

Visualizing the Mobility Gradient Over Time and in Relation to Poverty

Justin Snider
New York University
New York, USA
js10853@nyu.edu

Anchit Srivastava
New York University
New York, USA
as14022@nyu.edu

Diksha Chouhan
New York University
New York, USA
dc4454@nyu.edu

ABSTRACT

We compared the mobility patterns of people in Chicago, New York City and Los Angeles metropolitan area in 2020 as they respond to the COVID-19 pandemic. Then, within each city, we compared the mobility patterns of census block groups with different levels of poverty. Using the SafeGraph Social Distancing Data mobile phone metadata, generated from mobile phone records, we were able to observe daily behaviors such as full-time work, part-time work and full-time stay at home behavior. The daily statistics are provided for each Census Block Group (CBG). The data we use to find the population size of each CBG, the number of people in poverty in each CBG and calculate the percentage of the population in poverty is the American Community Survey (2016) 5-year estimate available from SafeGraph. We compare the mobility of Census Block Groups with high and low percentages of individuals in poverty. We visualize and explain who has been able to effectively decrease their mobility. Understanding who has not been able to decrease their mobility can help us understand who is most at risk and who is most in need of assistance to reduce their exposure to COVID-19 [2].

CCS CONCEPTS

- Human-centered computing → Visualization.

KEYWORDS

COVID-19, mobility

1 INTRODUCTION

Many recent papers, including "Mobility network models of COVID-19 explain inequities and inform reopening" [1] and "Coronavirus infections and deaths by poverty status: The effects of social distancing" [3] have established an essential link between mobility and COVID-19 infection rates and deaths. We study the daily change in mobility of the general population during 2019 and 2020 in the three most populous cities in the US: New York, Los Angeles and Chicago. Using the SafeGraph Social Distancing Metrics data[10], we analyze mobility at the resolution of the census block group level. We use attributes such as the number of devices that remain

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

..
© 2021 Association for Computing Machinery.

at home, the count of devices with full-time work behavior, the median distance traveled by devices and other vital actions relevant to mobility. We then visualize each city's mobility using graphs and maps produced with the GeoJSON file provided by SafeGraph of the Census Blocks.

We compare the mobility of census blocks with the largest and smallest percentage of people in poverty based on the Census American Survey data provided by SafeGraph[9]. Our objective is to better understand and visualize to what degree different groups have adopted limited mobility to mitigate the documented threat of COVID-19 exposure created by higher mobility.

All code used to extract, clean, integrate and visualize the data is available on the project Github repository <https://github.com/chouhandiksha/bigdataproject>. All the steps used for acquiring, extracting, integrating and cleaning the data used for the project are described in sections 3 through 6.

2 RELATED WORK

In the *Nature* article "Mobility network models of COVID-19 explain inequities and inform reopening" [1] Chang et al. produce a model that predicts the number of anticipated SARS-CoV-2 infections. The model takes as an input mobility data derived from mobile phone data. The model tracks the predicted number of susceptible, exposed, infectious and removed (SEIR) people anticipated to be in the population. The model is able to predict real case trajectories on held out data. The model correctly predicts higher infection rates among disadvantaged racial and socioeconomic groups solely as the result of differences in mobility.

Several months have passed since the publication of this article. All the 2020 data is now available from SafeGraph. There is now an opportunity to evaluate visualize the relationships between mobility, income, race and the impact of COVID-19 in our communities.[8]

"Coronavirus infections and deaths by poverty status: The effects of social distancing" [3] have shown the highest initial number of cases are in both the wealthiest and poorest countries. However, there is a great difference between the wealthy and poor countries when stay at home policies are put in place. The wealthiest countries have been able to show much greater reductions in mobility. As a result the wealthiest countries are able to curb the infection rates more effectively than poorer countries.

3 DATA SOURCES

We use the SafeGraph COVID-19 Response Social Distancing Metrics [10], the SafeGraph Demographic American Survey Census Data[9] and The New York Times COVID-19 Data. [12] We have

verified all three datasets are available to researchers, secured access to and downloaded the datasets.

The SafeGraph COVID-19 Response Social Distancing Metrics [10] schema can be viewed online at <https://docs.safegraph.com/docs>. The dataset is organized by Census Block and includes mobility statistics generated from cell phone data such as median distance travel from home, completely at-home device count, full-time work behavior and many more mobility aggregate statistics.

The schema for the SafeGraph Census Block Group Data [9] can be viewed online at <https://docs.safegraph.com/docs/open-census-data>. The dataset is organized by Census Block and contains demographic information, including statistics about race, household income and poverty.

The New York Times COVID-19 Data [12] can be accessed online at <https://github.com/nytimes/covid-19-data>. The dataset contains reported daily COVID cases and deaths organized at the Country, State and county levels.

4 DATA ACQUISITION

4.1 New York Times COVID-19 Data

The New York Times COVID-19 data can easily be downloaded directly from the Github repository at <https://github.com/nytimes/covid-19-data>. We are using the us-counties.csv file for county level daily case and death counts. The CSV file was about 46 MB, a manageable size.

4.2 SafeGraph Social Distancing Metrics and Census Block Group Data

The SafeGraph Social Distancing metrics include daily aggregate data for each CBG for all of 2019, 2020 and the beginning of 2021.

The SafeGraph Census Block Group Data include the American Community Survey (2016) 5-year estimate on the Census Block Group level.

You can find comprehensive instructions on how to download both the SafeGraph Social Distancing Metrics and Census Block Group Data in the code repository for the project: <https://github.com/chouhandiksha/bigdataproject>.

5 DATA EXTRACTION

The FIPS code is a set of integers used to uniquely identify a specific geographic area. Each Census Block Group (CBG) is represented by a 12-digit FIPS code. The first two digits of a FIPS code represents the state. The next three digits represent the county. Then, six digits give the Census Tract. Our final digit is the block group. Knowing the FIPS codes for the counties in a given metropolitan area allows us to extract the rows we need and leave the other behind.

5.1 New York Times COVID-19 Data

The New York Times COVID-19 file us-counties.csv has a FIPS column. After loading the file into a dataframe using Pandas we filter out just the rows with FIPS matching the counties in the metro areas. The result allowed us to get all the relevant row for the counties of Los Angeles and Chicago.

Filtering for the New York FIPS counties returned no rows. Upon investigation we discovered ironically the New York Times does

not maintain county level information for the counties of New York City. They just group all 5 counties of New York City together. Filtering for New York City in the county name column we were able to secure the city level cases and deaths for New York City.

5.2 SafeGraph Census Block Group Data

SafeGraph provides a download of the 2016 American Community Survey (ACS) 5-year estimate on the Census Block Group level. We were able to filter out just the counties of the New York, Los Angeles and Chicago Metro areas from the dataset one CSV file at a time using the FIPS code. We simply filter out CBGs where the first five digits of the CBG FIPS code matches a county from one of our 3 metro areas. Each row includes a Census Block Group that the row attributes describe. There are so many attributes in the ACS that the data is broken up into separate CSV files. The name of the file tells you what attributes are included in the file. Inside each file there is one row for each CBG.

We have extracted the following columns for all CBG in the three metro areas:

- **Column B01003e1:** the total estimate population
- **Column B02001e2** the white only estimated population
- **Column C17002e1** the total population for whom poverty status is determined
- **Column C17002e2** the population for whom the ratio of income to poverty level in the past 12 months is under 0.5 from the population for whom poverty status is determined
- **Column C17002e3** the population for whom the ratio of income to poverty level in the past 12 months is between 0.5 to 0.99 inclusive from the population for whom poverty status is determined

To better make comparisons between CBGs we calculated the white only and poverty percentages for each CBG from the given population values. The CBG white only percentage is calculated by $\frac{w}{t} \cdot 100$ where w is the white only population and t is the total population. The CBG percentage in poverty is calculated by $(a + b)/d * 100$ where $a + b$ is the total population with income below the poverty level and d is the total population for whom the poverty status is determined.

5.3 SafeGraph Social Distancing Metrics

The SafeGraph Social Distancing Metrics data is downloaded from AWS S3 storage to a local directory. The file structure provides the labels for the many separate gzip csv files. The folder structure pattern is: **YYYY/MM/DD.gz**. For instance, we have 12 folders representing each month i.e. from 01 to 12 in the “2019” folder, and in each of the folders a zipped file representing each day of the month were present. This was massive data that took us a significant amount of time to download.

There is one file for each day of 2019 and 2020. There are also daily files for the first few months of 2021. The data set is still being updated every day with about a one week lag time between when data is captured and when it is added to the data set.

We looped through all the files in the data set reading them into memory and keeping just the CBG rows from the 3 metro areas using the FIPS code column. For convenience once selected we save the filtered rows for each day to a csv file. Our new files follow a

different format. The name of the file gives the year, month and date by using the format **YYY-MM-DD-social-distancing.csv** and are all stored in a single directory for the entire year. For analysis we just read any needed daily data into memory by looping through the new filtered data files.

In addition to filtering the metro area rows, we have calculated several critical percentage values: the percentage of devices exhibiting part-time work behavior, percentage of devices exhibiting full-time work behavior and percentage of devices exhibiting completely home status. Each of these percentages was taken using $b/d * 100$ where b is the population exhibiting the behavior in the CBG and d is the total number of devices in the CBG.

Finally, for each of these three percentages we have also calculated the normalized value by using $(v - \mu)/\sigma$ where v is the value to normalize, μ is the mean and σ is the standard deviation.

6 DATA CLEANING

6.1 New York Times COVID-19 Data

In order to clean the data we have used Spark, Python, Pandas and OpenRefine to take the following steps:

- Verify the dates fall between January 24th, 2020 and March 18th, 2021 using Pandas min and max function.
- We have verified the extracted rows only include the desired rows from the 3 metro areas using Pandas.
- We have verified all unique date and CBG combos have only one row by comparing the total number of rows to the number of unique combos. We found the count matches for all extracted data.
- Verify the cases and deaths are integer values as expected using the Pandas min and max function.
- We have verified there are no empty, Null or NaN values in the extracted data we are using.

6.2 SafeGraph Census Block Group Data

In order to clean the ACS data we have used Python, Pandas and OpenRefine to take the following steps:

- We have checked that each CBG is unique and is not duplicated.
- We have checked to make sure all percentages fall within the valid range from zero to a hundred using the Pandas min and max function.
- We have produced histograms for each attribute to verify the shape of the data and range of values are valid.
- We have visualized the CBG values on maps to give an intuition into the contents of the data and look for outliers.
- Using Pandas ‘isna’ function we verified there are no empty, NaN, or Null entries in any of the Social Distancing CSV files.

6.3 SafeGraph Social Distancing Metrics

In order to clean Social Distancing data we have used Python, Pandas and OpenRefine to take the following steps:

- OpenRefine we have checked for duplicate CBG on a subset of the hundreds of csv files. We found no duplicates in any files we checked.

- Using Pandas we checked the minimum and maximum date values to ensure all the dates fall within the expected time frame. We found no values outside the expected time frame for all Social Distancing CSV files.
- Using the Pandas max and min function we checked that all the percentage values fall within the expected range for all percentage columns. We found no values outside the expected range for all Social Distancing CSV files.
- Using the Pandas max and min function we checked that all integer data types fall within the expected range, which includes no negative values. We found no values outside the expected range for all Social Distancing CSV files.
- Using Pandas ‘isna’ function we verified there are no empty, NaN, or Null entries in any of the Social Distancing CSV files.
- For each city we visualized the mean daily and mean monthly values for all three percentage values
percentage_completely_home,
percentage_part_time_work and
percentage_full_time_work in both 2019 and 2020. All three cities show an increase in completely home individuals during 2020. There is also a visible decrease in part-time and full-time work behaviors in 2020. Given the documented decrease in mobility during 2020 these are the trends we expect to find.

7 FINDINGS

During our study includes identical analysis on the cities of New York, Los Angeles and Chicago. Starting in March of 2020 much of the population in Chicago, New York City and Los Angeles metropolitan area drastically changed their mobility patterns. (See figures 1 and 2.) Using the SafeGraph Social Distancing data we visualize the scale of the change in population mobility through analysis of the amount of people staying at home, leaving home for full-time work and leaving home for part-time work. We compare the ability of the three cities to adopt new mobility patterns revealing both similarities and differences.

Comparisons are also made between the mobility patterns of census block groups with the highest and lowest levels of poverty in all three cities. We find areas with low levels of poverty reduce their mobility much more than those areas with high poverty. Specifically vulnerable populations in census block groups where 80% to 100% of the population are living in poverty are less able to reduce their mobility by staying at home and working from home. (See figures 4, 5 and 6) The majority of the population exhibit a rapid change in behavior to stay at home in March and April of 2020. (See figure 1) However, for the poorest census block groups the response tends to be slower, less drastic, less lasting, or all three.

One of the most notable findings we observed was New York stands out from Chicago and Los Angeles at all stages of analysis. The disparity between the richest and poorest groups is seen to be the most prolonged and continuous in New York [6]. It is also clearly visible that the mobility between the two groups are distinct from each other throughout the year. As soon as the stay at home orders are imposed, the two lines that show our mobility parameters including completely home, full-time and part-time work begin to

separate. For the entire 2020, we see that this disparity continues. This finding is also presented and talked about in a studies published in the International Journal of Population Research. [7]

7.1 Stay at Home Behavior

In all three cities the fully at home percentage of the population starts below the 2019 mean. However, in mid-march the percentage increases drastically with the trend continuing into April where all three cities peak. (See figures 1 and 2.)

The stay at home percentage reaches an extreme peak during April. New York has the highest peak at about 55 percent of individuals staying at home. Los Angeles just barely touches 50 percent and Chicago peaks below 50 percent. (See figure 1.)

We get an even clearer picture of the change of behavior when we look at the 2020 stay at home percentage deviation from the 2019 mean. We can see both New York and LA were able to change their behavior to a much greater extent than Chicago. Specifically, New York and LA increase their stay at home percentage by about 25% from mid-March to mid-April. Chicago on the other hand exhibits an increase close to 20% and is not able to keep the stay at home percentage above the 2019 mean through the summer. New York and LA maintain slightly elevated numbers. (See figure 2.)

There are also consistent trends when it comes to the relationship between poverty and the fully at home percentage. When we group census block groups by the percentage of the population in poverty we find areas with highest levels of poverty, 90%-100%, have the lowest fully at home percentages in all three cities. (See figure 3.)

The mobility disparity between the wealthy and the poor is often most pronounced in March and April of 2020, which is a trend we see in all three cities. (See figure 4.) In areas with high poverty less people stay fully at home to reduce their exposure.

We see in all three cities that areas with 90%-100% of the population below the poverty line approximately 5% less of the population do not stay at home. (See figure 3.) A subtle trend that we see is that the wealthiest and poorest CBGs stay home the least on average in 2020. A larger percentage of those in the middle of the spectrum stayed home reducing their exposure. This surprising pattern is in parallel to the findings of Jung et al.[3] who found COVID infection rates to be the highest in the wealthiest and poorest countries.

We plotted the map of Chicago, LA and New York for each day of 2020 (See figures 7, 8, 12, 13, 17 and 18.). The color of the CBG is set by the 2020 stay at home percentage deviation from the 2019 mean. A deep red represents over a 12% or more drop in the number of people at home from the 2019 mean. A deep blue represents an 11% or larger increase from the 2019 mean. 2020 up through the month of March starts out primarily red with a lot of activity. A relatively large percentage of people are out and not at home, especially during the weekdays. In all three cities we see a clear shift in behavior the days of March 14th, 15th and 16th. Monday, March 16th marked the first weekday we see a large number of people staying at home from work. Then, throughout the second half of March we see the majority of all three cities transition to a deep blue as many people stay at home. However, in all three cities you can still see deep red CBGs that are not staying at home.

7.1.1 Chicago. In Chicago, we see that the wealthiest people stayed at home significantly more than the poorest people for the entire

period from March to June. (See figure 4.) In Chicago the values for the wealthy and the poor areas both begin to increase at the same time. The two lines sometimes track closely together. The CBGs with the greatest levels of poverty have a substantially larger percentage of the population not staying at home when compared with the CBGs having the lowest levels of poverty. From March 17 the two lines deviate. We commonly see 10% less of the population at home and sometimes up to 20%. (See figure 9.) Despite the mobility disparity, the wealthiest and poorest areas fully at home percentage values tend to track more closely than in the other cities, usually within about 5%. Part of the reason the groups track closely together is Chicago does not increase the stay at home percentage as much as New York and Los Angeles. The Chicago mean daily fully at home percentage peaks below 50%. We tracked down the dates of the major announcements related to lock downs in Chicago in 2020, and analyzed their effect in the people's mobility.[11]

7.1.2 Los Angeles. In the Los Angeles at fully at home percentages we see a large mobility disparity between the behavior of the areas with the highest and lowest levels of poverty. Areas with low poverty rapidly changed their behavior in March and April increasing the fully at home percentage of the population by almost 30%. However, we do not see the areas with low poverty making a substantial change in behavior for over a month and a half. The delay in changed behavior may have contributed substantially to the increased exposure of the poor to possible infection. (See figures 4 and 19.) In no other cities do we see such a slow response. While the wealthiest areas reach a peak stay at home percentage, the poorest have about 20% more of their population not staying fully at home. However, in May they is a rapid increase in the number of people in poor areas staying home that allows the two groups to predominantly converge on similar values.

In Los Angeles, the Governor imposed stay at home restrictions on March 19 but we already see an existing gap in mobility between the areas with the highest and lowest poverty levels [4].The gap remains the same throughout March and further in summer(See figure 19.). During the second half of 2020, we see that the gap between the two groups minimizes but by the end of the year, it is evident from the visualisation (See figure 19.) that the difference in the mobility increases again.

