Descriptive Statistics

The data in the following descriptive statistics and analysis were accessed on April 15, 2020 at about 2:53pm EST. First, I used Minitab to draw some pictures of the data and run some descriptive statistics.

Descriptive Statistics

Variable	Mean	Standard Error of Mean	Standard Deviation	Minimum	Median	Maximum	Mode	Number of times Mode
Deaths	529	229	1652	1	120	11,586	9	3
Deaths per 1M pop	58.2	13.9	99.9	1.7	23.9	595.6	*	*
Tests	60,640	11,036	79,578	6,129	35,925	499,143	*	*
Confirmed Cases	11,855	4,107	2,9618	285	3,445	203,123	*	*
Confirmed Case Fatality Rate	3%	0.1%	1%	.3%	3%	7%	3%	3
Confirmed Cases per 1M pop	1,476	261	1,886	305	768	10,441	*	*
Seasonal Flu Deaths (10Yavg)	797	120	864	43	542	4,701	*	*

As we launch into my discussion of these numbers, I feel called to give a huge shout out to my major doctoral advisor, Edward Kifer. Anyone who has worked with Kifer and/or taken a class with him and/or drank wine with him may be able to hear these words as clearly as I do each time I review descriptive statistics, “never a center without a spread” (Kifer, 1970s+).

“Never a center without a spread” seems like a simple statement. It’s simple to remember and to restate. For me, it is seminal. The subsequent sections reflect my attempt to make sense of these data in terms of Kifer’s keenly profound contributions.

The Confirmed Case Fatality Rate makes me hopeful. These numbers represent human beings. It looks as if our states are ranging between a .3% to 7% in Confirmed Case Fatality Rates. Nationally, we seem to be hovering around an average fatality rate of 3%. I’m suggesting we are hovering because the standard deviation at 1% suggests our fatality rate across the states is not very spread out contrasted to some of the other standard deviations for this dataset.

Now to the crucial and strange data for the variable, Tests.

One state seems to have tested 6,129 human beings; another state has tested 499,143 human beings (as of April 15, 2020 at 2:53pm EST). On the face of it, these numbers seem spread apart.

In a forthcoming blog I will be sharing Test numbers in terms of per capita data. Because these data lead me to wonder the reasons why the spread is so huge. It might be that the state that has tested 6,129 human beings happens to be a less populated state than the state that has tested 499,143 human beings. Because population size is one factor that would seem reasonable for the range of tests that have been conducted.

The median number of tests across our states is 35,925. This median, 35,925, is the number directly in the middle of all the numbers spread out from 6,129 to 499,143.

Or, maybe these numbers are spot on if we are testing people in a similar fashion as what Stephen J. Rose teaches me about our social strata (Rose, 2015).

The average number of tests we have given is 60,640 with a standard deviation of 79,578. Does it matter to us that our standard deviation is higher than the average number of tests we have given?

This means we are 68% confident that most states have tested between 6,129 (our lowest number of tests… it doesn't make sense to subtract 79,578 from 60,640 because that will give us -18,938. It's nonsensical to offer a negative number for the lower end of the 68% confidence interval because we know the lower end of this spread. And, like Kifer suggests, I am the boss person of this blog, so I am using the lower number of the range of tests conducted here not -18,938.) and 140,218 human beings for COVID-19.

We are 95% confident that most states have tested between 6,129 and 219,796 human beings for COVID-19.

We explore more about the crucial and strange data for the variable, Tests, in upcoming sections.

Boxplots

Boxplots depict a lot of data in a way that visually shows us how common and/or uncommon the numbers are across the states. We see how close together and/or how spread apart the states are in terms of the variables, Deaths, Deaths/1Mpop, Tests, Confirmed Cases, Confirmed Case Fatality Rate, Confirmed Cases/1Mpop and Seasonal Flu Deaths (CDC10-Year Avg).

These boxplots seem to show us that New York tends to be the state with the highest number of deaths, confirmed cases, confirmed cases per 1 million people and seasonal flu deaths (10 year average).

