Your site hasn't been launched yet. Only you can see it until it is launched.

KW Data Stories

Data Science Journalism

US COVID-19 Data Analysis

Project analysis of the New York Times COVID-19 data. The story that we get from this data is difficult to take. It shows that COVID-19 is more deadly depending on where you are in the US and isolation lowers the chances of infection and death. It's a tough story to wrap our heads around because infection and death should be dependent features.

Project Notebook

The data is puzzling.¶

The Covid-19 pandemic has become a point of high controversy in the US. Opinions have been muddled by different practices of classification and over saturation of data. Some resources claim the entire response to Covid-19 is a political hoax. Fox News suggests that the deaths reported in the US are inflated. However, the entire world seems to have gone into lockdown. Here we will attempt to gain insight on the integrity of the Covid-19 data provided graciously by the New York Times. We will also reference the 2019 US Census data for perspective.

US Population by State
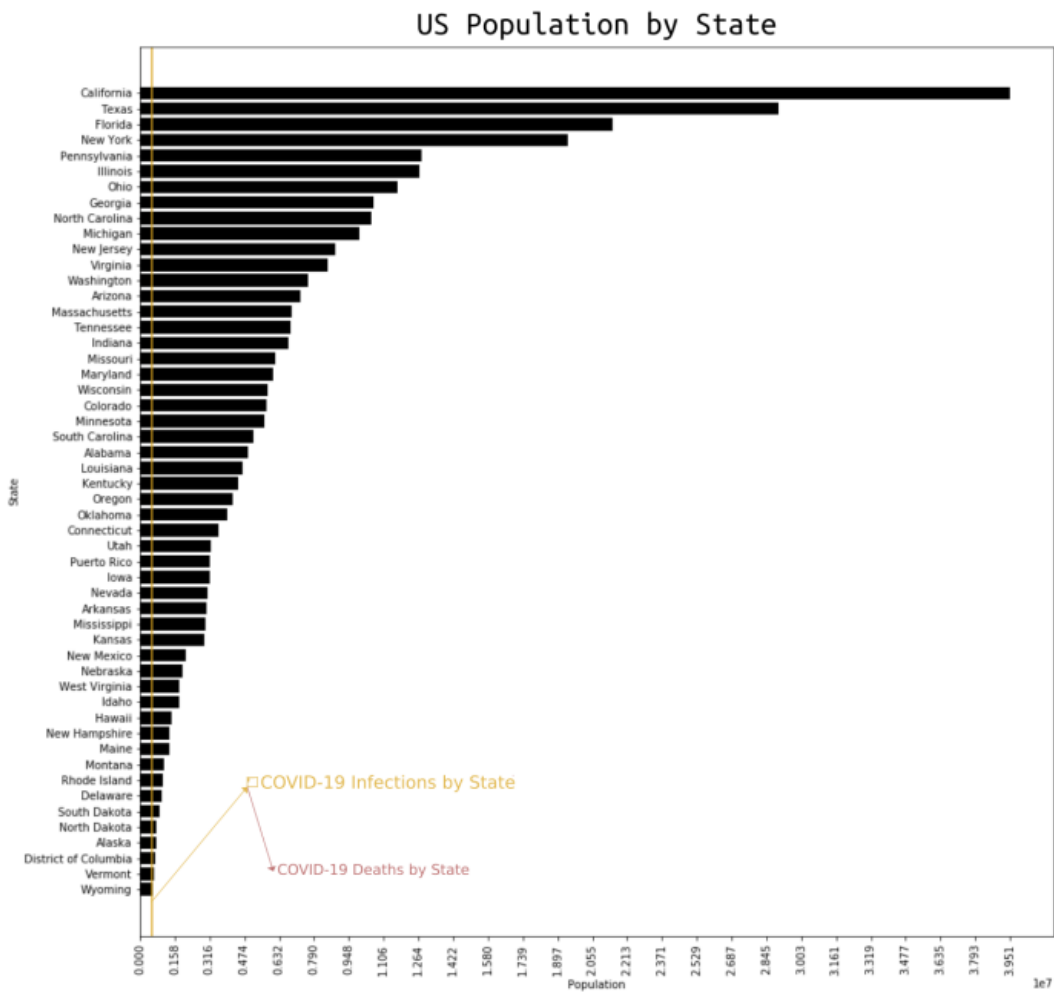
2019 US Census data by state ranked by population. The COVID-19 Infections Graph begins at the yellow line. COVID-19 deaths graph are shown at the red line. These graphs are shown by the yellow and red arrows to explicitly show that the following graphs are not measured by the same x-axis. The red line does not appear on the populations graph.

We can see by the vertical yellow line that we begin to measure infections at a relatively low number when compared to the entire population. Be aware that the vertical yellow line is denoted by the state with the highest amount of infections in the US, which we will be able to see clearly in the plot below. According to our data, the average amount of infections by population is 1.3%. Another point that must

be made is we are not measuring time. The number of people that have been recorded as infected are not the same as the number of people that are currently infectious. Although these numbers are many, this data could support that the CDC preventative measures are working or the virus is not being properly measured. Perhaps a mix of both.

COVID-19 confirmed and probable infections in the US by state from the New York Times data set.

The infections plot carries a similar distribution to its parent populations plot. At a glance, we could say that the data that shows a higher population contributes to a higher number of people infected, until we get to Washington state. Washington appears to have less COVID-19 activity when compared to other states with similar populations. If we check the states with the most aggressive counter measures to COVID-19, Washington state ranks in 5th place. Some aggressive measures from Washington state include a 4 step data-driven plan with excellent feature definitions. Unfortunately, our New York Times data set does not show evidence that these aggressive measures have helped reduce the amount of COVID-19 activity when compared to the other 9 states on the list.