7.1.3 New York. The poorest areas in New York also have a delayed response, but the delay is shorter than we see in Los Angeles. The poorest areas took an additional 2 week to match the response of the rich when individuals began to stay home in March in New York. However, the poorest areas daily fully at home percentage stays 5% to 15% below the wealthiest for most of 2020. The mobility disparity is clear. In neither of the other two cities do we see such a sustained difference between the two groups.

In New York, there is a significant drop in the stay-at-home percentage in the CBGs with the poorest population from March 9. The areas with the richest population on the other hand continuously increased their stay at home percentage and the gap between the two groups remained throughout the year.(See figure 14.) Following the month of March, the CBGs with the wealthiest population maintained their stay at home over 25% in March- April and above 15% later during summer. The CBGs with the poorest population show many irregularities in their behavior. We see that they stayed

at home during the weekends but often during the weekdays they do not stay at home despite the lockdown impositions becoming stricter with time [5].

7.2 Full-time and Part-time Work Behavior

The wealthiest areas in all three cities decrease the amount of the population commuting to full-time and part-time in mid-March. Then, they all maintain the low mobility throughout the rest of the year with only a slight gradual increase. (See figures 5 and 6.) The poorest areas in all three cities did not change their behavior in such a complete and consistent way. In all three cities the poor often leave home to go to work in larger numbers than those in more wealthy areas.

7.2.1 Chicago. In Chicago, we see an inversion of the work behavior of the those with the greatest levels of poverty and those with the least. (See figures 5 and 6.) We see the behavior of the wealthy and the poor tracking more closely together than in New York and Los Angeles.

7.2.2 Los Angeles. Unlike Chicago and New York, in Los Angeles the areas with the highest levels of poverty do not show a drastic decrease in full-time work in March and April. (See figure 5) In fact the poorest areas show an increase in full-time work during all of 2020, including March and April. During the same period the wealthiest areas are steeply decreasing their commuting to work behavior. Even the average commuting to work behavior in all cities, including LA, shows a steep decline during this period. The poor areas in Los Angeles are initially low, often around 2%-3%, which may explain the lack of downward change. Perhaps the wealthy avoiding in-person work and many in poor areas not having work contributed to the increasing at-work numbers we see in Los Angeles through 2020.

The part-time work behavior the poorest areas does decline in March. (See figure 6) However, the magnitude of the change is small in comparison with the radically changing behavior of the wealthy. Furthermore, the poorest areas are back to high percentages of at-work patterns of behavior by August of 2020.

The lack of rapid change in the fully at home home percentage among those living in poor areas is partly explained by the lack of substantial change in part-time work behavior and the increase in at-work behavior. We see a steep increase in the number of fully at home people at the end of April. At the same time full-time work is increasing steeply and part-time work is relatively stable. April is when the first round of stimulus checks went out, which may have been a major contributing factor to allow the increased at home percentage. Especially since full-time work is increasing at the same time.

7.2.3 New York. In New York, wealthy people went out for full-time work more than the poor people, but this trend inverts sharply during March. (See figure 5 and 6.) Furthermore, we see that during subsequent months, there are some time periods when rich people go out more than the poor. However, these time periods are very rare and it is mostly poor people who couldn't avoid going out for work. In fact, there are big peaks in the trends for poor people on some weekdays. This directly implies that while rich people were able to continue their jobs from home, poor people do not have the

luxury to do the same.(See figure 15.) A similar observation can be made for the part-time work behavior. Both the groups increase their participation in the part-time work increases during weekdays and decreases during weekends, but until December it is mostly poor people that go out of home for part-time. Only until December, we see both the groups leaving homes for their respective jobs.

7.3 Sampling Bias

We divided the average number of devices by the population size in each census block group, which was multiplied by a 100 to get the percentage of the population included in the SafeGraph Social Distancing in each census block group. We found in Chicago and Los Angeles the areas with the least poverty have a larger sample percentage than those with more poverty. (See figure 22) In New York the sample percentage is smaller for the areas with the highest poverty.

8 ISSUES FACED

The engineers at SafeGraph advocate transparency in data so that their customers face minimal issues in analyzing their different datasets. However, due to large dataset sizes and encoding of the columns, we faced several challenges while extracting the valuable information from the whole set. These challenges are discussed as below:

- Google Drive is being used by the team as the data storage solution while collaborating on this project. Our Google Colab notebooks sometimes require over the 10^9 limit of allowed daily file reads by Google Drive. This has happened several times while doing analysis that includes all CBGs for every day of one or more years. It has even happened after running several different analyses of a subset of the CBGs on the same day. To address the issue we cloned the Google Drive allowing us to spread out the usage across multiple Google Drive accounts.
- The social distancing metrics file size was 81 GB at the time of our download. It took more than 4 to 6 hours to download. Further, such a big file suffered download failed multiple times.
- The Census Block Group Data was 10 GB at the time of our download. It took 3 hours to download.
- The folder structure for the social distancing metrics dataset for a year is divided based on months and further based on days. Hence, for the years 2019 and 2020, we found 12 sub-folders (each representing a month) and archive files inside these sub-folders for each day of the month. It caused difficulty in integrating the dataset in a single data frame. Therefore, we ran a 'loop' for every month and day to extract the data rows and combine them into one data frame to perform analysis and visualization.
- We tried to write the combined dataset for 2020 in one CSV file, which caused memory overload and crashed our local machine.
- Each column in open census data is a particular census attribute estimated by the US Government. It includes 7500 *table_ids* (columns) for the 220,000 + census block groups in the USA. it was quite overwhelming and took time to wrap

our heads around. After following the specific guidelines (and cheat sheet) from the safegraph, we identified the type of population represented by each *table_id*.

- The SafeGraph Social Distancing Dataset uses location data collected from cell phone metadata. In order to be included in the data an individual must own a cell phone. There is a risk that an analysis based on cell phone data will miss trends related to groups not using cell phones in large numbers. Looking at the number of individuals we can see a few clear trends. CBGs having populations with greater than 50% of the population are relatively rare. So naturally we will have less information about trends in these areas. First, we find the average daily number of devices in each CBG over all of 2020. Then, we divide the number of devices by the CBG population size from the census data. We found several issues. First in both Chicago and Los Angeles we see there tends to be a larger percentage of the population in the wealthier CBGs. Second, in New York there are several CBGs where the 60-80 poverty range where the number of devices greatly exceeds the size of the population recorded in the census data.

9 CONCLUSIONS

New York was most able to reduce their fully at home percentage peaking at 55%, followed by Los Angeles peaking around 50%. (See figure 4) Chicago's population shows the least amount of change in their fully at home percentage peaking below 50%. Chicago has the most unified response, both the wealthiest and the poorest census block groups often stay with 5% of each other throughout 2020.

New York shows the most consistent mobility disparity where the poorest areas have a fully at home percentage 5% to 15% below the wealthiest areas for most of 2020. In New York the disparity is also apparent in the working patterns. (See figures 5 and 6.) The areas with the most poverty have full-time and part-time that often rises 5% to 10% above the wealthier areas. The wealthier areas on the other hand had a very stable low value with none of the erratic increase seen with the poor.

The disparity between the areas with high and low poverty plays out differently in Los Angeles. While the wealthy immediately and drastically change their stay at home behavior in mid-March the poorest areas do not see a substantial increase for a full month and a half. Only in May did the poorer areas begin to change their behavior and stay at home in larger numbers. In Los Angeles we see another surprising event. The areas with high poverty actually increase their at-work behavior throughout most of the year, peaking after a 4% increase in September. The striking increase in mobility above the baseline moves counter to the behavior in Chicago and New York. Also, the increase in mobility moves counter to the wealthy areas in Los Angeles, which reduce their at-work behavior starting in March and lasting through the rest of 2020.

Out of all three cities Chicago had the largest mean percentage of the population included in the sample when compared with Los Angeles and New York. (See figure 22.)

We found a bias in the percentage of the population sampled in the SafeGraph Social Distancing data relative to the amount of poverty in the census block group. (See figure 22.) In all three cities

higher areas with higher poverty had a lower percentage of the population included in the sample. In Chicago and Los Angeles areas with the least poverty, 0% to 20% of the population in poverty, had the largest percentage of the population of the included. All areas with higher poverty, 20% to 100% of the population in poverty, the percentage of the population included in the sample is lower. In New York the areas with 80% to 100% of the population in poverty have a lower percentage of the population included in the sample than all other areas. We should be careful about the assumptions we make when using data such as cell phone data. It is possible for some groups to be overrepresented and other to be left out all together.

The SafeGraph data is of high quality. As a result, our comprehensive data quality study found no major issues. The Census Block Group data was unique, all the values were found to fall within the expected range and there were no null values.

The datasets are huge and require a lot of computational time. The data extraction process can be run on a modest machine, but very slowly. Especially the very large SafeGraph Social Distancing dataset, which can take two hours to extract the data for just one city. However, after the extraction process loading data for one year of one city into memory only takes 15 minutes. Once in memory exploration and analysis can be performed quickly.

Using Google Colab with Google Drive the data can be extracted from the raw data set. However, occasionally the Google API call limit is reached. The issue only occurred when extracting the very large SafeGraph Social Distancing dataset. When the API limit is reached we found that closing all notebooks and waiting 24 hours allowed us to finish running the extraction notebooks.

The semantics of working with the Census data is not very intuitive. You must understand the FIPS code system, which is the primary method used for encoding the geological entities throughout the United States. The ACS columns naming system is not intuitive. In addition, the column descriptions in the documentation are not always easily understood either. Fortunately, the ACS documentation is very rigorous and there are many resources online to provide additional instruction on how to work with Census data. In addition, once you figure out the semantics and structure of the data there is very little cleaning to do because the quality of data is very good.

We anticipated a spike in the number of individuals staying at home starting in March due to many factors, including the issuing of stay at home orders. Inversely, we expected to see the number of people exhibiting part-time and full-time work behavior decline in March. We visualized the mean percentage values for each of these three attributes for each of the three cities in our study.³⁶ The line graphs conform to our expectations, which gives us confidence in the data. In comparison, the line graphs of 2019 show gradual change.³⁵

10 CITATIONS AND BIBLIOGRAPHIES

REFERENCES

- [1] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. 2021. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 589, 7840 (01 Jan 2021), 82–87. <https://doi.org/10.1038/s41586-020-2923-3>
- [2] Denise Lu Jennifer Valentino-DeVries and Gabriel J.X. Dance. 2020. Location Data Says It All: Staying at Home During Coronavirus Is a Luxury. <https://www>.

- nytimes.com/interactive/2020/04/03/us/coronavirus-stay-home-rich-poor.html
- [3] Juergen Jung, James Manley, and Vinish Shrestha. 2021. Coronavirus infections and deaths by poverty status: The effects of social distancing. *Journal of Economic Behavior & Organization* 182 (2021), 311–330. <https://doi.org/10.1016/j.jebo.2020.12.019>
 - [4] Jason Kandel. 2021. Timeline: A Look at Key Coronavirus Pandemic Events and Milestones in California. <https://www.nbclosangeles.com/news/coronavirus/california-coronavirus-pandemic-timeline-key-events/2334100/>
 - [5] Alexandra Kerr. 2021. A Historical Timeline of COVID-19 in New York City. <https://www.investopedia.com/historical-timeline-of-covid-19-in-new-york-city-5071986>
 - [6] Bijan Kiniagar and Jack Mullan. 2020. Opinion: NYC's Growing Income Inequality Threatens Pandemic Recovery. <https://citylimits.org/2020/10/06/opinion-nycs-growing-income-inequality-threatens-pandemic-recovery/>
 - [7] Robert G. Mogull. 2011. A Contrast of U.S. Metropolitan Demographic Poverty: Chicago, Los Angeles, and New York. <https://www.hindawi.com/journals/ijpr/2011/860684>
 - [8] Richard V. Reeves and Jonathan Rothwell. 2020. Class and COVID: How the Less Affluent face Double Risks. <https://www.brookings.edu/blog/up-front/2020/03/27/class-and-covid-how-the-less-affluent-face-double-risks>
 - [9] SafeGraph. 2016. Census Block Group Data. <https://docs.safegraph.com/docs/open-census-data> (All data from 2016 American Community Survey by Census Block Group).
 - [10] SafeGraph. 2021. Social Distancing Metrics. <https://docs.safegraph.com/docs/social-distancing-metrics>
 - [11] TRIBUNE STAFF WITH NEWS SERVICES. 2020. 6 months of COVID-19: Timeline of the outbreak and how politics, sports, entertainment and the economy changed. <https://www.chicagotribune.com/coronavirus/ct-viz-coronavirus-timeline-20200507-uvrzs32nljabrpn6vkzq7m2fpq-story.html>
 - [12] The New York Times. 2021. Coronavirus (COVID-19) Data in the United States. <https://github.com/nytimes/covid-19-data>

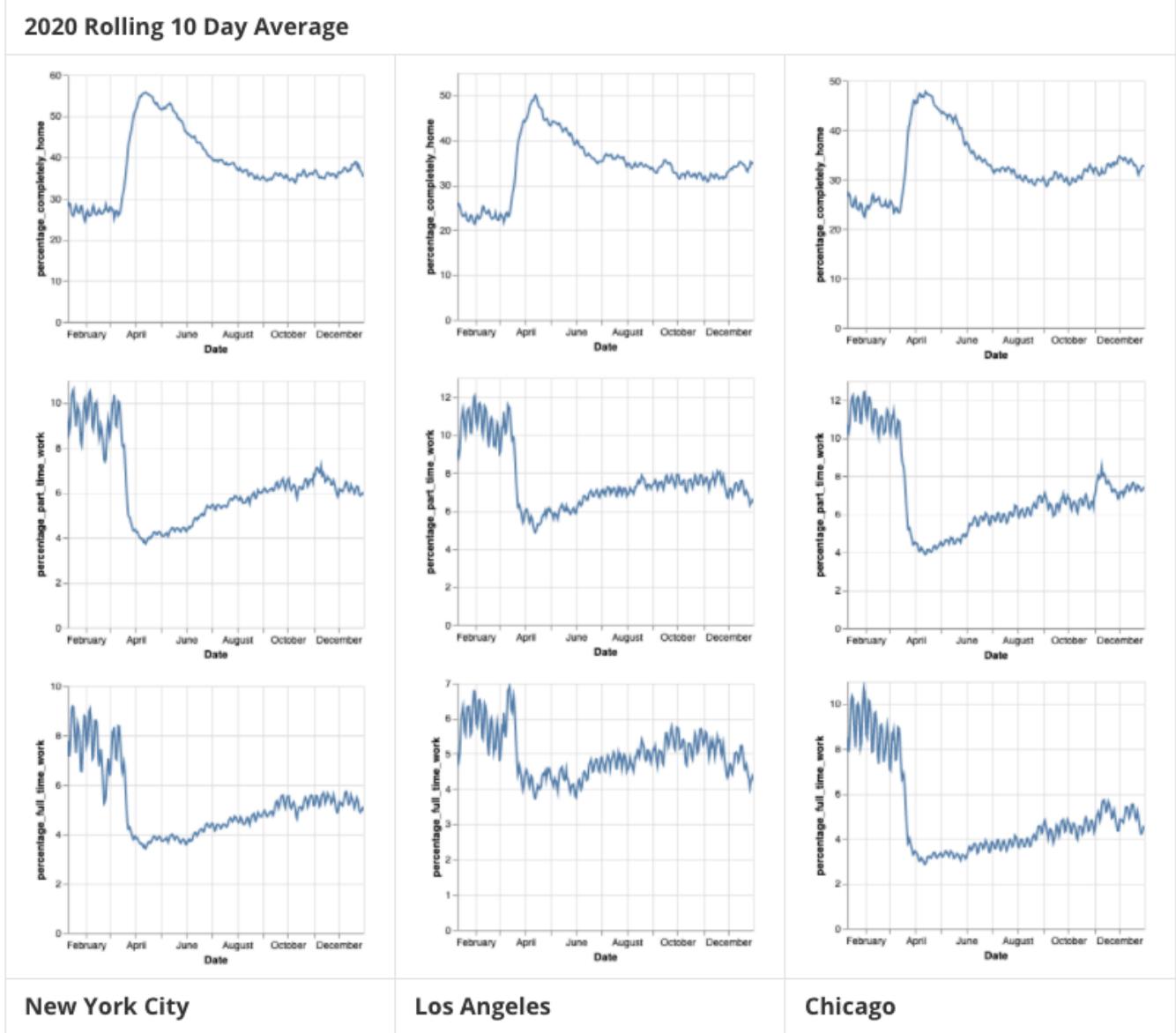


Figure 1: New York, Los Angeles and Chicago 2020 rolling 10 day average for the percentage of individuals at home all day, leaving home for full-time work and leaving home for part-time work.

