FIT3152 Data analytics

Assignment 1

Name: Lim Yu-Shan

Student ID: 32685467

Title: Analysis of country-level predictors of pro-social behaviours to reduce the spread of COVID-19 during the early stages of the pandemic

Notes to marker:

- The main body of this report is just over 14 pages, with some long code blocks and outputs taking up much of the length. All other pages are the appendix, which include repeated code and outputs.
- Some lines of the code output (significant predictors and coefficients for rest-of-the-world models) on page 9 are too long and flow off the page. The full list of significant predictors and coefficients have been manually copied and pasted below the original output, and can also be seen in the visualisations on pages 10 and 14.

Section 1

1(a)

The data in the file `PsyCoronaBaselineExtract.csv` is a reduced version of the data collected for the PsyCorona baseline study, a psychological survey investigating pro-social behaviours in different countries during the COVID-19 pandemic, by Van Lissa et al. (2002).

The following code is run to generate an individual subset of the data for my analysis. The data is then attached to the R search path for more convenient access to variables.

```
rm(list = ls())
set.seed(32685467)
cvbase <- read.csv("PsyCoronaBaselineExtract.csv")
cvbase <- cvbase[sample(nrow(cvbase), 40000), ]
attach(cvbase)
```

Important libraries to be used for the analysis is imported.

```
library(ggplot2)
library(dplyr)
library(tidyr)
```

To get a good initial understanding of the dataset, the following code is run to learn about its features and properties.

```
dim(cvbase)
as.data.frame(sapply(cvbase, class))  # get data types of each column
summary(cvbase, na.rm = TRUE)
```

From the first two outputs, we learn that my individual dataset has 40,000 rows/entries (as specified in my parameters for `sample()`) and 54 columns. All columns contain integer data except for `coded_country`, which contains character data (full strings of country names), making it the only text attribute.

Based on the codebook extract, all columns except `employstatus`, `gender`, `age`, `edu` and `coded_country` columns contain ordinal data in the form of integers that code for degrees such as level of agreement, age group and education level. The integer values of the `gender`, `age` and `edu` columns code for different gender, age and education categories respectively. Only a maximum of one `employstatus` column can have a value of 1 in each entry, denoting that that is the employment status for that individual.

From the output of `summary()`, we learn that the numerical attributes have varied ranges. Survey questions that measure a one-sided degree of agreement range from 1 to a higher positive number such as 5 and 6, while those that evaluate a two-sided degree of agreement range from a negative number to its modulus.

For the only text attribute, `coded_country`, running the following code

```
sort(unique(cvbase$coded_country))  # get all country names
table(cvbase$coded_country)          # get number of entries for each country
# get maximum and minimum number of entries, and their corresponding countries
max(table(cvbase$coded_country))
which(table(cvbase$coded_country) == max(table(cvbase$coded_country)))
max(table(cvbase$coded_country))
which(table(cvbase$coded_country) == min(table(cvbase$coded_country)))
```

reveals that there are 110 unique country names (including NA) in this dataset, and that each country has varied numbers of entries (the United States of America has the most with 6952, while 18 countries only have 1).

There are missing values in each column, though this is expected as each question in the survey is optional to answer. The `employstatus` columns have the most missing values among them as each participant only chooses one of 10 categories. For my dataset, `employstatus_3` has the fewest missing values whereas `employstatus_8` has the most. This implies that most of the participants are employed and working at least 40 hours per week, whereas the smallest minority in terms of employment status is disabled people.

One interesting observation is that the mean of the age groups in this dataset is 2.893, which means most participants are aged 35-44 years. This may be because most working-class adults with stable lifestyles fall into this category, and hence are studied more to better understand relationships between the pandemic and societal and job insecurity.

1(b)

No pre-processing is necessary as this dataset is tidy, with no faulty values or entries. The NA values in the `employstatus` columns, however, can be replaced with 0 as these columns are different answers to the same question, and the only other possible value being 1. This makes it easier for linear regression to be performed on these attributes later on. The head of `cvbase` is included in the **Appendix** to keep this report concise.

```
for (i in 21:30) {
  cvbase[, i][is.na(cvbase[, i])] <- 0
}
```

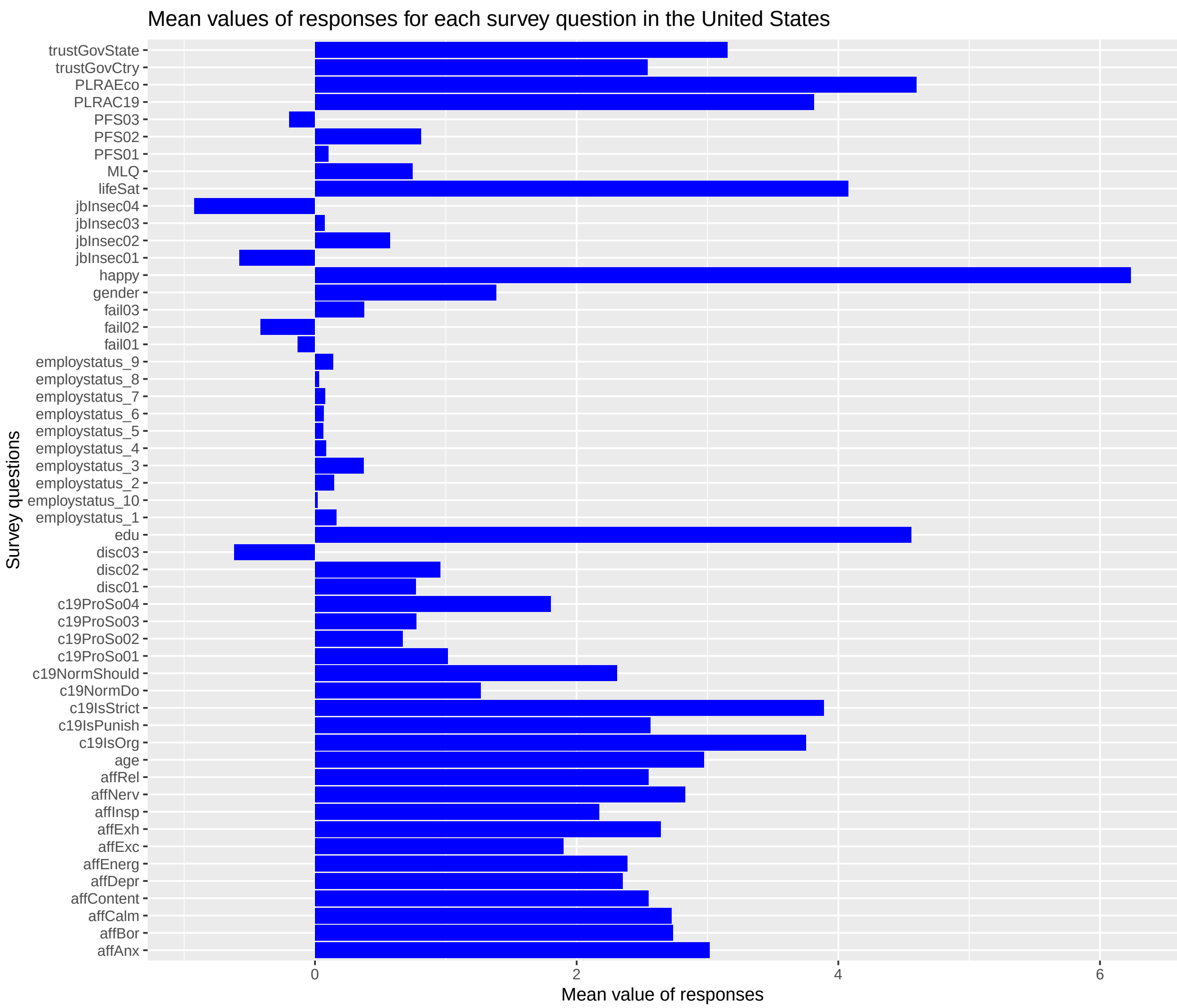
Section 2

2(a)

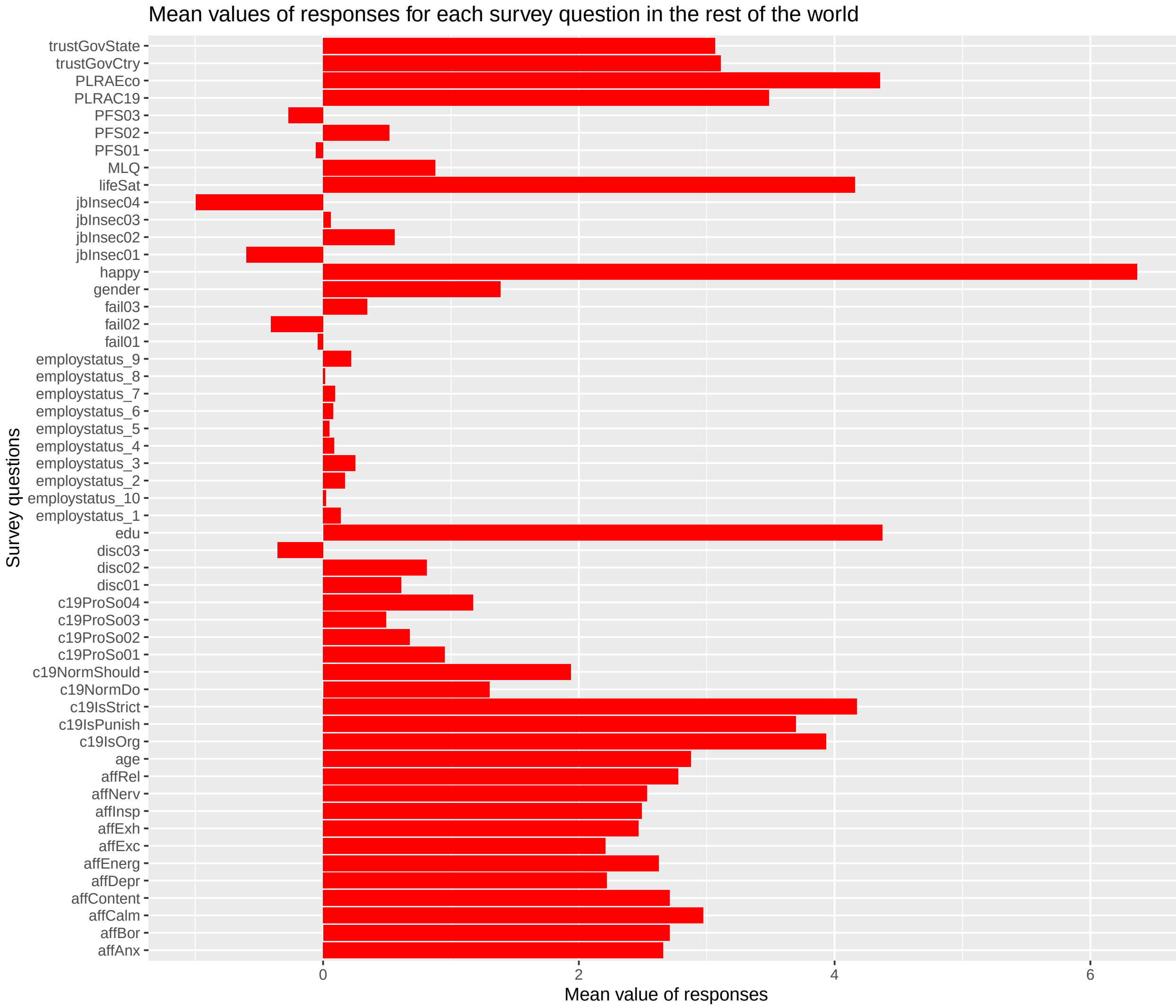
My focus country is the United States of America. To get a better view of how responses for the United States differ from other countries, bar charts are created for each group. The y-axis of each bar chart contains the survey question variables while the x-axis consists of the mean values of each question's responses. The following code creates the data frames for the mean values and plots the bar charts using `ggplot2`. `coded_country` is excluded as it is not a numerical attribute.

```
usa <- cvbase[coded_country == "United States of America", ]
rem <- anti_join(cvbase, usa)

means <- colMeans(usa[, !names(usa) %in% c("coded_country")], na.rm = TRUE)
usa_means <- data.frame(mean = means)

means <- colMeans(rem[, !names(rem) %in% c("coded_country")], na.rm = TRUE)
rem_means <- data.frame(mean = means)
```

```
usa_plot <- ggplot(usa_means) +
  geom_bar(mapping = aes(x = rownames(usa_means), y = mean), stat = "identity",
    fill = "blue") +
  coord_flip() +
  labs(x = "Survey questions", y = "Mean value of responses",
    title = "Mean values of responses for each survey question in the United States")

rem_plot <- ggplot(rem_means) +
  geom_bar(mapping = aes(x = rownames(rem_means), y = mean), stat = "identity",
    fill = "red") +
  coord_flip() +
  labs(x = "Survey questions", y = "Mean value of responses",
    title = "Mean values of responses for each survey question in the rest of the world")

usa_plot
```




```
rem_plot
```



