COVID STATE DATA ANALYSIS

Nicholas Machuga

**All of the code for this project can be found on my GitHub using this URL:** <https://github.com/NicholasMachuga/CovidDataAnalysisR>

Overview

In this project I take openly published data from Florida and Connecticut related to Covid-19 through csv files. I was able to use the data obtained to create new data and graph it in a more understandable way. The reason I chose these two states is because they took two completely different approaches to the pandemic and I wanted to see how their data compared.

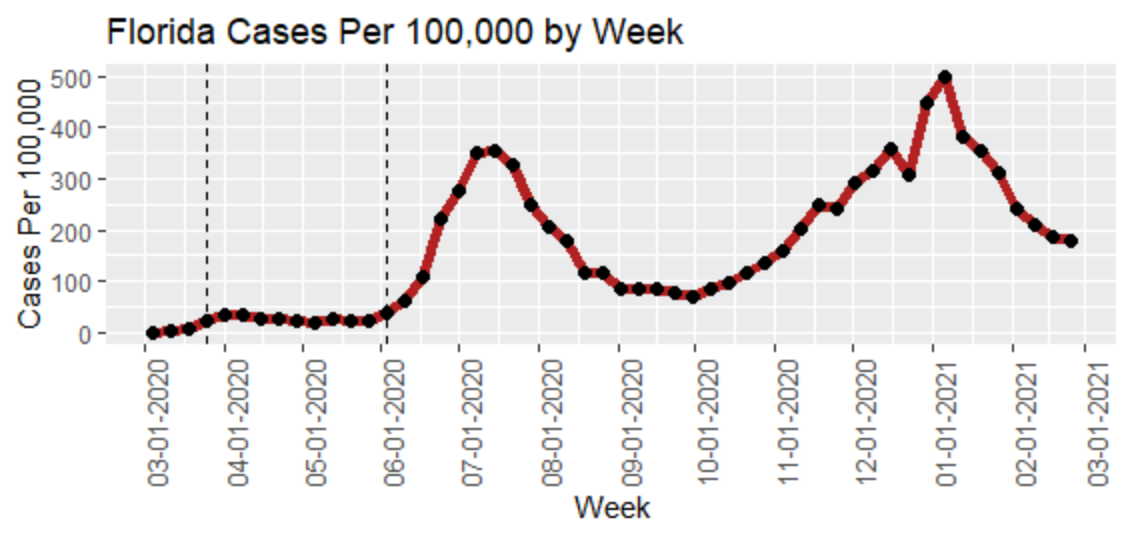
The Process

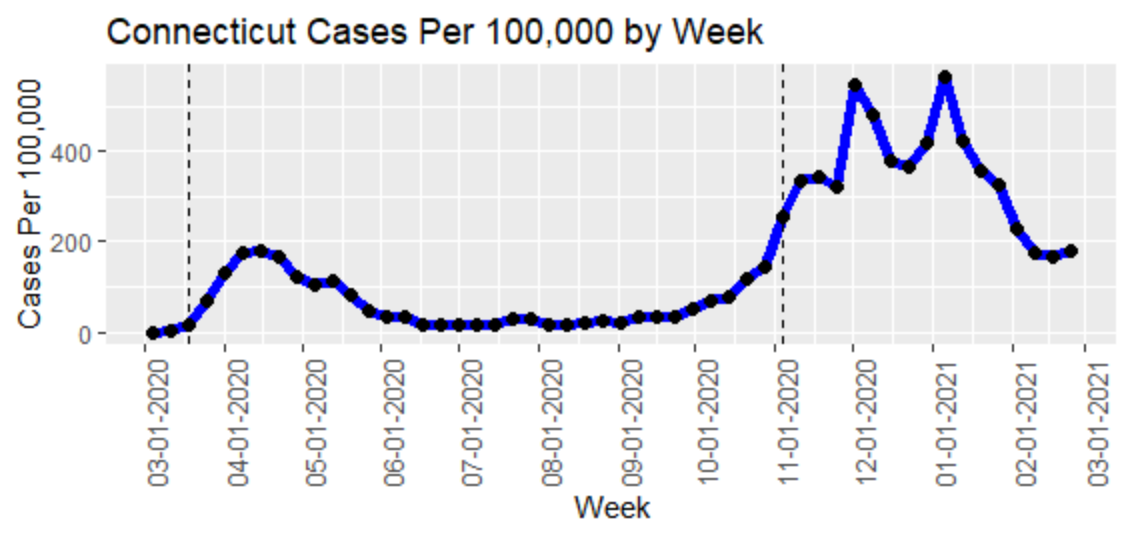
For this project I downloaded data from <https://covidtracking.com/data/download/> as csv files and imported them into R. I then cleansed the data because there were a lot of columns that either weren’t complete or contained data that I did not desire to use. I then created new data such as 7 day cases and death averages. For most new data points, I converted the raw values to values per hundred thousand. Moving average cases and deaths into per hundred thousand helped me make the data more understandable because with Florida’s population being much larger than Connecticut’s it was difficult to get a grasp on what the numbers meant.

