

Explanatory Data Analysis

COVID-19 has prompted unprecedented government action around the world. The Oxford COVID-19 Government Response Tracker (OxCGRT), a data set on government policy measures has been used in this study. Data covers daily and weekly new cases, new deaths and different Government measures taken like closure and containment, health and economic policy for 170 Country to fight the pandemic over the period of 3rd January, 2020 to 31st January 2021. Policy responses are recorded on ordinal or continuous scales for 15 policy areas, capturing variation in degree of response, while the daily new cases and deaths are measured on numerical scale. Details of the variables used in the dataset are shown in the following table.

Working data comprises of 24 variables.

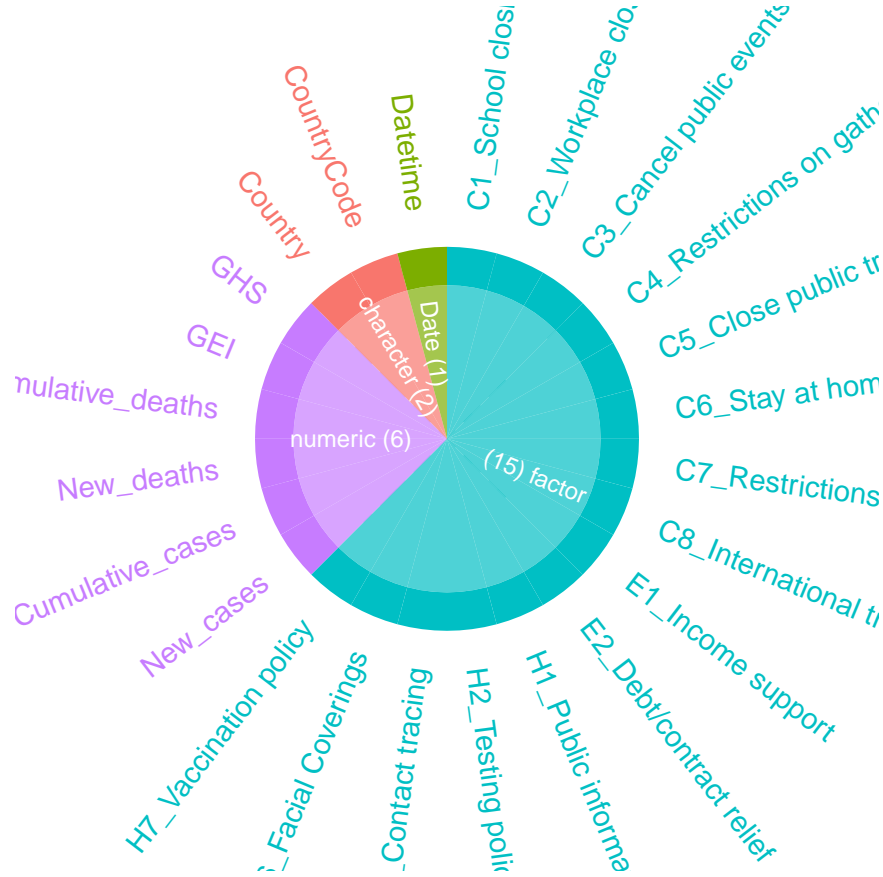


Figure 1: Plot showing categories of different variables in the data.

Here in the data there are 4 types of data. The details of the data are included in the following table.

From the above figure we see that the data contains Missing values and for only complete cases of the data were used. Also for a correct interpretation of the data detecting and eliminating outliers is crucial. Here boxplot was used to detect outleis in numerical variables of the data.

