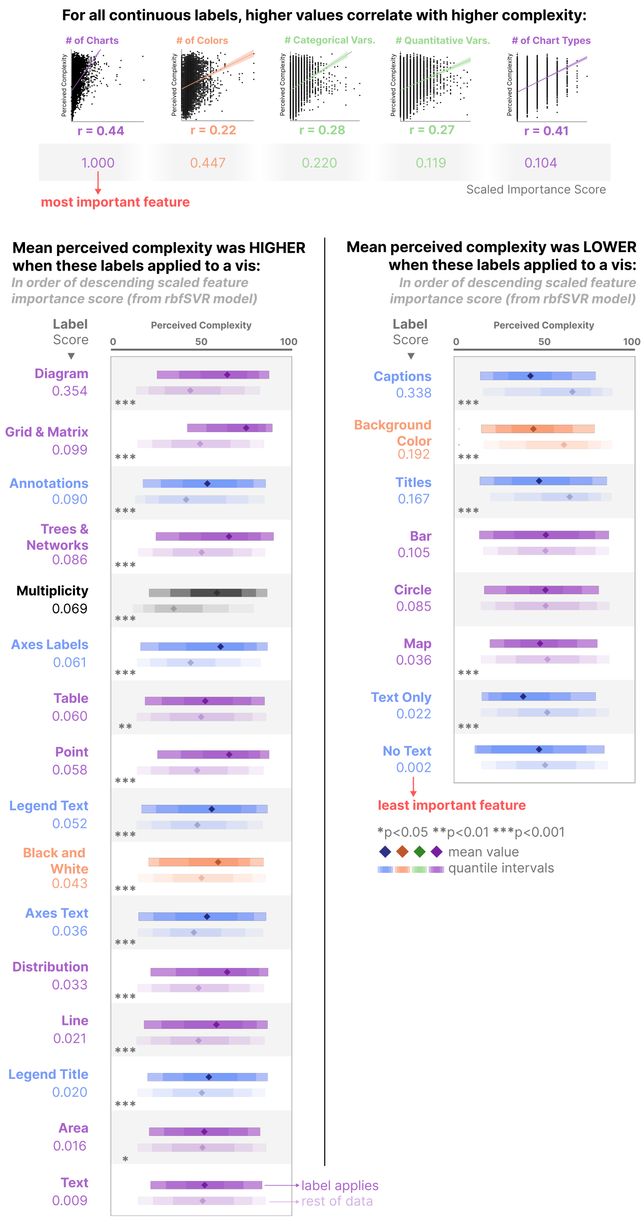

Figure 2: Distributions of and correlations with mean perceived visual complexity ratings by Text, Color, Data, Design, and Multiplicity.

5 Model Training and Evaluation

We trained and evaluated a series of ML models to predict the perceived visual complexity of a visualization using our design feature labeling scheme for the MASSVIS dataset. We derived the importance of the features from the best-performing models to attribute to the design features.

5.1 Model Training

We first describe the ML models that we used in this work and then provide information about how we prepared the design feature labels to make them suitable for training these models and the training details. We use interpretable machine learning models to easily determine how each visualization feature affects predicted perceived complexity. The three models used are Linear Regression, Support Vector Regression (SVR), and Decision Tree Regression [42].

Linear Regression predicts outcomes as a linear combination of input features, with weights referred to as coefficients. We also tested two variants: Ridge Regression, which penalizes large coefficients to reduce overfitting, and Lasso, which promotes sparsity in the coefficients. **Support Vector Regression (SVR)** applies Support Vector Machines (SVMs) to regression tasks, leveraging support vectors to predict new data points. SVR can utilize nonlinear kernels for capturing complex relationships. We tested three SVR variants: linearSVR (linear kernel), polySVR (polynomial kernel), and rbfSVR (Radial Basis Function kernel), with rbfSVR identified as the best-performing variant. **Decision Trees** use a tree structure to predict target values based on simple decision rules. We employed a Decision Tree Regressor as perceived visual complexity is a continuous variable (0–100). Decision trees are straightforward to interpret and visualize.

