

Visualizing the Mobility Gradient

Over Time
and in Relation to Poverty

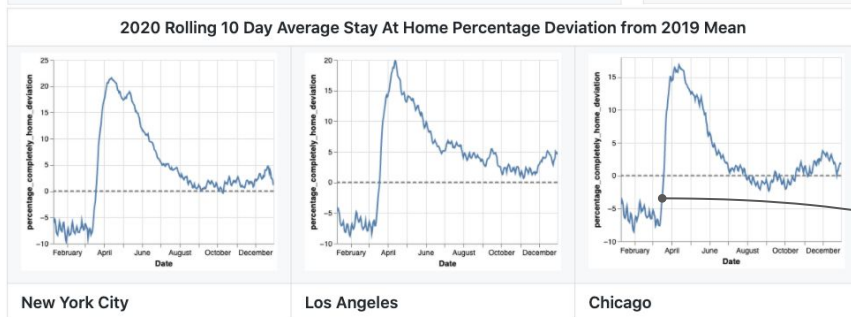
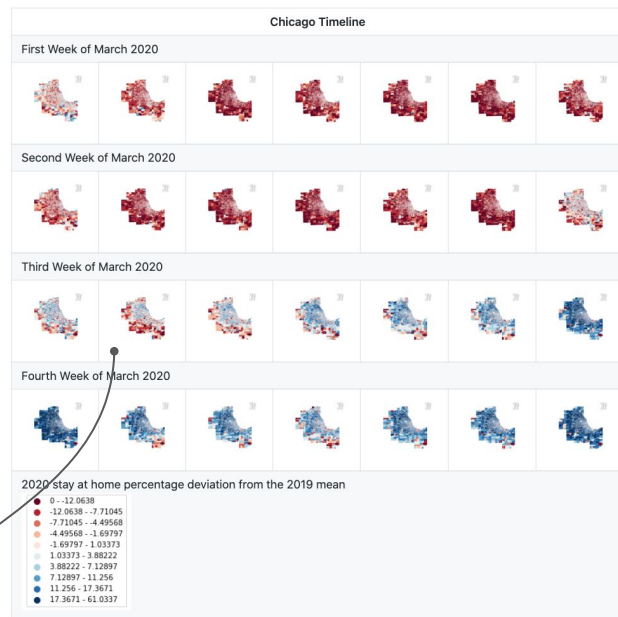
Justin Snider (js10853)

Anchit Srivastava (as14022)

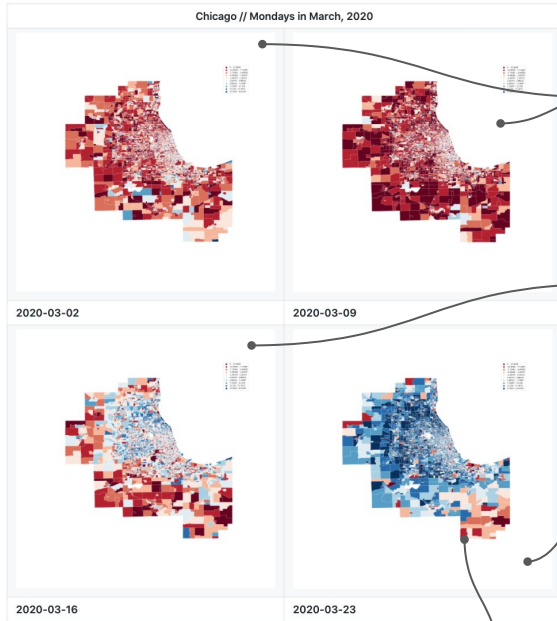
Diksha Chouhan (dc4454)

<https://github.com/chouhandiksha/bigdataportfolio>

Urban Mobility Gradient



In all three cities March 16 was the major turning point as the first weekday the at home percentage begins to increase

