dsaf_final_project_covid

R. Buessing

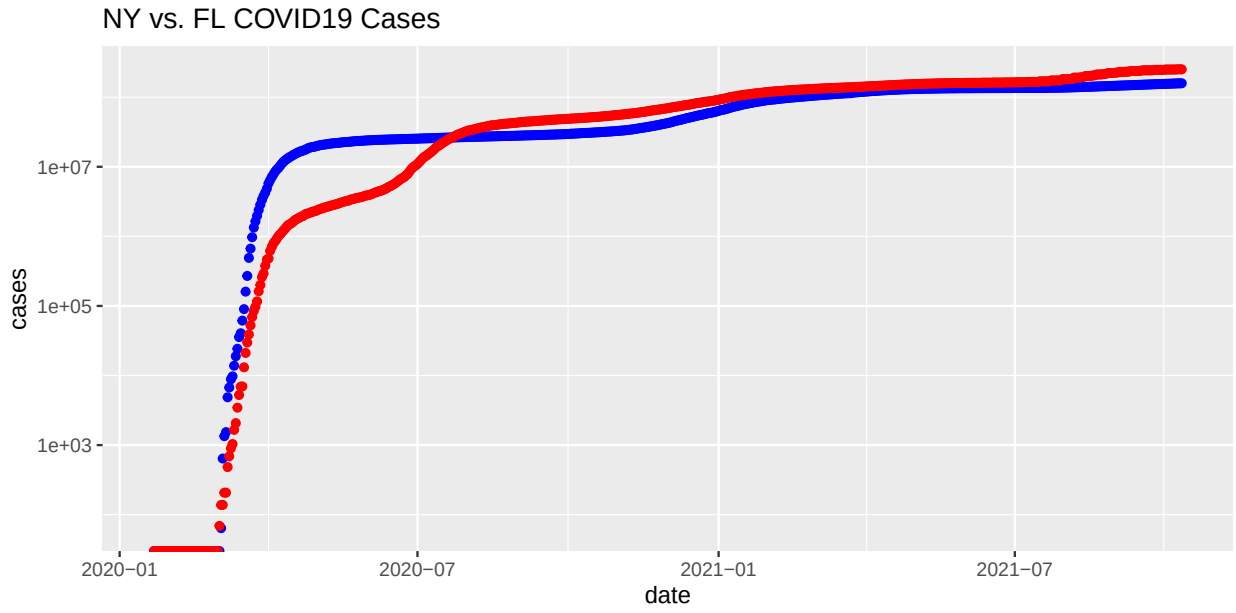
10/11/2021

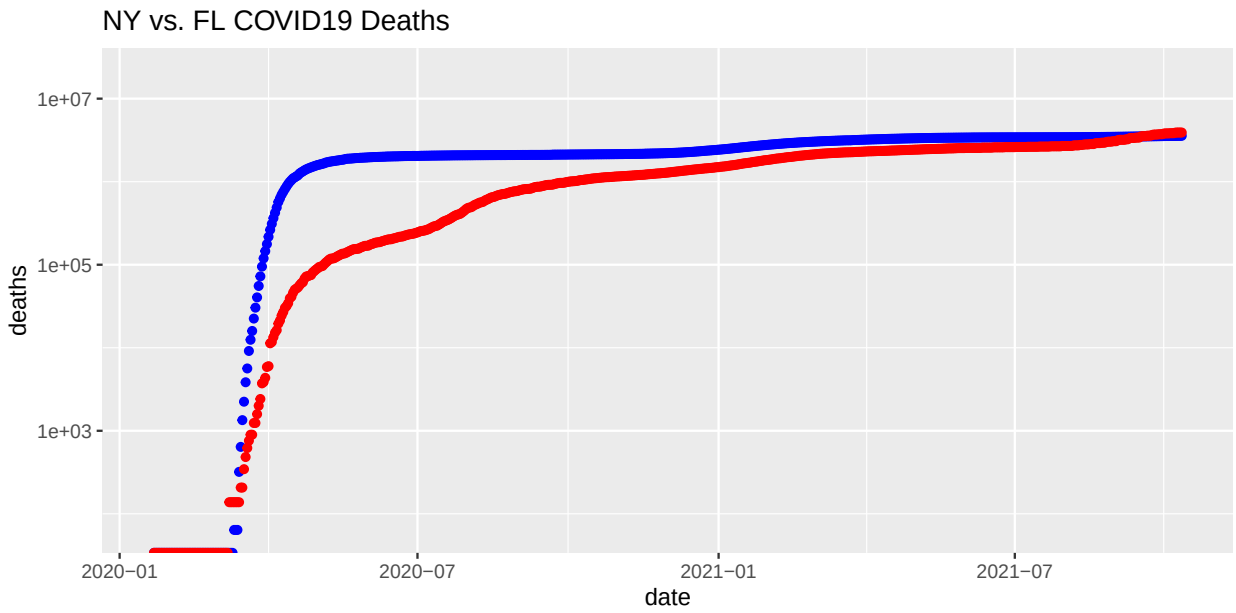
In this report we'll be comparing COVID19 cases between New York and Florida, two states with almost identical populations but also with two completely different approaches to handling the pandemic.