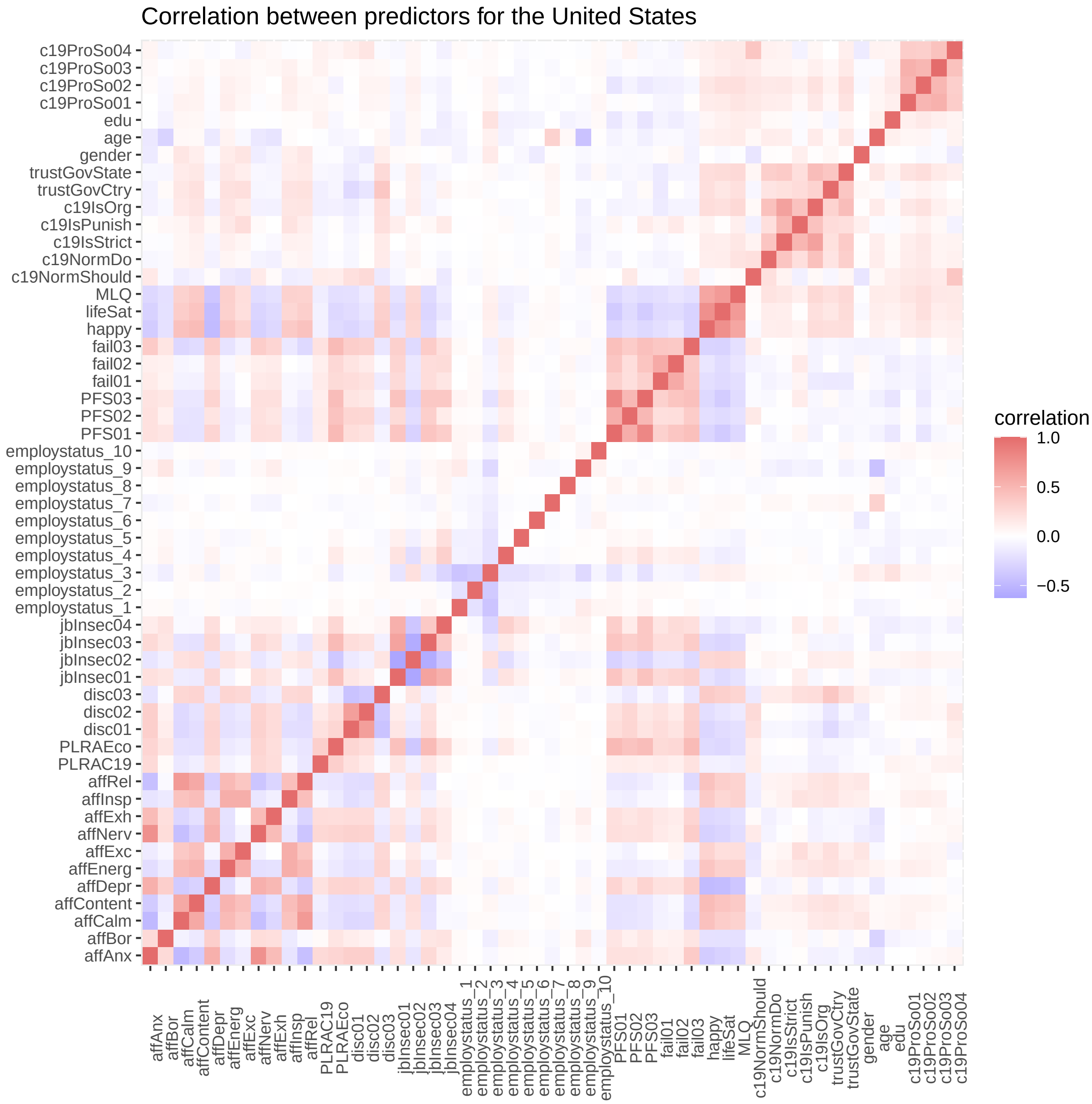
At first glance, participant responses in the United States is similar to that of the remaining countries as a group. The most notable difference is in the responses for `PFS01`. The possible responses for this question range from -2 (strongly disagree) to 2 (strongly agree). In the United States, the mean response is a positive number, while the same for other countries is a negative number, the only such difference among all variables.

2(b)

An initial look is taken at the correlation between predictor and pro-social attitude for the United States. The `cor()` function is used and the correlation matrix is visualised with a heatmap.

```r
usa_cor <- cor(subset(usa, select = -coded_country), use = "complete.obs")

# reshapes matrix to long format for plotting
usa_melted <- reshape2::melt(usa_cor)

usa_cor_plot <- ggplot(data = usa_melted) +
  geom_tile(mapping = aes(x = Var1, y = Var2, fill = value)) +
  scale_fill_gradient2(low = "#6b74ff", mid = "white", high = "#e46c6c", midpoint = 0) +
  labs(title = "Correlation between predictors for the United States", x = "", y = "",
    fill = "correlation") +
  theme(axis.text.x = element_text(angle = 90))
```

```
usa_cor_plot
```



Red and blue tiles indicate positive and negative correlation respectively, with tiles becoming white as correlation approaches 0. From this heatmap, there are many instances of strong correlation between predictors, but the subsection of the heatmap showing the correlation between pro-social attitudes and all other attributes is fairly light. This indicates that the attributes may not predict pro-social attitudes extremely well for the United States.

By fitting a linear regression model of each pro-social attitude against the attributes, we can see how well the responses predict the response to the pro-social attitude question. We will also be able to find out which predictors are the best.

The following code fits a linear model of each pro=social attitude against the attributes. A function and for loop is used to summarise each model, including their R-squared values, significant predictors with p-value less than 0.001, and their respective coefficients. **prds** and **mdl** are vectors that will be used to compare the strong predictors for each model in a table later.

```
prds <- NULL
mdl <- NULL

model_eval <- function(model) {
  rsqr <- summary(model)$r.squared
  a_rsqr <- summary(model)$adj.r.squared
  sig <- which(summary(model)$coefficients[-1, 4] < 0.001) + 1
  preds <- rownames(summary(model)$coefficients[sig, , drop = FALSE])
  coefs <- summary(model)$coefficients[sig, 1]

  return(list(rsqr, a_rsqr, preds, coefs))
}

fitted_usa1 <- lm(c19ProSo01 ~ .,
  data = subset(usa, select = -c(coded_country, c19ProSo02, c19ProSo03, c19ProSo04)))
fitted_usa2 <- lm(c19ProSo02 ~ .,
  data = subset(usa, select = -c(coded_country, c19ProSo01, c19ProSo03, c19ProSo04)))
fitted_usa3 <- lm(c19ProSo03 ~ .,
  data = subset(usa, select = -c(coded_country, c19ProSo01, c19ProSo02, c19ProSo04)))
fitted_usa4 <- lm(c19ProSo04 ~ .,
  data = subset(usa, select = -c(coded_country, c19ProSo01, c19ProSo02, c19ProSo03)))

cat("Summary of models for predicting pro-social attitudes in the United States\n\n")
```

```
## Summary of models for predicting pro-social attitudes in the United States
```

```
counter <- 1
for (model in list(fitted_usa1, fitted_usa2, fitted_usa3, fitted_usa4)) {
  cat("C19ProSo0", counter, "\n", sep = "")
  res <- model_eval(model)
  cat("R-squared value:", res[[1]], "\n")
  cat("Adjusted R-squared value:", res[[2]], "\n")
  cat("Significant predictors with p-value < 0.001:\n")
  cat(res[[3]], "\n")
  cat("Coefficients of predictors:\n")
  cat(res[[4]], "\n")
  cat("\n")
  for (pred in res[[3]]) {
    mdl <- c(mdl, paste0("USA C19ProSo0", counter))
  }
  prds <- c(prds, res[[3]])
  counter <- counter + 1
}
```

```
## C19ProSo01
## R-squared value: 0.1224774
## Adjusted R-squared value: 0.1089812
## Significant predictors with p-value < 0.001:
## disc02 MLQ c19NormShould trustGovState
## Coefficients of predictors:
## 0.1351951 0.08717715 0.1441734 0.1642502
##
## C19ProSo02
```

```
## R-squared value: 0.1674081
## Adjusted R-squared value: 0.154603
## Significant predictors with p-value < 0.001:
## disc02 PFS01 c19NormShould trustGovState edu
## Coefficients of predictors:
## 0.1542413 -0.1637085 0.1788572 0.1799418 0.08186903
##
## C19ProSo03
## R-squared value: 0.09581283
## Adjusted R-squared value: 0.08190663
## Significant predictors with p-value < 0.001:
## PLRAC19 MLQ c19NormShould trustGovState edu
## Coefficients of predictors:
## 0.08652849 0.1013942 0.1728413 0.156691 0.07075706
##
## C19ProSo04
## R-squared value: 0.2232844
## Adjusted R-squared value: 0.2113349
## Significant predictors with p-value < 0.001:
## disc02 MLQ c19NormShould c19IsPunish
## Coefficients of predictors:
## 0.1299587 0.06540818 0.3805587 -0.07275021
```

The responses best predict `C19ProSo04`, as evident from its adjusted R-squared value of 0.2113349, which is the highest among all models. Its best predictors are `disc02`, `MLQ`, `c19NormShould` and `c19IsPunish`. The model for `C19ProSo03` has the lowest adjusted R-squared value - 0.08190663 - with its best predictors being `PLRAC19`, `MLQ`, `c19NormShould`, `trustGovState` and `edu`.

