Total Deaths 2020

In this analysis, we will analyze the following:

- How different the number of deaths were in the United States for the years 2015 through 2020 and compare them.
- The total number of deaths per jurisdiction, per cause, per year.
- The deaths per capita rate per jurisdiction.
- If mask mandates helped reduce the deaths per jurisdiction.
- If travel restrictions helped reduce the deaths per jurisdiction.
- If gatherings banned helped reduce the deaths per jurisdiction.
- The jurisdiction severity levels.

How was the data collected?

- Weekly Counts of Death by Jurisdiction and Cause of Death
    - Provided by the CDC. Downloaded and accessed as a CSV. Can be accessed here.
    - 329988 rows × 15 columns. Variables included death counts, reason for death, state, and date.
    - Used to calculate the number of deaths per U.S. state per year, from 2015-2020.
- U.S. 2010 – 2019 Population Counts
    - Provided by the US Census. Downloaded and accessed as a CSV. Can be accessed here.
    - 57 rows × 151 columns. Variables include state/territory population estimates per year from 2010 census.
    - Used to calculate each state's death rate, controlling for population size.
- U.S. 2020 Population Estimates
    - Web scraped from Wikipedia's U.S. 2020 Census page. URL is: https://en.wikipedia.org/wiki/2020_United_States_census
    - Used to complete the population count data from the Census. As of the time of this project, the 2020 census data has not been released.

- State Lockdown Responses
    - Web scraped from Wikipedia's U.S. State Response to COVID-19 Pandemic page. URL: https://en.wikipedia.org/wiki/U.S._state_and_local_government_responses_to_the_COVID-19_pandemic
    - Two tables were scraped from this webpage.
        - The first contained information on individual states lockdown procedures, including if they had "Stay at Home" orders, if masks where required in public, and travel restrictions.
        - The second provided start and end dates to "Stay at Home" orders for states that had a specific end date.
    - Used to calculate each state's severity of lockdown, to see how that impacted death rates.

Data Cleaning

The final dataset involved gathering, cleaning and merging four smaller datasets.

The first dataset involved weekly death counts of each state and U.S. jurisdiction from the beginning of 2015 through the end of 2020. North Carolina, Indiana, and Connecticut all had some weeks with missing death counts for certain weeks. We filled those values with the median number of deaths for that state, in 2020, for that disease. This is because these values did not tend to have a high variance, when controlling for the three categories, so the median would be a good measure of central tendency. Then we type converted the "State" and "Cause of Death" to categorical variables.

The population data was created by combining two datasets, one for 2015-2019 and one for 2020. The 2015-2019 data was a CSV from the U.S. Census Bureau, which did not contain any strange or missing

population values. The main cleaning challenge was reformatting the tables to have a row for every state and every year, instead of having a row per state with a column for each year's population.

The 2020 population estimates were web scraped from Wikipedia, and also did not contain any strange or missing values.

Figure 1 below displays the total deaths per year. We can see that the total number of deaths increase every year. Table 1 is the numerical representation of figure 1.

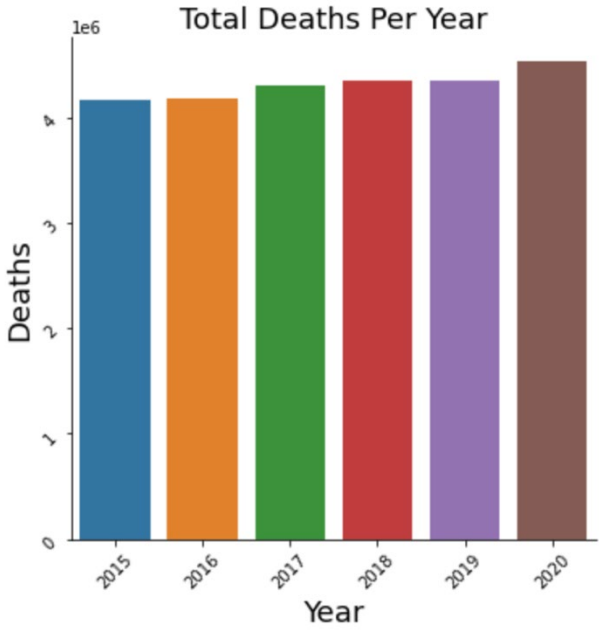

Figure 1

	Year	Deaths	Rate
0	2015	4169620	0.16115
1	2016	4185800	0.16178
2	2017	4301084	0.16623
3	2018	4343758	0.16788
4	2019	4341462	0.16780
5	2020	4531878	0.17515

Table 1

Figure 2 below displays the number of deaths per year per cause.

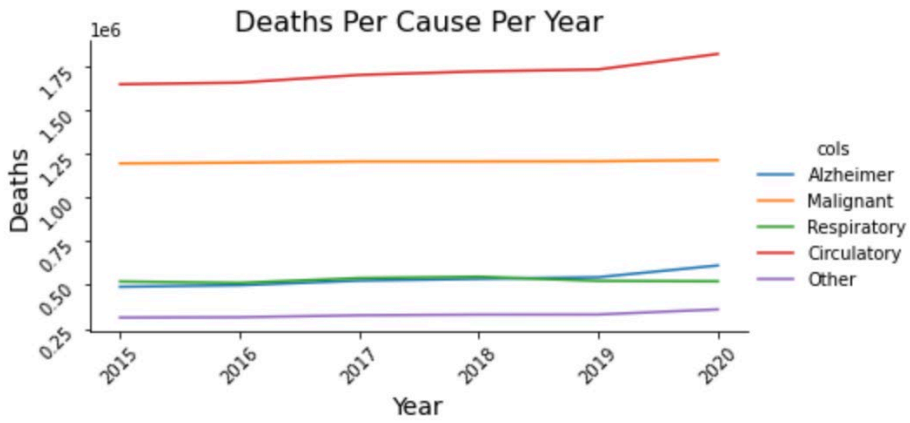

Figure 2

According to figure 2, the most deaths in all observed years is caused by circulatory problems, followed by malignant, then Alzheimer's and respiratory problems.

Let's look at total deaths by cause for all observed years combined in figure 3 below.

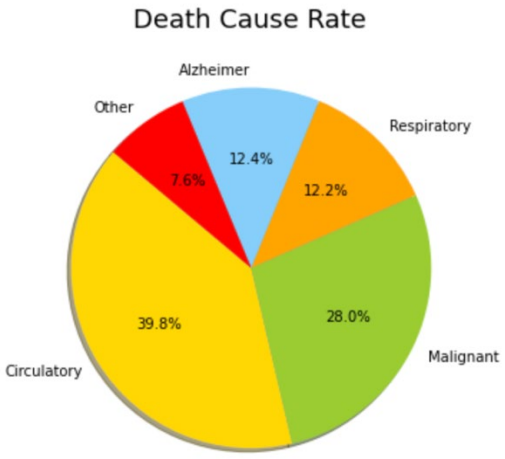

Figure 3 clearly displays that the total deaths for all 6 years combined came from circulatory complications at 39.8%, followed by malignant complications at 28%. Together, theses two causes almost add up to almost 70%.

