Covid-19 Response: Socio-Political Analysis

Justin Helfman

6/3/2020

Introduction:

Throughout the beginning of 2020 we as a world have experienced a global pandemic with the outbreak of Covid-19. Throughout this, many governments have received a great amount of criticism from numerous different perspectives of their response.

Going off of this idea, I was curious to know just how governments around the world performed as compared with one another. Expanding on this, I wondered if some political metrics would be able to predict the success of a response, given a country's ranking in the given statistic.

Data:

To create models on the response to Covid-19 for each country, I used coronavirus data on deaths and confirmed cases in the US and the world (source: Johns Hopkins Center for Systems Science and Engineering). Each set provided a daily death or confirmed infection count, ranging from 1/22/2020 to the present day, with the data being updated daily on Github.

In an attempt to normalize the Covid-19 data, so that countries with a larger population don't appear to have worse response than countries with less people, I used population data by country (source: United Nations). I used this data to get an estimated percent of population infected and killed by Covid-19 for analysis.

For political metrics, I chose to test three slightly different statistics, who I thought may have notable impacts on coronavirus response. Firstly, I chose the Human Development Index (source: United Nations), which serves as an all-encompassing representation of countries. The index ranges from 1 (best) to 0 (worst), and I expected countries with better indexes to have better responses, as their government should be more prepared than those with lower indexes.

The second statistic I chose to use was the Press Freedom Index (source: The World Bank). The index is scaled so that the best indexes are the lowest, and the worst indexes are the highest. With this data, I expected the worst Press Freedom Indexed countries to report very successful, yet likely falsified, numbers. For countries who are not at the very bottom of the index, I expect response to improve with the Press Freedom Index, as information may be able to be presented as quickly as possible.

The final metric I used was on the average years of schooling for each country (source: UNESCO). I expected the amount of education one receives to affect how citizens react and respond to the threat of Covid-19.

With some of the statistics I used, I chose to use the most recent reporting numbers. It is worth noting that some numbers are not extremely recent, through because I used statistics within a few years, I believe this to not have any effect on my analysis.

Methods:

To analyze the confirmed cases of Covid-19, I wanted to attempt to fit an equation to the current data and try and extrapolate the total confirmed cases and deaths. I wanted to choose an accurate model, but I was willing to sacrifice the complexity of the model for simplicity. After some research, I chose the model for carrying capacity, as I believed various factors like improved treatment, prevention, and/or herd immunity would eventually slow the spread of Covid-19. It is also worth noting that this model can only be used as a good fit for an individual wave of Covid-19 and cannot be used to estimate possible future waves of Covid-19.

To model carrying capacity, I used a sigmoid function $f(x) = \frac{L}{1+e^{-k(x-x_0)}}$. Where an equilibrium is reached at population L. Because a fixed percentage of the amount of people who get the virus will die, I used the same model to represent deaths, but with different coefficients.

In order to perform a regression, and determine the model's coefficients, I used the scipy.optimize library, primarily the curve_fit function.

To analyze the final results, I chose to take the equilibrium value, L, and divide this by the population of a given country to obtain the cases or deaths per population (as mentioned in the 'Data' section). I then classified each country into an upper, middle, or lower tertile, for each of the three political metrics. Finally, I plotted confirmed cases per population vs the deaths per population, and for each metric, I color coded the points into their appropriate tertile.

Errors:

Before analyzing the results of this project, it is worth noting a few errors that affect the outcome of the project. Firstly, as you may see in the plots, some of the plots have points well outside of the range of 0-1, inferring that the rates of infection or death per population are more than the population itself. I am unsure as to why this error is happening, though I infer that it may have to do with the sigmoidal regression. Admittedly, I am not too familiar with regression outside of linear and polynomial regression, so I relied on a temporary solution, using a post related to sigmoid regression (cited in sources). Though this was a good fit to the present data, it did not seem to be very good for future data, always inferring that infections would flatten out nearly immediately. This method also was inconsistent in fitting models to some countries, forcing them to not be used in plotting. At a point where I am more familiar with this type of regression, I plan to fix this process, and hopefully have a much more correct plot.

