

'Using Uber Movement to explore the Social Determinants of Traffic Flow': Methods and Initial Results *

Laurence Warner[†]

*working title

[†]lpwarner@uchicago.edu

1 Theory

I am interested in the extent to which traffic flow is dominated by commuting flow, particularly in the modern age in which the 9-5 job is on the decline. However, my hypothesis is still that commuting is the main cause of congestion. A nifty way of operationalizing traffic is to look at average travel time of rides along a given route.

2 Data

2.1 Uber Movement

Uber Movement provides anonymized data from over 2 billion trips in 21 cities around the world (6 in North America). The data is relatively limited. The main variable of interest is mean travel time between two zones. However, what is lacking is data on the volume of rides between each zone pair.

2.2 Data Collection

I focus on my analysis on the case of San Francisco. To analyze the data I download CSV files from the Uber Movement website.

3 Different avenues of Analysis

I have done some simple preliminary EDA. I have simplified analysis by choosing two arbitrary census zones in San Francisco. This analysis could be scaled up by downloading this data for multiple census tract pairs.

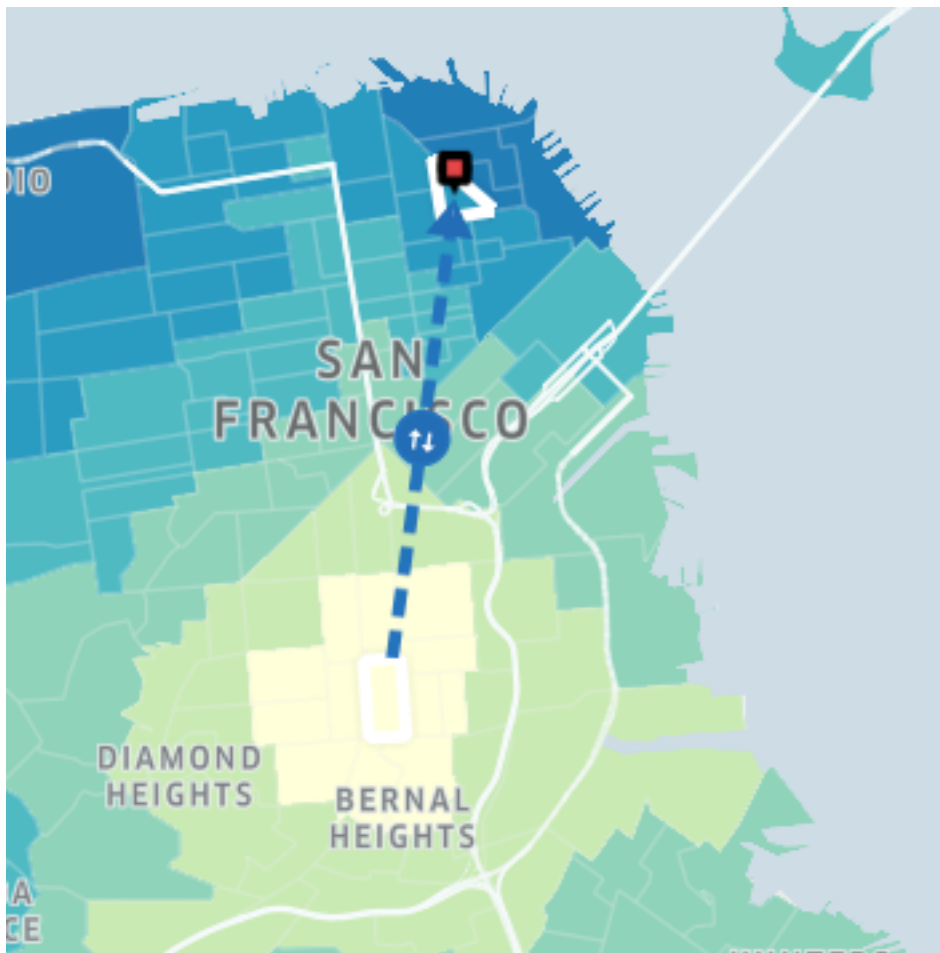
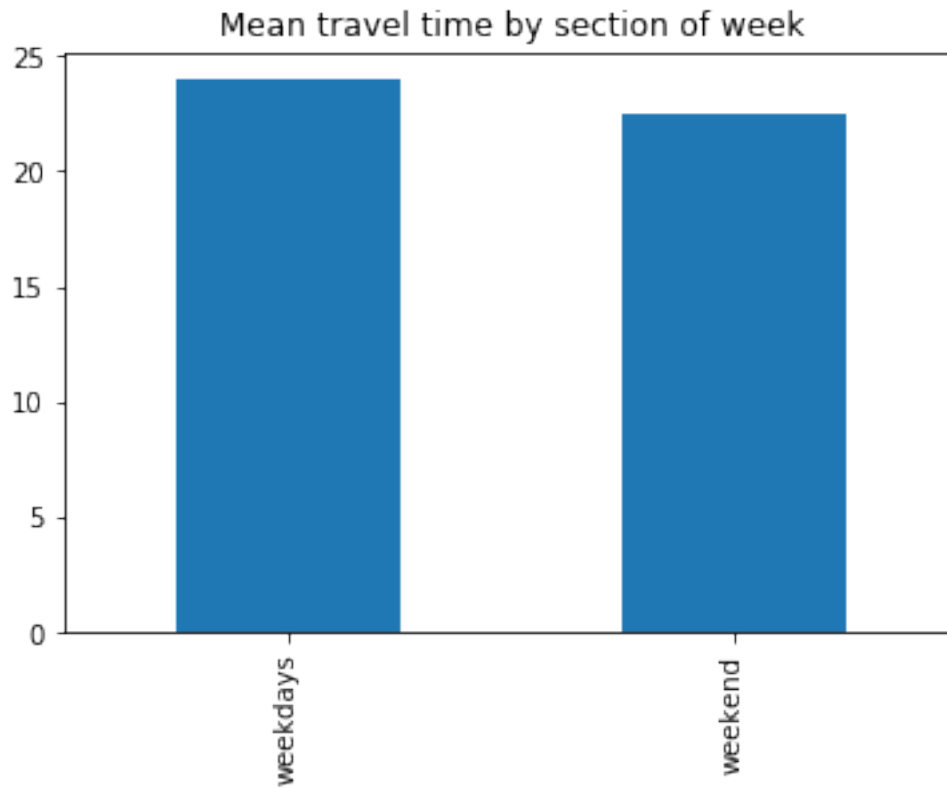