Dataset Preparation: Categorical data were one-hot encoded into binary columns for machine learning. We included the Multiplicity feature (single- vs. multi-panel) from the MASSVIS dataset, as it likely influences perceived complexity. Each visualization was represented as a 30-dimensional vector comprising design, text, color, data features, multiplicity, and mean perceived complexity ratings.

Training Details: We split the dataset into training (80%) and testing (20%) subsets, using the former for training and the latter for evaluation. Continuous variables were log-transformed to address skewness, with a constant of 1 added to values before transformation. A grid search was conducted to optimize hyperparameters for each model, with results shown in table 2. The rbfSVR model was identified as the highest-performing model and is the focus of detailed reporting. Implementation was conducted using the Python library scikit-learn [25].

Models	Features	Importance
Linear	# of Chart Types	21.35
	Circle	-10.22
	Black and White	9.88
	Bar	-9.32
	# of Charts	8.33
Lasso	# of Chart Types	13.54
	Black and White	9.53
	# of Charts	8.34
	Circle	-7.95
	Bar	-7.04
Ridge	# of Chart Types	19.67
	Black and White	9.84
	Circle	-9.74
	Bar	-8.83
	# of Charts	8.33
rbfSVR	# of Charts	72.35
	# of Colors	32.32
	Diagram(s)	25.62
	Caption(s)	24.49
	# of Categorical Vars.	15.89
Decision Tree	# of Charts	0.54
	Caption(s)	0.20
	# of Colors	0.07
	Title(s)	0.07
	Background Color	0.02

Table 1: Five most significant feature importance scores (original and scaled) for each of our trained models. For all feature importance scores, see supplemental materials.

Models	Hyperparameters	R2	MAE / MSE
Multiple Linear	–	0.591	10.73 / 179.70
Lasso	$\alpha = 0.0046$	0.589	10.75 / 180.30
Ridge	$\alpha = 1$	0.590	10.73 / 179.79
rbfSVR	$C = 10$, $\gamma =$ scale, $\varepsilon = 1$	0.644	9.86 / 156.33
Decision Tree	depth=6, split=85, leaf=11	0.525	11.19 / 208.38

Table 2: Overview of trained models and performance metrics (R-squared, Mean Absolute Error, and Mean Squared Error). MAE and MSE values are shown together for brevity.

5.2 Model Evaluation

Evaluation of Prediction Performance: We evaluate the model performance on the test split using three commonly used metrics: **Mean Squared Error (MSE)**, calculated as the mean or average of the squared differences between predicted and expected target values in a dataset, **Mean Absolute Error (MAE)**, calculated as the average of the absolute error values, and **R-squared score (R2)**, calculated as 1 - the sum of the residuals squared divided by the total sum of squares.

Table 2 presents the prediction performance of the ML models. The best performing model was the rbfSVR model ($R2 = 0.644$, $MAE = 9.855$, $MSE = 156.331$). These metrics reveal that our models only do a middling job at predicting complexity, as even the highest R2 score of 0.644 is a rather low value. Future work could leverage less interpretable machine learning models (such as deep neural networks) to improve the prediction performance.

Analysis of Feature Importance: Using all of our fitted models, we derived feature importance scores for each design feature to reveal the extent to which the models rely on a given feature. We used these scores to *quantify* how significant a design feature is to actual notions of perceived visual complexity. For Linear Regression models, we examine the coefficients assigned to each design feature, which can be either negative or positive in value, indicating whether the feature contributes to increased or decreased predicted ratings of complexity. The absolute value or magnitude of the coefficient can be used to infer the importance of the features. For the Decision Tree Regression model, we used scikit-learn's built-in method to collect feature importance, in which importance is calculated as "the (normalized) total reduction of the criterion brought by that feature" [25]. A higher score indicates higher importance for a given feature. For SVR models, we used permutation feature importance, a technique well-suited for nonlinear models. The technique observes and reports the extent to which perturbations to each input feature affect the output of the model.

Across all models, we found that the *Number of Charts* was consistently represented as a feature with relatively high importance. All Linear Regression models had the same top five labels (*Number of Chart Types*, *Black and White*, *Circle*, and *Bar*) as important features in addition to the number of charts, though the ordering of features varied. These labels were strictly from the **Design** and **Color** categories. For SVR and Decision Trees, however, captions were of high importance, introducing **Text**. The **Data** category was also found to be of high importance (5th) for the SVR model. See Figure 1 for the feature importance scores.