Figure 3

Let’s plot the deaths per cause per year separately.

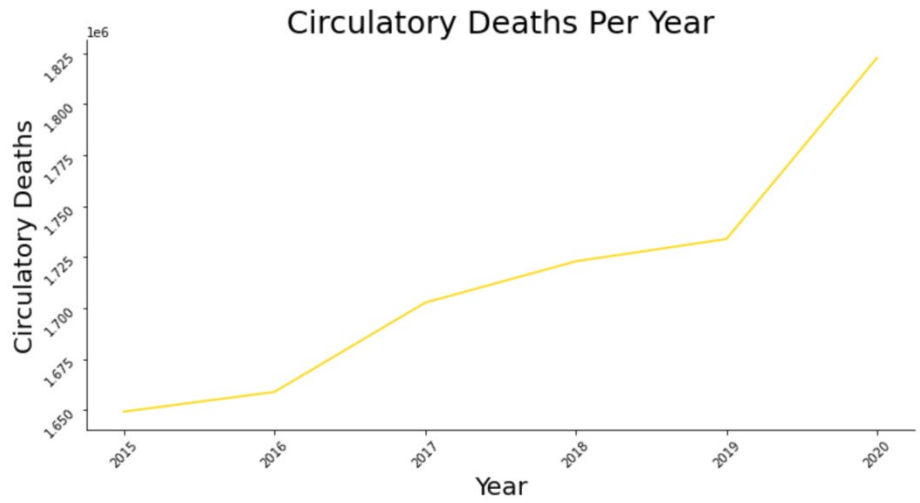

Figure 4

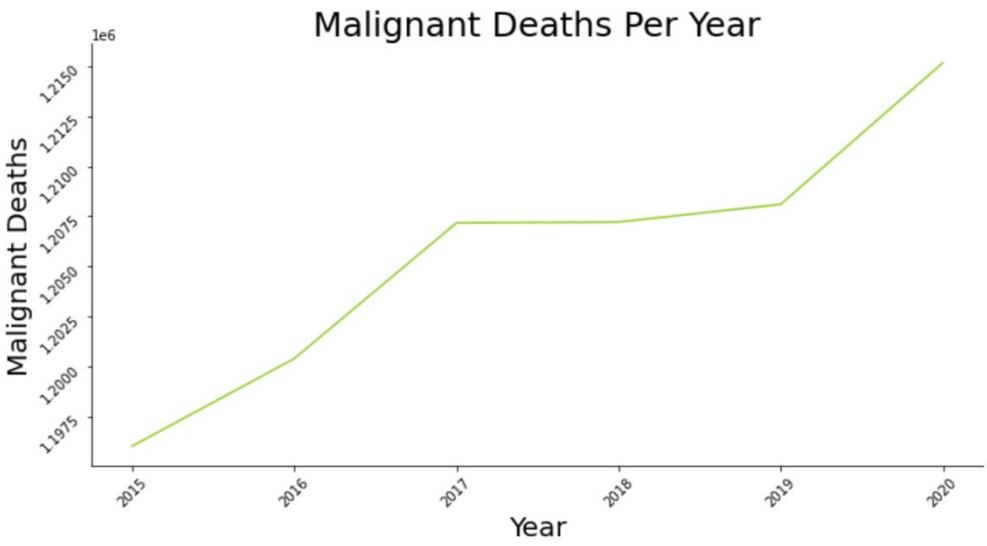

Figure 5

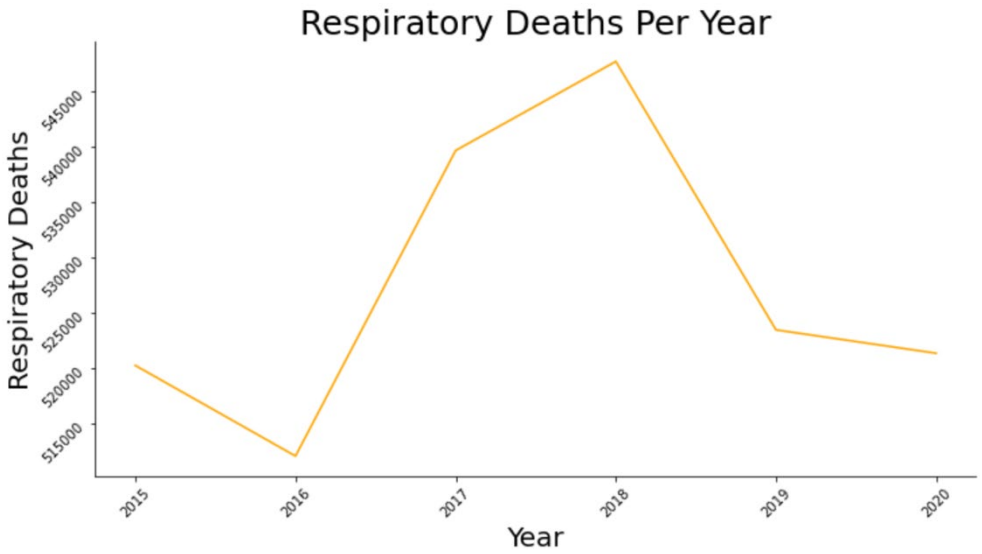

Figure 6

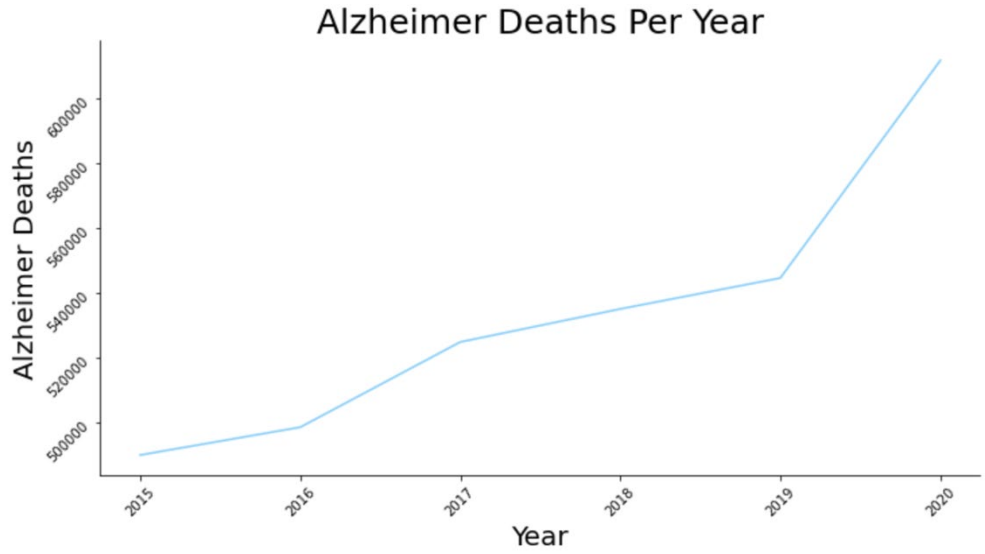

Figure 7

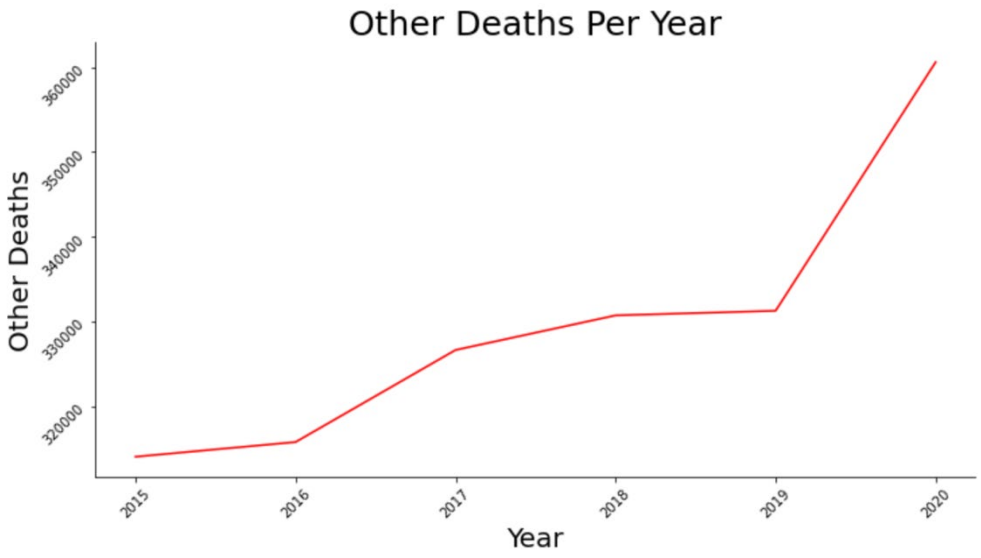

Figure 8

According to the visualizations above, you can see the deaths of every category increase every year except respiratory. This is interesting because in the year 2020, most of the country had lockdown restrictions for the COVID-19 pandemic.

Let's look at the total number of deaths per jurisdiction. Figure 9 shows the ten jurisdictions with the most deaths and figure 10 shows the 10 jurisdictions with the least deaths.

