Monash University

FIT3152 - Data Analytics

Assignment 1, Semester 1, 2024

Rhyme Bulbul

31865224

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

AI statement: Generative AI was used in this assignment

Task 1: Descriptive analysis and pre-processing

1(a) A condensed extract of the PsyCorona baseline study is contained in the file `PsyCoronaBaselineExtract.csv`.

We start by taking a unique sample of the data set based on my student ID, and attaching the data to the R search path for ease of variable use

```
rm(list = ls())
set.seed(31865224)
cvbase = read.csv("PsyCoronaBaselineExtract.csv")
cvbase <- cvbase[sample(nrow(cvbase), 40000), ]
```

We will take assistance from a few libraries, so let's start by importing those.

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

It is helpful to learn about a dataset's features and properties before working on it.

```
dim(cvbase)
as.data.frame(sapply(cvbase, class))
summary(cvbase, na.rm = TRUE)
```

In this case, we will run the dimensions method, to find out the data frame has 40,000 rows, given we have sampled it to be so when reading it in, and 52 columns.

The only text attributes in the dataframe appear to be `coded_country` and the `Rank Order of Life` columns, while the rest are integer data

According to the codebook extract, every column aside from `employstatus`, `gender`, `age`, `edu`, and `coded_country` contains ordinal numbers representing degrees of agreement for things like age group, education level, and level of agreement. Different `gender`, `age`, and `education` categories are coded by the integer values in their respective columns. Each record may only have a maximum of one employstatus column with a value of 1, indicating the employee's employment status.

We may infer that the numerical attributes have different ranges from the `summary()` output. Survey questions evaluating a two-sided degree of agreement vary from a negative number to its modulus, while those measuring a one-sided degree of agreement go from 1 to a larger positive number, such as 4, 5 or 6.

We are able to use the following, to understand `coded_country` better

```
sort(unique(cvbase$coded_country))
table(cvbase$coded_country)
max(table(cvbase$coded_country))
which(table(cvbase$coded_country) == max(table(cvbase$coded_country)))
```

```
min(table(cvbase$coded_country))
which(table(cvbase$coded_country) == min(table(cvbase$coded_country)))
```

Based on the outputs, there seems to be 110 unique countries inclusive of NA values, with each country having a widely different number of entries, the Croatia being the highest at 6987, and host of other countries having much less.

Missing values are present in all columns, however this is the norm as surveys of this nature do not require participants to answer all questions. The `employstatus` columns appear to be the biggest culprit, given each participant will only pick one out of the 10 categories. In this dataframe, `employstatus_3` seems to have the least missing values, while `employstatus_8` has the highest number. Potentially, this could imply that the majority of participants are working 40 hours or more, while a small number of people who are disabled may be out of work.

Another point to note is the mean age group in this dataframe amounts to 2.905, which indicates that the majority of participants are likely to be aged between 35-44 years. This could indicate working-class adults with concise lifestyle, and are studied accordingly to research the outcomes covid had on people.

1(b)

This dataset is relatively tidy and without too many missing values, hence preprocessing should not be required. However, the missing values in the `employstatus` column should be replaced with 0, as it would be easier to transform and process the data in binary format, which might be required for linear regression involving these attributes down the track.

```
cvbase[is.na(cvbase)] <- 0
```