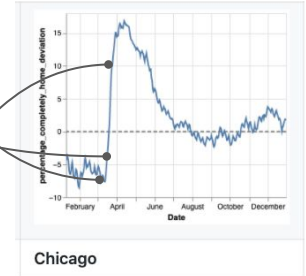


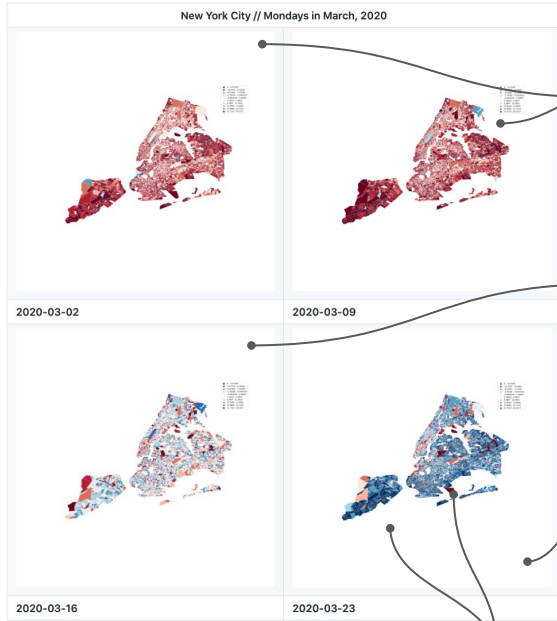
Typical pre-increase Mondays

Initial at home percentage increase

Near peak at home percentage

Some low density and high poverty areas maintain low stay at home percentages



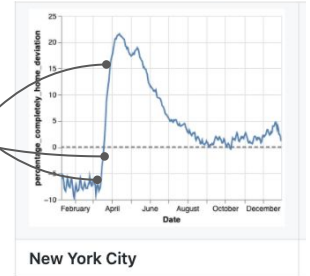


Typical pre-increase Mondays

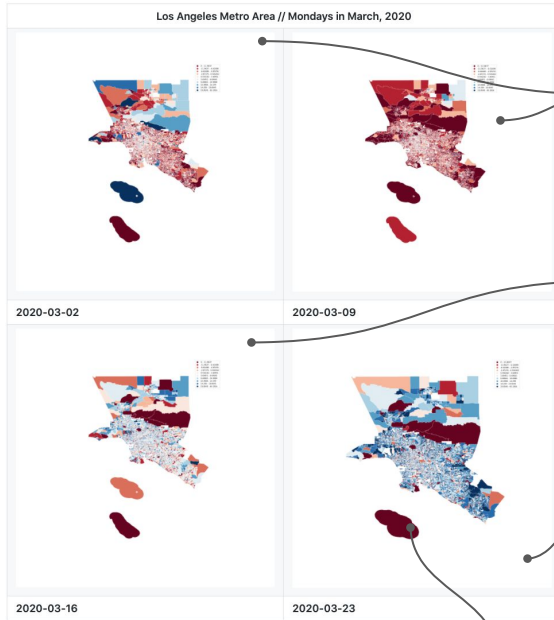
Initial at home percentage increase

Near peak at home percentage

Some low density and high poverty areas maintain low stay at home percentages



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

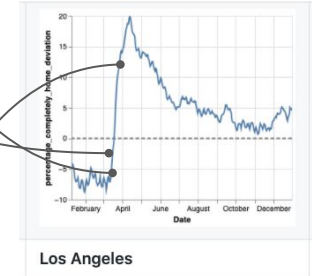


Typical pre-increase Mondays

Initial at home percentage increase

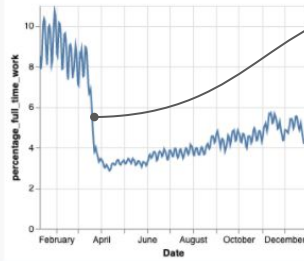
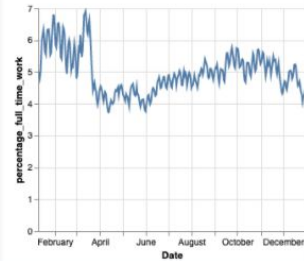
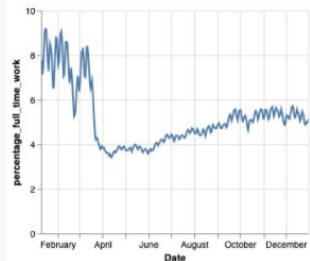
