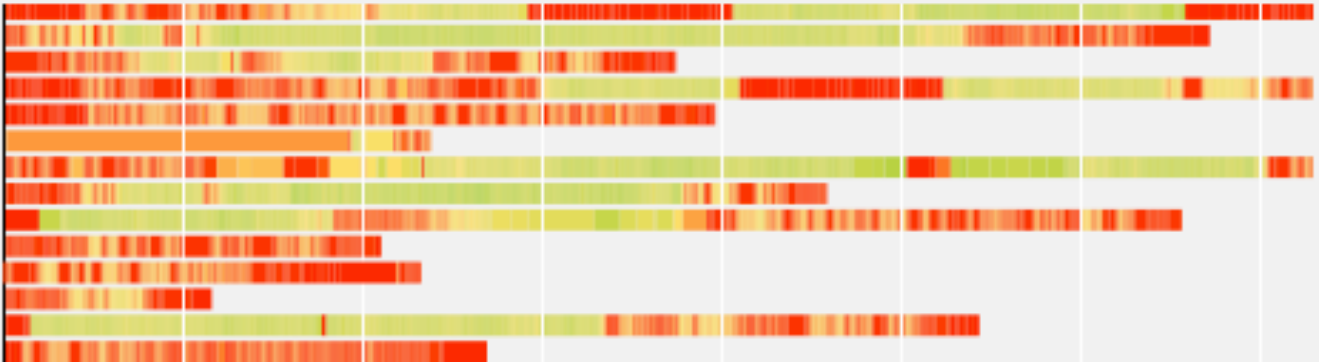


Visualizing People's Commuting Patterns

Michael Porath, UC Berkeley School of Information, mike@ischool.berkeley.edu

Trip Lengths and Velocity



Abstract— Commutes are a part of many people's lives. This paper shows patterns in the commutes of an individual, with the goal of changing driving behavior in the long run. An interactive visualization allows for brushing-and-linking of three different panels: a map showing commutes, when commutes take place during the day, and the length of each commute. A color coding indicates the velocity over the course of a commute. In addition to the visualization, the paper suggests different ways of visualizing commutes in general, velocity, acceleration/breaking patterns, as well as gas consumption and costs.

Index Terms— Driving Commute, Traffic, Velocity, Acceleration, Visualization, Pattern, Segmented, Behavior

1 INTRODUCTION

Commutes are a part of many people's lives. In that context, personal transportation has a big impact on the environment and individuals. The time spent in traffic is directly related with a loss in productivity, money spent on gas, and the emission of green house gases. This project was done with the goal of optimizing people's commutes. Looking at commuting data it becomes apparent that there is a striking regularity in people's commuting patterns. By visualizing those patterns and thus giving a way to analyze behavior, the project hopes to contribute to optimizations in people's commutes

This project is part of a larger research project. The data is exposed over an API. Data collection and the API is not subject of this paper, but required in understanding problems and challenges when dealing with the data. The larger goal of the research is to influence people to drive less and use more sustainable modes of transportation.

2 RELATED WORK

In their paper *Using GPS Data to Understand Driving Behavior*, Grengs et al. explore the possibilities in visualizing automatically collected GPS data [1]. Their focus is on driving in general, and not restricted on regular commutes. Their paper visualized the data in two ways. On one hand, the research analyzes temporal patterns in commutes showing distribution graphs across weeks and within days. On the other hand, they computationally assess start and end points of trips.

Stopher et al. takes a more general approach in collecting and processing GPS data from mobile technologies [2]. Their paper

outlines different approaches and problems in collecting multi-day driving data and processing it. They also assess the usefulness of the automatic collection of driving data.

3 DATA

3.1 Data Collection

A handful of individual participating in the study of the larger research project installed an application on their mobile phone that automatically collects their data as they drive. The application collects periodical discrete data points with a frequency of five seconds. One data point includes

- Geodata [lat/lng]
- Timestamp [s]
- Accuracy horizontally and vertically [m]
- Velocity [mph]

3.2 API

The data is exposed via a RESTful API that provides the data in JSON format, as an array with one object per data point.