The arguably small R-squared values among the models are unsurprising as most of the survey questions are subjective. For example, different participants perceive different levels of calmness differently, and interpret financial strain differently. As a vast and populous country with many working classes and standards of life, different parts of the United States are like separate countries on their own, with their own economies, healthcare and overall happiness. This makes it hard for the pro-social attitude responses to be predicted consistently.

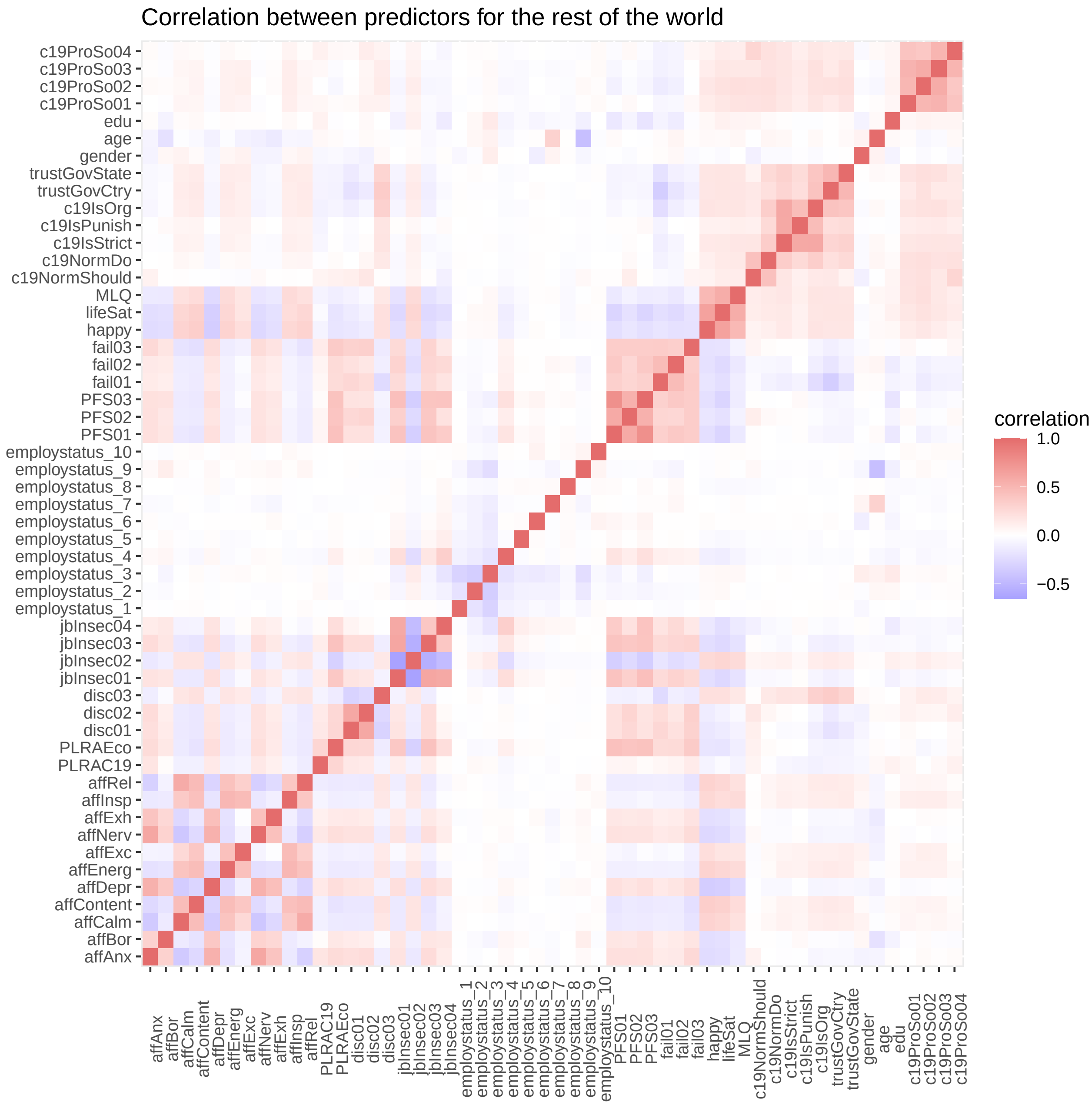
Each model has its own list of significant predictors, but some predictors can be considered more reliable overall as they appear more often across the models. The prime example would be `c19NormShould`, which is a strong predictor for all four models. This makes sense as someone who is willing to assist society during the pandemic would want the best for it, and thus encourage members of society to self-isolate and socially distance. These measures of curbing viral spread are suggested by the United States' own Centers for Disease Control and Prevention (CDC), and as a developed nation with a well-educated population, individuals with pro-social intentions tend to follow these guidelines. On the other hand, someone without pro-social attitudes would be indifferent towards societal behaviours and not be bothered to follow new norms. The predictive strength of `c19NormShould` may also be affected by individuals who think that social distancing is bad for society, and that they are helping others by opposing these measures. Protests against lockdowns were common in the United States during the pandemic, proving that this belief does exist.

Other variables that predict three of the models well are `disc02`, `MLQ` and `trustGovState`. Individuals would tend to be more pro-social based on their concern about the society's future, their sense of purpose in life, and their belief on whether they can find common ground with society in dealing with the pandemic.

2(c)

To repeat the same task for the rest of the world, previous code is reused, but with the `rem` dataset instead of `usa`. The correlation matrix for this dataset is first visualised. From this point onwards, variants of reused code will appear in the **Appendix** to keep this report concise.

```r
rem_cor_plot
```



Comparing both heatmaps we have thus far, we observe that `usa_cor_plot` has more darker-coloured tiles, indicating stronger correlation between predictors overall. In addition to having lighter tiles, `rem_cor_plot` looks "cleaner" with less scatter of coloured tiles. However, focusing on the subsections of the heatmaps that show the correlation between pro-social attitudes and all other attributes allows us to make an initial guess that the attributes for both groups of data should predict pro-social attitudes with roughly similar performance, as the subsections in both plots look fairly similar.

```
## Summary of models for predicting pro-social attitudes in the rest of the world
```

```
## C19ProSo01
## R-squared value: 0.1268007
## Adjusted R-squared value: 0.1240447
## Significant predictors with p-value < 0.001:
```

```
## affInsp PLRAC19 disc02 employstatus_10 fail03 lifeSat MLQ c19NormShould c19NormDo c19IsOrg trustGovSt
## Coefficients of predictors:
## 0.06154919 0.0652525 0.1005851 0.3293603 0.06135636 0.05886684 0.08799994 0.1034505 0.0727743 0.0577
##
## C19ProSo02
## R-squared value: 0.1684555
## Adjusted R-squared value: 0.1658316
## Significant predictors with p-value < 0.001:
## affAnx affBor affExc affExh affInsp PLRAEco disc02 disc03 jbInsec02 PFS01 fail01 lifeSat MLQ c19Norm
## Coefficients of predictors:
## 0.04854233 0.06002516 0.06926414 0.04935493 0.04810877 -0.03626116 0.1502406 0.0650781 0.06608606 -0
##
## C19ProSo03
## R-squared value: 0.1243751
## Adjusted R-squared value: 0.121612
## Significant predictors with p-value < 0.001:
## affExc affExh affInsp PLRAC19 disc02 disc03 employstatus_10 lifeSat MLQ c19NormShould c19NormDo c19I
## Coefficients of predictors:
## 0.05042581 0.04464075 0.06039361 0.07098512 0.1348678 0.07466859 0.3418363 0.09238179 0.05582841 0.0
##
## C19ProSo04
## R-squared value: 0.1445334
## Adjusted R-squared value: 0.1418332
## Significant predictors with p-value < 0.001:
## affInsp PLRAC19 disc02 disc03 jbInsec01 employstatus_10 PFS02 fail01 fail02 fail03 lifeSat c19NormSh
## Coefficients of predictors:
## 0.07070157 0.08621845 0.1716264 0.04722402 0.06153325 0.348768 0.04779941 -0.06649288 -0.05701471 0.0
```

Note: the lines for predictor names and their coefficients are too long and were cut off instead of wrapped when this PDF was knitted from my R Markdown file. The cut-off lines are, in order, as follows:

`affInsp PLRAC19 disc02 employstatus_10 fail03 lifeSat MLQ c19NormShould c19NormDo c19IsOrg trustGovState edu`

`0.06154919 0.0652525 0.1005851 0.3293603 0.06135636 0.05886684 0.08799994 0.1034505 0.0727743 0.05776118 0.142124 0.02873322`

`affAnx affBor affExc affExh affInsp PLRAEco disc02 disc03 jbInsec02 PFS01 fail01 lifeSat MLQ c19NormShould c19NormDo trustGovCtry trustGovState age edu`

`0.04854233 0.06002516 0.06926414 0.04935493 0.04810877 -0.03626116 0.1502406 0.0650781 0.06608606 -0.0901943 -0.07940344 0.06139211 0.1117521 0.1321387 0.08202972 0.06169359 0.1511349 -0.05740599 0.05113085`

`affExc affExh affInsp PLRAC19 disc02 disc03 employstatus_10 lifeSat MLQ c19NormShould c19NormDo c19IsOrg trustGovState age edu`

`0.05042581 0.04464075 0.06039361 0.07098512 0.1348678 0.07466859 0.3418363 0.09238179 0.05582841 0.08829495 0.07615688 0.0830227 0.1792569 -0.05783799 0.03418924`