Boxplot showing outliers in the data

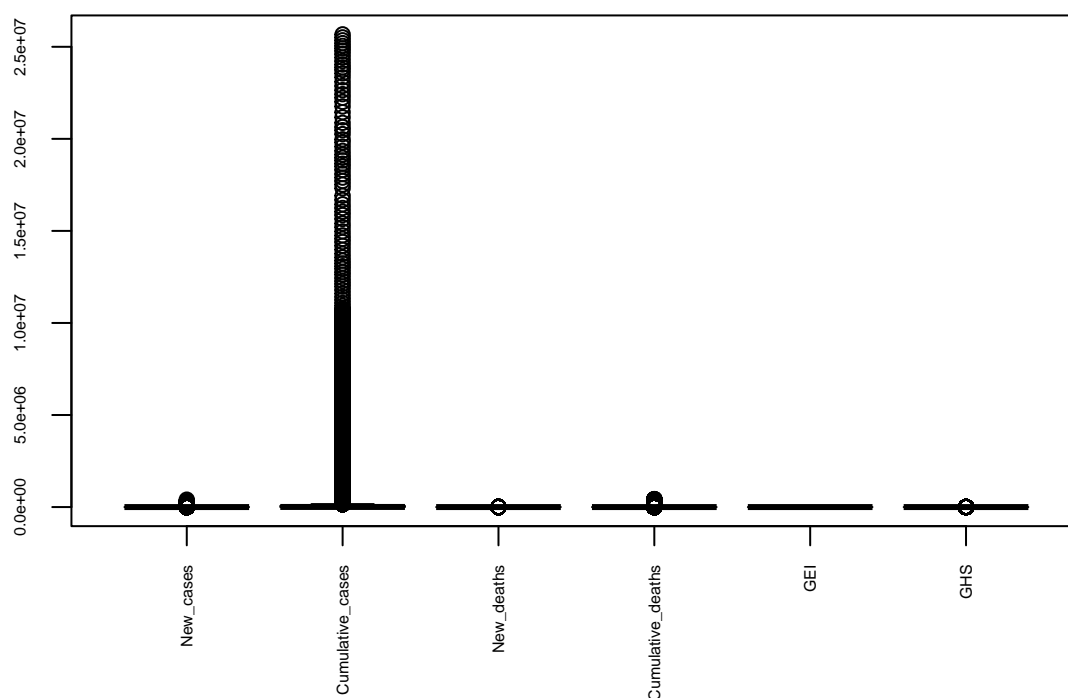


Figure 2: Boxplot showing outliers in the numerical variables in the data.

```
#removing outlier from data.

detect_outlier <- function(x){
  # calculate first quantile
  Quantile1 <- quantile(x, probs=.25)
  # calculate third quantile
  Quantile3 <- quantile(x, probs=.75)
  # calculate inter quartile range
  IQR = Quantile3-Quantile1
  # return true or false
  x > Quantile3 + (IQR*1.5) | x < Quantile1 - (IQR*1.5)
}

# create remove outlier function
remove_outlier <- function(dataframe,
                             columns=names(dataframe)) {
  # for loop to traverse in columns vector
  for (col in columns) {
    # remove observation if it satisfies outlier function
    dataframe <- dataframe[!detect_outlier(dataframe[[col]]), ]
  }
}
```

```

# return dataframe
print("Remove outliers")
print(dataframe)
}

remove_outlier(dat, c(4:7,23,24))

## [1] "Remove outliers"
## # A tibble: 29,591 x 24
##   Country      Count~1 Datetime   New_c~2 Cumul~3 New_d~4 Cumul~5 C1_Sc~6 C2_Wo~7
##   <chr>      <chr>    <date>     <dbl>    <dbl>    <dbl>    <dbl> <fct>    <fct>
## 1 Afghanist~ AFG      2020-02-24      5      5      0      0 0      0
## 2 Afghanist~ AFG      2020-02-25      0      5      0      0 0      0
## 3 Afghanist~ AFG      2020-02-26      0      5      0      0 0      0
## 4 Afghanist~ AFG      2020-02-27      0      5      0      0 0      0
## 5 Afghanist~ AFG      2020-02-28      0      5      0      0 0      0
## 6 Afghanist~ AFG      2020-02-29      0      5      0      0 0      0
## 7 Afghanist~ AFG      2020-03-01      0      5      0      0 0      0
## 8 Afghanist~ AFG      2020-03-02      0      5      0      0 0      0
## 9 Afghanist~ AFG      2020-03-03      0      5      0      0 0      0
## 10 Afghanist~ AFG      2020-03-04      0      5      0      0 0      0
## # ... with 29,581 more rows, 15 more variables:
## #   'C3_Cancel public events' <fct>, 'C4_Restrictions on gatherings' <fct>,
## #   'C5_Close public transport' <fct>, 'C6_Stay at home requirements' <fct>,
## #   'C7_Restrictions on internal movement' <fct>,
## #   'C8_International travel controls' <fct>, 'E1_Income support' <fct>,
## #   'E2_Debt/contract relief' <fct>, 'H1_Public information campaigns' <fct>,
## #   'H2_Testing policy' <fct>, 'H3_Contact tracing' <fct>, ...

