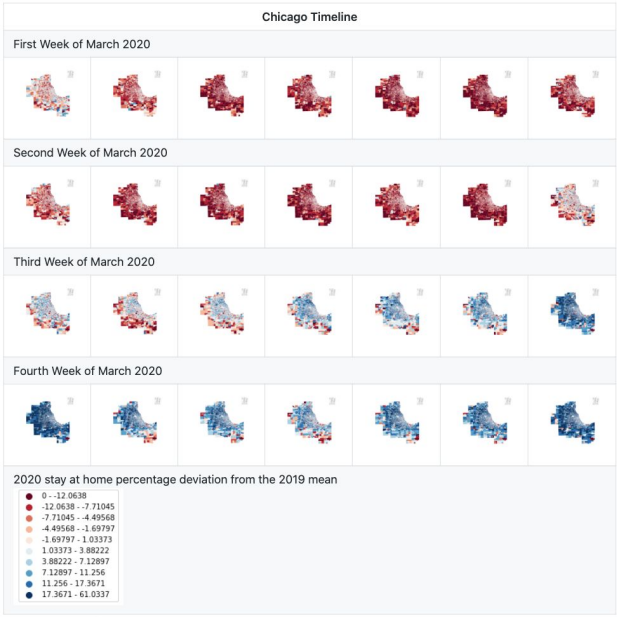
Visualizing the Mobility Gradient

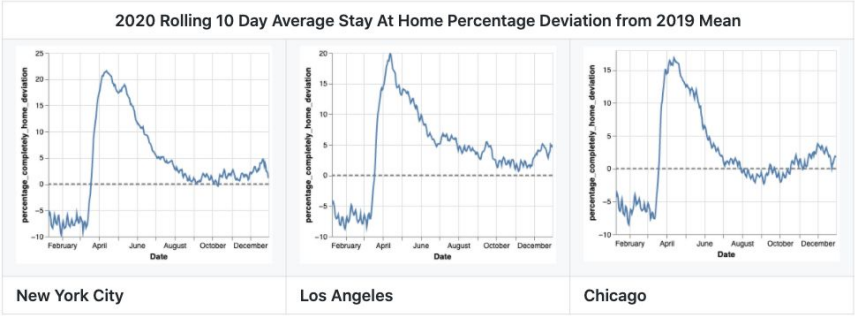
Over Time
and in Relation to Poverty

Justin Snider (js10853)
Anchit Srivastava (as14022)
Diksha Chouhan (dc4454)

