

# COVID-19 in the World

Nathan Gaynor, Kayleigh Habib, Zoya King, George Mandl

12/12/2020

- Introduction to COVID-19
  - How Countries Responded
  - Factors Analysed
  - Dates Selected and Data Sets
- Political
  - Data Set Analysis
  - Approach Used
  - Analysis
  - Findings
- Economic
  - Data Set Analysis
  - Single Variable Analysis
  - Two Variable Analysis
  - Findings
- Employment
  - Data Set Analysis
  - Approach Used
  - Findings
- Technology
  - Data Set Analysis
  - Comparisons and Relationships
  - Findings
- Combining Factors
- Conclusion and Recommendations
- Appendix
  - Data Sources

**Team Name : TrendSetters**

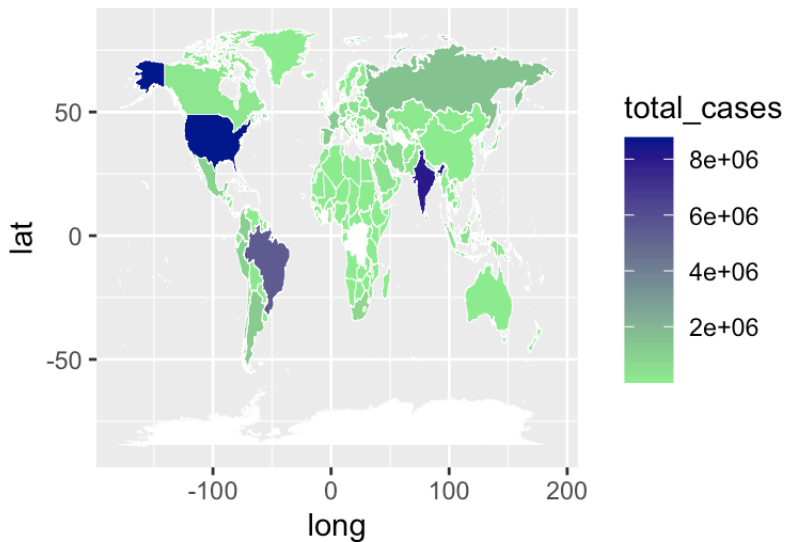
Last Name (Family)	First Name (Given)	Student ID
Gaynor	Nathan	170586250
<b>Habib</b>	Kayleigh	200370580
King	Zoya	203240570
Mandl	George	200655330

# Introduction to COVID-19

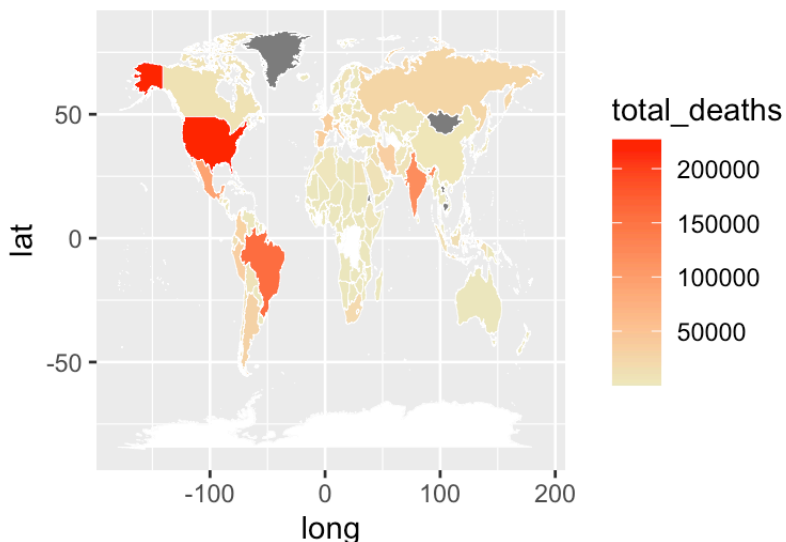
In late 2019, reports of an infectious respiratory illness began to emerge from China. While its origin is unclear, and is outside the scope of this analysis, the lack of information and methods of treatment raised alarms in many other parts of the world. Caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the disease has become widely referred to as COVID-19.

On March 11, 2020, the World Health Organization (WHO) officially declared COVID-19 to be a pandemic, based on its definition of “worldwide spread of a new disease”. At that time, there were just 118,000 cases worldwide, with 90% of the cases concentrated in 4 countries. Over the past 8 months, COVID19 has infected over 53 million people, and caused over 1.3 million deaths in over 220 countries worldwide.

Total COVID-19 Cases - Oct 28, 2020



Total COVID-19 Deaths - Oct 28, 2020



The maps show that while this is a global health crisis, the levels of infection as well as the fatality rates vary widely by region, country, and even within countries. For example, there is a significant difference in experience among provinces in Canada, based on demographics and level of urbanization.

# How Countries Responded

In response to the pandemic declaration, many governments around the world began to implement various measures to control the spread of the disease, to reduce the strain on their health systems, and to provide economic relief for citizens adversely impacted by these measures.

Detection of the extent of the virus is dependent on the level of testing within a country. Effective testing allows public health authorities to identify COVID-19 cases and outbreaks and implement contact tracing and quarantining procedures to limit the spread. These statistics also help frame appropriate government response measures.

The extent and effectiveness of any of these measures depended on level of compliance among the population, status of the healthcare system within a country, and availability of resources to be able to implement such measures.

## Factors Analysed

While there are many different factors that impact or are impacted by the infection rate and the fatality rate of the virus within different regions, this report will be limited to the following:

- **Political :** Given the different types of government around the world, the impact of COVID-19 infection and fatality rates will also be varied. It is expected that countries in which any government responses are complied with, either forcefully or voluntarily, would have more favourable experience. On the other hand, if government systems are unstable, or the population is more inclined to reject government-imposed restrictions, experience would be adverse.
- **Economic:** The COVID-19 pandemic has had a significant impact on the economies of nearly all countries of the world. It can also be expected that different economic situations could affect COVID infection and fatality rates of various countries. Some countries may have been hit harder by the pandemic than others due to their expenditures of their GDPs. It could also be possible that countries with large economies could be reluctant to go into lockdown since it could result in economic recession and be potentially devastating to its economy. We hope to discover connections between COVID-19 and economy through trends in the datasets healthexp and educationexp.
- **Employment :** The COVID-19 pandemic has greatly impacted the employment rate of many countries around the world. It can be assumed that employment could be affected by the COVID infection and fatality rates in each country. Different countries have been impacted by the pandemic in various ways, as some may have been hit harder than others. In the case of employment, some countries shutdown their economy faster than others which may have dropped the infection/fatality rates quicker leading to less job loss. While other countries are still struggling to lower the infection/fatality rates causing longer economic shutdowns and more job losses as many businesses have had to close.
- **Technology :** The impact of COVID-19 has played a major role as it relates to technology. It may be assumed that countries with not that much access to the internet or technology may have a challenge as it may be difficult for them to be updated on fatality rates and recoveries within their country and by extension the world. One can assume that for smaller, less developed countries, persons might be unaware of some of the updates across the world, or unable to adjust to remote working due to technology limitations. Hence playing a major role in the technology aspect of things.

# Dates Selected and Data Sets

For this analysis we chose to look at infection and fatality rates for the period starting **March 12, 2020** to **October 28, 2020**. The starting date of March 12 for our study provides a glimpse of the state of the virus across the world just after the WHO declared the disease a pandemic, and allows for the measure of its progress almost from the starting point for many countries. The ending date of October 28 allows us to examine the course of the virus over a period that is long enough to determine and trends, but also provides a cut-off date that takes into account lags in reporting of data.

- COVID-19 cases
  - This dataset contains 10 variables and provides information on the number of COVID cases and deaths reported by country, and is updated daily from December 31, 2019 to the current date. The numeric values include cumulative totals, as well as daily, weekly and bi-weekly new cases and deaths. The data was obtained from the World Health Organization (WHO), the European Centre for Disease Prevention and Control (ECDC) and more recently, the Johns Hopkins University, all of which are recognized and reliable data sources.
- COVID-19 testing
  - This dataset contains 14 variables and provides information on the number of COVID tests by country, for the period starting January 1, 2020 to the current date. The data includes numeric values for the cumulative tests and the daily change in the number of tests, the cumulative testing rate per thousand of population and the daily change in that rate, the 7-day average testing rate per thousand of population and the change in that rate. The data also includes a short-term positivity rate and the number of tests per case, which provide an indication of the proportion of total tests that have a positive result. This information is provided by country with additional columns for country, date and source of data. Where available, information is typically sourced from government agencies within each country, and as a result, reporting methods and frequency are varied. The dataset included information on the source, as well as any relevant notes which indicate limitations or provide additional clarity on the data.
- COVID-19 government responses
  - This dataset contains 49 variables that facilitate comparing government responses. It includes a boolean indicator for each country on whether or not a measure was implemented, and if so, a ranking to indicate the severity or extent of that measure (i.e., whether localized or national). The data is collected by the University of Oxford, and includes data from January 1, 2020 to the current date. The measures are grouped as follows:
    1. Containment measures e.g., school and workplace closures, public event cancellation and restrictions on gatherings and internal movement
    2. Economic measures e.g., income support, debt/contract relief and other fiscal measures
    3. Health measures e.g., public information campaigns, testing policy, contact tracing, and facial coverings.
  - Several numerical indices are calculated for each country and date, based on the above measures, and converted to a value between 1 and 100:
    1. Government Response Index – overall level of stringency based on a combination of all measures taken and their severity
    2. Containment and Health Index – based on the containment and health measures only
    3. Economic Support Index – measures the level of stimulus provided
    4. Stringency Index – records the strictness of ‘lockdown’ measures taken.