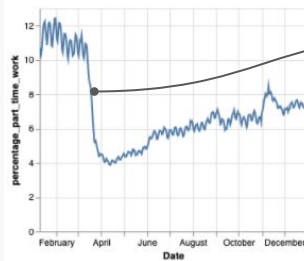
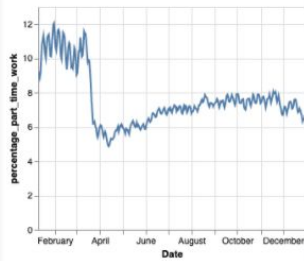
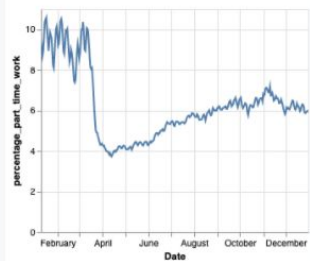
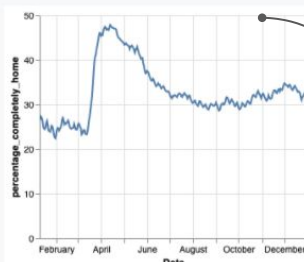
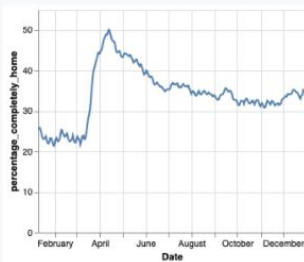
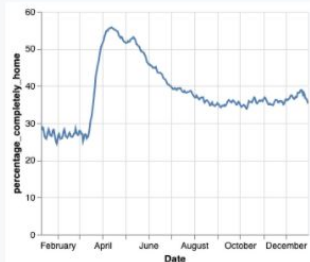
Near peak at home percentage

Some low density and high poverty areas maintain low stay at home percentages



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

2020 Rolling 10 Day Average



New York City

Los Angeles

Chicago

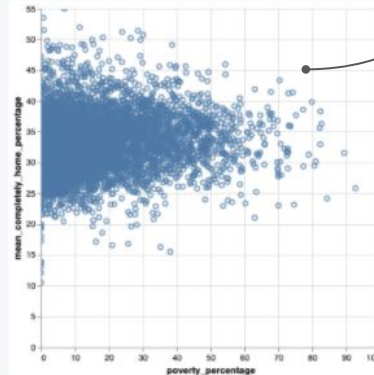
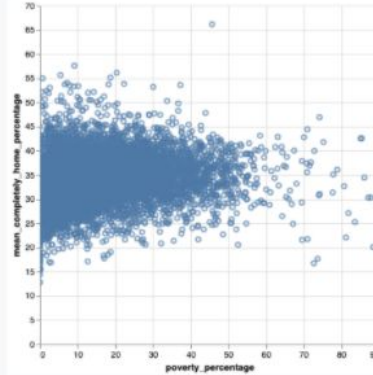
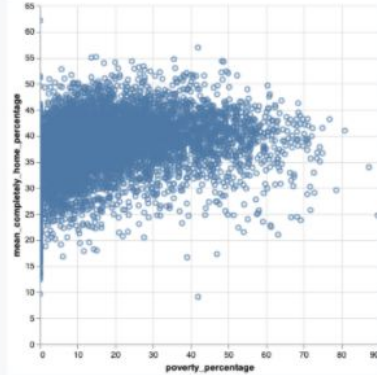
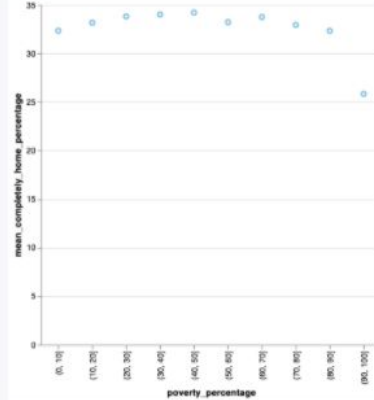
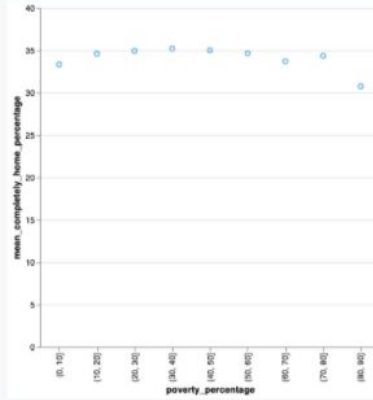
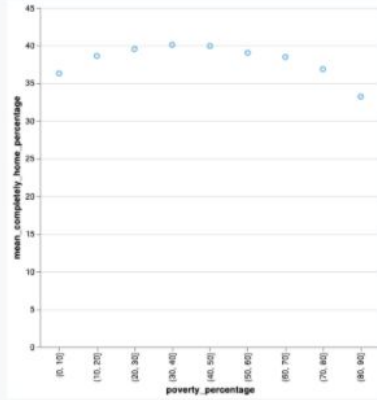
Chicago peaks under 50%
Los Angeles peaks at 50%
New York peaks above 50%