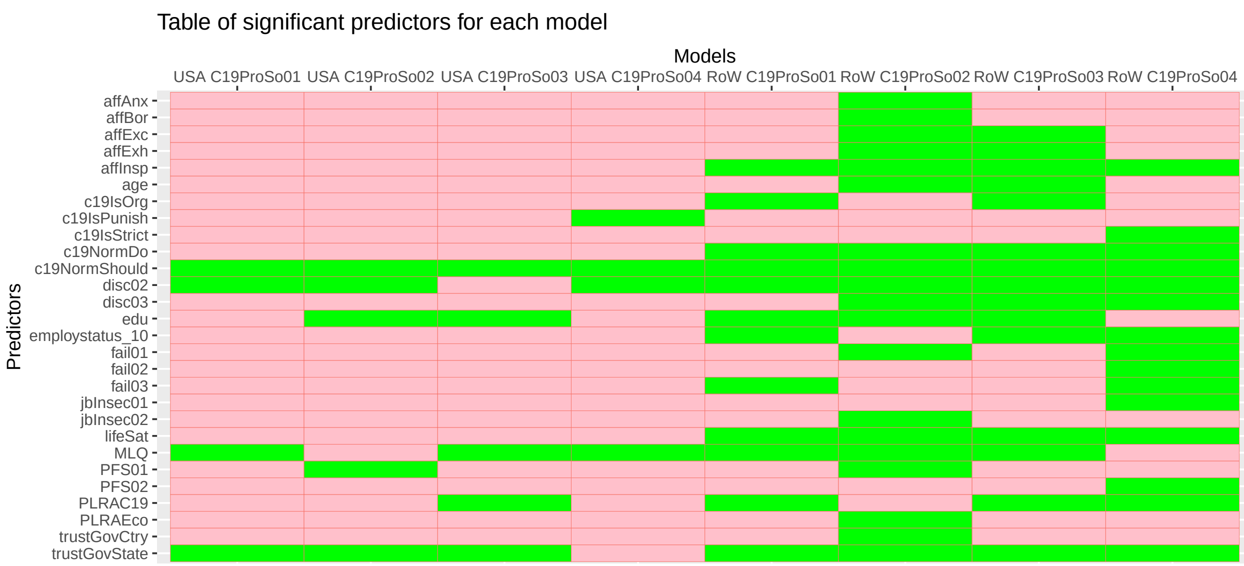
`affInsp PLRAC19 disc02 disc03 jbInsec01 employstatus_10 PFS02 fail01 fail02 fail03 lifeSat c19NormShould c19NormDo c19IsStrict trustGovState`

`0.07070157 0.08621845 0.1716264 0.04722402 0.06153325 0.348768 0.04779941 -0.06649288 -0.05701471 0.07318474 0.08059154 0.2330507 0.04257108 0.05811818 0.09548559`

Based on the summary for the rest of the world, all four models have roughly the same adjusted R-squared values between 0.12 and 0.17, which is narrower than the corresponding range for the US dataset (0.08 - 0.21). The models have many more significant predictors compared to the `usa` models. Strong predictors that predict all four models well are `disc02`, `lifeSat`, `c19NormShould`, `c19NormDo` and `trustGovState`. These predictors include most of those that had good performance across the four `usa` models, which are `c19NormShould`, `disc02` and `trustGovState`. As previously mentioned, the United States by itself resembles a collection of separate countries due to its size and diversity. Hence, it is no surprise that strong predictors for the United States would apply to other countries as a group as well.

The findings of the best predictors for each pro-social attitude for the United States and other countries as a group can be visualised in a table as shown below, generated using `ggplot2`.

```
summ_table <- table(predictors = prds, models = mdl)

# reorder the columns
summ_table <- summ_table[, c("USA C19ProSo01", "USA C19ProSo02", "USA C19ProSo03",
  "USA C19ProSo04", "RoW C19ProSo01", "RoW C19ProSo02", "RoW C19ProSo03",
  "RoW C19ProSo04")]

summ_table_vis <- ggplot(data = as.data.frame(summ_table)) +
  geom_tile(mapping = aes(x = models, y = predictors, fill = Freq, colour = "black")) +
  scale_fill_gradientn(colours = c("pink", "green")) +
  theme(legend.position = "none") +
  scale_x_discrete(position = "top") +
  scale_y_discrete(limits = rev) +
  labs(x = "Models", y = "Predictors",
    title = "Table of significant predictors for each model")

summ_table_vis
```



Section 3

3(a)

In addition to the indicators found in the sources listed in the references, some other socioeconomic and health data have been sourced from other websites as well. The final data table (in **Appendix**) that I have compiled for use in clustering consists of 8 indicators: `HDI`, `GHS`, `freedom`, `political_stability`, `happiness`, `total_vax_per_hundred`, `total_cases_per_mil` and `total_deaths_per_mil`. Details and explanations about each indicator and their sources are included in the **Appendix**.

To identify countries similar to the United States, k-means clustering is performed. Countries with NA values are first removed for the `kmeans()` function to work. This has minimal impact on our results as most of these countries are very different from the United States in terms of development and data transparency (eg. Afghanistan, Syria), and also do not appear in the baseline data in the first place (eg. Solomon Islands, Cuba). The data is then scaled and K-means clustering is performed with 15 random starts.

```
collected <- read.csv("task3.csv")
collected_clean <- na.omit(collected)
collected_clean[, 2:9] <- scale(collected_clean[, 2:9])

kfit <- kmeans(collected_clean[, 2:9], round(nrow(collected_clean) / 5), nstart = 15)
clusters <- data.frame(country = collected_clean[[1]], cluster = kfit$cluster)

target <- filter(clusters, country == "United States of America")$cluster
similar <- filter(clusters, cluster == target)
similar
```

```
##                      country cluster
## 17                   Belgium       1
## 45            Czech Republic       1
## 100                Lithuania       1
## 156                 Slovenia       1
## 182           United Kingdom       1
## 183 United States of America       1
```

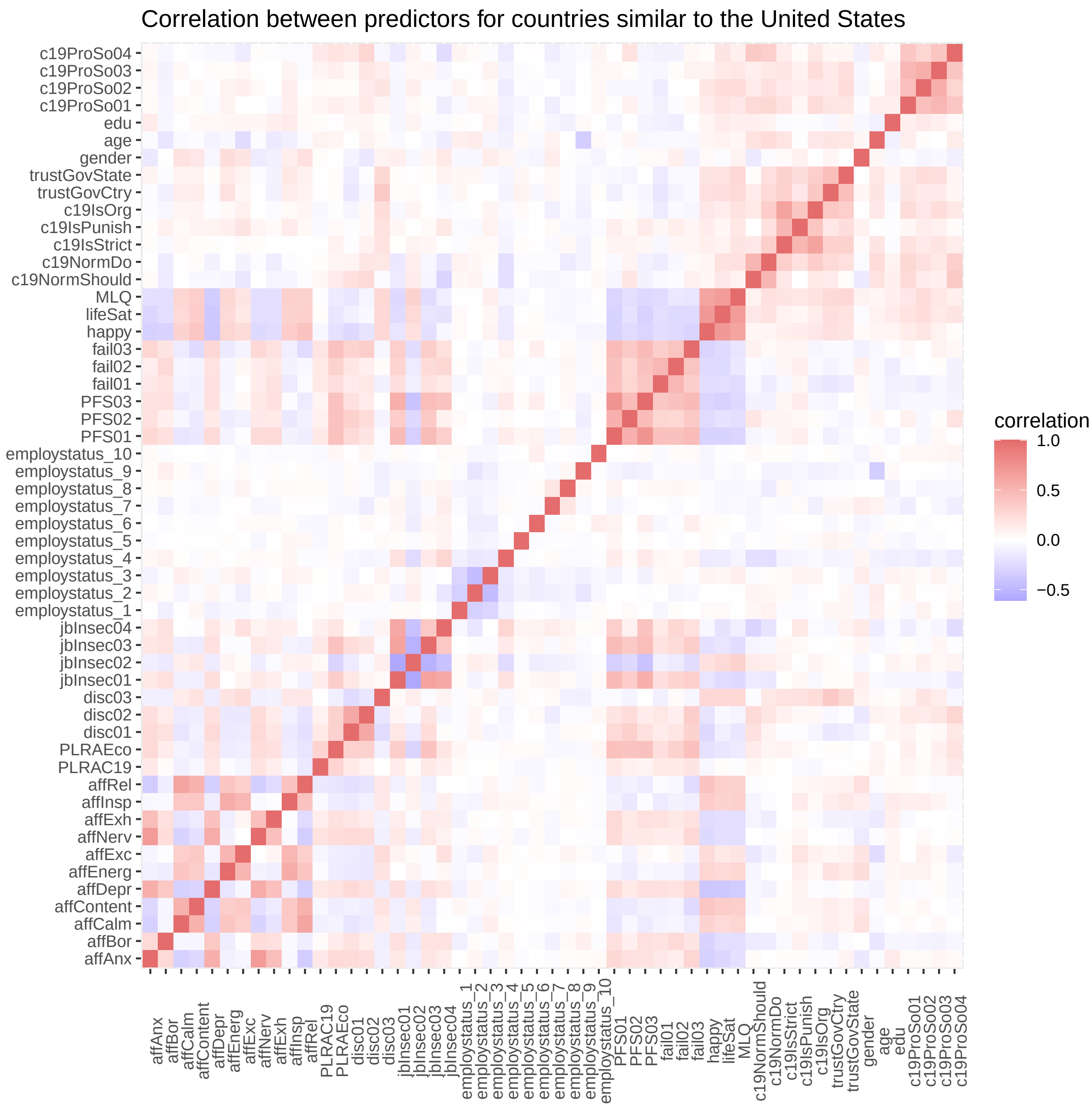
Based on the clustering, countries similar to the United States are Belgium, Czech Republic, Lithuania, Slovenia and the United Kingdom.

3(b)

Baseline data of the countries belonging to the cluster are first extracted through an inner join of `cvbase` and `similar`, with the United States data removed. A visualisation of the correlation matrix for this subset of data is then plotted, just as for `usa` and `rem`.

```
colnames(similar)[colnames(similar) == "country"] <- "coded_country"
intersect <- merge(cvbase, similar, by = "coded_country", all = FALSE)
intersect <- intersect[, -ncol(intersect)]
clus <- filter(intersect, coded_country != "United States of America")

clus_cor <- cor(subset(clus, select = -coded_country), use = "complete.obs")
clus_melted <- reshape2::melt(clus_cor)
```

```
clus_cor_plot
```



The scatter of coloured tiles for this heatmap resembles that of the United States heatmap, illustrating the similarity between these countries. The subsection of tiles showing correlation between predictors and pro-social attitudes are overall darker compared to the previous plots, indicating that the predictors for this cluster of countries might have better predictive performance compared to the previous two groups of data.

To find out how participant responses predict pro-social attitudes for this cluster of similar countries, the same code as in 2(b) and 2(c) is reused to print a formatted summary of the four models.

```
## Summary of models for predicting pro-social attitudes in countries similar to the US
```

```
## C19ProSo01
## R-squared value: 0.2135323
## Adjusted R-squared value: 0.1284619
## Significant predictors with p-value < 0.001:
##
```

```
## Coefficients of predictors:
##
##
## C19ProSo02
## R-squared value: 0.1949902
## Adjusted R-squared value: 0.107914
## Significant predictors with p-value < 0.001:
##
## Coefficients of predictors:
##
##
## C19ProSo03
## R-squared value: 0.2164434
## Adjusted R-squared value: 0.1316878
## Significant predictors with p-value < 0.001:
##
## Coefficients of predictors:
##
##
## C19ProSo04
## R-squared value: 0.3212664
## Adjusted R-squared value: 0.2478493
## Significant predictors with p-value < 0.001:
## disc02 PFS02
## Coefficients of predictors:
## 0.3093201 0.2396266
```

From the output, the models for these similar countries generally have roughly the same adjusted R-squared values as the models for the United States and all other countries as a group. The highest adjusted R-squared value is seen in the model for `C19ProSo04` (0.2478493), just as with the United States models. However, unlike the previous eight models, none of these models have significant predictors with p-values less than 0.001 except the model for `C19ProSo04`, whose significant predictors are `disc02` and `PFS02`. `disc02` also appears as a strong predictor in the United States model for `C19ProSo04`, but not `PFS02`. The rest-of-the-world model for `C19ProSo04`, however, has both `disc02` and `PFS02` as strong predictors.