# Political

## Data Set Analysis

For this factor, the following additional databases were used:

- Democracyindex
  - This dataset contains 167 observations and 12 variables, and provides a point-in-time view of the state of democracy in the world. The data includes for each country, a numeric score based on responses to 60 questions grouped in 5 categories which is then converted to an overall numeric Democracy Index to classify countries by type of Regime as follows:
    - Full Democracy with democracy index greater than 8
    - Flawed Democracy with democracy index between 6 and 8
    - Hybrid Regimes with democracy index between 4 and 6
    - Authoritarian Regimes with democracy index lower than or equal to 4.
  - The data and scores were collected and calculated by the Economist Intelligence Unit, a division of the world recognized Economist Group which provides research and analysis on economic and political affairs.
- Population
  - This dataset contains 238 observations and 4 variables and provides estimates of the total population of each country based on population census data, as well as surveys and projections based on past statistics. For the majority of countries, the most recent data contains population estimates for July 2020. The data is compiled by the Central Intelligence Agency and presented as part of the World Factbook project.

*System of Government* Using the Democracy Index data collected by the Economist Intelligence Unit, governments can be classified into the following regime types, based on their overall “democracy score”.

### Systems of Government

RegimeType	Num_Countries	TotalPop	Prop_Countries	Prop_Population
Full democracy	22	438035679	13.2 %	5.9 %
Hybrid regime	37	1226735377	22.2 %	16.6 %
Authoritarian	54	2543710570	32.3 %	34.5 %
Flawed democracy	54	3162255047	32.3 %	42.9 %

As can be seen, there are 4 main types of political regimes, with the majority of the world’s population belonging to Flawed Democracies.

In *authoritarian* countries, there is strong enforcement of government impose measures, with heavy penalties imposed on individuals who do not comply. In such regions, personal freedoms are denied in favour of social benefits.

In *fully democratic* countries, the degree of enforcement is much lower. However, due to the level of trust and openness within society, citizens are more likely to voluntarily comply with restrictions as they clearly see the benefit of these measures to their overall well-being.

In *flawed democracies*, where misinformation is widespread and there is a high level of distrust of the government and the media, a large portion of citizens believe these restrictions are an infringement on their personal freedoms.

*Hybrid regimes* represent a combination of authoritarian rule and democratic systems, and usually have a high level of corruption and lack of free and fair elections.

Given the differences in systems of government, it is likely that the incidence rates of COVID-19 within a country would depend on the government structure, its ability to respond effectively to the pandemic, and the level of public support.

## Approach Used

To determine the relationship between government system and COVID-19 infection and mortality rates, several data sources were used. As some of the information in these sources was inconsistent or incomplete, the analysis is based on the top 60 countries based on total COVID-19 reported cases at October 28, 2020. The dataframe was developed by combining information from the COVID-19 cases data, with population counts, government response data, COVID-19 testing statistics, and democracy indicator.

## Analysis

### Extent of COVID-19 by System of Government

Infection rates varied by country, with many parts of the world experiencing relatively high cases counts. The following summary shows the summary of descriptive statistics for the overall Infection Rate. As the mean is greater than the median, the data is right-skewed, implying that there are some outliers with larger infection rates.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.06543	5.99736	11.25986	14.74681	21.83321	53.87873

Infection and Mortality Rates by Government Type

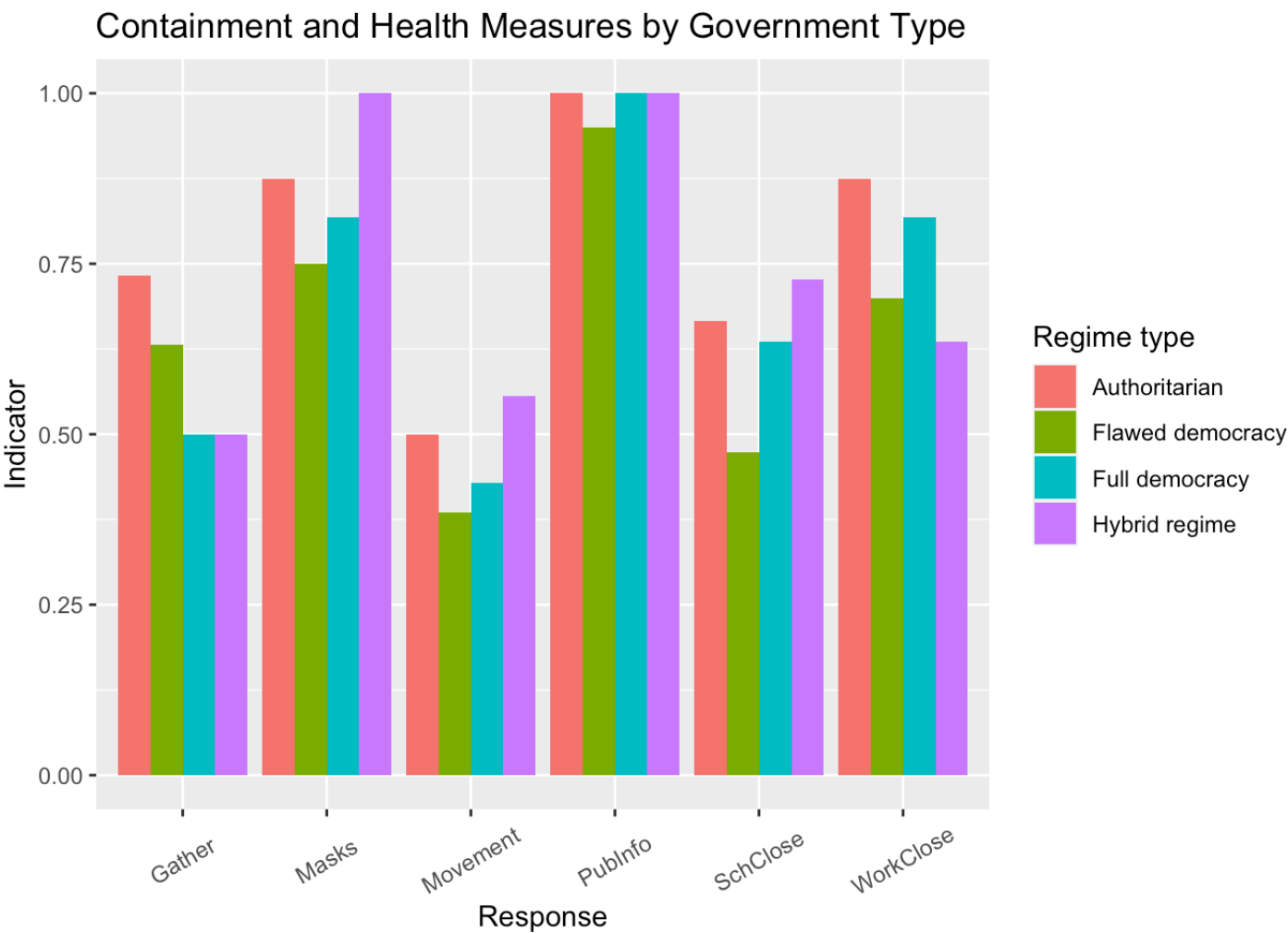
Regime type	InfRate_perThousand	MortRate
Authoritarian	2.09	2.28 %
Flawed democracy	10.53	2.69 %
Full democracy	14.00	3.2 %
Hybrid regime	3.74	2.18 %

The table above illustrates the infection and mortality rates by system of government. This shows that countries with more democratic systems of government fared worse during the pandemic, as the infection rate per thousand of population was almost 5 times higher than for Authoritarian and Hybrid regimes. Mortality rates (number of COVID-19 tests / number of cases) were more consistent, although still higher for democratic countries.

## COVID-19 Response by System of Government

The rate of spread of the virus can partly be attributed to the types of containment and health measures put in place by each country. Government responses to the pandemic varied not only by types of measures, but also by the degree or extent of application.

