The Impact of State Governor's Political Party on the Reopening Policy and the Number of Covid-19 Cases in the US

Nadia Fitriani

Yuwen Xiong

I. Executive Summary

In this project, we aim to analyze the impact of reopening policy on the number of reported Covid-19 cases in the US by state during Mar 2020-Nov 2020. We want to see if state Governor's party will have an impact on his/her decision of reopening policy and the number of new cases in the specific state.

The project contains two parts:

1. Web scraping text and downloading pre-existing datasets. (downloading_and_webscraping.py)
2. Do analysis and draw graphs to show results analysis_and_graph.py)

Our datasets include:

1. Dataset for stay-at-home order expiration dates by state (web scraped 3 websites (see part 3 for detailed explanation) and tidied the data to ensure the accuracy)
2. Dataset for Governor's Party by state (downloaded)
3. Dataset for Covid cases (downloaded and tidied)

Our final dataset used for follow-up analysis is "*data.csv*" in the folder, where we merged and tidied all three datasets above.

We find that most of the states implement reopening policy on mid May 2020 on average which was only two months after most states reported their first Covid-19 cases (early March 2020). After the reopening policy, the trend of the number of new daily cases decreased only until June 2020, then increasing again around July-August 2020, and after October 2020, the number of cases soared, creating a third wave in the US. This finding tells us that reopening does have a positive effect on the number of new Covid cases. It took some time to see the increase of case number as people slowly return to normal life. It also tells us that most states are not ready

for getting back to normal yet. Evaluation needs to be done and Government Executive Order for reopening needs to be cautiously issued.

The coding skills that we used are:

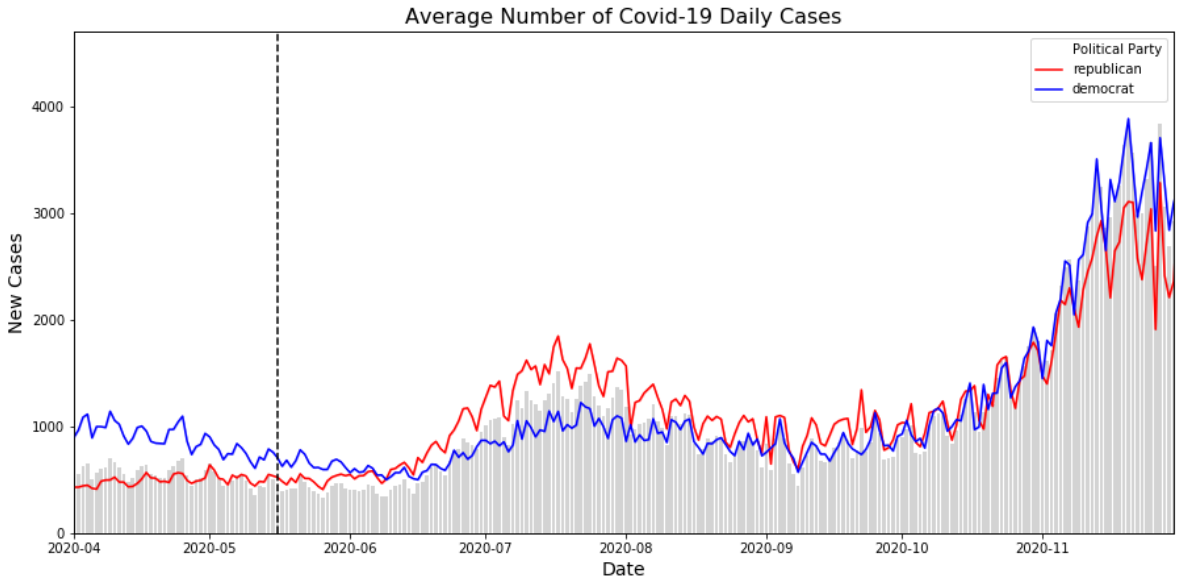
1. Web scraping (a lot & different structured websites)
2. Data cleaning & merging & analysis
3. Create informative and elaborate plots

Room for improvement of this project:

In this project, we did not capture the inconsistencies of reopening policy implemented by each state. We also only use the initial date of reopening policy implementation. Therefore, we believe future study should capture the inconsistencies to understand how the second and the third waves of Covid-19 occur. We need to continue working on more sophisticated skills such as Nature Language Processing to capture all types of reopen policy and its accurate dates. In that case, we might be able to see how different states react to the increase of new cases and see whether party plays a role in it. We also only use naïve before-and-after estimator to understand the change after reopening policy implementation which does not reflect the counterfactuals. We believe that future study can use event study to better understand the impact of the policy. Moreover, we only took into account the political party, but other important factors are missing in our project, such as population, state industry, etc. Those can be added to further analysis.