It's quite interesting that New York does not stick out, with its own asterisk data point, on the Boxplot of Confirmed Case Fatality Rate. For these data, Michigan had the highest Confirmed Case Fatality Rate at 6.55%. New York's Confirmed Case Fatality Rate was 5.7% which is a rate within a common range of other state's Confirmed Case Fatality Rate. And, could be why New York doesn't stand out in this boxplot.

I spent quite a bit of time questioning New York's data in relation to other states. New York seems higher than other states in most of the pictures below. When we see data points far away from the rest of the other data, such as New York looks, we wonder if we have an outlier.

Outliers can cause problems in the analyses. Outliers may suggest some sort of measurement error or data collection error or some contextual factor(s) specific to the outlier. In this exploration, New York seems so far away from the rest of the states it could indicate that New York is unlike the majority of the states. I needed to consider the possibility that New York was unlike most of the other states.

Because I'm a nerdy type who gets exhilarated by finding "something unexpected that sometimes results in restructuring the whole direction of a study," I called a few other measurement folks on my iphone and we had some delightful banter about whether to consider New York an outlier or an extreme case. The banter turned more intense as we discussed whether to keep New York in the data set or remove New York. I ran the descriptives and forthcoming analysis with and without New York. The data and the context within which we find ourselves leads me to keep New York in the dataset. New Jersey's data brought on similar questions.

New York City, New York, is one of our most densely populated cities with (before shelter in place orders) many human beings moving around a lot within the city. If memory serves, New York was one of the first in our country to deal with multiple COVID-19 cases. New York seems to have been one of the first to start calling for national help. These "firsts" matter because they could show us what the rest of our states could have looked like had we not taken medical and health sciences recommendations when we did.

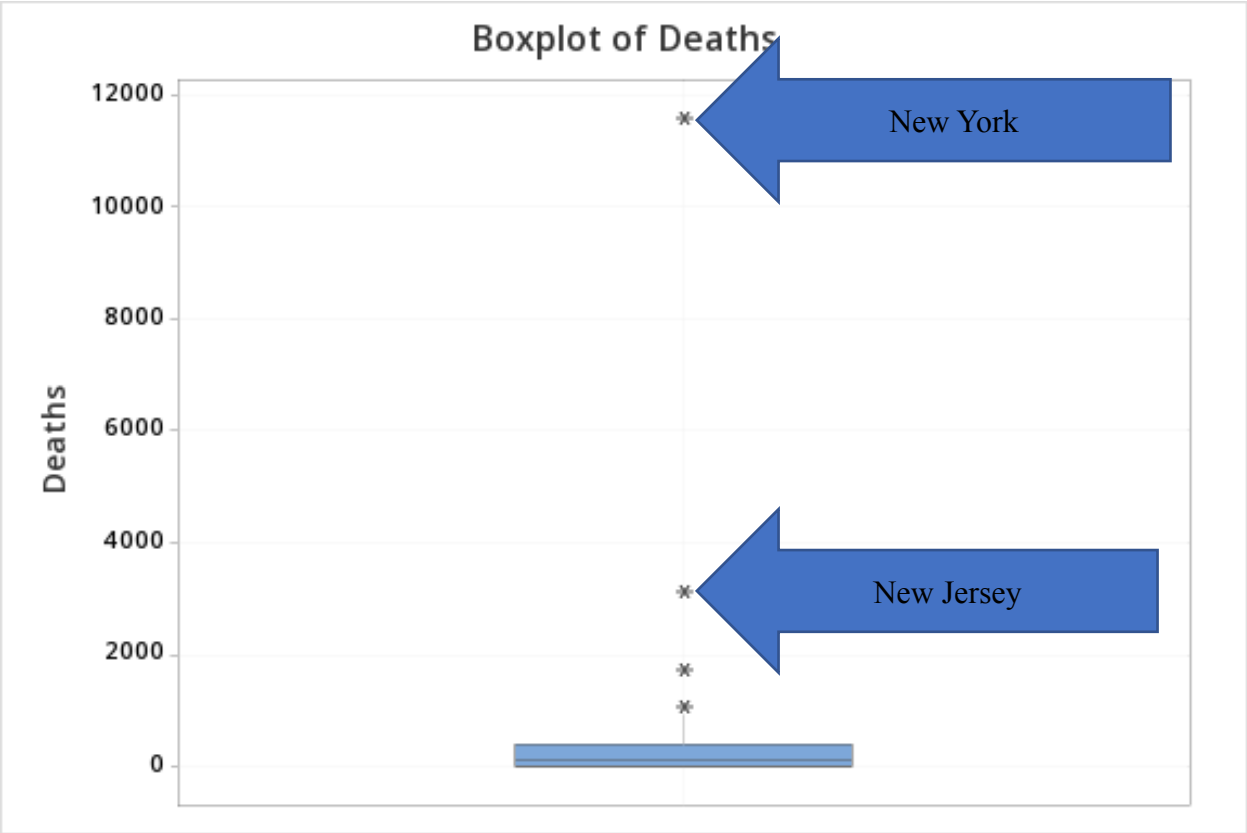
If this was being submitted to an academic journal for publication, I would have to consider labeling New York an extreme case. Even though the reality New York is living with may be extreme contrasted to the data for our other states, we all may be feeling that we are living with extremely challenging circumstances.

