# DATA3888 Group Project

Covid C2

2022-05-31

# Contents

```r
# Load required packages
library(tidyverse)
library(dplyr)
library(tidyr)
```

# Introduction

The Coronavirus disease (COVID-19) is a virus that was declared as a pandemic in early March of 2020. Since then, various different variants with different traits have mutated, which has affected its contagiousness. Governments have been tasked with limiting the spread and the impacts the virus causes on society, which may include financial/economic impacts, as well as physical and even mental.

As the virus has continued to evolve, so have combative measures such as the rapid development of vaccinations to reduce the effects of serious illness or even death from the virus. As such, governments have employed various measures to limit the demands on hospitals, with measures such as mandatory facial coverings, stay-at-home orders and also a strong recommendation for vaccinations.

This leads into the next section, which aims to understand whether the development of a country affects its response and how successful its measures tackle the pandemic, and at what cost these responses affect the freedoms of its citizens.

## Motivation

Realistically, there is no perfect response to a situation like this, and there is also no perfect metric that defines success or failure. As such, a method to create our own metric to determine success will be created.

Since the pandemic began, various governments have been battling with their own country's various issues, which has included economic damage, death and most importantly, illness. The motivation to better understand how the varying HDI score and GDP of each nation which has needed to respond to this pandemic, and identifying the best performing countries for certain clusters to assist in future pandemic response planning would greatly benefit nations with similar financial and developed resources such as hospitals and infrastructure.

# Stakeholder

The main stakeholder identified for this product is the World Health Organisation (WHO) as the group that provides advice and global health orders. This product is crucial in the decision making process as it allows to better provision certain resources in the near future, such as vaccines, ventilators and items of necessity such as face masks to provision certain imposed restrictions for a variety of nations, with the ability to prioritise nations that may be underdeveloped or less developed than others.

# Data and Context

Data provided as a baseline for this investigation includes the Our World in Data COVID Dataset. This was extracted directly from the GitHub repository, where the data is updated daily with the first date updated in March 2020. This also includes an amalgamation of data from John Hopkins University, the UN, the WHO and various provided from the OWID team based off worldwide government sources.

Furthermore, data provided by the University of Oxford regarding the stringency and actions of governments has been used, which details certain attributes such as stay-at-home orders, mask mandates, vaccination mandates and similar stringency measures. This was also extracted from GitHub, more specifically from the repository of Oxford.

## Data Cleaning

### OWID Covid Dataset

```
# OWID Covid Data Path
owidURL = "https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv"

# Read COVID Data
covidData = as.data.frame(read_csv(owidURL))

# Identify Dimensions of Covid Data
dim(covidData)
```

```
## [1] 190220     67
```