https://github.com/chouhandiksha/bigdataproject

Conclusions

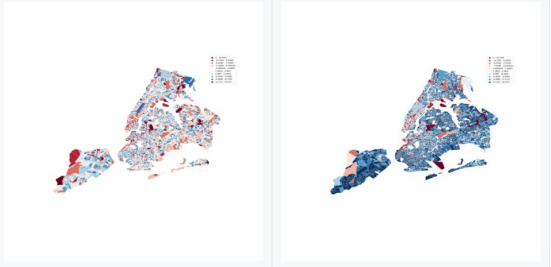




Conclusions // Transformation in Daily Fully at Home Behavior // Similarities

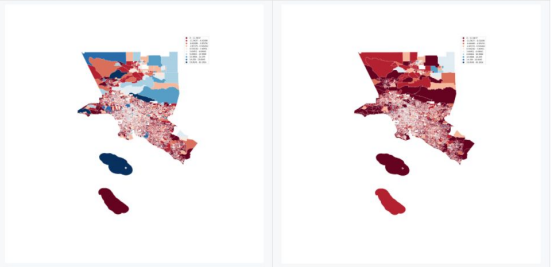
New York City // Mondays in March, 2020

2020-03-02 | 2020-03-09

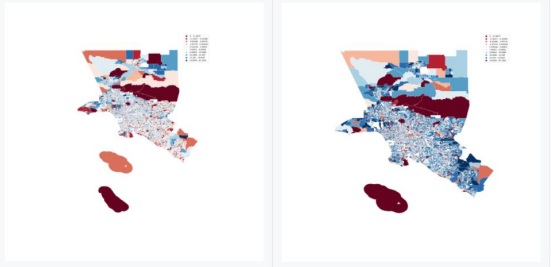


2020-03-16 | 2020-03-23

Los Angeles Metro Area // Mondays in March, 2020



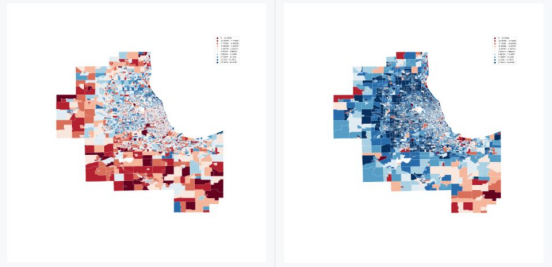
2020-03-02 | 2020-03-09



2020-03-16 | 2020-03-23

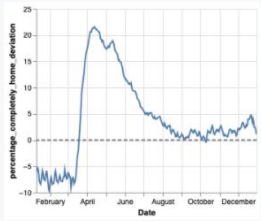
Chicago // Mondays in March, 2020

2020-03-02 | 2020-03-09

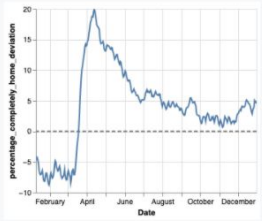


2020-03-16 | 2020-03-23

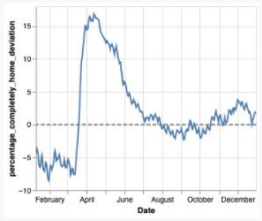
2020 Rolling 10 Day Average Stay At Home Percentage Deviation from 2019 Mean



