

3300 Project 2

Justin Kuang (zk65) , Ellin Hu (eh479) , Stephanie Mark (srm276)

*If the files don't work, project is also hosted here: <https://uber-nyc.herokuapp.com/index.html>*

# Income Patterns and Uber Rides in NYC

## The Data

- Uber Data from **FiveThirtyEight**: <https://github.com/fivethirtyeight/uber-tlc-foil-response>
- Income Data from **University of Michigan**:  
<https://www.psc.isr.umich.edu/dis/census/Features/tract2zip/>
- NYC Map from **NYC Open Data**:  
<https://data.cityofnewyork.us/Business/Zip-Code-Boundaries/i8iw-xf4u/data>

We specifically found 2 datasets that were of interest:

1. A sample of around 20 million Uber pickups (specific data attributes being location, date, and time) in 2014 and 2015, broken down by month (found from FiveThirtyEight, who got it from the NYC Taxi and Limousine Commission). Since the dataset contained an overwhelming number of actual data points, we decided (after a bit of trial and error) to focus on a single month- we picked April of 2014. We preprocessed this data using Excel, especially in separating the combined date/time field into two different columns- this allowed us to easier parse the data later. In our actual JavaScript implementation, we stored this Uber pickup data in separate objects that represent each day of the week (Mon-Sun), and furthermore each "day" object contained 24 different arrays that mapped to each hour of the day. By adding each data point to the correct corresponding day object/time array, we had easy  $O(1)$  access to the aggregated data separated by time (hour) and day of week.
2. Our second dataset contained data on the median and mean income of each zip code of the United States- this data was collected by the University of Michigan, and is based on the 2009 American Community Survey. We decided to use median numbers instead of mean because New York City contains a large number of high net worth individuals, which would impact the mean income more than the median.

Since we were looking to visually map the data, we also sourced a geojson file that divided NYC by federal zip codes. The file we found was sourced directly from the City of New York, and is in JSON format.

## Visual Elements

We mapped colors in the map of NYC to the income data from each zip code. The color scale uses `scaleLinear` to show areas with high income as having high amounts of green pigment, whereas areas with lower income have more white pigment.

Additionally, we graph Uber pick-ups during the selected day and time using orange dots. This uses a `geoAlbers` scale to map longitude and latitude to positions on the map. However, since we also implemented zooming and panning for the map (with boundaries within our box), we regenerate the map with a translating scale when there is mouse movement.

This data is also summarized on the graph to the right of the map. The line graph shows the average number of pickups per hour at different times of a selected day. The selected time is reflected as an orange dot on the plot, while the other dots are blue. This creates a focus on the specific number of pickups in a certain hour.

## The Story

Our visualization tells us that zipcodes with higher incomes tend to use Uber more, regardless of time or day. This is true despite the fact that there are also many public transportation options available in areas with higher incomes. We can additionally see the activity levels of people at different hours and how that varies by day. For example, the plots show that the Uber pickup patterns of people are very similar across weekdays. On the weekend, however, people are much more active on Saturday nights (and Sunday mornings), but tend to call fewer Ubers during the day on Sundays. It was surprising because we expected Uber pick ups to be more evenly distributed across the city, especially since Uber is often marketed as an affordable alternative to taxis and other methods of transport.