```

Table1: Table containing details of the variables.

ID	Name	Type	Coding
Country	Name of the country	Character	
Country Code	Small abbreviation used for country names	Character	
Datetime	Daily Date	Date	
New_cases	Daily new case	Numeric	
New_deaths	Daily new deaths	Numeric	
Containment and closure			
C1	School closing	Ordinal	0 - no measures 1 - recommend closing or all schools open with alterations resulting in significant differences compared to non-Covid-19 operations

ID	Name	Type	Coding
C2	Workplace closing	Ordinal	<p>2 - require closing (only some levels or categories, eg just high school, or just public schools)</p> <p>3 - require closing all levels</p> <p>Blank - no data</p> <p>0 - no measures</p> <p>1 - recommend closing (or recommend work from home) or all businesses open with alterations resulting in significant differences compared to non-Covid-19 operation</p> <p>2 - require closing (or work from home) for some sectors or categories of workers</p> <p>3 - require closing (or work from home) for all-but-essential workplaces (eg grocery stores, doctors)</p>
C3	Cancel public events	Ordinal	<p>Blank - no data</p> <p>0 - no measures</p> <p>1 - recommend cancelling</p> <p>2 - require cancelling</p>
C4	Restrictions on gathering size	Ordinal	<p>Blank - no data</p> <p>0 - no restrictions</p> <p>1-restrictions on very large gatherings(the limit is above 1000 people)</p> <p>2-restrictions on gatherings between 101-1000 people</p> <p>3-restrictions on gatherings between 11-100 people</p> <p>4-restrictions on gatherings of 10 people or less</p>

ID	Name	Type	Coding
C5	Close public transport	Ordinal	Blank-no data 0 - no measures 1-recommend closing(or significantly reduce volume/route/means of transport available) 2-require closing (or prohibit most citizens from using it)
C6	Stay-at-home requirements	Ordinal	Blank - no data 0 - no measures 1-recommend not leaving house 2-require not leaving house with exceptions for daily exercise, grocery shopping, and 'essential' trips 3-require not leaving house with minimal exceptions (eg allowed to leave once a week, or only one person can leave at a time, etc)
C7	Restrictions on internal movement	Ordinal	Blank - no data 0 - no measures 1-recommend not to travel between regions/cities 2-internal movement restrictions in place
C8	Restrictions on international travel	Ordinal	Blank - no data 0 - no restrictions 1-screening arrivals 2 - quarantine arrivals from some or all regions 3 - ban arrivals from some regions 4 - ban on all regions or total border closure Blank - no data

ID	Name	Type	Coding
Economic response			
E1	Income support	Ordinal	0 - no income support 1-government is replacing less than 50% of lost salary (or if a flat sum, it is less than 50% median salary) 2-government is replacing 50% or more of lost salary (or if a flat sum, it is greater than 50% median salary) Blank - no data
E2	Debt/contract relief for households	Ordinal	0-no debt/contract relief 1-narrow relief, specific to one kind of contract 2-broad debt/contract relief
Health systems			
H1	Public information campaign	Ordinal	0-no Covid-19 public information campaign 1-public officials urging caution about Covid-19 2- coordinated public information campaign (eg across traditional and social media) Blank - no data
H2	Testing policy	Ordinal	0 - no testing policy 1 - only those who both (a) have symptoms AND (b) meet specific criteria (eg key workers, admitted to hospital, came into contact with a known case, returned from overseas) 2 - testing of anyone showing Covid-19 symptoms

ID	Name	Type	Coding
H3	Contact tracing	Ordinal	3 - open public testing (eg “drive through” testing available to asymptomatic people) Blank - no data 0 - no contact tracing 1-limited contact tracing; not done for all cases 2-comprehensive contact tracing;done for all identified cases
H6	Facial coverings	Ordinal	0 - No policy 1 - Recommended 2 - Required in some specified shared/public spaces outside the home with other people present, or some situations when social distancing not possible 3 - Required in all shared/public spaces outside the home with other people present or all situations when social distancing not possible 4 - Required outside the home at all times regardless of location or presence of other people
H7	Vaccination policy	Ordinal	0-No availability 1-Availability for ONE of following: key workers/ clinically vulnerable groups (non elderly) / elderly groups 2-Availability for TWO of following: key workers/ clinically vulnerable groups (non elderly) / elderly groups

