**Analyzing COVID-19 Spread in Relation to State-Enforced Mask Mandates Using Julia**

Wendy Zhu, Tue Do, Costa Nikols

## Purpose & Hypothesis

The purpose of this project was to use the Julia programming language to organize, analyze, and present data. We chose to examine COVID-19 data from the Johns Hopkins Data Hub. Using Julia, our goal was to look at the trends in COVID cases per day in different states and compare those numbers between states that passed mask mandates and those that did not. Prior to completing our analysis, we expected to find that states that *did* have a mask mandate would show a significantly slower rate of spread and lower cases per capita overall than those that did not have a mask mandate.

## Method/Code

For this analysis, we mainly used tools from the *StatsPlots* and *DataFrames* Julia packages. We found them very analogous to the *pandas* package from Python, and the language was very intuitive given prior experience with Python.

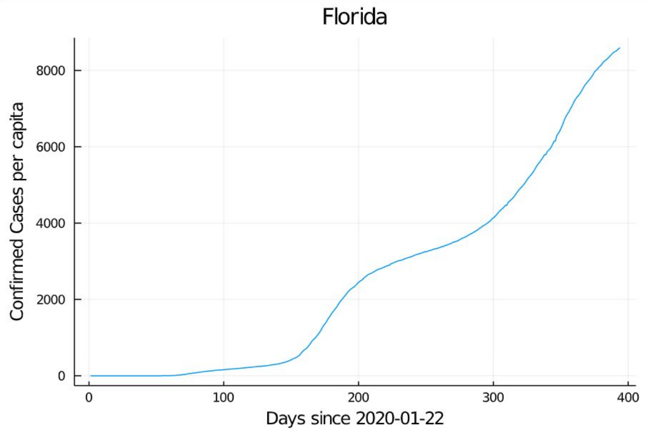
The CSV file we used for our analysis of confirmed cases came from John Hopkins's Data Hub, and contained daily data for each county in the United States for confirmed cases and deaths. We chose to focus on confirmed cases per state, and the first part of our data analysis involved shaping the data for it to be more easily workable for our purposes.

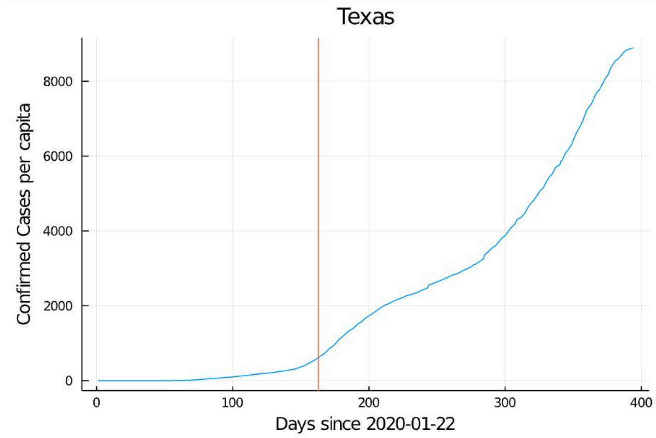
While the syntax isn't identical to *pandas*, example code that we were able to find analyzing other datasets had a similar feel, and it wasn't too much trouble to create DataFrames to suit our purposes. We first eliminated columns unrelated to our analysis, including county names and fips codes associated with them. We then split our DataFrame into a 51-part Group DataFrame, in order to sum up cases over all counties in a state for each day. A particularly neat bit involved writing a generator to print code to create new columns for each state.

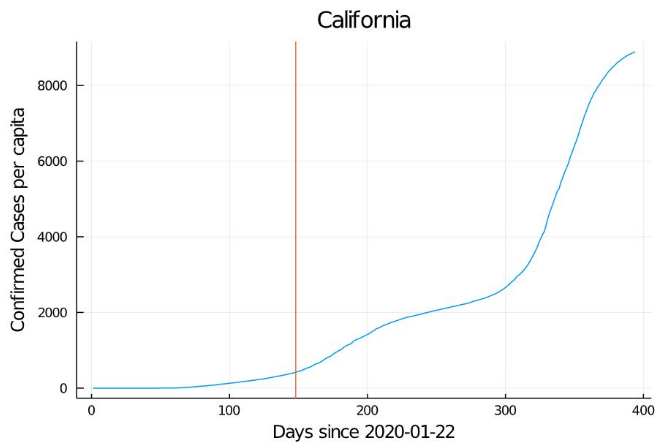
We then used a different CSV file, also from John Hopkins, to get a table for dates of mask mandates in public space, which varied from state to state. To relate the two datasets, we used a handy *Dates* Julia package to convert our data into an integer number of days after January 22, 2020, which was the first day of data that the other CSV file considered. You will see in the graphs below that for states that enforced mask mandates, a vertical line is drawn at the date of an announced mask mandate.

