Machine Learning for Geospatial Visualisation Recommendation

Anthony Joseph Jungo, Kaiza Kunonu Ilomo, John Waithaka

Carnegie Mellon University

Kigali, Rwanda

{ajosephj, kilomo, jwaithak}@andrew.cmu.edu

# Topic And Motivation

This research is focused on systems for recommending effective data visualizations. An effective data visualisation is one that accurately represents patterns in data, efficiently communicates these patterns and is useful for achieving a certain goal [1]. Following Hu et al. [2], we define the act of data visualisation as the process of making design choices such as visualisation type, based on the properties of a dataset, that maximize the effectiveness of the resulting visualisation. Therefore, the recommendation of effective data visualisations involves making design choices that most effectively visualise data. We focus on the visualisation of geospatial data, which is data with a geographic location dimension [3].

We believe the automation of the recommendation of effective geospatial visualisation is a key to the benefits of geospatial data analysis in Africa. Automated visualisation may overcome the barrier of the lack of data visualisation expertise in Africa and the complexity of data visualisation even for experts.

# Literature Review

Most of the early research on automated visualisation recommendation focused on rule-based systems. Mackinlay’s A Presentation Tool (APT) [4] exemplifies such systems. The development of APT involved two parts. The first was the codification of visualisation guidelines (from Semiology of Graphics [5], a seminal work on visualisation) into programmatic form and using the resulting rules to rank visualisations by their effectiveness. The second was the efficient generation of a set of diverse visualisations which were then ordered and recommended according to their effectiveness rank. This work demonstrated the ability of rule-based systems to recommend effective visualisations and was a foundation for later research like Show Me [6] and SAGE [7]. However, Hu et al. [2] point out the following limitations of rule-based systems. First, rule-based models may be too simple to model the complex process of effective visualisation. Two, since the rule sets in rule-based models usually have a combinatorial nature, increasing the dimensions in input data would result in an exponential increase in the rule set size. Third, the identification of visualisation guidelines and their codification is a complex process requiring considerable expertise.

Hu et al. [2] propose an ML model for visualisation recommendation. They define visualisation recommendation as the process of making a set of design choices, such as visualisation type, to effectively visualise a dataset. They use a neural network trained on publicly accessible visualisation descriptions and corresponding datasets to model the design choices that effectively visualise a dataset. The resulting neural network model performed comparably to people with domain knowledge and who spent a lot of time visualising data. This work showed that machine learning can recommend sufficiently effective data visualisations. This work, however, only considered typical tabular data and visualisations like bar charts, line charts and scatter plots. It does not consider geospatial data and visualisations.

Dibia and Demiralp [8] propose Data2Vis, an ML-based system to automatically create effective visualizations. They formulate the problem of visualisation generation as a machine translation problem, where data specifications are mapped to visualization specifications. They use a deep neural translation model to map JSON-formatted datasets to Vega-lite visualization specifications. However, the study did not consider geospatial data and its visualisation techniques.

Other similar works, like Luo et al.’s [9], which use ML models to recommend effective visualisation also do not consider geospatial data and visualisation. This raises the question of whether such works apply to geospatial data and visualisation.

Wu et al. [10], presented MobileVizFixer, a tool based on reinforcement learning to automate and optimize visualization in mobile devices. This involves deconstructing SVG visualisations into a declarative format and then making design choices that optimise the original visualisations to be mobile-friendly based on a pre-defined but extensible set of criteria. This work, however, is not relevant to our research scope in the sense that it is a corrective system. It does not recommend effective visualisations given a dataset, but instead recommends effective visualisations given another visualisation.

A range of studies have investigated alternative visualization methods for large geospatial datasets. Koua [1] proposed leveraging computational algorithms and graphical representations to uncover patterns and relationships within the data. Considering Koua's emphasis on computational algorithms, ML/AI techniques could be applied to automate the process of pattern recognition and relationship identification within geospatial datasets. For example, supervised machine learning algorithms could be trained on labelled datasets to automatically detect and classify patterns in spatial data.