ID	Name	Type	Coding
			3 - Availability for ALL of following: key workers/ clinically vulnerable groups (non elderly) / elderly groups
			4 - Availability for all three plus partial additional availability (select broad groups/ages)
			5 - Universal availability

After data cleaning several visualization techniques were used to explore few features of the data.

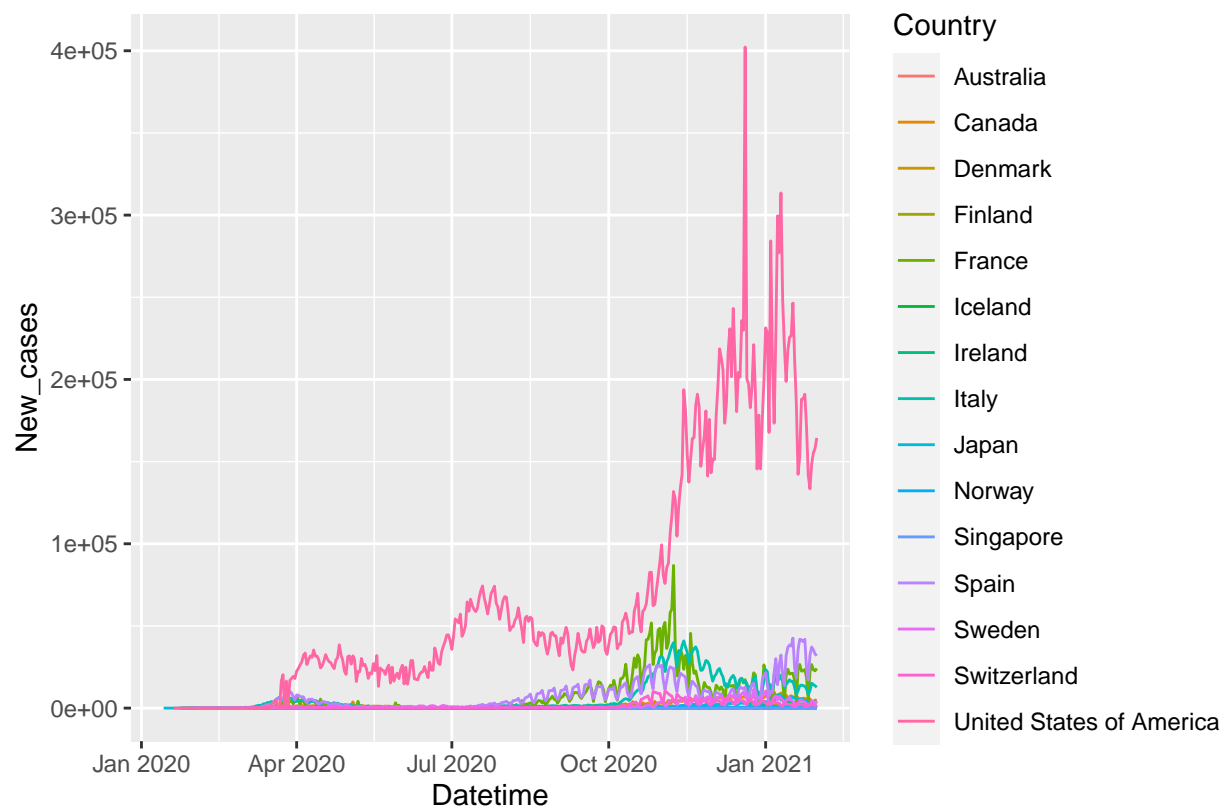


Figure 3:Plot showing Daily New cases for developed Country over study period

Among selected developed Country USA has attained highest number of daily cases and the growth sparked after October 2020 and again came under control after mid January. Also Italy,France and Spain have higher daily cases compared to other countries.