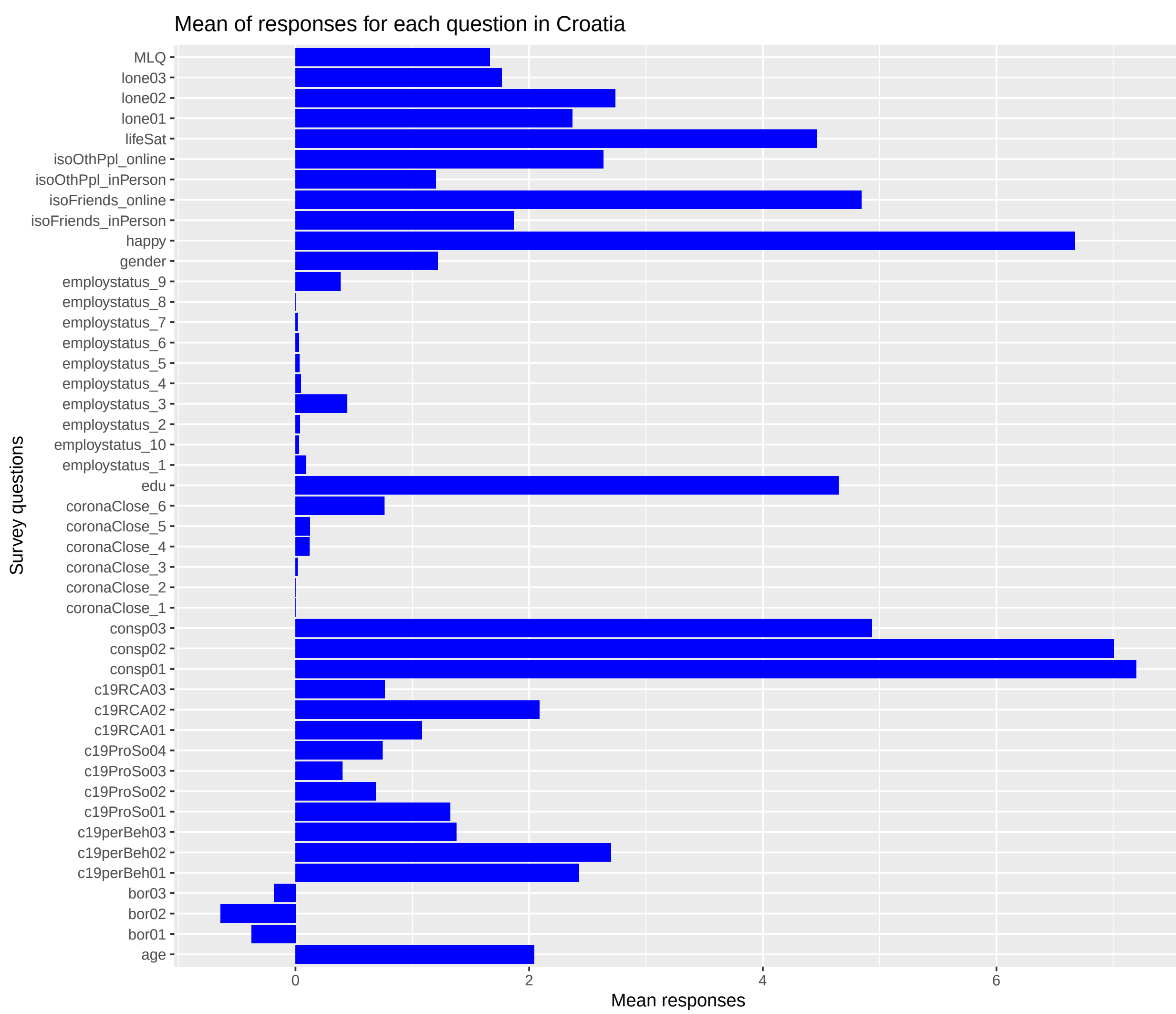
Task 2: Focus country vs all other countries as a group

2(a)