6 What Makes a Visualization Complex?

Design features were categorized based on whether they increase or decrease complexity ratings. In fig. 2, we visualize the mean complexity ratings for visualizations with and without a given feature, highlighting statistically significant differences. The most influential feature, with a scaled importance score of 1.000, is the *Number of Charts*, which has a moderate positive correlation with perceived complexity. Other features, such as *Number of Colors* and *Chart Types*, show weaker correlations but still contribute to complexity.

In general, more visual elements result in higher perceived complexity, with 16 features associated with increased complexity and 8 with reduced complexity. Prior work suggests that reducing visual elements, or "decluttering," enhances memorability and professionalism [1], and our findings align with this. However, reducing complexity may also reduce critical thinking [16, 17]. Popular designs like small multiples and composite dashboards often involve many charts, increasing perceived complexity. As these designs are commonly used in decision-making tools, future research should explore how higher complexity impacts trust and comprehension in data. In the next section, we present a case study to explore these questions further.

7 Applications for Model & Relabeled Dataset

We showcase potential applications of our models and our augmented dataset for predicting complexity and understanding human cognition.

7.1 Model Applications: Predicting Complexity

As a preliminary exploration, we implemented a simple function in Python (using the rbfSVR model from section 5) that predicts the perceived visual complexity of a visualization on a 0-100 scale based on feature labels that can be passed as arguments to the function (see the supplementary materials for details). We present a scenario showcasing the everyday benefits of predictive complexity modeling and its potential for future research.

Imagine a student, Akira, who has designed a visualization for their Film & Media final project. The visualization is a bar chart with many embellishments, as shown on the left of fig. 3 [37]. Akira wants to know whether the design might be too complex, so they extract features from the visualization and input them to our function as the following: multiplicity (single panel), text types (title, annotation, axes labels, axes text, caption), black and white (0), background color (0), number of colors (4), chart types (Bar), the number of quantitative (0) and categorical (1) variables, and the number of charts (1). The function returns the predicted perceived visual complexity score of **32.74**.

This isn't a bad score to Akira, but they would prefer it to be a little lower since the visualization is part of a short, fast-paced presentation. Akira redesigns the visualization, removing annotations and the images overlaid on the bars, which reduces the number of colors. The right side of fig. 3 shows the final redesign [39], which has a new set of design feature labels associated with it: multiplicity (single panel), text types (title, axes labels, axes text, legend text, caption), black and white (0), background color (0), number of colors (1), number of quantitative variables (1), number of categorical variables (1), number of charts (1), chart types (Bar). Akira runs the new feature labels through our function, which returns the predicted perceived visual complexity score of **25.74**. Satisfied with the simpler design, Akira includes the redesign in their presentation.

Building on this example, we argue that there is a vast range of future work in the space of modeling complexity. Future research could enhance the function's usability and generalizability by contextualizing complexity scores, clarifying *for whom* and *for what purpose* they are generated and calibrated.

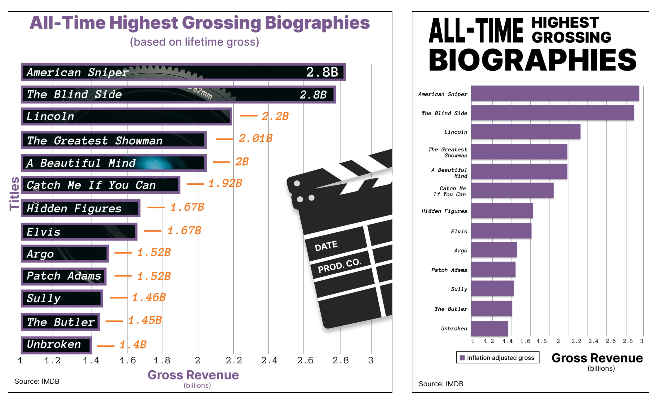

Figure 3: Bar charts for model applications with differing complexity predictions. (Left: 32.74. Right: 25.74.)