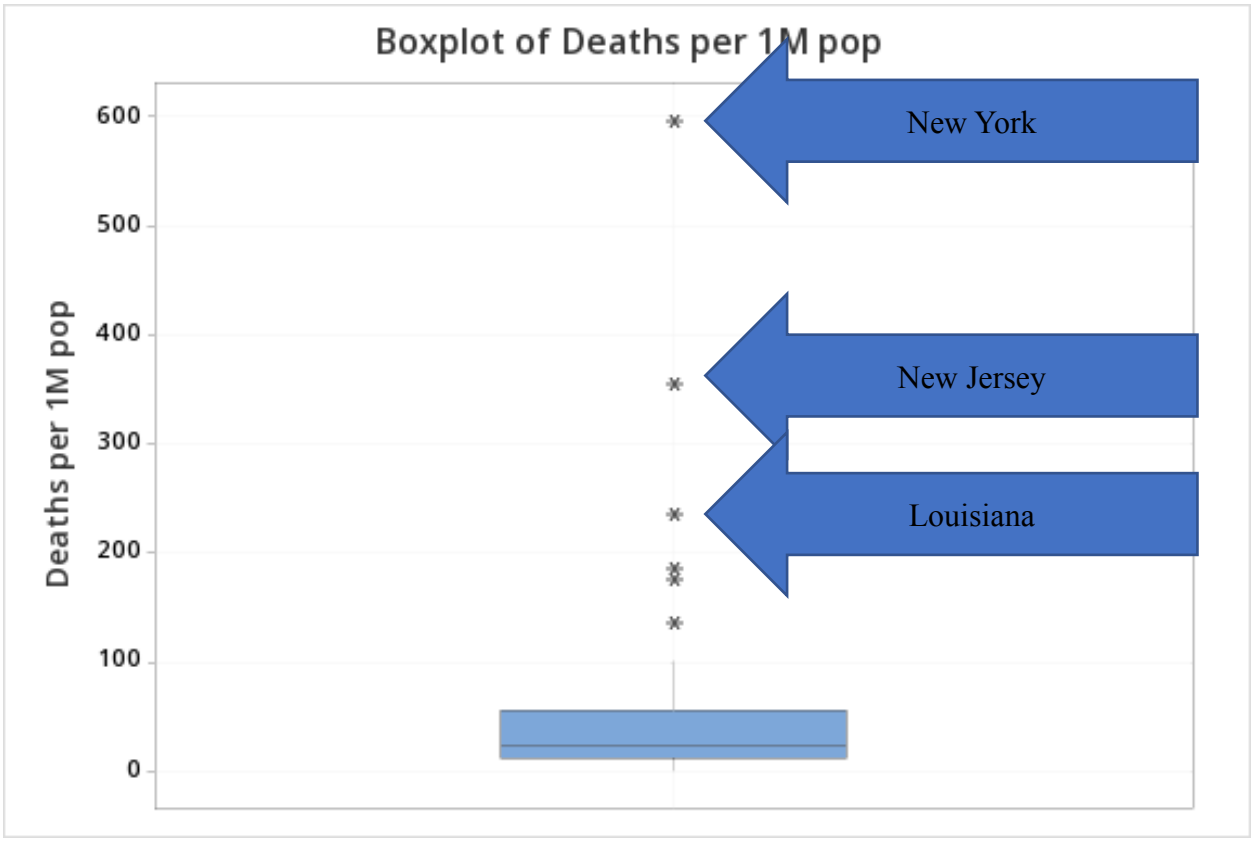
Additionally, New York shows us the reality we could all be living if we had not implemented social distancing and shelter in place orders.