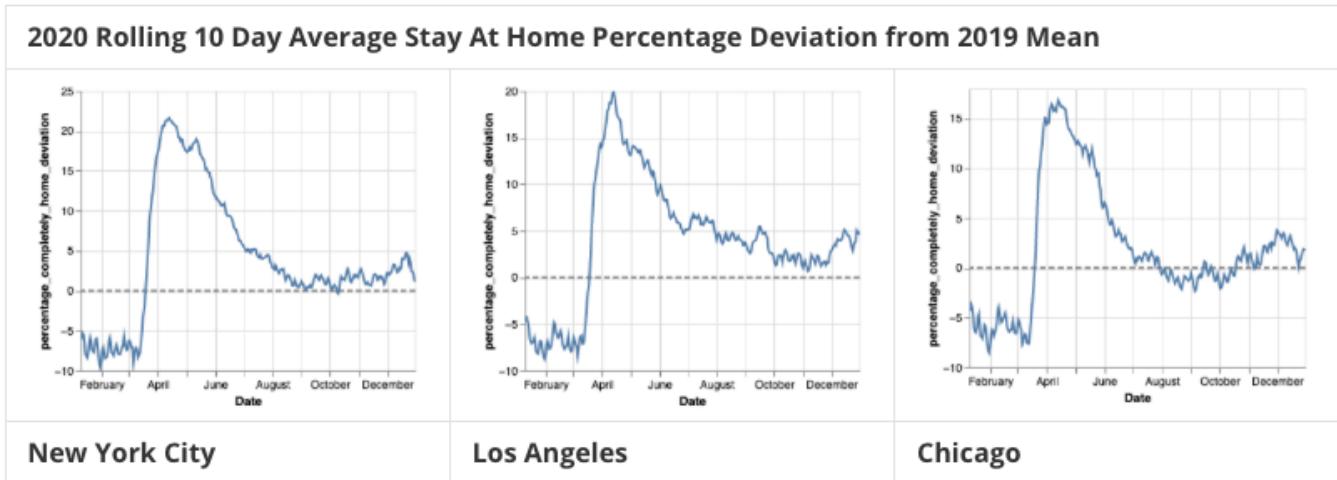


Figure 2: New York, Los Angeles and Chicago rolling 10 day average for the 2020 daily percentage of individuals at home all day deviation from the 2019 mean.

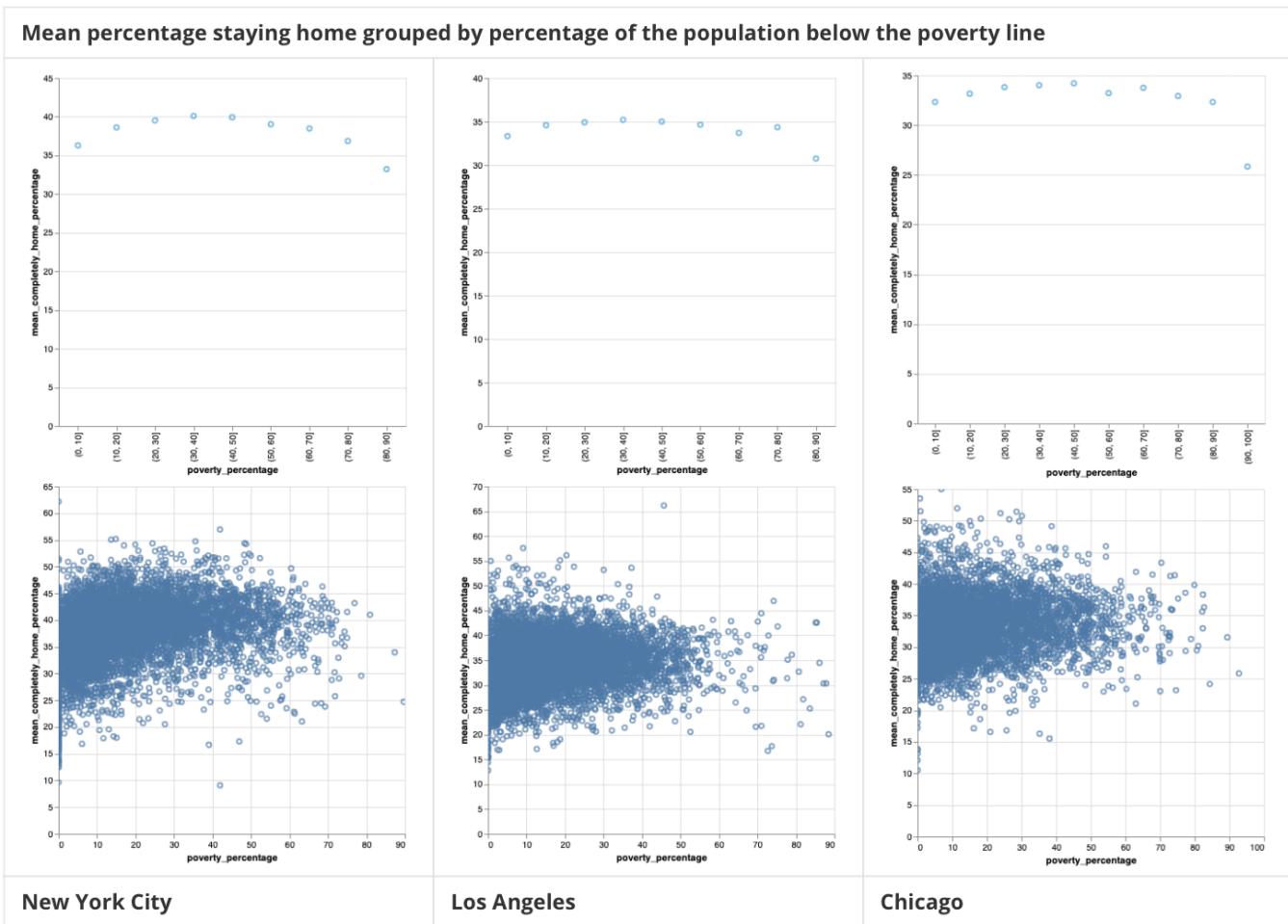


Figure 3: 2020 New York, Los Angeles and Chicago mean percentage at home grouped by percentage of the population below the poverty line (top) and grouped by individual CBGs (bottom).

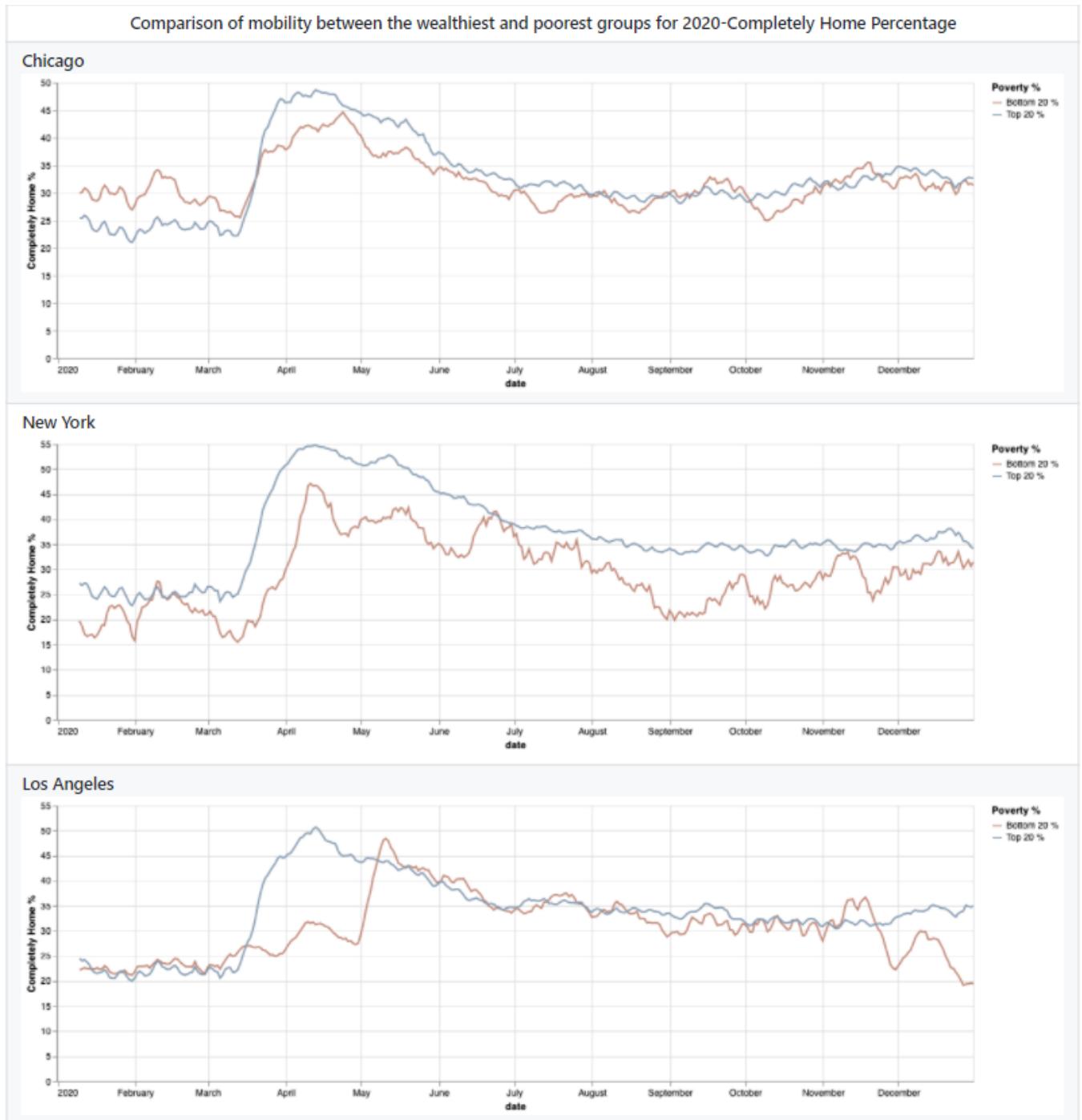


Figure 4: 2020 New York, Los Angeles and Chicago daily completely at home percentage comparing the lowest (top 20%) and highest (bottom 20%) poverty quintile.

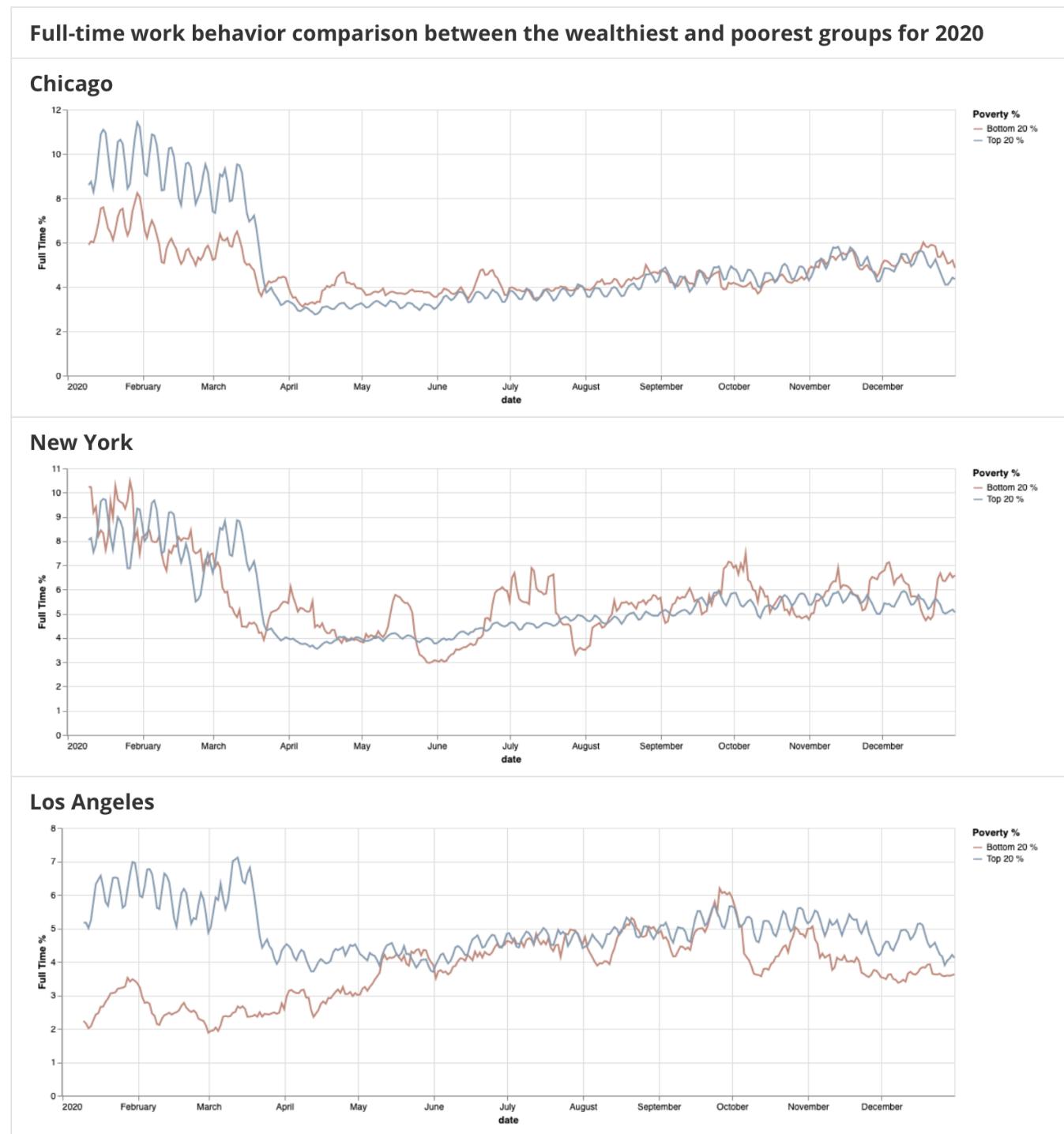
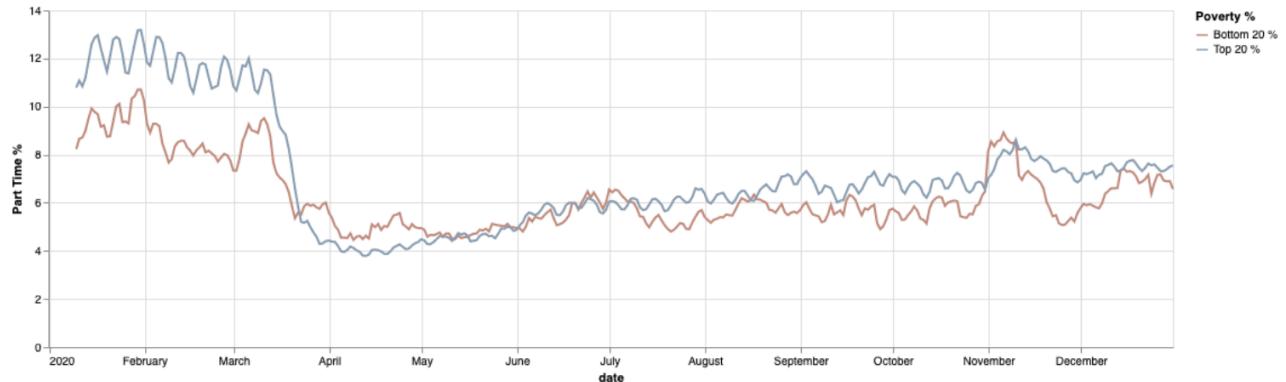


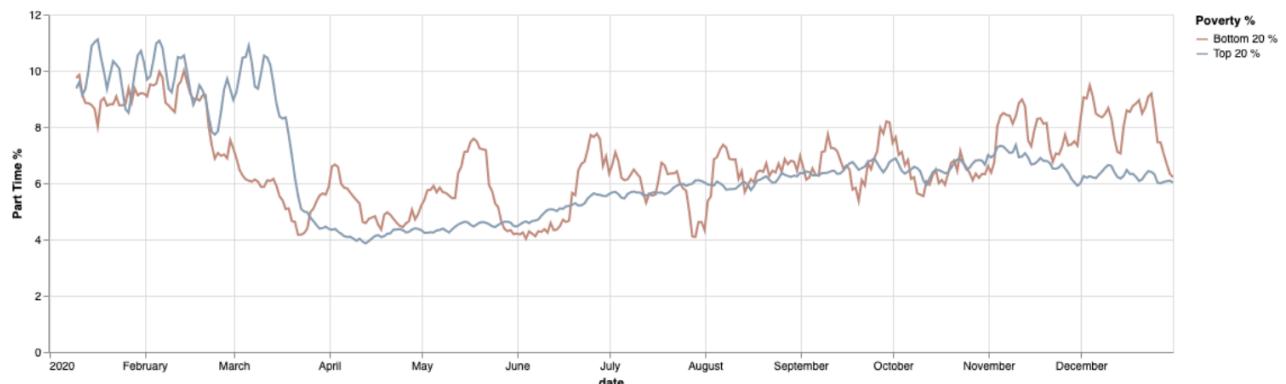
Figure 5: 2020 New York, Los Angeles and Chicago Full-time Work Percentage comparing the lowest (top 20%) and highest (bottom 20%) poverty quintile.