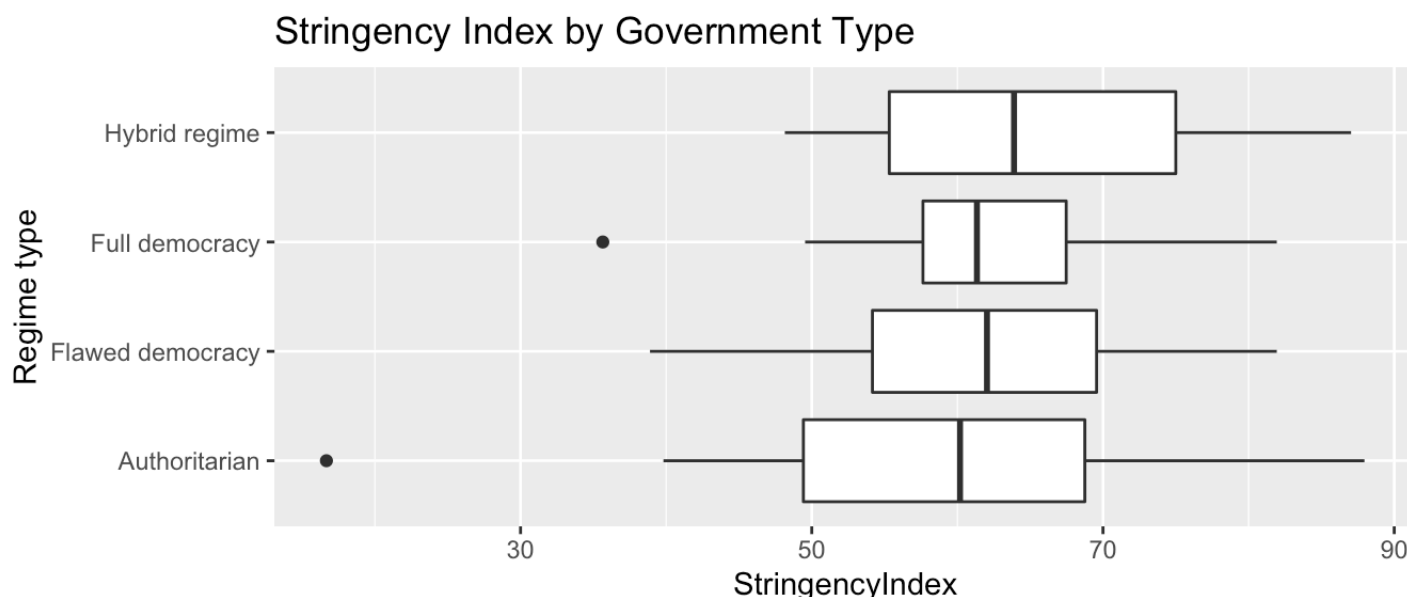
Countries introduced a combination of measures to curb the spread of the disease in order to protect their citizens. The following table summarizes the number of countries implementing specific measures.



Containment measures that included School and Workplace Closures, were put in place by at least 75% of Authoritarian countries, compared to less than 60% of Flawed democracies. While public information campaigns were adopted by almost all countries, only 75% of democratic countries introduced mandatory facial coverings, while this proportion was almost 90% for Authoritarian and Hybrid regimes combined.

The Government Response data collected by the University of Oxford records presence of specific measures as well as their severity. For example, if a government implemented restrictions on mobility, they may apply to the entire country, or for a specific region within that country where COVID-19 cases were more prevalent. The

types of measures also varied at different points during the study period, depending on the level of infection. The severity of combined measures is determined by the Stringency Index, which applies weightings to each measure based on level of intensity. Similar indices are also provided for Containment and Health measures, Economic measures, and overall Government response. The following boxplot shows the stringency index by type of government system.



These charts show the variation in stringency levels by regime type: 1. The median stringency index was lowest in Authoritarian regimes. 2. The interquartile range was widest for Authoritarian and Hybrid Regimes, indicating wider variability in response. 3. While Authoritarian, Hybrid and Full Democracies are more right-skewed, Flawed Democracies are the opposite.

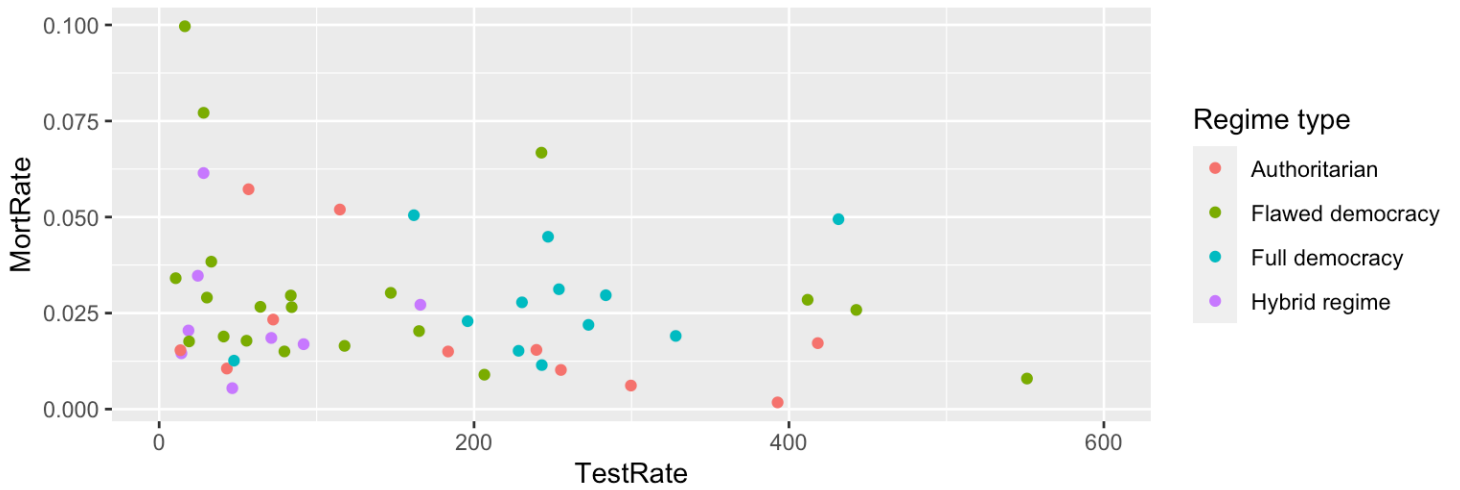
## COVID-19 Testing by System of Government

COVID-19 testing statistics are based on data collected by the World in Data project, which records reports provided by government agencies within each country. As these sources are not consistent, there is some variation in the completeness and frequency of reporting, as well as what the underlying data represents. For example, countries may report number of unique tests, while other provide total tests, including multiple tests on the same individual.

The accurate reporting of the incidence of COVID-19 within a country depends on the level of testing being carried out. Initially, these levels would have been low as countries worked to develop and distribute testing kits, and resources to perform tests and process results. Detection of COVID-19 infection rates and effective contact-tracing would be measures that would help in determining the appropriate government response.



## COVID-19 Mortality vs Testing Rate by Government Type



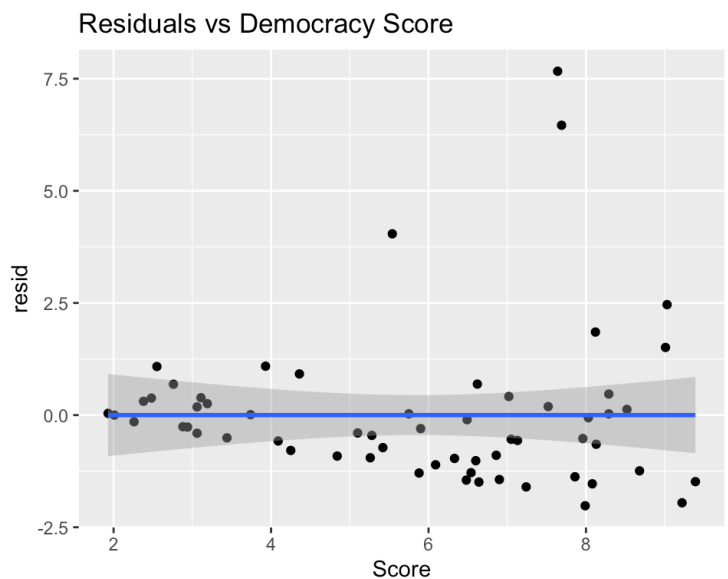
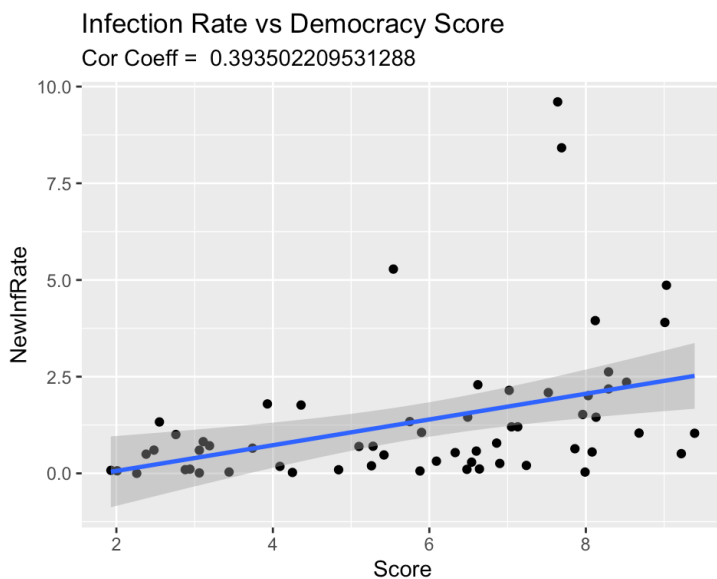
Testing rates are much higher in Authoritarian and Full democracies compared to the other 2 systems of government. Adequate testing helps identify positive COVID-19 cases, and allows public health authorities to implement contact tracing and quarantine measures that can help curb the spread of the virus.