Hence, the predictive performance of attributes for this cluster of countries is not significantly better than that of the United States nor the rest of the world, with similar R-squared values and predictors with overall higher p-values. The strong correlation we observed earlier may be due to chance or a small sample size, instead of actual statistically significant relationships between attribute and pro-social attitude.

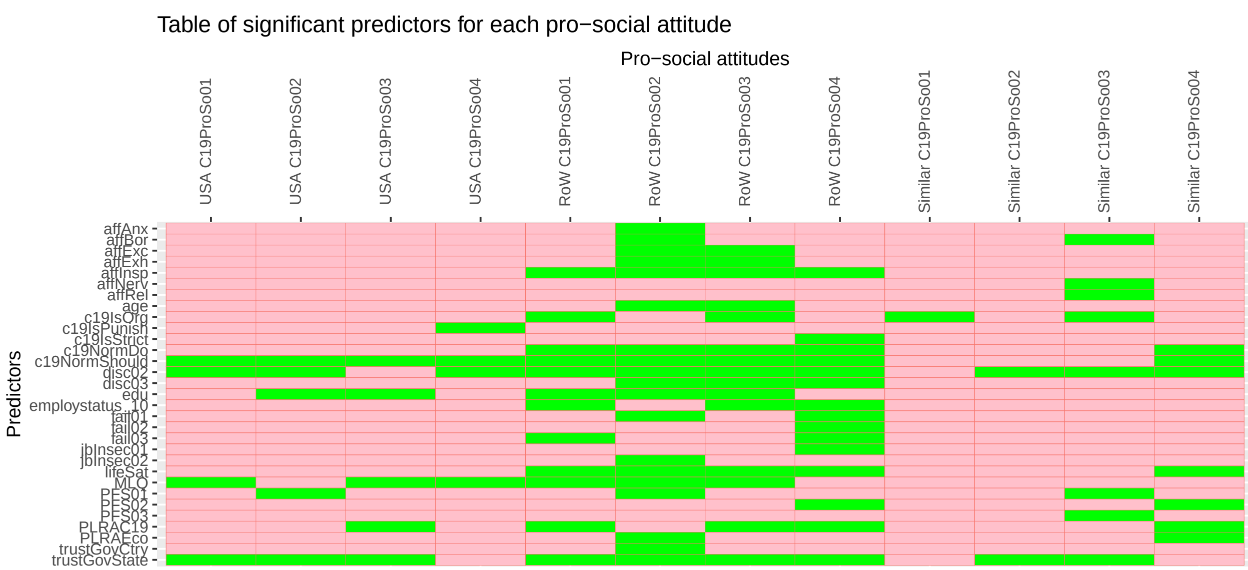
For the sake of comparison, we can set the definition of a strong predictor relative to the overall p-values in a model. We define a strong predictor for these new cluster models as a predictor with a p-value less than 0.05 (a commonly used threshold). The `model_eval` function is updated to reflect this (see **Appendix**) and a new visualisation table is created.

```
## Summary of models for predicting pro-social attitudes in countries similar to the US
```

```
## C19ProSo01
## R-squared value: 0.2135323
## Adjusted R-squared value: 0.1284619
## Significant predictors with p-value < 0.05:
## c19IsOrg
## Coefficients of predictors:
## 0.1637241
##
## C19ProSo02
```

```
## R-squared value: 0.1949902
## Adjusted R-squared value: 0.107914
## Significant predictors with p-value < 0.05:
## disc02 trustGovState
## Coefficients of predictors:
## 0.2962495 0.2245365
##
## C19ProSo03
## R-squared value: 0.2164434
## Adjusted R-squared value: 0.1316878
## Significant predictors with p-value < 0.05:
## affBor affNerv affRel disc02 PFS01 PFS03 c19IsOrg trustGovState
## Coefficients of predictors:
## -0.1300093 -0.2459104 -0.1962942 0.2401793 0.3136709 -0.2414336 0.2064041 0.2300887
##
## C19ProSo04
## R-squared value: 0.3212664
## Adjusted R-squared value: 0.2478493
## Significant predictors with p-value < 0.05:
## PLRAC19 PLRAEco disc02 PFS02 lifeSat c19NormShould c19NormDo
## Coefficients of predictors:
## 0.1410079 0.09845499 0.3093201 0.2396266 0.183763 0.1646776 0.1250363
```

```
summ_table_vis_2
```



We observe that the distribution of strong predictors of the similar countries' models is more alike to that of the United States models (ie. they look as "sparse" as the US models), with a few common significant predictors shared. The models of the group of all other countries share many more common significant predictors with the United States models, with more similar p-values. However, these models also have many strong predictors which are not as strong in the United States models. Therefore, the cluster of similar countries might give a better match to the important attributes for predicting pro-social attitudes. The higher p-values and fewer shared common strong predictors seen in their models may no longer be observed when further analysis is done or a larger sample size is introduced.

A possible explanation is that, despite being similar to the United States, each country in the cluster are slightly different in terms of socioeconomic factors outside the indicators used for clustering. When these slight differences are aggregated as a group, their performance in predicting pro-social attitudes deviates more from that of the United States alone. On the other hand, the United States models share many common strong predictors with the models of the group of all countries, due to the complexity of its politics, culture and other features of society, akin to a group of many countries. The group of all other countries may be too large and complex, and hence its models may report many significant predictors that are actually not significant in reality.

Appendix

Head of `cvbase` at the end of 1(b).

```
head(cvbase)
```

```
##       affAnx affBor affCalm affContent affDepr affEnerg affExc affNerv affExh
## 30480      3      4       3          2       2        1      1       4      1
## 34061      2      1       4          1       5        3      3       2      1
## 16871      3      2       3          3       2        4      2       2      1
## 21638      2      3       2          2       2        1      1       1      4
## 53709      4      3       3          3       2        3      2       2      2
## 49621      2      1       4          3       2        1      2       1      1
##       affInsp affRel PLRAC19 PLRAEco disc01 disc02 disc03 jbInsec01 jbInsec02
## 30480       1      2       5       6      1      1      1         1        -1
## 34061       3      4       3       6      1      1     -1         0        NA
## 16871       2      3       3       3      0      1     -1         0         1
## 21638       1      2       4       8      2      1     -2        NA        NA
## 53709       3      1       4       7      1      1      1        -1         1
## 49621       2      3       6       5      1      1     -2        -1         1
##       jbInsec03 jbInsec04 employstatus_1 employstatus_2 employstatus_3
## 30480         1         0              0              0              0
## 34061         2        NA              0              0              0
## 16871         0         0              1              0              0
## 21638        NA        NA              0              0              0
## 53709         0        -1              0              0              1
## 49621         1        -2              0              0              1
##       employstatus_4 employstatus_5 employstatus_6 employstatus_7
## 30480              0              0              0              0
## 34061              0              0              0              0
## 16871              0              0              0              0
## 21638              0              1              0              0
## 53709              0              0              0              0
## 49621              0              0              0              0
##       employstatus_8 employstatus_9 employstatus_10 PFS01 PFS02 PFS03 fail01
## 30480              0              1               0    -1    -1    -1      1
## 34061              0              1               0     1     2    -1      0
## 16871              0              0               0     0     1     0     -1
## 21638              0              1               0     0     2     2      2
## 53709              0              0               0     0     1     0     -1
## 49621              0              0               0    -1     1    -1      1
##       fail02 fail03 happy lifeSat MLQ c19NormShould c19NormDo c19IsStrict
## 30480      0      0     6       5   0             3         2           4
## 34061     -2     -1     7       2   2             3         3           5
## 16871     -2      0     4       3  -2             2        -1           2
## 21638      1      1     5       3   0             3        -1           1
## 53709      0      1     6       4   2             2         1           5
## 49621     -1      1     8       5   1             2        -2           5
##       c19IsPunish c19IsOrg trustGovCtry trustGovState gender age edu
```

```
## 30480           4         5            NA             NA       2   1   4
## 34061           5         4            NA             NA       1   2   5
## 16871           1         2             2              2       2   2   5
## 21638           1         1             1              1       2   1   5
## 53709           5         4             3              3       2   3   7
## 49621           2         2             3              3       2   3   4
##                  coded_country c19ProSo01 c19ProSo02 c19ProSo03 c19ProSo04
## 30480                    Spain         -1          0          1          1
## 34061                    Spain          2          1         -2          3
## 16871 United States of America          2         -2          2          1
## 21638               Bangladesh          3          1         -3         -3
## 53709               Kazakhstan         -1          1          0          0
## 49621                   Brazil          2          2          2          2
```

Code for correlation matrix of `rem` from 2(c).

```
rem_cor <- cor(subset(rem, select = -coded_country), use = "complete.obs")
rem_melted <- reshape2::melt(rem_cor)

rem_cor_plot <- ggplot(data = rem_melted) +
  geom_tile(mapping = aes(x = Var1, y = Var2, fill = value)) +
  scale_fill_gradient2(low = "#6b74ff", mid = "white", high = "#e46c6c", midpoint = 0) +
  labs(title = "Correlation between predictors for the rest of the world", x = "", y = "",
    fill = "correlation") +
  theme(axis.text.x = element_text(angle = 90))
```

Code for summary results of `rem` models from 2(c).

```
fitted_rem1 <- lm(c19ProSo01 ~ .,
  data = subset(rem, select = -c(coded_country, c19ProSo02, c19ProSo03, c19ProSo04)))
fitted_rem2 <- lm(c19ProSo02 ~ .,
  data = subset(rem, select = -c(coded_country, c19ProSo01, c19ProSo03, c19ProSo04)))
fitted_rem3 <- lm(c19ProSo03 ~ .,
  data = subset(rem, select = -c(coded_country, c19ProSo01, c19ProSo02, c19ProSo04)))
fitted_rem4 <- lm(c19ProSo04 ~ .,
  data = subset(rem, select = -c(coded_country, c19ProSo01, c19ProSo02, c19ProSo03)))

cat("Summary of models for predicting pro-social attitudes in the rest of the world\n\n")
counter <- 1
for (model in list(fitted_rem1, fitted_rem2, fitted_rem3, fitted_rem4)) {
  cat("C19ProSo0", counter, "\n", sep = "")
  res <- model_eval(model)
  cat("R-squared value:", res[[1]], "\n")
  cat("Adjusted R-squared value:", res[[2]], "\n")
  cat("Significant predictors with p-value < 0.001:\n")
  cat(res[[3]], "\n")
  cat("Coefficients of predictors:\n")
  cat(res[[4]], "\n")
  cat("\n")
  for (pred in res[[3]]) {
    mdl <- c(mdl, paste0("RoW C19ProSo0", counter))
  }
  prds <- c(prds, res[[3]])
  counter <- counter + 1
}
```

