Samuel Andrews

03/18/2020

DSCI 430

Professor Iverson

Project Part 1 Data Curation

My primary goal with choosing indicators for this project was to find variables that could help one understand how the level of industrial development and international economic policy prior to the pandemic affected an individual country’s Covid-19 cases, recoveries, and deaths. In addition, the living conditions of the general populace where also considered through a few variables to see if Covid-19 was greatly impacted by day-to-day sanitation measures used in varying degrees across the world. In order to prevent the actions caused by the emergence of the virus affecting data collection and to avoid changes done in the short term by Nations affecting the input, all of the data was taken from 2018 with the exception of a single variable from 2017. This is because this endeavor aimed to compare policy before to results after and thus wanted to avoid operational changes that occurred as a result of the virus beginning to impact the world in 2019.

Therefore, in addition to obvious development indicators such as population growth, variables tracking air freight and passenger moving through the country in any capacity where chosen to see if countries utilizing alternative travel methods overall performed better. As an extension of such, the cost to import in compliance with border regulations was also looked at, as such can be used a proxy for how much a country encourage international trade with its neighbors and it may be interesting to see if those that promoted interdependence suffered more spreading than others. These three variables also were almost complete as far as data for each country went, which resulted in them being chosen as opposed to other that may have looked at similar ideas.

To look at the industrial ubiquity across the population of each country, it was decided that variables such as the percent access to electricity should be looked at, both due to its obvious implications on the quality of life and its night completeness in being kept track of. Other indicators related to it, such as the number of ATMs per 100,000 adults and the percentage of the population practicing open defecation. The former aims to indicate how modernized the financial system of a country is, with the added fact that ATMs found new usage in a pandemic as a way to interact monetarily with others without physical contact, which may have made an impact of the virus’s spread ability. Meanwhile, the latter aims to act as way to understand how the strength of civic sanitation systems impacted the spread of a virus. The implementation of widespread plumbing is known to have greatly slowed down the ravaging impacts of cholera and dysentery in many parts of the world, and it might prove interesting to see if it helped at all with Covid-19. This variable also took its data from 2017 as in 2018 much of the information was missing.

Finally, the health standards of the population were considered, and a trio of indicators were selected specifically to explore that aspect. Starting off, life expectancy was chosen due to it being a common way for individuals to relate the health of a population in way that is easy to relate for the common individual. It is also very tightly tracked, making widespread comparisons easier. Following this, the number of nurses and midwives per 1,000 people was also tracked, as the commonly discussed problems of nursing shortages during this ongoing pandemic often was related to additional deaths in various media outlets and it would be interested to look at that correlation. Finally, the indicator of total alcohol consumption per capita in liters for individuals over the age of 15 was chosen. It is no secret that a healthier lifestyle among the populace can strengthen the chances of one’s survival in the face of illness, and high alcohol consumption can contribute to negative effects. Compared to variables like smoking and hard drug use, this variable was used as it is much easier to track across various nations and also does not necessarily thought to correlate directly with a country’s economic status.

All of the above variable was combined together with the information on Covid-19 cases, recoveries, and deaths from the provided GitHub page at [https://coviddata.github.io/coviddata/#csvs](https://coviddata.github.io/coviddata/%23csvs). Unfortunately, not every country in the World Bank was tracked, leaving 199 areas left to track. In addition, not every country was separated the same way across both sets, with French departments being one example where the World Bank tracks them as one entity where the pandemic data treats the separately. In most cases, these problems were solved by recoding them to match the World Bank categories and performing aggregation to combine the information. These cases are handled and described in greater detail in handled in the python notebook also found within this repository. In general, the idea was to stay consistent with how such problems were handled, but ultimately concession needed to be made in some areas. Despite this, the dataset should be able to allow us to make the thought-provoking correlations one may seek.