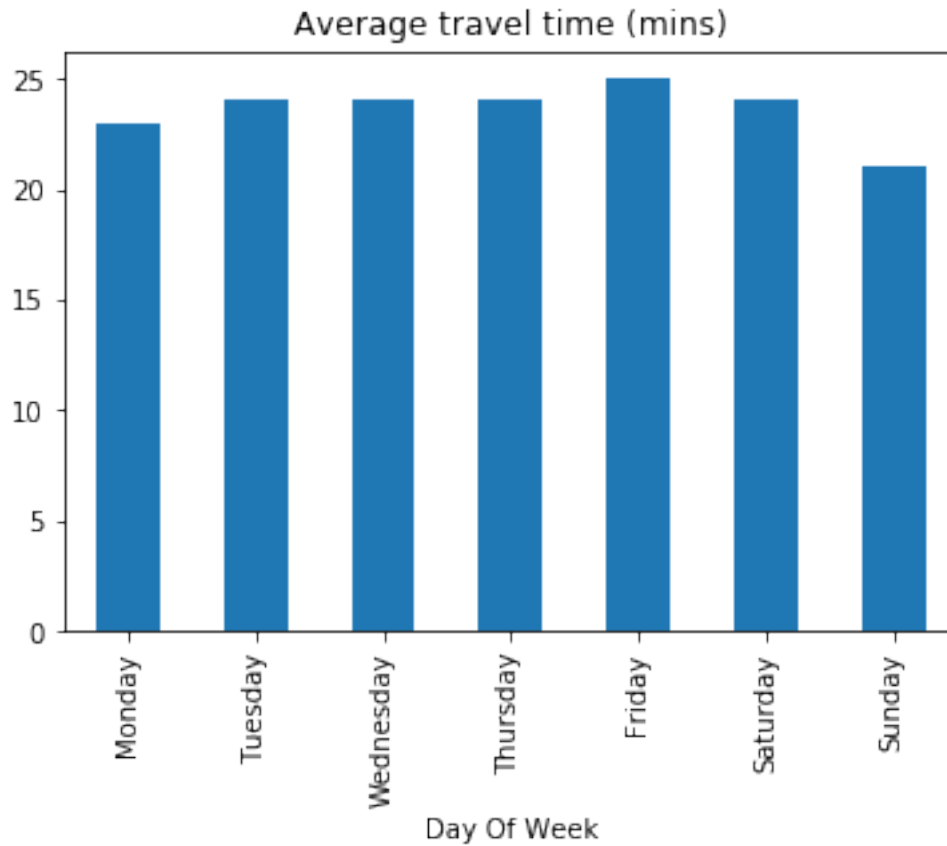
Figure 1: Example zones: 28th & Mission and 1400 Powell St