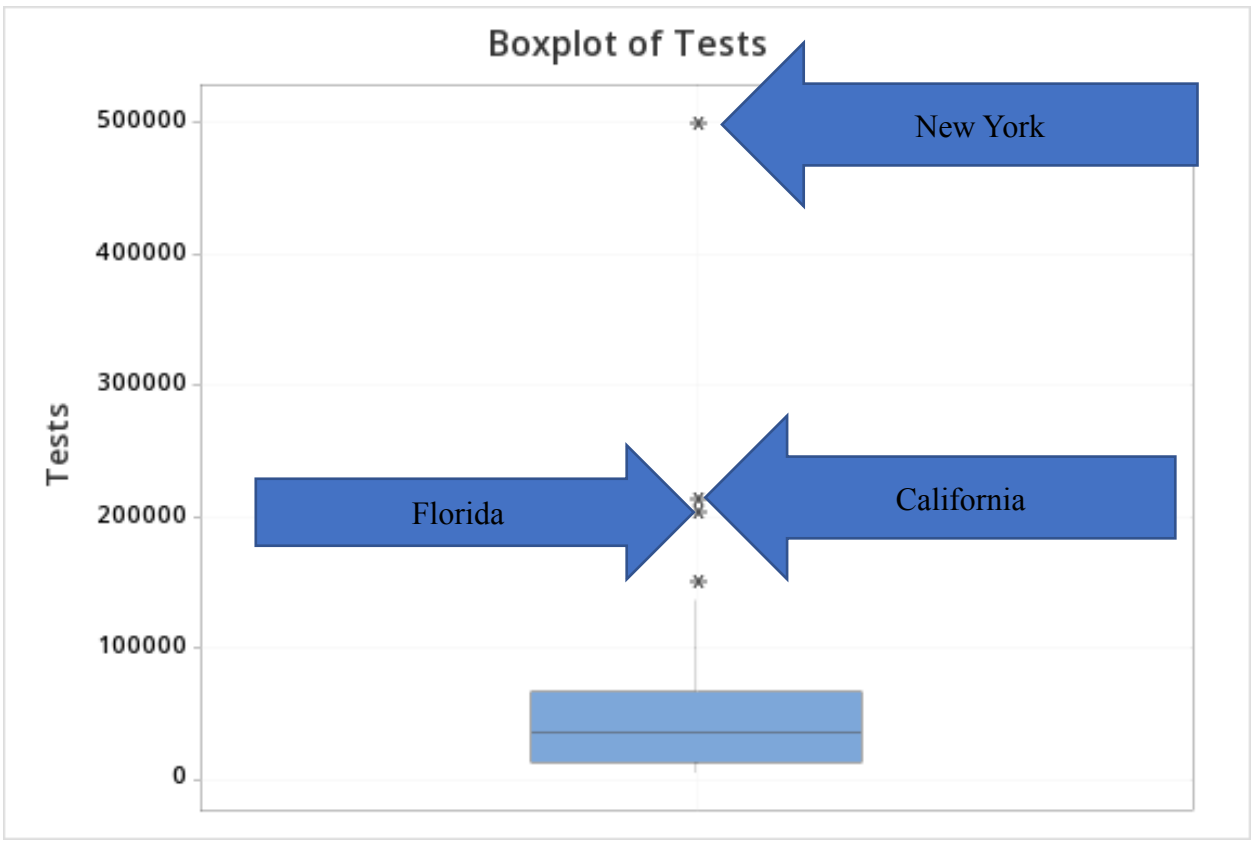
As we continue to consider New York, it seems relevant to remind us that the data analyzed here are dynamic and change from moment to moment. My bias is that we need to view New York as reality. Reality because if a state decides to open up too soon, that state could see numbers like New York. This exploration is real time and the stakes are our lives.

