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

# Motivation

Data analytics is known to significantly improve decision-making and performance in areas such as business, public administration, and healthcare. Geospatial data analytics, specifically, has big potential benefits in agriculture in applications such as yield optimization and prediction, and early warning systems for famine and crop failure. This is especially important for Africa, where the economy and people’s livelihoods are highly dependent on agriculture.

Data visualization is an essential part of data analysis since it enables the identification and communication of insights drawn from data, which enables data-driven decision-making. However, effective visualization is a difficult task – there are many ways to visualize a dataset but only a few produce effective visualizations. This, along with the shortage of ICT skills in Africa, keeps Africa from reaping the potential benefits of geospatial data analytics. We believe that providing tools for automating effective geospatial data visualization will increase the use of geospatial data analysis in Africa. This will in turn yield significant benefits for Africans.

# Research Question

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

Research on the automation of data visualization goes back to 1986 [1]. The rise of Machine Learning (ML) and AI increased focus in this area [2]. For example, Hu et al. built an ML model trained on data and graphs from Plotly Community Feed [3] to recommend between bar, line or scatter plots, given a dataset [4]. Notably, there is very little research on automating effective visualization of geospatial data. Most work focuses on visualizations like bar charts, scatter plots and the like.

The scope of this research will be to investigate whether there is a simple ML model that can appropriately recommend between a path map (Fig. 1) and a point distribution map (Fig. 2), given geospatial datasets of the same format. We chose these two visualization types because their applications are easily distinguishable to a non-expert (but not necessarily to an ML model). Furthermore, the data for both visualization types can take the same format.

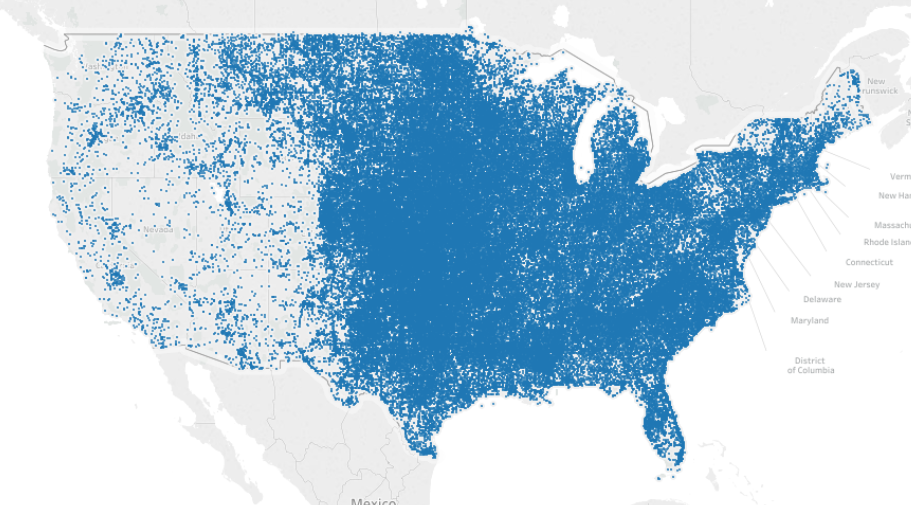


Fig. 1. Point Distribution map [5]

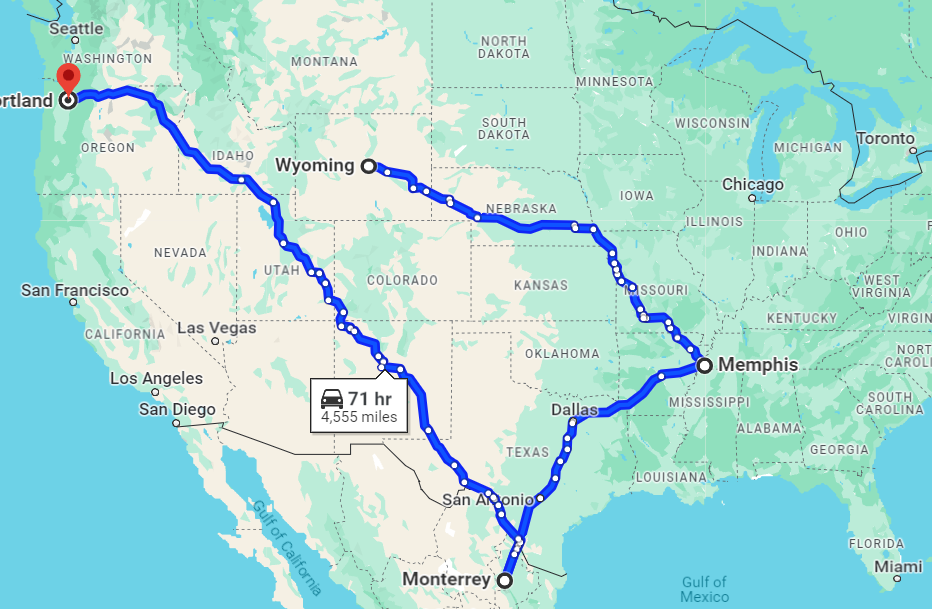


Fig. 2. Path Map [6]

##### References

[1] 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.

[2] A. Wu et al., ‘AI4VIS: Survey on Artificial Intelligence Approaches for Data Visualization,’ *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 12, pp. 5049–5070, Dec. 2022, doi: 10.1109/TVCG.2021.3099002.

[3] ‘Plotly | Make charts and dashboards online’. Accessed: Mar. 05, 2024. [Online]. Available: https://chart-studio.plotly.com/feed/#/

[4] K. Z. Hu, M. A. Bakker, S. Li, T. Kraska, and C. A. Hidalgo, ‘VizML: A Machine Learning Approach to Visualization Recommendation’. *arXiv*, Aug. 14, 2018. doi: 10.48550/arXiv.1808.04819.

[5] ‘Create Maps that Highlight Visual Clusters of Data in Tableau - Tableau’. Accessed: Mar. 18, 2024. [Online]. Available: https://help.tableau.com/current/pro/desktop/en-us/maps\_howto\_pointdistribution.htm

[6] ‘Google Maps’. Accessed: Mar. 18, 2024. [Online]. Available: https://www.google.com/maps