The Our World in Data contains 67 observable variables of interest, with the number of rows increasing daily as new data is added. Each observation contains the `date`, `location` and `iso_code` of each observation, with the relevant metrics of case numbers, vaccinations, stringency index and HDI. The format of this data can be found below.

```
# Convert date column to actual date values
covidData$date = as.Date(covidData$date, "%Y-%m-%d")
# Show top observations of dataset
head(covidData)
```

```
##   iso_code continent    location       date total_cases new_cases
## 1      AFG      Asia Afghanistan 2020-02-24           5         5
## 2      AFG      Asia Afghanistan 2020-02-25           5         0
## 3      AFG      Asia Afghanistan 2020-02-26           5         0
## 4      AFG      Asia Afghanistan 2020-02-27           5         0
## 5      AFG      Asia Afghanistan 2020-02-28           5         0
## 6      AFG      Asia Afghanistan 2020-02-29           5         0
##   new_cases_smoothed total_deaths new_deaths new_deaths_smoothed
## 1                 NA           NA         NA                  NA
## 2                 NA           NA         NA                  NA
## 3                 NA           NA         NA                  NA
## 4                 NA           NA         NA                  NA
## 5                 NA           NA         NA                  NA
## 6              0.714           NA         NA                  NA
##   total_cases_per_million new_cases_per_million new_cases_smoothed_per_million
## 1                   0.126                 0.126                             NA
## 2                   0.126                 0.000                             NA
## 3                   0.126                 0.000                             NA
## 4                   0.126                 0.000                             NA
## 5                   0.126                 0.000                             NA
## 6                   0.126                 0.000                          0.018
##   total_deaths_per_million new_deaths_per_million
## 1                       NA                     NA
```

```
## 2                       NA                     NA
## 3                       NA                     NA
## 4                       NA                     NA
## 5                       NA                     NA
## 6                       NA                     NA
##   new_deaths_smoothed_per_million reproduction_rate icu_patients
## 1                              NA                NA           NA
## 2                              NA                NA           NA
## 3                              NA                NA           NA
## 4                              NA                NA           NA
## 5                              NA                NA           NA
## 6                              NA                NA           NA
##   icu_patients_per_million hosp_patients hosp_patients_per_million
## 1                       NA            NA                        NA
## 2                       NA            NA                        NA
## 3                       NA            NA                        NA
## 4                       NA            NA                        NA
## 5                       NA            NA                        NA
## 6                       NA            NA                        NA
##   weekly_icu_admissions weekly_icu_admissions_per_million
## 1                    NA                                NA
## 2                    NA                                NA
## 3                    NA                                NA
## 4                    NA                                NA
## 5                    NA                                NA
## 6                    NA                                NA
##   weekly_hosp_admissions weekly_hosp_admissions_per_million total_tests
## 1                     NA                                 NA          NA
## 2                     NA                                 NA          NA
## 3                     NA                                 NA          NA
## 4                     NA                                 NA          NA
## 5                     NA                                 NA          NA
## 6                     NA                                 NA          NA
##   new_tests total_tests_per_thousand new_tests_per_thousand new_tests_smoothed
## 1        NA                       NA                     NA                 NA
## 2        NA                       NA                     NA                 NA
## 3        NA                       NA                     NA                 NA
## 4        NA                       NA                     NA                 NA
## 5        NA                       NA                     NA                 NA
## 6        NA                       NA                     NA                 NA
##   new_tests_smoothed_per_thousand positive_rate tests_per_case tests_units
## 1                              NA            NA             NA        <NA>
## 2                              NA            NA             NA        <NA>
## 3                              NA            NA             NA        <NA>
## 4                              NA            NA             NA        <NA>
## 5                              NA            NA             NA        <NA>
## 6                              NA            NA             NA        <NA>
##   total_vaccinations people_vaccinated people_fully_vaccinated total_boosters
## 1                 NA                NA                      NA             NA
## 2                 NA                NA                      NA             NA
## 3                 NA                NA                      NA             NA
## 4                 NA                NA                      NA             NA
## 5                 NA                NA                      NA             NA
## 6                 NA                NA                      NA             NA
```

```
##   new_vaccinations new_vaccinations_smoothed total_vaccinations_per_hundred
## 1               NA                        NA                             NA
## 2               NA                        NA                             NA
## 3               NA                        NA                             NA
## 4               NA                        NA                             NA
## 5               NA                        NA                             NA
## 6               NA                        NA                             NA
##   people_vaccinated_per_hundred people_fully_vaccinated_per_hundred
## 1                            NA                                  NA
## 2                            NA                                  NA
## 3                            NA                                  NA
## 4                            NA                                  NA
## 5                            NA                                  NA
## 6                            NA                                  NA
##   total_boosters_per_hundred new_vaccinations_smoothed_per_million
## 1                         NA                                    NA
## 2                         NA                                    NA
## 3                         NA                                    NA
## 4                         NA                                    NA
## 5                         NA                                    NA
## 6                         NA                                    NA
##   new_people_vaccinated_smoothed new_people_vaccinated_smoothed_per_hundred
## 1                             NA                                         NA
## 2                             NA                                         NA
## 3                             NA                                         NA
## 4                             NA                                         NA
## 5                             NA                                         NA
## 6                             NA                                         NA
##   stringency_index population population_density median_age aged_65_older
## 1             8.33   39835428             54.422       18.6         2.581
## 2             8.33   39835428             54.422       18.6         2.581
## 3             8.33   39835428             54.422       18.6         2.581
## 4             8.33   39835428             54.422       18.6         2.581
## 5             8.33   39835428             54.422       18.6         2.581
## 6             8.33   39835428             54.422       18.6         2.581
##   aged_70_older gdp_per_capita extreme_poverty cardiovasc_death_rate
## 1         1.337       1803.987              NA               597.029
## 2         1.337       1803.987              NA               597.029
## 3         1.337       1803.987              NA               597.029
## 4         1.337       1803.987              NA               597.029
## 5         1.337       1803.987              NA               597.029
## 6         1.337       1803.987              NA               597.029
##   diabetes_prevalence female_smokers male_smokers handwashing_facilities
## 1                9.59             NA           NA                 37.746
## 2                9.59             NA           NA                 37.746
## 3                9.59             NA           NA                 37.746
## 4                9.59             NA           NA                 37.746
## 5                9.59             NA           NA                 37.746
## 6                9.59             NA           NA                 37.746
##   hospital_beds_per_thousand life_expectancy human_development_index
## 1                        0.5           64.83                   0.511
## 2                        0.5           64.83                   0.511
## 3                        0.5           64.83                   0.511
## 4                        0.5           64.83                   0.511
```

```
## 5                        0.5          64.83                    0.511
## 6                        0.5          64.83                    0.511
##   excess_mortality_cumulative_absolute excess_mortality_cumulative
## 1                                   NA                          NA
## 2                                   NA                          NA
## 3                                   NA                          NA
## 4                                   NA                          NA
## 5                                   NA                          NA
## 6                                   NA                          NA
##   excess_mortality excess_mortality_cumulative_per_million
## 1               NA                                      NA
## 2               NA                                      NA
## 3               NA                                      NA
## 4               NA                                      NA
## 5               NA                                      NA
## 6               NA                                      NA
```

Some processing will be required to ensure that data in both datasets are identical to ensure joining. By using the OWID data as the baseline, this ensures that the Oxford data can be processed to follow the same format to ensure the inner join is computed correctly.

