Julia Ingram | jmingram
CS 102
Project 1

**Introduction**
This project will analyze when states imposed and lifted stay-at-home or shelter-in-place restrictions in response to the COVID-19 pandemic and when they lifted them against their case counts. It will also examine mobility data to identify disparities, if any exist, in how states' populations adhered to these restrictions.

**Datasets**
- State level data on COVID-19 case and death counts in each state, compiled by The New York Times [Downloaded version updated as of May 15, 2020]
  - Source link: https://github.com/nytimes/covid-19-data/blob/master/us-states.csv
- COVID-19 Community Mobility Reports, compiled by Google. [Downloaded version updated as of May 14, 2020]
  - Source link: https://www.google.com/covid19/mobility/
- Data on state's shelter-in-place orders, compiled by me from The New York Times
  - Source link: https://bit.ly/36daX9d

**Goals**
The overarching question the project seeks to answer is whether there are disparities in states' quarantine rules and behavior. Using the above datasets, this project will seek to explore this through the following questions:

1. Which states were the quickest to enact shelter-in-place orders, with respect to their relative outbreaks?
2. Did cases spike in states that reopened?
3. Did states that enacted stay-at-home orders earlier peak sooner?
4. Which states were more likely to "break quarantine"?
5. Did states with higher mobility have higher rates of infection?
6. Did state stay-at-home policy or mobility behavior differ across party lines?

**Data Processing**

*Data Cleaning*

The first step was to clean and/or compiled the data into my three sheets.

The New York Times' case count data (case-counts.csv) required minimal processing. I did not have to delete any rows, but I deleted the FIPS column since it was unnecessary and reordered the states column to be leftmost, as it is in the other two datasets.

The Community Mobility Reports (mobility-report.csv) dataset was extremely large, because it contains global information and regional information with entries for each date across multiple months. I reduced it to a smaller dataset to make it easier to work with.

The first thing I did was eliminate all rows that contained data on countries other than the United States. Then I eliminated data on the U.S. as a whole, since I was only interested in state-level data, which meant deleting rows without state data. Then, because I am focusing on the state level rather than the county level, I deleted county-level data by deleting all rows that contained information in a second the sub-region field. I did these three steps using filters, opting to display the data I didn't want and selecting all but the header row before deleting it. This reduced the dataset down to 4,336 rows, which was much more manageable than the original dataset.

I then deleted unnecessary columns. I deleted the column that used to contain county names, as well as columns that indicate the data is in the U.S., since all the data is in the U.S.

The data on states' shelter-in-place orders (shelter-in-place.csv) was the most intensive in terms of data processing, because I compiled it myself. I could not find a spreadsheet that displayed the data in the format I wanted it, I made my own using an article from The New York Times. The fields I included were state, a true/false on whether they enacted a stay-at-home order, the data that was enacted, the date that was lifted or set to expire (if there was one stated) and the party of the governor.

Across all three datasets, I standardized the state and date fields, to simplify joining and comparison. This included renaming all columns with states as 'state' and presenting dates in the same format. I chose YYYY-MM-DD.

*Data Analysis*

I performed most of my analysis in a Jupyter notebook using a mix of SQL, python and pandas techniques. I answered most questions in SQL, but modified data and created new columns using lists of dictionaries. When I thought the output of a SQL query would be useful to visualize, I wrote it into a new CSV using pandas to carry over later to Tableau. I created the tables in the following section using SQL queries.

*Data Visualization*

After completing my analysis in Jupyter notebook, I imported into Tableau both the three sheets I started with and some of the new sheets that I had produced in my Jupyter notebook with truncated and joined data. (See description of files used for more detail). I created the visualizations in the following section with Tableau.