It is also worth noting that when running the code, there will be some overflow errors to occur. These originate from trying to raise eulers number to a very small magnitude number (typically much greater than a magnitude of 10^{-10}). As $e^0=1$, I included a try-except statement that simplifies the equation to $\frac{L}{2}$, but this does not have any major effect on the outcome of this project.

Analysis:

To take into account the errors of some failed regressions, I chose to enlarge in various times to show the most populated clusters before performing analysis. To analyze this data, I chose to compare each cluster's closeness to the origin, as well as the overall spread of each tertile's cluster. The origin is intended to be the most desirable location in the plots, as it signifies a very low death rate and infection rate per population.

For the Human Development Index (HDI) figure 1.1 shows a couple of interesting results. Firstly, and most surprisingly, lesser performing countries appear to cluster closest to the origin. Though surprising, there may be a reasonable explanation for why this is. In simple terms, the HDI is an index that broadly represents opportunity within countries. Because of this, it is reasonable to assume that many people may chose to emigrate their country for one with a higher HDI. This would allow not only for migrants to potentially carry or contract disease en route to a high HDI country, but it would also be removing potentially infected people from a lower HDI country.

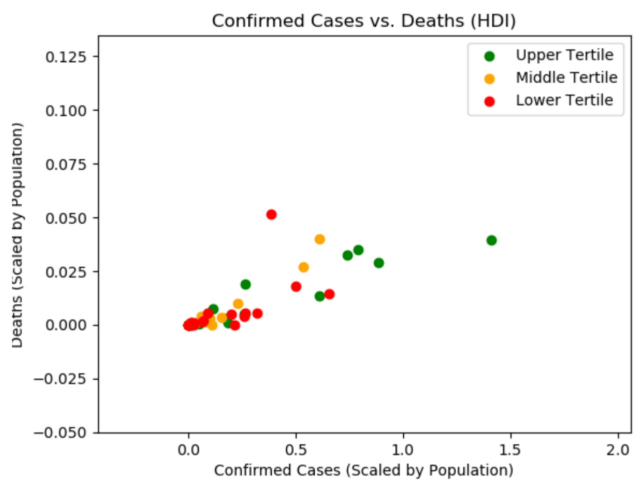

Figure 1.1 – HDI with 2 enlargements

Another interesting thing to note in figure 1.2 is that if you were to roughly make a linear trendline originating from the origin for each of the three clusters, it would appear that the slope of the upper tertile would be less than that of the middle tertile, whose slope would be similar to, if not less than the lowest tertile. In simpler terms, the slope would signify deaths/cases, or the death rate, most likely due to the success of the country's hospital system. This point seems very intuitive and reveals generally that as a country's HDI increases, so does the successes of its hospitals. Given that the HDI takes into account the life expectancy of a country this is very unsurprising, as a high life expectancy is typically a sign of a good healthcare system.

