Troy Siegler

INFO 3401

Set 5 Response Doc

1. Pulled Data set off of the internet at <https://github.com/CSSEGISandData/COVID-19>
2. The National Health Commission of the People’s Republic of China (NHC) is a government-run organization whose main purpose is to draft laws, regulations, and policies in order to maintain overall health services in China. Considering this is a government organization that has access to millions of medical records, I believe that the data coming from this organization should be very reliable.

BNO News is a news organization that mainly operates in the United States but covers news from all across the globe. The origin of the data that they are receiving are unknown to viewers, but the organization claims that each figured is 100% verified by their local health departments or local media. Honestly, I see BNO as being reliable. Nonetheless, we still are not sure exactly where all of these numbers and figures are coming from, so we have to be careful when analyzing data like this.

The WorldoMeters website is an internal organization that collects data on the coronavirus from “thousands of sources in real time”. Even though they do not specify where exactly their data comes from, their visualizations, graphics, and statistics are used by other organizations all across the world. I would say their data is reliable.

1. In this problem, we are plotting the “Death”, “Confirmed”, and “Recovered” values from the data sets. All of these data columns are matched up with their respective “Date” and then plotted on the graph. Obviously, as time as gone on the growth rate for deaths and confirmed cases are beginning to slow down (even though they had been exponentially growing for over a month). And the recovered values continue to grow as time moves on.
2. This problem is doing the same thing as the code for numbers 10-11 except this time we are adding in a new variable. We add in the hue value in order to change the color of the lines/variables within our graph. The data looks exactly as one would expect it with recovery rates going up and the correlated death rates (barely) starting to dip down. The Confirmed cases are continuing to rise in numbers.
3. For this problem, I decided to look at and compare the growth rates between the two “early in the outbreak countries” which were South Korea and China and two later ones: Italy and the United States. The US and Italy’s growth rates absolutely blow China and South Korea’s out the water. Based on these results, we can say that the data from the early outbreak countries did not help the regions that experienced the virus later. If the data from China and Korea were more *drastic,* meaning growth rate and numbers were higher, then I’m sure countries like Italy and the US would have prepared better for the epidemic.
4. Most of these rates are strongly correlated one way or another.

Hypotheses:

As the Confirmed Cases go up, so does the Death Rate

As Death Rates go up, so do the recovery rate

As recovery rate goes up, so does confirmed cases

I believe that all of these are positively correlated. The fact that confirmed cases and death goes up as recovery rates go up are contradictory. This happens because so many mire individuals are getting tested now, so the cases and deaths are bound to increase dramatically. And people recovering from the virus will have constant growth at the epidemic progresses.