**Visualizations**

*Shelter-in-place orders' correlations with outbreaks*

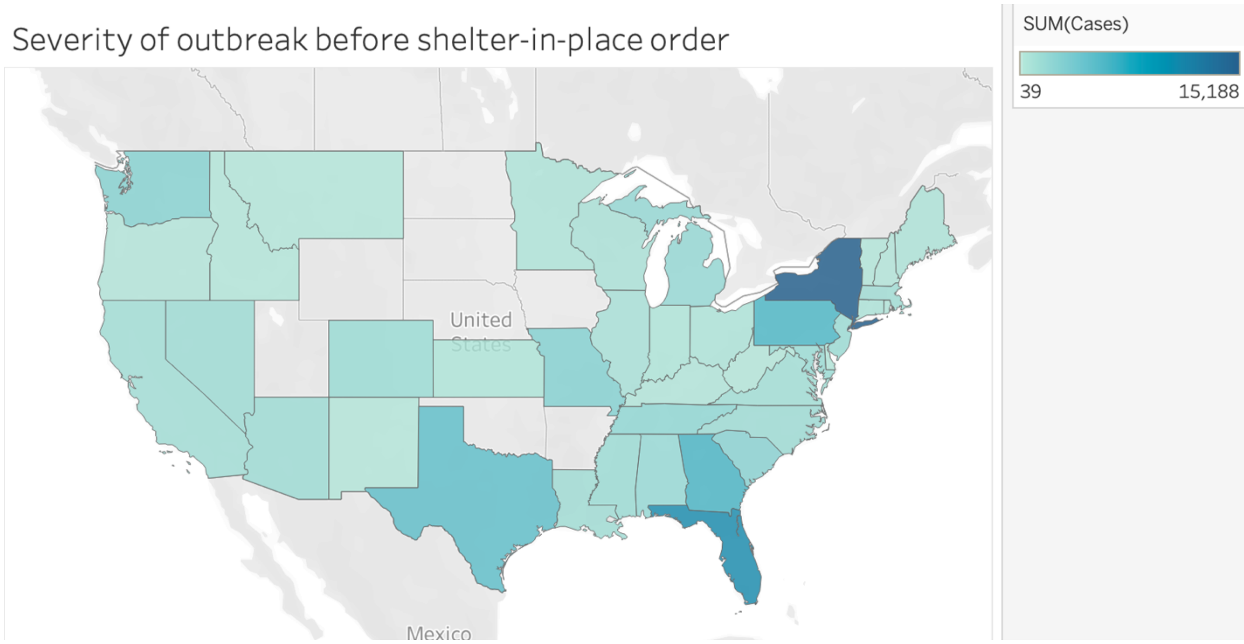

*Figure 1*

States that had the **most** cases on the day they enacted shelter-in-place orders

| state | cases | date |
|---|---|---|
| New York | 15188 | 2020-03-22 |
| Florida | 10260 | 2020-04-03 |
| Georgia | 5967 | 2020-04-03 |
| Pennsylvania | 5805 | 2020-04-01 |
| Texas | 4880 | 2020-04-02 |
| Missouri | 2722 | 2020-04-06 |
| Washington | 2585 | 2020-03-25 |
| South Carolina | 2417 | 2020-04-07 |
| Tennessee | 2049 | 2020-03-31 |
| Michigan | 1791 | 2020-03-24 |

*Figure 2*

States that had the **least** cases on the day they enacted shelter-in-place orders

| state | cases | date |
|---|---|---|
| West Virginia | 39 | 2020-03-24 |
| Hawaii | 95 | 2020-03-25 |
| New Mexico | 100 | 2020-03-24 |
| Alaska | 102 | 2020-03-28 |
| Delaware | 104 | 2020-03-24 |
| Idaho | 123 | 2020-03-25 |
| Vermont | 123 | 2020-03-25 |
| Montana | 147 | 2020-03-28 |
| New Hampshire | 187 | 2020-03-27 |
| Oregon | 191 | 2020-03-23 |