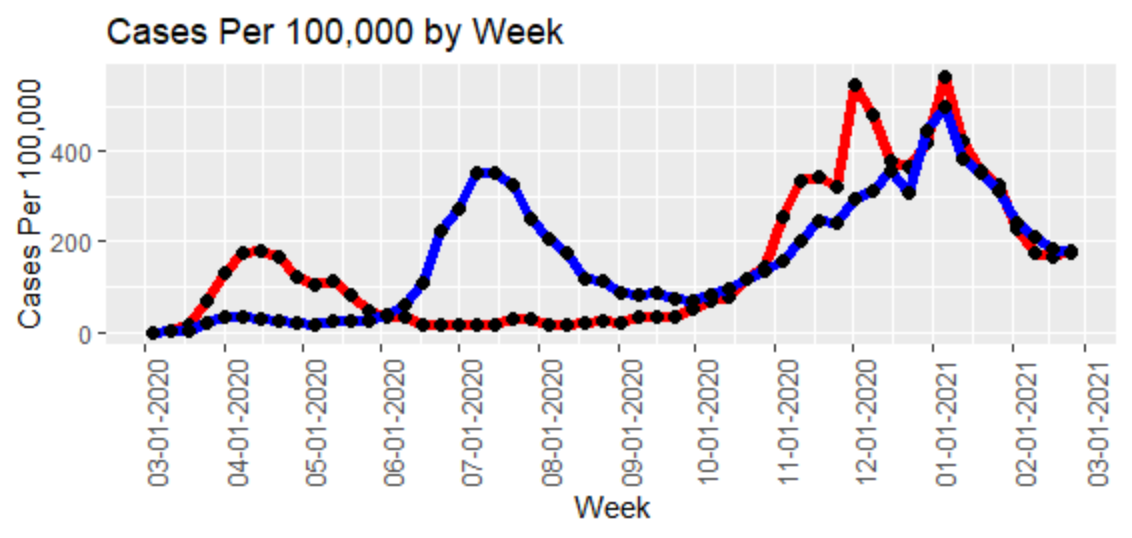
For each set of data I created, I graphed it for each state using the ggplot2 package and made graphs that I thought showed the data in the best way. I also made plots with both lines of data, from Connecticut and Florida on the same plot for a better comparison.