## Relationship between Type of Government and COVID rates

### Model Statistics

term	estimate	std.error	statistic	p.value
(Intercept)	-0.6023645	0.6364618	-0.9464268	0.3478574
Score	0.3325508	0.1020153	3.2598133	0.0018679

```
## [1] "R-squared value = 0.15"
```



This shows that there is a positive linear relationship between the rate of new infections and the system of government as determined by the democracy score, with a correlation coefficient of 0.39. The p-value is 0.018, indicating that it is a moderately significant variable. Based on the R-squared value, this model shows that about 15% of the variation in infection rate can be attributed to the democracy score. Therefore, there are other factors that impact infection rates.

## Findings

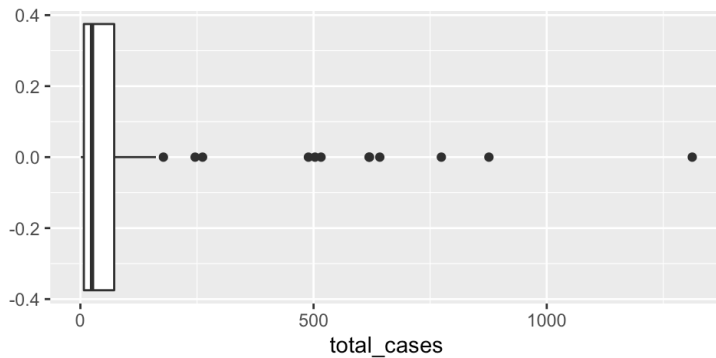
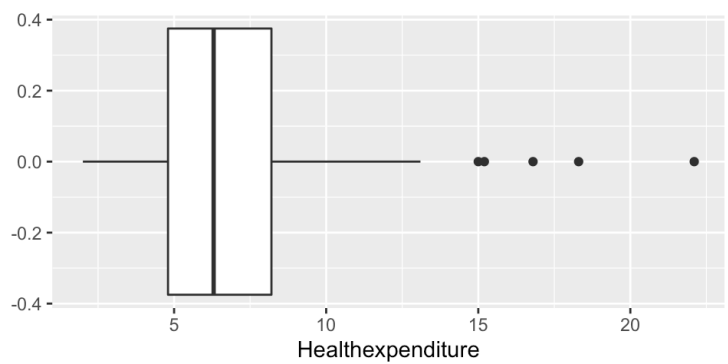
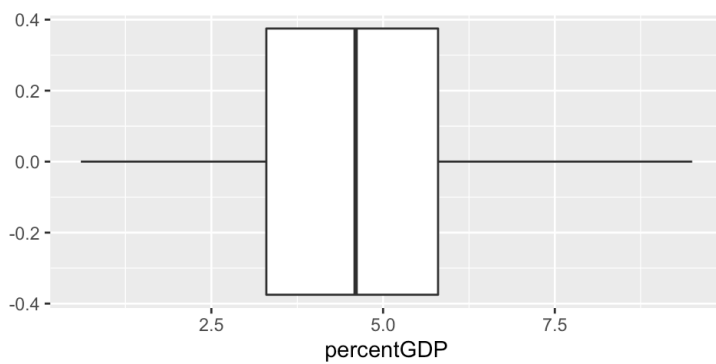
Based on the analysis above, it appears that there is a weak relationship between government system and level of COVID-19 infection. Authoritarian government systems appear to have been more successful at controlling the level of new infections, likely because they exercise more control over their citizens. In contrast, democratic countries have been less successful at controlling the virus levels despite the containment measures that were put into place. For these countries, it is likely that citizens are more accustomed to a certain level of freedom, and enforcement of strict measures was more difficult. While the model can detect patterns in data, it is difficult to account for individual social behaviours, particularly when personal choice is highly prevalent.

# Economic

## Data Set Analysis

- healthexp
  - The data set 'healthexp' contains 3 variables and 190 observations. The first variable, 'Country', is of type character and contains the full name of the country. The second variable is 'Current Health Expenditure' which contains the health expenditure of the country as a percentage of its economy and is a character vector. The last variable is 'year', which contains the year in which the observation was recorded and is type double. My assumption would be that higher health expenditure would result in less significant COVID implications. Research question: Does higher health expenditure cause lower COVID cases and deaths and higher recovery rates?
- educationexp
  - A data set 'educationexp' contains 4 variables and 173 observations. 'Country' is a character variable and contains the full names of the observation's countries. '(% of GDP)' is a double variable and contains the education expenditure as a percent of the total GDP of an observation. The variable 'Rank' is type double and contains the rank of the observations based on the '(% of GDP)' variable. The 'Date of Information' is a double variable and contains the year in which the observation was last updated. One may assume that as the education expenditure increases then they would have less money for health services. Therefore, a higher education expenditure may result in larger COVID implications. Research question: Does higher education expenditure cause higher COVID cases and deaths and lower recovery rates?

## Single Variable Analysis



### *educationexp*

The set contains 4 outliers all of which are on the right/top of the data. These points skew the data and therefore will be removed for more accurate conclusions. From visual analysis of the boxplot the middle quartiles of the set seem to be almost even, and the whisker of the box plots seem fairly even too. The standard deviation of education expenditure is 2.14 including the outliers and 1.81 excluding. This gives further justifications on why they should be removed, the standard deviation decreases by 15%, meaning the data becomes less spread out, when the 4 outliers are removed. The mean of the data is 4.85 and the median is 4.7 which means that the data has a slight positive skew. The outliers mentioned before seem to have some responsibility for this.

### *healthexp*

The data set contains 6 outliers one of which is an extreme outlier that will be removed since it is unrealistic and provide little insight for future trends and conclusions. This is the observation of the Democratic Republic of Congo and its health expenditure is 403% of its GDP which is extremely unrealistic and skews the data incredibly. The observations contain the country Iran is also removed since it does not have any data for Current Health Expenditure. After these removals, the mean of the Current Health Expenditure is 6.81 and the median is 6.3. The data has a slight positive skew since its mean is greater than its median which is likely caused by the outliers on the right end of the data. The standard deviation of the Current Health Expenditure is approximately 2.9 which is relatively large given that the IQR is only 3.4.

### *COVID\_cases: March 12*

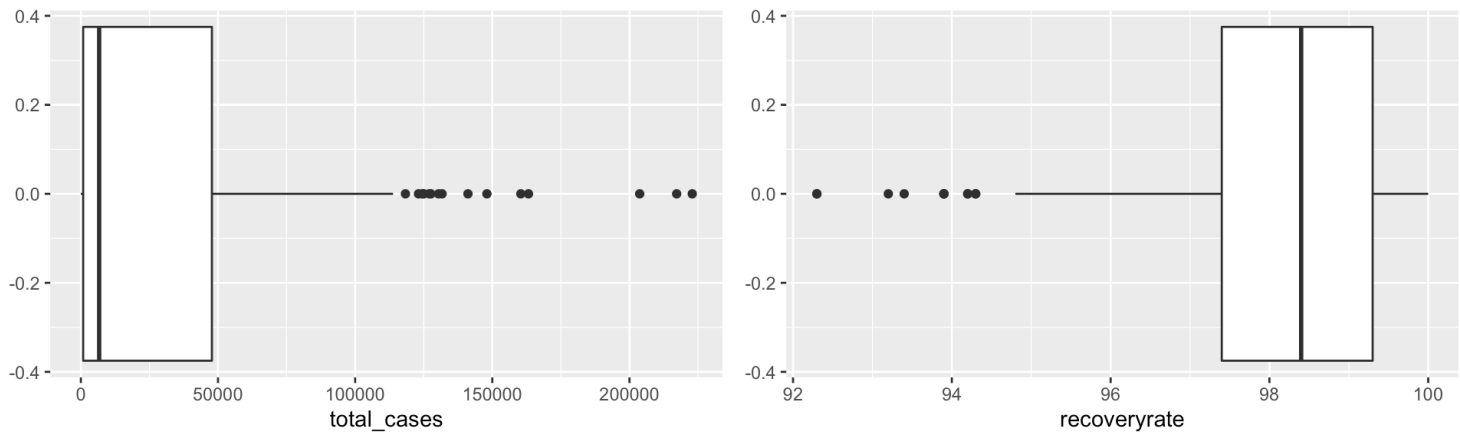
The total cases on the other will be a more accurate than total deaths at this stage in the pandemic since most countries will have cases but few, if any, deaths. There are some countries in which the pandemic was already in a different stage of the pandemic which would result in extreme outliers in the dataset. The aim of using the March date is to see a country's early COVID-19 trends. Therefore, some of these outliers should be removed so we are able to focus more on the core of the data. I have decided to remove the observations with more than 1,500 total cases. The remaining data is positive skewed with a median of 25 and a mean of 114.1. The

data is very spread out with a standard deviation of 228.57 which is likely caused by the outliers. The recoveryrate variable is inaccurate for the March 12th day since it is are dependent on total\_deaths, which at this stage in the pandemic is not accurate.

#### *COVID\_cases: October 28*

There are 3 observations in the total\_cases variable that are extreme outliers and are skews the data immensely. These observations should be removed to increase the accuracy of the trends. By removing these observations, the data becomes less skewed.

The recovery rate contains one usual data point which contains a country that has a nearly 30% deathrate which is significantly different from the other observations. Removing this observation results in the data being less skewed.