*Figure 3*

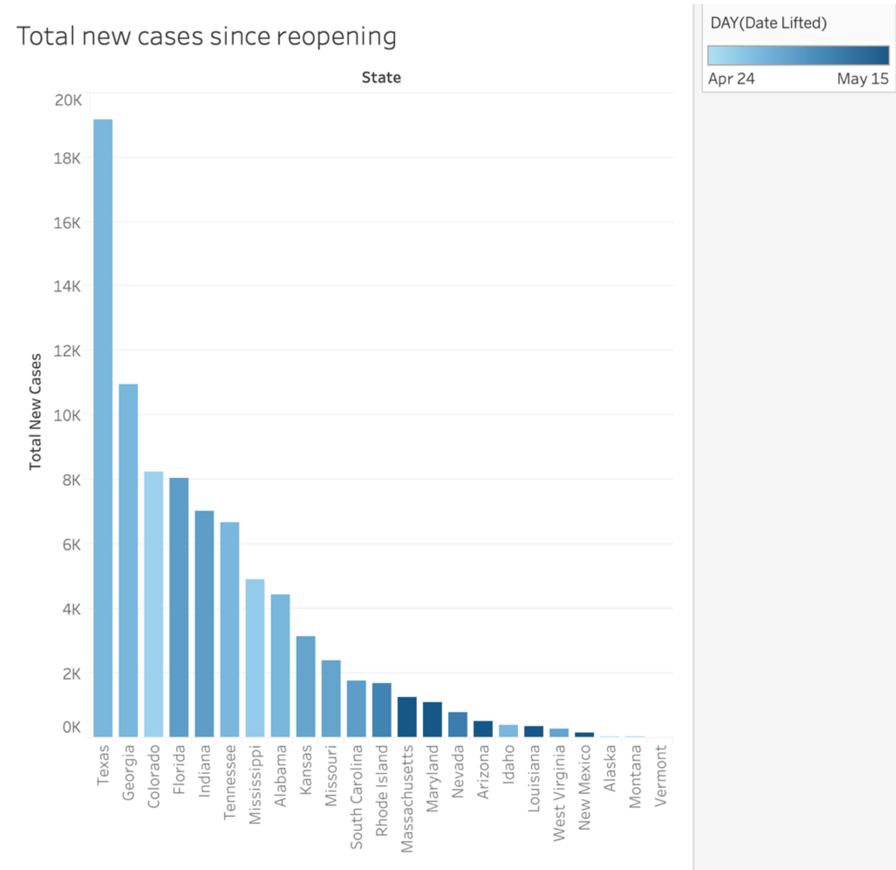


*Figure 4*

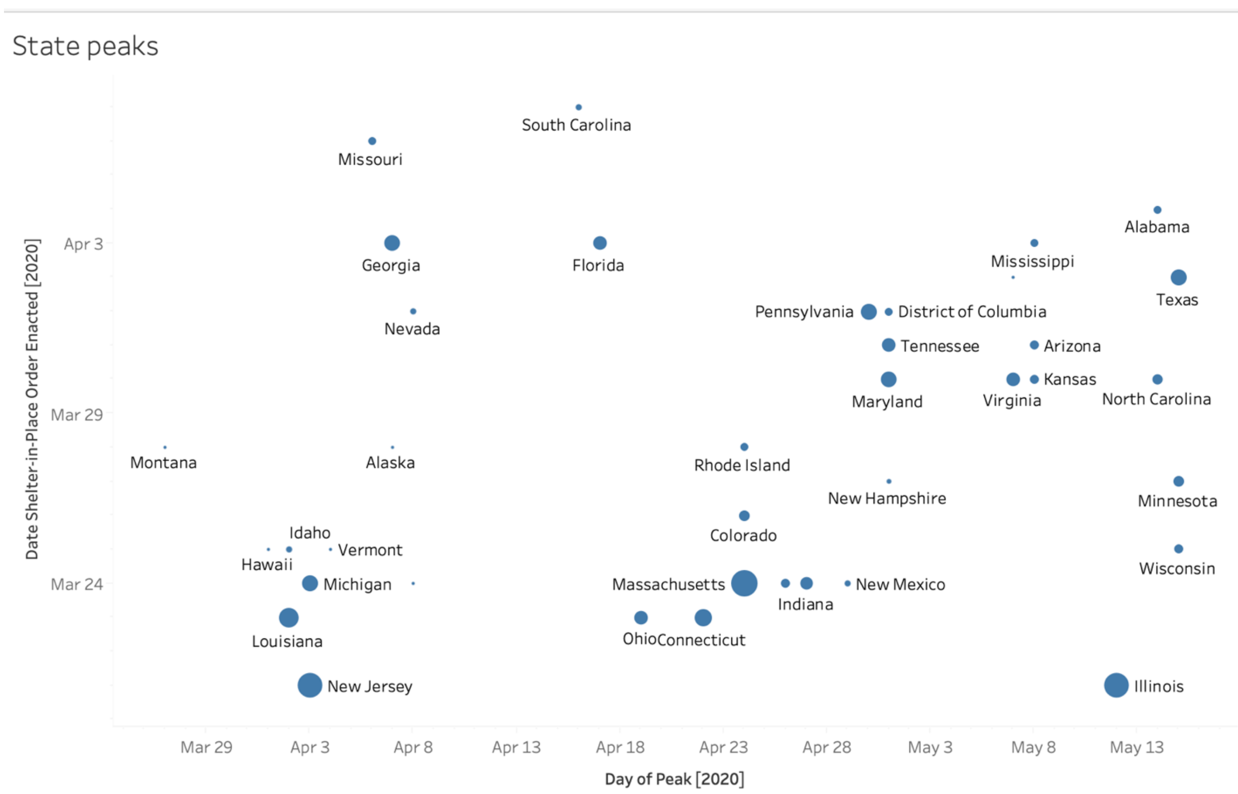


*Figure 5 (Size of point represents magnitude of peak in terms of case count)*