7.2 Dataset Applications: Studying Cognition

This work's labels and complexity ratings for the MASSVIS dataset can enhance understanding of human cognition during visualization experiences. Through a proof-of-concept case study, we explore how visual complexity affects data comprehension, confidence, and trust. We recruited 40 participants (M_{age} = 32.3, SD_{age} = 13.36) via Prolific

[24] to evaluate two MASSVIS visualizations: a high-complexity bar chart (*rating = 60.30, SD = 25.52*) and a low-complexity pie chart (*rating = 27.65, SD = 27.48*). Participants completed visual literacy assessment test questions [19], rated confidence and trust using a 6-point Likert scale, and shared insights on chart comprehension.

We found **higher visual complexity can hinder comprehension**. Participants performed worse on the more complex bar chart (VLAT accuracy = 0.64, $SD = 0.30$) than on the simpler pie chart (accuracy = 0.79, $SD = 0.22$). We also found evidence of miscalibrated confidence. Participants **reported higher confidence interpreting the complex bar chart** ($M = 3.7$, $SD = 1.30$) compared to the pie chart ($M = 3.2$, $SD = 1.40$), despite performing worse. This aligns with findings that overconfidence often correlates with lower actual ability [18, 31], underscoring the need for deeper exploration of confidence calibration in visualizations. Participants showed slightly **lower trust in the more complex bar chart** ($M = 4.25$, $SD = 1.21$) than in the simpler pie chart ($M = 4.45$, $SD = 0.89$). This supports theories linking clarity and trust [9, 30, 44], suggesting that reducing complexity may enhance trustworthiness. Future work could identify design features that foster trust in complex visualizations.

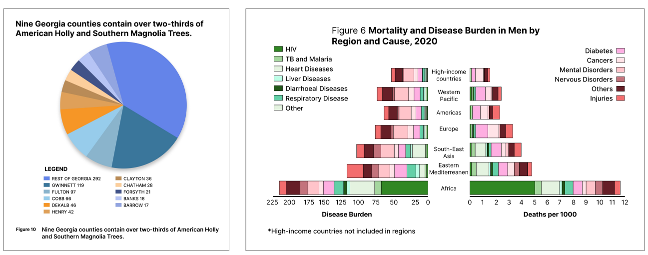

Figure 4: Pie and bar charts (modified to comply with copyright restrictions) for dataset applications case study, showing contrasting perceived complexities. Left: pie chart (27.65). Right: bar chart (60.30).

8 Conclusion

Generally, the more visual elements a visualization contains, the greater its perceived complexity. Future work should closely examine potential trade-offs of adding more or less elements in a visualization. For example, an appropriately complex visualization with the right number of elements might force people to slow down and think analytically about the presented data [16]. On the other hand, popular visualization designs such as small multiples and composite visualization dashboards often include a large number of charts. Our results predict these visualizations as more perceptually complex. Considering these visualizations are often used in visual analytic tools to assist decision-making, future work can investigate whether higher perceived complexity influences people's trust and comprehension in the data.

Acknowledgments

The authors thank Abhinav Vinod, Alex Soong, Saaliha Allauddin, Pranit Dodda, Carina Zeng, Kunal Mohindra, Mahathi Gumudavelli, and Reid Laughton for assistance with this research. This work is partially supported by NSF IIS-2237585 and IIS-2311575.

References

[1] Kiran Ajani, Elsie Lee, Cindy Xiong, Cole Nussbaumer Knaflic, William Kemper, and Steven Franconeri. 2021. Declutter and focus: Empirically evaluating design guidelines for effective data communication. *IEEE Transactions on Visualization and Computer Graphics* 28, 10 (2021), 3351–3364.

[2] Derya Akbaba, Jack Wilburn, Main T. Nance, and Miriah Meyer. 2021. Manifesto for Putting 'Chartjunk' in the Trash 2021! *alt.VIS, a workshop co-located with IEEE VIS* (2021).