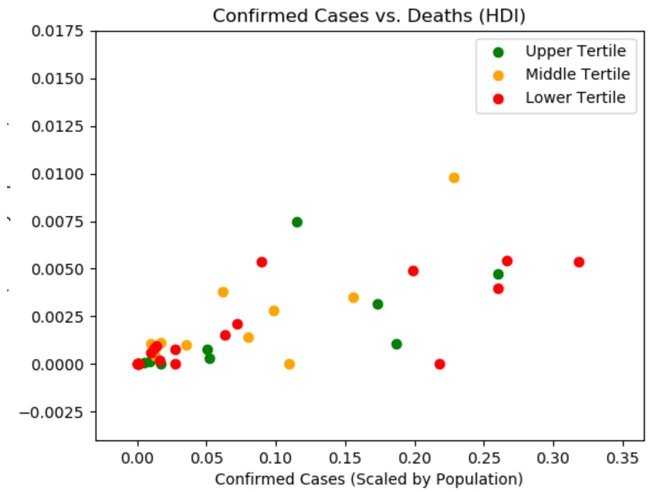

Figure 1.2 – HDI with 3 enlargements

Moving on to the Press Freedom Index (PFI), figures 2.1 and 2.2 confirmed exactly what I hypothesized in the 'Data' section. As expected, with a poorer PFI typically a country would have a higher death and infection rate than that of a country with a better PFI. Though there are many reasons as to why this happens, I theorize that the more freely information is allowed to spread, the quicker a country and its people responded to the threat of Covid-19.

This idea holds true for most data however, as expected, there is a very dense cluster of poor PFI countries very close to the origin, signifying a successful response to Covid-19. Though it is entirely possible that this is the case, I believe it to be more likely that correct information is not being provided by these countries. Being that these countries typically censor their media in attempts to improve how their people, and the world view them. Extrapolating this idea it wouldn't be too unreasonable to assume that these countries may provide incorrect information about their Covid-19 response in an attempt to boost their reputation in the international community.

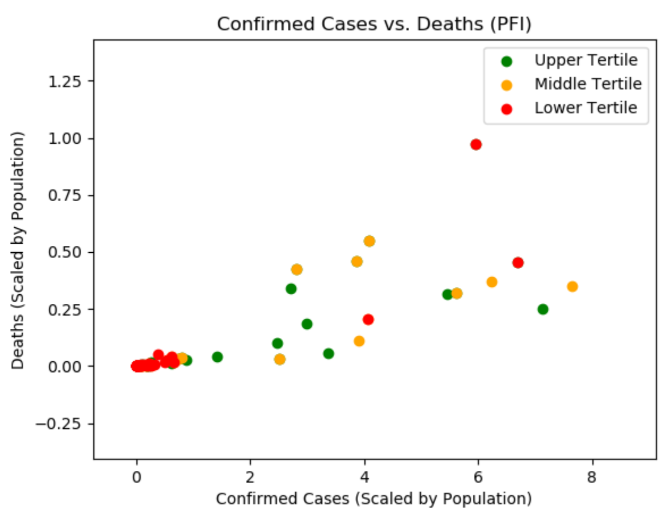

Figure 2.1 – PFI with 2 enlargements

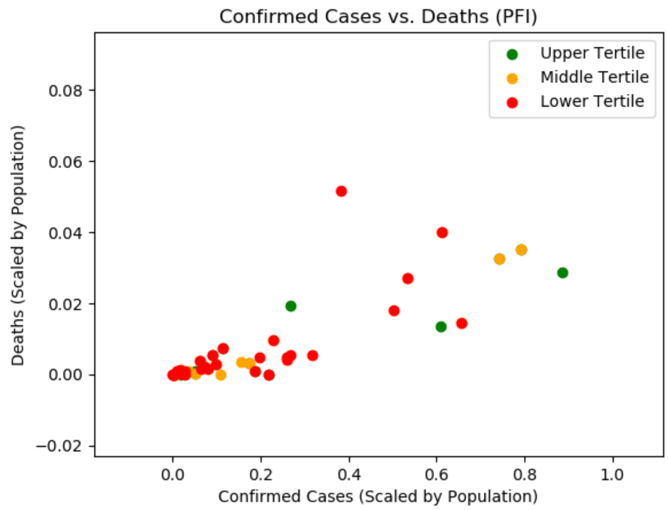

Figure 2.2 – PFI with 3 enlargements

Finally, with the case of the mean years of schooling (EDU) in figures 3.1 and 3.2, there were some fairly interesting results with the data. As expected, countries in the highest tertile were closer to the origin than countries in the middle tertile, which is expected, as a higher level of education would likely allow people to understand and react to the spread of Covid-19 quicker. However, this trend is not the case for countries in the lowest tertile, as they appear to be clustered the closest to the origin.