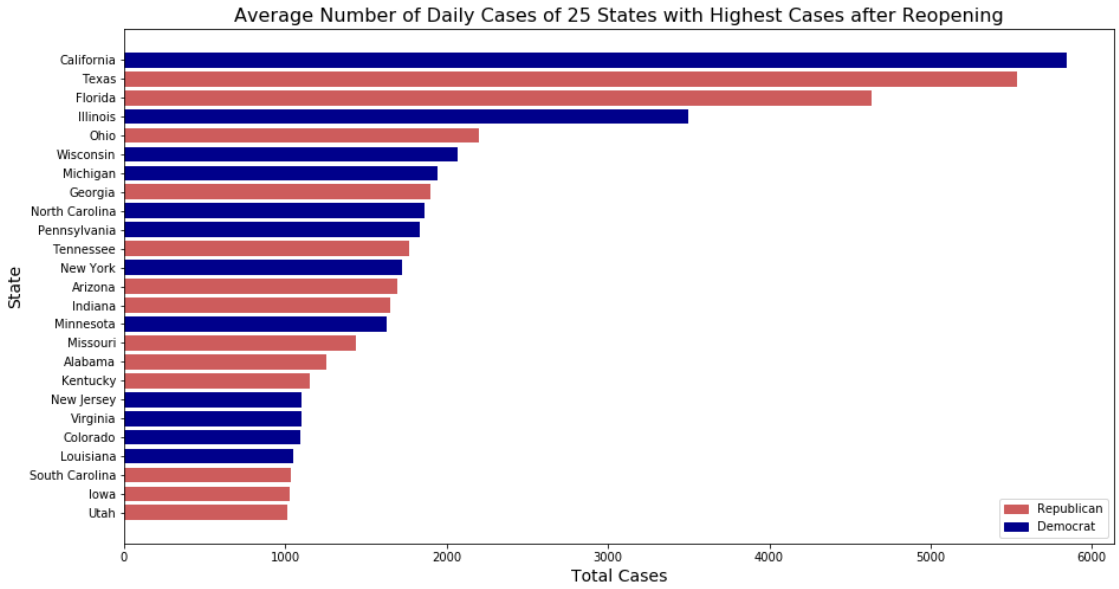
II. Results

Graph 1. Average Number of Daily Cases in All States by Governors' Affiliated Party



As we have mentioned earlier, we observed three waves of Covid-19 daily new cases in all US' states. We try to understand whether the number of cases between states differ based on the Governors' affiliated parties in which divided into Republican (red-states) and Democrat (blue-states). As can be observed from Graph 1 above, all states have similar trend on the number of new daily Covid-19 cases throughout the period. Observe that during the first and third waves, blue-states have relatively higher number of Covid-19 cases. However, on the second wave, red-states had relatively higher number of new cases. It implies that it is unclear whether political views affect Covid mitigation.