## Two Variable Analysis

#### *Healthexp & COVIDMAR12*

The graph shows the relationship between Healthexpenditure and the resulting total cases upon until March 12. There seems to be a positive linear relationship: as the percentage of GDP allocated to health increases so do the number of COVID cases. This information seems confusing since one might assume that more money spent on health would result in fewer COVID cases. The correlation of the variables approximately 0.59 which shows a moderate positive correlation as seen in the plot. The covariation is 360, which is exceptionally high due to some outliers in the total\_cases variable. The linear model of the variable appears significant since the p-value is very close to 0 but the fit is not very good as the R-squared value is only 0.34.

#### *Healthexp & COVIDOCT28*

The graph of Healthexpenditure and the total\_cases on October 28 still shows a positive relationship but to less of a degree and less accuracy. The correlation coefficient is only 0.05 meaning there is a weak positive relationship and the covariation a lot higher than it was in March. The p-value is also much larger whereas the R-squared value is much smaller, which signifies that the linear model is far less accurate. Overall, the trends from March to October have changed significantly. The relationship was much stronger in March and there was less covariation in the variables.

The graph of Healthexpenditure and recovery rate of COVID cases shows a negative linear correlation. The correlation coefficient is -0.23 which means that there is a weak negative correlation meaning that as healthexpenditure increases the recovery rate decreases which is once again unusual. The covariance in

-0.95, the p-value is 0.05 and the R-squared is 0.05 which shows that there could be other information that may improve the model fit.

Overall, the healthexpnditure shows that the higher percentage of a country's GDP that is spent on healthcare results in negative effects on its COVID cases, deaths, and recovery rates. Although all variables show the same pattern but many of the correlation coefficients show weak relationship and there are inaccuracies shown in the linear models and covariation.

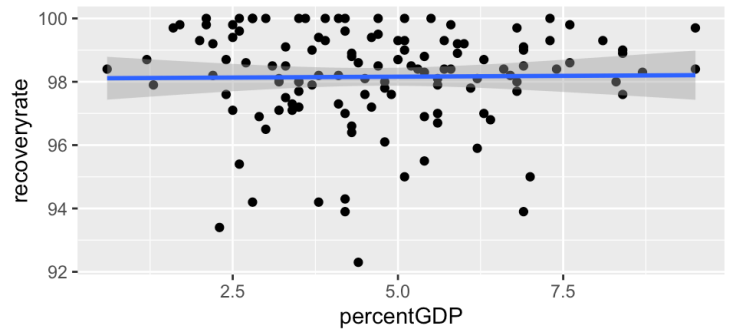
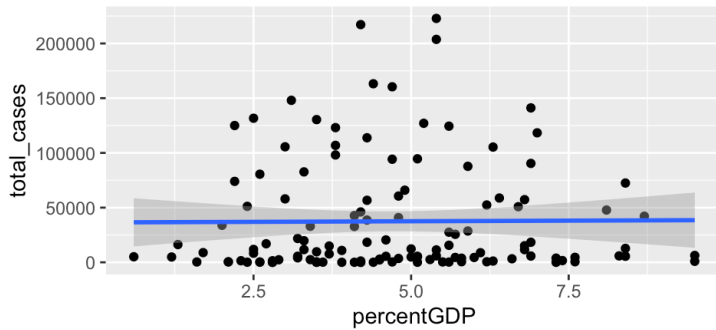
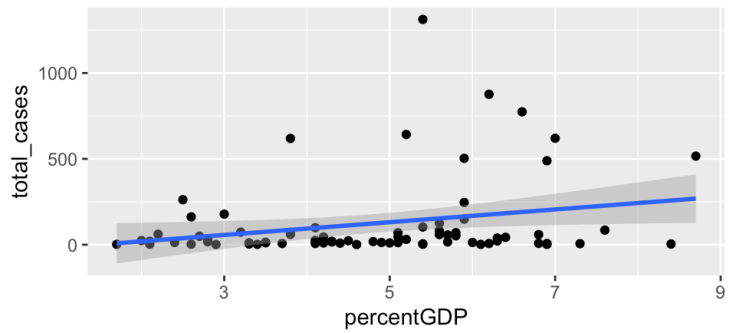
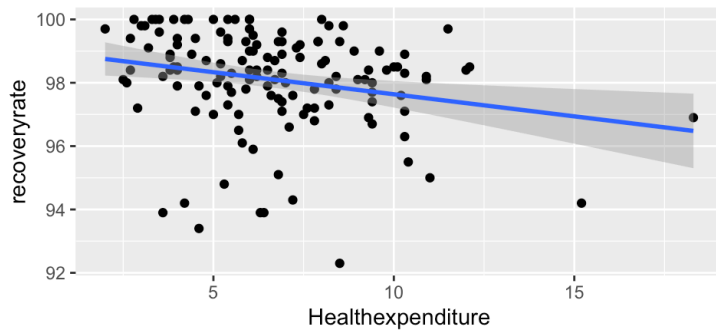
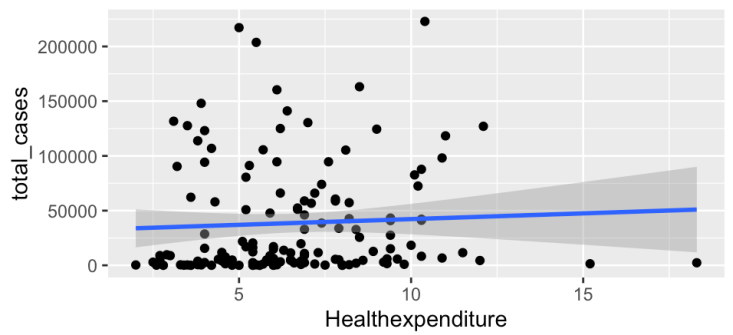
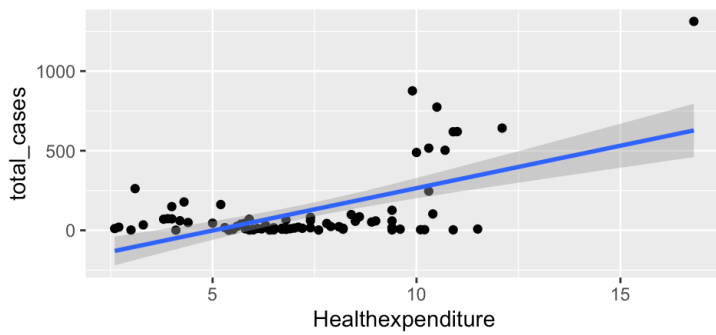
#### *Educationexp & COVIDMAR12*

The graph between the variables percentGDP and total\_cases shows a positive linear relationship. A correlation coefficient of 0.25 signifies a slightly weak positive linear relationship between the two variables which is understandable because if a country spends more money on education then it has less to spend on health which could causes higher COVID infection and cases. The covariance is 99.39 and shows the spread of the data points is fairly small. The linear model is a poor fit since its R-squared is very low but so is the p-value.

#### *Educationexp & COVIDOCT28*

From the graph it is difficult to see whether the linear relationship between percentGDP and total\_cases is positive and negative. The correlation coefficient is nearly 0 which shows an extremely weak positive correlation. The covariance is also fairly large which shows that the data points are pretty spread out. The linear model does not seem to explain all the variation as the R-squared very close to 0 and a p-value close 1. The difference between the relationship of these variables in October vs March is that the strength of the correlation has decreased by a lot and the covariance has increased as well.

The graph of percentGDP and recovery shows a very small positive linear correlation. The correlation coefficient is 0.01 which shows a very weak positive linear relationship. This relationship is not very accurate but nevertheless, one would assume a negative relationship since higher education expenditure would result in a lower recovery rate. The data's spread is not very large, which is good, as its covariance in only 0.3. The linear model is inaccurate since the R-squared is very close to 0 and the p-value is fairly close to 1.



## Findings

Based on the results of the data analysis, there seems to be very little correlation between economic trends in relation to COVID-19. When comparing the COVID data on the dates March 12 to October 28, one is able to conclude that any trends seen in the early stages of COVID weakened or become irrelevant in the later COVID data. In the initial stages of COVID there appears to be a slight linear relationship between education expenditure of a country in relation to its COVID cases. The larger of a percentage of its GDP a country spends on education, the higher its cases were in the beginning stages of COVID. There are no recommendations that can be made based on this trend since this occurred in the initial stages and at the moment we are entering the height of our second wave. The health expenditure data for the March 12 shows an interesting trend. It shows a moderate positive correlation between COVID cases as a result of the health expenditure of a Country. The reason behind this could be that countries that spend more money on health may feel more confident in their health systems and as a result took the initial stages of COVID too lightly. In the data of second date, October 28, there seems to be absolutely no relationship between COVID cases and education exponential nor the health exponential. The correlation coefficients are all very weak and the linear models do not full explain the differences in the infection rates by contry. Overall, there do not seem to be any specific recommendations to the leaders of countries for response to COVID 19 based of economic factors and the data sets explored in this report.

# Employment

## Data Set Analysis

- Laborforce
  - This dataset contains 16104 observations and 4 variables. The 4 variables are Country Name, Country Code, Year, and Labor Force. The integer variable 'Year' dates the years 1960-2020, we only examine the years after to 1989 because that is when labor force numbers started to be recorded which cuts the number of observations down to 8184. The Country Name variable is of type character and Country Code is a character variable and contains abbreviated versions from the Country Name column. The integer variable Labor force contains the value of the corresponding countries labor force in the specified year. If data is missing in the Labor Force variable for a year it is filled in as NA.
- Unemp
  - This dataset contains 218 observations and 4 variables. The character type variable 'Country' carries the full name of countries listed in the observations. The variable Rank is of type integer and ranks the countries from lowest to highest percentage of unemployment in order. The column Percent is of type double and lists the unemployment rate in each country. The character variable 'Date of Information' gives us an estimated year that the data for each country's unemployment rate was taken in.
- Unempyouth
  - This dataset contains 181 observations and 3 variables. The character variable country lists the full name of countries in the observations. The variable Rank is of type integer and ranks the countries from highest to lowest percentage of youth unemployment (ages 15-24) in order. The variable Percent is type double and gives us the youth unemployment rate for each country listed in the observations.