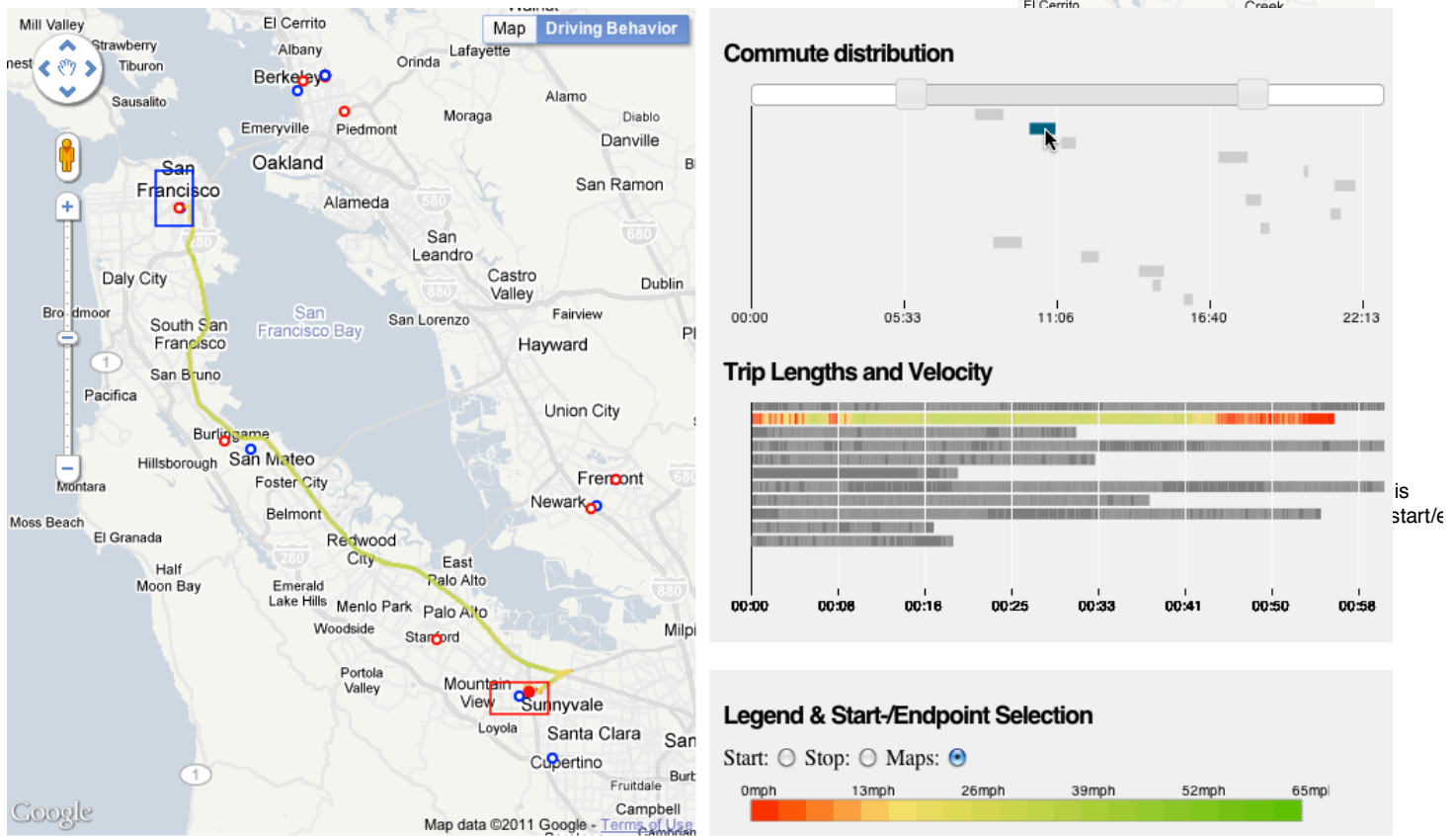


Fig. 1. Visualizing Commuting Patterns. The left hand panel on the left shows the trips start and end points, and commute routes. The panel on the right shows the trips distributed by the time of day (top) and by their length (middle) color coded by speed.