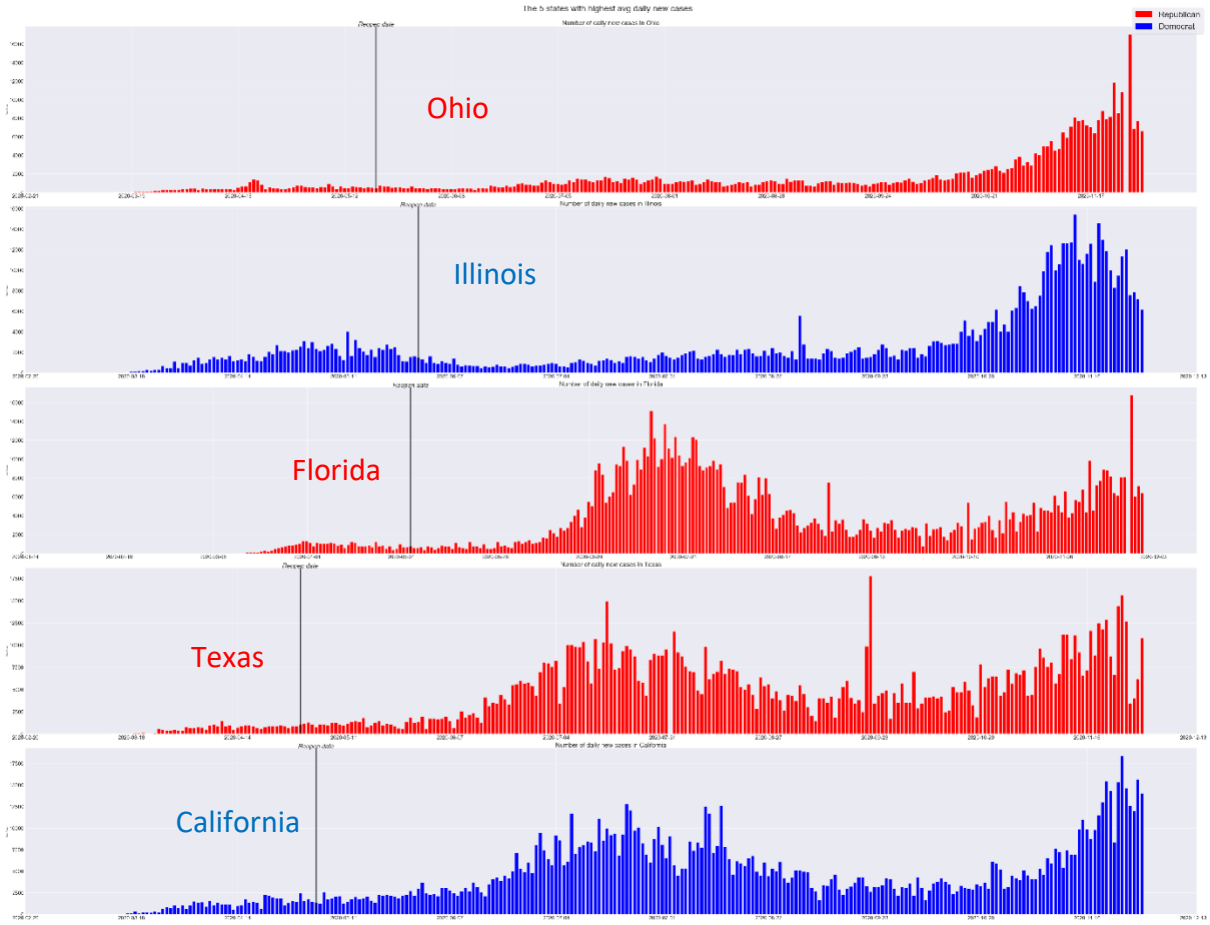
Graph 2. Average Number of Daily Cases in Top 25 States with Highest Number of Cases



We then rank each states based on the average number of new daily cases after reopening policy implemented at each state and come up with Graph 2. As can be seen, there are three red-states and two blue states with the highest number of new daily cases among the highest 5. Out of the top 10 states, there are 6 blue-states. Out of top 25 states, there are 13 red-states and 12 blue states. This finding supports our previous claim that the effects of political views are unclear in mitigating Covid-19 in the US.

In order to see the trend for each state more clearly, we create Graph 3 including the top 5 states with the highest number of Covid cases after reopening. The graph is in ascending order. The last graph is for the top state who has the largest number of average daily new cases after reopen, which is California. Color indicates the party of the current State Governor.