### Oxford Covid Government Response Tracker

The Oxford CGRT data provided includes multiple files of data for each location. It was decided that an inner join was to be computed using the `iso_code` from the OWID data, and `country_code` in the Oxford datasets. Furthermore, the dates represented as columns in the Oxford data would have to be transformed into columns to ensure that joining would be flawless.

```
# Create string array of all files to load in from Oxford Dataset
oxSuffixes = c("c1_school_closing.csv","c2_workplace_closing.csv","c3_cancel_public_events.csv",
               "c4_restrictions_on_gatherings.csv","c5_close_public_transport.csv",
               "c6_stay_at_home_requirements.csv","c7_movementrestrictions.csv","c8_internationaltravel
               "e1_income_support.csv","e2_debtrelief.csv","h1_public_information_campaigns.csv",
               "h2_testing_policy.csv", "h3_contact_tracing.csv", "h6_facial_coverings.csv",
               "h7_vaccination_policy.csv", "h8_protection_of_elderly_people.csv")

# Identify location of Oxford Data on Github so that the file suffixes above can be iterated through
prefixURL = "https://raw.githubusercontent.com/OxCGRT/covid-policy-tracker/master/data/timeseries"
```

The code above sources the Oxford datasets directly from its GitHub repository, and a loop iterates through each entry to join it to the OWID dataset.

### Data Join

To join the data, the two variables decided to have the inner join computed on included `iso_code` and `date`. For this to occur, the processing of the Oxford data had to be computed within the loop. These processes included dropping parts of the file name to create the variable name, in addition to extracting columns from the dataset due to it being unnecessary.

```
vec = c()
#
```

```r
for(i in 1:length(oxSuffixes)){
  # Concatenate the link to create the URL to load data
  link = paste(prefixURL,oxSuffixes[i],sep = "/")
  # Convert concatenated strings into URL
  dataURL = url(link)
  # Load data using read_csv
  oxData = read_csv(dataURL)
  # Drop specific parts of the file name
  stringSub = str_sub(oxSuffixes[i],4,-5)
  # Create File Name
  varName = append(vec,stringSub)
  # Extract from all rows of data, the column `...1` and `country_name`
  oxData = subset(oxData, select = -c(...1,country_name))

  # Pivot Oxford Data to the same format as the OWID Data
  mergeData = oxData %>%
    pivot_longer(!country_code, names_to = "date", values_to = stringSub)
  # Amend the date values to ensure that it is in the same format
  mergeData$date <- as.Date(mergeData$date,"%d%b%Y")
  # Convert variable names to be identical to the OWID Data for merging
  mergeData = rename(mergeData,iso_code = country_code)
  # Compute a join of the data, by `iso_code` followed by `date`. Disregard values that do not sit with
  covidData = merge(x = covidData, y = mergeData, by = c("iso_code","date"))
}
```

Further renaming of columns was required, as well as processing of the date values from strings to dates to ensure that joining was computed without issues and properly. `iso_code` was selected as a merge variable due to locations having potentially different names in the data listings, notably locations with suffixes or prefixes such as "Republic of", "Kingdom of" or "S.A.R." etc. As such, using the `iso_code` eliminates the potential effect that these suffixes and prefixes have on the overall data. The date is used in the join to compute the correct values of the government response for the specific location on a specific day. As such, once the data is joined the following dimensions can be found.

```r
dim(covidData)
```

```
## [1] 150657     83
```

There are now 83 variables observed, which adds on the 16 different Oxford datasets available, which includes one observation per dataset.

## Data Clustering

In order to answer the overarching question of how HDI impacts government response, and identifying the ideal response combination for different groups of locations. It was found that six clusters led to an ideal grouping system which provided logical results, where the groups were of a decent size and comparable.

HDI refers to Human Development Index, and it indicates the development of a location based on human capability rather than based on economic growth. As such, it is a more realistic indicator for a country's development when compared to GDP per capita. The HDI is calculated based on health and educational data, in addition to economic data. This includes 4 core metrics, GNI per capita, mean and expected years of schooling and life expectancy at birth. These together combine to form the HDI, which is a considerably stronger metric to judge a location's development compared to soley economic output.

The clustering processes chosen was hierarchical clustering, which was found to contain six clusters of a logical nature..

# Processes

### Freedom Score

The self-derived freedom score was used as a metric to analyse the freedoms that certain locations were afforded during the pandemic, with a score of 100 depicting a restriction-free stance from the government, and the opposite having a score of 0. Furthermore, the Oxford stringency score takes these restrictions with equal weighting values, however it was felt that some restrictions bore more weight on how free or not a society could be deemed, and as such a reclassification of weights was deemed necessary. The weightings for this new freedom score was computed through incremental weighting, in addition to the use of small-scale survey data to determine an aggregate score to correctly determine the weight values. By aggregating the scores and normalising through the maximum aggregated score, the new freedom score was computed and displayed from a scale of 0-100. This process can be expressed in the equation below.

$$FS = 100 \times \left(1 - \left(\frac{\sum_1^n W_n \times AS_n}{\sum_1^n W_n \times AS_{n_{max}}}\right)\right)$$

$AS$ refers to the attribute score, and $W$ refers to the relevant weighting for each attribute considered for the freedom score.

### Performance Index

In similar fashion to the freedom score, a performance index was crucial to be developed to ensure that a metric exists to compare locations in their responses to the pandemic. These performance values were explored through the use of normalised data, provided in the form of new cases and new deaths per million of the population. The values were scaled with deaths being slightly heavier than cases, due to deaths resulting in loss of human life, but reduces infectivity. These combined values were then normalised and transformed using a logarithmic transformation, to introduce greater disparity due to the data being left-skewed. This disparity allows for greater separation of the higher-performing locations for better visualisation and understanding. The equations to depcit this score calculation can be found below:

$$PSC = \log(nCasePM + 1.25 \times nDeathPM)$$

PSC is the calculated performance score calculated to determine the normalising factors to fully compute the final performance score, with new cases per million and new deaths per million included.