All three cities shift by about -6% in March
Afterwards NY and LA drift up 2%
Chicago drifts up over 3%

All three cities shift by about -4% in March
Afterwards NY and Chicago drift up 1%
Los Angeles drifts up and back down 1%

Poverty and Mobility

Mean percentage staying home grouped by percentage of the population below the poverty line



New York City

Los Angeles

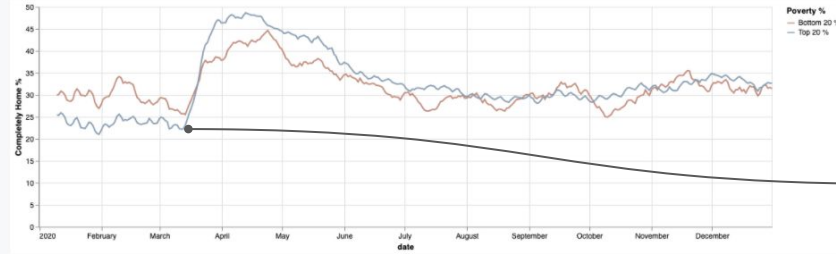
Chicago

On average, in all three cities areas with the highest poverty also have the lowest at home percentage

In all three cities there is a clear decline in the maximum at home percentage value as the poverty level increases

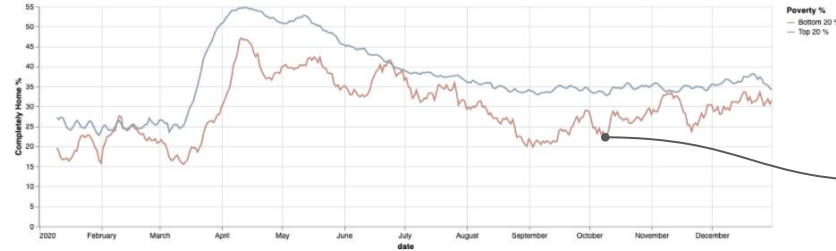
Comparison of mobility between the wealthiest and poorest groups for 2020-Completely Home Percentage

Chicago



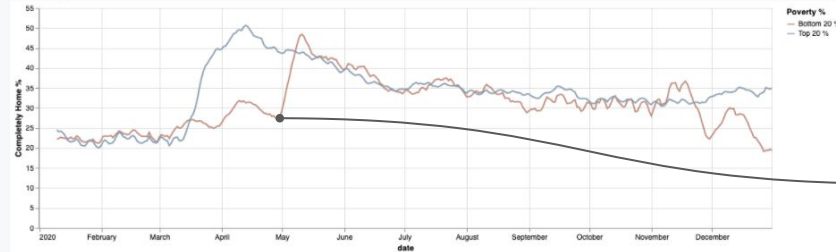
- Both areas with high and low poverty respond quickly and stay within about 5% for 2020