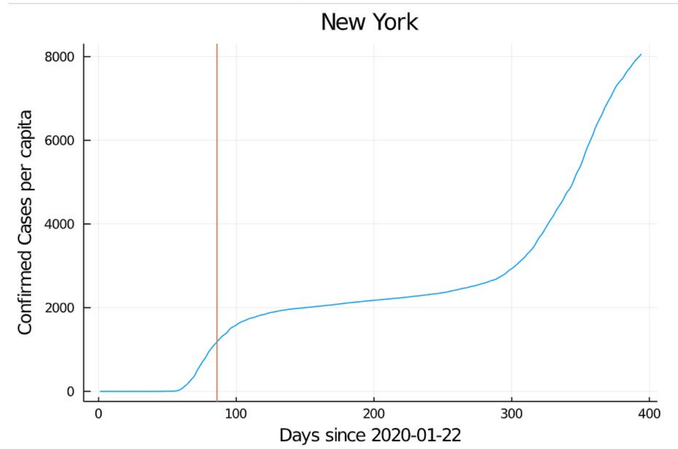
## Results and Discussion

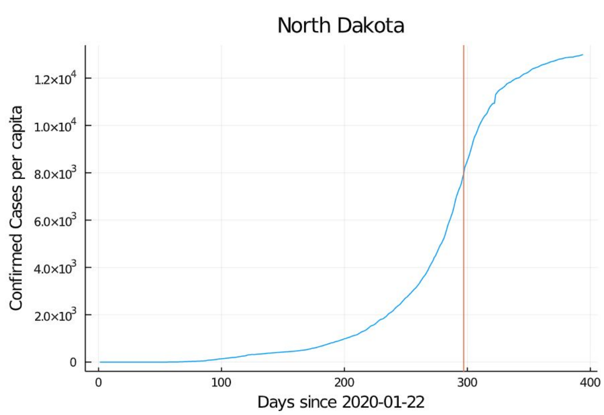
The following are the charts that we created comparing days after January 22, 2020 and the number of confirmed COVID cases per capita in different states: Florida, which did not pass a mask mandate, alongside Texas, California, North Dakota and New York, which did pass mask mandates on the days indicated by the vertical red lines.




At first glance, these results were rather surprising because it appears all states had roughly the same trend in new cases, especially after around day 300. Also, all states had around 8000+ cases per capita by the last day represented on the charts, regardless of whether they had a mask mandate or not. Another interesting observation was that all of the charts had a similar "hump" around the middle of the graph, representing where COVID cases slowed for some time before infection rates increased again (around day 300). These sections of the graph likely represent when measures like social distancing and shelter-in-place were put in place, even in

Florida. New York, however, had some notable differences compared to the other states: its mask mandate was passed much earlier, before day 100, and the curve in the chart is notably flatter from days 100 to 300, implying that the mask mandate, along with other COVID-19 response measures, successfully slowed the infection rate for a while. While mask mandates slowed the spread we can only guess as to why on near the 300th day cases seemed to spike in every state. Our main theory is that the winter holidays were around this time period, and with people hosting or going to parties along with disregarding the use of masks perhaps caused the spike due to covid fatigue. However there is no way to confirm this with the data that we had analyzed at this time. In summary, mask mandates have seemed to have helped curve the spread of the virus early in its life span, but as time went on states with a mask mandate versus those who didn't have not seen a large difference in terms of confirmed cases per capita.

**Conclusion**

This fellowship was an amazing experience in learning about data organization, analysis, and presentation of data. In this project we learned how to properly use *DataFrames* and *StatsPlots* packages in Julia along with learning about the language itself. Furthermore we had learned how to more efficiently parse the data from the CSVs into a DataFrame and how to create graphs in a concise and easily readable way. Through these skills we learned more about how mask mandates affect the spread of COVID-19. It was a unique and insightful experience to learn more about a once in a lifetime historical event that we would not have otherwise seen from just looking at the raw data. We had found the language extremely intuitive along with the speed in which the language is able to run. We want to give special thanks to our mentor, Ms. Kelly Shen, for overseeing this project over the course of our fellowship.

**Sources**

John Hopkins Data Hub: [https://popcenter.jhu.edu/data-hub/](https://popcenter.jhu.edu/data-hub/)