Hey ChatGPT, can you visualize my data? – A Multi-Dimensional Study on using an LLM for Constructing Data Visualizations

M. Ströbel[1], K. Eckert[1], T. Nagel[1]

[1]Mannheim University of Applied Sciences, Germany

Abstract

This paper explores the effectiveness of an LLM in creating data visualizations across a spectrum of scenarios, characterized by three key dimensions: the complexity of the underlying data, the user's data visualization competencies, and the requirements of the resulting visualization. Based on an empirical study, we offer insights into the potential role of LLMs as tools for empowering users with varied expertise to effectively visualize data.

CCS Concepts

• ***Human-centered computing*** → ***Visualization;*** • ***Computing methodologies*** → *Natural language processing;*

1. Introduction

Creating visualizations that effectively communicate complex data insights is a challenge, particularly for individuals with varying levels of expertise and analytical needs [GTS10]. While user interface-based and low-code visualization tools have made strides in democratizing data visualization [GBTS13], they still require users to surmount a significant learning curve and possess certain technical skills [BLC*23]. Recent advancements in large language models (LLMs) have introduced a promising avenue for overcoming these barriers. LLMs can create and learn from visualization descriptions [KJP*24], generate data visualizations directly from natural-language instructions [CPB*22, Dib23, CLB23, MS23], or support visualization engagement [LM22], potentially bridging the gap for those who lack traditional data visualization skills [SNL*21].

This paper explores the effectiveness of LLMs in creating data visualizations across a spectrum of scenarios, characterized by three key dimensions: the complexity of the underlying data, the user's data visualization competencies, and the requirements of the resulting visualization. We designed an empirical study where LLMs generate data visualizations for a diverse set of combinations in these dimensions. The quality of these LLM-generated visualizations was assessed by an expert reviewer using established heuristics, focusing on time efficiency, supported insights, and design principles. Our research aims to explore the capabilities and limitations of LLMs in the field of data visualization, offering insights into their potential role as tools for empowering users with varied expertise to effectively visualize data.

Our contribution is three-fold: We synthesized literature-based aspects of data visualization construction, generated and evaluated visualizations for diverse scenarios, and documented the creation process alongside our key insights.

2. Methodology

For our study, we generated data visualizations for pre-selected datasets through specific prompts to an LLM. The effectiveness of the resulting visualizations is evaluated using two relevant heuristics, aimed at measuring the visualizations' utility in enhancing dataset understanding and their adherence to established design principles. Through this, we aim to assess the capabilities of LLMs in producing high-quality data visualizations that are both informative and aesthetically coherent.

We structured the experiments around three relevant factors: the *dataset complexity (D)*, the *user compentencies (U)*, and the *visualization requirements (V)*. Each of the factors is categorized into three levels, leading to $3 \times 3 \times 3 = 27$ combinations. *Data complexity (D)* was determined based on four attributes from data literacy literature [KS19] classifying data complexity of real-world data sets: scope, curation, size, and messiness. To test visualization construction for different *user competencies (U)* in a more controlled study environment, we did not recruit actual participants but instead, based on literature [KJP*24, BLC*23], identified three levels of user competencies. Beginners are familiar with everyday visualizations but lack foundational principles. Intermediate users understand basic visualization software but lack programming experience. Advanced users have deep knowledge, including design, and seek to refine LLM-generated visualizations. Lastly, we specified three *visualization requirements (V)*: a simple data visualization, an visualization with no specified complexity requirements, or a complex data visualization. We created prompts for each combination of the three factors, resulting in 27 experiments. This experimental design aims to systematically explore how variations in user skill levels and dataset complexity influence the effectiveness of LLM-generated visualizations.

DOI: 10.2312/evp.20241083

3. Study Design and Implementation

3.1. Design

In each experiment for one of the 27 combinations, ChatGPT 4 was asked to generate Python code for the visualization. The code was then used to render a static image of visualization. We selected three datasets based on the criteria from above, and used the Titanic dataset (D1, simple), a dataset about pop star Taylor Swift (D2, medium), and a dataset about Lego sets (D3, complex). In an experiment, a novice user (U1) might ask for a basic visualization (V1) of a dataset with medium complexity (D2). Prompts for user group U1 were kept broad (*"Create a simple, meaningful data visualization"*) to explore visualization generation for users with no prior skills, without specifying data tasks or visualization properties. For comparability, the data attributes selected by ChatGPT for U1 were reused and detailed for experiments with U2 and U3. U2 had the option to request one general improvement, while U3 could seek multiple specific enhancements (e.g. *"Check the result again carefully according to relevant principles for good and meaningful data visualizations"*).

