Analysis of Uber travel times across major cities in the world

Vihang Godbole, Shane Pearlman

***Abstract*—Spatio-temporal information is one of the richest kinds of information since it allows us to work with multiple dimensions of data. This project aims at analyzing the publicly available *Uber travel times* dataset to extract valuable information about usage trends and traffic patterns in Uber rides. We use some domain specific techniques to visualize and convey it in a meaningful, yet intuitive manner.**

I. Motivation

THE failure of public transport systems to be an efficient, reliable, and most importantly versatile solution to the commuter lifestyle has paved way for innovations in ride sharing, driver less cars etc. Uber and other shared taxi services have become a ubiquitous mode of transport, especially in larger cities. Analyzing and understanding their usage trends gives us a chance to make informed decisions both as consumers and policy makers. Moreover, there is a history of interesting applications of traffic management in data visualization. General managers and City Ops teams need to have up to date information of the current supply-demand and traffic distribution. Access to aggregated, graphical data helps better understand a city market. Another possible use is in data science, which needs exploratory interfaces for multidimensional data (broken down by product, time, and geo). Thus, there are many use cases of having a visual representation of the Uber dataset.

II. Related Work

Much work has been done in the domain of traffic analysis and visualization. For example, work done by Adrienko et. al. [1] for analyzing flight trajectories was used to visualize traffic along motorways as well [2]. Some analysis has been done by news outlets and tech journals as well. Notably among them, FiveThirtyEight [3] has studied Uber usage according to the demographics, geographic areas etc. In fact, data visualization is such an important tool in this domain that Uber themselves have a large group dedicated to it [4]. They have investigated various events like the impact of road closures in London, traffic conditions after floods in Nairobi etc. They have also launched open source multiple frameworks called kepler.gl and deck.gl for visual exploratory analysis.

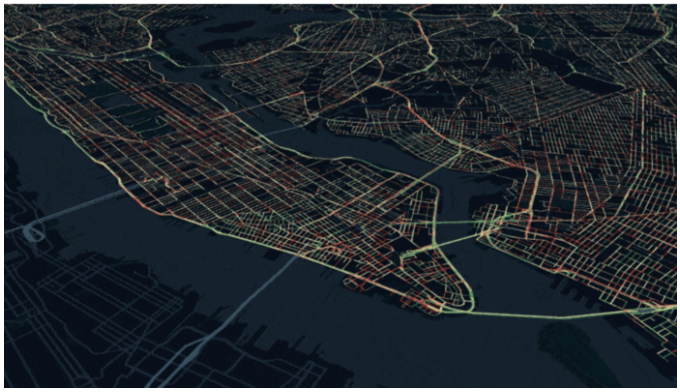

Fig. 1. Uber Traffic visualization

III. Proposed research directions

In our project our main aim is to provide a comprehensive set of visualizations to easily explore trends and patterns in Uber usage. Since the data spans across continents, we believe we will obtain some very interesting trends depending on the country and location. Since we have some data specific to New York City, we plan on using that as a case study and giving an in depth analysis of various other aspects of Uber usage as well. Even though traffic analysis is not a new idea, we hope to use techniques learned in class and analysis frameworks developed in Adrienko et. al. [1] to specifically work with Uber data. Some such techniques might include data transformations, using artificial spaces, interactive filtering and data aggregation etc. Some

Vihang Godbole - CMPS261, Shane Pearlman - CMPS161.

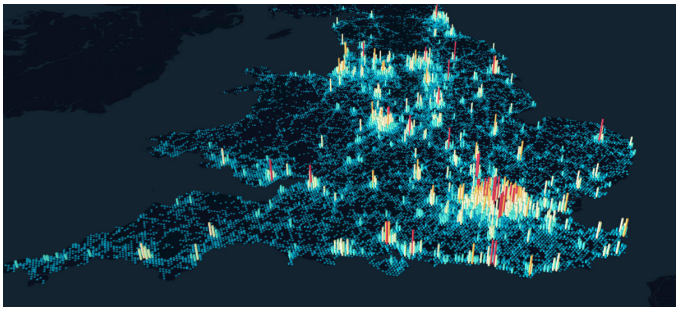

Fig. 2. Uber Map overlay

interesting correlations and trends to look at might be -

1) Correlations between availability of public transport and Uber usage
2) Busiest times of the year
3) Major areas of cities with traffic - airports, restaurants etc.

IV. Further information

Since two people are going to be working on this project, it is important to clearly state the work distribution beforehand. As per the course requirements for CMPS161, the project is worth 10% of the total grade. Thus, Shane will work mainly with only the implementation of the final visualizations. Over two assignments, he will be familiar with geographical mapping and data viz tools. I will carry out most of, if not all of the prior work. This mainly includes data analysis, data transformation to artificial spaces, calculation of similarity and other metrics etc.

The data consists of Uber travel times across 10+ major cities around the world. This means that for every city, we have the mean travel time between various ”sectors” (smaller areas of the city). This data is by the hour, day, month, quarter, year and also by peak times (morning/evening rush). Additionally, we have even more data for NYC with latitude, longitude for each Uber pickup and drop. We hope to get insight into various other aspects as well for example - using number of pickups from a particular ”sector”, we can find high traffic areas, or if the mean travel time to the northern sectors is more during evenings, this might show that there are many places of interest in the northern parts of the city. We currently do not have specific information about Uber rides to airports, restaurants etc. However, we hope to identify these POI (places of interest) by overlaying a map on top of the available data. We also believe Google Places API does provide information about the location of these POIs. We will look into this to complement our available data set and get richer visualizations.

References

[1] Andrienko, N., Andrienko, G., Garcia, J. M. C. and Scarlatti, D. *Analysis of Flight Variability: a Systematic Approach*, IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2018.2864811

[2] Andrienko, N., Andrienko, G., Garcia, J. M. C. and Scarlatti, D. *Analysis of Flight Variability: a Systematic Approach - Appendix to VAST 2018 submission* , IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2018.2864811

[3] Reuben Fischer-Baum and Carl Bialik *Uber Is Taking Millions Of Manhattan Rides Away From Taxis* [online] Available at: https://fivethirtyeight.com/features/uber-is-taking-millions-of-manhattan-rides-away-from-taxis/

[4] Uber Engineering Blog. (2016). *Engineering Intelligence Through Data Visualization at Uber.* [online] Available at: https://eng.uber.com/data-viz-intel/