Part-time work behavior comparison between the wealthiest and poorest groups for 2020

Chicago



New York



Los Angeles

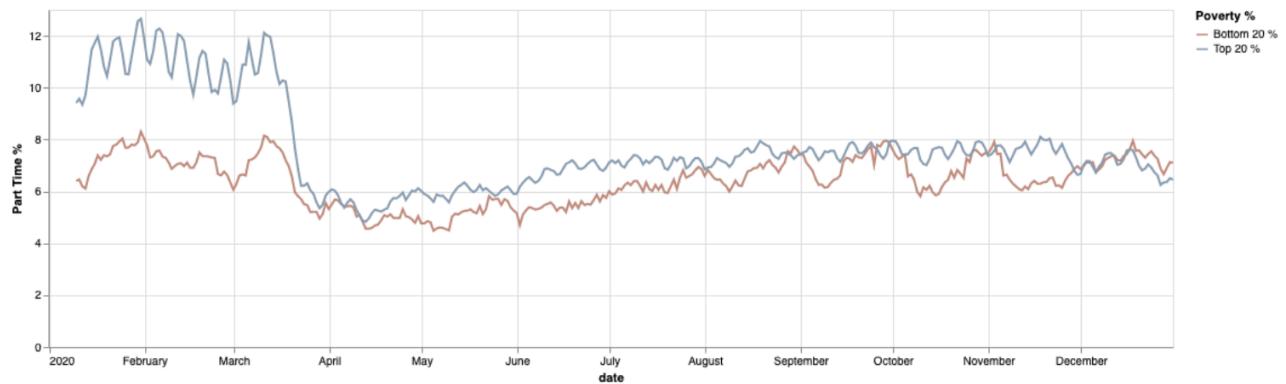


Figure 6: 2020 New York, Los Angeles and Chicago Part-time Work Percentage comparing the lowest (top 20%) and highest (bottom 20%) poverty quintile.

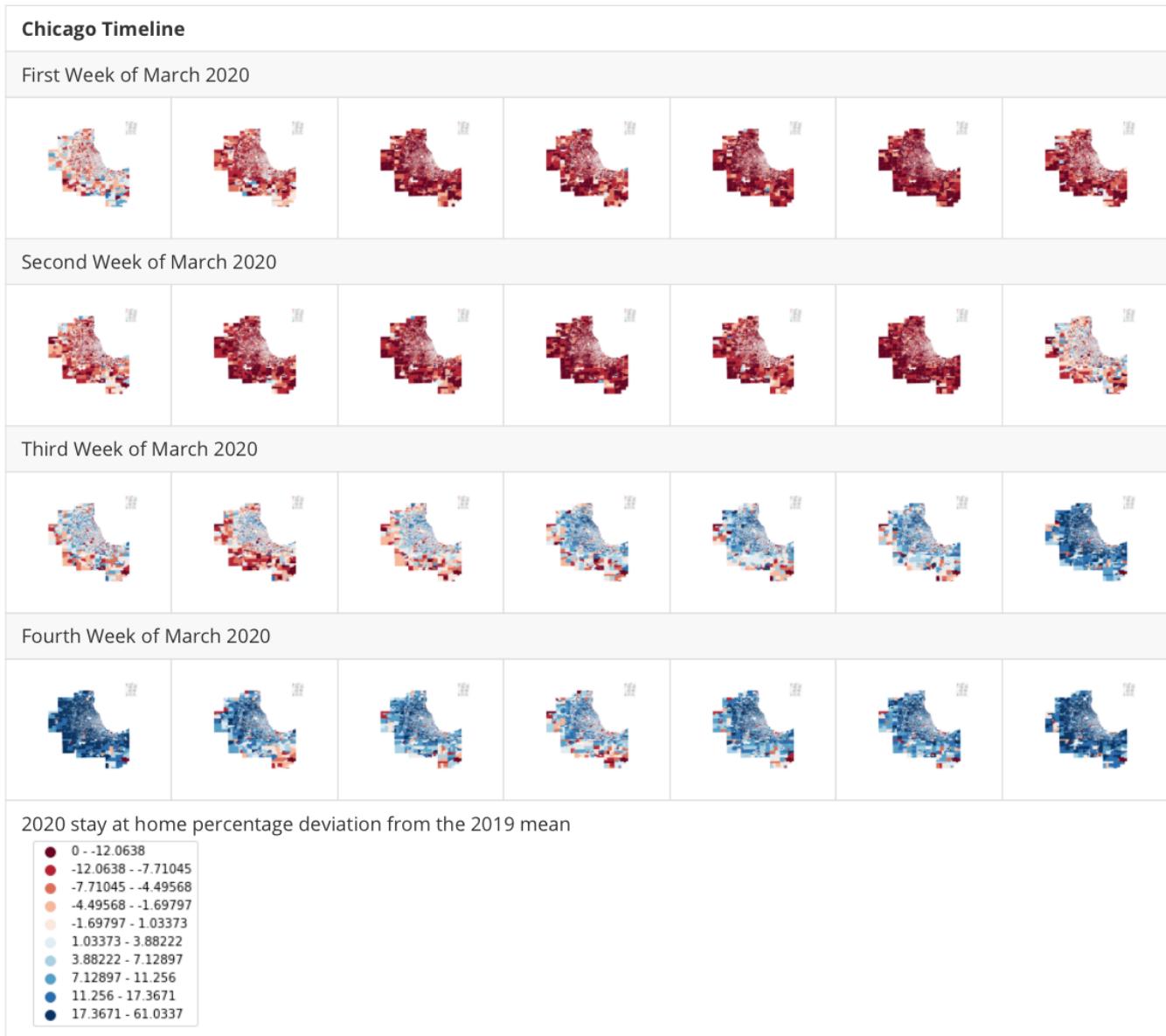


Figure 7: Chicago census block groups daily at home percentage deviation from the 2019 mean for the first four weeks of March 2020.

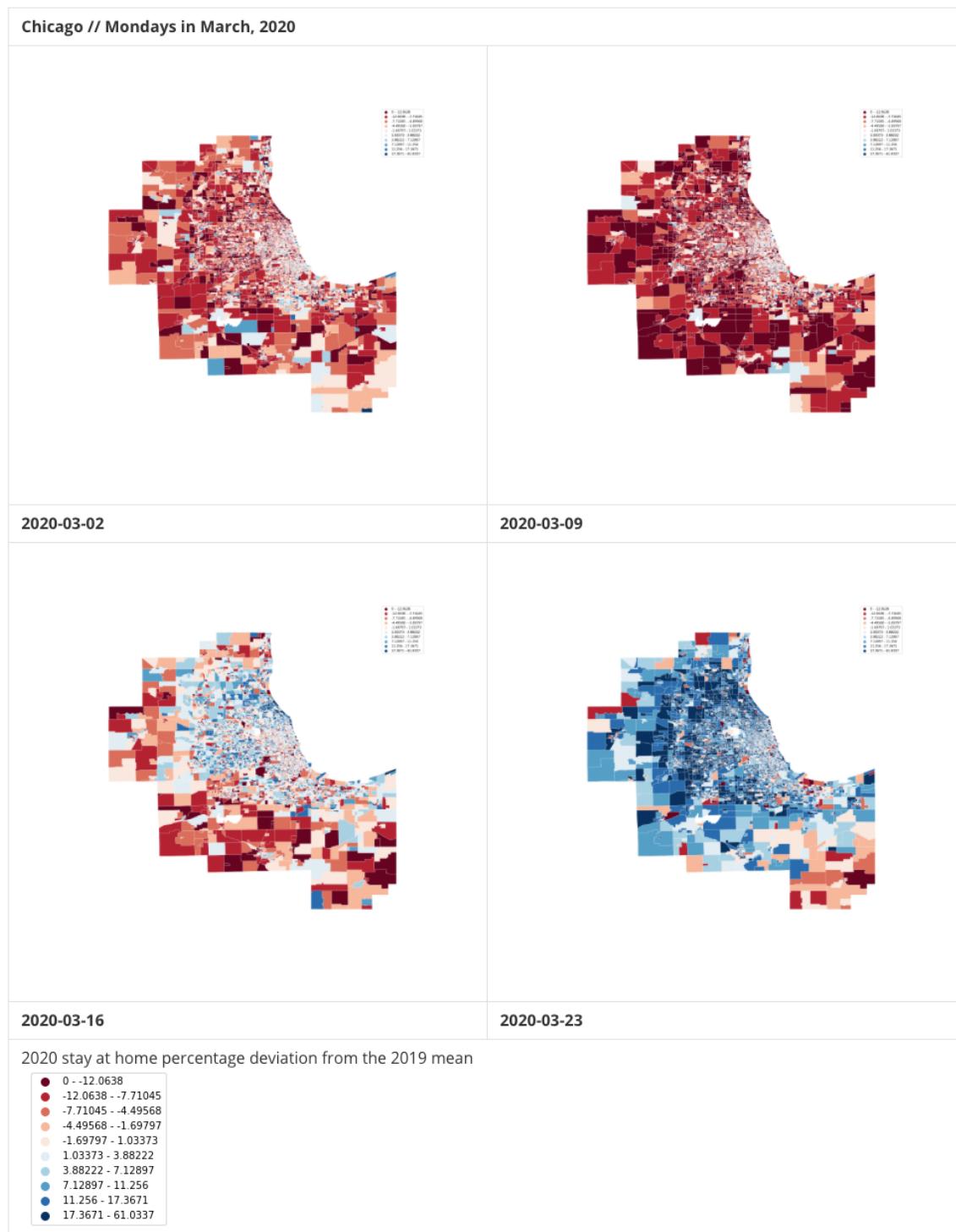


Figure 8: Enlarged view of the Chicago census block groups daily at home percentage deviation from the 2019 mean for the first four Mondays of March 2020.

Visualizing the Mobility Gradient Over Time and in Relation to Poverty

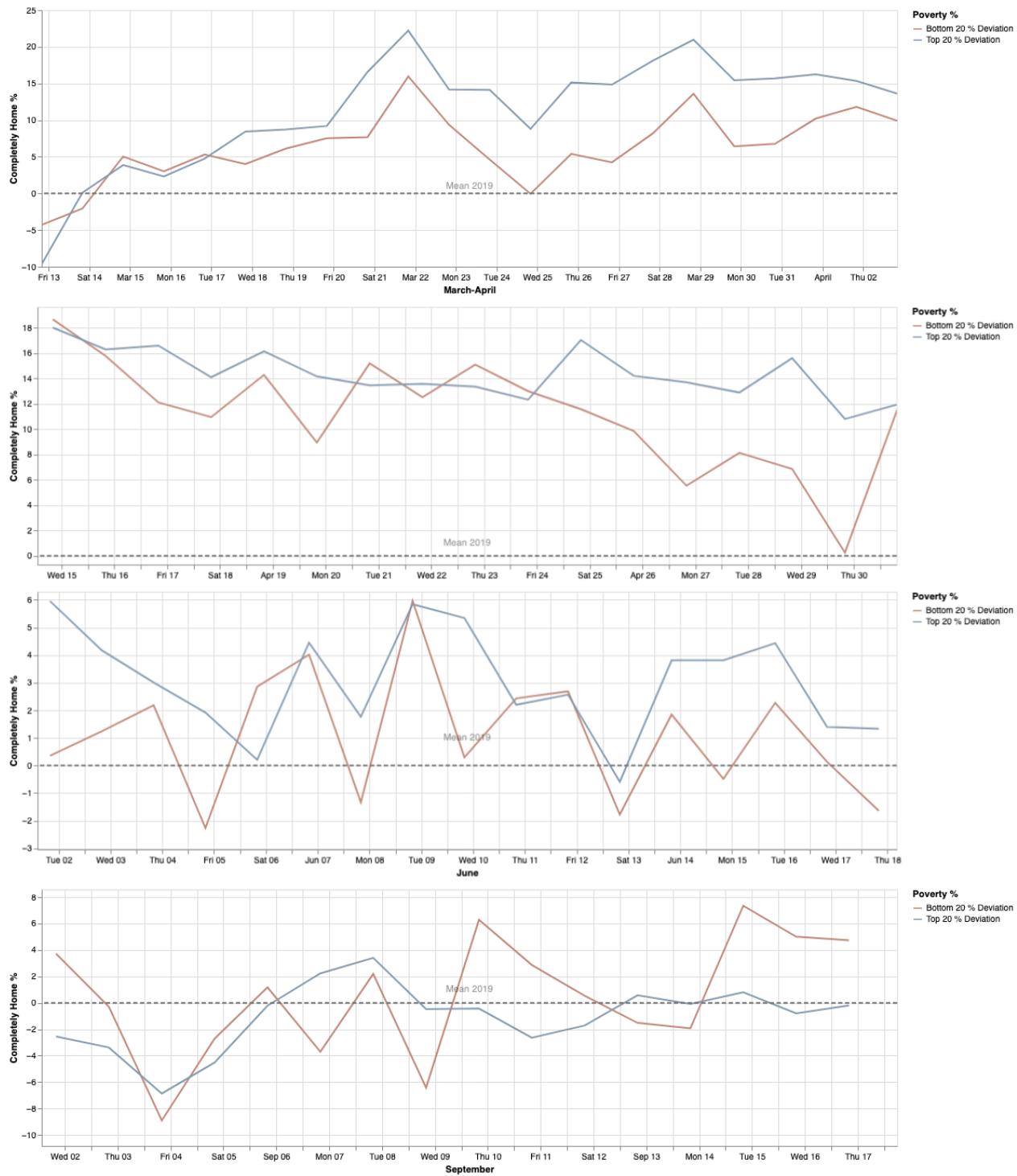


Figure 9: Comparison of Completely Home behavior between the wealthiest and poorest groups for 2020 for specific timelines in Chicago: March, April, June & September.

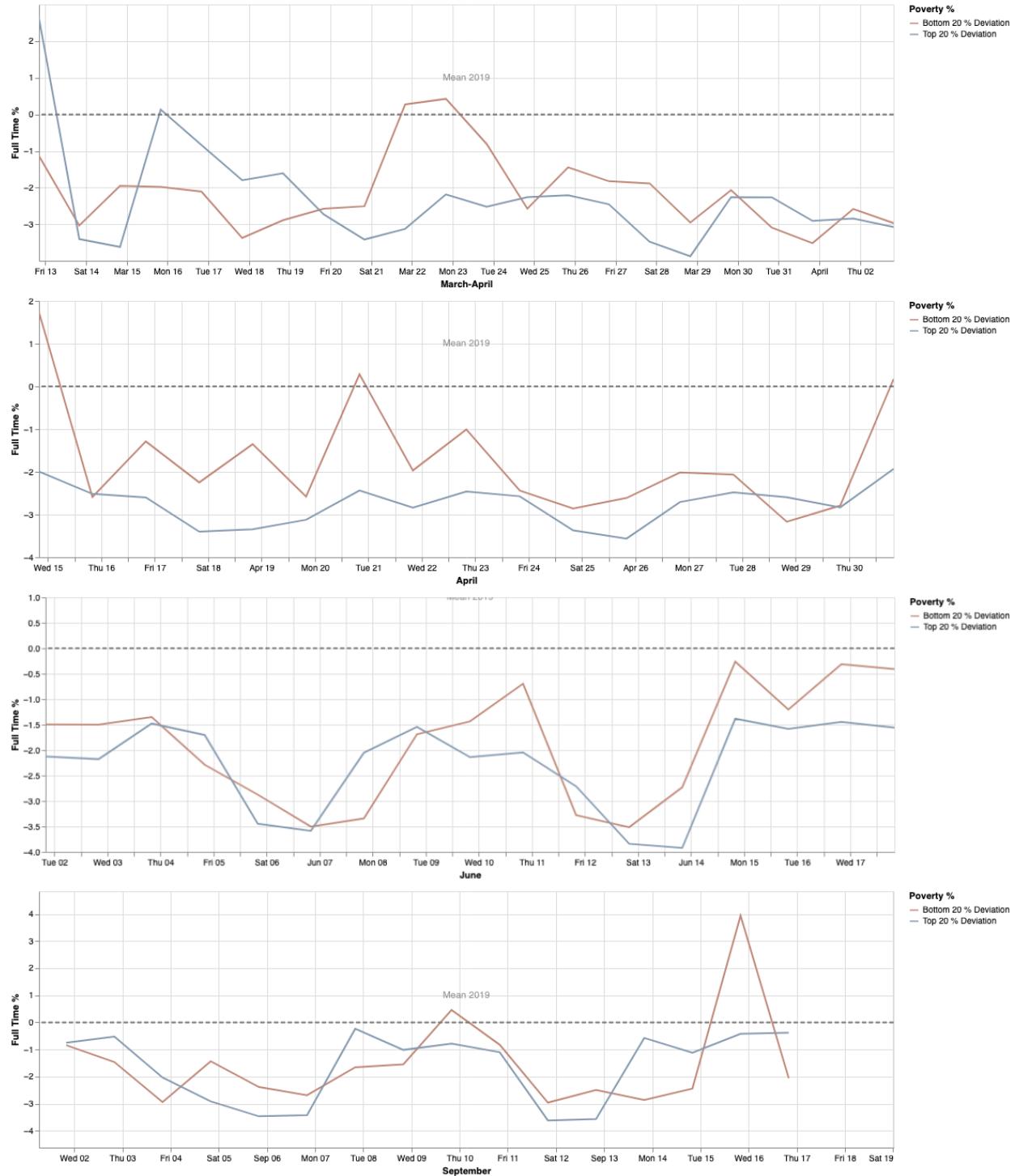


Figure 10: Comparison of Full-Time Work behavior between the wealthiest and poorest groups for 2020 for specific timelines in Chicago: March, April, June & September.

