Machine Learning for Geospatial Visualisation Recommendation

Anthony Joseph Jungo, Kaiza Kunonu Ilomo, John Waithaka

Carnegie Mellon University

Kigali, Rwanda

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

I. 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 component of data analytics. It enables intelligible communication of patterns extracted from large volumes of data, enabling data-driven decision-making. However, effectively communicating patterns in data through visualizations is not easy. There are many types of data visualizations, each revealing (sometimes subtly) different patterns in data. Choosing the most effective visualization to reveal hidden patterns in a dataset is often not a simple task. 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.

II. RESEARCH QUESTION

Can machine learning automate the effective visualization of geospatial data? 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 (figure 1) and a point distribution map (figure 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); path maps are used to visualize routes or movement and point distribution maps are used to visualize location or density. Furthermore, the data for both visualization types can take the same format.

The geospatial datasets will be CSV files where each record is indexed by coordinate data and may contain a categorical feature characterizing the record.

III. APPROACH

The first step will be to obtain geospatial data - both movement or route data and point distribution data. If necessary, synthetic data will be generated. We will then preprocess this data to ensure all the obtained geospatial datasets have the same format. This will ensure the ML models cannot identify the appropriate visualization from the dataset format but instead from the data itself. Each geospatial dataset will then be manually labelled as being better visualized by a path map or a point distribution map.

The next step will be to identify an effective way to feed a geospatial dataset as input to an ML model. Lastly, we will develop and test different ML models to determine their ability to effectively recommend between path map and point distribution map visualizations.
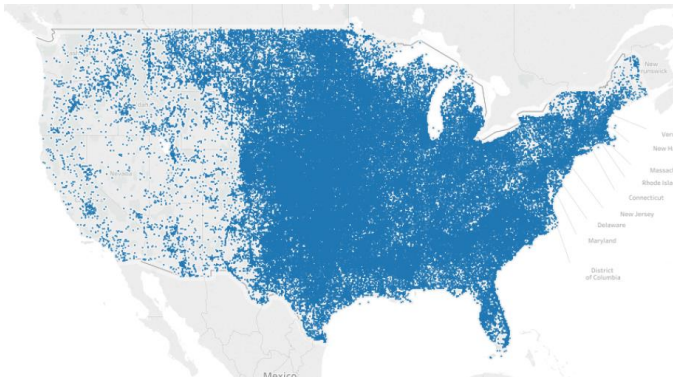
Figure 1: Point Distribution Map (Source: https://help.tableau.com/current/pro/desktop/en-us/maps_howto_pointdistribution.htm)
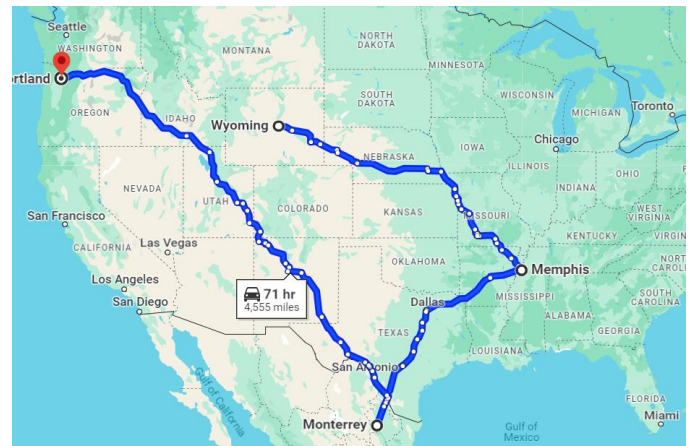

Figure 2: Path **Map** (Source: https://www.google.com/maps)

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.