3.1 Day of Week

In support of the hypothesis, weekends are about a minute and a half quicker on average.



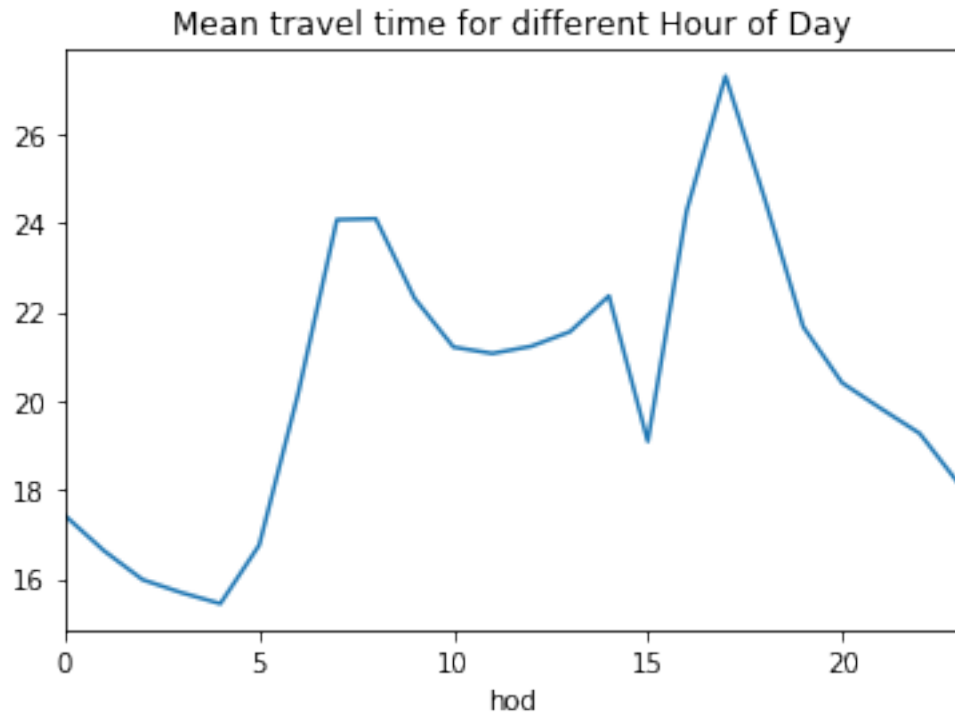
However, when we break it down by day, we see that Saturdays are actually suprisingly slow.



This implies that there are other dynamics at play here.