New York City

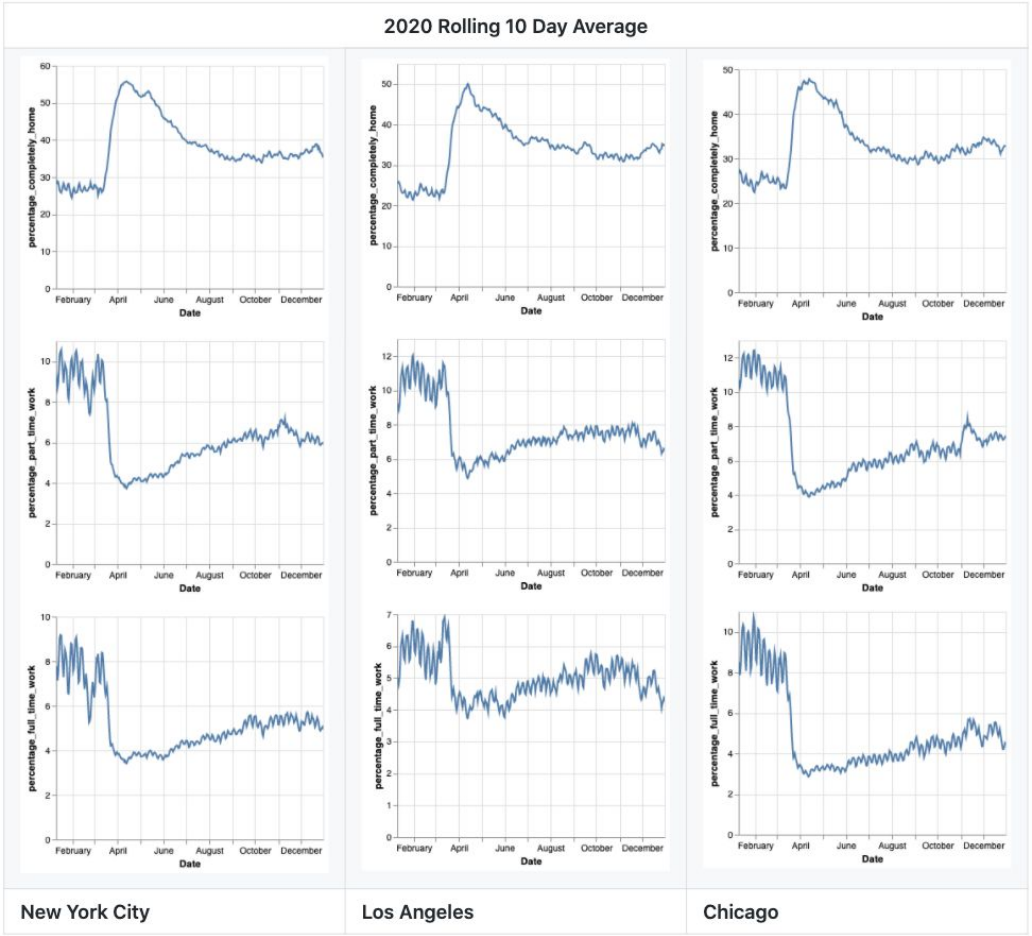


Los Angeles

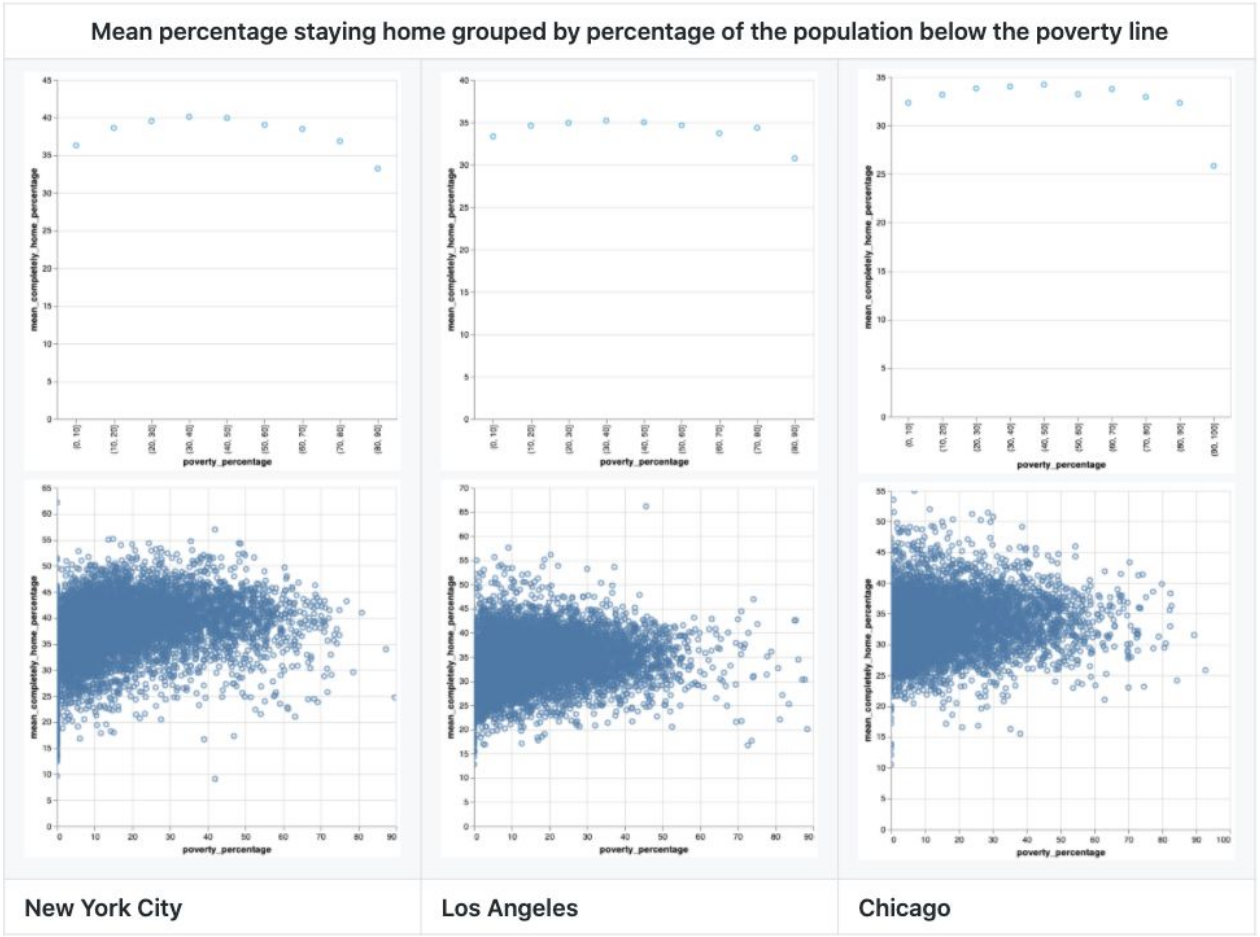


Chicago