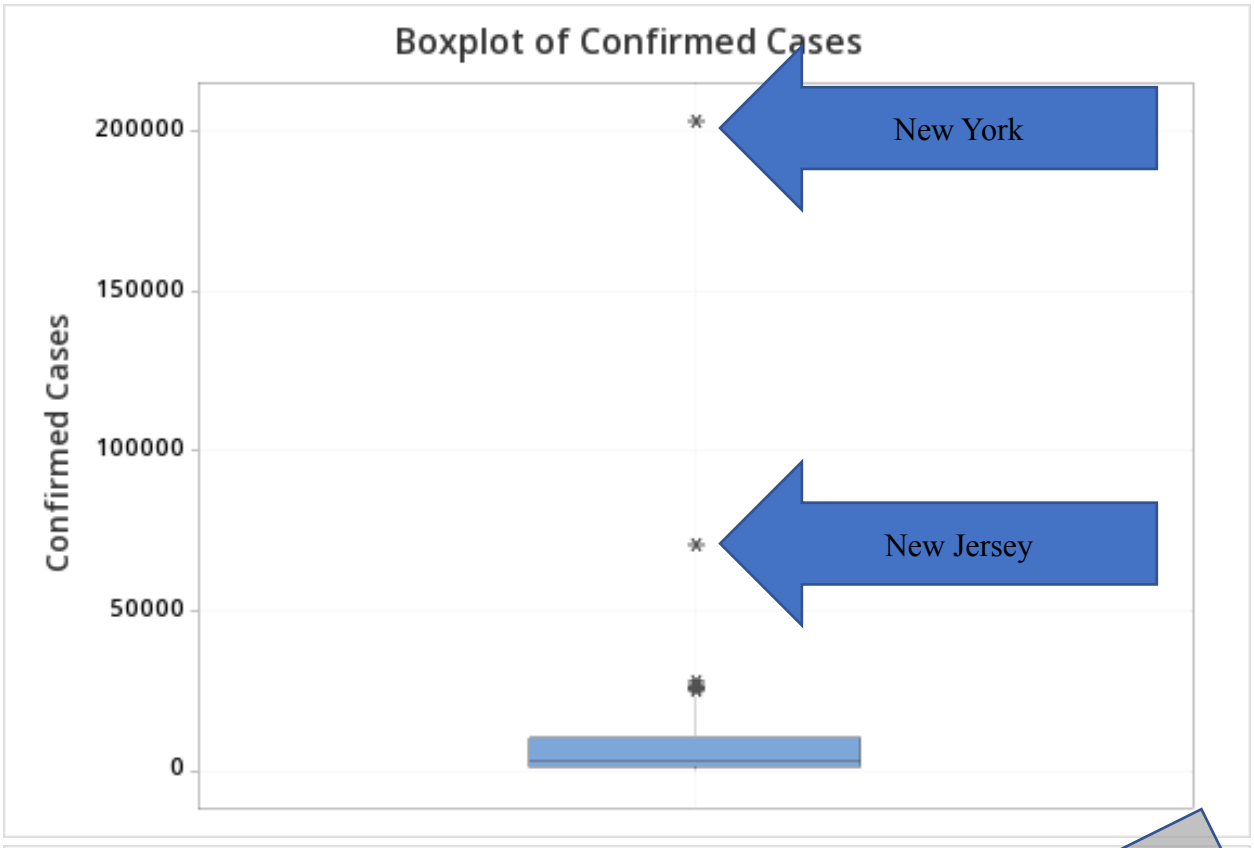
Plus, according to Kifer, I'm the boss person! It's my blog. Hence, the label I give New York is: reality.