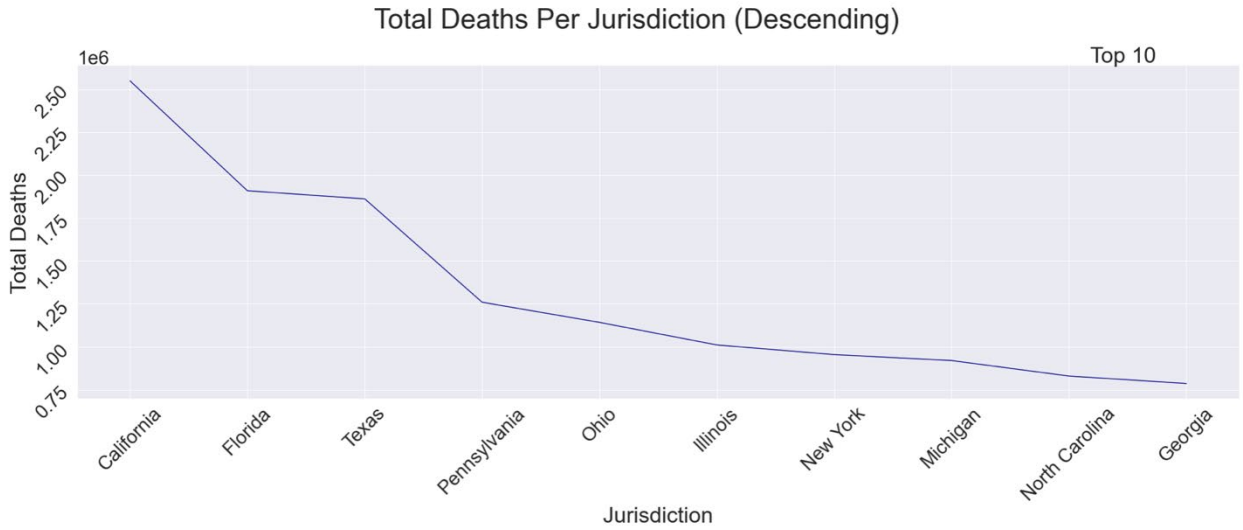

Figure 9

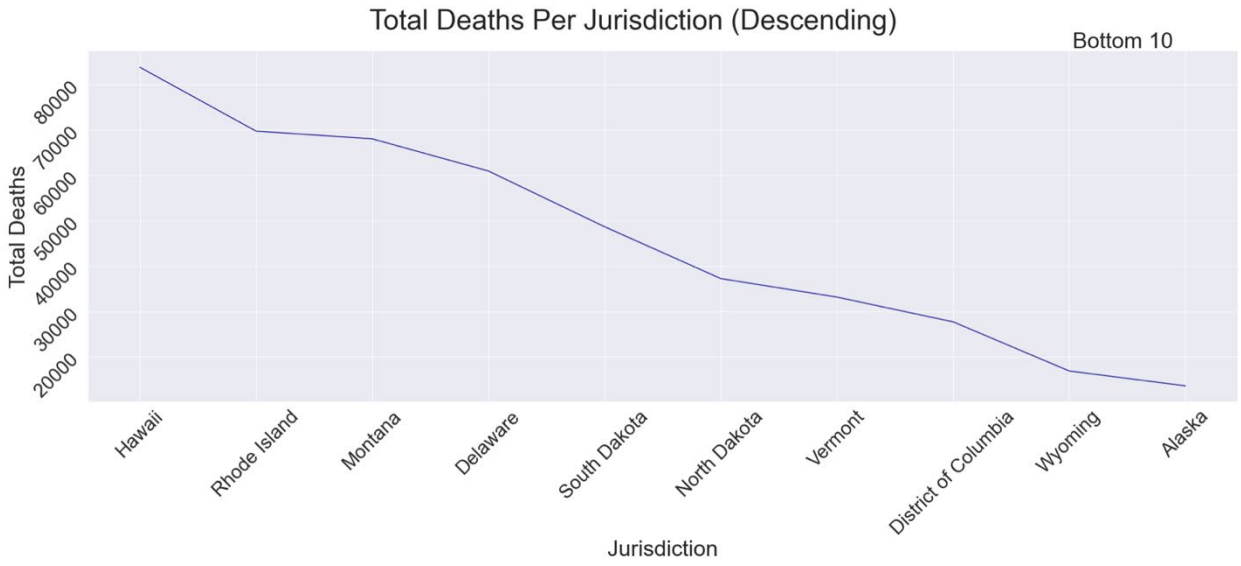

Figure 10

Figure 11

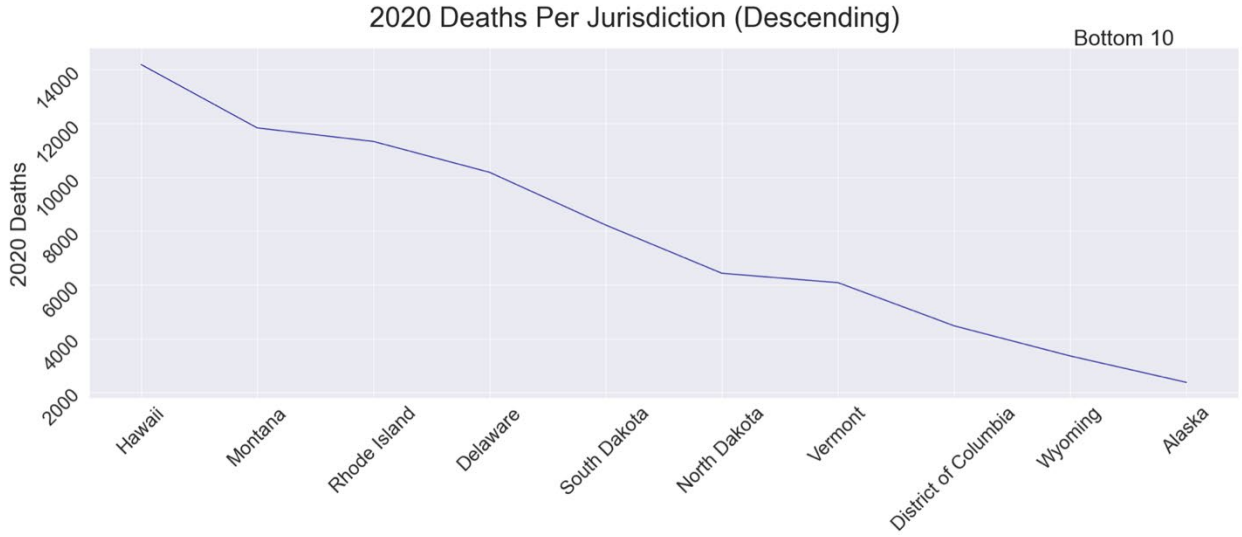

Figure 12

From figure 9 and figure 11 above, California is the jurisdiction with the most deaths but is this because California has the most population? We will analyze deaths per capita next.