The Results

**Cases per 100,000:**

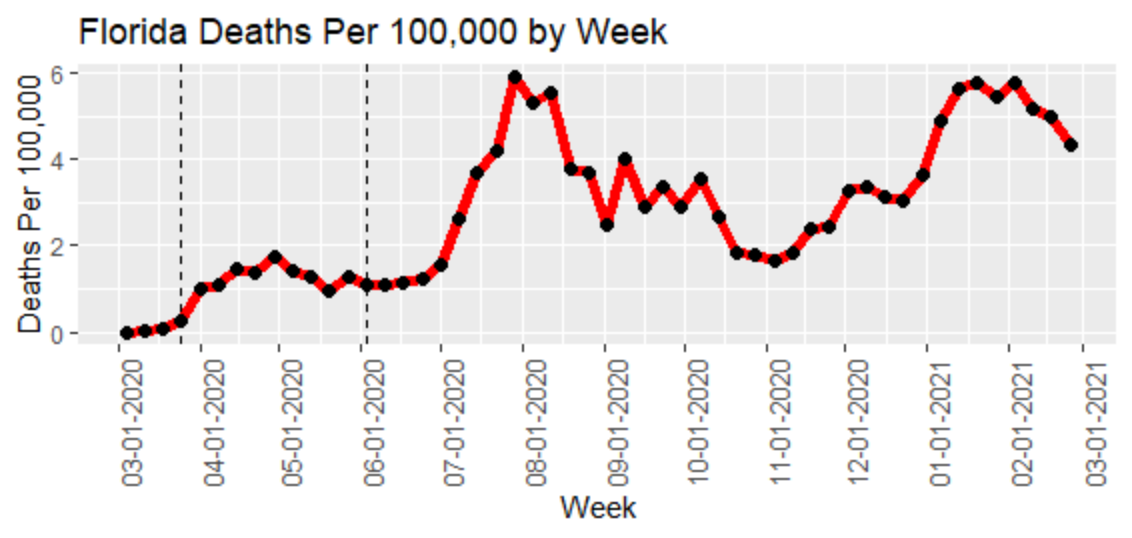
****

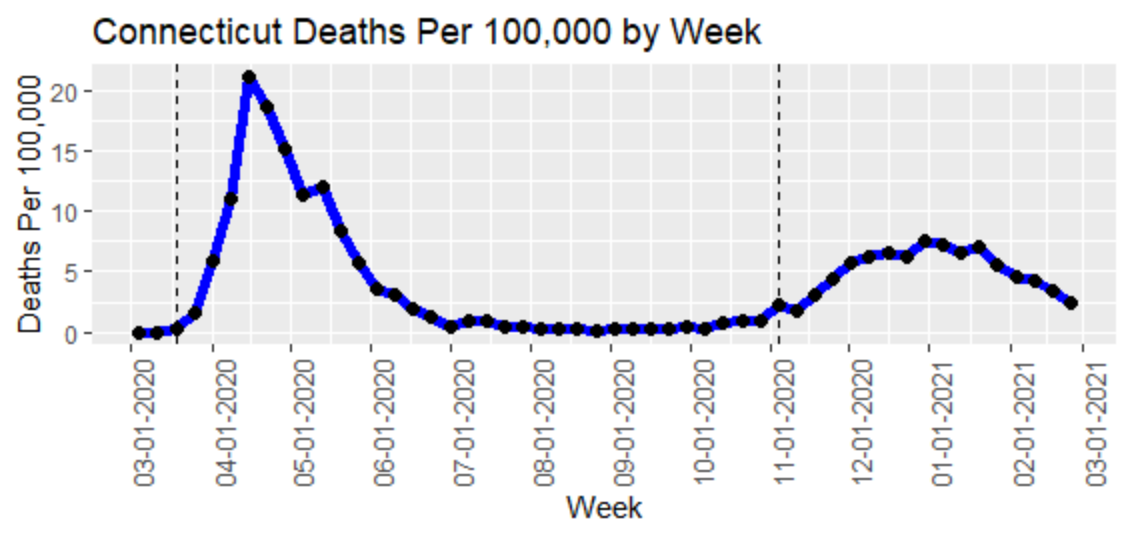
****

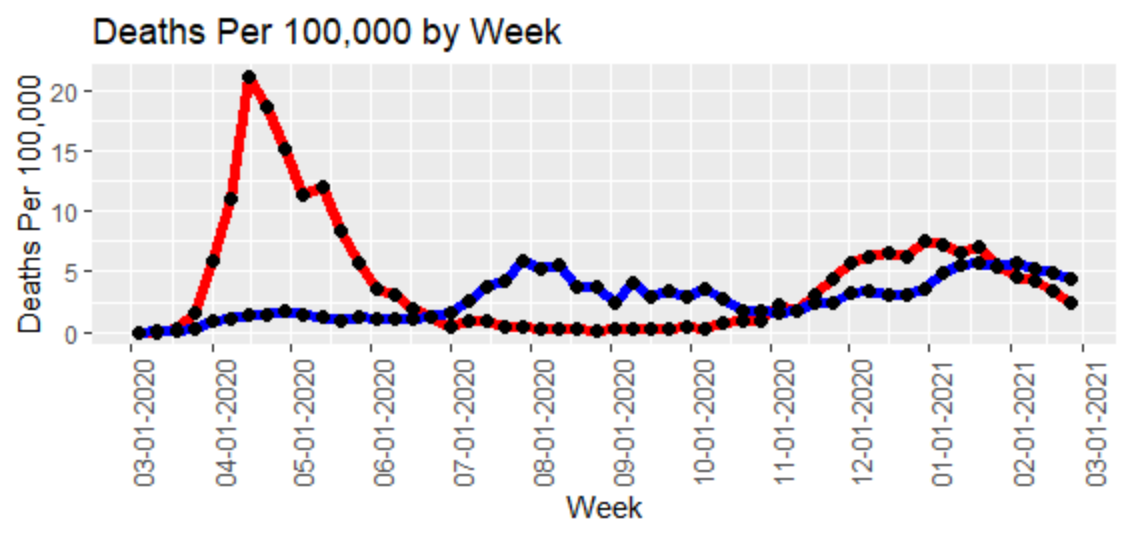
****

The vertical lines in these plots represent when the state went into lockdown and when the state reopened. For the sake of consistency, I considered reopening as the state allowing businesses to operate at 50% capacity. As we can see Florida had a large spike of cases when the pandemic first started and then flattened out during lockdown. However, upon reopening their cases increased substantially and did not go back down until the vaccine started to become available. Connecticut on the other hand experienced a large spike as soon as they shut down, most likely from activities that took place the week before it. We then see Connecticut’s cases stay low throughout their lockdown but then increase as soon as the lockdown ended. These graphs imply that a shutdown does work while in effect, however in both states we saw an extreme increase as their respective shutdowns ended.

