Ross Davidson, Aaron Lewis and Tyler Ward

**COVID-19 in the USA: Things you didn’t know.**

What has happened so far and how exactly we arrived here - Situational understanding

Goal:

Investigate the differences in infection and death rates between regions, states and individual counties, and the significance of these differences.

Question 1: How did we get here? What brought us to this point in the proliferation of the virus?

1. Cases began to rise toward the end of March with an increase in deaths lagging 10-14 days behind.

A screenshot of a map

Description automatically generated

1. When we take a closer look at the daily variation in the virus reporting, we can see that there is a slight plateau in the daily increase in virus reporting.

A picture containing screenshot, map

Description automatically generated

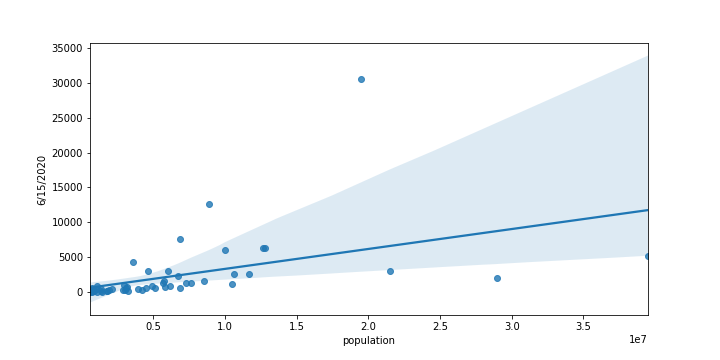
1. Upon examination of which counties are reporting virus cases, it can be noted that the gross number of counties are growing more asymptote adjacent.

A picture containing screenshot, map

Description automatically generated

Question 2: Do states report similar outbreak patterns and mortality rates?

1. When we look at state population, we see that just having a larger number of citizens does not directly indicate that a high number of cases will follow.



1. A close up of a piece of paper

   Description automatically generatedThe outcome when the mortality data is examined by state paints a more shocking picture. Contrary to what might be indicated by news, it is not just New York that has a startlingly high mortality rate. New York actually falls in 3rd after Michigan and Connecticut.
2. A screenshot of a social media post

   Description automatically generatedWhen looking at the upper end of the mortality rate spectrum, we would want to see the significance of the outliers and to do that, we will create a box and whisker plot.

Question 3: What is the geographical importance?

1. When mapped nationwide, the virus looks fairly extensive.

A picture containing text, map

Description automatically generated

1. When we examine the locations of the highest mortality rates, we can see a pattern indicating that the Northeast is home to the top six.

A close up of a map

Description automatically generated

1. However, even though we can visually determine roughly where the case hot-spots are located, it is of significant value to perform an Anova analysis to see if any of the regions are significantly different from the others.

What we learned through this particular method of analyses, was that the p value approaches zero and therefore we can say with confidence that the proportion of cases varies regionally in the United States.

A screenshot of a social media post

Description automatically generated

1. A screenshot of a cell phone

   Description automatically generatedJudging by the above data, we would like to see if location alone plays an important role in the deaths of COVID victims or if there are larger forces at play. When we analyze the latitude of the states however, we can see that it is more than geography.
2. If it isnt location affecting these locations, what can it be?

Question 4: Is access to healthcare affecting the mortality rates?

1. An HPSA score is given to each county (and a large number of clinics) in the United States and is a measure of each location’s need for clinicians. The score for primary care ranges from 1-25 with the higher number indicating that location has a higher need for medical staff.
2. A picture containing white

   Description automatically generatedAs one would expect, counties with a higher HPSA score tend to generally have higher total deaths per capita when compared to their more well-equipped counterparts.
3. However, what does seem counterintuitive is that the HPSA score is not indicative of a higher mortality rate.

A screenshot of a social media post

Description automatically generated

Question 5: Is Age or access to health insurance the driving factor behind COVID morbidity and mortality?

1. The only region to show any correlation between age, health insurance status and infection rates is the Northeast, whereas the South, Midwest and West have little to no relationship

A screenshot of a computer

Description automatically generated

Question 6: Is poverty a more convincing model to determine COVID infection and deaths?

1. Poverty levels seem to be a lager predictor of one’s susceptibility to the virus with strong relationships seen across all 4 regions.

A close up of a map

Description automatically generated

1. This would seem to indicate that rather than age or health insurance status being a predictor, the economic status of the individual is a bigger driver.

Takeaways:

We know more than we did before!

1. Primary finding: there is a statistically significant difference across the USA.
2. Some of that difference can be explained by demographic factors.
3. Contrary to our expectation, access to health care or age do not seem to be significant factors.
4. However, poverty levels appear to affect cases. Whether that is an indication of a larger influencer remains to be seen.
5. There are also clearly other factors at work, and the data is still evolving!
6. This has implications for how we understand and address what has happened.