To take a better look at the distribution of Labor Force globally, the data was split into 4 levels: - Very High: given to countries with a Labor force greater than 70 million people - High: given to countries with a Labor force between 30-70 million people - Mid: given to countries with a Labor force between 10-30 million people - Low: given to countries with a Labor force lower than 10 million people

Separating the countries into categories like this makes it easier for us to see how the COVID-19 pandemic affects countries with different sized Labor Forces and can analyze its effect on employment.

### Labor force Levels

Tier	NumCountries	TotalLaborForce
Very High	6	1058604324
High	12	558612139
Mid	27	483141622
Low	106	345082766

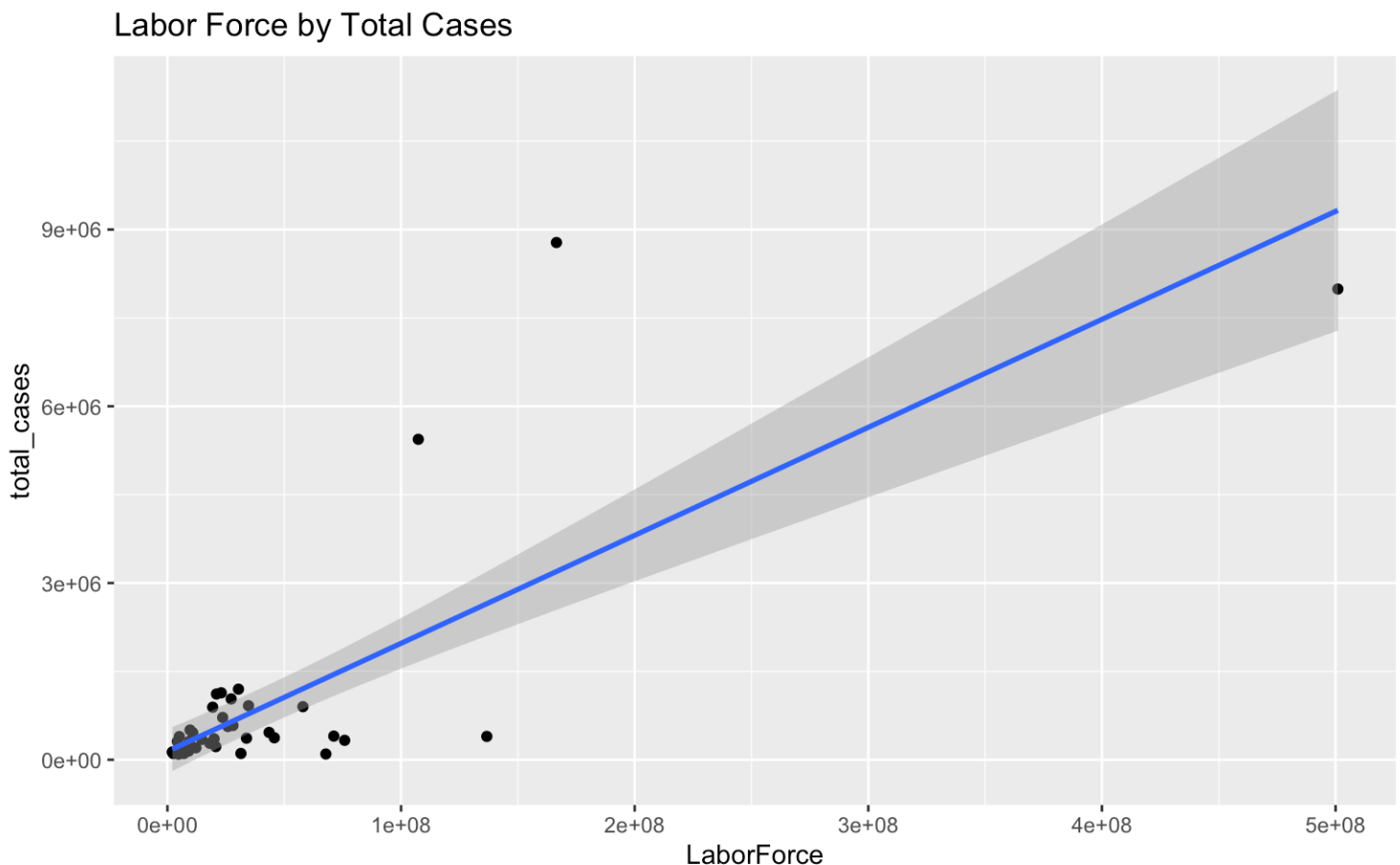
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	54360	1326936	4729920	16194973	13000922	500995152

As you can see from the Labor Force Levels table, the majority of countries in our data have a Labor Force that is below 10 million people. The majority of countries did not exceed a labor force greater than 20 million as the 5-point summary shows. With this in mind, we are now going to look for a relationship between COVID-19 infection/mortality rates and employment by examining if countries with a larger labor force have higher infection and mortality rates than those with a smaller labor force.

## Approach Used

To determine the relationship between COVID-19 infection/mortality rates and employment, we examined several different data sources. Due to the fact that some of the information provided in the given data sources is incomplete, our analysis is based on the top 50 countries total COVID-19 cases as of October 28, 2020. The dataframe was developed by combining information from the COVID-19 cases data, with Labor Force levels, government response data, COVID-19 testing statistics, and unemployment percentages.

The COVID-19 pandemic is affecting some regions of the world more severely than others, with some countries total cases being significantly higher than others. The following graph illustrates how a country's labor force contributes to the total number of COVID-19 infections they are experiencing as of October 28, 2020.



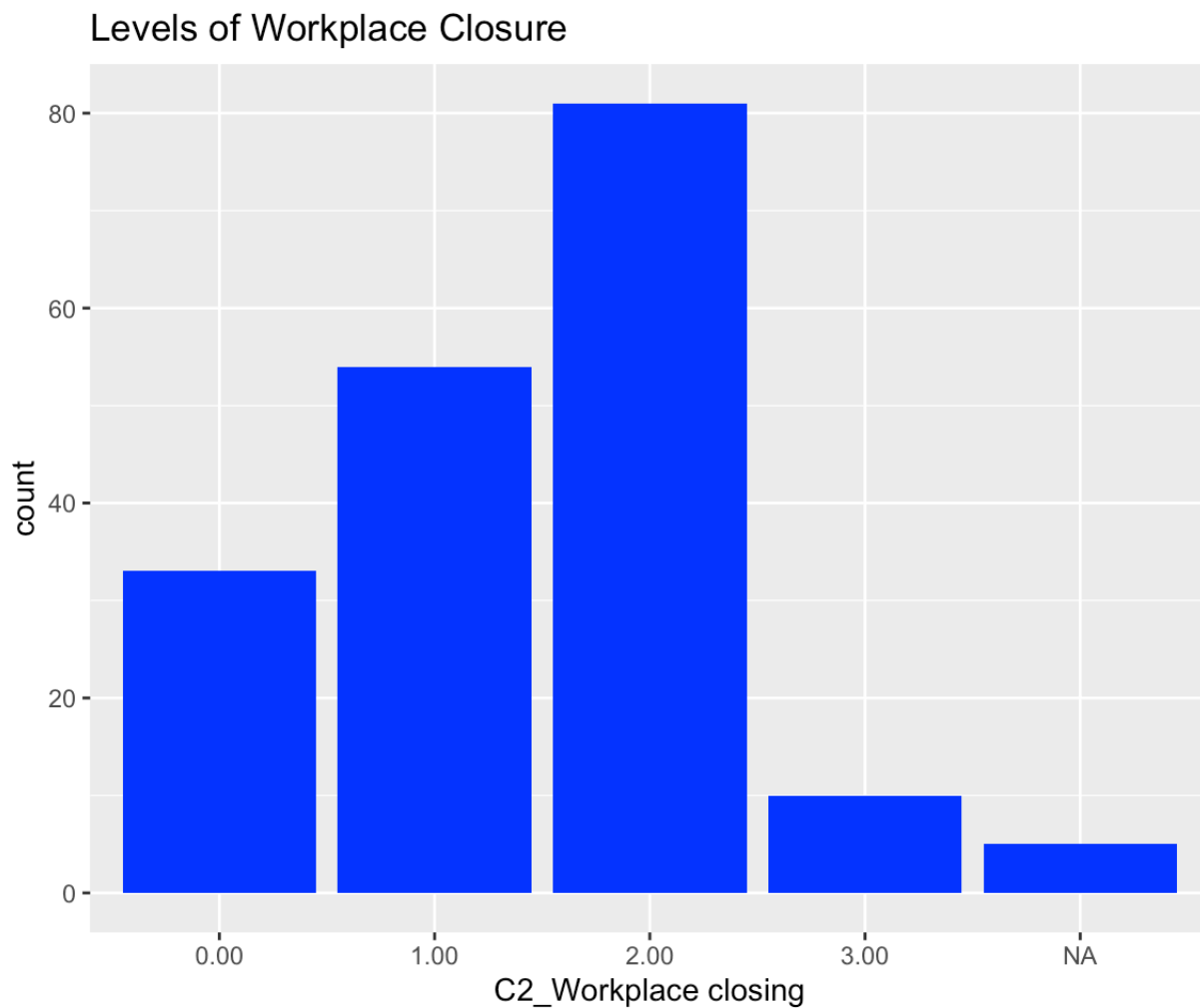


```
##
## Call:
## lm(formula = Oct28covid_cases_final$total_cases ~ Oct28covid_cases_final$LaborForce)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2252366  -187909  -108416   111705   5584485
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.432e+05  1.865e+05   0.768   0.447
## Oct28covid_cases_final$LaborForce 1.833e-02  2.162e-03   8.479 8.37e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1131000 on 44 degrees of freedom
## (4 observations deleted due to missingness)
## Multiple R-squared:  0.6203, Adjusted R-squared:  0.6117
## F-statistic: 71.89 on 1 and 44 DF, p-value: 8.373e-11
```