Let's make a visualization of DeathsPerCapita by pulling a few random jurisdictions and look at the change over time.

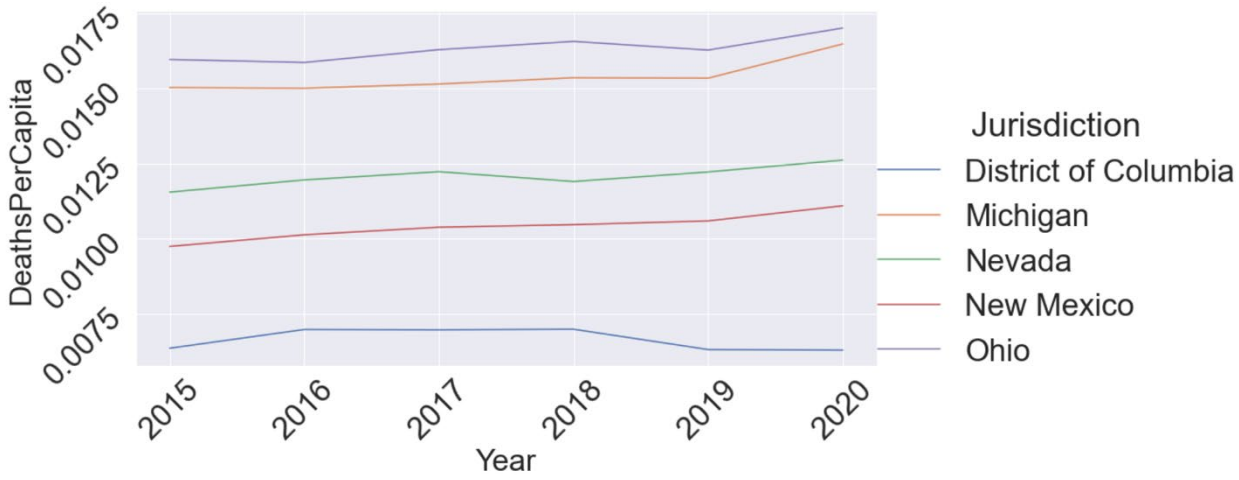

Figure 13

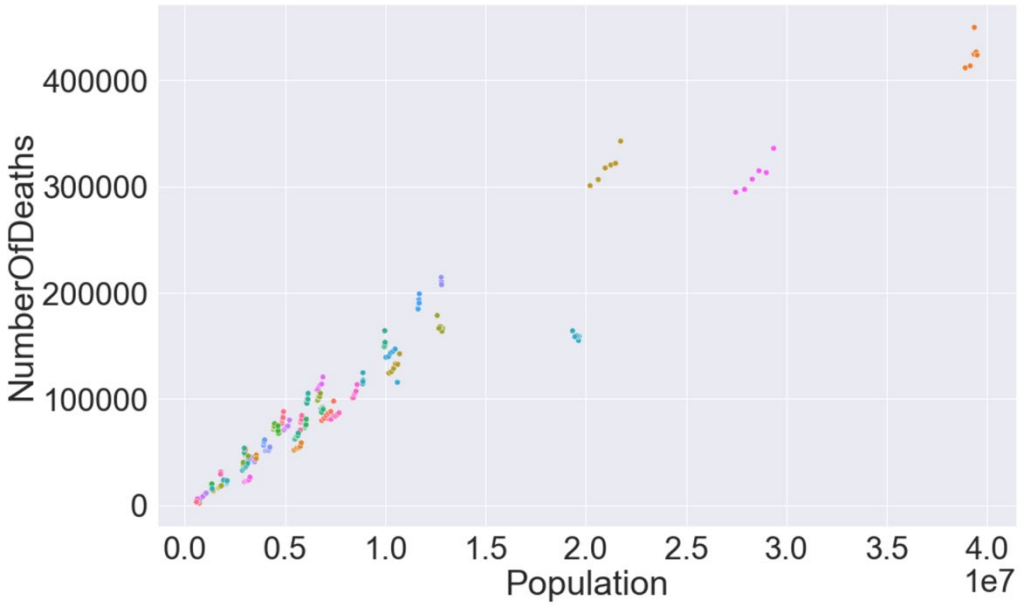

Figure 14

Figure 14 above shows the number of deaths per population. It also shows that as the population increase, the number of deaths also increase.