*Mobility trends while under shelter-in*

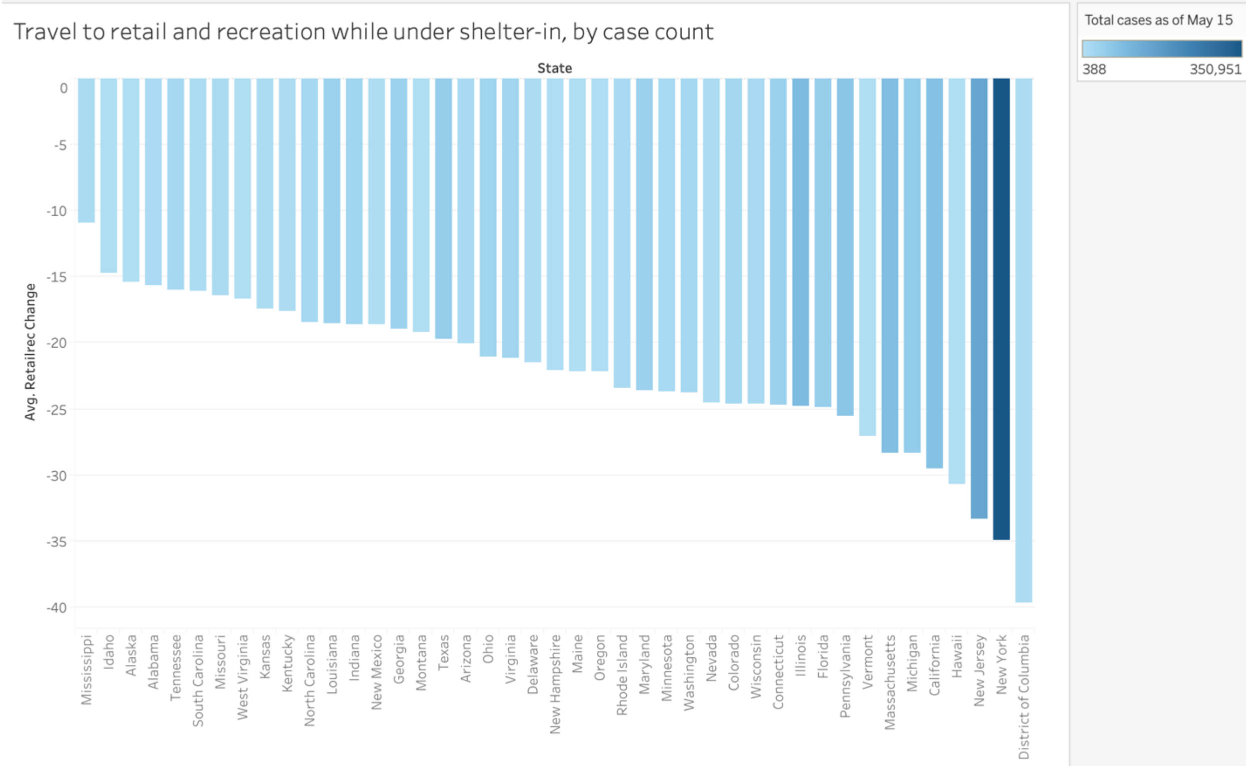


*Figure 6*

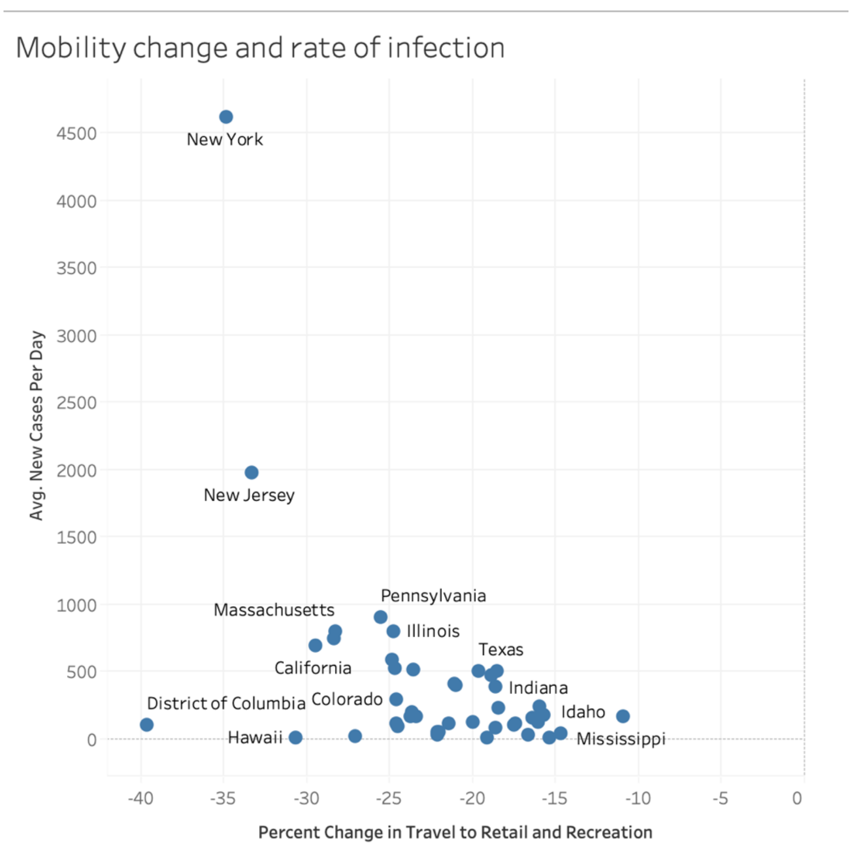


*Figure 7*

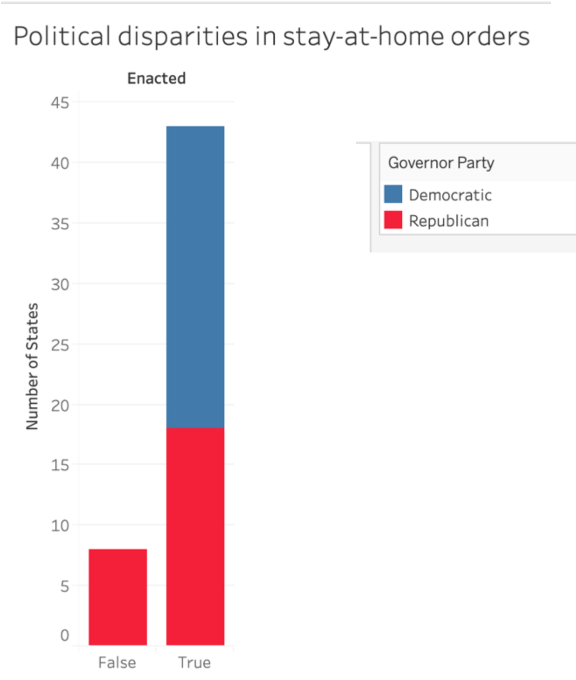


*Figure 8*

Average percent change in travel to retail and recreation while under lockdown

| average percent mobility change to retail and recreation | party of governor |
|---|---|
| -24.5049411765 | Democratic |
| -19.2287581699 | Republican |

