

Final Project: COVID-19 Dataset

Due Friday, November 20, 11:59 PM

The Lads: Frankie Willard, Manny Mokel, Alex Katopodis, Parker Dingman

Introduction

Throughout the year 2020, the COVID-19 pandemic took the world by storm, deeply impacting every country on the planet, albeit with differing degrees of severity. As cases continued to rise, families suffered from the loss of family members, jobs, social interactions, disposable income, and more.

This public health crisis became severe enough such that many countries took decisive action, shutting down their economies to prioritize the lives of citizens. Meanwhile, other countries were less strict in their policies, attempting to preserve their economy at the potential expense of their citizen's lives. The difference in each country's characteristics, demographics, public health capacities, and the strictness of COVID-19 policies led to vastly different effects of the pandemic on different countries. Given our personal connections to the effects of the pandemic through our lives, our friends, and our families, we wanted to determine what led to the pandemic affecting some places worse than others.

We are interested in investigating how a country's demographics impact the domestic severity of COVID-19. More specifically, we would like to see which demographics lead to higher cases per capita and deaths per case. We are also interested in analyzing how effective lockdowns and COVID-19 related policies have been in mitigating the spread of the virus.

We hypothesize that stringency-index, GDP per capita, population density, human development index, and latitude will have a strong impact on cases per capita. We also hypothesize that deaths per case will be largely determined by GDP per capita, the number of citizens aged 65+, hospital beds per thousand, and prevalence of pre-existing conditions (ex. diabetes prevalence, cardiovascular death rate, etc.). Finally, we expect that strict COVID-19 policy has effectively slowed the transmission of the virus.

These hypotheses are based on prior experiences and research. There is evidence that a high stringency index, a composite score based on how strict a country's restrictions are, slows the spread of COVID-19 [1]. Cold climates, which are caused by varying latitudes, have been associated with faster viral spread of COVID-19 [2, NEED A SOURCE!!!!]. We also know that patients with pre-existing conditions face a higher COVID-19 mortality rate [3, NEED A SOURCE!!!!].

Data Description

We selected a data set from "Our World in Data." Each observation in the data set shows relevant COVID-19 data for a particular country on a given date. The COVID-19 data in the data set includes total deaths, total cases, new deaths, new cases, total cases per million, total deaths per million, total tests, new tests, total tests per thousand, positive rate, as well as telling country numbers such as stringency index (composite measure of government strictness policy) and hospital beds per thousand. Additionally, the data set includes country characteristics including population density, median age, GDP per capita, diabetes prevalence, life expectancy, and extreme poverty rate. While the previous variables are quantitative, the data set also includes categorical variables when it comes to geography such as the country and continent.

"Our World In Data" uses data from the European Center for Disease Prevention and Control (ECDC), a world leader for COVID-19 data. The ECDC has a team of epidemiologists that works every day to screen up to 500 sources to get the latest figures. These sources include ministries of health (43%), websites of

public health institutes (9%), websites of public health institutes (6%), World Health Organization (WHO) websites, WHO situation reports (2%), and official dashboards and interactive maps from national and international institutions (10%). The EDEC also utilizes social media accounts maintained by national authorities, ministries of health, and official media outlets (30%). These social media sources are screened and validated by the other sources mentioned previously. The data is recorded daily, and we will be using the data set updated as of October 9, 2020 (10:30, London time).

We used latitude data from the website “Kaggle” to supplement the COVID-19 data set. This data was collected from a Google data set and was merged with the “Our World in Data” COVID-19 data set to create one larger data set.

Sources: <https://ourworldindata.org/coronavirus-source-data> <https://www.ecdc.europa.eu/en/covid-19/data-collection> <https://www.kaggle.com/paultimothymooney/latitude-and-longitude-for-every-country-and-state>

Methodology

Results

Discussion

References

- [1] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7418951/#:~:text=Our%20model%20implies%20that%20social,at%2021%>
- [2] [3]