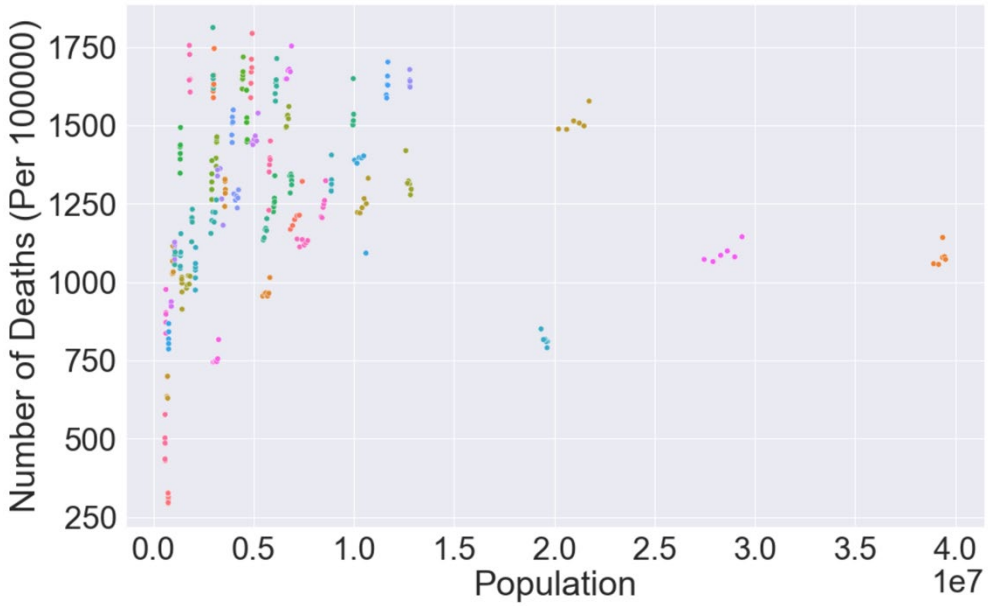

Figure 15

Figure 15 displays the deaths per capita.

Now, let's look at state lockdowns. We will feature engineer to get a lockdown severity of each jurisdiction during the 2020 lockdown. We will use travel restrictions, gathering bands, and stay at home order.

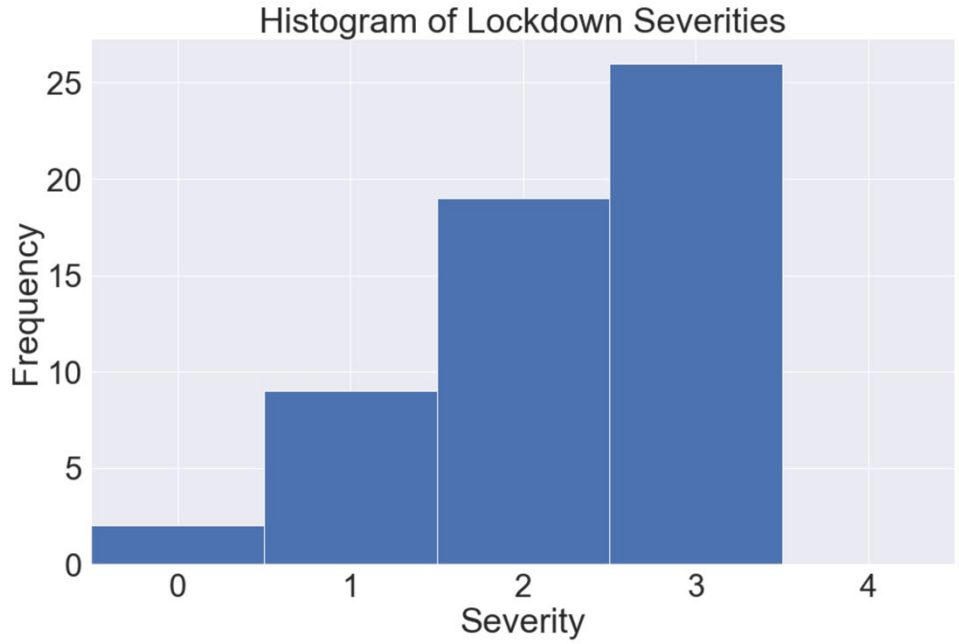

Figure 16

Figure 16 above shows that most level 3 severity had the highest frequency.

Let's continue with our analysis and analyze if mask mandates made a difference with deaths per capita.

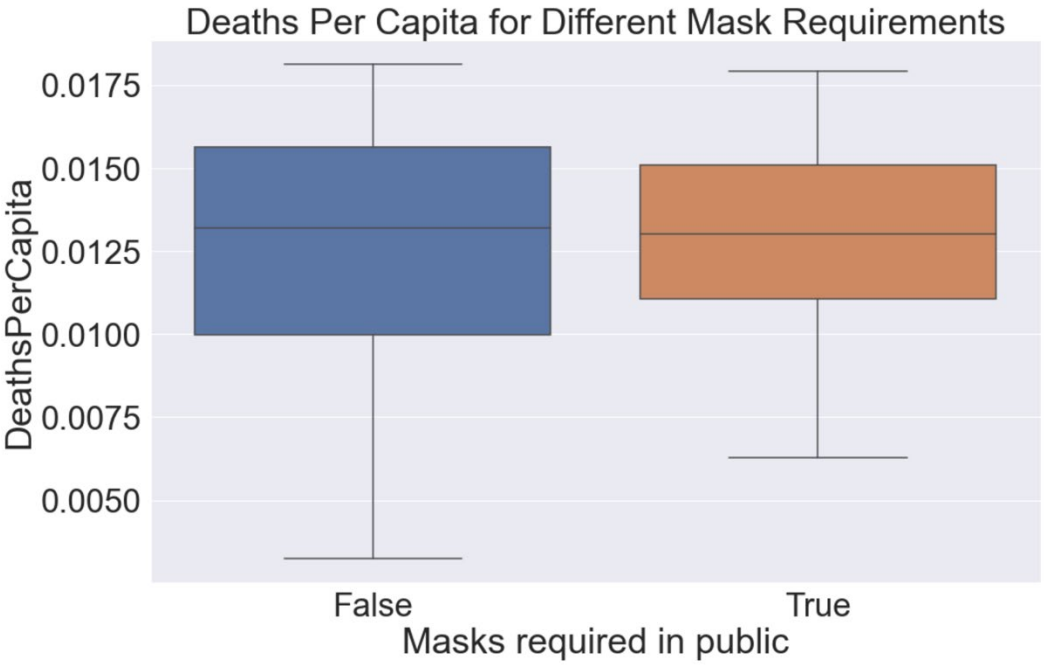

Figure 17

According to figure 17, there doesn't appear to be a significant difference between these two box plots. Let's use ANOVA to determine if there is a significant difference between the groups. We will use a significance level of alpha = 0.05.