*Figure 9*

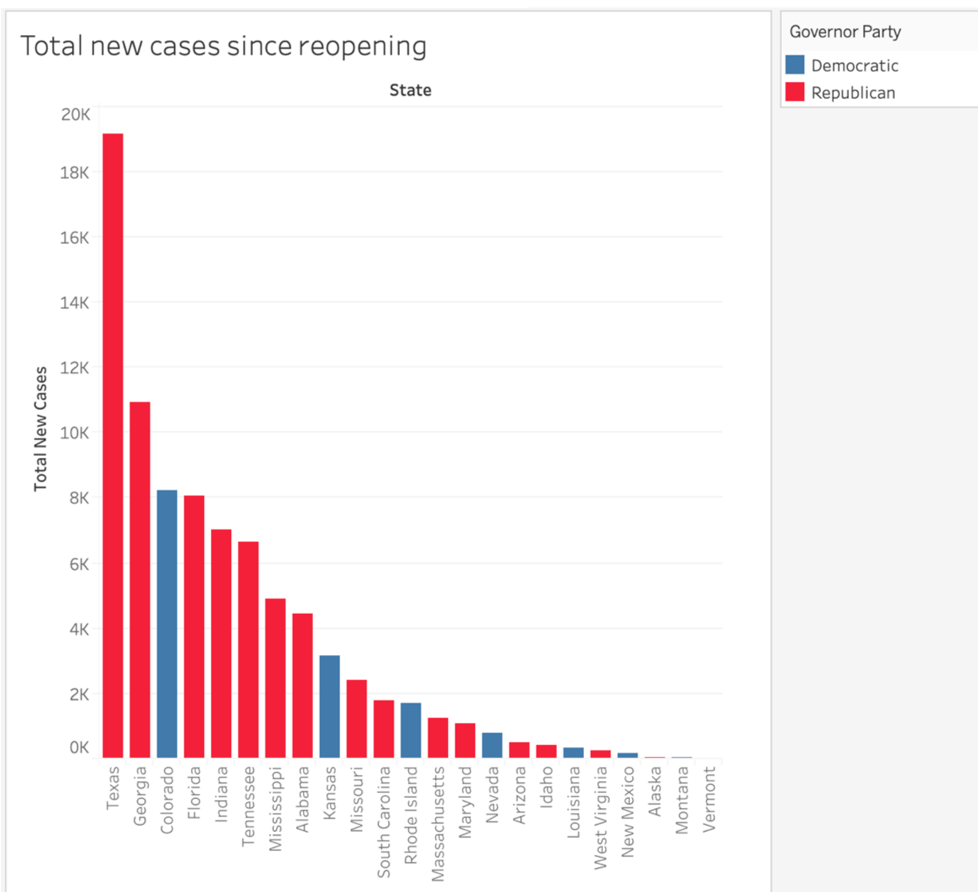


*Figure 10*

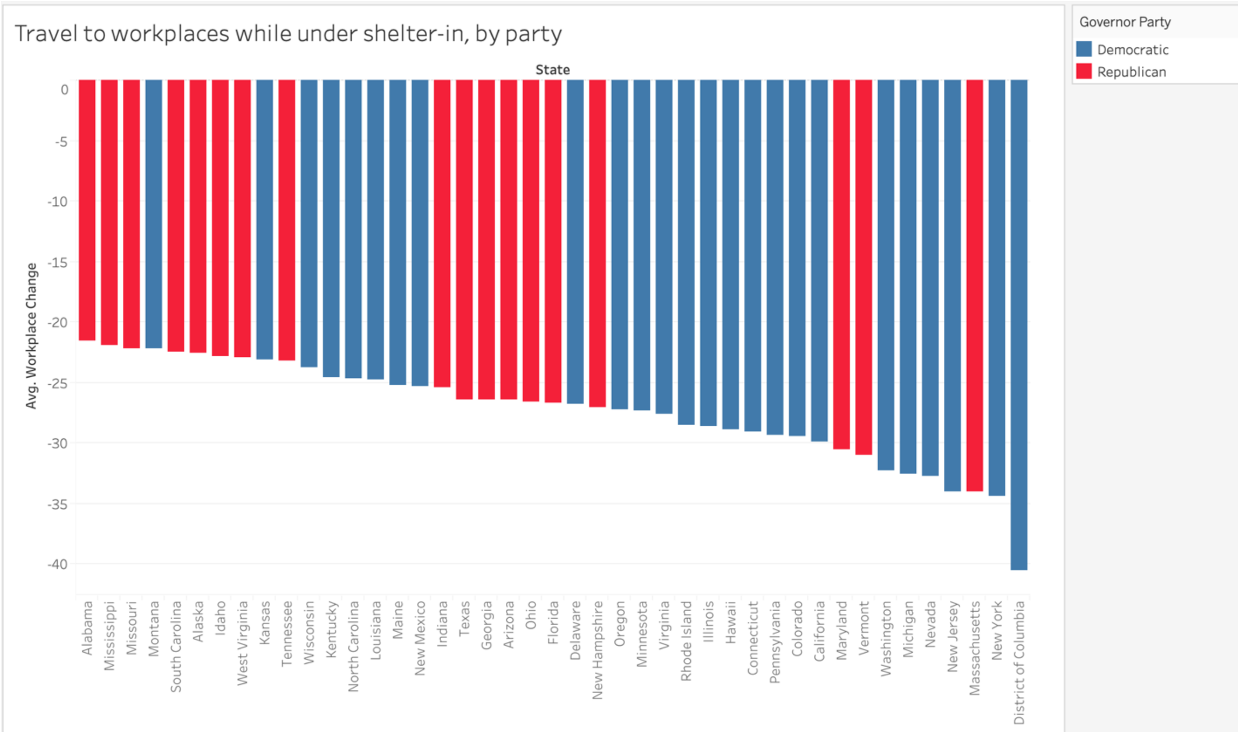


*Figure 11*

**Conclusions**

1. *Which states were the quickest to enact shelter-in-place orders, with respect to their relative outbreaks?*

For this question, I used cumulative number of cases in a state before they enacted a shelter-in-place order as a metric for how "quickly" a state reacted to their relative outbreak. By this measure, West Virginia, Hawaii and New Mexico acted the fastest, while New York, Florida, and Georgia acted the slowest (Figure 2 and Figure 3). All of the top 10 fastest states closed down in late March (Figure 3) while some of the slowest did so in early April (Figure 2). The fastest states were also ones that COVID-19 took longer to reach, suggesting that as COVID-19 rapidly spread in urban or other early hotspots, faster states had the benefit of observing the spread of COVID-19 in other U.S. states, allowing them to get ahead of the virus.

Visualizing the speed of all states against each other (Figure 1), we can see how some of the states that didn't make the top or bottom 10 reacted to their outbreaks. Of note are some states with worst outbreaks as of mid-May, such as California and Illinois, reacted faster than some states that had fewer total cases in their state by mid-May.

2. *Did cases spike in states that reopened?*

Yes, in some states. Cases spiked most dramatically in Texas, which reopened on April 30, 2020, and less so but still significantly in Georgia, which reopened on the same day (Figure 4). There isn't a perfect correlation between how long ago a state reopened and how many new cases have been recorded since their reopening, but states that opened in mid-May, with the exception of Idaho and West Virginia, two very rural states that shut down quickly, have seen fewer new cases since reopening. This is likely due to the short timeframe for which number of cases is measured, however. Other factors are also clearly at play than reopening date; for example, both Montana and Colorado opened on April 26, 2020, but Colorado, a more urban state, has had a much larger growth in cases since reopening.

3. *Did states that enacted stay-at-home orders earlier peak sooner?*

Not necessarily. There is a very weak positive correlation between the date a state enacted a shelter-in-place order and the date it peaked. (Figure 5). The cluster of peaks around May 13, 2020 and after, in which states have shelter-in-place enaction dates spanning from March 21, 2020 to April 4, 2020, shows that an early shutdown did not imply an early peak.