This was a very unexpected outcome, though I have come up with a possible explanation as to why this happened. In many lower education countries, agriculture may play a much more significant role in both the economy and daily lives of people than in countries with a higher mean year of schooling. This would support why the mean years of schooling would be very low, as farming is typically taught through experience and work instead of schooling. Finally, because rural areas have a much lower population density than that of an urban area, Covid-19 is less likely to spread.

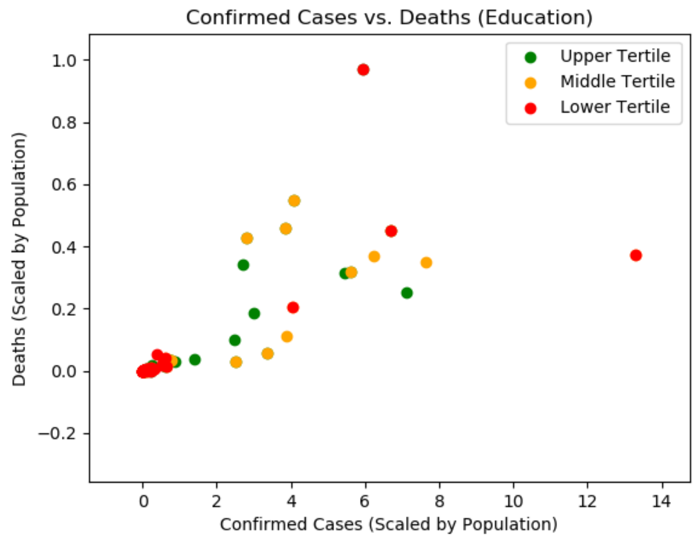

Figure 3.1 – EDU with 2 enlargements

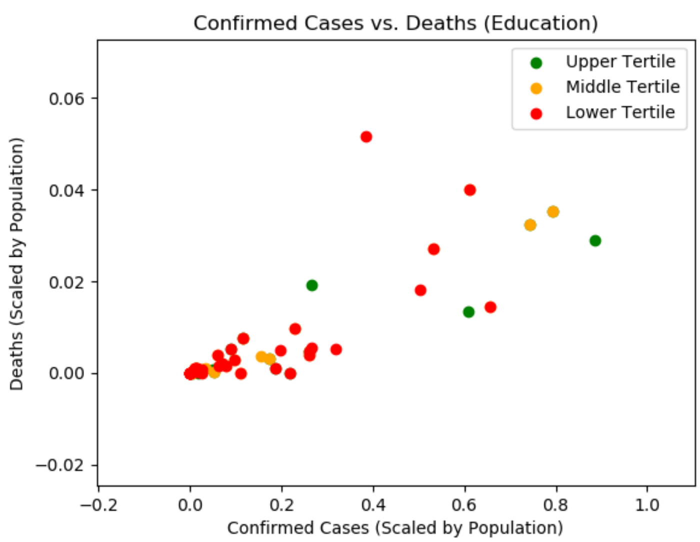

Figure 3.2 – EDU with 3 enlargements

Conclusion:

In conclusion, there were some fairly interesting trends and patterns to be seen in this data. I viewed this project to be a good exercise in attempting to explain Covid-19 response with respect to some socio-political metrics. In the future, I hope to improve my code and model making, as well as to include more metrics to help analyze and explain global responses to Covid-19.

Sources:

- Covid-19 deaths & confirmed infection data: https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series
- World Population data: https://population.un.org/wpp/Download/Standard/CSV/
- Human Development Index: http://hdr.undp.org/en/data#
- Press Freedom Index: https://tcdata360.worldbank.org/indicators/h3f86901f?indicator=32416&viz=line_chart&years=2001,2019
- Mean Years of Schooling: http://data.uis.unesco.org/index.aspx?queryid=242