[3] Cindy Xiong Bearfield, Chase Stokes, Andrew Lovett, and Steven Franconeri. 2024. What Does the Chart Say? Grouping Cues Guide Viewer Comparisons and Conclusions in Bar Charts. *IEEE Transactions on Visualization and Computer Graphics* 30, 8 (2024), 5097–5110. doi:10.1109/TVCG.2023.3289292

[4] Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2015. Beyond memorability: Visualization recognition and recall. *IEEE transactions on visualization and computer graphics* 22, 1 (2015), 519–528.

[5] Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What makes a visualization memorable? *IEEE transactions on visualization and computer graphics* 19, 12 (2013), 2306–2315.

[6] 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:10.1109/TVCG.2024.3456318

[7] Romain Collaud, Irene Reppa, Lara Défayes, Sine McDougall, Nicolas Henchoz, and Andreas Sonderegger. 2022. Design standards for icons: The independent role of aesthetics, visual complexity and concreteness in icon design and icon understanding. *Displays* 74 (2022), 102290.

[8] Don C Donderi, PWGSC Contract No, and Sharon McFadden. 2003. *A complexity measure for electronic displays: Final report on the experiments*. Department of National Defence, Defence Research & Development Canada-Toronto, North York, ON, Canada.

[9] Hamza Elhamdadi, Adam Stefkovics, Johanna Beyer, Eric Moerth, Hanspeter Pfister, Cindy Xiong Bearfield, and Carolina Nobre. 2024. Vistrust: a Multidimensional Framework and Empirical Study of Trust in Data Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 348–358. doi:10.1109/TVCG.2023.3326579

[10] Steven L Franconeri, Lace M Padilla, Priti Shah, Jeffrey M Zacks, and Jessica Hullman. 2021. The science of visual data communication: What works. *Psychological Science in the public interest* 22, 3 (2021), 110–161.

[11] Kary Främling. 2023. Feature Importance versus Feature Influence and What It Signifies for Explainable AI. In *Explainable Artificial Intelligence*, Luca Longo (Ed.). Springer Nature Switzerland, Cham, 241–259.

[12] Racquel Fygenson, Steven Franconeri, and Enrico Bertini. 2024. The Arrangement of Marks Impacts Afforded Messages: Ordering, Partitioning, Spacing, and Coloring in Bar Charts. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 1008–1018. doi:10.1109/TVCG.2023.3326590

[13] Mariano García, Albert N Badre, and John T Stasko. 1994. Development and validation of icons varying in their abstractness. *Interacting with computers* 6, 2 (1994), 191–211.

[14] Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky. 2010. A tour through the visualization zoo. *Commun. ACM* 53, 6 (2010), 59–67.

[15] Mary Hegarty. 2005. Multimedia learning about physical systems. *The Cambridge handbook of multimedia learning* (2005), 447–465.

[16] Jessica Hullman, Eytan Adar, and Priti Shah. 2011. Benefitting infovis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2213–2222.

[17] Daniel Kahneman. 2011. *Thinking, fast and slow*. macmillan, New York, NY, United States.

[18] Justin Kruger and David Dunning. 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology* 77, 6 (1999), 1121.

[19] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2016. Vlat: Development of a visualization literacy assessment test. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 551–560.

[20] Sharon Lin, Julie Fortuna, Chinmay Kulkarni, Maureen Stone, and Jeffrey Heer. 2013. Selecting Semantically-Resonant Colors for Data Visualization. *Computer Graphics Forum* 32, 3pt4 (2013), 401–410. doi:10.1111/cgf.12127 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12127

[21] Giorgia Lupi and Stefanie Posavec. 2016. *Dear data*. Chronicle books, San Francisco, CA, United States.

[22] Laura E Matzen, Michael J Haass, Kristin M Divis, Zhiyuan Wang, and Andrew T Wilson. 2017. Data visualization saliency model: A tool for evaluating abstract data visualizations. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 563–573.