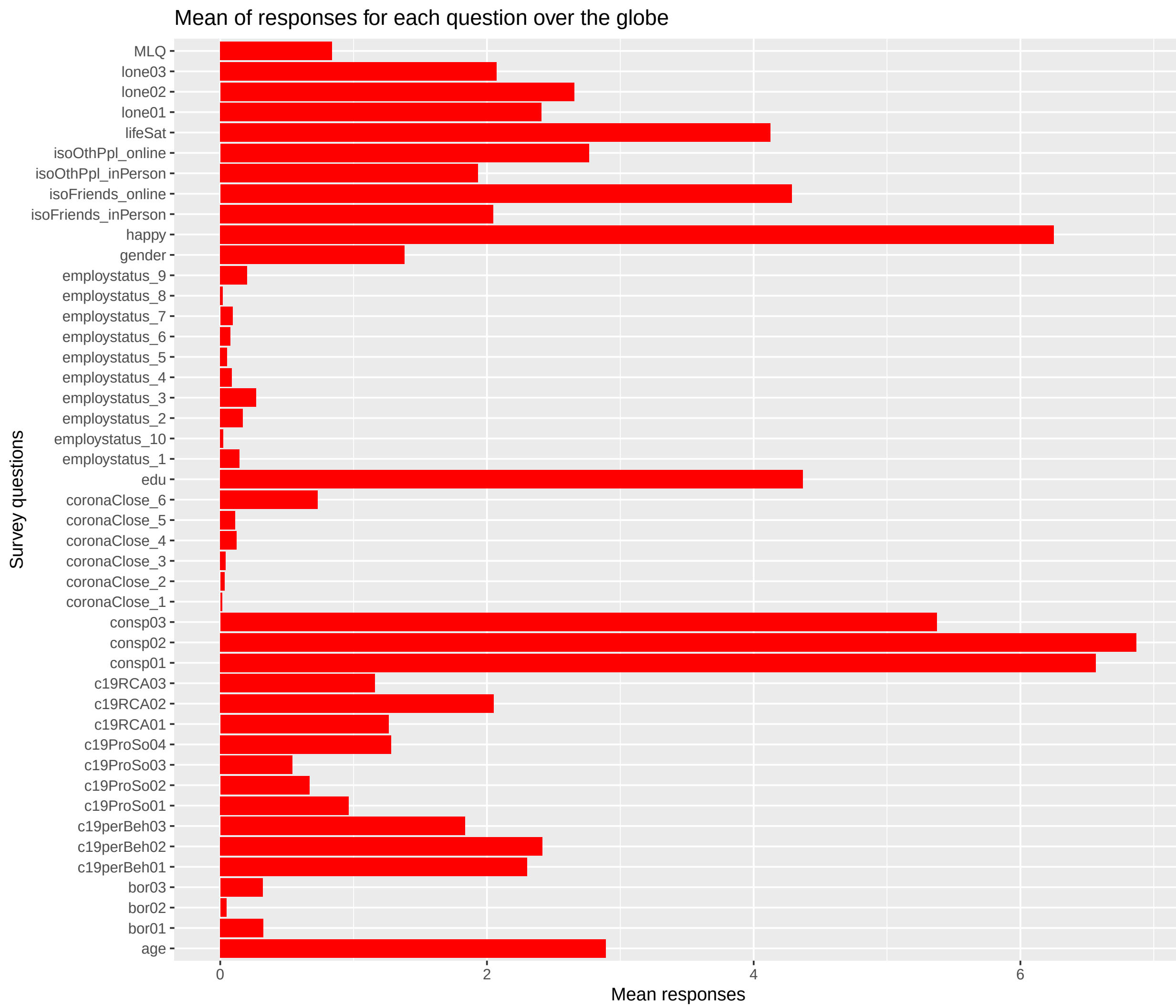
I will be analyzing Croatia in this assignment. We will start by creating bar charts for each group of countries, where the y-axis represents the survey questions, and the x-axis represents the mean of each questions' response. Below, we will create data frames for the mean values utilizing `ggplot2`, excluding non-numerical attributes such as `coded_country`.

```
croatia <- cvbase[cvbase$coded_country == "Croatia", ]
others <- anti_join(cvbase, croatia)

numeric_cols <- sapply(croatia, is.numeric)
means <- colMeans(croatia[, numeric_cols], na.rm = TRUE)

croatia_means <- data.frame(mean = means)

numeric_cols <- sapply(others, is.numeric)
others_means <- colMeans(others[, numeric_cols], na.rm = TRUE)
```

```r
others_means <- data.frame(mean = others_means)

croatia_plotted <- ggplot(croatia_means) +
  geom_bar(mapping = aes(x = rownames(croatia_means), y = mean), stat = "identity",
    fill = "blue") +
  coord_flip() +
  labs(x = "Survey questions", y = "Mean responses",
    title = "Mean of responses for each question in Croatia")

world_plotted <- ggplot(others_means) +
  geom_bar(mapping = aes(x = rownames(others_means), y = mean), stat = "identity",
    fill = "red") +
  coord_flip() +
  labs(x = "Survey questions", y = "Mean responses",
    title = "Mean of responses for each question over the globe")

croatia_plotted
```



```
world_plotted
```



