

Florida COVID-19 Trends and Insights

Milestone 3

Anna Harvey
Spring 2021

<https://github.com/anhar421/Portfolio>

Abstract

Understanding the rates of COVID-19 cases and deaths in Florida has been a continually complex problem. Florida is a unique state with widely varying population demographics that are largely divided by county. Since the beginning of the pandemic, many of the counties have functioned under their own authorities, despite receiving different guidance from state and federal levels. This study investigated the extent COVID-19 has affected Florida counties differently and if there is any variance between counties with different dominating demographic trends. Initial analysis of case percentage by population and deaths percentage by population shows that there is no strong correlation between population and cases or deaths. In fact, some of the lowest populated counties had the highest rates of cases and deaths. Further investigation into county demographics and COVID-19 mitigation policies is required to determine if any key variable affected these values.

Introduction

State-level data about the COVID-19 pandemic has been essential for state governments in their decisions about policy changes and safety restrictions. There has been little focus, however, on the differences of COVID-19 case and positivity rates between counties within the same state. States like Florida operate in a similar fashion to the nation as a whole, where there is an overarching authority that determines high-level policies, but the counties make many of their own decisions within state limitations. Data at a county level is crucial to analyze if local governments are meant to make decisions appropriate for their own communities.

The presumed driving factors behind the speed of COVID-19 spread and the severity of cases are population centers and age. Larger populations are assumed to have a higher rate of spread. Areas with older citizens are assumed to have a higher rate of death. Florida has 67 counties that range in population between approximately 8300 and 2.7 million¹. There are several areas of the state with expansive retirement communities and elder care facilities. Additionally, as a political swing state, Florida is well-known to have a wide range of political beliefs. Since COVID-19 mitigation decisions have been largely influenced by the main political parties, pandemic behavior in different counties has also likely been influenced by the main party represented in each county.

Problem Statement

If population and age distribution play a significant factor in COVID-19 spread through communities, a strong correlation should be found between those variables. Should such a correlation exist, it would provide a more targeted strategy for different counties based on the demographics of the area. Alternatively, the lack of a strong correlation could suggest that other factors may be the driving forces behind spread and death rates.

Methods

Datasets were utilized from US Census data as well as the compiled county-level data collected by the New York Times and provided on GitHub^{1,2}. The original county-level data includes daily reports from every county in the US from every day data was reported since early 2020². Florida data was isolated from that dataset. The case and death numbers in the data are cumulative and each day's numbers are added to create a running total for each county over the course of the year. This data does not reflect the daily changes to case and death rates. The cumulative data for the most recent reported date was isolated to evaluate the overall case and death rates for each county. Features were created to standardize the case and death rates by population. The percentage was calculated by dividing the variables by the population. A standardized number per 10,000 people was also created by multiplying the percentage by 10,000.

Visualizations were created for the cumulative totals for the year to explore any possible correlations or outliers (Figure 1; Figure 2). It was determined that Lafayette County is an outlier for case rates with a significantly high percentage of 19% of the population infected since March 2020. Investigation into this situation revealed that there was a significant COVID-19 outbreak at the Mayo Correctional Institution in August 2020³. Miami-Dade County also could be considered an outlier for case rates at 16%. It is a surprisingly high number when compared to other large counties. Union County is the only outlier for death rates at approximately 46 of every 10,000 people dying from COVID-19.

Correlation matrices were calculated for all variables. Both Pearson and Spearman correlations were calculated, however, it is unlikely that any of the variables have a linear relationship therefore Spearman correlations should be primarily utilized. There are significant positive correlations between case percentage/death percentage (0.91) and standardized cases and deaths (0.91). Strong correlations between percentage and standardizations with their contributing variables (cases and deaths) were ignored. There are weak negative correlations between case percentage and population (-0.16). That was the strongest negative correlation between variables.