4 METHODS

The visualization shows commuting data for an individual. It consists of three parts. A map shows her commute start and end points, as well as the commute traces. A second panel shows more details about the commutes. At the top, the panel lists commutes and arranges them on an x-axis by the time of day. The trips that don't match the start and end points are grayed out. As a result, the visualization shows when during the day these commutes are taking place.

The bottom shows the same trips color-coded by velocity. This allows the exploration of patterns when it comes down to travel bottlenecks.

The three visualizations are linked to each other to allow brushing-and-linking of the data across all panels.

The visualization uses Protovis for visualizing the data, which includes the marks that appear on the map. To achieve this, however, the visualization makes extensive use of the Google Maps API (V3) in order to convert geospatial data into pixels on the canvas. The same method is used to get the correct bounding box that contains all the start and end points on the map.

4.1 Data Processing and Cleaning

Because the technology used for collecting data for this research project, the quality is not quite where it has to be for robust data analysis. For this visualization we had to filter out data that was not accurate enough to be shown on a map. More specifically, each data point with accuracy of more than 30 meters is removed from the data domain. The visualization does not interpolate these data points, but simply omits them.

4.2 Identifying Commute Start and End Points

Grengs et al. assess start and end trips algorithmically. This makes sense in the context of their research, because it focuses on a wider range of trips. This project focuses on commutes alone. The assumption goes that people know best what start and end points their commutes are. As a conclusion, the tool lets them draw bounding boxes to identify start and end trips (Fig. 2). The start and end points are white dots with blue or red outlines, depending if the location was a trip start or trip end. The start and end point for trips that fall into the bounding boxes appear filled.

4.3 Filtering Data by Time of Day

In a data collection for multi-day commutes the number of commute quickly grow to a size where you can't perceive the details really well any more. For this reason the commute distribution panel allows for filtering over time. When the slider handles are changed, the trip length visualization are filtered down to only show the commutes that happened during the indicated times of days.

As described above, the data points have a time stamp, which indicates seconds since Jan 01, 1970. In order to indicate the time of day, an algorithm determines the seconds passed since 12AM the day the commute takes place. With this algorithm in place, the data domain for the filtering panel is the amount of seconds of a day.

4.4 Color Coding Velocity

One main aspect of the visualization is the color-coding of the individual's velocity. The color-coding appears in two places. On one hand, it shows up in the Trip Length visualization showing the changes in velocity in a linear way. On the other hand, the same changes are plotted on the map indicating where the changes

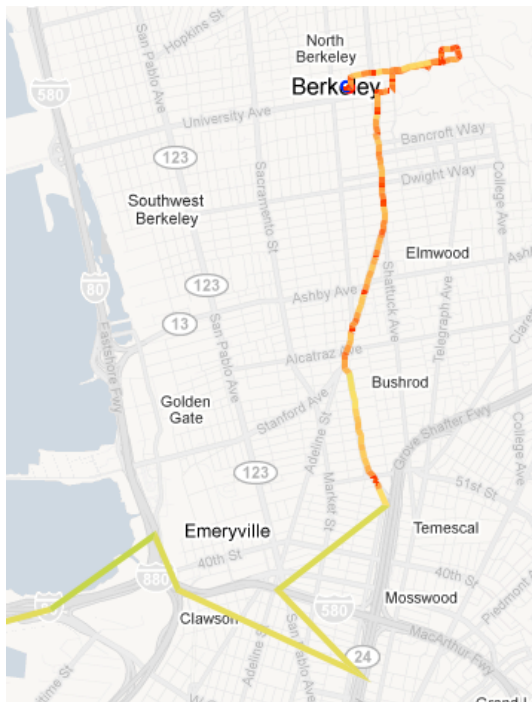


Fig. 3. Color-coded commute. Notice the frequent stops in city-traffic

happened. This is particularly useful to see where there was congestion, and where traffic could flow.