Looking at both graphs of Croatia, and all other countries in comparison, most responses seem to be quite similar, except for Boredom, `bor01`,`bor02` and `bor03`. While the worldwide mean is between 0 (Neither agree nor disagree) and 1 (Somewhat agree), in Croatia, the mean is negative and closer to 1 (Somewhat disagree). This leads us to believe people in Croatia could be less bored than other countries worldwide, albeit slightly. Additionally, there seems to be no Corona Proximity for participants themselves, as well as members of their family in this dataset.

2(b)

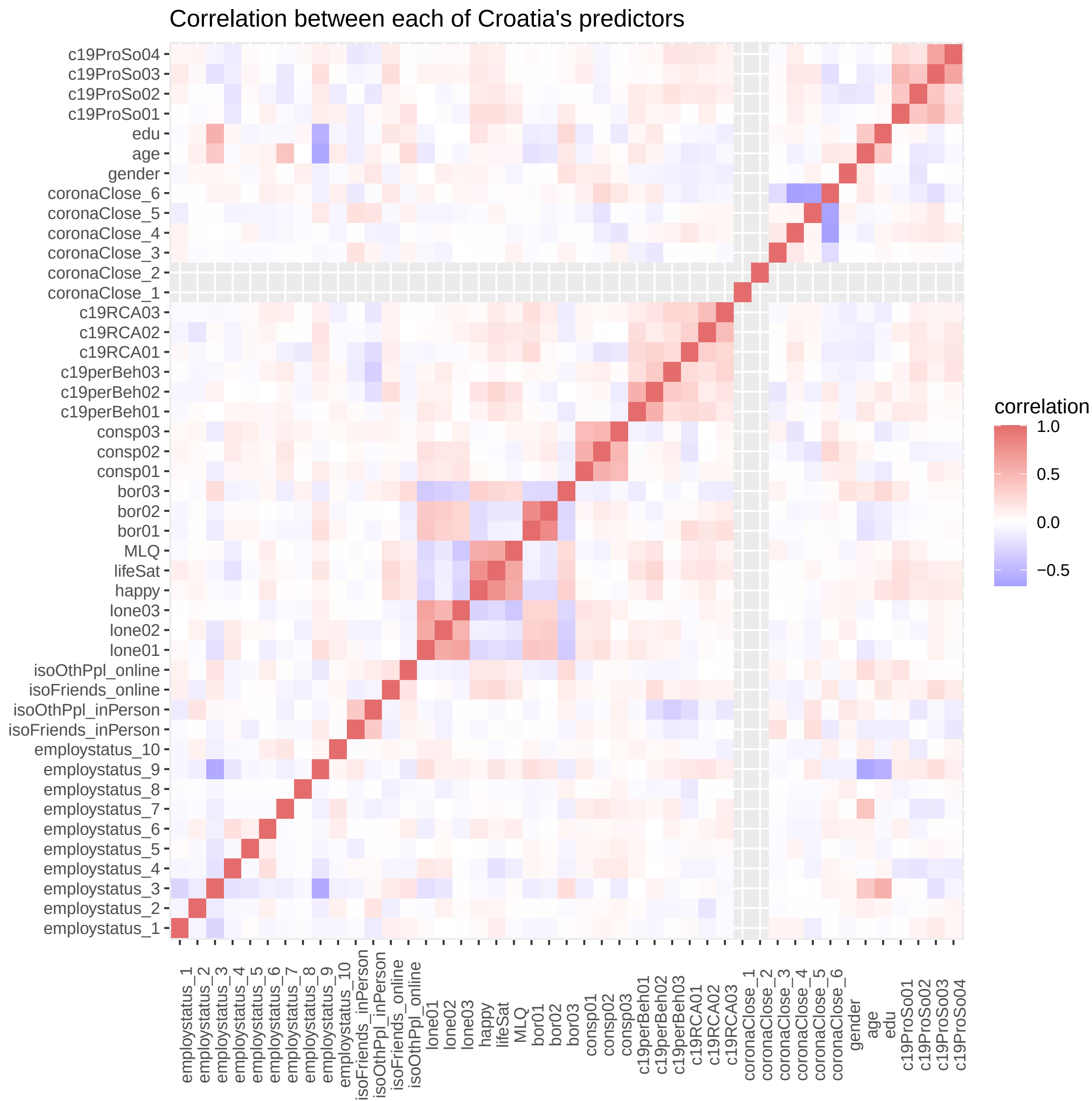
Let's start by taking a peek through the correlation of each predictor for pro-social attitude in Croatia. We can visualise the correlation matrix through a heatmap using `cor()`.

```
numeric_croatia <- croatia[sapply(croatia, is.numeric)]

croatia_correlation <- cor(numeric_croatia, use = "complete.obs")
```

```
## Warning in cor(numeric_croatia, use = "complete.obs"): the standard deviation
```

```
## is zero
```

```
#diag(croatia_cor) <- NA
croatia_melted <- reshape2::melt(croatia_correlation, na.rm = TRUE)

croatia_correlation_plot <- ggplot(data = croatia_melted, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "#6b74ff", mid = "white", high = "#e46c6c", midpoint = 0) +
  labs(title = "Correlation between each of Croatia's predictors", x = "", y = "",
       fill = "correlation") +
  theme(axis.text.x = element_text(angle = 90))

croatia_correlation_plot
```