**Deaths:**

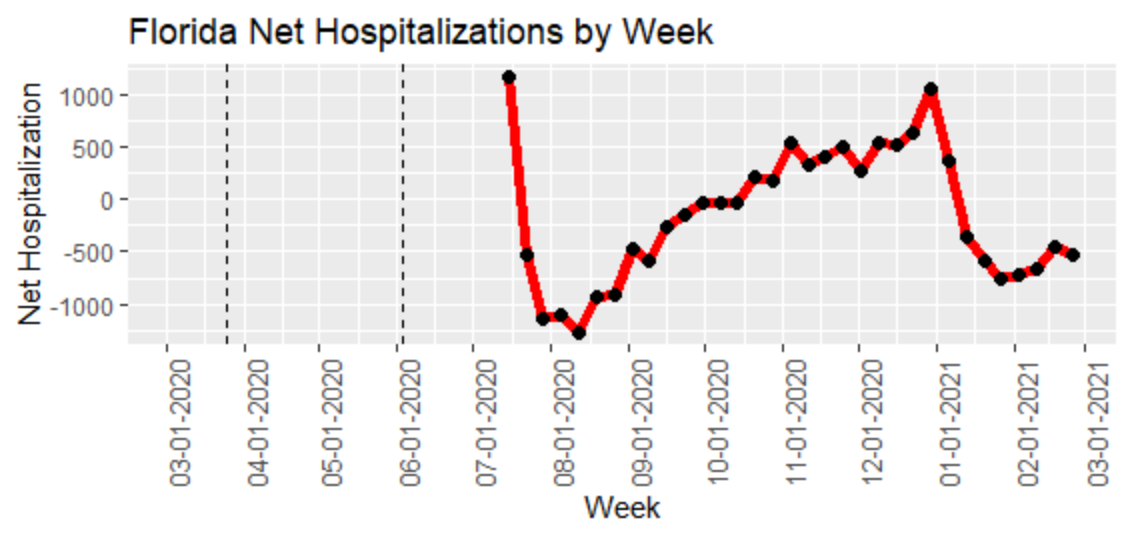
****

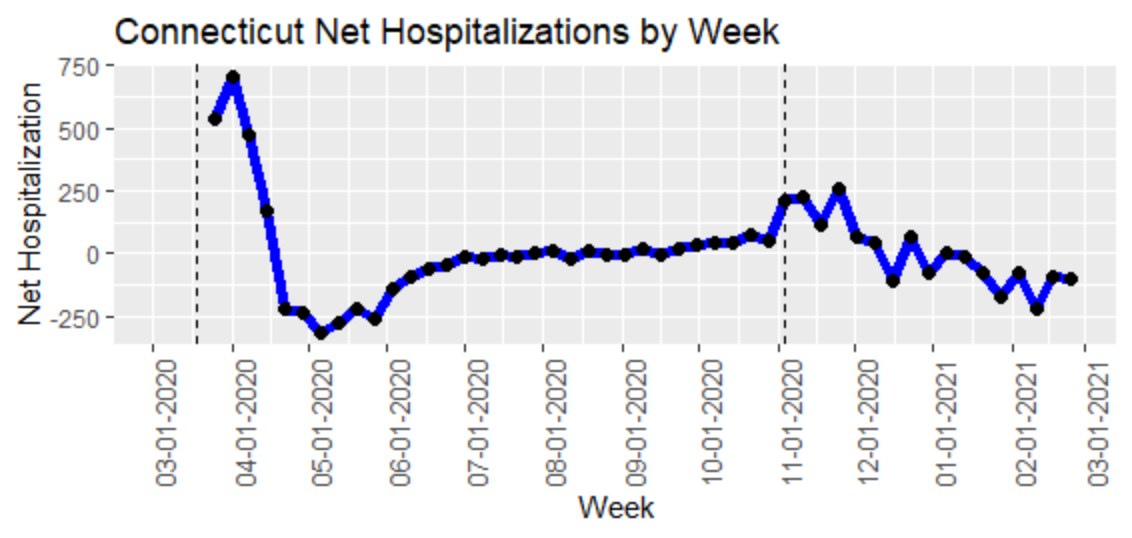
****

****

The results of plotting the deaths per 100,000 were what surprised me the most in this analysis. I expected Florida’s numbers to be significantly higher than Connecticuts. I hypothesised this because Florida’s percentage of people over 65 years old is much higher than Connecticut’s and we know that COVID is much more deadly to the elderly. Florida has a huge spike of deaths as the pandemic hit very hard initially and doctor’s didn’t know how to treat it well. SInce that point, even with cases being high, deaths have stayed lower. In Connecticut we don’t see much variation in deaths per 100,000 from the beginning to now. The cases have risen in Connecticut recently, creating higher death numbers as well.

**Hospitalizations:**

****

****

This was the toughest data to compare because both states didn’t record their hospitalizations from the start. However, we can see that Florida’s net hospitalizations per week have increased in waves. They seem to align with the spikes in cases which makes sense. In Connecticut's case, the net hospitalization stays closer to 0 because their general population is lower than Florida’s.

**Conclusion**

In conclusion, we can see how different strategies when it came to controlling the pandemic have played out so far. In Florida’s case, a huge spike of cases took place at the beginning and then the shutdown helped control the outbreak until the state reopened. When the state reopened, the outbreak resumed and the cases have come in waves. Connecticut on the other hand had the outbreak and decided to shut down for a much longer period of time. The pandemic was under control for that whole time but when the state moved to reopen the cases and deaths went back up and have stayed stable since then.