3.2. Evaluation

To evaluate the generated visualizations (see Appendix), two assessment methods have been employed: the ICE-T heuristics [WAM*19] for determining visualization value (excluding factors around interaction due to the static nature of the visualizations produced), and the Data Visualization Checklist (DVC) by Evergreen and Emery [EE16] for design aspects. A data visualization expert conducted the evaluations, scoring each visualization based on these methodologies. Visualizations are deemed *valuable and good* if their overall ICE-T score is five or higher and *well-formatted* if they achieve more than 90% of the possible points on the DVC.

3.3. Results

We aggregated and visualized the evaluation results to identify potential patterns, as well as the analysis of the visualization techniques used. The scatter plot in Fig. 1 shows the distribution of evaluation results for each experiment. Out of 27 visualizations generated by ChatGPT and assessed by an expert, three were deemed high-quality and valuable according to the ICE-T evaluation framework. Most remaining visualizations scored above 4 points and no visualization below 3. In the design evaluation using the DVC, none of the visualizations met the threshold for being considered *well-formatted*. Most visualizations (25/27, 92.6%) scored between 60% and 80%, with two achieving only 50%.

4. Discussion

Given that most visualization (18/27, 66.7%) achieved a score higher than 4, and considering the ICE-T method recommends redesigning only for scores below 4, we see this as an indicator that ChatGPT, despite not ensuring uniformly high quality, is an effective tool for visualization creation for all levels of user expertise and visualization requirements. This utility is underscored by reports where participants requested generic visualizations instead of specific, task-driven prompts [SS23].

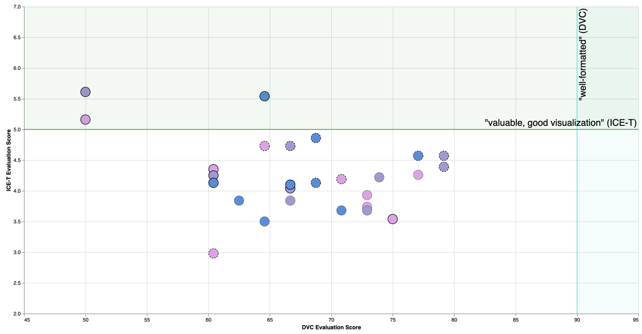

Figure 1: *Evaluation results for the 27 LLM-generated visualizations according to ICE-T and Data Visualization Checklist scores. Colors of circles encode user competencies (pink=low, violet=medium, blue=high). Green and turquoise lines mark the threshold for 'good' and 'well-formatted' visualizations.*

Overall, the visualizations scored from average to poor in design aspects, potentially due to the Data Visualization Checklist's specific requirements. While improvement prompts (for U3) were partly derived from this checklist, ChatGPT struggled with many of these design aspects, such as text placement, annotations, or labels.

Seven out of nine visualizations utilizing the Titanic dataset (D1) achieved favorable outcomes, with the top three visualizations all based on this dataset. It remains unclear whether these results stem from the dataset's low complexity or whether ChatGPT more easily creates valuable visualizations with familiar and frequently used datasets. As a mitigation strategy, generating three new artificial datasets of varied complexity could provide further insights.

To limit the scope of our experiments, we differentiated user groups solely by their data visualization experience, omitting programming skill levels. While opting for code generation allowed for flexibility beyond the libraries and environments ChatGPT supports in-system, this approach may not align well with users possessing minimal coding skills.

5. Conclusions

In conclusion, our study demonstrates that LLMs for generating data visualizations present a promising tool suitable for users across all levels of competencies, including those without prior experience in data visualization. Throughout our experiments, ChatGPT-generated visualizations reliably achieved a basic level of quality both in supporting data analysis as well as design. However, they often lack the analytics excellence and refined formatting necessary for final, publication-ready visualizations. Given these findings, we advocate for the use of LLM-generated visualizations primarily in the exploratory data analysis (EDA) phase, where the ability to rapidly generate and iterate on visualizations can enhance productivity and insight discovery. Still, for this phase and beyond, caution is essential with ChatGPT visualizations due to their variable and non-deterministic quality. Our study highlights the role of LLMs as a valuable starting point in the data visualization process, pointing towards a future where further advancements could broaden their applicability and effectiveness.