4. *Which states were more likely to "break quarantine"?*

For my analysis of mobility data, I looked exclusively at the percent change from baseline in travel to places of retail and recreation, since these are the locations most dramatically forbade by shelter-in-place orders. By tendency to "break quarantine," I mean having a smaller percent change in mobility when compared to other states. Mississippi, Idaho and Alaska were the top 3 states with the lowest percent change in travel to retail and recreation while under shelter-in-place (Figure 6). However, all three of these states, as well as the those that closely follow them, did not have outbreaks nearly as severe as other states. The states with the top 5 most cases by mid-May were all clustered toward having the greatest percent change in mobility, or the most likely to stay at home. Interestingly, Washington D.C. clearly topped both New York and New Jersey by having the greatest percent change in mobility, despite having a less severe outbreak than both of those states.

5. *Did states with higher mobility have higher rates of infection?*

Roughly, yes There is a medium negative correlation between percent change in mobility to recreation and retail and average number of new cases per day while under shelter-in-place (Figure 7). States with fewer average cases per day had lower percent changes in mobility in many, but not all cases — explaining the outcomes in Figure 6 as discussed above.

6. *Did state stay-at-home policy or mobility behavior differ across party lines?*

For all political analysis, I used the party of the state's governor as a metric for politics, since they are the primary individuals making shelter-in-place decisions. Admittedly, this is not as great of a metric for mobility data, since that measures an entire population, but being that governors are elected officials it still carries weight.

My findings fall in line with a perceived trend in Democrats being more stringent with enacting and following shelter-in-place orders than Republicans. I found that while nearly all states enacted statewide shelter-in-place orders eventually, the 8 states that did not have Republican governors (Figure 8). Of the states that reopened as of May 15, 2020, the states with the highest number of new cases also skew Republican (Figure 10), suggesting they may not have been as cautious with their decision to reopen as some Democratic states were. The mobility data echoed this, but the difference was less dramatic. Residents of Democratic-run states on average had an approximately 5.28% greater change from baseline movement in travel to recreation and retail while under shelter-in-place orders than Republican-run states (Figure 9). This number would presumably be greater if the 8 Republican states that did not enact a statewide a stay-at-home order were factored into this mobility analysis. Looking at the state-by-state breakdown of this data (Figure 11), we see the 7 states with the lowest percent change in movement to retail and recreation are Republican.

It is worth noting that the states with the worst outbreaks — New York, New Jersey, Illinois, and California, for example — all have Democratic governors, which works to explain some Democratic states' orders and mobility numbers suggest they took the situation more seriously.

**Description of Files Used**

*shelter-in-place.csv*
Compiled data on when states enacted shelter-in-place orders.

*case-counts.csv*
Cleaned New York Times state case data. The latest version of the sheet contains an additional column for number of new cases each day as calculated and added in the Jupyter notebook. The original data only had cumulative cases for each state on a given day. Data is up to date as of May 15, 2020.

*mobility-reports.csv*
Cleaned Google Mobility Reports data. Data is up to date as of May 14, 2020.

*project1-notebook.ipynb*
Juypter notebook containing SQL queries, calculations, data modification and manipulation. The following sheets were created from the Juypter notebook from the above three CSV files:

> *preshutdown.csv*
> Contains cumulative case count for each state up until and including the day they enacted their shelter-in-place order. It does not include the 8 states that did not enact statewide orders. Data was drawn from shelter-in-place.csv and case-counts.csv.
> *state-peaks.csv*
> Contains each state's peak, i.e. the maximum number of new cases they had in one day, as well as the date that peak occurred. Data was drawn from case-counts.csv.
> *open-states.csv*
> Contains the total number of new cases in a state since it lifted its shelter-in-place order for each of the 23 states whose shelter-in-place orders were lifted on or before May 15, 2020. It also contains the date the order was lifted. Data was drawn from shelter-in-place.csv and case-counts.csv.
> *quarantine-breakers.csv*

Contains the average percent change from baseline for travel to retail and recreation for each state that enacted a statewide shelter-in-place order. A helper SQL table, called underLockdown, was needed to limit the relevant days to calculate average change in movement to only those in which a state was under a shelter-in-place order. Data was drawn from mobility-report.csv and case-counts.csv, while underLockdown was created by querying a joined case-counts.csv with shelter-in-place.csv.

*projec1-viz.twb*
Tableau workbook where I created the visualizations contained in this report. Most visualizations were created with data from a combination of the above seven CSV files.