ANOVA TABLE

	df	sum_sq	mean_sq	F	PR(>F)
masks	1.0	0.000002	0.000002	0.188835	0.665796
Residual	49.0	0.000567	0.000012	NaN	NaN

Table 3

The p-value from our ANOVA table is 0.67 which is greater than the significant level, so we fail to reject the null hypothesis. Therefore, there is not a statistically significant difference between the groups, meaning jurisdictions that required masks in public did not have significantly reduced death rates per capita than states that did not require masks.

Let's look at the amount of travel restrictions that were put in place.

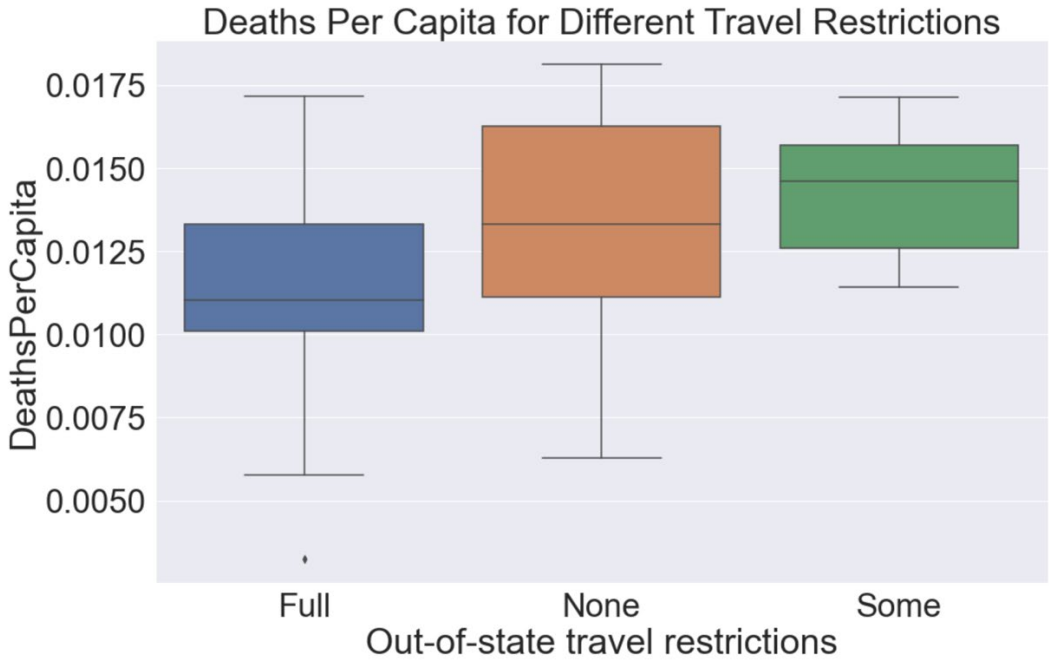

Figure 18

Let's use ANOVA again to determine if there is a significant difference between these groups. Again, we will use a significance level of alpha = 0.05.

ANOVA TABLE

	df	sum_sq	mean_sq	F	PR(>F)
travel	2.0	0.000067	0.000034	3.20711	0.049282
Residual	48.0	0.000502	0.000010	NaN	NaN

Table 4

From the ANOVA results, we get a p-value of 0.049 which is just below the significance level, therefore is significantly different.

Let's look at the lockdown severity feature.

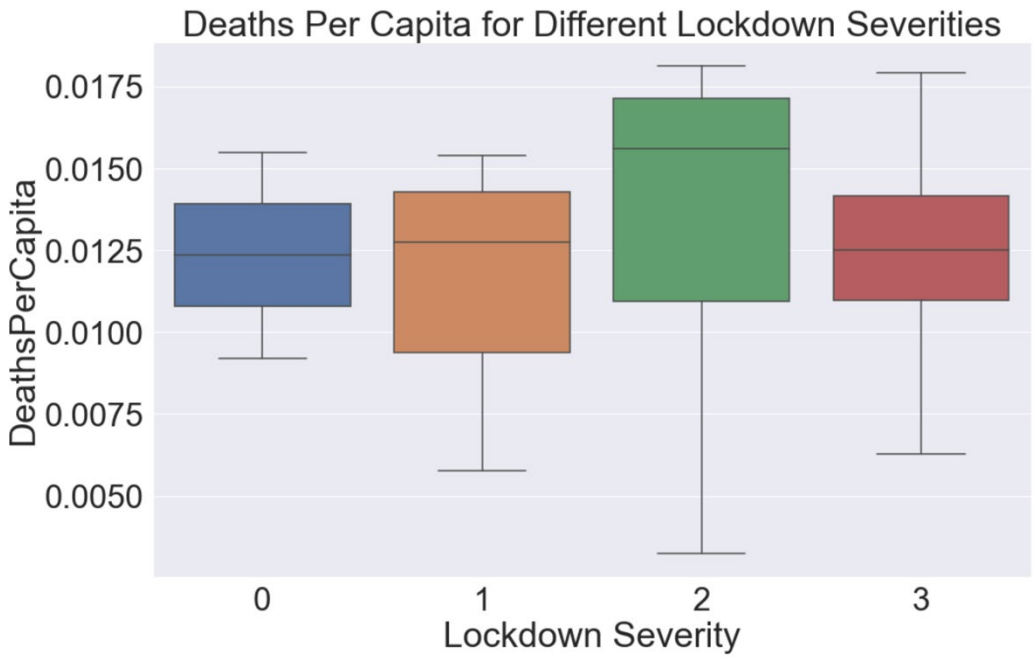

Figure 19

ANOVA TABLE

	df	sum_sq	mean_sq	F	PR(>F)
severity	1.0	3.957225e-07	3.957225e-07	0.034084	0.854291
Residual	49.0	5.689057e-04	1.161032e-05	NaN	NaN

Table 5

From the ANOVA results, we get a p-value of 0.854. Therefore, we fail to reject the null hypothesis and assert that there is not a statistically significant difference between the deaths per capita of jurisdictions that had different lockdown severity levels.

Now let's look at some more useful plots.