From the illustration we can see that as many countries fall into the low tier category of total labor force the number of total cases also is low. Though the graph appears to show a somewhat linear relationship between labor force and total cases, our regression model tells us otherwise. The model's R-squared value shows that 62% of the variation between total cases can be explained by labor force. 62% is more than half but we cannot assume that a higher labor force causes an increase in the total number of cases for a country from this analysis. There are a few countries that have high counts of total cases even though their labor force numbers are not necessarily high. This difference may be able to be explained by the government response of their country and if they decided to close their workplaces.

### *Workplace Closing as of Oct28*

With COVID-19 numbers still increasing during October, stricter rules were enforced to ensure people are staying home and avoiding contact with others. This includes the closing of workplaces to reduce the need to leave home. The chart below represents the number of countries that either recommended closing, required closing in some regions or for some workers, or required closing for everyone other than essential workplaces. - 0 represents the countries that didn't close any workplaces - 1 represents the countries that recommended that workplaces close or for their staff to work from home - 2 represents the countries that required some workplaces to close or just some staff to work from home - 3 represents the countries that closed all workplaces except for the essential workplaces(i.e., grocery stores, hospitals) - NA is was given to the countries that there wasn't data for



With COVID-19 incidence now the highest it has ever been, the majority of countries implemented some level of workplace closure to reduce the climbing amount of. If the COVID-19 cases continue to increase in the future, the required workplace closures will likely be continued. This could prove to be costly for many businesses and some will have to close permanently resulting in a rise in the unemployment rate.

## Findings

Based on the results of the data analysis, there seems to be a correlation between employment trends in relation to COVID-19. When examining the COVID data from October 28 and the given data sets, it is hard to draw an accurate conclusion based on the information provided. An assumption that we can make from the trends seen in COVID-19 cases and government responses is that the overall employment rate will decrease as a result of measures taken to control the COVID infection and mortality rates. This occurs because the longer workplaces stay closed, the more money their businesses will lose and eventually they may not be able to afford to stay open. In the data provided we are only given outdated unemployment and youth unemployment rates from before COVID-19 arrived. To be able to come to an accurate conclusion we would need to review how unemployment rates have changed during the time between when the pandemic started and October 28. Therefore, there are no specific recommendations to be made based on this factor alone.

## Technology

The impact of COVID-19 has played a major role as it relates to technology. It may be assumed that countries with not that much access to the internet or technology may have a challenge as it may be difficult for them to be updated on fatality rates and recoveries within their country and by extension the world. One can assume that for smaller countries persons might be unaware of some of the updates across the world. Hence playing a major role in the technology aspect of things.

## Data Set Analysis

- **Accesstoelectricity**
  - This dataset contains 264 observations and 8 variables. The character variable 'Country Name' carries the full name of countries listed in the observations. The character variable "Country Code" has all the abbreviations of the countries that are listed in this dataset. The integer variable "Year" dates the years from 2015-2020 with the last two years containing N/A values and type double variables listed from 2015-2018. Within these variables are columns of persons that have access to electricity as percentage of the population.
- **Internetuser**
  - This dataset contains 264 observations and 8 variables. This dataset has a similar layout to accesstoelectricity. The character variable 'Country Name' contains the full name of countries listed in the observations that use the internet. The character variable 'Country Code' has the abbreviations of names within the 'Country Name' dataset. What is different with the two datasets is that the integer variable 'Year' dates from 2015-2019 have columns with persons who use the internet as a percentage of the population. The remaining year 2020 has N/A values.

Diving deeper into the data, we can see which countries have full access and the opportunity to use the internet, those that are in the middle and those who are at the lowest end. For those with low access to electricity, there would expect to be low internet use.

By examining this access, we can determine during the pandemic who maybe have been getting alerts and news updates within their country.

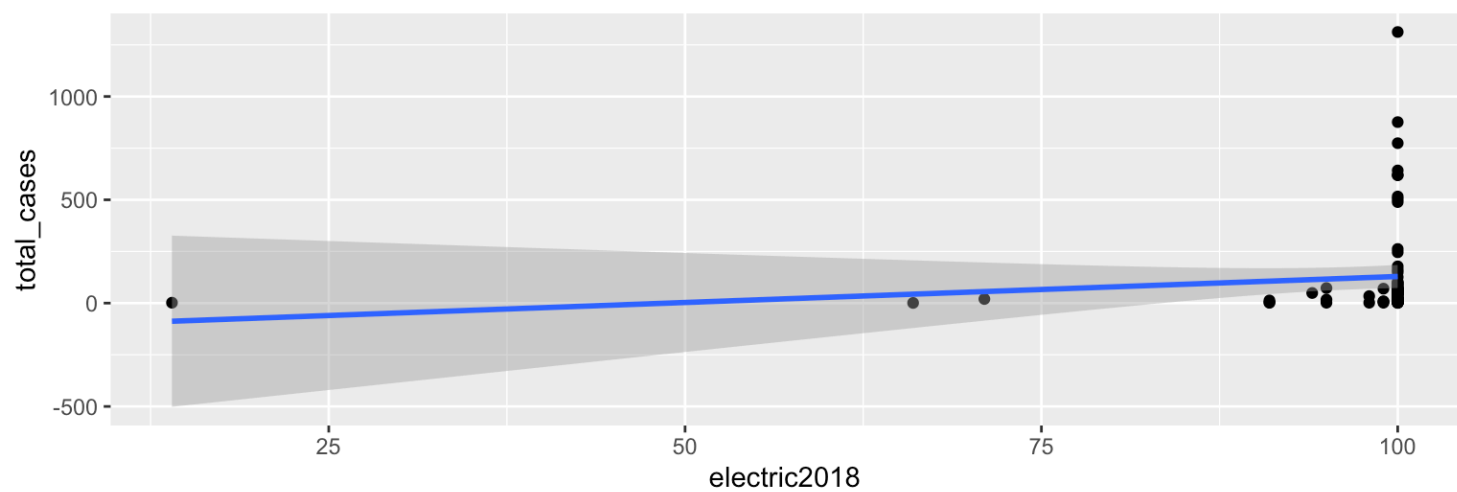
### Access to Electricity Levels

Level	NumCountries	meanAccesstoelectricity
Very High	97	99.10
High	1	89.00
Mid	1	66.00
Low	4	36.25

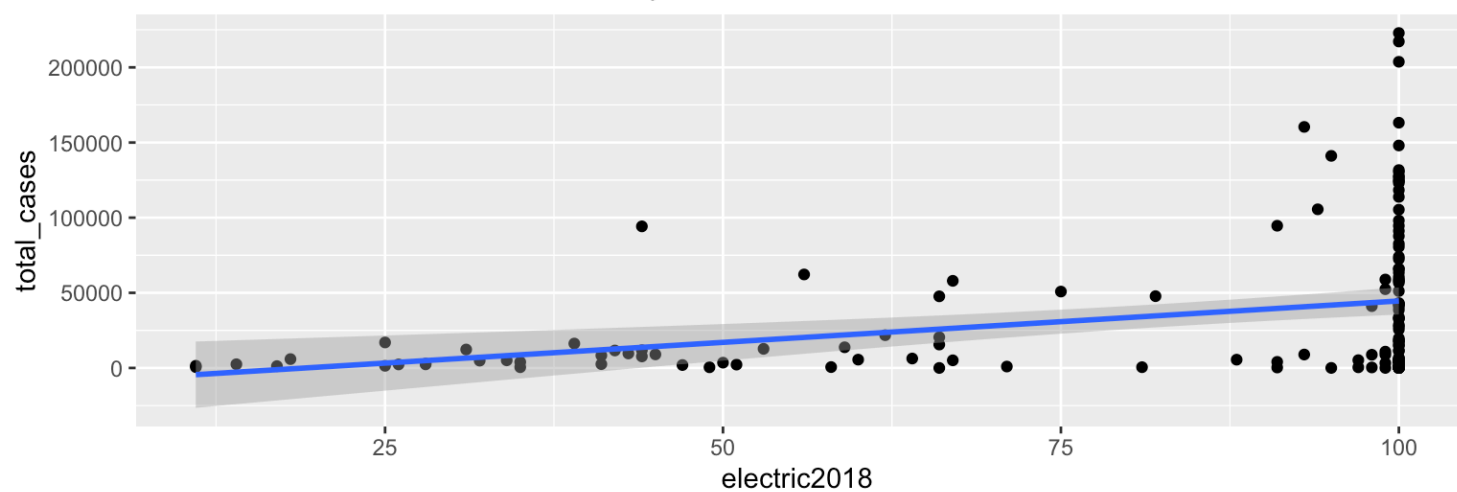
As we can see from the table most countries have a very high level which means they have access to electricity, which should correlate with the opportunity to be high internet users. We can now take this information and investigate how this has impacted the ability of individuals to receive news and COVID-19 updates within their country, or to be able to work or attend classes remotely in order to reduce mobility within the country.

