Impact of Novel Coronavirus on Public Policy (Research)

10/22/2020

Team Members

Christopher Hopkins 412

Santiago Nule 412

Jose T.Muci Lander 612

Jaehee Lee 412

Overview

On december 2019, the first case of the novel coronavirus was identified in Wuhan, China. In early 2020s, the virus would eventually find its way out of China. As of today, 40 million people have contracted the disease and over 1.1 million fatalities have been recorded worldwide.

Although the virus has had a vast global influence, there are still noticeable differences in the impacts produced by the virus due to geographical, sociological and political disparities world-wide.

This analytical research focuses on why different areas around the world have been affected differently from covid. Looking at the data for covid-19 from the last 8 months.

Data source and Status of Data collection and tidying

We found a data package that is live updated from a variety of sources including the Oxford COVID-19 Government Response Tracker, and has intergration with the World Bank Open Data, Google Mobility Reports, and Apple Mobility Reports.

Status of the literature review and the initial sources

Main source of data: https://covid19datahub.io/index.html

Codebook for the school_closing, workplace_closing, etc: https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/codebook.md

How to interpret the codebook: https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/interpretation_guide.md

Another data repository: https://github.com/OxCGRT/covid-policy-scratchpad/tree/master/risk_of_openness_index

Our initial sources for data are backed by detailed github repositories that explain the organization and meaning of the data. This has proven useful in narrowing down what questions we can answer using the data. What is especially useful is a codebook explaining the ordinal data in the data set.

initial Questions of Interest

What is the lag/delay between changes in data and policy.

It is speculated that during summer vacations, tourism would help propagate the virus. Looking at countries with loose travel and gathering restrictions we can see if the total amount of cases increased during that time.

Some people said that the virus could have a more catastrophic effect in the winter. Looking at the countries in the southern hemisphere (lower latitude) that experienced winter during the months of may-august look compared with the countries that were during summer at that time.

Places with a higher population density should be expected to have more cases. Create a variable called population density. See if the higher population density relates to more cases.

Is there a correlation between countries that spend more money in a vaccine H5 and the amounts of deaths that they have had from covid. Plot. Likewise, as cases have increased, have countries spent more money on health care investments?

Have stay at home requirements shown to have a decrease in the amount of cases. By looking at the percentage change in recorded cases from before the restriction and after.Then, plotting the amounts of months with lockdown requirements as a categorical variable. And create a variable for continents. Then, color by continent.

```
library(COVID19)
```

```
## Warning: package 'COVID19' was built under R version 4.0.3
```

To access the data a search function is used. A country ID must be used to pull that specific countries information, but a woldwide data set is avaliable as well.

```
covid19()
```

```
## We have invested a lot of time and effort in creating COVID-19 Data Hub, please cite the following wh
##
##   Guidotti, E., Ardia, D., (2020), "COVID-19 Data Hub", Journal of Open
##   Source Software 5(51):2376, doi: 10.21105/joss.02376.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {COVID-19 Data Hub},
##     year = {2020},
##     doi = {10.21105/joss.02376},
##     author = {Emanuele Guidotti and David Ardia},
##     journal = {Journal of Open Source Software},
##     volume = {5},
##     number = {51},
##     pages = {2376},
##   }
##
## To retrieve citation and metadata of the data sources see ?covid19cite. To hide this message use 'ver
```

```
## # A tibble: 54,193 x 35
## # Groups:   id [197]
```

```
##    id    date       tests confirmed recovered deaths  hosp  vent   icu
##    <chr> <date>     <int>     <int>     <int>  <int> <int> <int> <int>
##  1 AFG   2020-01-22    NA        NA        NA     NA    NA    NA    NA
##  2 AFG   2020-01-23    NA        NA        NA     NA    NA    NA    NA
##  3 AFG   2020-01-24    NA        NA        NA     NA    NA    NA    NA
##  4 AFG   2020-01-25    NA        NA        NA     NA    NA    NA    NA
##  5 AFG   2020-01-26    NA        NA        NA     NA    NA    NA    NA
##  6 AFG   2020-01-27    NA        NA        NA     NA    NA    NA    NA
##  7 AFG   2020-01-28    NA        NA        NA     NA    NA    NA    NA
##  8 AFG   2020-01-29    NA        NA        NA     NA    NA    NA    NA
##  9 AFG   2020-01-30    NA        NA        NA     NA    NA    NA    NA
## 10 AFG   2020-01-31    NA        NA        NA     NA    NA    NA    NA
## # ... with 54,183 more rows, and 26 more variables: population <int>,
## #   school_closing <int>, workplace_closing <int>, cancel_events <int>,
## #   gatherings_restrictions <int>, transport_closing <int>,
## #   stay_home_restrictions <int>, internal_movement_restrictions <int>,
## #   international_movement_restrictions <int>, information_campaigns <int>,
## #   testing_policy <int>, contact_tracing <int>, stringency_index <dbl>,
## #   iso_alpha_3 <chr>, iso_alpha_2 <chr>, iso_numeric <int>, currency <chr>,
## #   administrative_area_level <int>, administrative_area_level_1 <chr>,
## #   administrative_area_level_2 <lgl>, administrative_area_level_3 <lgl>,
## #   latitude <dbl>, longitude <dbl>, key <lgl>, key_apple_mobility <chr>,
## #   key_google_mobility <chr>
```

```
covid19("US")
```

```
## # A tibble: 275 x 35
## # Groups:   id [1]
##    id    date       tests confirmed recovered deaths  hosp  vent   icu
##    <chr> <date>     <int>     <int>     <int>  <int> <int> <int> <int>
##  1 USA   2020-01-22    NA         1        NA     NA    NA    NA    NA
##  2 USA   2020-01-23    NA         1        NA     NA    NA    NA    NA
##  3 USA   2020-01-24    NA         2        NA     NA    NA    NA    NA
##  4 USA   2020-01-25    NA         2        NA     NA    NA    NA    NA
##  5 USA   2020-01-26    NA         5        NA     NA    NA    NA    NA
##  6 USA   2020-01-27    NA         5        NA     NA    NA    NA    NA
##  7 USA   2020-01-28    NA         5        NA     NA    NA    NA    NA
##  8 USA   2020-01-29    NA         6        NA     NA    NA    NA    NA
##  9 USA   2020-01-30    NA         6        NA     NA    NA    NA    NA
## 10 USA   2020-01-31    NA         8        NA     NA    NA    NA    NA
## # ... with 265 more rows, and 26 more variables: population <int>,
## #   school_closing <int>, workplace_closing <int>, cancel_events <int>,
## #   gatherings_restrictions <int>, transport_closing <int>,
## #   stay_home_restrictions <int>, internal_movement_restrictions <int>,
## #   international_movement_restrictions <int>, information_campaigns <int>,
## #   testing_policy <int>, contact_tracing <int>, stringency_index <dbl>,
## #   iso_alpha_3 <chr>, iso_alpha_2 <chr>, iso_numeric <int>, currency <chr>,
## #   administrative_area_level <int>, administrative_area_level_1 <chr>,
## #   administrative_area_level_2 <lgl>, administrative_area_level_3 <lgl>,
## #   latitude <dbl>, longitude <dbl>, key <lgl>, key_apple_mobility <chr>,
## #   key_google_mobility <chr>
```