Let's plot the most jurisdiction deaths mask mandate in 2020.

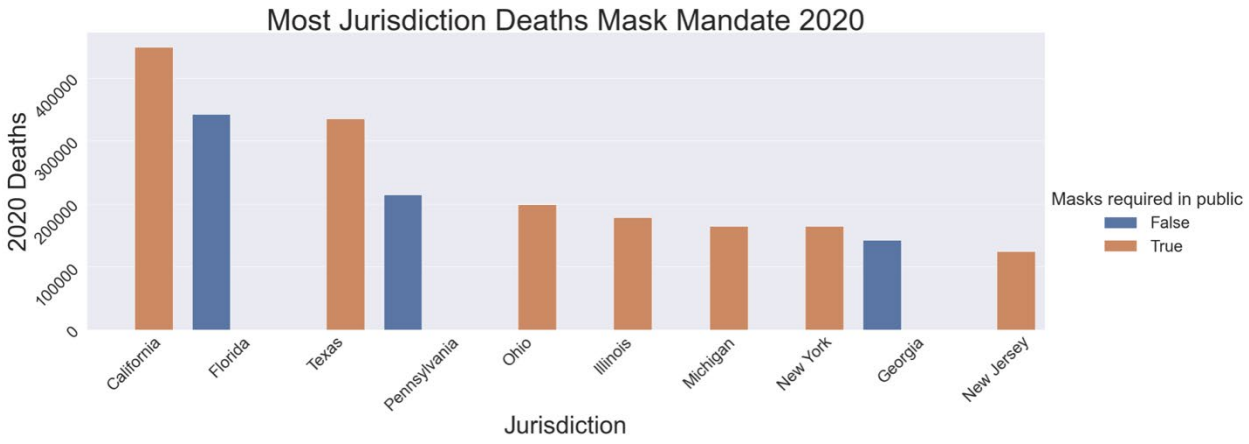

Figure 20

The bar plot visualization above in figure 20 displays data that shows how effective mask mandates were in the top 10 Jurisdictions with the most deaths. Brown bars indicate masks required in public and blue bars indicate no masks required in public. According to the plot, 7 of the top death jurisdictions required masks in public and 3 of the top death jurisdictions did not require masks in public. The number one death state (California) required masks in public and the number two death state (Florida) did not require masks in public. The plot indicates that mask mandates did not make a difference.

Now, let's plot the least jurisdiction deaths mask mandate in 2020.

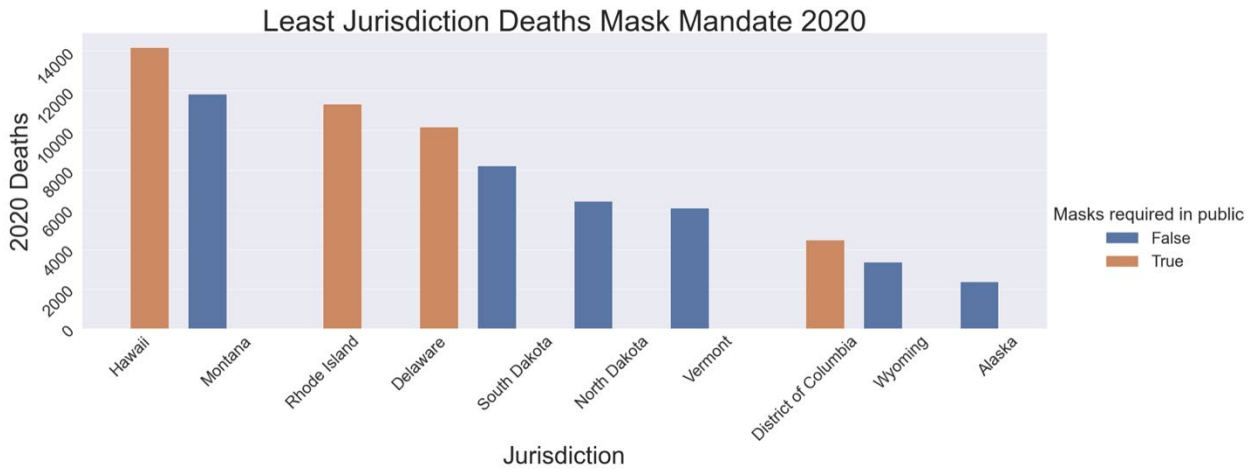

Figure 21

Figure 21 displays the least jurisdiction deaths mask mandate in 2020. It shows that 4 of the least jurisdictions required masks and 6 of them did not require masks.

Now, let's plot the deaths per capita instead of deaths.

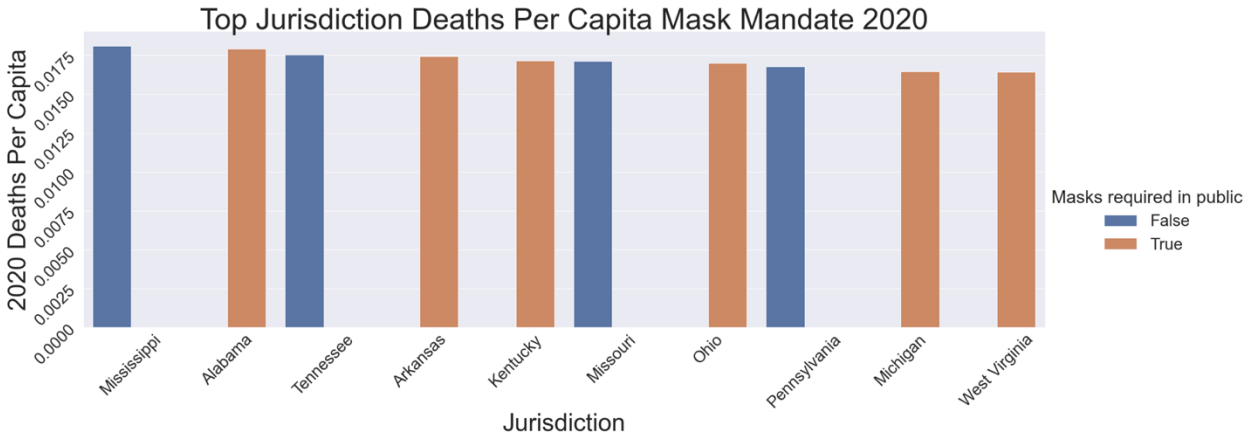

Figure 22

Figure 22 above displays the top 10 jurisdictions with the highest deaths per capita rate. The visualization shows that 6 of the top deaths per capita jurisdictions required masks and 4 of them did not require masks.

Deaths per capita differ from deaths such that California was the jurisdiction with the most deaths and Mississippi was the jurisdiction with the highest deaths per capita rate.