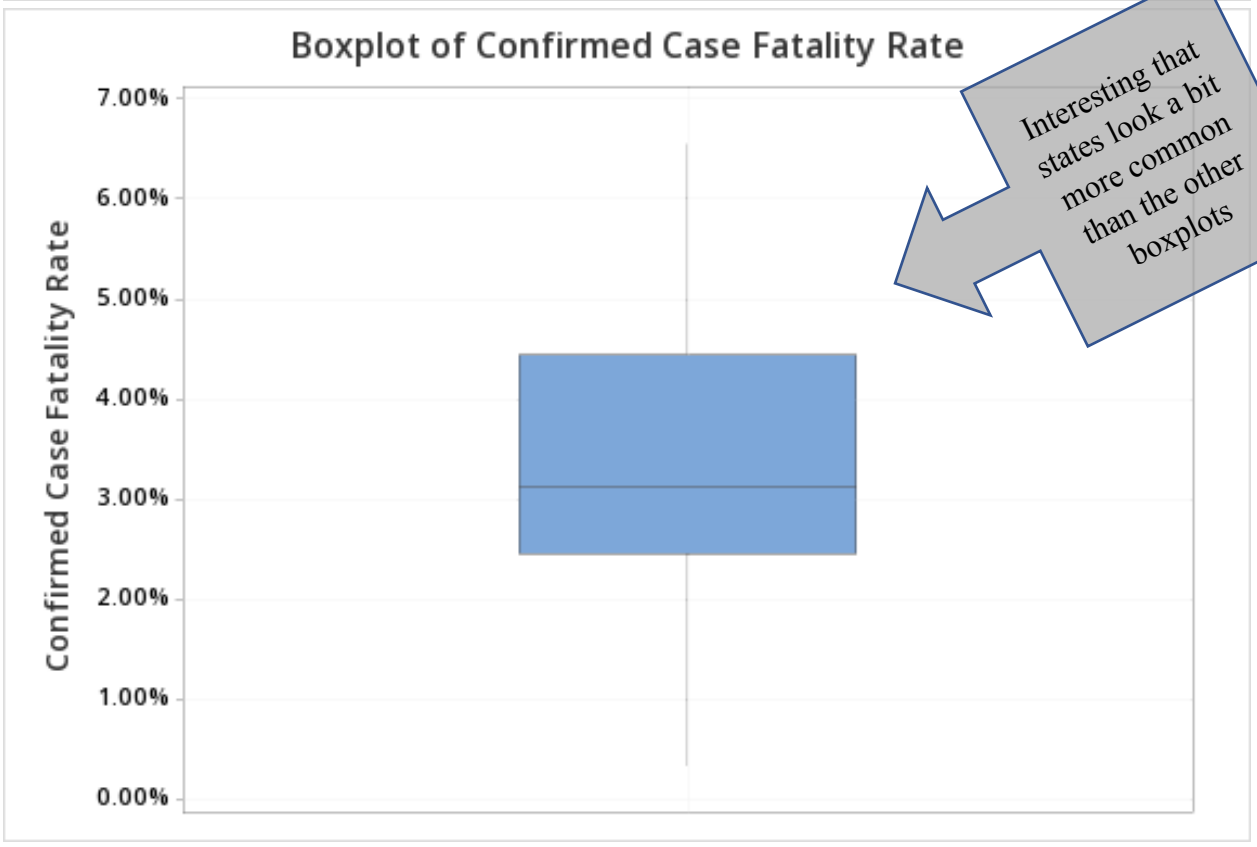
More about New York in forthcoming blogs in this series.