Graph 3 Top 5 States with the Highest Number of Average Daily New Cases

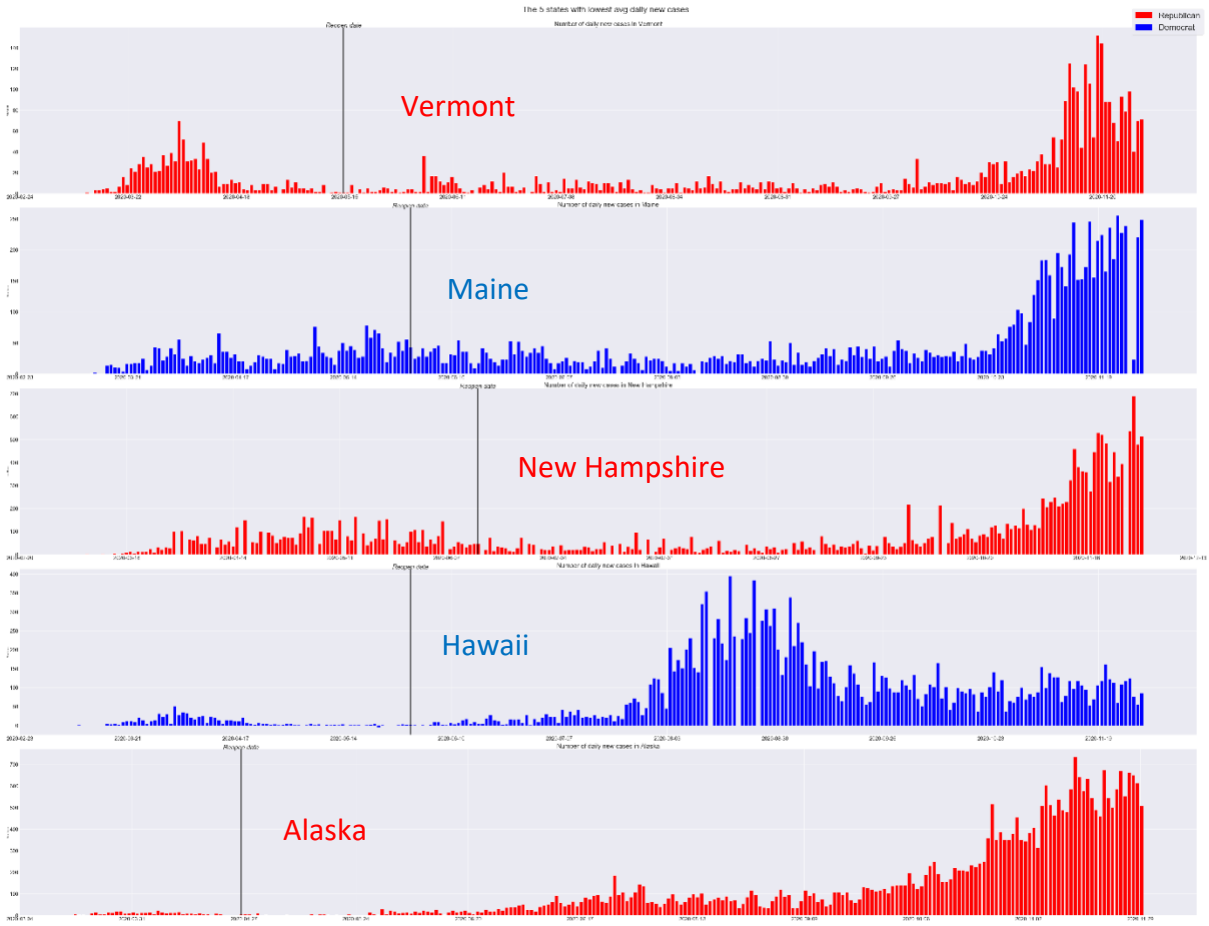


From this graph, we can see that the top three states, California, Texas and Florida have a huge case increase two months after the state is reopen. Ohio and Illinois have a much better control. But starting from October, situation is getting worse again in all top 5 states. There are actually

many reasons that can explain this situation, including location, how much economic activities those states have, the main industry, the size and the population of the states, and people' attitudes toward self-quarantine, etc. The top 3 states are actually the states with the highest population. Illinois and Ohio rank 6 and 7 respectively.