Visualizing the Mobility Gradient Over Time and in Relation to Poverty

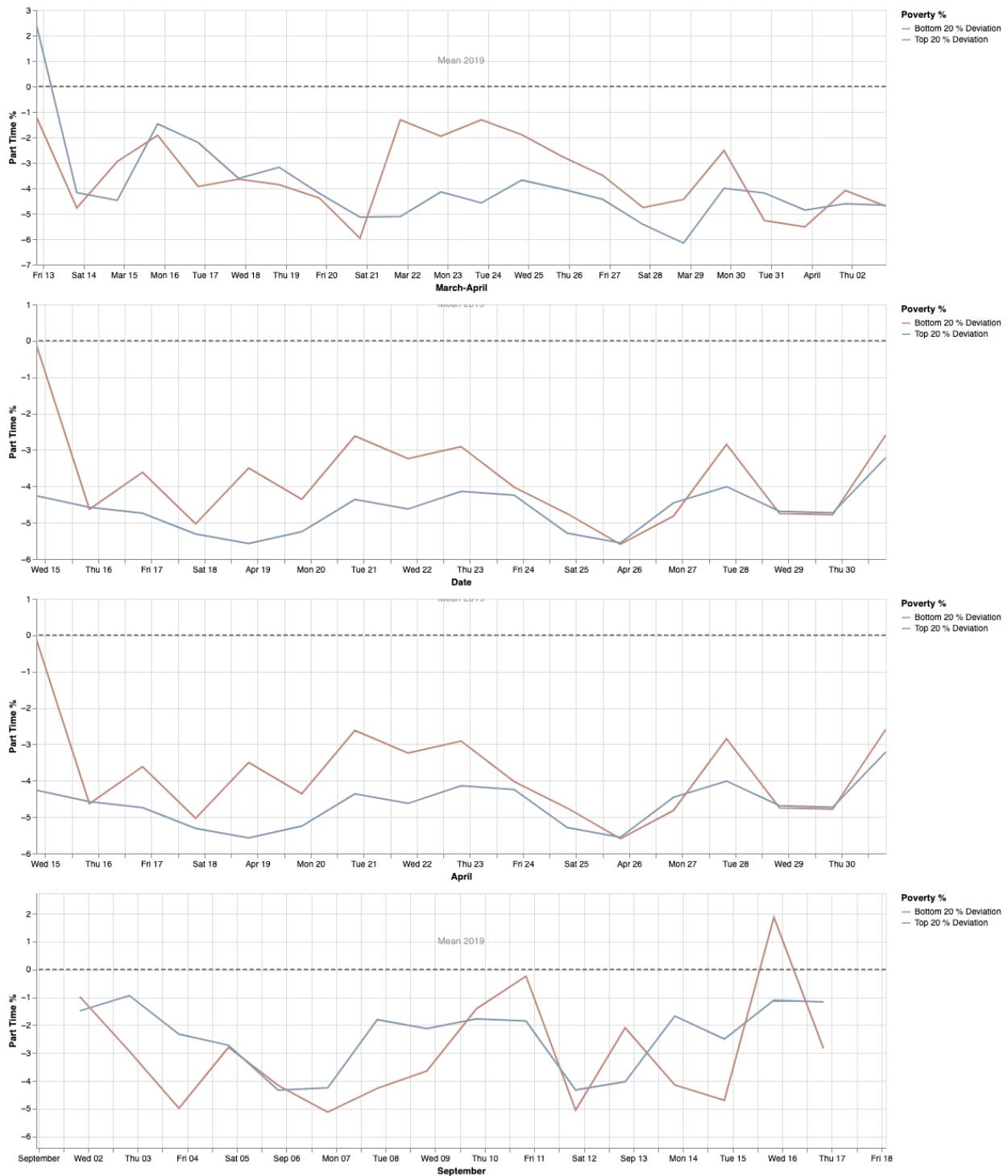


Figure 11: Comparison of part-time Work behavior between the wealthiest and poorest groups for 2020 for specific timelines in Chicago: March, April, June & September.

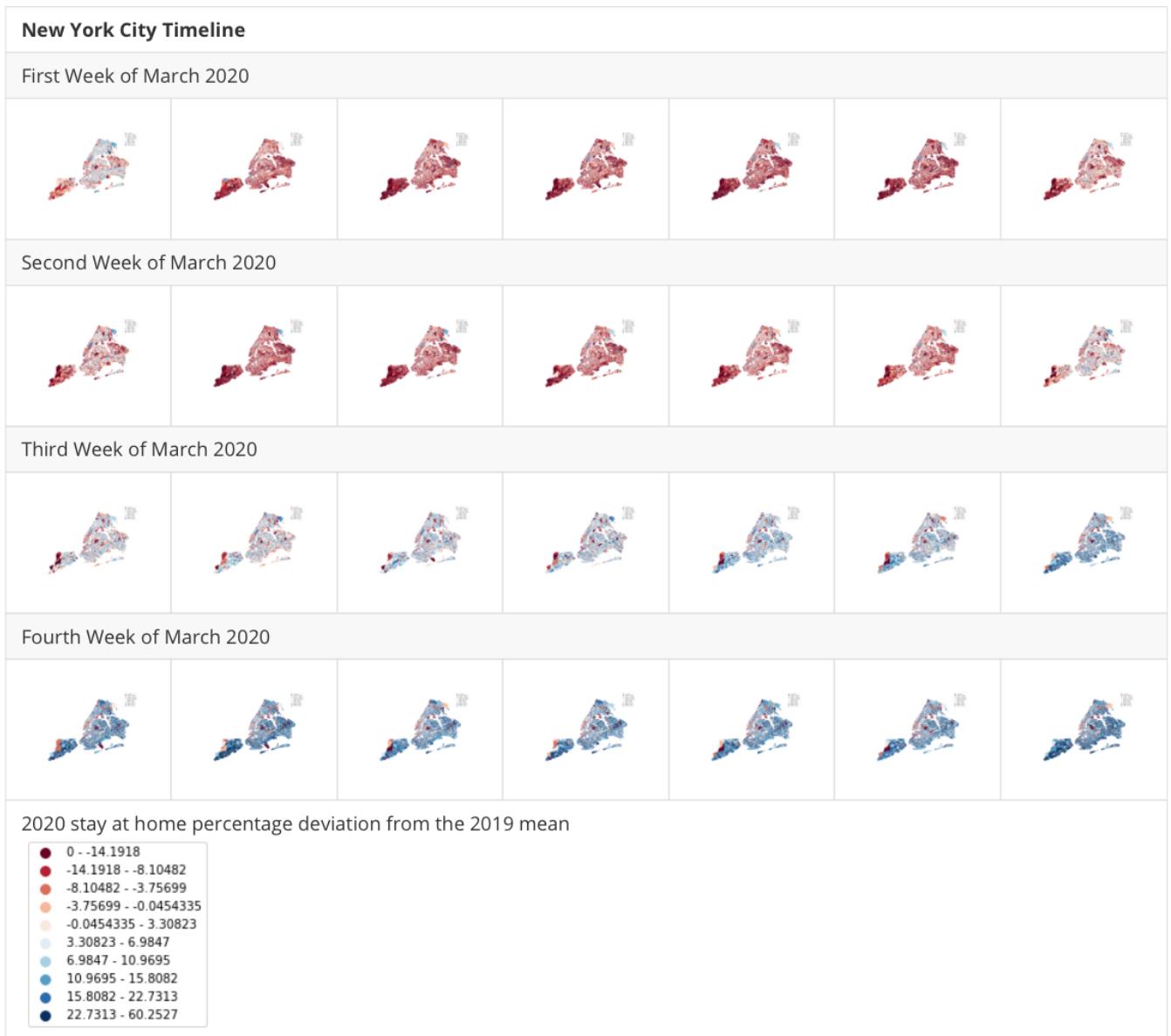


Figure 12: New York City census block groups daily at home percentage deviation from the 2019 mean for the first four weeks of March 2020.

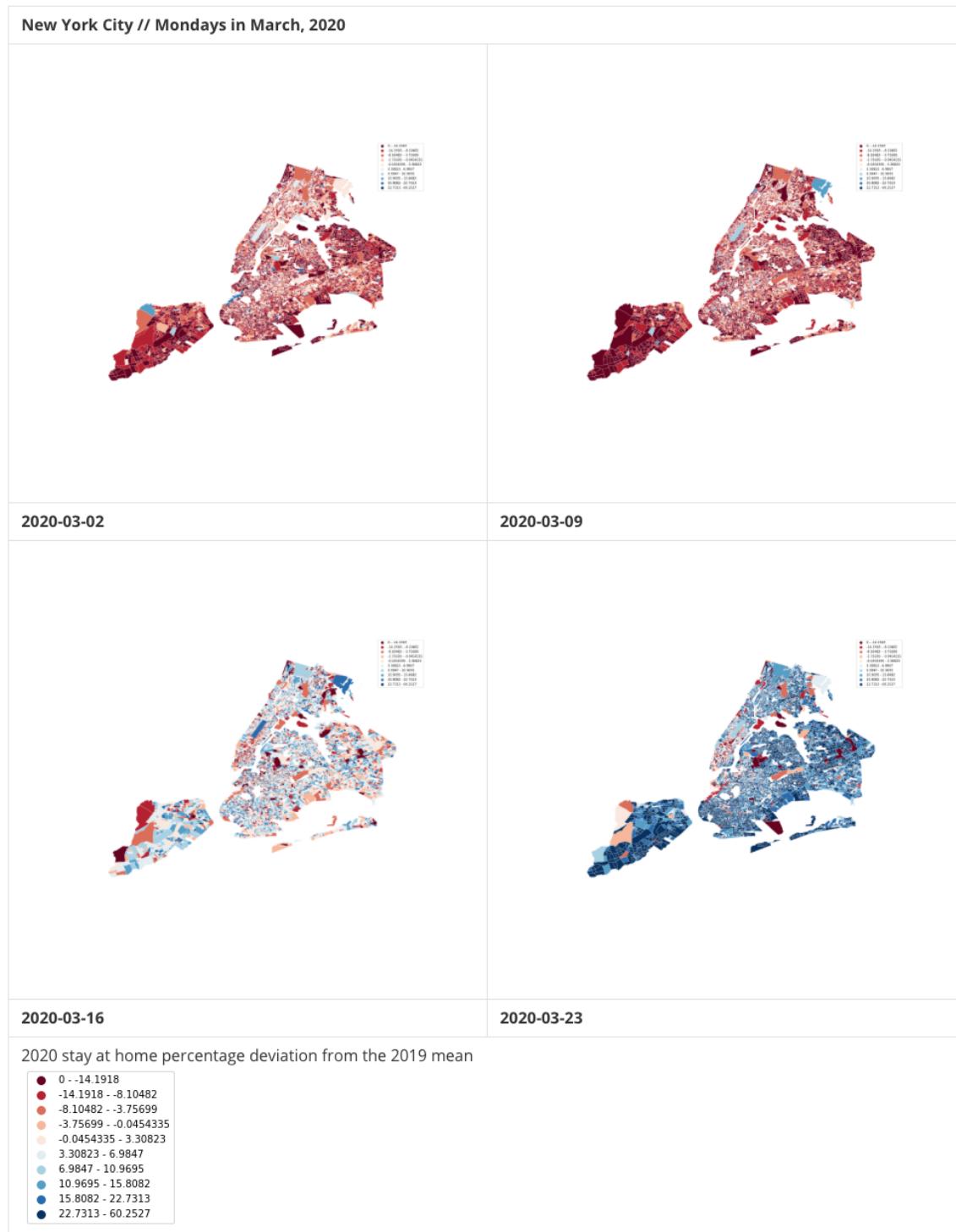


Figure 13: Enlarged view of the New York City census block groups daily at home percentage deviation from the 2019 mean for the first four Mondays of March 2020.

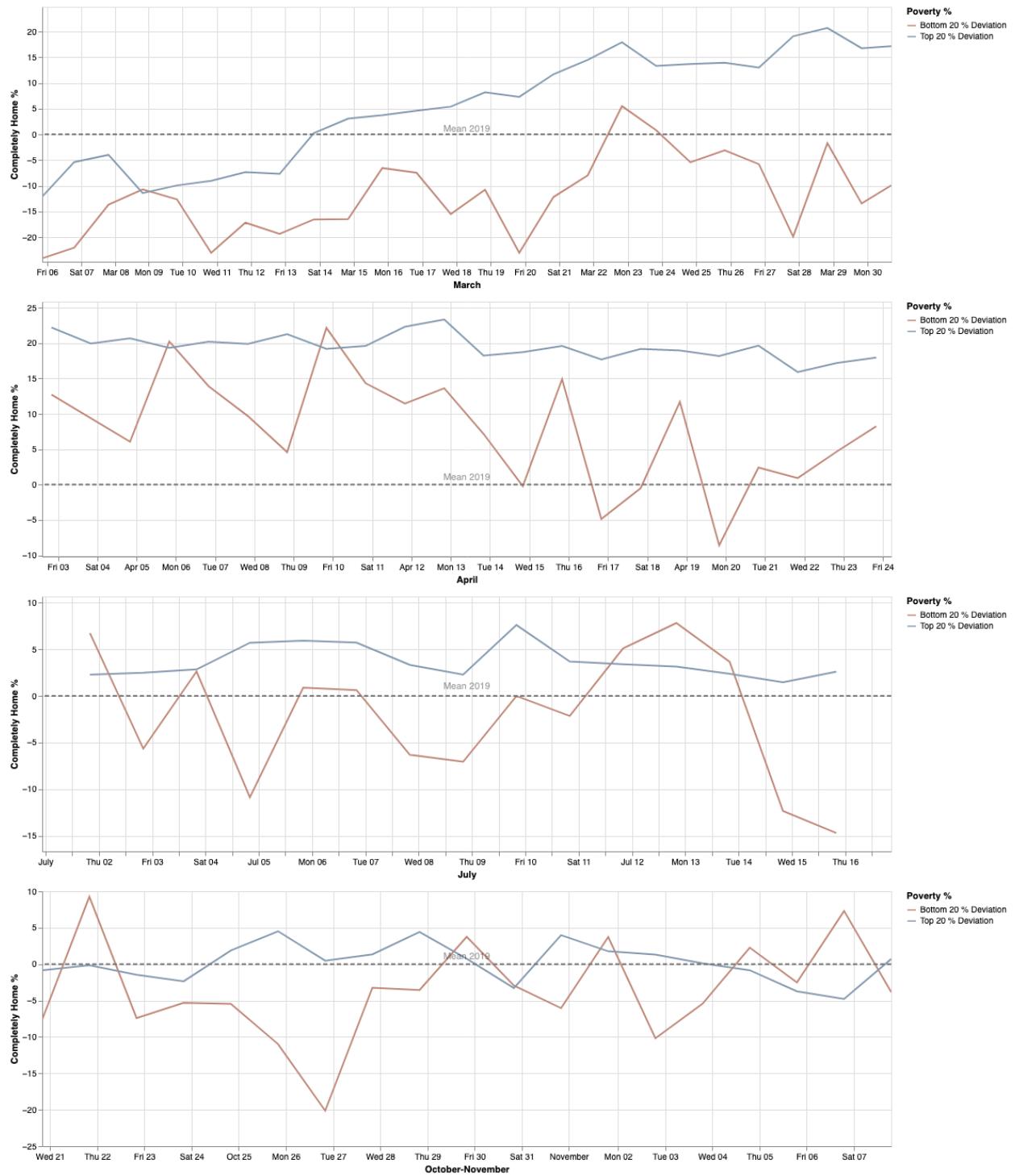


Figure 14: Comparison of Completely Home behavior between the wealthiest and poorest groups for 2020 for specific timelines in New York: March, April, July & October.

Visualizing the Mobility Gradient Over Time and in Relation to Poverty



Figure 15: Comparison of Full-time Work behavior between the wealthiest and poorest groups for 2020 for specific timelines in New York: March, April, July & October.



Figure 16: Comparison of part-time Work behavior between the wealthiest and poorest groups for 2020 for specific timelines in New York: March, April, July & October.

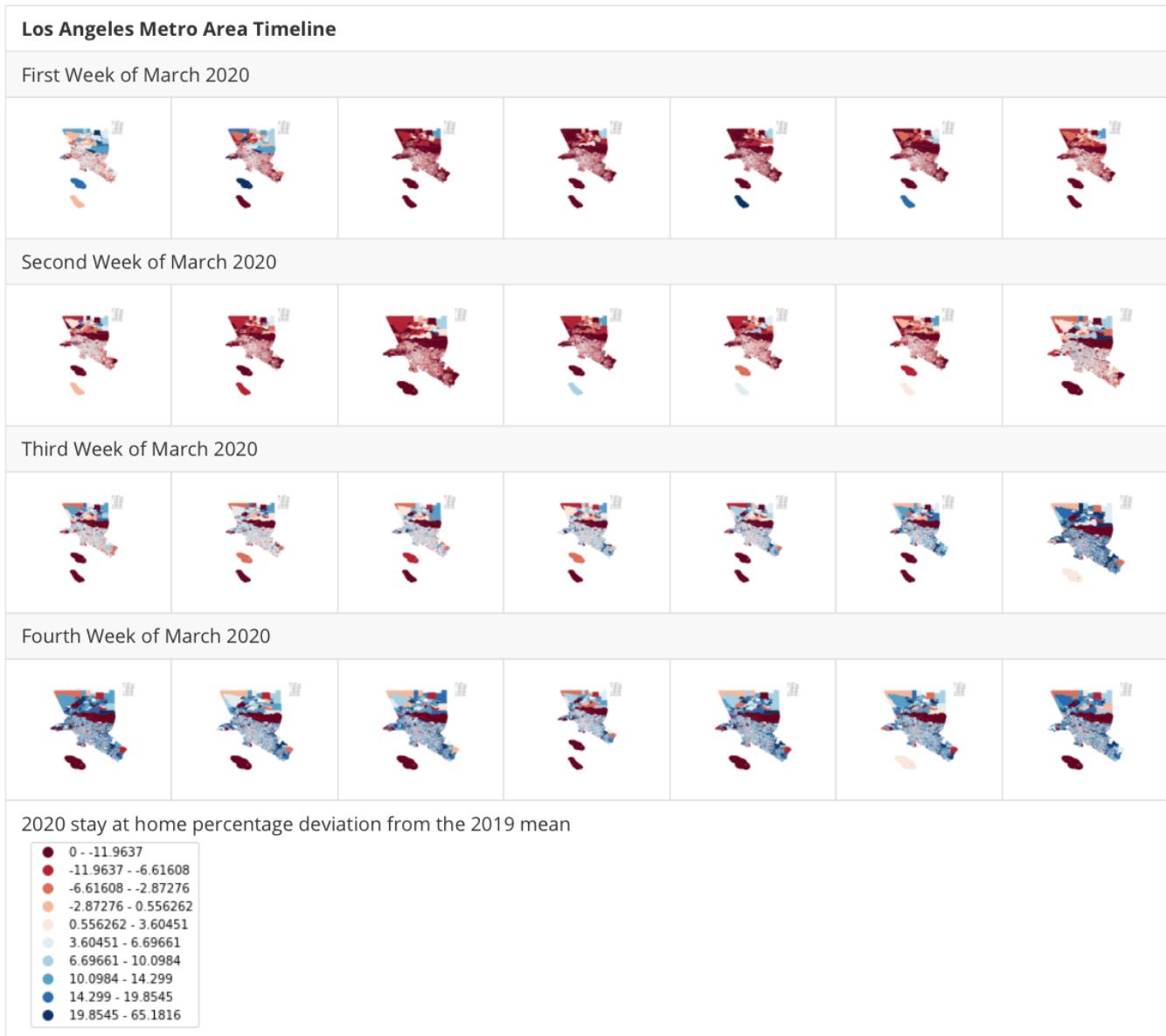


Figure 17: Los Angeles Metro area census block groups daily at home percentage deviation from the 2019 mean for the first four weeks of March 2020.