The corresponding maximum and minimum calculated PS values were collated to use for normalisation and were denoted as $PSC_{min}$ and $PSC_{max}$. The formula for the final performance score can be found below.

$$PS_{final} = 100 \times \left(1 - \frac{(\log(nCPM + 1.25 \times nDPM) + |PSCmin|}{PSC_{max} - PSC_{min}}\right)$$

Noting that the logarithm of 0 is not obtainable, this value is manually decoded to equal 100, as there were no recorded cases for that day and as such the performance is ideal.

### Overall Index

The overall index was a summation of the freedom score and the performance index, with the performance index given preferential weighting as it could be defined as a greater contributing factor towards a location's performance. However, it shouldn't be retracted that freedom doesn't affect performance.

$$Overall = FS + 1.25 \times PS$$

# Results

```
Because our performance score is composed of death cases and new cases, and our freedom score is compose
```

```
library(lattice)
```

```
## Warning: package 'lattice' was built under R version 4.1.1
```

```
library(ggplot2)
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.1.3
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(rpart)
library(C50)
```

```
## Warning: package 'C50' was built under R version 4.1.3
```

```
data=read.csv("covidC2DataFinal.csv")
set.seed(3888)

# repeated vc
ind = sample(2,nrow(data),replace = TRUE,prob = c(0.7,0.3))
trainset = data[ind == 1,]
testset = data[ind == 2,]
control = trainControl(method = "repeatedcv",number = 10,repeats = 3)
model = train(performanceScore~school_closing+workplace_closing+cancel_public_events+restrictions_on_ga
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :
## There were missing values in resampled performance measures.
```

```
importance = varImp(model,scale = FALSE)
maxC1 = max(data$school_closing,na.rm = TRUE)
maxC2 = max(data$workplace_closing,na.rm = TRUE)
maxC3 = max(data$cancel_public_events,na.rm = TRUE)
maxC4 = max(data$restrictions_on_gatherings,na.rm = TRUE)
maxC5 = max(data$close_public_transport,na.rm = TRUE)
maxC6 = max(data$stay_at_home_requirements,na.rm = TRUE)
maxC7 = max(data$movementrestrictions,na.rm = TRUE)
maxC8 = max(data$internationaltravel,na.rm = TRUE)
maxtestScore1 = 14*importance$importance[2,]/sum(importance$importance)*maxC8 + 14*importance$importanc
+ 14*importance$importance[6,]/sum(importance$importance) * maxC2 + 14*importance$importance[1,]/sum(im
```

```
## [1] 30.43187
```

```
+ 14*importance$importance[7,] /sum(importance$importance)* maxC7 + 14*importance$importance[5,] /sum(i
```

```
## [1] 2.463058
```

```
Testscore1=(1 -
              (14*importance$importance[4,]/sum(importance$importance) * data$school_closing + 14*impor
               + 14*importance$importance[1,]/sum(importance$importance) * data$cancel_public_events+ 14
               + 14*importance$importance[8,]/sum(importance$importance) * data$close_public_transport +
               + 14*importance$importance[7,] /sum(importance$importance)* data$movementrestrictions + 





# logistic regression
logistic_function=glm(performanceScore~school_closing+workplace_closing+cancel_public_events+restrictio
summary(logistic_function)
```

```
##
## Call:
## glm(formula = performanceScore ~ school_closing + workplace_closing +
##     cancel_public_events + restrictions_on_gatherings + close_public_transport +
##     stay_at_home_requirements + movementrestrictions + internationaltravel,
##     data = trainset, na.action = na.omit)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -65.448  -16.346   -5.173   10.372   61.583
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                68.01892    0.23714 286.831  < 2e-16 ***
## school_closing             -1.83008    0.09598 -19.067  < 2e-16 ***
## workplace_closing          -5.01425    0.12195 -41.117  < 2e-16 ***
## cancel_public_events       -2.07995    0.16499 -12.606  < 2e-16 ***
## restrictions_on_gatherings -2.31970    0.07628 -30.409  < 2e-16 ***
## close_public_transport      0.25277    0.13682   1.847   0.0647 .
```

```
## stay_at_home_requirements   1.48727    0.11078  13.425  < 2e-16 ***
## movementrestrictions        0.54938    0.11523   4.768 1.86e-06 ***
## internationaltravel         0.72061    0.07815   9.221  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 558.7121)
##
##     Null deviance: 55008908  on 88643  degrees of freedom
## Residual deviance: 49521449  on 88635  degrees of freedom
##   (8739 observations deleted due to missingness)
## AIC: 812301
##
## Number of Fisher Scoring iterations: 2
```

```
importance_logistic=varImp(logistic_function,scale = FALSE)
maxtestScore2 = 14*importance_logistic[8,]/sum(importance_logistic)*maxC8 + 14*importance_logistic[1,]/s
+ 14*importance_logistic[2,]/sum(importance_logistic) * maxC2 + 14*importance_logistic[3,]/sum(importan
```

```
## [1] 28.9484
```

```
+ 14*importance_logistic[7,] /sum(importance_logistic)* maxC7 + 14*importance_logistic[6,] /sum(importa
```

```
## [1] 5.26456
```

```
Testscore2=(1 -
              (14*importance_logistic[1,]/sum(importance_logistic) * data$school_closing + 14*importance
               + 14*importance_logistic[3,]/sum(importance_logistic) * data$cancel_public_events+ 14*imp
               + 14*importance_logistic[5,]/sum(importance_logistic) * data$close_public_transport + 14*
               + 14*importance_logistic[7,] /sum(importance_logistic)* data$movementrestrictions + 14*im




# Random forest
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 4.1.2
```

```
## randomForest 4.7-1
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
rf<-randomForest(performanceScore~school_closing+workplace_closing
                 +cancel_public_events+restrictions_on_gatherings
                 +close_public_transport+stay_at_home_requirements
                 +movementrestrictions+internationaltravel, data=trainset, mtry=7, ntree=100, importance
rf_sum=sum(rf$importance[1:8])
maxtestScore3 = 14*rf$importance[8]/rf_sum*maxC8
+ 14*rf$importance[1]/rf_sum * maxC1
```

```
## [1] 5.869542
```

```
+ 14*rf$importance[2]/rf_sum* maxC2 + 14*rf$importance[3]/rf_sum* maxC3
```

```
## [1] 9.334891
```

```
+ 14*rf$importance[4]/rf_sum* maxC4 +14*rf$importance[5]/rf_sum* maxC5
```

```
## [1] 11.96109
```