Final table of data compiled and used for clustering in 3(a).

```
collected
```

```
##                              country   HDI  GHS freedom political_stability
## 1                        Afghanistan 0.478 28.8      NA               -2.53
## 2                            Albania 0.796 45.0    8.14                0.11
## 3                            Algeria 0.745 26.2    5.26               -0.88
## 4                            Andorra 0.858 34.7      NA                1.63
## 5                             Angola 0.586 29.1    6.09               -0.71
## 6                Antigua and Barbuda 0.788 30.0      NA                0.96
## 7                          Argentina 0.842 54.4    7.38               -0.11
## 8                            Armenia 0.759 61.8    8.20               -0.84
## 9                          Australia 0.951 71.1    8.84                0.85
## 10                           Austria 0.916 56.9    8.67                0.91
## 11                        Azerbaijan 0.745 34.7    6.16               -0.85
## 12                           Bahamas 0.812 30.1    8.22                0.88
## 13                           Bahrain 0.875 36.3    5.73               -0.51
## 14                        Bangladesh 0.661 35.5    5.75               -0.97
## 15                          Barbados 0.790 34.9    7.92                1.12
## 16                           Belarus 0.808 43.9    6.73               -0.74
## 17                           Belgium 0.937 59.3    8.61                0.61
## 18                            Belize 0.683 29.7    7.64                0.46
## 19                             Benin 0.525 25.4    7.32               -0.30
## 20                            Bhutan 0.666 39.8    6.86                0.97
## 21                           Bolivia 0.692 29.9    6.94               -0.32
## 22            Bosnia and Herzegovina 0.780 35.4    7.54               -0.38
## 23                          Botswana 0.693 33.6    7.90                0.98
## 24                            Brazil 0.754 51.2    7.22               -0.49
## 25                            Brunei 0.829 43.5    6.46                1.17
## 26                          Bulgaria 0.795 59.9    8.08                0.46
## 27                      Burkina Faso 0.449 29.8    6.85               -1.64
## 28                           Burundi 0.426 22.1    5.02               -1.36
## 29                        Cape Verde 0.662 34.1      NA                0.90
## 30                          Cambodia 0.593 31.1    6.47               -0.13
## 31                          Cameroon 0.576 28.6    5.63               -1.41
## 32                            Canada 0.936 69.8    8.85                0.94
## 33          Central African Republic 0.404 18.6    5.62               -2.10
## 34                              Chad 0.394 23.9    5.57               -1.34
## 35                             Chile 0.855 56.2    8.44                0.06
## 36                             China 0.768 47.5    5.57               -0.48
## 37                          Colombia 0.752 53.2    7.01               -0.91
## 38                           Comoros 0.558 24.9    6.07               -0.23
## 39                             Congo 0.571 26.3    5.55               -0.61
## 40                        Costa Rica 0.809 40.8    8.25                0.87
## 41                     Côte d'Ivoire 0.550 31.2    6.90               -0.95
## 42                           Croatia 0.858 48.8    8.16                0.71
## 43                              Cuba 0.764 30.5      NA                0.43
## 44                            Cyprus 0.896 41.9    8.42                0.44
## 45                    Czech Republic 0.889 52.8    8.61                0.96
## 46                       D.R. Congo 0.479 26.1    5.62               -1.61
## 47                           Denmark 0.948 64.4    8.98                0.95
## 48                          Djibouti 0.509 25.2    5.84               -0.71
## 49                          Dominica 0.720 26.4      NA                1.39
## 50                Dominican Republic 0.767 34.5    7.88                0.14
```

```
## 51                              Ecuador 0.740 50.8    7.43                 -0.27
## 52                                Egypt 0.731 28.0    4.49                 -1.02
## 53                          El Salvador 0.675 40.8    7.39                 -0.21
## 54                    Equatorial Guinea 0.596 17.4      NA                 -0.29
## 55                              Eritrea 0.492 21.4      NA                 -1.01
## 56                              Estonia 0.890 55.5    8.91                  0.76
## 57                             Eswatini 0.597 29.3    5.79                 -0.03
## 58                             Ethiopia 0.498 37.8    5.95                 -2.07
## 59                                 Fiji 0.730 25.8    7.36                  0.67
## 60                              Finland 0.940 70.9    8.85                  0.98
## 61                               France 0.903 61.9    8.34                  0.37
## 62                                Gabon 0.706 21.8    6.80                 -0.09
## 63                               Gambia 0.500 28.7    6.88                  0.18
## 64                              Georgia 0.802 52.6    8.20                 -0.42
## 65                              Germany 0.942 65.5    8.73                  0.76
## 66                                Ghana 0.632 34.3    7.49                  0.07
## 67                               Greece 0.887 51.5    7.86                  0.15
## 68                              Grenada 0.795 26.7      NA                  1.04
## 69                            Guatemala 0.627 29.1    7.63                 -0.39
## 70                               Guinea 0.465 26.8    5.82                 -0.97
## 71                        Guinea-Bissau 0.483 21.4      NA                 -0.28
## 72                               Guyana 0.714 30.8    7.49                 -0.14
## 73                                Haiti 0.535 30.4    7.21                 -1.10
## 74                             Honduras 0.621 26.2    7.09                 -0.61
## 75                     Hong Kong S.A.R. 0.952   NA    8.41                  0.26
## 76                              Hungary 0.846 54.4    7.73                  0.86
## 77                              Iceland 0.959 48.5    8.77                  1.37
## 78                                India 0.633 42.8    6.39                 -0.62
## 79                            Indonesia 0.705 50.4    7.10                 -0.51
## 80                                 Iran 0.774 36.5    4.53                 -1.62
## 81                                 Iraq 0.686 24.0    5.02                 -2.40
## 82                              Ireland 0.945 55.3    8.90                  0.86
## 83                               Israel 0.919 47.2    7.66                 -1.06
## 84                                Italy 0.895 51.9    8.49                  0.58
## 85                              Jamaica 0.709 31.8    7.91                  0.22
## 86                                Japan 0.925 60.5    8.73                  1.03
## 87                               Jordan 0.720 42.8    6.91                 -0.28
## 88                           Kazakhstan 0.811 46.1    6.77                 -0.25
## 89                                Kenya 0.575 38.8    6.73                 -1.09
## 90                             Kiribati 0.624 26.2      NA                  1.19
## 91                               Kuwait 0.831 36.8    6.34                  0.30
## 92                           Kyrgyzstan 0.692 42.4    7.18                 -0.43
## 93                                 Laos 0.607 34.8    5.85                  0.73
## 94                               Latvia 0.863 61.9    8.67                  0.69
## 95                              Lebanon 0.706 33.4    6.76                 -1.49
## 96                              Lesotho 0.514 30.9    7.01                 -0.22
## 97                              Liberia 0.481 35.7    6.81                 -0.24
## 98                                Libya 0.718 25.3    5.05                 -2.37
## 99                        Liechtenstein 0.935 46.4      NA                  1.64
## 100                           Lithuania 0.875 59.5    8.68                  0.82
## 101                          Luxembourg 0.930 48.4    8.80                  1.21
## 102                          Madagascar 0.501 30.4    7.02                 -0.64
## 103                              Malawi 0.512 28.5    6.99                 -0.11
## 104                            Malaysia 0.803 56.4    7.17                  0.14
```

```
## 105                         Maldives 0.747 32.0       NA               0.50
## 106                             Mali 0.428 29.0     6.25              -2.35
## 107                            Malta 0.918 40.2     8.45               0.97
## 108                 Marshall Islands 0.639 24.6       NA               0.61
## 109                       Mauritania 0.556 26.2     5.73              -0.67
## 110                        Mauritius 0.802 39.7     8.07               0.86
## 111                           Mexico 0.758 57.0     6.92              -0.64
## 112                       Micronesia 0.628 28.5       NA               1.11
## 113                          Moldova 0.767 41.0     7.68              -0.21
## 114                         Mongolia 0.739 41.0     8.00               0.65
## 115                       Montenegro 0.832 44.1     7.88              -0.15
## 116                          Morocco 0.683 33.6     5.90              -0.40
## 117                       Mozambique 0.446 30.4     6.80              -1.23
## 118                          Myanmar 0.585 38.3     5.78              -2.07
## 119                          Namibia 0.615 30.3     7.56               0.55
## 120                            Nepal 0.602 34.0     7.12              -0.24
## 121                      Netherlands 0.941 64.7     8.78               0.92
## 122                      New Zealand 0.937 62.5     9.01               1.44
## 123                        Nicaragua 0.667 36.3     6.24              -0.47
## 124                            Niger 0.400 28.7     6.41              -1.62
## 125                          Nigeria 0.535 38.0     6.28              -1.78
## 126                  North Macedonia 0.770 42.2     7.75               0.12
## 127                           Norway 0.961 60.2     8.76               1.10
## 128                             Oman 0.816 39.1     5.92               0.51
## 129                         Pakistan 0.544 30.4     5.63              -1.67
## 130                            Palau 0.767 25.5       NA               0.95
## 131                        Palestine 0.715   NA       NA                 NA
## 132                           Panama 0.805 53.5     8.12               0.29
## 133                 Papua New Guinea 0.558 25.0     7.17              -0.58
## 134                         Paraguay 0.717 40.3     7.54               0.00
## 135                             Peru 0.762 54.9     7.93              -0.41
## 136                      Philippines 0.699 45.7     6.83              -0.93
## 137                           Poland 0.876 55.7     7.96               0.51
## 138                         Portugal 0.866 54.7     8.69               0.95
## 139                            Qatar 0.855 48.7     6.15               0.96
## 140                          Romania 0.821 45.7     8.33               0.53
## 141                           Russia 0.822 49.1     6.23              -0.65
## 142                           Rwanda 0.534 33.1     6.36               0.17
## 143            Saint Kitts and Nevis 0.777 31.7       NA               0.96
## 144                      Saint Lucia 0.715 34.7       NA               0.85
## 145 Saint Vincent and the Grenadines 0.751 33.5       NA               1.04
## 146                            Samoa 0.707 28.8       NA               1.11
## 147                       San Marino 0.853 32.9       NA               0.91
## 148            Sao Tome and Principe 0.618 26.6       NA               0.60
## 149                     Saudi Arabia 0.875 44.9     5.12              -0.58
## 150                          Senegal 0.511 32.8     7.07              -0.17
## 151                           Serbia 0.802 45.0     7.54              -0.13
## 152                       Seychelles 0.785 31.8     7.84               0.76
## 153                     Sierra Leone 0.477 32.7     6.70              -0.16
## 154                        Singapore 0.939 57.4     7.98               1.49
## 155                         Slovakia 0.848 54.4     8.21               0.56
## 156                         Slovenia 0.918 67.8     8.37               0.76
## 157                  Solomon Islands 0.564 23.3       NA               0.49
## 158                     South Africa 0.713 45.8     7.30              -0.71
```

```
## 159                      South Korea 0.925 65.4    8.39                0.66
## 160                      South Sudan 0.385 21.3      NA               -2.30
## 161                            Spain 0.905 60.9    8.56                0.58
## 162                        Sri Lanka 0.782 34.1    6.58               -0.32
## 163                            Sudan 0.508 28.3    4.48               -1.94
## 164                         Suriname 0.730 35.0    7.64                0.37
## 165                           Sweden 0.947 64.9    8.83                1.03
## 166                      Switzerland 0.962 58.8    9.11                1.13
## 167                            Syria 0.577 16.7    3.66               -2.66
## 168                       Tajikistan 0.685 29.3    5.52               -0.61
## 169                         Tanzania 0.549 31.3    6.48               -0.44
## 170                         Thailand 0.800 68.2    6.89               -0.55
## 171                      Timor-Leste 0.607 27.8    7.22                0.17
## 172                             Togo 0.539 27.8    6.50               -0.80
## 173                            Tonga 0.745 26.4      NA                1.07
## 174              Trinidad and Tobago 0.810 36.8    7.70                0.15
## 175                          Tunisia 0.731 31.5    6.46               -0.70
## 176                           Turkey 0.838 50.0    5.79               -1.10
## 177                     Turkmenistan 0.745 31.9      NA               -0.32
## 178                           Tuvalu 0.641 20.0      NA                1.28
## 179                           Uganda 0.525 36.5    6.32               -0.86
## 180                          Ukraine 0.773 38.9    6.86               -1.10
## 181             United Arab Emirates 0.911 39.6    6.06                0.65
## 182                   United Kingdom 0.929 67.2    8.75                0.54
## 183         United States of America 0.921 75.9    8.73                0.00
## 184                          Uruguay 0.809 40.3    8.36                1.05
## 185                       Uzbekistan 0.727 39.0      NA               -0.24
## 186                          Vanuatu 0.607 25.9      NA                0.79
## 187                        Venezuela 0.691 20.9    4.03               -1.53
## 188                          Vietnam 0.703 42.9    5.90               -0.11
## 189                            Yemen 0.455 16.1    4.08               -2.59
## 190                           Zambia 0.565 26.5    6.82                0.06
## 191                         Zimbabwe 0.593 32.4    5.60               -1.03
##     happiness total_vax_per_hundred total_cases_per_mil total_deaths_per_mil
## 1       2.523                 11.37            3843.027              178.853
## 2       5.117                 81.50           73495.999             1130.064
## 3       4.887                 27.94            4855.709              139.656
## 4          NA                146.85          289593.327             1753.441
## 5          NA                 32.64            2157.605               49.369
## 6          NA                129.19           45802.585             1269.036
## 7       5.929                172.04          127015.620             2596.686
## 8       5.283                 58.51          124054.477             2867.139
## 9       7.183                162.66           13850.033               92.790
## 10      7.268                186.55          141452.592             1866.187
## 11      5.171                109.54           59504.476              805.748
## 12         NA                 73.22           59699.163             1748.827
## 13      6.647                219.14          191141.779              946.858
## 14      5.025                 62.26            9262.063              163.985
## 15         NA                106.38          100516.251              923.145
## 16      5.534                 80.84           73162.372              583.222
## 17      6.834                186.45          179883.824             2432.755
## 18         NA                104.77           79122.099             1473.037
## 19      5.045                 13.28            1875.553               12.057
## 20         NA                147.59            3399.548                3.834
```

```
## 21      5.716              80.11     48410.298          1607.478
## 22      5.813              48.06     89830.928          4152.737
## 23      3.467              42.89     84421.169           932.213
## 24      6.330             153.86    103401.940          2874.028
## 25         NA             200.09     34454.190           135.857
## 26      5.266              54.57    109746.821          4554.734
## 27      4.834               4.65       777.639            14.025
## 28      3.775               0.06      2370.131             1.086
## 29         NA              96.29     68679.383           593.430
## 30      4.830             181.64      7185.596           179.629
## 31      5.142               3.65      3928.633            66.381
## 32      7.103             179.01     54674.470           779.054
## 33         NA               7.83      2232.240            18.103
## 34      4.355               1.61       321.667            10.213
## 35      6.172             226.05     92058.065          1994.314
## 36      5.339             198.85        92.420             3.997
## 37      6.012             124.71     99059.263          2503.488
## 38      4.289              69.50      7785.770           187.623
## 39      5.342              12.71      3563.730            61.805
## 40      7.069             149.71    110206.152          1419.462
## 41      5.306              25.26      2419.910            25.284
## 42      5.882             117.35    176082.986          3099.722
## 43         NA             275.36     86117.905           742.227
## 44      6.223             172.00    180555.509           720.977
## 45      6.965             147.62    239885.878          3462.077
## 46         NA               0.34       800.655            12.372
## 47      7.620             203.62    133231.468           553.529
## 48         NA               5.77     12162.187           168.622
## 49         NA              78.40     93652.932           645.977
## 50      5.545             125.45     37160.446           378.134
## 51      5.764             153.14     30320.534          1870.396
## 52      4.283              47.59      3466.327           195.756
## 53      6.061             151.83     19212.981           603.340
## 54         NA              27.03      8185.485           104.483
## 55         NA                 NA      2166.643            20.358
## 56      6.189             136.41    182347.157          1456.943
## 57      4.308              33.25     54783.303          1080.987
## 58      4.275               8.85      3367.185            56.136
## 59         NA             136.28     57360.484           750.724
## 60      7.842             173.57     47621.033           307.720
## 61      6.690             183.22    146728.723          1871.705
## 62      4.852              16.45     17496.045           120.553
## 63      5.051              10.89      3758.322           126.756
## 64      4.891              67.11    249638.058          3685.518
## 65      7.155             184.68     85942.734          1420.562
## 66      5.088              23.17      4364.905            39.013
## 67      5.723             168.22    112691.012          1994.035
## 68         NA              62.55     48406.252          1594.146
## 69      6.435              63.39     35119.817           902.380
## 70      4.984              21.30      2341.236            28.212
## 71         NA              19.66      3079.437            70.764
## 72         NA              88.50     48518.227          1299.573
## 73      3.615               1.70      2258.869            66.724
## 74      5.919              91.91     36379.485          1000.109
```

```
## 75      5.477                132.54          NA          NA
## 76      5.992                151.24  126053.645    3931.454
## 77      7.554                192.26   75853.506      96.540
## 78      3.819                102.24   24583.308     339.465
## 79      5.345                 99.73   15472.592     523.025
## 80      4.721                131.24   69934.029    1485.840
## 81      4.854                 31.78   47047.604     542.834
## 82      7.085                196.18  149793.912    1211.999
## 83      7.157                177.65  146252.196     874.061
## 84      6.483                188.66  101315.788    2324.744
## 85      6.309                 42.75   33101.647     873.600
## 86      5.940                162.94   13983.875     148.388
## 87      4.395                 73.24   94060.939    1118.212
## 88      6.152                 90.22   55265.342     939.633
## 89      4.607                 18.51    5409.043      99.505
## 90         NA                 62.61          NA          NA
## 91      6.106                162.63   97597.125     578.137
## 92      5.744                 34.01   27853.952     422.585
## 93      5.030                 77.43   14616.420      47.812
## 94      6.032                138.09  149500.663    2469.397
## 95      4.584                 79.78  131816.711    1658.001
## 96      3.512                 37.21   12859.600     291.002
## 97      4.625                 16.60    1241.634      54.123
## 98      5.410                 39.34   56982.296     836.129
## 99         NA                160.12  159827.214    1753.272
## 100     6.255                150.25  190696.342    2689.761
## 101     7.324                166.23  158256.396    1412.907
## 102     4.208                  2.51    1697.943      34.682
## 103     3.600                  8.83    3636.356     115.411
## 104     5.384                170.46   81162.575     927.038
## 105     5.198                150.85  182704.019     500.193
## 106     4.723                  4.68     914.861      29.123
## 107     6.602                201.07   98388.691     894.443
## 108        NA                    NA      96.170          NA
## 109     4.227                 40.90    8689.344     182.216
## 110     6.049                156.75   70100.456     604.858
## 111     6.317                116.71   31644.640    2382.841
## 112        NA                    NA          NA          NA
## 113     5.766                 54.28  114812.345    3137.495
## 114     5.677                157.15  203809.588     584.397
## 115     5.581                101.19  268446.551    3828.845
## 116     4.918                134.19   25656.966     396.284
## 117     4.794                 44.64    5587.555      60.541
## 118     4.426                 58.80    9797.725     355.634
## 119     4.574                 24.55   57981.149    1419.932
## 120     5.269                 71.98   27119.361     379.539
## 121     7.464                162.32  177345.164    1189.079
## 122     7.277                157.86    2650.961       9.836
## 123     5.972                112.04    1951.962      31.230
## 124     5.074                  3.71     281.021      10.455
## 125     4.759                  6.79    1105.114      13.865
## 126     5.101                 83.83  107493.483    3803.963
## 127     7.392                180.14   72669.388     256.518
## 128        NA                133.69   66754.583     979.612
```

```
## 129     4.934                66.41          5490.774           122.638
## 130        NA                   NA           552.975                NA
## 131     4.517                64.44         89580.227           939.224
## 132     6.180               140.49        111383.433          1684.215
## 133        NA                 4.97          3564.955            58.170
## 134     5.653               100.78         68738.907          2451.648
## 135     5.840               150.06         67181.253          5949.676
## 136     5.880                93.92         24584.998           444.561
## 137     6.166               117.89        103098.230          2435.147
## 138     5.929               194.34        132070.771          1843.760
## 139        NA               193.15         92680.838           228.931
## 140     6.140                80.50         91927.269          2986.581
## 141     5.477               101.14         72557.126          2134.289
## 142     3.415                91.38          8024.998            97.919
## 143        NA               115.07         61198.381           587.236
## 144        NA                58.37         74903.265          1640.055
## 145        NA                58.87         57253.340           798.392
## 146        NA               118.35             8.993                NA
## 147        NA               160.11        244909.469          2938.557
## 148        NA                60.52         17049.777           250.667
## 149     6.494               139.79         15255.011           243.760
## 150     5.132                10.99          4323.634           109.145
## 151     6.078               119.91        188770.738          1846.455
## 152        NA               171.25        231371.634          1176.086
## 153     3.849                10.09           811.437            14.293
## 154     6.377               209.25         49505.040           146.709
## 155     6.331                88.61        149152.071          2947.662
## 156     6.461               130.29        218941.214          2891.252
## 157        NA                32.57            33.137                NA
## 158     4.956                46.59         57543.972          1520.372
## 159     5.845               200.63         12174.566           107.361
## 160        NA                 2.46          1431.848            12.462
## 161     6.491               175.64        136797.480          1927.894
## 162     4.325               155.01         26898.175           686.098
## 163        NA                 6.99           998.950            71.191
## 164        NA                79.25         84186.290          1923.805
## 165     7.363               167.00        124623.804          1453.644
## 166     7.571               158.42        152718.543          1363.885
## 167        NA                 7.93          2270.845           130.756
## 168     5.466                66.37          1757.598            12.559
## 169     3.623                 3.71           447.435            11.252
## 170     5.985               146.67         31011.538           302.635
## 171        NA                   NA         14789.405            90.957
## 172     4.107                27.28          3408.749            28.027
## 173        NA               121.87             9.357                NA
## 174        NA                91.99         59324.918          1845.147
## 175     4.596                98.39         58743.540          2068.935
## 176     4.948               154.26        110635.410           962.067
## 177     5.066                 0.80                NA                NA
## 178        NA               106.87                NA                NA
## 179     4.636                20.66          3019.518            69.778
## 180     4.875                71.68         92380.048          2415.486
## 181     6.561               237.33         80446.976           228.998
## 182     7.064               197.47        199109.448          2220.847
```

```
## 183     6.951                     157.08          158249.753               2421.163
## 184     6.431                     203.91          119875.973               1802.036
## 185     6.179                     112.73            5744.052                 42.885
## 186        NA                      46.74              21.423                     NA
## 187     4.892                     106.18           15694.676                188.010
## 188     5.411                     153.72           17464.069                327.620
## 189     3.658                       1.62             300.505                 58.878
## 190     4.073                       8.64           12448.652                186.335
## 191     3.145                      44.51           12973.101                306.179
```