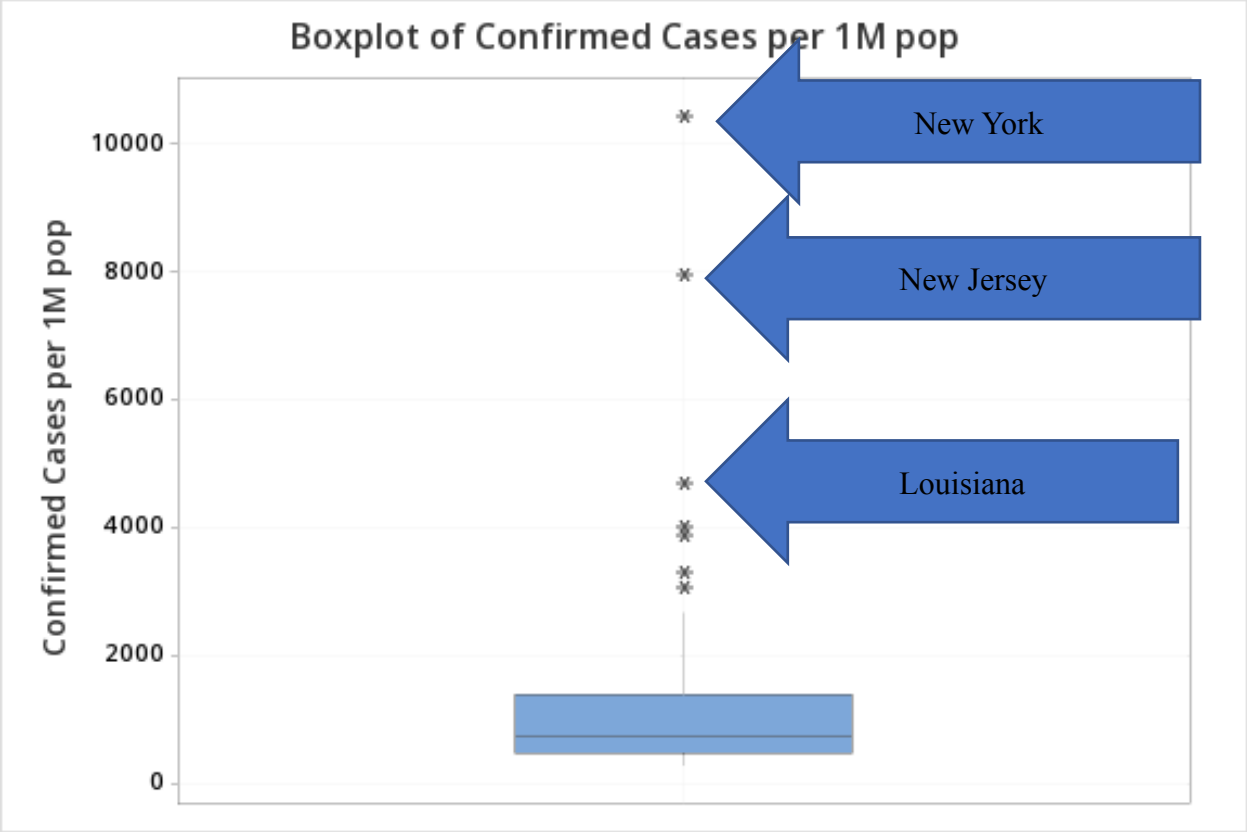


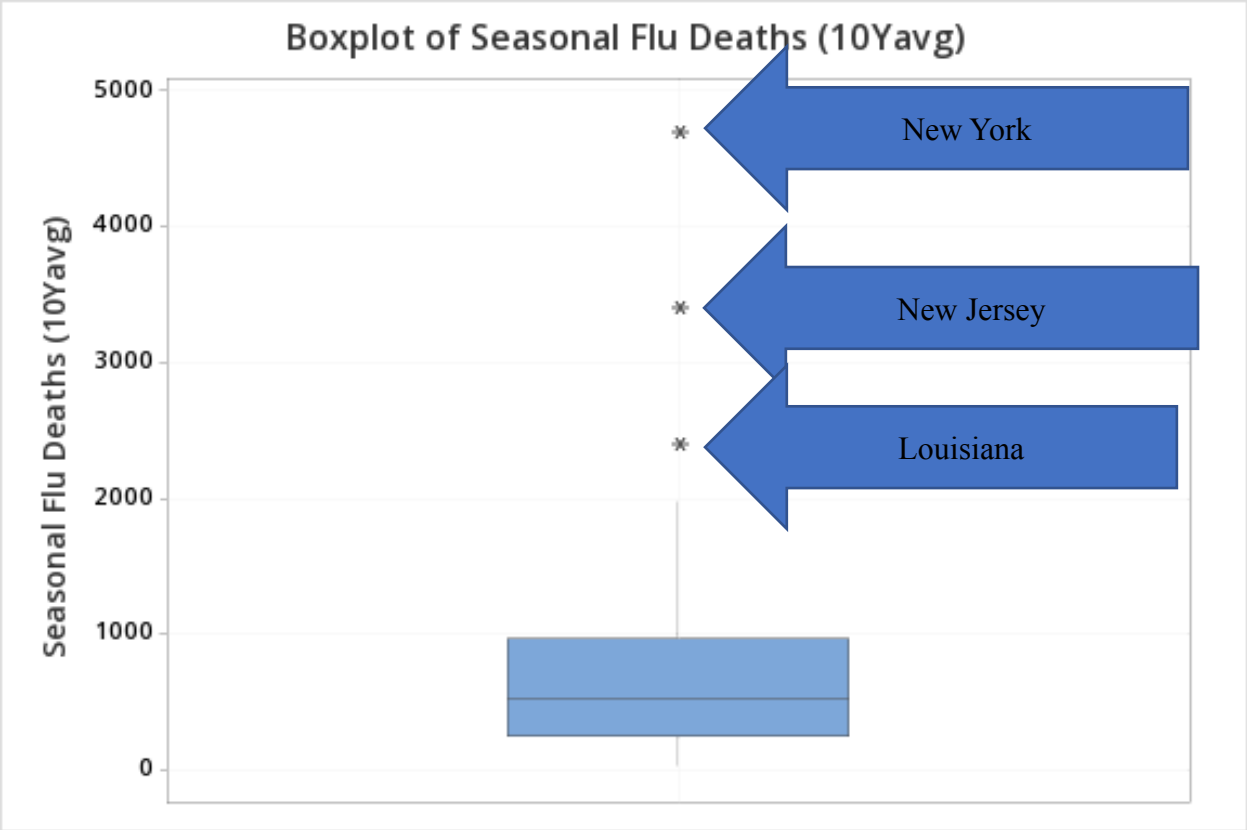

Boxplot of Deaths per 1M pop
Deaths per 1M pop
600
500
400
300
200
100
0
New York
New Jersey
Louisiana


Boxplot of Tests
Tests
500000
400000
300000
200000
100000
0
New York
Florida
California


Boxplot of Confirmed Cases
Confirmed Cases
200000
150000
100000
50000
0
New York
New Jersey


Boxplot of Confirmed Case Fatality Rate
Confirmed Case Fatality Rate
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
Interesting that states look a bit more common than the other boxplots


Boxplot of Confirmed Cases per 1M pop
Confirmed Cases per 1M pop
10000
8000
6000
4000
2000
0
New York
New Jersey
Louisiana


Boxplot of Seasonal Flu Deaths (10Yavg)
Seasonal Flu Deaths (10Yavg)
5000
4000
3000
2000
1000
0
New York
New Jersey
Louisiana