[23] Joachim Meyer, David Shinar, and David Leiser. 1997. Multiple factors that determine performance with tables and graphs. *Human factors* 39, 2 (1997),

268–286.
[24] Stefan Palan and Christian Schitter. 2018. Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17 (2018), 22–27.
[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[26] Md Dilshadur Rahman, Ghulam Jilani Quadri, Bhavana Doppalapudi, Danielle Albers Szafir, and Paul Rosen. 2025. A Qualitative Analysis of Common Practices in Annotations: A Taxonomy and Design Space. *IEEE Transactions on Visualization and Computer Graphics* 31, 1 (2025), 360–370. doi:10.1109/TVCG.2024.3456359
[27] Katharina Reinecke, Tom Yeh, Luke Miratrix, Rahmatri Mardiko, Yuechen Zhao, Jenny Liu, and Krzysztof Z. Gajos. 2013. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2049–2058. doi:10.1145/2470654.2481281
[28] Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. 2007. Measuring visual clutter. *Journal of vision* 7, 2 (2007), 17–17.
[29] Gabriel Ryan, Abigail Mosca, Remco Chang, and Eugene Wu. 2018. At a glance: Pixel approximate entropy as a measure of line chart complexity. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 872–881.
[30] Andrew K Schnackenberg and Edward C Tomlinson. 2016. Organizational transparency: A new perspective on managing trust in organization-stakeholder relationships. *Journal of management* 42, 7 (2016), 1784–1810.
[31] Corwin Senko, Andrew H Perry, and Melissa Greiser. 2022. Does triggering learners' interest make them overconfident? *Journal of Educational Psychology* 114, 3 (2022), 482.
[32] Priti Shah and Eric G Freedman. 2011. Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in cognitive science* 3, 3 (2011), 560–578.
[33] Priti Shah, Richard E Mayer, and Mary Hegarty. 1999. Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of educational psychology* 91, 4 (1999), 690.
[34] Sungbok Shin, Sunghyo Chung, Sanghyun Hong, and Niklas Elmqvist. 2022. A scanner deeply: Predicting gaze heatmaps on visualizations using crowdsourced eye movement data. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 396–406.
[35] Tarique Siddiqui, Paul Luh, Zesheng Wang, Karrie Karahalios, and Aditya Parameswaran. 2018. Shapesearch: flexible pattern-based querying of trend line visualizations. *Proc. VLDB Endow.* 11, 12 (Aug. 2018), 1962–1965. doi:10.14778/3229863.3236235
[36] Joan G Snodgrass and Mary Vanderwart. 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory* 6, 2 (1980), 174.
[37] StatsPanda. 2022. The Highest Grossing Movies of All Time. https://www.instagram.com/statspanda/p/Ci79Qv5PNOd/.
[38] Edward R Tufte. 2001. *The visual display of quantitative information*. Vol. 2. Graphics press, Cheshire, CT.
[39] Adam B. Vary. 2022. All-Time Domestic Box Office Adjust for Inflation. https://www.buzzfeednews.com/article/adambvary/the-force-awakens-avatar-box-office.
[40] Yong Wang, Daniel Archambault, Hammad Haleem, Torsten Moeller, Yanhong Wu, and Huamin Qu. 2019. Nonuniform Timeslicing of Dynamic Graphs Based on Visual Complexity. In *2019 IEEE Visualization Conference (VIS)*. IEEE, New York, NY, United States, 1–5. doi:10.1109/VISUAL.2019.8933748
[41] Yong Wang, Zhihua Jin, Qianwen Wang, Weiwei Cui, Tengfei Ma, and Huamin Qu. 2019. Deepdrawing: A deep learning approach to graph drawing. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 676–686.
[42] Jeremy Watt, Reza Borhani, and Aggelos K Katsaggelos. 2020. *Machine learning refined: Foundations, algorithms, and applications*. Cambridge University Press, Cambridge, CB2 8EA.
[43] Florian Windhager, Alfie Abdul-Rahman, Mark-Jan Bludau, Nicole Hengesbach, Houda Lamqaddam, Isabel Meirelles, Bettina Speckmann, and Michael Correll. 2024. Complexity as Design Material Position Paper. In *2024 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*. IEEE, New York, NY, 71–80. doi:10.1109/BELIV64461.2024.00013
[44] Bernhard Winkler. 2000. *Which Kind of Transparency? On the Need for Clarity in Monetary Policy-Making*. Working Paper 26. European Central Bank.
[45] Cindy Xiong, Lisanne Van Weelden, and Steven Franconeri. 2019. The curse of knowledge in visual data communication. *IEEE transactions on visualization and computer graphics* 26, 10 (2019), 3051–3062.