**Are we reopening safely?**

In the United States, the response to COVID-19 has differed from state to state. Some states, such as California and New York, were hit hard, early. Others are just now seeing the spikes that prompted widespread lockdowns earlier in the year. But are these new spikes the result of controlled reopening or a precursor to a new onslaught of COVID-related deaths?

Put simply: as states reopen at different rates, can we tell if they have learned the hard lessons from the spring? Do states with spikes now have the same death rates and states with spikes earlier?

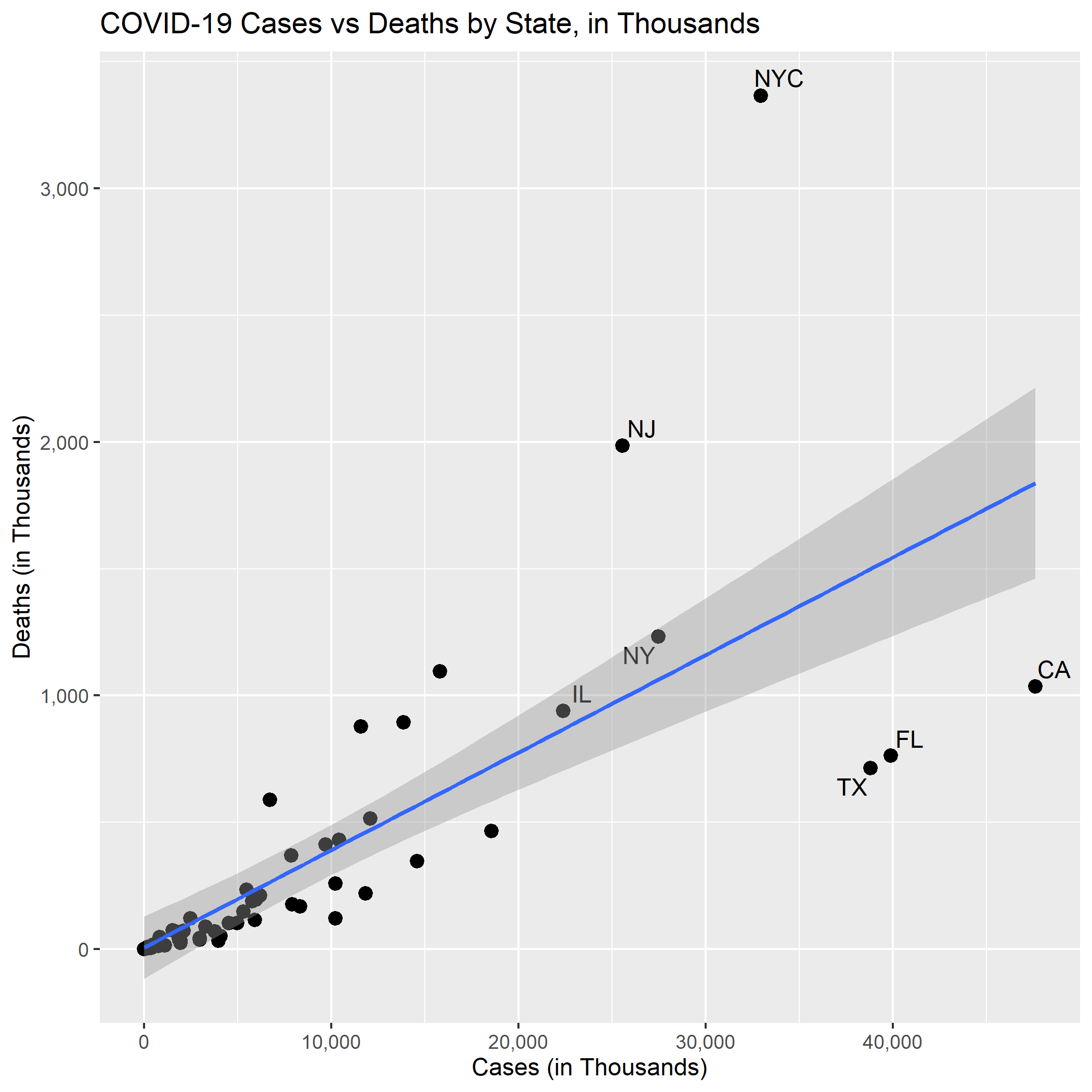
Let’s take a look.

First, we download case and death data from the Center for Disease Control (CDC) website: <https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36>. Note that the CDC breaks out some cities, like New York City, into different “states” because of the prevalence of the virus in those locations. For the purposes of this analysis, we will consider these cities separately from their states, per the CDC data.