New York



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

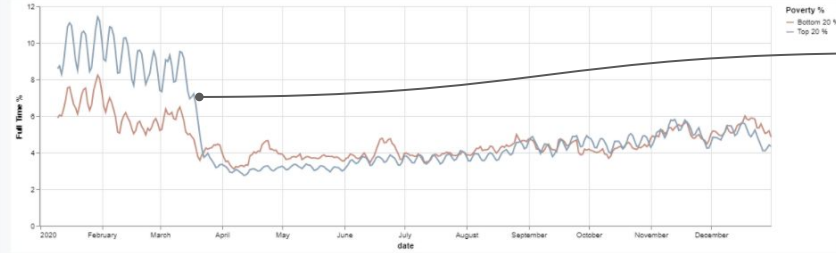
Los Angeles



- 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.

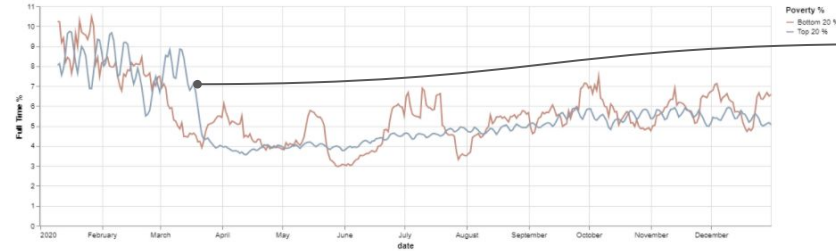
Comparison of mobility between the wealthiest and poorest groups for 2020 - Full Time Work Percentage

Chicago



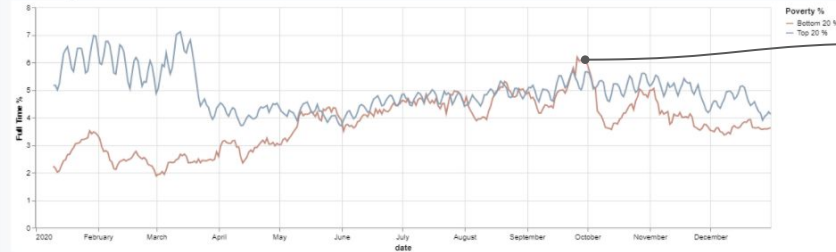
Both areas with high and low poverty respond quickly and maintain similar full-time work behavior for 2020. A possible contributing factor for the lack of disparity is the average response in Chicago has a low variance.

New York



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.

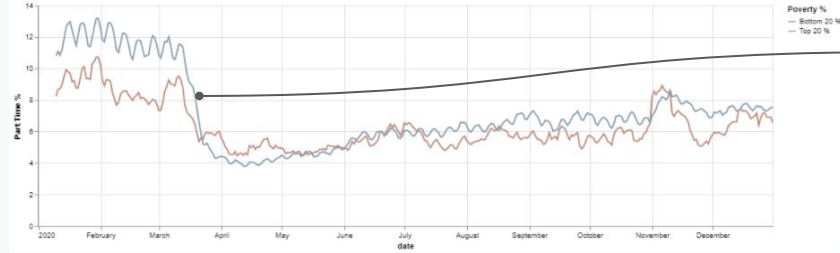
Los Angeles



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.

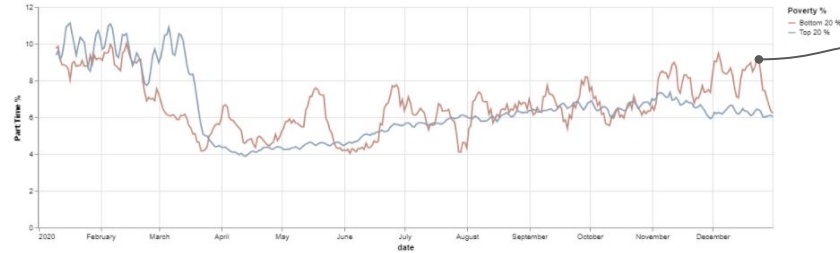
Comparison of mobility between the wealthiest and poorest groups for 2020 - Part Time Work Percentage

