Connor Saari

Problem Set 5

Open Response Problems

2.

* International Organizations
  + I believe WHO, the World Health Organization will have relatively accurate data because they have been a functioning international health organization for some time now and have dealt with other similar major pandemic. For this reason I expect them to understand some of the disparities in the data, and are making active responses to organize and legitimize their datasets.
* Government Organizations
  + The US CDC is going to have some problems aggregating and cleaning their data, because of the lackluster response from the American Government. Masks, tests, and cases all seem to be areas where the US CDC may fall short because US domestic policy has made it difficult to keep the pandemic at large, realizing the most cases for a country in the world. For this reason I believe that the US CDC will have a difficult time managing their Covid-19 data, but they are an experienced and funded organization allowing them the means and the resources to succeed in aggregating unbiased datasets.
* News Organizations
  + The US Tracking Project may have some interesting data to look at particularly because they offer differing analytical perspectives of US data from hospitals to federal datasets. They have the ability to create a different bias within the data as well as see the data from a different lense in that they can aggregate between both private and public facilities, whereas governmental data will lack some of this data.

11.

* Confirmed
  + The visualization for the confirmed cases displays a gradual exponential curve gaining gradually with a large spike in confirmed cases toward the center. This may be because of varying factors including a jump in available testing kits, more intervention for a particular country, or simply the disease may have mutated and is growing and spreading faster than before.
* Deaths
  + The deaths visualize a steady exponential increase at the beginning and through most of the visualization, but with a slight plateau toward the end. The reason for this may be because the months are slowly becoming warmer possibly increasing recovery times and decreasing spreadability, or because the major spread of the disease has already killed the most vulnerable population, and the more resistant ones are not dying as often.
* Recovery
  + This visualization shows little to no increase at the beginning of the plot, but with a significant increase in cases toward the end. This is probable because there have been more advances in combating the disease as well as regular immunity and recovery of previous patients have been realized making them spike in recovered cases.

13.

* Recovered vs. Deaths
  + Understandably, there is a lot more recovered data because the amount of people who have completely recovered covers all age groups including those most immune to the disease, whereas official deaths are hard to be confirmed in some situations and only represent the population of people that were the most susceptible.
* Confirmed vs. Deaths
  + There is much more data for the confirmed cases over deaths because the death rate compared to the rate of infection is much lower. This makes it possible to see so many more confirmed cases.
* Confirmed vs. Recovered
  + This graph shows the total number of confirmed cases compared to the total number of recovered. As expected the number of cases for recovered slowly but exponentially increased because of more vaccines and knowledge on the virus, and confirmed in a similar trend because of the increased amount of testing availability.

17.

* The top 5 most correlated pairs of countries in order from most correlated to least correlated is Germany/Spain, Czechia/Germany, Czechia/Spain, Morocco/Romania, Belgium/Portugal. I believe these pairs to be the most correlated because of their very similar sizes and corresponding governmental response. Even though Germany’s Prime Minister greatly outlined the risks to Corona, its spread in such a large country was said to be above 70% overall infected, as well Spain’s weak government and economy represented conditions for the virus to spread as easily as possible. This correlation is seen with the rest of the pairs, specifically with the smaller countries whose infrastructure may not be as strong as other countries.
* 18
  + Along with the corresponding correlations that pairs these countries together, the growth rates of each represent increased similarity in each countries response and the effect the virus has had on their populations. The growth rates among Czechia and other similar countries like Spain or Germany show that the growth rate among similar countries for this pandemic represents similar situations and establishments. Countries in Europe, having such a close proximity to one another, are prone to spread virus and pandemic quicker than other countries with physical borders and that more more spread out.