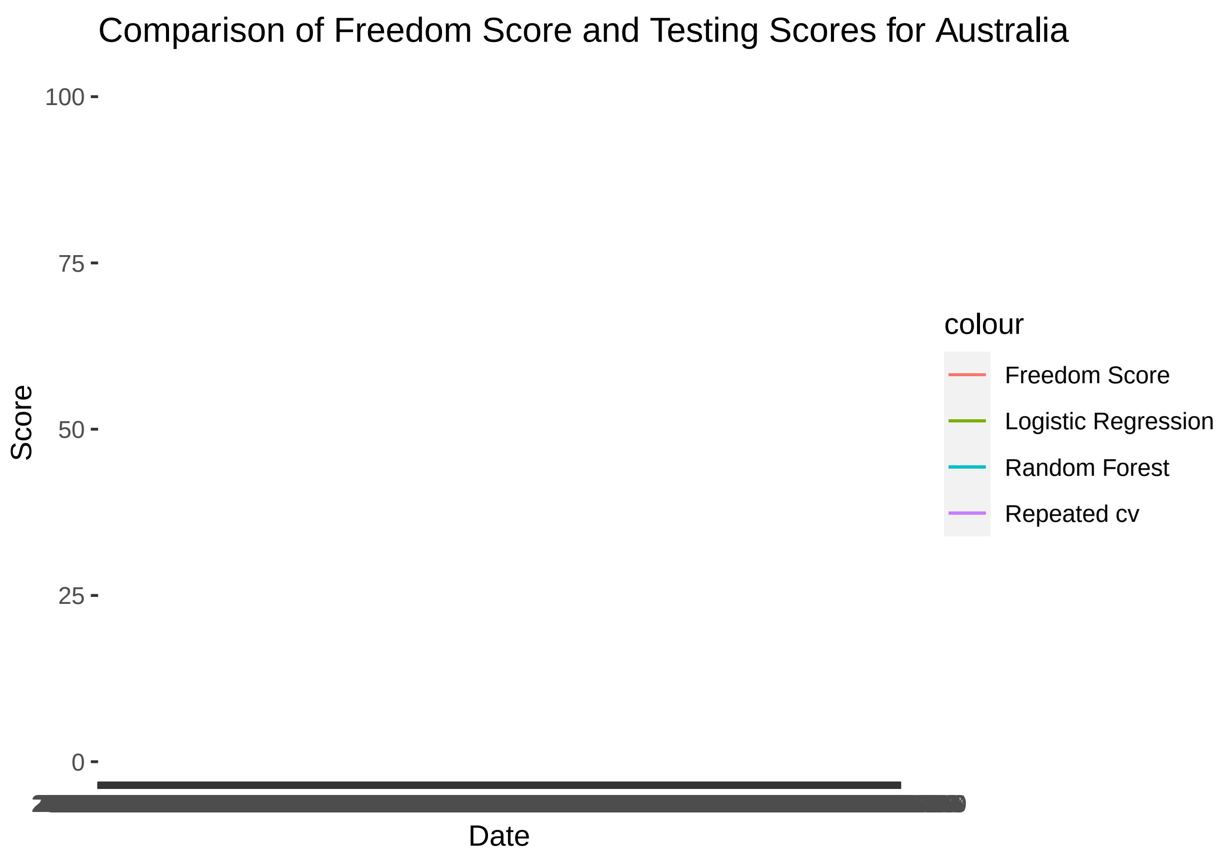
```
+ 14*rf$importance[7]/rf_sum* maxC7 + 14*rf$importance[6]/rf_sum* maxC6
```

```
## [1] 7.50787
```

```
Testscore3=(1 -
              (14*rf$importance[1]/rf_sum* data$school_closing + 14*rf$importance[2]/rf_sum* data$workp
               + 14*rf$importance[3]/rf_sum* data$cancel_public_events
               + 14*rf$importance[4]/rf_sum* data$restrictions_on_gatherings
               + 14*rf$importance[5]/rf_sum* data$close_public_transport + 14*rf$importance[6]/rf_sum* d
               + 14*rf$importance[7]/rf_sum* data$movementrestrictions+ 14*rf$importance[8]/rf_sum*data$

# normalization using Linear transformation(because of the existence of negative numbers, log cannot be
t1=(Testscore1-min(Testscore1,na.rm = TRUE))/(max(Testscore1,na.rm = TRUE)-min(Testscore1,na.rm = TRUE)
t2=(Testscore2-min(Testscore2,na.rm = TRUE))/(max(Testscore2,na.rm = TRUE)-min(Testscore2,na.rm = TRUE)
t3=(Testscore3-min(Testscore3,na.rm = TRUE))/(max(Testscore3,na.rm = TRUE)-min(Testscore3,na.rm = TRUE)

data0=data.frame(data$date,data$location,data$freeScore,t1,t2,t3)
data1=data0[data0$data.location=="Australia",]
data2=data0[data0$data.location=="Canada",]
data3=data0[data0$data.location=="United States",]

p1<-ggplot(data1, aes(x=data.date))+geom_line(aes(y=data.freeScore, color="Freedom Score"))+geom_line(a
p2<-ggplot(data2, aes(x=data.date))+geom_line(aes(y=data.freeScore, color="Freedom Score"))+geom_line(a
p3<-ggplot(data3, aes(x=data.date))+geom_line(aes(y=data.freeScore, color="Freedom Score"))+geom_line(a
p1
```

```
## Warning: Removed 8 row(s) containing missing values (geom_path).
```

