PPOL564 Data Science I: Foundations

Project Proposal

Zixun Hao zh210

I. Topic summary, Background, Scope of Research

The topic I want to explore in this project is the relations between government's response to Covid-19 by country and the spread of Covid-19 as well as the economic performances in the country.

Covid-19 pandemic in 2020 caused 45.2 million cases of infection and more than 1 million deaths globally, resulting in GDP contraction, high unemployment, and economic recessions around the world. Are there governments who do better than others in protecting the country's public health and/or economy during the pandemic? Are there potential trade-offs between protecting public health and guarding economic growth? Which policies and policy combinations are most effective in controlling the disease and conducive to pushing the economy back to pre-pandemic levels?

The scope of research needs to be narrowed and more clearly defined as the availability of Covid-19 data and economic information vary considerably. One possible way to specify the scope of countries is to only focus on OECD countries.

II. Data sources

1, In terms of "government's response to Covid-19", I will use the "COVID-19 Government Response Event Dataset (CoronaNet)" created by Cheng, C., Barceló, J., Hartnett, A.S. et al. that was published on Nature Human Behavior in June 2020. They classify the governments' policy actions in response to Covid-19 into 19 types, including closure and regulation of schools, declaration of emergency, lockdown, restriction of non-essential businesses, external border restrictions, internal border restrictions, anti-disinformation measures, etc. This dataset was generated from more than 200,000 news articles around Covid-19 and it covers policy actions taken in 201 countries and territories, recording start and end dates of the policies. I plan to continue using the policy classifications in the CoronaNet dataset with minor changes, for example, merging similar policy types into one, and creating parent groups to combine policy types.

Source: (Downloadable) Cheng, C., Barcelo, J., Hartnett, A.S., Kubinec, R., & Messerschmidt, L. (2020). *CoronaNet—Tracking governments COVID-19 responses*. https://www.coronanet-project.org

2, We measure the spread and public health outcome of Covid-19 in a country by the aggregated number of the country's Covid-19 cases and deaths per one million people by October 31, 2020, and the daily (or weekly, depending on data availability of countries) reported new cases and deaths per million people, from January through October. The data should also be scaled by the country's population. The data is mainly scraped from Johns Hopkins Coronavirus Resource Center and Johns Hopkins University Center for Systems Science and Engineering. Population data is from United Nations.

Sources: (Web scraping) Johns Hopkins Coronavirus Resource Center, "Morality Analysis: Cases and mortality by country". https://coronavirus.jhu.edu/data/mortality

(Downloadable) COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://github.com/CSSEGISandData/COVID-19

3, We measure economic performances by each country's percent change of GDP in Q2 2020 from Q2 2019, percent change of GDP in Q2 2020 from Q1 2020, and percent change of unemployment rate in Q2 2020 from Q2 2019. Many countries' economic information will not be available until the annual report of GDP and employment that is released at the end of the fiscal year 2020; Eurostat and OECD have Q2 2020 quarterly GDP and unemployment data of their member countries available.

Sources: (Downloadable) Eurostat, "Quarterly national accounts - GDP and employment", October 19 2020, https://ec.europa.eu/eurostat/statistics-explained/

(Downloadable) OECD, "Quarterly GDP", https://data.oecd.org/gdp/quarterly-gdp.htm. "Unemployment Rate", https://data.oecd.org/unemp/unemployment-rate.htm#indicator-chart

(Web scraping) Joe Hasell, "Which countries have protected both health and the economy in the pandemic?", September 1, 2020, https://ourworldindata.org/covid-health-economy

4, Include the pandemic preparedness by country, measured by global health security (GHS) index.

Source: (Web scraping) GHS Index, 2019 Global Health Security Index, https://www.ghsindex.org/

III. Methods to use

1, Data wrangling: Group the CoronaNet government Covid-19 response dataset by countries to see what actions each government took at what time; incorporate time-series, consider the start dates and duration of certain policies and the public health outcomes. Combine the policy reaction data with the health and economic data of each country to analyze the effects of certain policy combinations; understand the features of effective policies: do they tend to be national or provincial, mandatory or voluntary, for a longer or shorter period of time? Also, incorporate the GHS index into the analysis, as countries don't have the same level of capacities to effectively respond to health emergencies.

2, Data visualization: Use ggplot2 and matplotlib to create plots for the cumulative occurrences of certain policies globally over time. Plot the relation between the duration of economic shutdown ("restriction of non-essential businesses" type) and Covid-19 cases across countries. Use Folium package to visualize cumulative Covid-19 cases and deaths, unemployment rate and GDP on maps.

3, Machine learning with Scikit-Learn: Use K-nearest neighbors to make prediction of a country's class based on training patterns of other countries' Covid-19 policies; Use decision tree algorithm to model decisions of Covid-19 policies taken and the consequences.

IV. Goals and How to gauge success

The goal is to evaluate country's Covid-19 policies and outcomes, find the policy combinations that are most effective in controlling the disease and protecting the economy. A successful project should have solid modeling and analysis, produce meaningful insights, and detailed, nuanced findings.

References

Cheng, C., Barceló, J., Hartnett, A.S. et al. COVID-19 Government Response Event Dataset (CoronaNet v.1.0). Nature Human Behavior 4, 756–768 (2020). https://doi.org/10.1038/s41562-020-0909-7

Pettersson, H., Manley, B. and Hernandez, S. "Tracking coronavirus' global spread," CNN (Updated October 30, 2020). https://www.cnn.com/interactive/2020/health/coronavirus-maps-and-cases/

Hasell, J. "Which countries have protected both health and the economy in the pandemic?", *Our World in Data,* September 1, 2020, https://ourworldindata.org/covid-health-economy