To make sure that we are considering the severity of cases across different population areas, we want to define what a “significant” number of cases are. For the purposes of this analysis, a state will reach a “significant” number of cases when the case rate is more than 1% of the population. To do this analysis, we need to download state population data from the US Census Bureau: <https://www.census.gov/newsroom/press-kits/2019/national-state-estimates.html>.

But the Census Bureau writes out states where the CDC writes the abbreviation! So have to download another list to combine the two from the US Postal Service: <https://pe.usps.com/text/pub28/28apb.htm>.

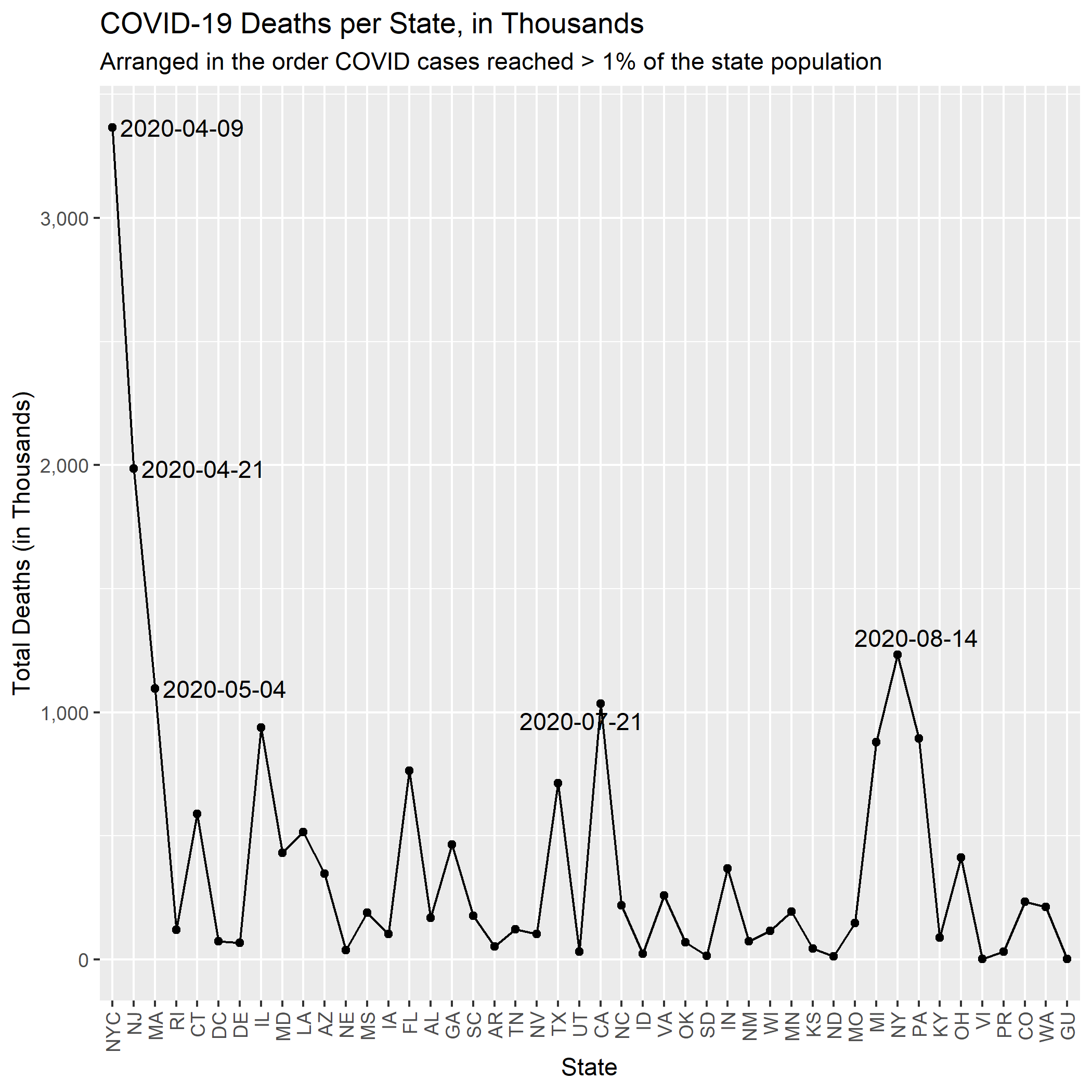
First of all, let’s take an initial look at the Case vs Death rate in different states



Already there are clear outliers in the data. NYC, for example, has the unfortunate distinction of the most number of deaths, even despite it being broken out from the rest of the state. New Jersey, too, has deaths much higher than the proportion of cases it had. Texas, Florida, and California, on the other hand, have high numbers of cases but relatively few numbers of deaths.