Chicago



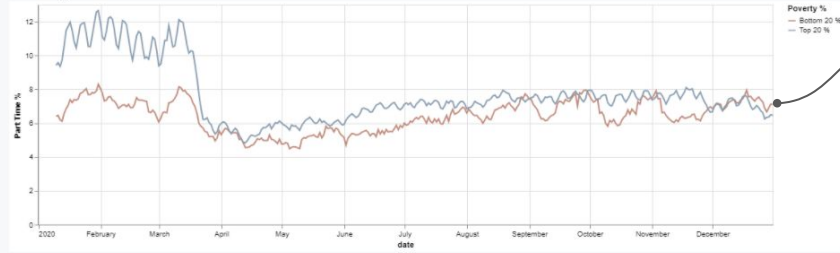
Both areas with high and low poverty respond quickly and maintain similar part-time work behavior for 2020. A possible contributing factor for the lack of disparity is the average response in Chicago has a low variance.

New York

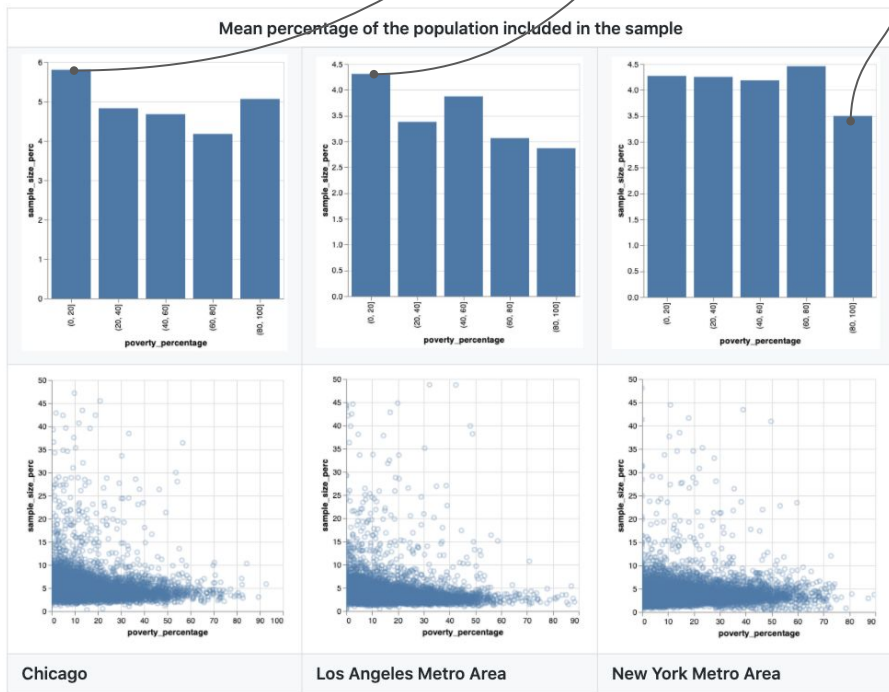


While there is a decrease from the starting value the areas with high poverty have surges of increased part-time work starting in mid-March and continuing throughout 2020, while areas with low poverty maintain relatively stable and low values.

Los Angeles



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



• 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.

Conclusion

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.

Key Findings

Mobility

- Chicago show the lowest overall change in mobility.
- New York showed the highest change in mobility.

Poverty

- New York has the most sustained disparity between areas with high and low poverty.
- Areas with high poverty in Los Angeles have the most delayed response to taking a month and a half to increase their stay at home percentage.
- Chicago had the least disparity at least in part because the wealthy show relatively low levels of change in mobility compared to NY and LA.

Sampling Bias

- A sampling bias exists in the SafeGraph Social Distancing data. The percentage of the population included in the SafeGraph dataset is higher in areas with less poverty in all three cities.