Results and Conclusions

There does not seem to be a strong correlation between population and the number of COVID-19 cases or deaths. If the number of people in an area were the primary factor in determining how quickly COVID-19 spreads in a community, we would see a higher percentage of cases in the highly populated counties. In fact, most of the counties with higher case rates are primarily lower population counties. We particularly see this in the smaller counties along the panhandle. Overall, we do see most counties clustered within the same range of case rates (between 6%-11%), regardless of population (Figure 2). The two exceptions to this are Miami-Dade County and Lafayette County. As previously mentioned, Lafayette County's primary factor in its high percentage of cases is due to the outbreak at the Mayo Correctional Institution. Interestingly, there are major correctional facilities located in almost every county, but not every correctional facility had a major outbreak. Miami-Dade County seems to display the assumption that a higher population leads to higher case rates. However, when comparing Miami-Dade to other large counties, the differences in case percentages are 5%-9%, while the other large counties only vary from each other by 1%-4%.

The relationship between death rates and population theoretically is that the higher the population, the more cases there are, and therefore the more deaths. However, since we see that population and case rates do not have a strong correlation, it makes sense that death rates and population do not have strong correlations either. Death rates and case rates have strong correlations to each other, which does confirm that theoretical relationship. The one outlier of Union County falls within the mid-level case rates for counties, but has a difference of 14 more people per 10,000 dying from COVID-19 compared to the next highest death rate. Otherwise, the death rates for all counties stay between 6-32 per 10,000 people (Figure 2). That is a wide variance and further investigation into the variation between the main cluster is warranted.

Future Investigations

The affect of age distribution on case and death rates has yet to be investigated. The US Census data has some data missing from certain Florida counties and a more complete dataset needs to be found to accurately compare the counties' COVID-19 data to their age distribution. Miami-Dade County's high case rate does not have any variable strongly correlated to it within this dataset. Behavioral or other demographic information needs to be investigated to determine why the case rate there is so high. Union County also does not have a variable strongly correlated to its high death rate. Behavioral or demographic information need to be investigated for this county as well.

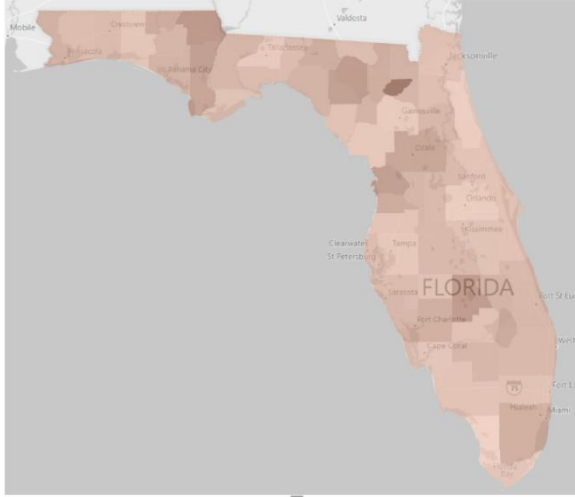
Were daily county data from the past year to become accessible, comparing the growth rates at different times of year could also lead to information regarding case and death rates. Additionally, political information about counties could be integrated into analysis. This however works on the assumption that counties behave in specific ways associated with different political parties. Mobility data about behavioral information of different counties

would be most helpful as an additional variable to focus on. The next likely variable known to affect case and death rates besides population and age distribution is social distancing.

APPENDIX

Figure 1. Filled maps showing case and death rate percentages by county.