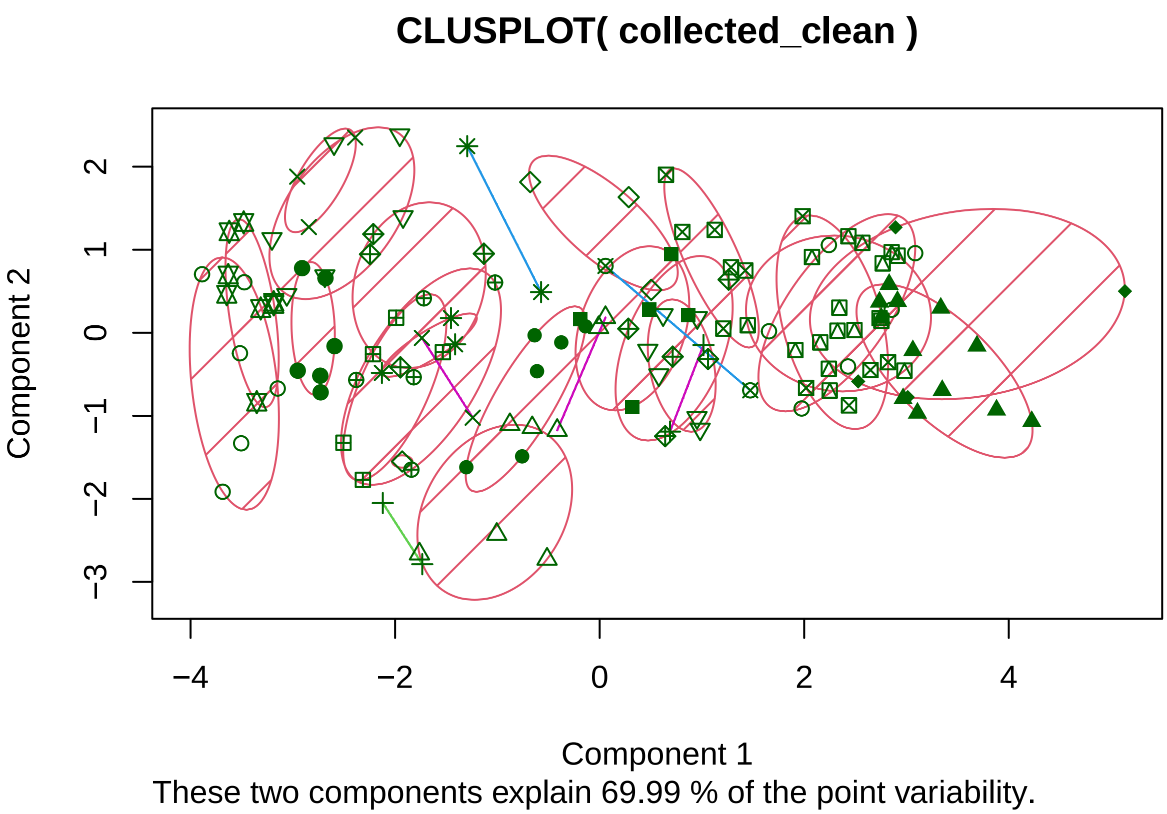
Explanation of each indicator used for clustering and their sources (from 3(a)).