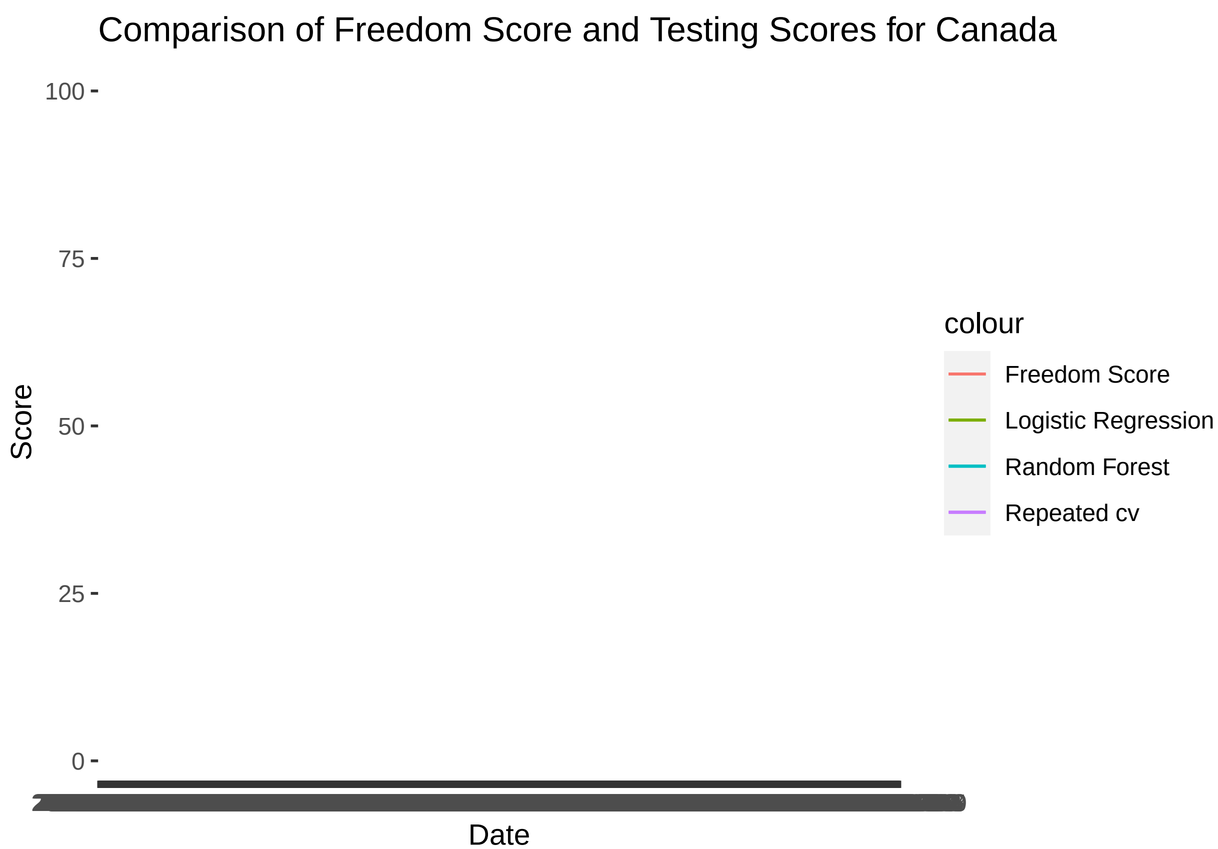
```
## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?

## Warning: Removed 8 row(s) containing missing values (geom_path).

## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?

## Warning: Removed 8 row(s) containing missing values (geom_path).

## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?

## Warning: Removed 8 row(s) containing missing values (geom_path).

## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```




```
p2
```

```
## Warning: Removed 3 row(s) containing missing values (geom_path).

## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```

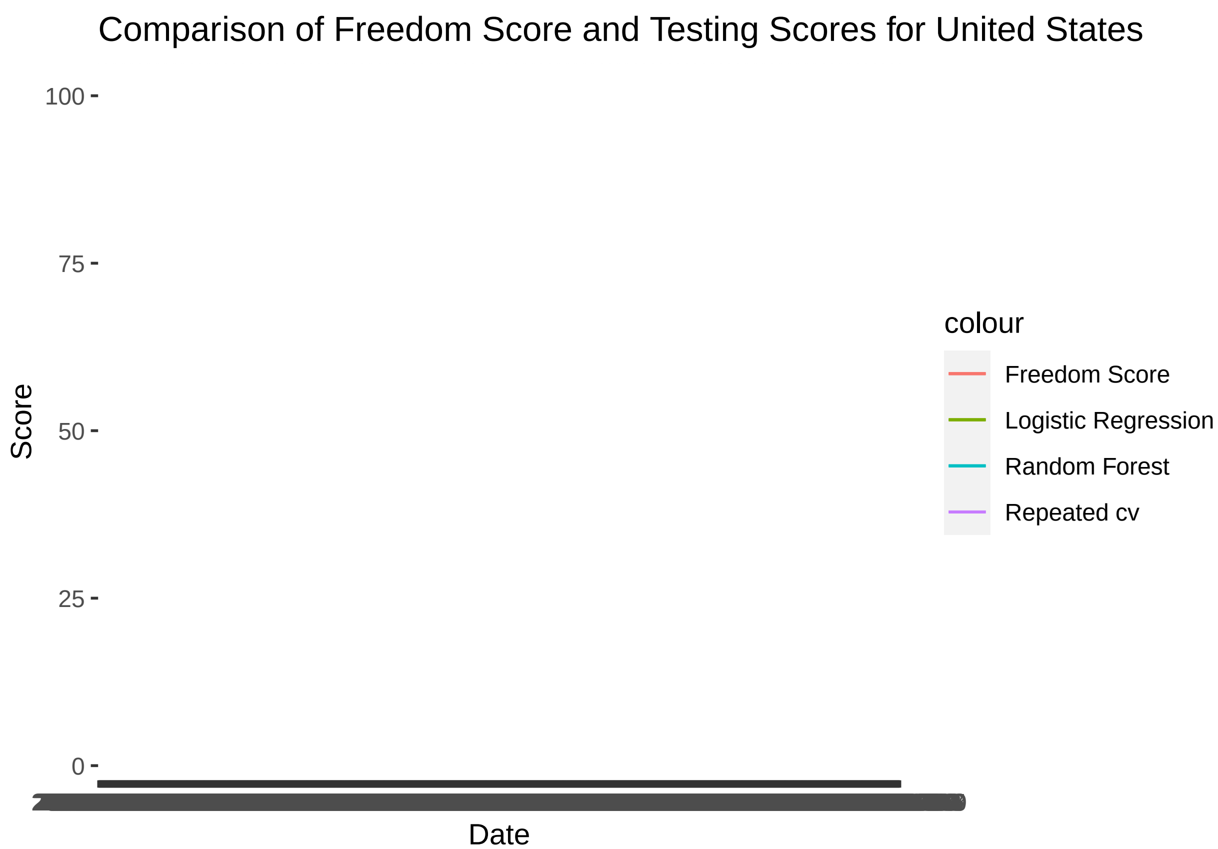
```
## Warning: Removed 3 row(s) containing missing values (geom_path).

## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?

## Warning: Removed 3 row(s) containing missing values (geom_path).

## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?

## Warning: Removed 3 row(s) containing missing values (geom_path).

## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```




```
p3
```

```
## Warning: Removed 24 row(s) containing missing values (geom_path).

## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?

## Warning: Removed 24 row(s) containing missing values (geom_path).
```

```
## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?

## Warning: Removed 24 row(s) containing missing values (geom_path).

## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?

## Warning: Removed 24 row(s) containing missing values (geom_path).

## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```




```
The above pictures take Australia, New Zealand and the United States as examples. According to the compa
```

### shiny app

This shiny app has a total of 5 pages, namely home, methodology, cluster, prediction and about page. In the home, methodology and about pages, users can learn most of the information about this shiny app, such as: the purpose of this project, the source of data, the analysis methods used in this project, etc.

Most of the functions of this shiny app are concentrated on the cluster and prediction pages, which are the core of the entire project as well. The Cluster page is divided into two sections, an input section on the left and an output section on the right. The user can select the interested country and the cluster corresponding to this country on the left, as well as a specific period that the user is interested in. After the user makes a selection, the output section on the right would display information related to the selected country. The first picture on the right will highlight all the countries included in the selected cluster in red on the world map, so that the user has a clearer understanding of the selected cluster. The second picture will show a line graph of the score of the selected country in the selected time period. Below these two pictures, the user can see some information related to the selected cluster, such as the average score of the cluster, The best and worst two performing countries within this cluster, and the average score for the selected countries. Users can choose to compare the country A they are interested in with the country B that performs better in the same cluster, and through the comparison in a period which some B countries have significantly higher scores than A country, the different policies implemented by the two countries can be used to find out the direction of progress of country A's policy. Moreover, In the prediction page, the government can input the population and the HDI of the country, and the Shiny app will give the score that the country should achieve according to the input data. Comparing this score with the actual score, the government can know whether they are More policies and work are needed to make their country perform better.