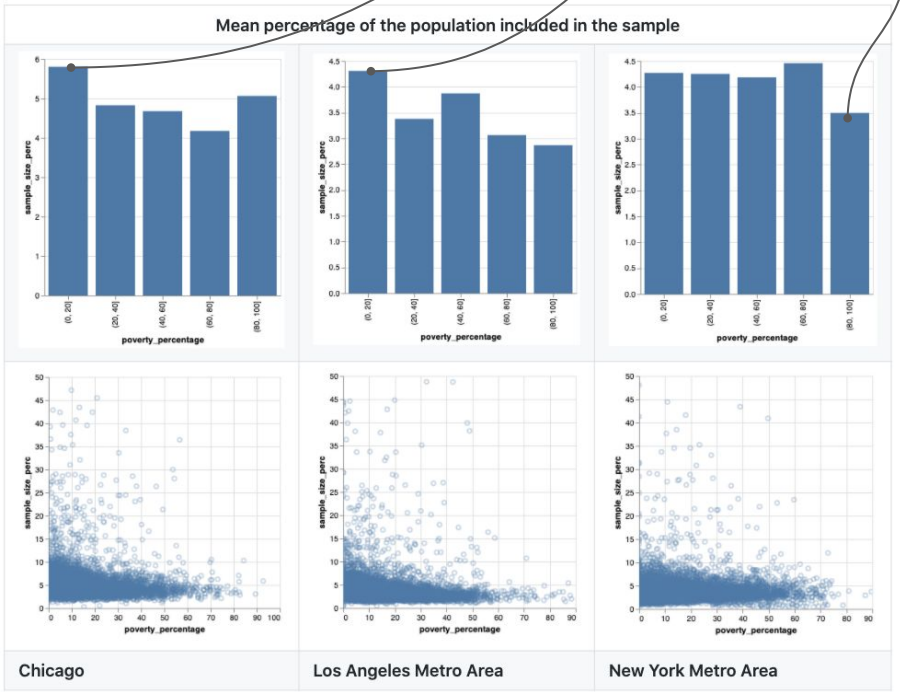
Conclusions // Transformation in Daily Fully at Home Behavior // Similarities