COVID-19 confirmed and probable deaths in the US by state from the New York Times data set.

When viewing the deaths chart, we should expect the amount of deaths reported to be strongly correlated with the amount of infections, however this is clearly not the case. Alaska reports deaths per infections as 0.5% and Connecticut has a reported death rate of 8.83%. Either something is causing COVID-19 deaths to have a wide range of survival rates from state to state or the reporting is incorrect, there are arguments for both sides. The data says that the average death rate in the US is nearly 3%, this is especially troubling when compared to Germany's metric of 0.4%.

The argument in this article is that Germany has tested more of the population and found more infections that are asymptomatic resulting in a higher infection rate and lower death rate. Fox News reports that people who die with COVID-19 are being reported as a victim of COVID-19 regardless of other factors contributing to death. Without an international standard for logging these deaths the above comparison carries little weight.

To dive deeper into the data, Below we will observe a table of the infection rate and death rate for each state as well as some summary statistics.

	Infections	Deaths		Infections	Deaths
California	1.44%	1.83%	Oregon	0.51%	1.69%
Texas	1.76%	1.77%	Oklahoma	1.11%	1.38%
Florida	2.50%	1.54%	Connecticut	1.41%	8.83%
New York	2.19%	7.60%	Utah	1.38%	0.78%
Pennsylvania	0.97%	5.94%	Puerto Rico	0.71%	1.22%
Illinois	1.56%	3.98%	Iowa	1.56%	1.90%
Ohio	0.87%	3.61%	Nevada	1.85%	1.69%
Georgia	1.91%	2.05%	Arkansas	1.64%	1.10%
North Carolina	1.31%	1.60%	Mississippi	2.27%	2.83%
Michigan	0.98%	6.70%	Kansas	1.10%	1.21%
New Jersey	2.10%	8.50%	New Mexico	1.06%	3.07%
Virginia	1.18%	2.31%	Nebraska	1.47%	1.24%
Washington	0.86%	2.72%	West Virginia	0.43%	1.82%
Arizona	2.58%	2.22%	Idaho	1.39%	0.96%
Massachusetts	1.76%	7.22%	Hawaii	0.25%	0.86%
Tennessee	1.77%	1.01%	New Hampshi	0.50%	6.13%
Indiana	1.14%	3.97%	Maine	0.30%	3.09%
Missouri	0.98%	2.32%	Montana	0.47%	1.49%
Maryland	1.60%	3.71%	Rhode Island	1.88%	5.09%
Wisconsin	1.12%	1.54%	Delaware	1.61%	3.78%
Colorado	0.88%	3.67%	South Dakota	1.09%	1.51%
Minnesota	1.09%	2.76%	North Dakota	1.01%	1.52%
South Carolina	1.96%	2.03%	Alaska	0.61%	0.54%
Alabama	2.10%	1.74%	District of Co	1.81%	4.61%
Louisiana	2.84%	3.25%	Vermont	0.23%	3.97%
Kentucky	0.82%	2.17%	Wyoming	0.53%	0.92%

Summary Stats	Infections	Deaths
Minimum	0.23%	0.54%
Maximum	2.84%	8.83%
Range	2.61%	8.29%
Average	1.32%	2.90%

Table of New York Times data set, COVID-19 infection and death ratios. Where infections = infections / population and deaths = deaths / infections. Ordered by descending population size.

The most interesting thing about this table is the range of deaths vs the range of infections. If we assume the behavior of the COVID-19 virus is a constant, this range should provide little variation. Instead, the data shows that there is a range of

8.29% death rate depending on which state you are in. This could be evidence that hospitals in states that are showing above a 5% death rate are unable to provide the same treatment as other states, or the data is not getting reported in the same manner.

So what states are getting the best COVID-19 metrics?

Both Alaska and Hawaii have infection rates and death rates lower than 1%. This could be an indicator that isolation is the best preventative measure to COVID-19.

The story that we get from this data is difficult to take. It shows that COVID-19 is more deadly depending on where you are in the US and isolation lowers the chances of infection and death. It's a tough story to wrap our heads around because infection and death should be dependent features. The range brings up many questions worth investigating.

Sources

Kaggle Covid-19 Dataset – Scientific community call to action

- Kaggle

New York Times COVID-19 data

- GitHub

Coronavirus myths: Don't believe these fake reports about the deadly virus

- Cnet

2019 Census Data

- Census.gov

Cases & Deaths by County

- CDC

Coronavirus hype biggest political hoax in history

- Washington Times

Fox News shares that the deaths reported in the US are inflated

- Fox News

government is classifying all deaths of patients with coronavirus as 'COVID-19' deaths, regardless of cause

- Fox News

Safe Start Washington

- Washington State

10 States with the Most Aggressive Response to Covid

- US News

Why is Covid-19 death rate so low in Germany?

- CNN

[Our ongoing list of how countries are reopening, and which ones remain under lockdown](#)

- Business Insider

[CDC's 'best estimate' is 40 percent COVID-19 infections are asymptomatic](#)

- Fox News

[Population density in the U.S. by federal states including the District of Columbia in 2019](#)

- Statista

Like

Be the first to like this.

Kalen Willits / August 11, 2020 /

KW Data Stories / Create a free website or blog at WordPress.com.