- `HDI`: Human Development Index (2021); a value between 0 and 1 that measures average achievement in human development based on three dimensions - life expectancy, education and standard of living. (Source: Human Development Reports)
- `GHS`: Global Health Security Index (2021); a value between 0 and 100 that benchmarks a country's health security and preparedness in preventing, detecting and responding to health emergencies. (Source: Global Health Security Index: Reports and Data)
- `freedom`: Human Freedom Index (2021); a value between 0 and 10 that assesses the level of human freedom in a country. Human freedom is a combination of two distinct dimensions - personal freedom (freedom of religion, speech, sexual orientation, etc.) and economic freedom (size of government, judicial impartiality, freedom to trade, etc.) (Source: World Population Review)
- `political_stability`: a value **approximately** between -2.5 and 2.5 that evaluates political stability and absence of violence/terrorism of each country in 2021. (Source: The World Bank Data Collections (and Governance Indicators))
- `happiness`: World Happiness Report score (2021); a value between 0 and 10 that represents happiness of a country's citizens based on several socioeconomic factors. (Source: World Happiness Report)
- `total_vax_per_hundred`: latest updated total number of COVID-19 vaccinations administered per 100 people before 2022.
- `total_cases_per_mil`: latest updated total number of COVID-19 cases per 1,000,000 people before 2022.
- `total_deaths_per_mil`: latest updated total number of COVID-19 cases per 1,000,000 people before 2022.

The last three indicators were sourced from Our World in Data's COVID-19 Github repository.

Visualisation of k-means clustering performed in 3(a) (cluster plot).

```
library(cluster)
clusplot(collected_clean, kfit$cluster, color = TRUE, shade = TRUE, labels = 0, lines = 0)
```



Code for correlation matrix of `rem` from 3(b).

```
clus_cor_plot <- ggplot(data = clus_melted) +
  geom_tile(mapping = aes(x = Var1, y = Var2, fill = value)) +
  scale_fill_gradient2(low = "#6b74ff", mid = "white", high = "#e46c6c", midpoint = 0) +
  labs(title = "Correlation between predictors for countries similar to the United States",
    x = "", y = "", fill = "correlation") +
  theme(axis.text.x = element_text(angle = 90))
```

Code for summary results of `clus` models from 3(b).

```
fitted_clus1 <- lm(c19ProSo01 ~ .,
  data = subset(clus, select = -c(coded_country, c19ProSo02, c19ProSo03, c19ProSo04)))
fitted_clus2 <- lm(c19ProSo02 ~ .,
  data = subset(clus, select = -c(coded_country, c19ProSo01, c19ProSo03, c19ProSo04)))
fitted_clus3 <- lm(c19ProSo03 ~ .,
  data = subset(clus, select = -c(coded_country, c19ProSo01, c19ProSo02, c19ProSo04)))
fitted_clus4 <- lm(c19ProSo04 ~ .,
  data = subset(clus, select = -c(coded_country, c19ProSo01, c19ProSo02, c19ProSo03)))

cat("Summary of models for predicting pro-social attitudes in countries similar to the US\n\n")
counter <- 1
for (model in list(fitted_clus1, fitted_clus2, fitted_clus3, fitted_clus4)) {
  cat("C19ProSo0", counter, "\n", sep = "")
  res <- model_eval(model)
  cat("R-squared value:", res[[1]], "\n")
  cat("Adjusted R-squared value:", res[[2]], "\n")
  cat("Significant predictors with p-value < 0.001:\n")
  cat(res[[3]], "\n")
```

```
  cat("Coefficients of predictors:\n")
  cat(res[[4]], "\n")
  cat("\n")
  counter <- counter + 1
}
```

Code for summary results of `clus` models from 3(b), with updated `model_eval` function such that significant predictors have p-value less than 0.05.

```
model_eval_2 <- function(model) {
  rsqr <- summary(model)$r.squared
  a_rsqr <- summary(model)$adj.r.squared
  sig <- which(summary(model)$coefficients[-1, 4] < 0.05) + 1
  preds <- rownames(summary(model)$coefficients[sig, , drop = FALSE])
  coefs <- summary(model)$coefficients[sig, 1]

  return(list(rsqr, a_rsqr, preds, coefs))
}

cat("Summary of models for predicting pro-social attitudes in countries similar to the US\n\n")
counter <- 1
for (model in list(fitted_clus1, fitted_clus2, fitted_clus3, fitted_clus4)) {
  cat("C19ProSo0", counter, "\n", sep = "")
  res <- model_eval_2(model)
  cat("R-squared value:", res[[1]], "\n")
  cat("Adjusted R-squared value:", res[[2]], "\n")
  cat("Significant predictors with p-value < 0.05:\n")
  cat(res[[3]], "\n")
  cat("Coefficients of predictors:\n")
  cat(res[[4]], "\n")
  cat("\n")
  for (pred in res[[3]]) {
    mdl <- c(mdl, paste0("Similar C19ProSo0", counter))
  }
  prds <- c(prds, res[[3]])
  counter <- counter + 1
}
```

Code for table of strong predictors of `usa`, `rem` and `clus` models from 3(b).

```
summ_table_2 <- table(predictors = prds, models = mdl)
summ_table_2 <- summ_table_2[, c("USA C19ProSo01", "USA C19ProSo02", "USA C19ProSo03",
  "USA C19ProSo04", "RoW C19ProSo01", "RoW C19ProSo02", "RoW C19ProSo03", "RoW C19ProSo04",
  "Similar C19ProSo01", "Similar C19ProSo02", "Similar C19ProSo03", "Similar C19ProSo04")]

summ_table_vis_2 <- ggplot(data = as.data.frame(summ_table_2)) +
  geom_tile(mapping = aes(x = models, y = predictors, fill = Freq, colour = "black")) +
  scale_fill_gradientn(colours = c("pink", "green")) +
  theme(legend.position = "none") +
  scale_x_discrete(position = "top") +
  scale_y_discrete(limits = rev) +
  labs(x = "Pro-social attitudes", y = "Predictors",
    title = "Table of significant predictors for each pro-social attitude") +
  theme(axis.text.x = element_text(angle = 90))
```