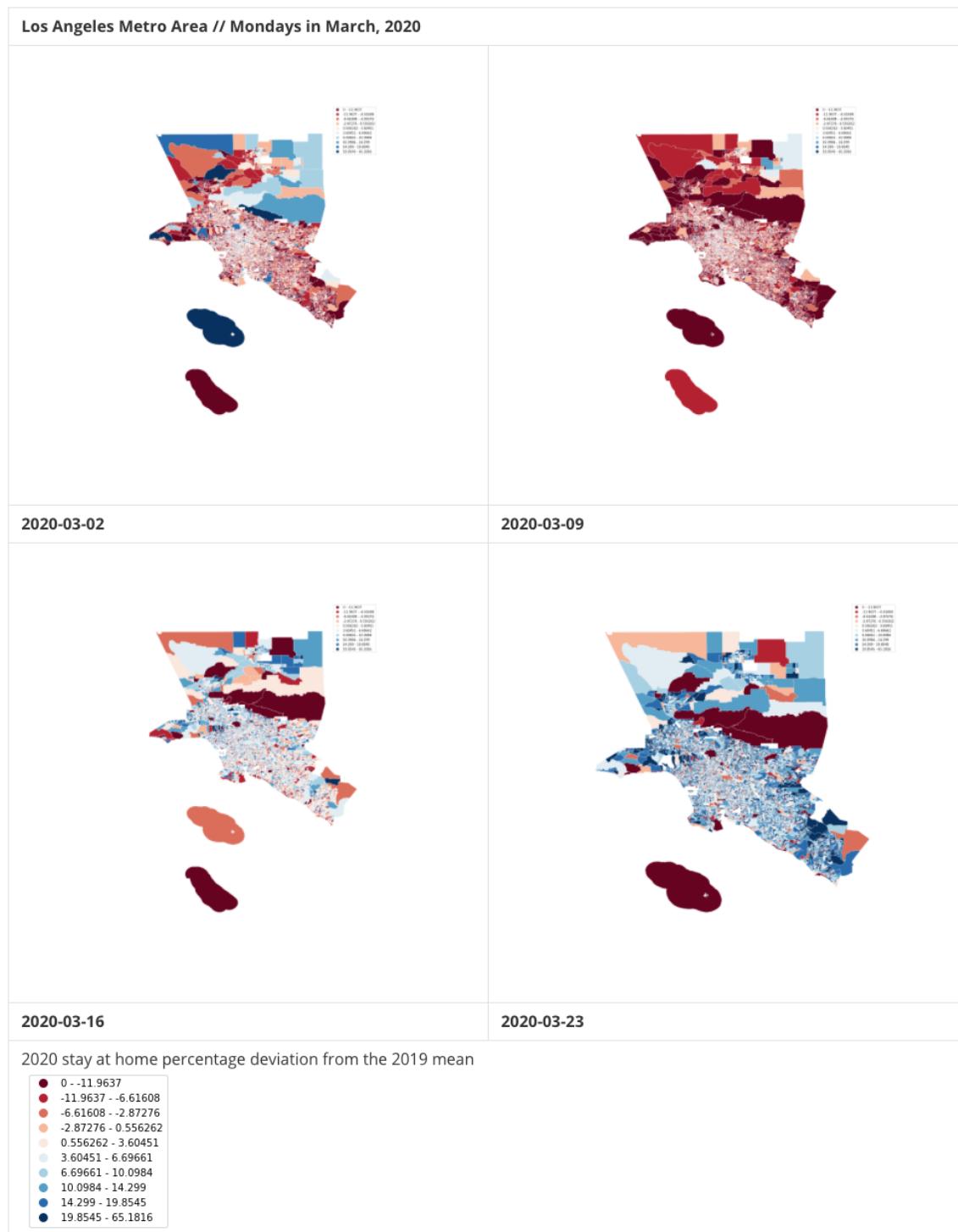


Figure 18: Enlarged view of the Los Angeles Metro area census block groups daily at home percentage deviation from the 2019 mean for the first four Mondays of March 2020.

Visualizing the Mobility Gradient Over Time and in Relation to Poverty

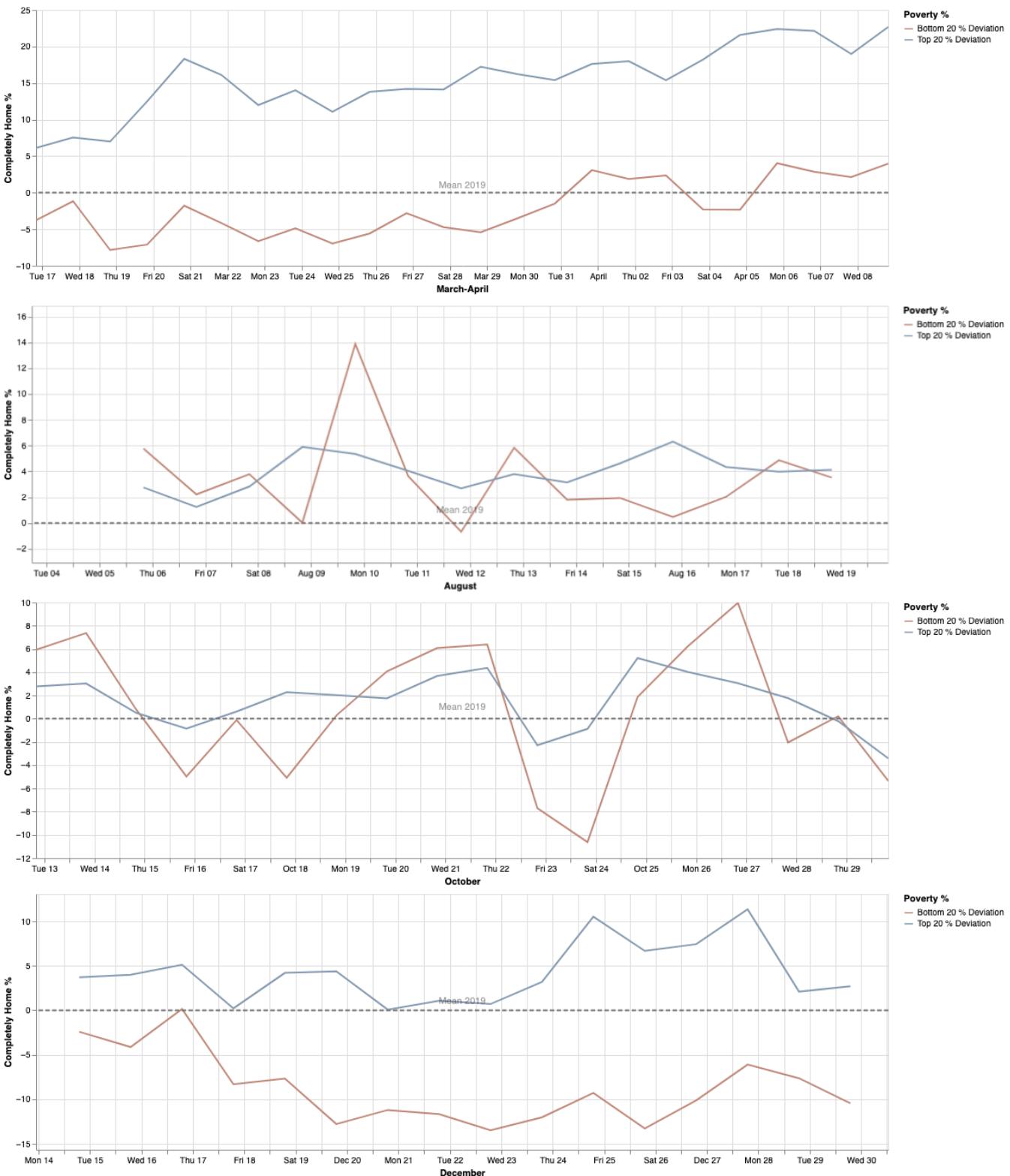


Figure 19: Comparison of Completely Home behavior between the wealthiest and poorest groups for 2020 for specific timelines in Los Angeles: March, April, August & October.

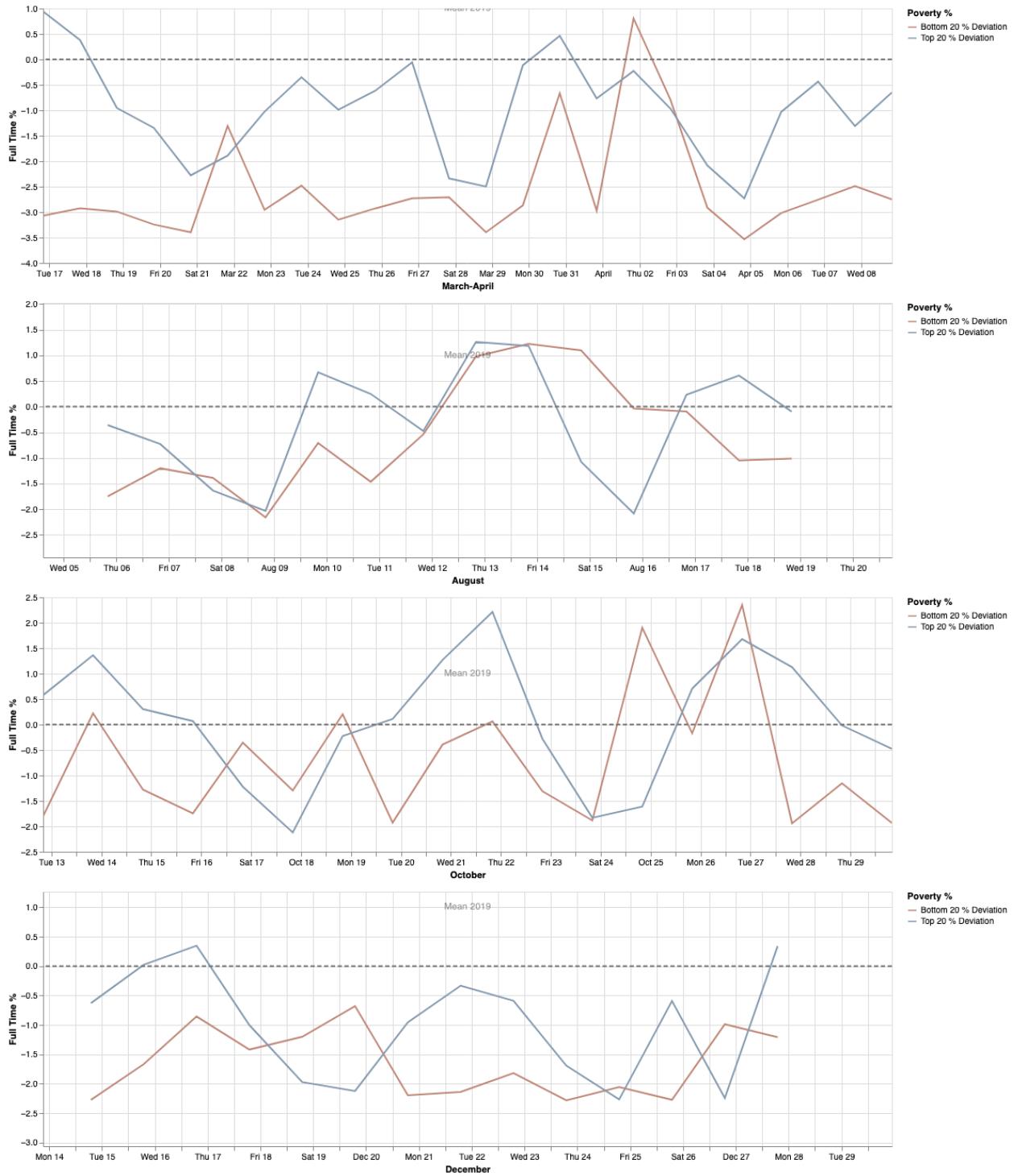


Figure 20: Comparison of Full-time Work behavior between the wealthiest and poorest groups for 2020 for specific timelines in Los Angeles: March, April, August & October.

Visualizing the Mobility Gradient Over Time and in Relation to Poverty

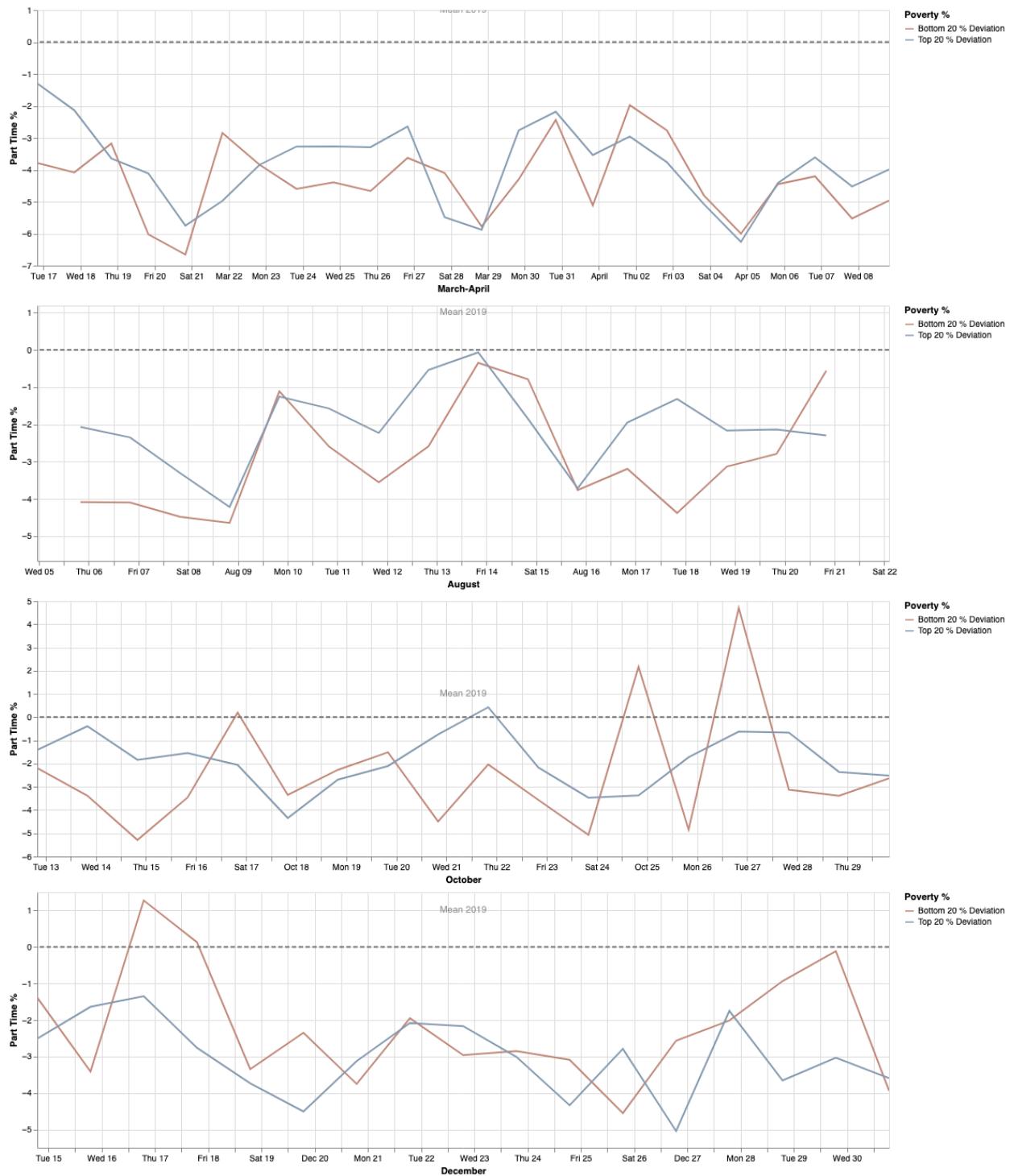


Figure 21: Comparison of part-time Work behavior between the wealthiest and poorest groups for 2020 for specific timelines in Los Angeles: March, April, August & October.