Although in this project, we did not cooperate with students from other disciplines, it is still a manifestation of multidisciplinary cooperation. We did not directly display the content of the data. By using the HDI which is a concept in commerce, we divided all countries into six parts according to the level of HDI. In this way, users can choose to compare only countries that have similar HDI, which makes the comparison more meaningful and makes the policies of excellent countries more feasible in their own countries.

# Discussion

# !need to add a graph of New Zealand's overall score there!

```
According to the graph of our overall score, we can find the changes in the realization of each country
In August 2020, the covid boomed again in New Zealand. This time, the New Zealand government did not tal
In mid-August 2021, the third covid epidemic happened in New Zealand. This time, the New Zealand governm
```

### Areas for Improvement

Upon marker's comments in the Presentation and Demonstration the product's following areas were questioned to improve the product when looking retrospectively and in the future:

1. Consider the different methods of what counts as a case and does not. How can we account for reporting inaccuracies for each location. To address this area, each of the 174 countries would need to have their COVID-19 case methodologies examined. Therefore, a smaller subset of locations would need to be considered as trying to accurately account for this would be unfeasible. A smaller subset method could include Western Europe as a geographic region, or more specifically the European Union. For example, the European Centre for Disease Prevention and Control (ECDC) defines a COVID-19 case a one containing at least a single symptom of cough, fever, shortness of breath, sudden onset of anosmia, ageusia or dysgeusia with detection of SARS-CoV-2 nucleic acid. This would add a layer of consistency regarding case numbers if the ECDC dataset was used as opposed to the OWID dataset.
2. Consider the different strains of COVID-19 when the Performance Scores were measured and how this is impacted with time across the different waves. Of course different strains should influence the requirements of citizens regarding social distancing, contact tracing and vaccinations. For instance the Omicron variant is more contagious than Delta and thus some alteration in the performance / freedom score is needed. One method that could be considered for future work is that Omicron cases should have less of an influence on freedom score due to its severity such that it does not decrease the total score. In essence, greater restrictions on movement due to increased Omicron cases should indeed increase the total score rather than decreasing it. In theory this sounds reasonable, however, cases are not reported according to strain but rather just by whether a person has COVID-19 or not.

One area that was considered during the early stage of the product was to have some measure of the minimum requirements needed to yield a user's inputted score given that the user chose a location. I.e. what combination of social distancing measures and vaccinations are needed to attain a score of 50. What this would aim to achieve is finding the appropriate balance between freedom and performance to achieve a total score. For instance a lack of freedom could cause civil unrest despite the measures being ineffective in the performance score. This idea was scrapped due to the complexity of it for instance given the equation from the overall index there are infinitely many solutions to attaining an overall/total score of 50 and it would be difficult to have a generic solution.

In retrospect another area that could be improved is to have a weekly data point on the graph rather than a daily data point. What this should do is smooth out some of the total scores observed per location in the app. When certain locations were analysed due to the style of reporting cases, the total score line tended to zig zag along the graph. Reasons behind this are reporting dates, on certain dates cases might not be entered in giving the false illusion that there are no cases for that day when there are. The trade-off between the weekly and daily method is that for those locations which already had accurate daily reporting this would reduce the accuracy of the total score whereas those with inaccurate data reporting would have a more accurate total score. Furthermore, given that the policy responses are daily, some "average" policy response would need to be made for the week despite policy being changed per day e.g. on Monday facial coverings are mandatory, then on Tuesday it is no longer.

# Conclusion

# References

European Centre for Disease Prevention and Control. (2020). *Case definition for coronavirus disease 2019 (COVID-19), as of 3 December 2020.* Retrieved from https://www.ecdc.europa.eu/en/covid-19/surveillance/case-definition

Lowry Institute. (2021). Covid Performance Index DECONSTRUCTING PANDEMIC RESPONSES What impact have geography, political systems, population size, and economic development had on COVID-19 outcomes around the world?. Retrieved from https://interactives.lowyinstitute.org/features/covid-performance/?fbclid=IwAR2KzlVzXeqneE8w3WZrm3uIdTgpEXd7v1JDPL7Iv7H_MrPi306kmK2mA34

Bloomberg. (2022). The Covid Resilience Ranking The Best and Worst Places to Be as Covid Flare-Ups Break Records. Retrieved from https://www.bloomberg.com/graphics/covid-resilience-ranking/?fbclid=IwAR2GOm-AKEcl0JSQ3KB3fic9x3Ft1cQQ0yAKxR6kwfEHBrfAwhcO8kZfpkI

Our World in Data. (2022). covid-19-data. Retrieved from https://github.com/owid/covid-19-data

University of Oxford. (2022). Oxford Covid-19 Government Response Tracker. Retrieved at https://github.com/OxCGRT