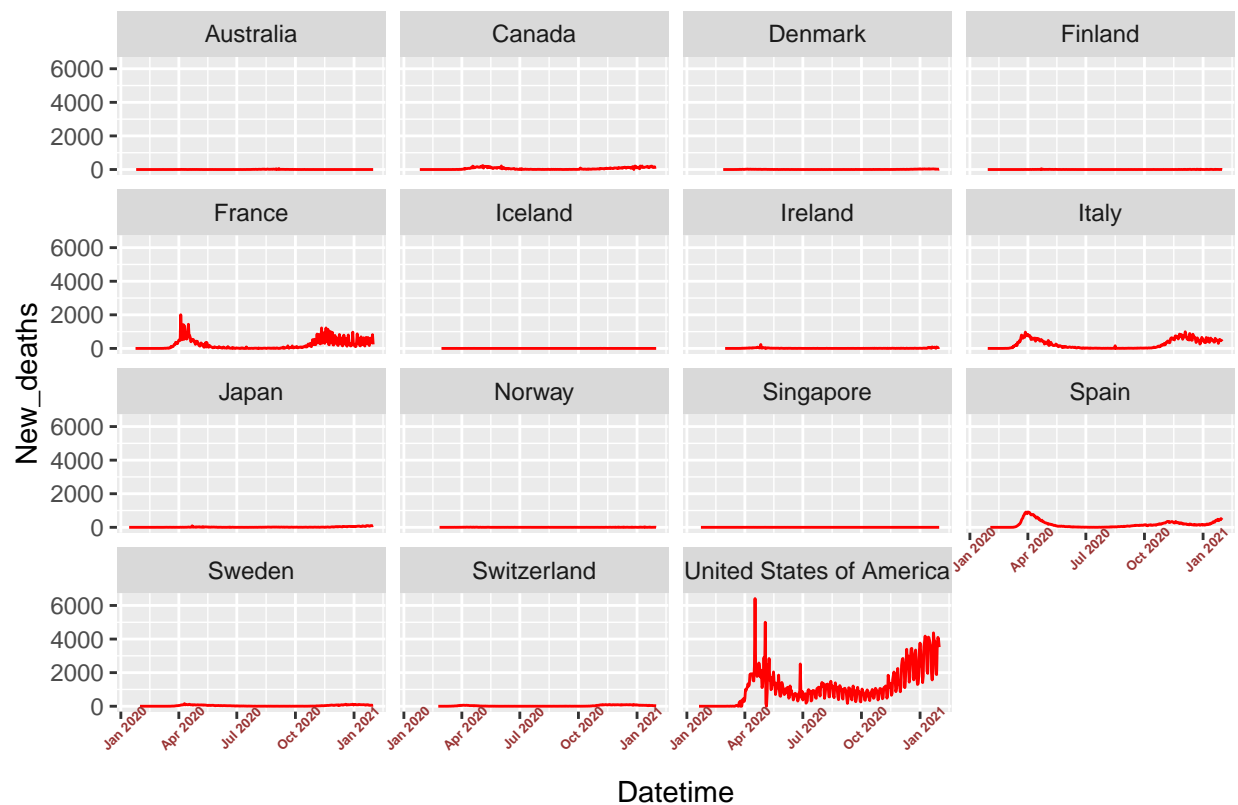


Figure 4:Plot showing Daily New deaths for developed Country over study period

USA has maximum daily number of deaths while Italy, France and Japan were also ahead compared to other countries.

For USA lets check how Government measures affected the daily cases.