The color-coding makes use of a linear scale with three reference points. 0mph is shown in red, 15mph in yellow, and 65mph in green, with gradients changing in between. This allows for really reiterating the nature of a trip, indicating stop-and-go, bottlenecks, flowing traffic, and high-speed roads vs. city-traffic (see Fig. 3).

A segmented line connects the data points on the map. Each segment is color-coded separately, which makes for the effective change in colors over the course of a commute.

5 RESULTS

We tested the visualization with a handful of people as a means of an informal usability test. With an introduction and some guidance on how to use the tool they were able to perform the three major tasks:

- Selecting bounding boxes for the start and end points of commutes
- Filtering commutes by time
- Selecting commutes to show on the map

6 DISCUSSION

The visualization is particularly useful to highlight bottlenecks in traffic and showing where they happen regularly versus accidentally. As one user pointed out, this could be even more useful if the visualization would let users select a specific part of the commute and show the changes in velocity only for this part.

Another aspect that has proven useful is to compare the trip length based on the time of day. The data set for one participant in particular has shown that his commute back home from where he works takes significantly longer because it's during rush hour, while his commute to work is slightly earlier.

7 CONCLUSION AND FUTURE WORK

In the process of this project, a lot of experiments with showing the data have been done. Some of it didn't make it into the artefact, but could be leveraged in future work.

7.1 Comparing Different Routes

Where there are different routes for a commute, visualizing the difference in traffic flowing, costs, and time spent for the different routes could be particularly helpful.

7.2 Inter-Personal Comparison

With enough data available, the data corpus to be shown could be extended to showing the commute for other participants. This would be useful in particular to indicate what in general the best times are for a particular commute.

Along the same lines, a side-by-side comparison for the same commute of different people would be useful to cause a change in behaviour.

7.3 Emissions And Costs

Fuel consumption and emission based on the car make and model are available for the whole industry. This data could be correlated with the commute behaviour data and shown as another indicator to change behaviour.

7.4 Visualizing Other Modes of Transportation

The technology used already allows for collection of data for other modes of transportation. Comparing side by side the differences in cost and time for these different modes is another helpful future path.

7.5 Acceleration/Breaking Profiles

A small part of the commuting traces had acceleration/breaking data. While the data was not sufficient to be visualized, it indicated some promising future work. The acceleration/breaking profile shown in fig. 4 shows a bar diagram with time on the x axis and acceleration on the y axis, collected with a rate of 1Hz and aggregated using a sliding window of 6.

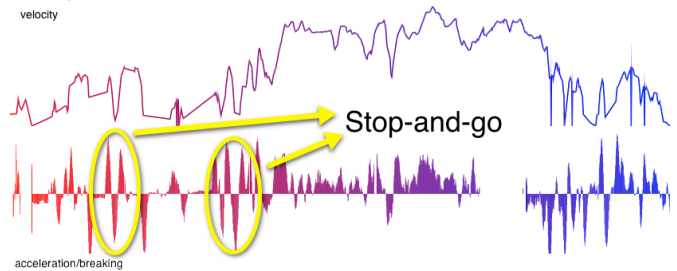


Fig. 4. Velocity (top) vs. acceleration (bottom). The acceleration profile shows stop-and-go behavior

7.6 Velocity as varying line width



Fig. 5. Velocity plotted with varying line width instead of color-coding

As mentioned above, data quality is key when it comes to plotting geospatial data on a map. Because the data quality was not optimal, color-coding was the right choice to indicate changes in velocity. However, given a trace with good data quality, plotting the velocity spatially as a varying line width is another effective way of visualizing congestion.

REFERENCES

- [1] Peter R. Stopher, "Collecting and Processing Data From Mobile Technologies", *8th International Conference on Survey Methods in Transport* (2008)
- [2] Joe Grengs, Xiaoguang Wang, and Lidia Kostyniuk, "Using GPS Data to Understand Driving Behavior.," *Journal of Urban Technology* 15, no. 2 (2008): 33-53.