Key information about the data set is as follows:

```
#str(covid19())
nrow(covid19())
```

```
## [1] 54193
```

```
colnames(covid19())
```

```
##  [1] "id"                                "date"
##  [3] "tests"                             "confirmed"
##  [5] "recovered"                         "deaths"
##  [7] "hosp"                              "vent"
##  [9] "icu"                               "population"
## [11] "school_closing"                    "workplace_closing"
## [13] "cancel_events"                     "gatherings_restrictions"
## [15] "transport_closing"                 "stay_home_restrictions"
## [17] "internal_movement_restrictions"    "international_movement_restrictions"
## [19] "information_campaigns"             "testing_policy"
## [21] "contact_tracing"                   "stringency_index"
## [23] "iso_alpha_3"                       "iso_alpha_2"
## [25] "iso_numeric"                       "currency"
## [27] "administrative_area_level"         "administrative_area_level_1"
## [29] "administrative_area_level_2"       "administrative_area_level_3"
## [31] "latitude"                          "longitude"
## [33] "key"                               "key_apple_mobility"
## [35] "key_google_mobility"
```

"id" is the country identifier, for example the United States of America is USA. The columns school closing, workplace closing, cancel events, gathering restrictions, transport closing, stay home restrictions, internal movement restrictions, international movement restrictions, information campaigns, testing policy, contact tracing all come from the Oxford Covid-19 Government Response Tracker(OxCGRT). Each column is a ordinal variable with a number that identifies the level of response. What each number represents can be found in the OxCGRT codebook (https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/codebook.md)

The stringency index is calculated based on governments response to the Covid-19 pandamec. The indices for the stringency index are: school closing, workplace closing, cancel events, gathering restrictions, transport closing, stay home restrictions, internal movement restrictions, international movement restrictions, information campaigns. A detailed account of how the stringency index is caculated can be found here: https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/index_methodology.md

Administrative area level refers to the level of data, that is, level 1 corresponds to country wide data, level 2 to state data, and level 3 to lower level or city data. The data becomes less accurate and complete as the level increases, with level 1 data, country level data, being nearly complete and highly accurate and level 3 data being inaccurate and often incomplete.

There is the option to integrate in World Bank Open Data, Google Mobility Reports, and Apple Mobility Reports.

```
#wb <- c("gdp" = "NY.GDP.MKTP.CD", "hosp_beds" = "SH.MED.BEDS.ZS")
#gmr <- "https://www.gstatic.com/covid19/mobility/Global_Mobility_Report.csv"
#amr <- "https://covid19-static.cdn-apple.com/covid19-mobility-data/"
#covid19("USA", wb = wb)
#covid19("USA", gmr = gmr)
#covid19("USA", amr = amr)
```

These additional data sets have not been reviewed yet, but the Google Mobility Reports and Apple Mobility Reports may be used to provide insight into population movements during the Covid-19 pandemic. A detail break down of the Google Mobility Reports can be found here: https://support.google.com/covid19-mobility/answer/9824897?hl=en&ref_topic=9822927. Currently the Apple Mobility Reports does not properly load and is not being used.

Credit for the Covid-19 data set belongs to:

```
#covid19cite(covid19())
```

The OxGRT database has additional data points not found in the covid19 dataframe. Some of our questions require the use of these data points. Fortunatly this data set is already compatible with the covid19 dataframe.

```
covid19_health_indicators<- read.csv(file = "OxCGRT_latest.csv")
#covid19_health_indicators
```

Planned next steps and schedule

- Continued meetings over zoom.
- Tidying the data and getting rid of variables that won't be used.
- Begin to visualize lag between policy changes and changes in covid rates
- Review of literature pertaining to Public Policy in response to Covid.
- Begin creating the final .Rmd file to submit

Planned roles for team members

Christopher Hopkins = Lead Coder

Santiago Nule = Analyst

Jose T. Muci Lander = Analyst

Jaehee Lee = Statistician

Any concerns or major challenges

Time zone differences (US East Coast, Spain, South Korea) makes it hard to find times for all group members to meet.

Accuracy of data reporting will be an issue, not all entities in the dataset will report complete or accurate data. We must take this into account under the assumptions.