Proceedings published by Eurographics - The European Association for Computer Graphics.

References

[BLC*23] BURNS A., LEE C., CHAWLA R., PECK E., MAHYAR N.: Who Do We Mean When We Talk About Visualization Novices? In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg Germany, Apr. 2023), ACM, pp. 1–16. URL: https://dl.acm.org/doi/10.1145/3544548.3581524, doi:10.1145/3544548.3581524. 1

[CLB23] CHENG L., LI X., BING L.: Is GPT-4 a good data analyst? In *Findings of the Association for Computational Linguistics: EMNLP 2023* (Singapore, Dec. 2023), Bouamor H., Pino J., Bali K., (Eds.), Association for Computational Linguistics, pp. 9496–9514. URL: https://aclanthology.org/2023.findings-emnlp.637, doi:10.18653/v1/2023.findings-emnlp.637. 1

[CPB*22] CHEN Q., PAILOOR S., BARNABY C., CRISWELL A., WANG C., DURRETT G., DILLIG I.: Type-directed synthesis of visualizations from natural language queries. *Proc. ACM Program. Lang. 6*, OOPSLA2 (oct 2022). URL: https://doi.org/10.1145/3563307, doi:10.1145/3563307. 1

[Dib23] DIBIA V.: LIDA: A tool for automatic generation of grammar-agnostic visualizations and infographics using large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)* (Toronto, Canada, July 2023), Bollegala D., Huang R., Ritter A., (Eds.), Association for Computational Linguistics, pp. 113–126. URL: https://aclanthology.org/2023.acl-demo.11, doi:10.18653/v1/2023.acl-demo.11. 1

[EE16] EVERGREEN S., EMERY A. K.: The Data Visualization Checklist, Oct. 2016. URL: https://stephanieevergreen.com/updated-data-visualization-checklist/. 2

[GBTS13] GRAMMEL L., BENNETT C., TORY M., STOREY M.-A.: A Survey of Visualization Construction User Interfaces. *EuroVis - Short Papers* (2013), 5 pages. URL: http://diglib.eg.org/handle/10.2312/PE.EuroVisShort.EuroVisShort2013.019-023, doi:10.2312/PE.EUROVISSHORT.EUROVISSHORT2013.019-023. 1

[GTS10] GRAMMEL L., TORY M., STOREY M.: How Information Visualization Novices Construct Visualizations. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov. 2010), 943–952. URL: http://ieeexplore.ieee.org/document/5613431/, doi:10.1109/TVCG.2010.164. 1

[KJP*24] KO H.-K., JEON H., PARK G., KIM D. H., KIM N. W., KIM J., SEO J.: Natural language dataset generation framework for visualizations powered by large language models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11-16, 2024, Honolulu, HI, USA* (2024). doi:10.48550/arXiv.2309.10245. 1

[KS19] KJELVIK M. K., SCHULTHEIS E. H.: Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE—Life Sciences Education 18*, 2 (2019). doi:10.1187/cbe.18-02-0023. 1

[LM22] LIEW A., MUELLER K.: Using Large Language Models to Generate Engaging Captions for Data Visualizations. In *IEEE VIS Workshop on NLVIZ* (2022). doi:10.48550/arXiv.2212.14047. 1

[MS23] MADDIGAN P., SUSNJAK T.: Chat2vis: Generating data visualizations via natural language using chatgpt, codex and gpt-3 large language models. *IEEE Access 11* (2023), 45181–45193. doi:10.1109/ACCESS.2023.3274199. 1

[SNL*21] SRINIVASAN A., NYAPATHY N., LEE B., DRUCKER S. M., STASKO J.: Collecting and characterizing natural language utterances for specifying data visualizations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2021), CHI '21, Association for Computing Machinery. doi:10.1145/3411764.3445400. 1

[SS23] SRINIVASAN A., SETLUR V.: BOLT: A Natural Language Interface for Dashboard Authoring. In *EuroVis 2023 - Short Papers* (2023), Hoellt T., Aigner W., Wang B., (Eds.), The Eurographics Association. doi:10.2312/evs.20231035. 2

[WAM*19] WALL E., AGNIHOTRI M., MATZEN L., DIVIS K., HAASS M., ENDERT A., STASKO J.: A Heuristic Approach to Value-Driven Evaluation of Visualizations. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (Jan. 2019), 491–500. URL: https://ieeexplore.ieee.org/document/8454343/, doi:10.1109/TVCG.2018.2865146. 2