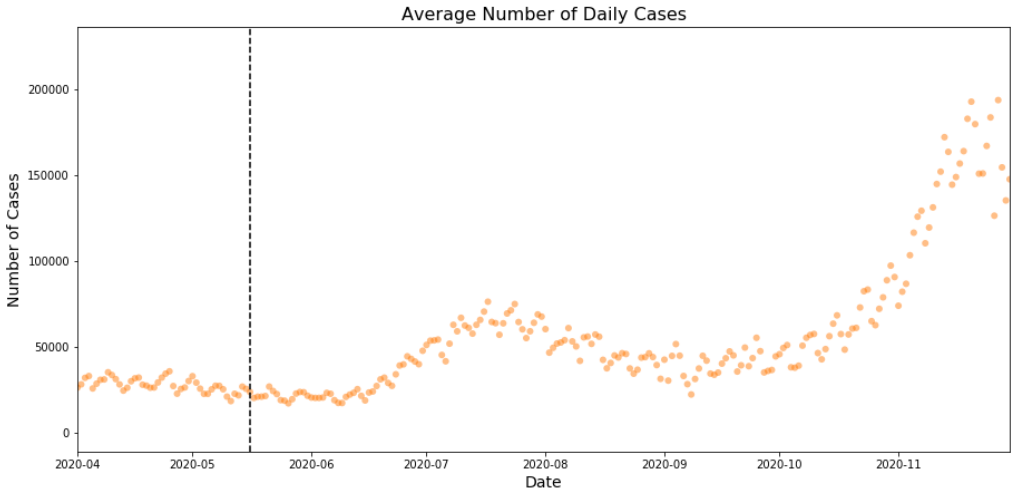
Since we already see the top 5 states, we also create a graph that show the 5 states with the lease average number of new cases. It is also in the ascending order. Vermont is the lowest among all states. Color indicates the party of the current State Governor.

Graph 4 Top 5 States with the Lowest Number of Average Daily New Cases



From this graph, we can see that those 5 lowest states have a fairly good control after reopen, except Hawaii experienced a huge increase one month after reopen. But Hawaii is not experiencing the increase after October while other four states both have a huge increase in the number of new cases.

Graph 5. Average Number of Daily Cases for All States



We use naïve before-and-after estimator to estimate the impact of reopening policy at all states using the mean date as the period of when the policy was implemented. We found and increase of 50,702 new daily cases on average after the reopening policy was implemented.

III. Explanation of Working Process

1. Web scraping text and downloading pre-existing datasets.

We scrapped text to get the date in which reopening policy was implemented. Here, we first need to define the reopening policy. Since each state has different policies/executive orders/phases, and the definition of reopening varies a lot across people, it is hard to find a general rule that tells us the extent of reopen in one state. Thus, instead of looking at the reopening policy directly, we paid attention to the date when the stay-at-home order expires. The expiration of the stay-at-home order indicates that people are not strictly quarantined at home. Instead, they can go to public places and resume regular social activities while still remain 6-feet social distance. It indicates that the economy of the state is generally on the track of getting back to normal. Most states have stay-at-home order that expired from Apr to Jun. Thus, it is more convenient to do the comparison as most stay-at-home orders are similar to each other.

We checked three websites to get the accurate results for the expiration of the stay-at-home order: HUSCHBLACKWELL, NGA and MULTISTATE. We first tried Huschblackwell, as it has very detailed explanation by date about how the reopen progressed in each state. However, it is very complicated to get the accurate reopen date due to the diverse definition of reopen

and the complicatedness of the text on this website. We tried to use NLP and keyword index, but found it is even hard to include all keywords that indicates reopen or stay-at-home order. Thus, we decided to jump to other sources. We find NGA but too many states' dates are missing. Thus, we changed to Multistate. It is very clearly organized and relatively easy to scrape. It worked pretty well, though there are a few states that do not have data. We manually checked their official websites, news and information on Huschblackwell and input those dates. In terms of data regarding Covid cases and Governor's political party, there are pre-existing dataset on the internet. Thus, we downloaded and tidied the data. Then we merged the two datasets with the dates that we generated by web scraping.

2. Do analysis and draw graphs to show results

After getting all the datasets we need, we started to do analysis and created graphs. It is easier to calculate average and do boxplot to make comparison. But we still think that simple average wastes all the necessary information we have. Thus, besides presenting the overall trend in Graph 1, we decided to keep time series data and show the trend for each states separately. We first take a look at the top 25 states, and then review states rank top 5 and bottom 5 average daily new cases. We expected to see political party playing a role in the results. But it didn't show up as what we expected. There might be other more important factors that we didn't take into account. But by creating graph for those ten states, we still find some meaningful results. At the end, we want to see if there is any change in the number of new cases after the reopen policy is implemented. Thus we create a naïve estimator to show the results. It might not be the best way, but it clearly shows that reopening does increase the number of new cases in all states in the US.

IV. Conclusion

The reopening policy increase the observed number of new daily cases on average at all States. It is unclear whether political view affect the mitigation of Covid-19 in the US. There might be other factors that have a more direct impact on the number of new Covid cases that we did not take into account in this project.