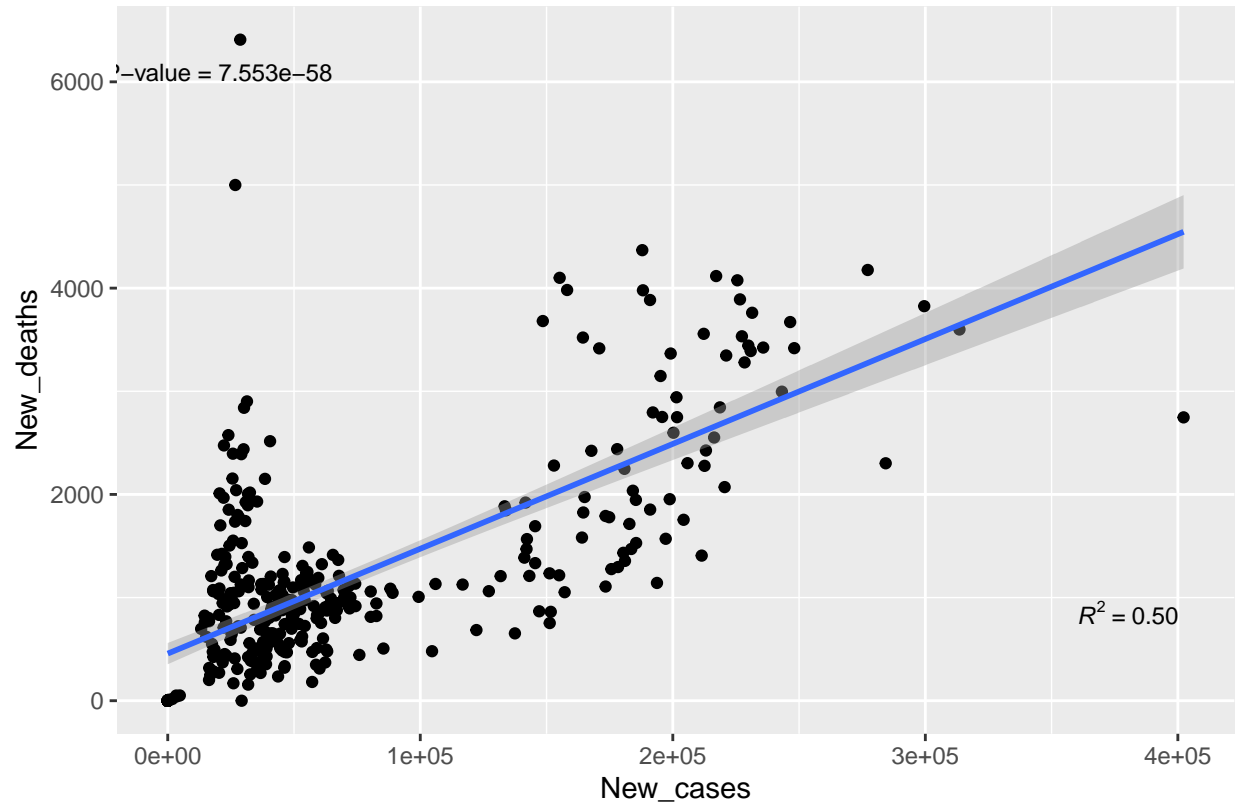


Figure 5: Plot showing dependence between Daily New deaths and daily new cases for USA

From Figure 3 it is clear that there is a significant relationship between New daily deaths and daily new cases in the United States of America with $p\text{-value} = 7.55e^{-58}$ and R^2 value of 0.50.



Figure 6:Plot showing different Government policies in USA

From the figure it is clear that in USA gatherings were restricted to maximum 10 people over the whole study period.

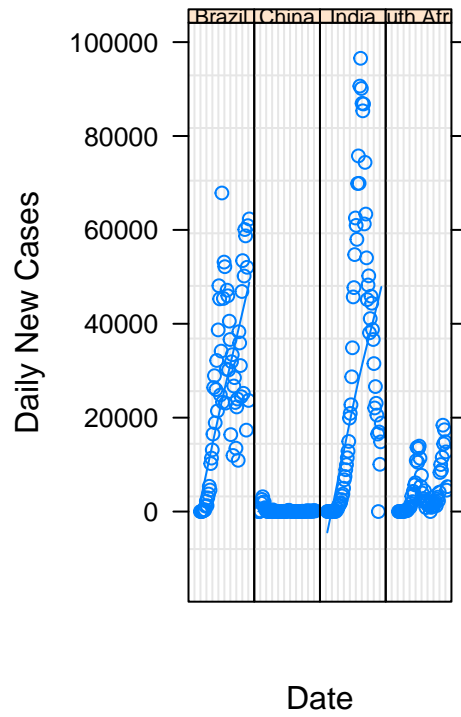


Figure 7: Plot showing number of daily new cases in countries Japan, India, China, Brazil, south Africa, Russia

Among the chosen emerging countries India has reported highest number of daily cases among all selected countries.