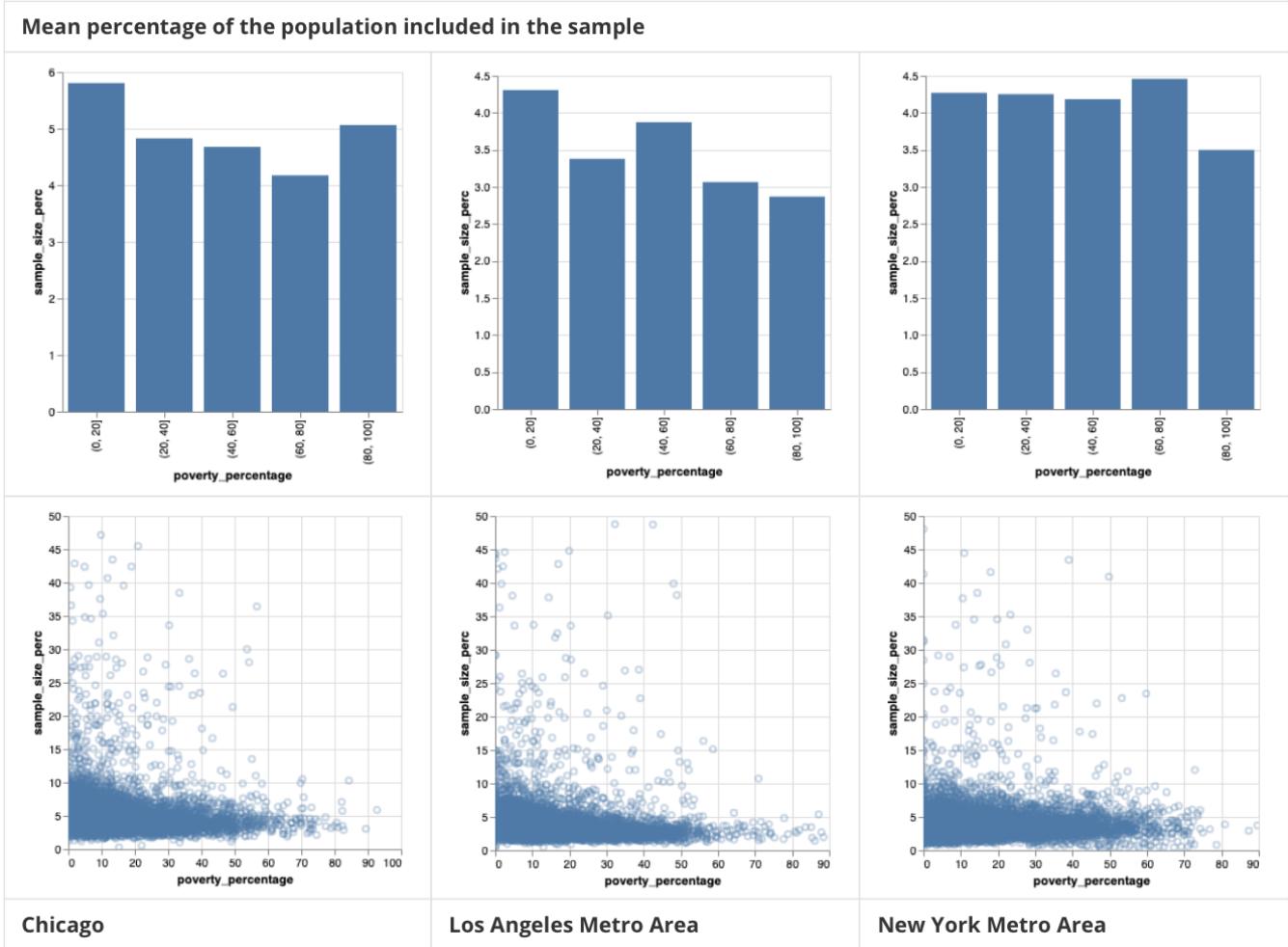


Figure 22: Chicago, Los Angeles and New York mean percentage of the population included in the SafeGraph Social Distancing daily data and percentage of the population in poverty by census block group (bottom) and grouped by quintile (top).

Visualizing the Mobility Gradient Over Time and in Relation to Poverty

..

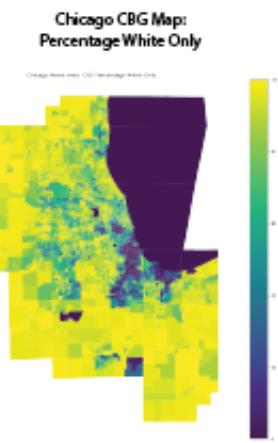
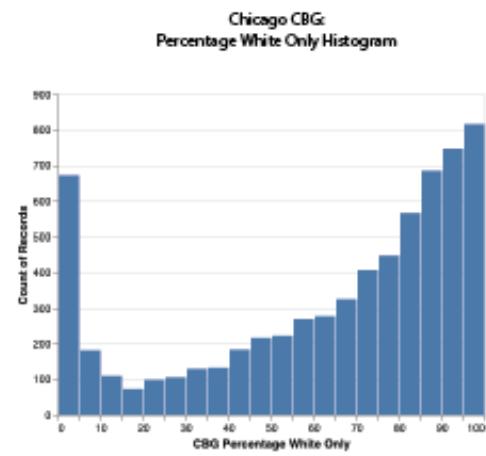
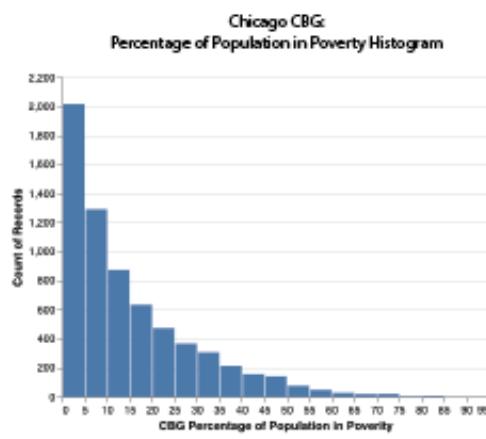
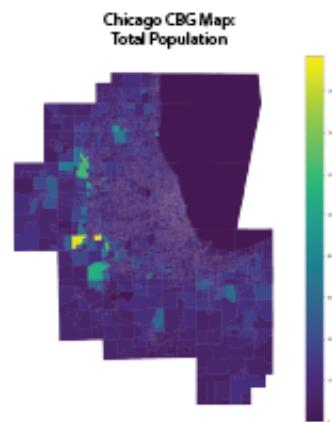
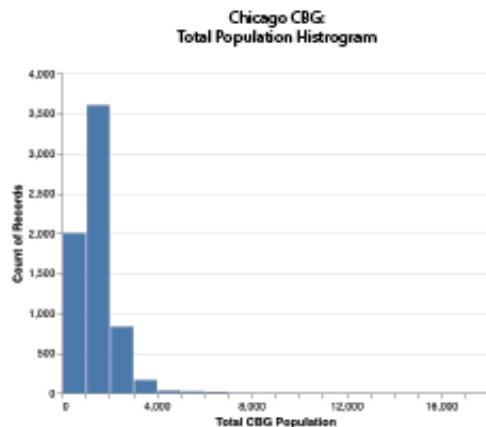


Figure 23: Chicago Metro Area CBG histograms for the total population, percentage below the poverty level and percentage white only from the 2016 American Community Survey 5-year Estimate on the Census Block Group level.

Figure 24: Chicago Metro Area CBG maps for the total population, percentage below the poverty level and percentage white only from the 2016 American Community Survey 5-year Estimate on the Census Block Group level.

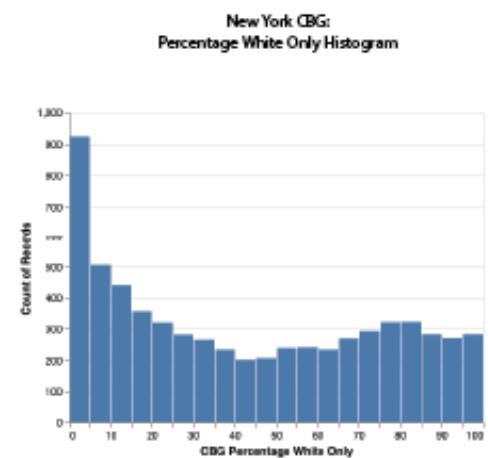
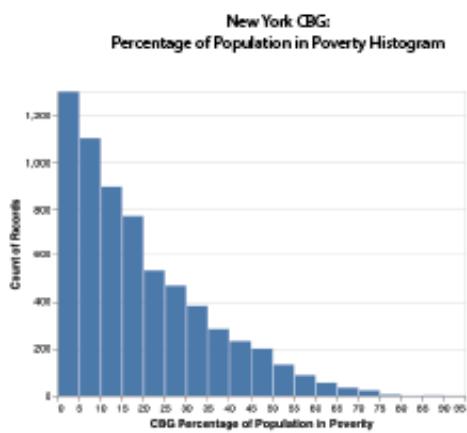
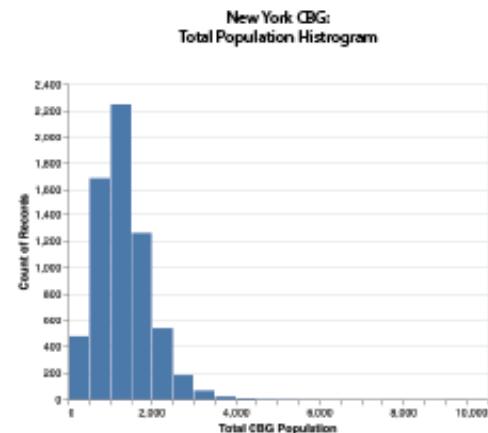


Figure 25: New York Metro Area CBG histograms for the total population, percentage below the poverty level and percentage white only from the 2016 American Community Survey 5-year Estimate on the Census Block Group level.

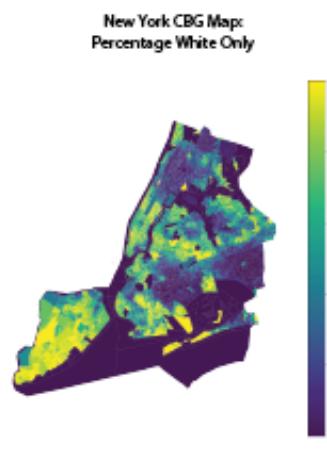
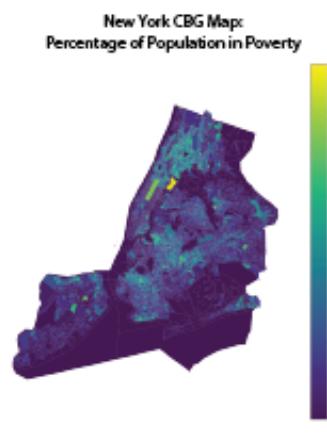
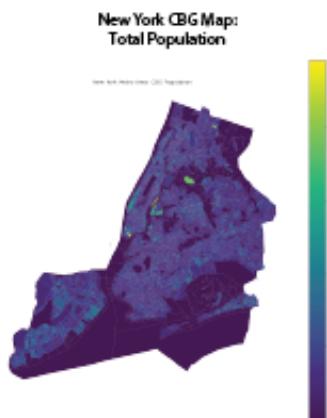


Figure 26: New York Metro Area CBG maps for the total population, percentage below the poverty level and percentage white only from the 2016 American Community Survey 5-year Estimate on the Census Block Group level.

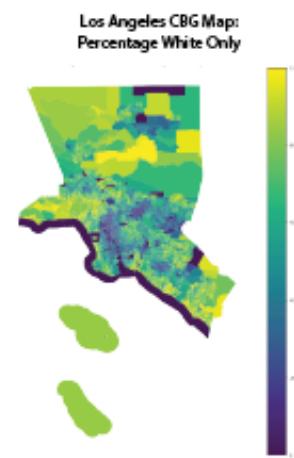
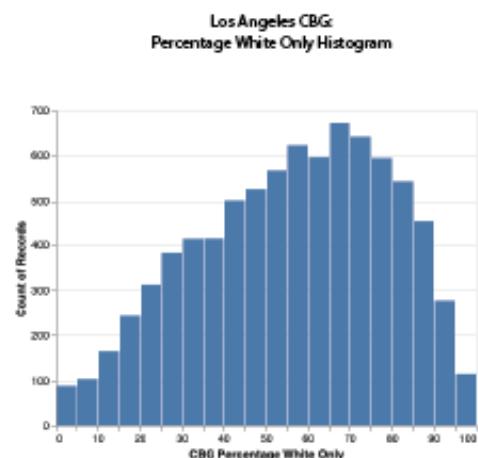
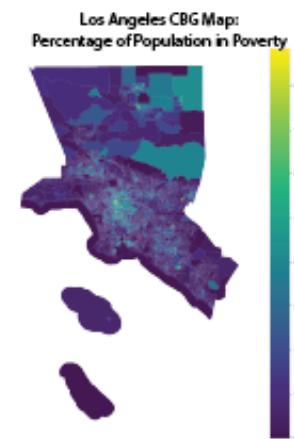
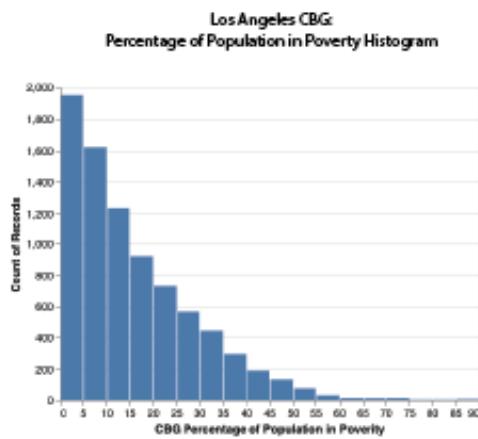
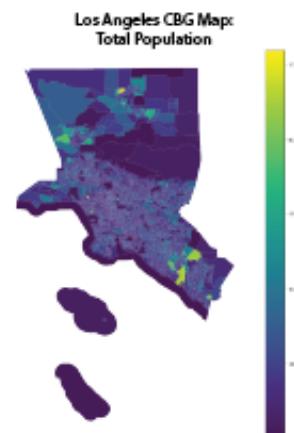
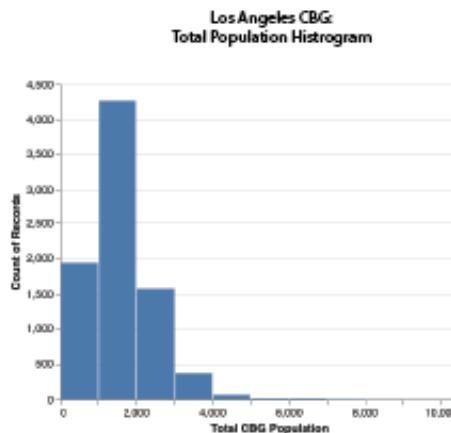


Figure 27: Los Angeles Metro Area CBG histograms for the total population, percentage below the poverty level and percentage white only from the 2016 American Community Survey 5-year Estimate on the Census Block Group level.

Figure 28: Los Angeles Metro Area CBG maps for the total population, percentage below the poverty level and percentage white only from the 2016 American Community Survey 5-year Estimate on the Census Block Group level.

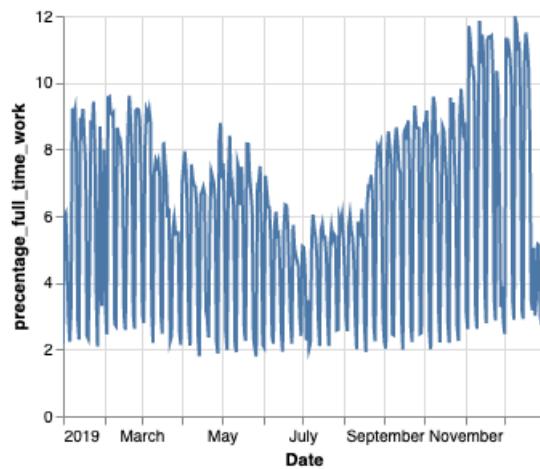
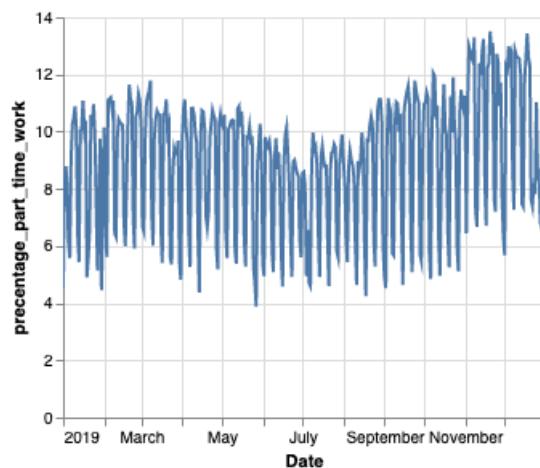
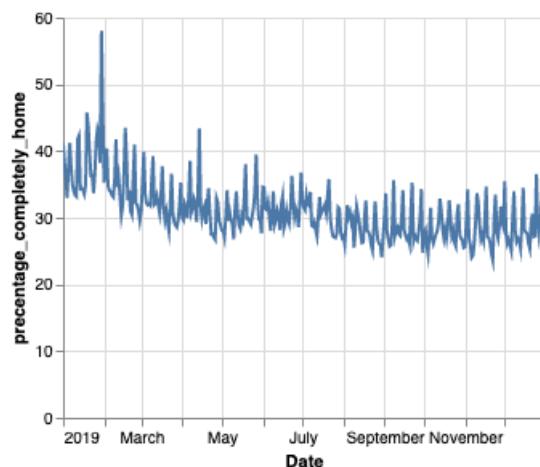


Figure 29: 2019 Chicago Metro Area daily mean percentages calculated using the SafeGraph Social Distancing data set.