Let's look at deaths per capita for every jurisdiction and sort them from highest deaths per capita to lowest deaths per capita.

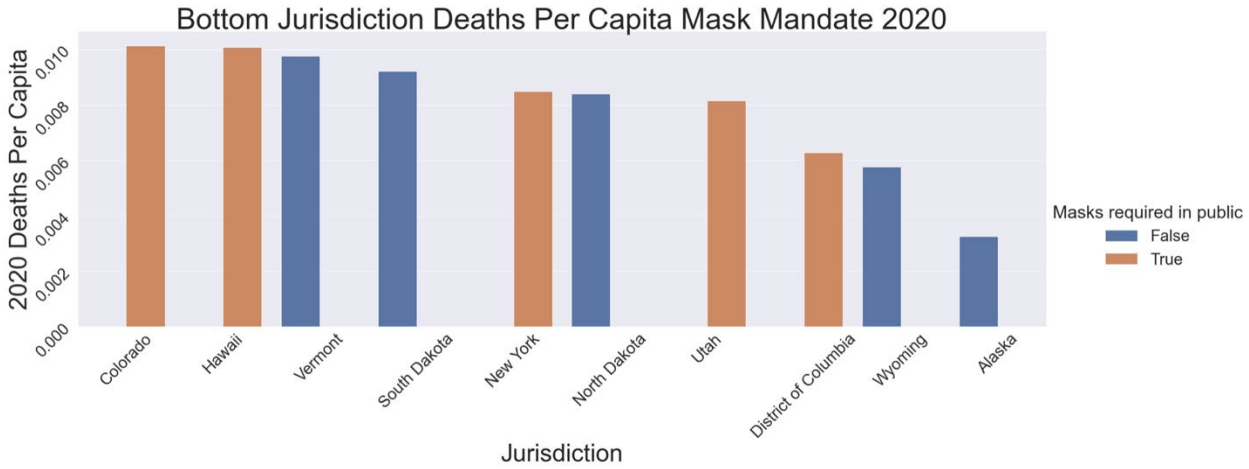

Figure 23

The plot Figure 23 shows the jurisdictions with the lowest deaths per capita rate. It visualizes that 5 of the jurisdictions required masks and the other 5 did not require masks.

Let’s look at gatherings banned.

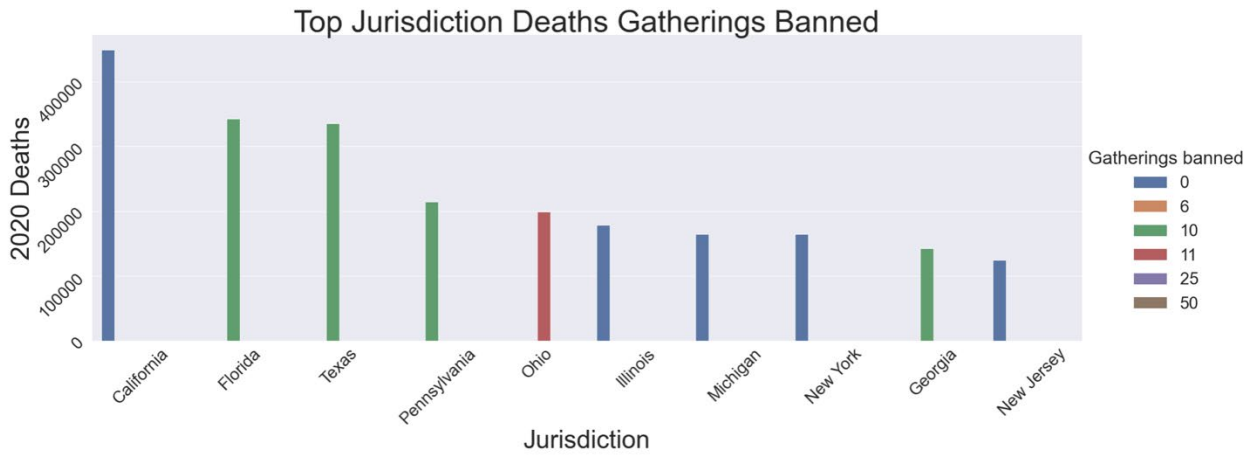

Figure 24

The top jurisdiction deaths gathering banned plot shows that California had the most deaths and also had all gatherings banned. Florida had the second most deaths and had groups of 10 gatherings banned.

Let’s plot gatherings banned by deaths per capita.

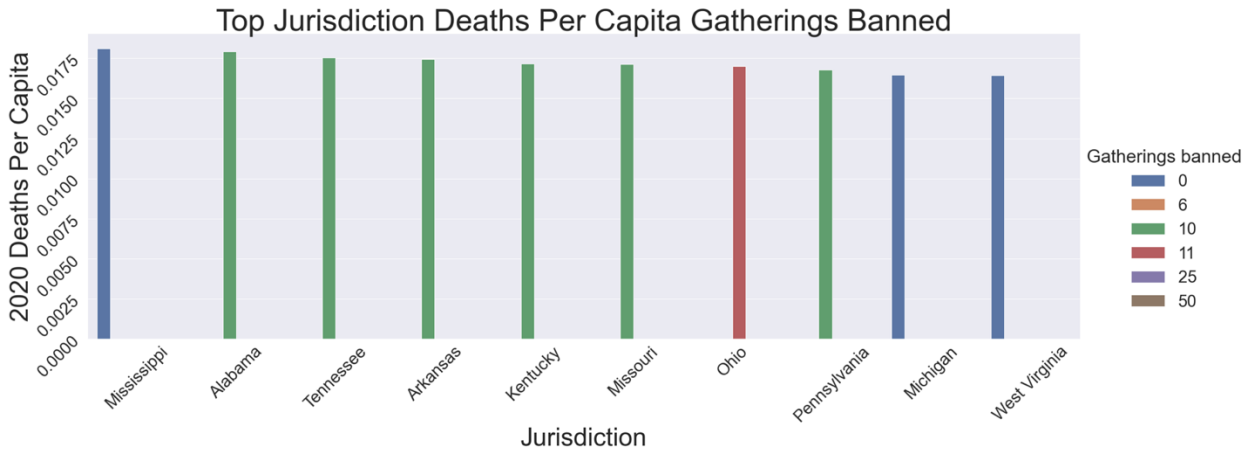

Figure 25

From the deaths per capita gatherings banned plot above, you can see that Mississippi had the highest deaths per capita rate and had all gatherings banned.

Conclusion

Overall, from our computations, plots and analyses, we can come to the conclusion that the different lockdown levels did not have a significant effect on the death rates.

One of our strangest findings was that states with less severe lockdowns also had the same deaths per capita rates as states with more severe lockdowns.

We found that a lot states with mask mandates had around the same number of deaths and deaths per capita rate as states that didn't have a mask mandate.