Conclusions // Change in Mean Fully at Home Percentage // Comparing Differences



Conclusions // Comparing Areas with Different Levels of Poverty

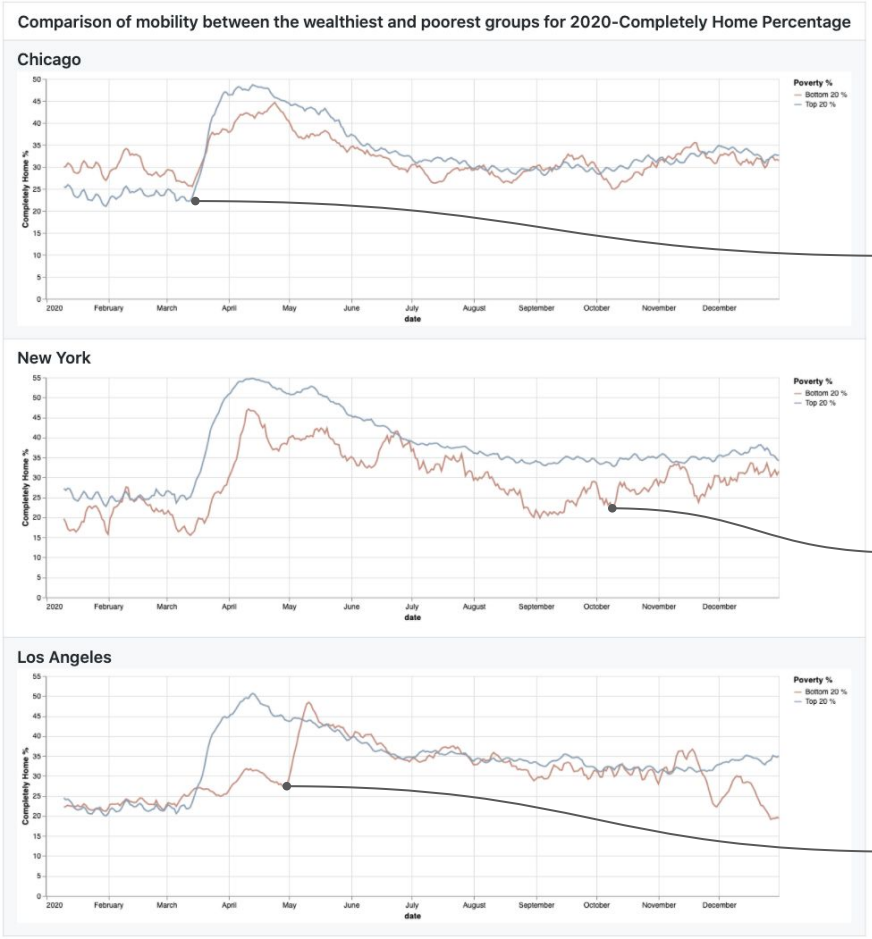


The areas with low poverty have the highest percentage of the population included in the SafeGraph Social Distancing data.

The areas with highest poverty have the lowest percentage of the population included in the SafeGraph Social Distancing data.

In general the data includes on average 2.5% to 6% of the population. So we should keep in mind those without phones and not included will not be represented in this dataset.

You can see specific census block groups have higher and lower representation in the data.

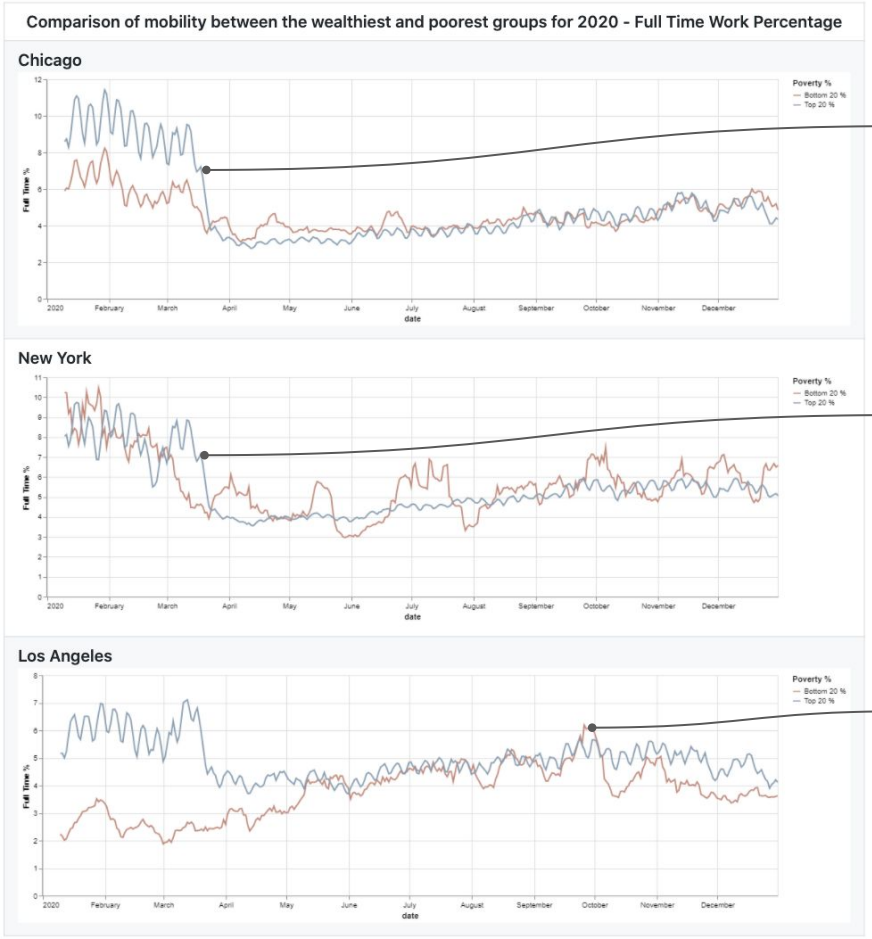


Both areas with high and low poverty respond quickly and stay within about 5% for 2020

The areas with high poverty have 5% to 15% lower fully-at-home percentage sustained over most of 2020.

The areas with high poverty take one and a half months to respond to the crisis. This is a few weeks after the first round of stimulus checks.