# Comparisons and Relationships

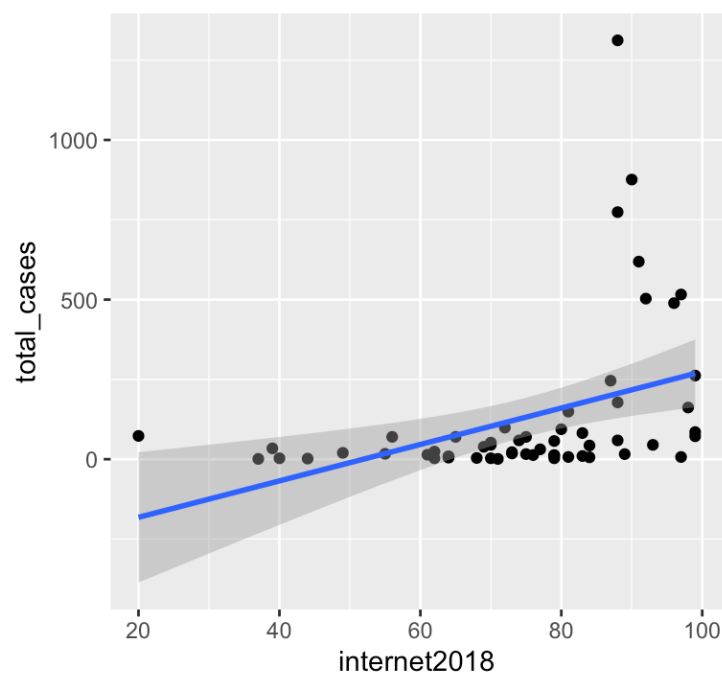
COVID Mar12 Cases and Electricity Access



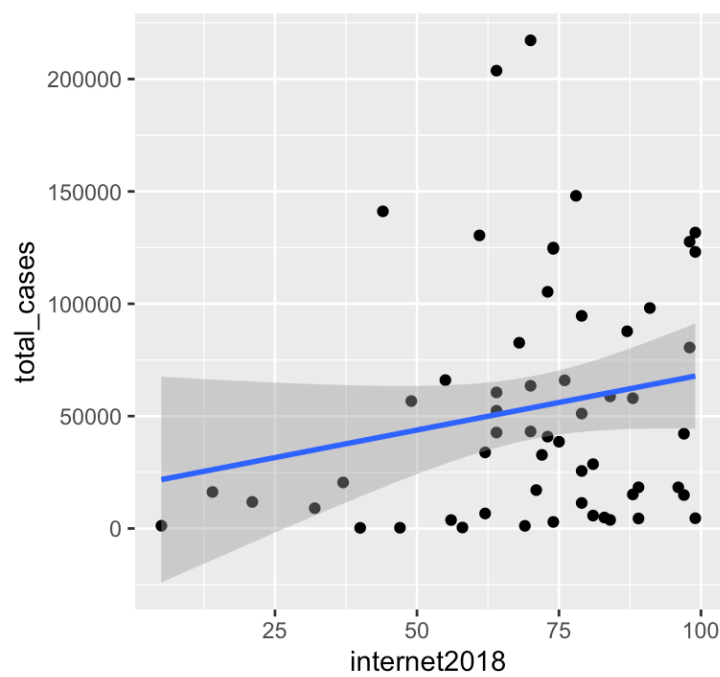
COVID Oct28 Cases and Electricity Access



COVID Mar12 Cases vs Internet Access



COVID Oct28 Cases vs Internet Access



# Findings

There appears to be a positive relationship between level of technology and incidence of COVID-19 cases. This is stronger in the early phase of the pandemic compared to the later periods. This may be due to the fact that information on the virus was limited at first but this improved over time. However, as people became better informed, they were able to take proper precautions.

# Combining Factors

While each of the factors analyzed may have had some relationship with overall infection rate, it is also possible that there is a combined impact.

Combined Model Statistics	
Statistic	Value
r.squared	0.3638136
adj.r.squared	0.2312747
sigma	1.6356481
statistic	2.7449581
p.value	0.0424569
df	5.0000000
logLik	-53.9821765
AIC	121.9643530
BIC	131.7727347
deviance	64.2082740
df.residual	24.0000000
nobs	30.0000000

When combining the several factors we discussed into a single model, including interactions between variables, there appears to be a slightly better, with the model explaining only about 36% of the variation in new infection rates among countries. This may be due to factors that are not captured in the model, such as population demographics.

# Conclusion and Recommendations

While there are some relationships between each of the factors explored and the level of infection within a country, it is not possible to conclude that any specific factor or combination of factors can accurately model this rate. One challenge we encountered in our analysis was with the data. Quite often data was incomplete or outdated, particularly for information on each country's economic and employment statistics. Additional factors can be explored to help improve the model.

However, the analysis showed that while different government systems responded in some way to the pandemic, the severity of these responses varied, as well as the level of enforcement. In addition, while some countries may have been in a more stable economic position prior to the pandemic, this did not mean that they were able to have much of an advantage in controlling its spread. The size of the labour force, as well as unemployment rates did not seem to impact the spread of the disease, although it is likely that unemployment rates will increase as a result. Finally, the level of technology as determined by availability of electricity and access to the internet impacted whether governments were able to broadcast information to the population in a timely and efficient manner.

# Appendix

## Data Sources

- <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>  
(<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>)
- Covid-19 cases : [https://covid.ourworldindata.org/data/ecdc/full\\_data.csv](https://covid.ourworldindata.org/data/ecdc/full_data.csv)  
([https://covid.ourworldindata.org/data/ecdc/full\\_data.csv](https://covid.ourworldindata.org/data/ecdc/full_data.csv))
- Covid-19 testing : <https://covid.ourworldindata.org/data/testing/covid-testing-all-observations.csv>  
(<https://covid.ourworldindata.org/data/testing/covid-testing-all-observations.csv>)
- COVID-19 government responses: <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker#data> (<https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker#data>) [codebook] [https://raw.githubusercontent.com/OxCGRT/covid-policy-tracker/master/data/OxCGRT\\_latest.csv](https://raw.githubusercontent.com/OxCGRT/covid-policy-tracker/master/data/OxCGRT_latest.csv) ([https://raw.githubusercontent.com/OxCGRT/covid-policy-tracker/master/data/OxCGRT\\_latest.csv](https://raw.githubusercontent.com/OxCGRT/covid-policy-tracker/master/data/OxCGRT_latest.csv)) to understand the meanings of the values
- Democracy index: <https://www.eiu.com/topic/democracy-index>  
(<https://www.eiu.com/topic/democracy-index>) developed by the [Economist Intelligence Unit]  
<http://www.eiu.com/home.aspx> (<http://www.eiu.com/home.aspx>) , which is contained in the table from the Wikipedia page ([https://en.wikipedia.org/wiki/Democracy\\_Index](https://en.wikipedia.org/wiki/Democracy_Index)  
([https://en.wikipedia.org/wiki/Democracy\\_Index](https://en.wikipedia.org/wiki/Democracy_Index))).
- Population index: <https://www.cia.gov/library/publications/resources/the-world-factbook/fields/335rank.html> (<https://www.cia.gov/library/publications/resources/the-world-factbook/fields/335rank.html>)
- Health <https://www.cia.gov/library/publications/resources/the-world-factbook/fields/409.html>  
(<https://www.cia.gov/library/publications/resources/the-world-factbook/fields/409.html>)

- Education <https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2206rank.html> (<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2206rank.html>)
- LaborForce <https://data.worldbank.org/indicator/SL.TLF.TOTL.IN> (<https://data.worldbank.org/indicator/SL.TLF.TOTL.IN>) : the up-to-date data we use. [CIA] <https://www.cia.gov/library/publications/resources/the-world-factbook/fields/218rank.html> (<https://www.cia.gov/library/publications/resources/the-world-factbook/fields/218rank.html>) : has a slightly outdated data set.
- Unemp <https://www.cia.gov/library/publications/resources/the-world-factbook/fields/220rank.html> (<https://www.cia.gov/library/publications/resources/the-world-factbook/fields/220rank.html>)
- Unempyouth <https://www.cia.gov/library/publications/resources/the-world-factbook/fields/373rank.html> (<https://www.cia.gov/library/publications/resources/the-world-factbook/fields/373rank.html>)
- AccesstoElectricity <https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS> (<https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS>)
- InternetUsers <https://data.worldbank.org/indicator/IT.NET.USER.ZS> (<https://data.worldbank.org/indicator/IT.NET.USER.ZS>)