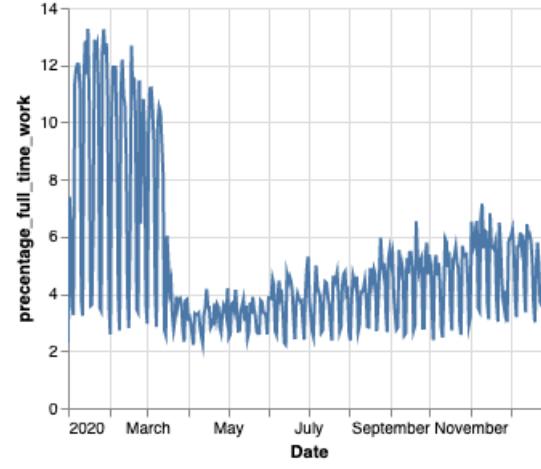
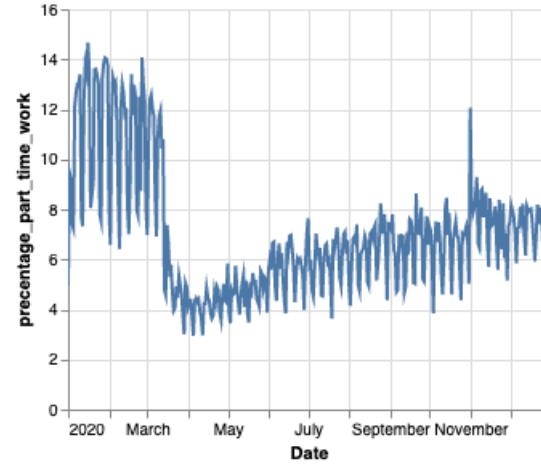
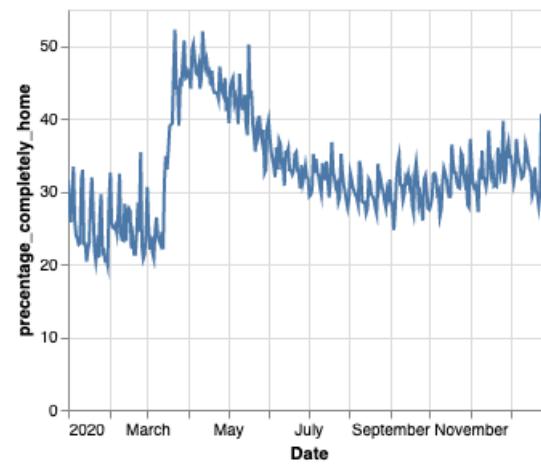


Figure 30: 2020 Chicago Metro Area daily mean percentages calculated using the SafeGraph Social Distancing data set.

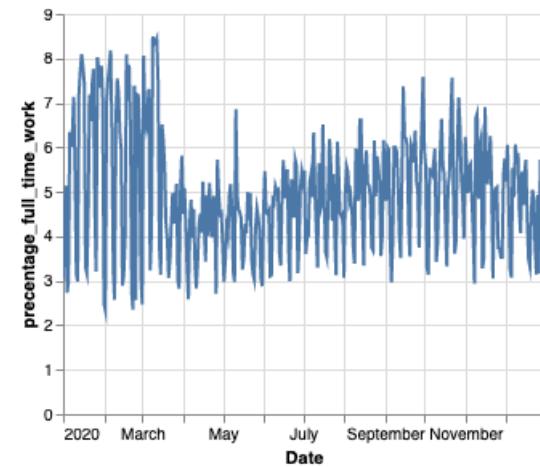
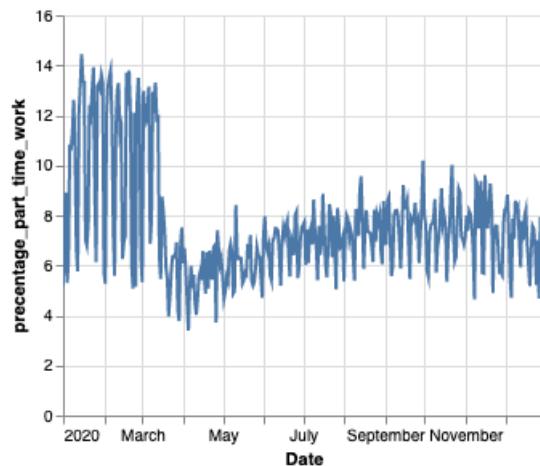
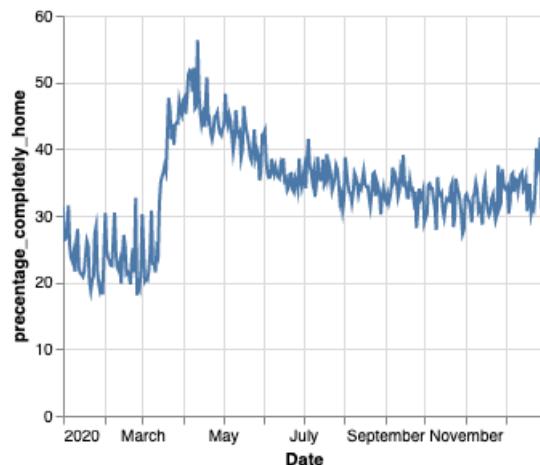
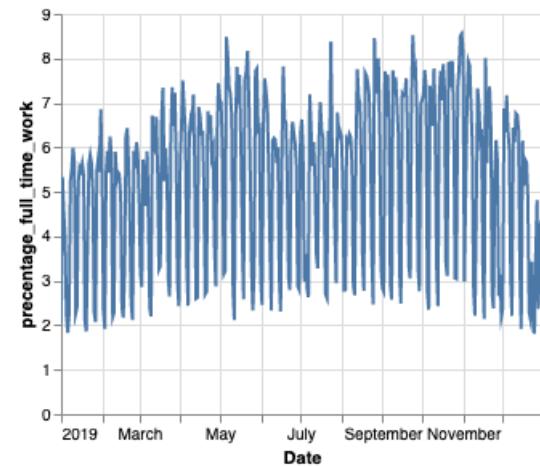
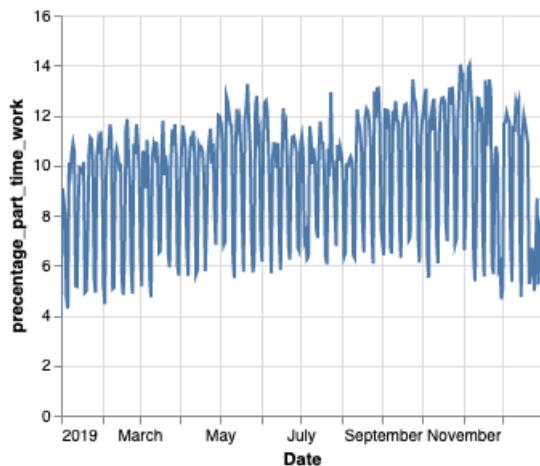
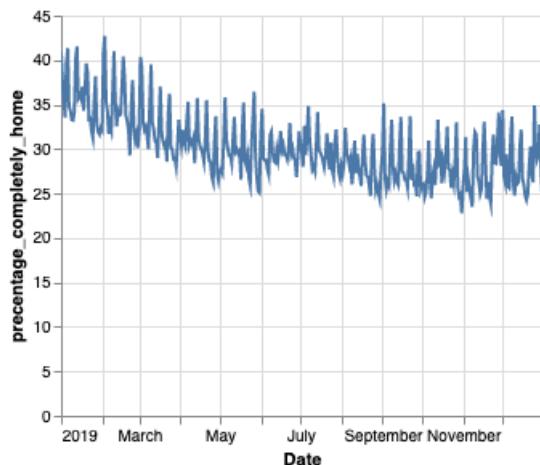


Figure 31: 2019 Los Angeles Metro Area daily mean percentages calculated using the SafeGraph Social Distancing data set.

Figure 32: 2020 Los Angeles Metro Area daily mean percentages calculated using the SafeGraph Social Distancing data set.

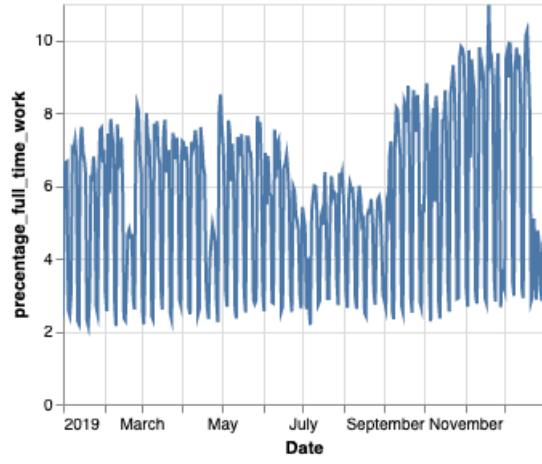
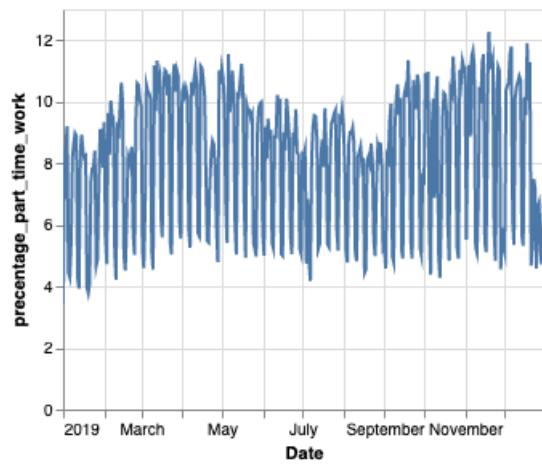
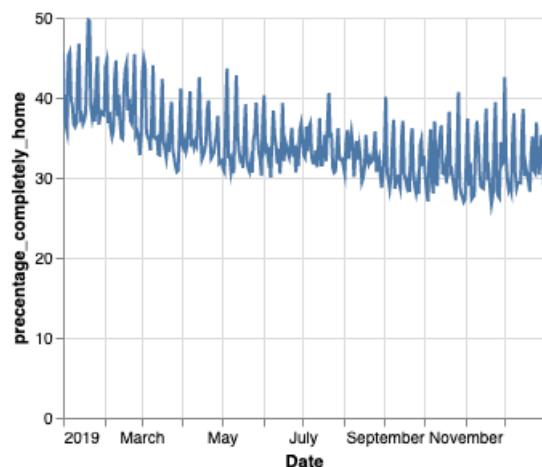


Figure 33: 2019 New York Metro Area daily mean percentages calculated using the SafeGraph Social Distancing data set.

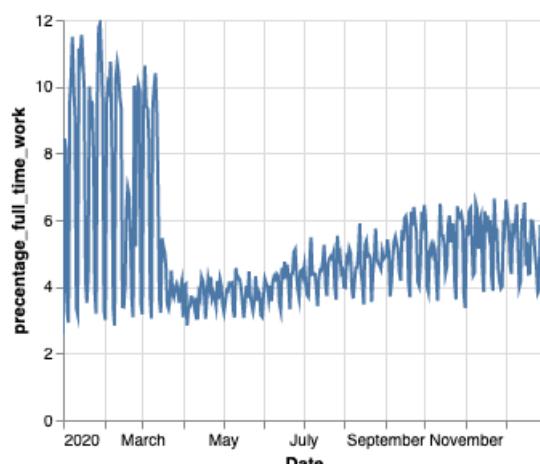
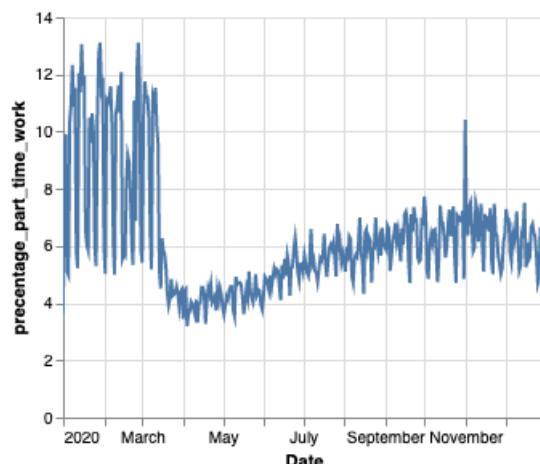
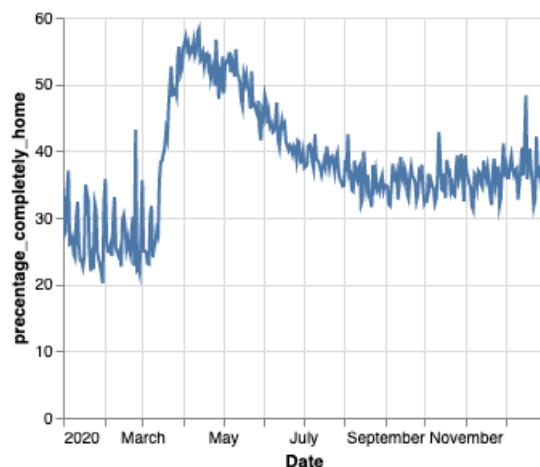


Figure 34: 2020 New York Metro Area daily mean percentages calculated using the SafeGraph Social Distancing data set.

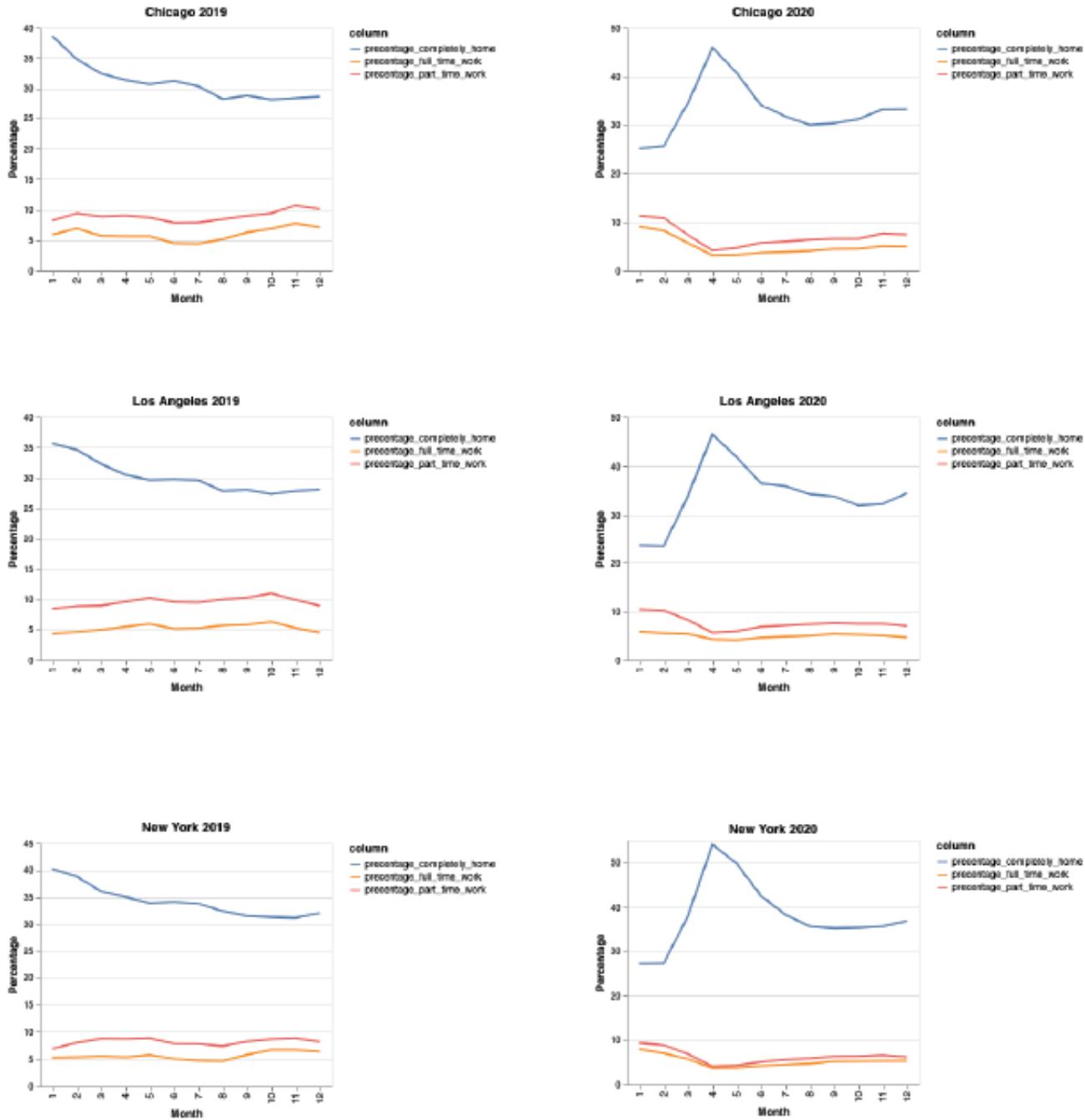


Figure 35: 2019 New York Metro Area monthly mean percentages calculated using the SafeGraph Social Distancing data set.

Figure 36: 2020 New York Metro Area monthly mean percentages calculated using the SafeGraph Social Distancing data set.