

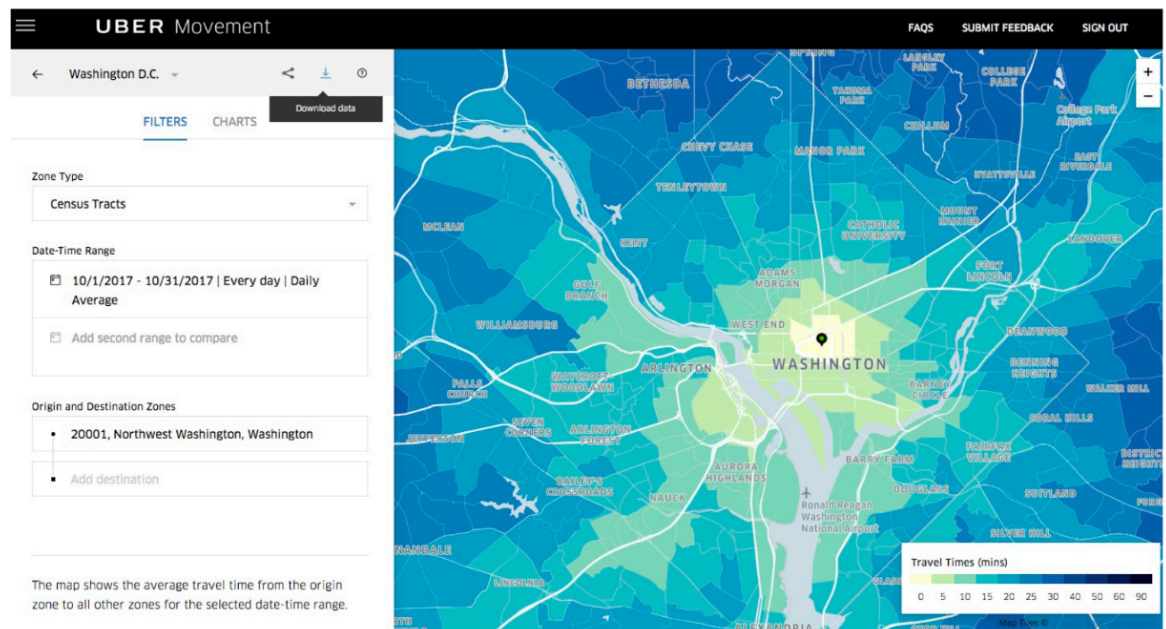
UML
Laurence (w4rner) Warner

Problem Set 1 Exploration & Computation

1. Obtain a dataset (preferably of substantive interest/domain expertise).

Uber Movement is a service providing anonymized data to the public on over 2 billion trips to help urban planners and the public better understand traffic patterns.

The following shows an example of the data that exists. It is semi-aggregated: it shows the average travel time between different city zones.

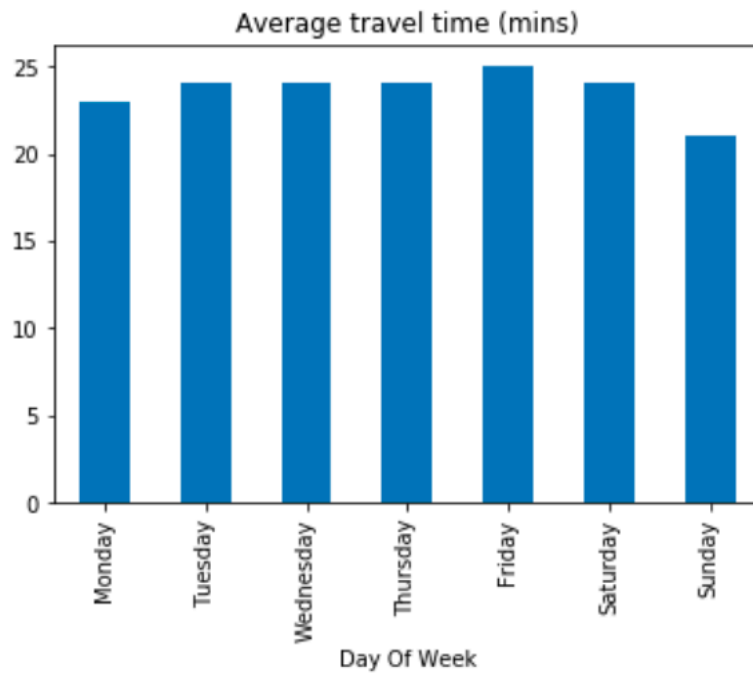


Let's look at the example of the links between two Census Tracts in this dataset and explore that: in San Francisco: 28th & Mission to 1400 Powell St.

2. Choose a visual technique to illustrate your data (e.g., barplot, histogram, scatterplot).

Barplot

3. Now generate and present the visualization and describe what you see.



We see that daily average travel time is pretty consistent, but lower on Sunday.

4. Calculate the common measures of central tendency and variation, and then display your results.

```
mean    23.571429
std      1.272418
min      21.000000
25%     23.500000
50%     24.000000
75%     24.000000
max     25.000000
```

5. Describe the numeric output in substantive terms, e.g., a. What do these numeric descriptions of data reveal?

Mean travel time is 23.5m, and the median is 24m.

- b. Why is this important?

It is useful to know the central tendency of the data specifically.

c. What might you infer about the distribution or spread of the data?

Standard deviation is low, which reveals that there is little variation.

d. Etc.

We can infer that day of week is not a major determinant of Uber travel time.