Death rate % by county



Case rate % by county

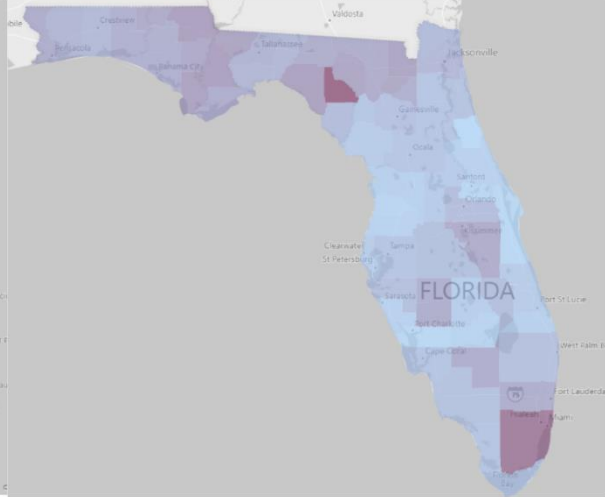
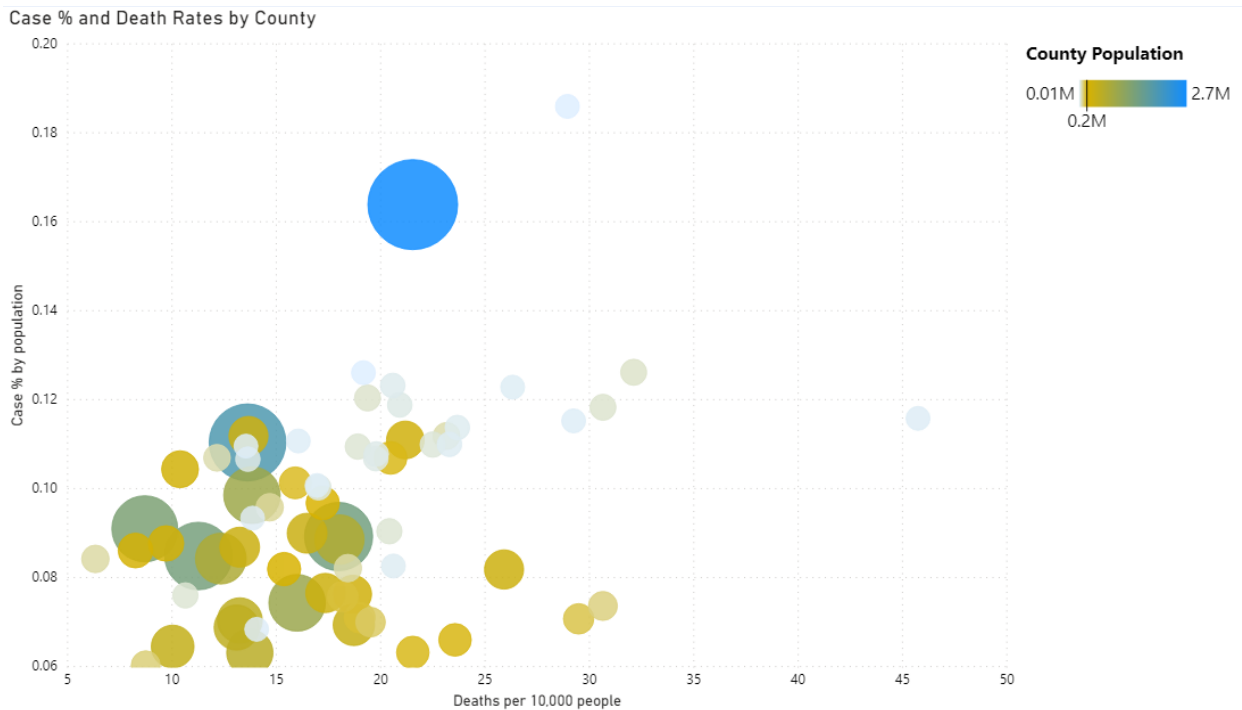


Figure 2: Scatter plot for case percentages and death rates with bubbles sized for population.



SOURCES

1. *Florida Counties by Population*. Florida Outline. (n.d.).
https://www.florida-demographics.com/counties_by_population .
2. Nytimes. (n.d.). nytimes/covid-19-data. <https://github.com/nytimes/covid-19-data>.
3. Renton, A., & Hayes, M. (2020, September 10). *Nearly 14% of the people in this Florida county have had Covid-19*. CNN. https://edition.cnn.com/world/live-news/coronavirus-pandemic-09-09-20-intl/h_1a0c7d2ff36b08ab4ca3ffbb1e304c8a .

10 Questions

1. Why did you use data compiled from Johns Hopkins University instead of data direct from the Florida Department of Health?
2. What makes Florida different than other states when it comes to COVID-19?
3. What domain knowledge is required to have a complete understanding of COVID-19 spread?
4. Which counties had surprising case and/or death rates? Why?
5. Why did you focus on population as a focal point?
6. What factors might contribute to population not being strongly correlated with case rates?
7. Why do you think county data has not come under more investigation when deciding policy?
8. Are COVID-19 case and death rates enough to make informed statements about the current state of the virus in a county?
9. Could COVID-19 case and/or death rates be predicted using predictive analytics?
10. What future paths could you investigate to further understand COVID-19 from data standpoint?