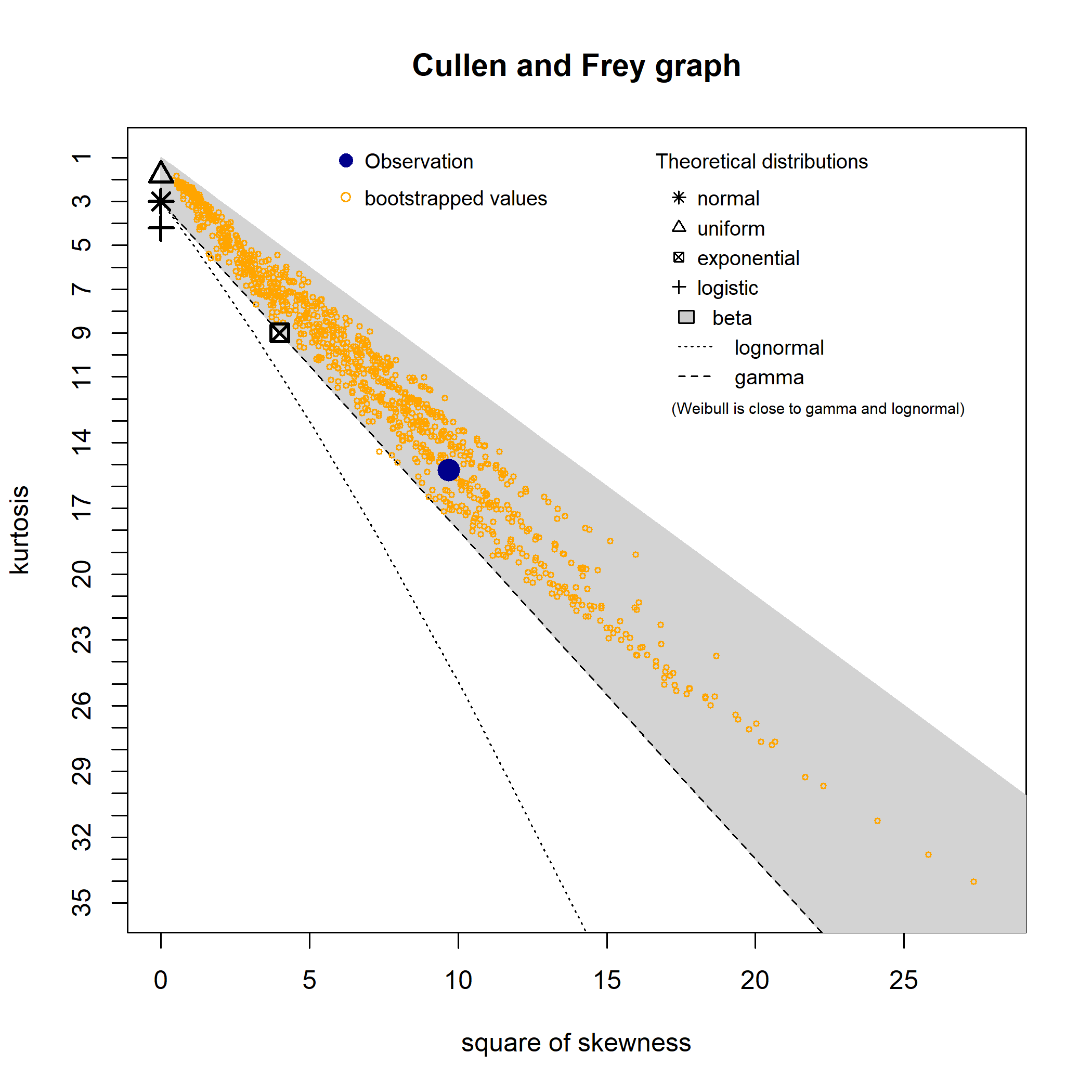
The graph includes a *very rough* linear regression based on just number of cases and number of deaths. Of course, it is much more complicated than this! But this fitted line gives a general understanding of which states have more deaths per cases than normal.

Next, we look at the death rates in different states, arranged in the order of when a state had a “significant” number of cases (>1% of population).



There does not seem to be a clear trend here in the number of deaths over time, either rising or falling. It is important to note that New York falls to the right side of the graph (reaching a significant number of cases on August 14, 2020) even though it had a large share of cases early on.

If states re-opening learned from the challenges faced by other states, we would expect to see a long right tail to this graph. But there is no clear distribution that these deaths over time follow. To check this intuition, we can attempt to fit a distribution to the data using a skewness-kurtosis graph. After bootstrapping 1,000 times, it is still clear that no easily-recognizable distribution will fit our data.



This analysis shows that it is not clear if states seeing surges now are learning the lessons from earlier and reopening safely.