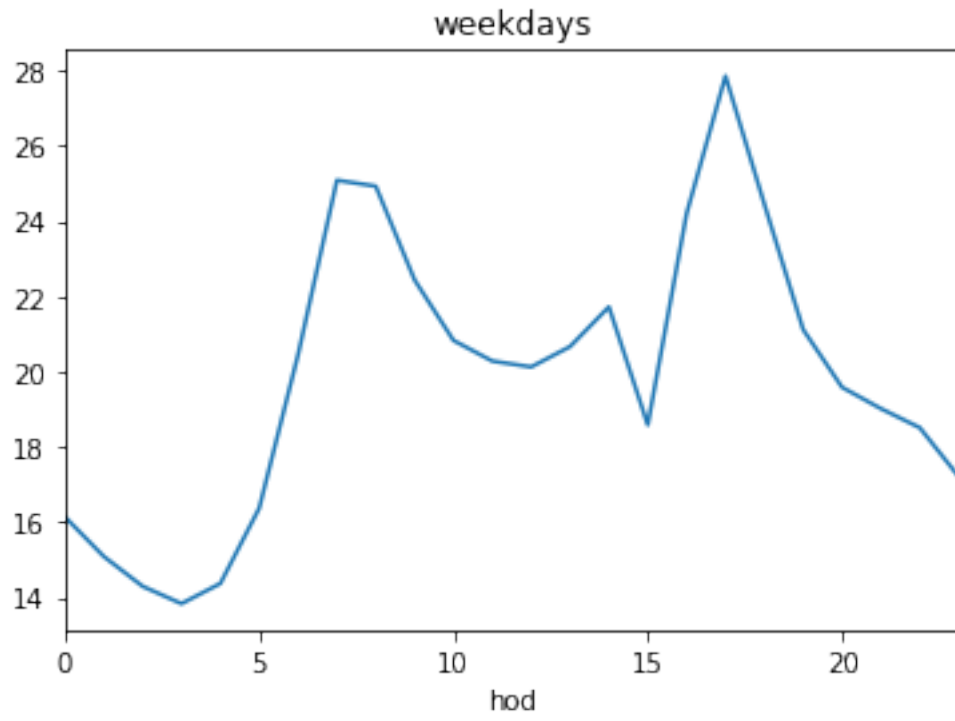
3.2 Hour of Day

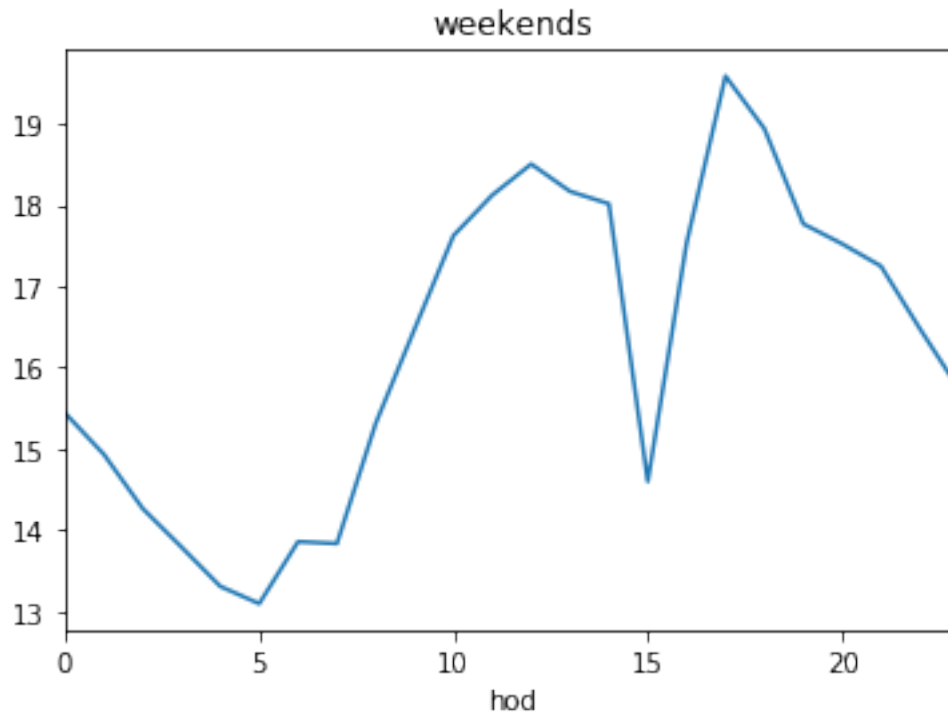
To further investigate whether work is contributing mostly to traffic, let's break down time into Hours of the Day. This data is already easily scaled up because there are files which consider all of San Francisco's 1760 census tracts (big 250MB file). However, one big limitation of this approach is that it gives all census tract pairs equal weight in averages, when some may actually have much higher volume!



It looks like work is a major contributor to mean travel time spikes, with highest spikes being 6-8a and 4-6p, although the evening spike is more pronounced.

Let's distinguish between weekdays and weekends.





In support of the hypothesis, weekends do lack the morning spike,. However, weekends have an evening spike. This is largely an artefact of a huge drop at 3p on weekends, the cause of which is unclear and warrants further investigation.

4 Avenues for further investigation

I hope to construct an index which represents commuter hour impact (e.g. weekday:weekend ratio) and compare it between two different cities. I hope to look at the special case of weekday holidays, to see if they are the same as weekends or if there is something special about them.