Conclusions // Daily Completely Home Percentage in Areas with High and Low Poverty

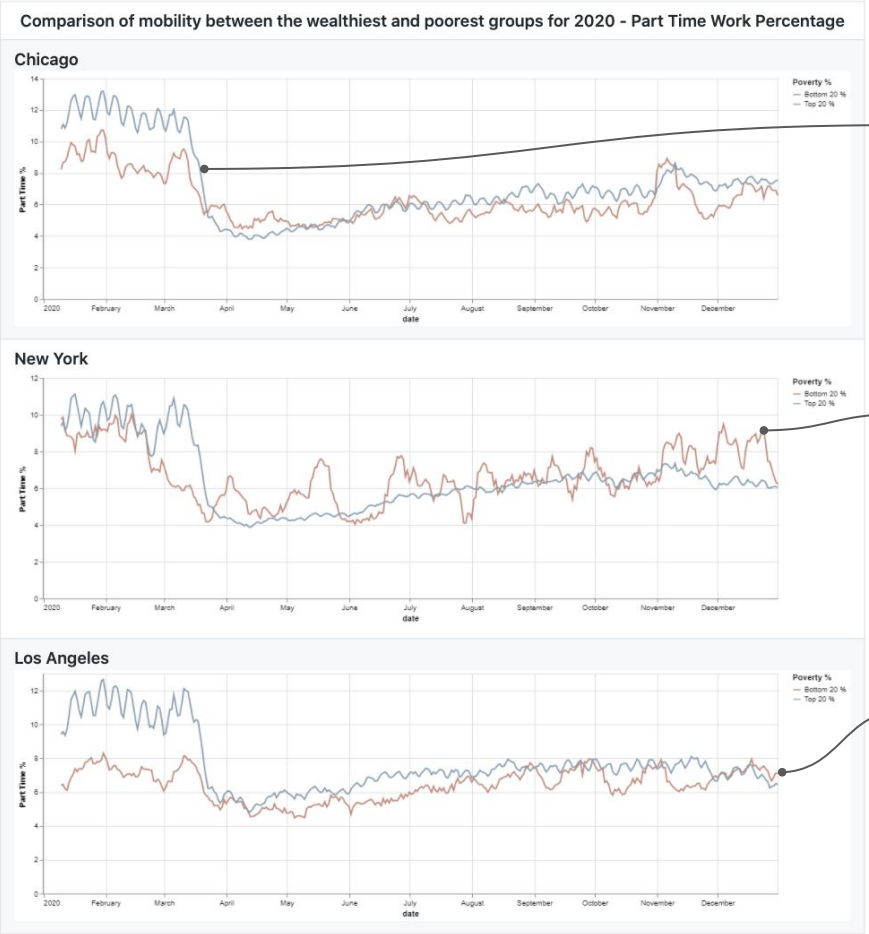


Both areas with high and low poverty respond quickly and maintain similar full-time work behavior for 2020. A possible contributing factor for the lack of disparity is the average response in Chicago has a low variance.

While there is a decrease from the starting value the areas with high poverty have surges of increased full-time work starting in mid-March and continuing throughout 2020, while areas with low poverty maintain relatively stable and low values.

The areas with high poverty increase full-time work behavior through most of 2020, while all other groups and cities decrease full-time work below January and February levels.

Conclusions // Daily Full-Time Work Percentage in Areas with High and Low Poverty



Both areas with high and low poverty respond quickly and maintain similar part-time work behavior for 2020. A possible contributing factor for the lack of disparity is the average response in Chicago has a low variance.

While there is a decrease from the starting value the areas with high poverty have surges of increased part-time work starting in mid-March and continuing throughout 2020, while areas with low poverty maintain relatively stable and low values.

The areas with high poverty show relatively little change in part-time work behavior throughout 2020.

Conclusions // Daily Part-Time Work Percentage in Areas with High and Low Poverty

Challenges

- Finding clear and useful information in such a large dataset was a substantial challenge.
- We repeatedly hit the Google Drive 1,000,000,000 file read limit while doing analysis. As a solution we created multiple clones of the Google Drive account to spread out usage.
- The SafeGraph Social Distancing Metrics data set is 81 GB and took 6 hours to download. The download sometimes fails due to the long download time. Download requires the use of the AWS command line interface.
- The 731 days of data for each the 11,078,297 census block groups in the United States was challenging to process. All together there are approximately 8,098,235,107 rows of data. Using Spark and filtering out just the census blocks for the three cities made analysis much more efficient.
- The American Census Survey data was not intuitive to understand. There data attributes do not have intuitive names and they provide population counts. We had to extract just the useful columns and transform them into percentages to make the data useful.