Before we begin, this report, the U.S. cases and deaths data from John Hopkins, and R code can be viewed at the following public GitHub repo: https://github.com/rabu4751/dsaf_final

Data Plots

Below are two charts that compare COVID19 cases and deaths between Florida and New York.



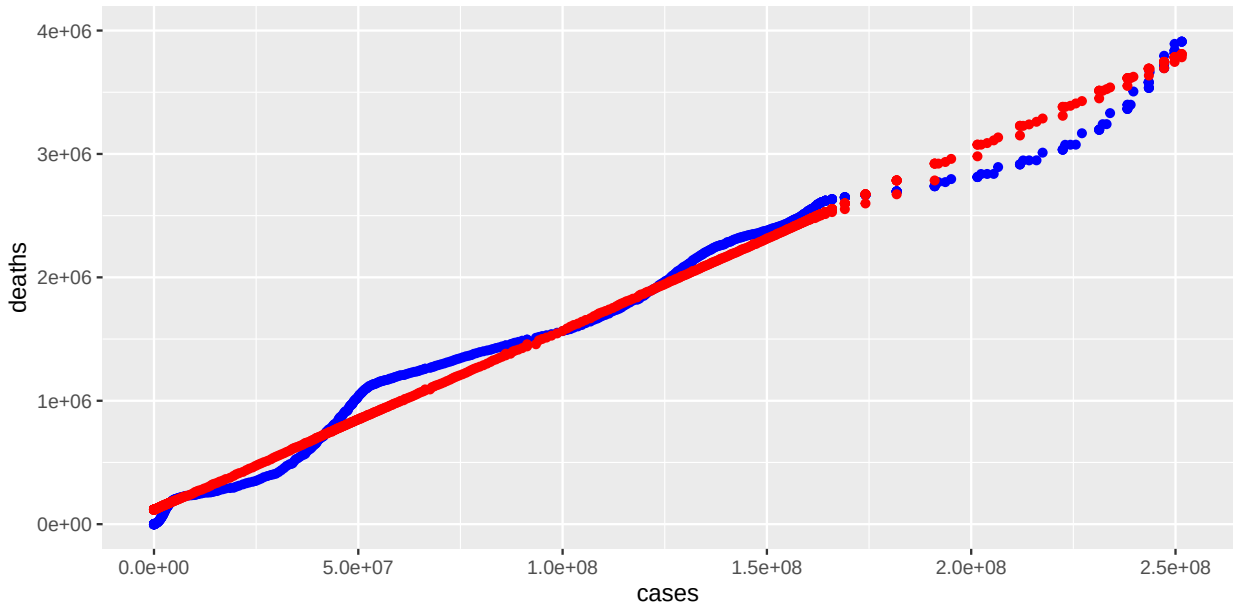


Findings

Interestingly enough, at the beginning of the pandemic, New York had a much quicker spike in the number of COVID19 cases in the State while Florida took a couple months to catch up and eventually pass New York consistently in overall cases. In regards to deaths, New York has actually outpaced Florida until just recently which could indicate the overall preparedness of each state as the pandemic largely begin for the U.S. in New York. Florida may have had the opportunity to learn from New York and prepare their health infrastructure before COVID19 had a stronger foothold in the state.It's important to note that due to the speed and ever changing nature of the virus thus far, data could possibly be lagging behind what is currently available, however this should not fuel biases or theories that the data is heavily skewed in any given direction. Measurement bias due to human error would likely be the most prevalent form of bias in this data.

Prediction

Using the same does as above, I created a linear regression model for Florida to see if the model would predict more or less deaths in Florida. The model is able to predict fairly well with the exception of the most recent data where the model was over predicting deaths but has begun to gradually level out.



Conclusion

Despite New York having what most would consider a stricter approach to the COVID19 pandemic, only recently has the state be surpassed by a more lenient state such as Florida. This could be due to the timing of the virus and when it struck each state as well as the overall medical infrastructure of each state. Bias in this data would likely be from measurement bias as a result of human error and inputting the data incorrectly or simply not having all of the available data due to how fast data changes on COVID19.