The exploration of geographical metadata has been approached through a combination of visual and automatic data mining techniques. Albertoni [2] advocates for the use of visualization-based tools to facilitate this exploration, with the latter also integrating automatic data mining methods. Albertoni's work suggests that ML/AI techniques could enhance the automatic data mining process by enabling algorithms to learn from the data and extract insights without explicit programming. For instance, unsupervised learning algorithms such as clustering could be used to identify patterns and correlations within geographical metadata.

# Research Question

Can machine learning automate the effective visualization of geospatial data? In particular, can it recommend geospatial visualizations that effectively communicate patterns in geospatial datasets?

# References

[1] Y. Zhu, ‘Measuring Effective Data Visualization,’ in *Advances in Visual Computing*, G. Bebis, R. Boyle, B. Parvin, D. Koracin, N. Paragios, S.-M. Tanveer, T. Ju, Z. Liu, S. Coquillart, C. Cruz-Neira, T. Müller, and T. Malzbender, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2007, pp. 652–661. doi: 10.1007/978-3-540-76856-2\_64.

[2] K. Hu, M. A. Bakker, S. Li, T. Kraska, and C. Hidalgo, ‘VizML: A Machine Learning Approach to Visualization Recommendation,’ in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI ’19). New York, NY, USA: Association for Computing Machinery, May 2019, pp. 1–12. doi: 10.1145/3290605.3300358.

[3] K. Stock and H. Guesgen, ‘Chapter 10 - Geospatial Reasoning With Open Data,’ in *Automating Open Source Intelligence*, R. Layton and P. A. Watters, Eds., Boston: Syngress, 2016, pp. 171–204. doi: https://doi.org/10.1016/B978-0-12-802916-9.00010-5.

[4] J. Mackinlay, ‘Automating the design of graphical presentations of relational information,’ *ACM Trans. Graph.*, vol. 5, no. 2, pp. 110–141, Apr. 1986, doi: 10.1145/22949.22950.

[5] J. Bertin, *Semiology of Graphics : Diagrams, Networks, Maps*. Madison, Wis: University of Wisconsin Press, 1983.

[6] J. Mackinlay, P. Hanrahan, and C. Stolte, ‘Show Me: Automatic Presentation for Visual Analysis,’ *IEEE Trans. Vis. Comput. Graph*., vol. 13, no. 6, pp. 1137–1144, Nov. 2007, doi: 10.1109/TVCG.2007.70594.

[7] S. F. Roth, J. Kolojejchick, J. Mattis, and J. Goldstein, ‘Interactive graphic design using automatic presentation knowledge,’ in *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence* (CHI ’94), Boston, Massachusetts, United States: ACM Press, 1994, pp. 112–117. doi: 10.1145/191666.191719.

[8] V. Dibia and Ç. Demiralp, ‘Data2Vis: Automatic Generation of Data Visualizations Using Sequence-to-Sequence Recurrent Neural Networks,’ *IEEE Comput. Graph*. Appl., vol. 39, no. 5, pp. 33–46, Sep. 2019, doi: 10.1109/MCG.2019.2924636.

[9] Y. Luo, X. Qin, N. Tang, and G. Li, ‘DeepEye: Towards Automatic Data Visualization,’ in *2018 IEEE 34th International Conference on Data Engineering* (ICDE), Apr. 2018, pp. 101–112. doi: 10.1109/ICDE.2018.00019.

[10] A. Wu, W. Tong, T. Dwyer, B. Lee, P. Isenberg, and H. Qu, ‘MobileVisFixer: Tailoring Web Visualizations for Mobile Phones Leveraging an Explainable Reinforcement Learning Framework,’ *IEEE Trans. Vis. Comput. Graph*., vol. 27, no. 2, pp. 464–474, Feb. 2021, doi: 10.1109/TVCG.2020.3030423.