Tiles that are red or blue denote positive or negative correlation, respectively, and turn white as correlation gets closer to zero. Numerous examples of substantial association between predictors can be seen in this heatmap; nevertheless, the portion displaying the correlation between pro-social views and all other traits is quite weak. This suggests that the characteristics might not be a very good indicator of pro-social sentiments in Croatia.

It is possible to determine how survey replies predict the pro-social attitude question through a linear regression model, fitted for each pro-social attitude based on the qualities. Once fitted, it is possible to find the most accurate predictions.

A linear model of each pro-social attitude versus the qualities is fitted by the code that follows. Each model's R-squared values, significant predictors with p-values less than 0.001, and corresponding coefficients are summarised using a function and a for loop. The vectors `predictors` and `each_model` will be utilised in a subsequent table to compare the strong predictors for every model.

```
predictors <- NULL
each_model <- NULL

model_evaluated <- function(model) {
  rsquared <- summary(model)$r.squared
  adjusted_rsquared <- summary(model)$adj.r.squared
  sig <- which(summary(model)$coefficients[-1, 4] < 0.001) + 1
  predictor <- rownames(summary(model)$coefficients[sig, , drop = FALSE])
  coefficient <- summary(model)$coefficients[sig, 1]

  return(list(rsquared, adjusted_rsquared, predictor, coefficient))
}

fitted_croatia1 <- lm(c19ProSo01 ~ .,
  data = subset(croatia, select = -c(coded_country, c19ProSo02, c19ProSo03, c19ProSo04)))
fitted_croatia2 <- lm(c19ProSo02 ~ .,
  data = subset(croatia, select = -c(coded_country, c19ProSo01, c19ProSo03, c19ProSo04)))
fitted_croatia3 <- lm(c19ProSo03 ~ .,
  data = subset(croatia, select = -c(coded_country, c19ProSo01, c19ProSo02, c19ProSo04)))
fitted_croatia4 <- lm(c19ProSo04 ~ .,
  data = subset(croatia, select = -c(coded_country, c19ProSo01, c19ProSo02, c19ProSo03)))

cat("Pro-social attitudes in Croatia predictors model summary\n\n")
```

```
## Pro-social attitudes in Croatia predictors model summary
```

```
i <- 1
for (model in list(fitted_croatia1, fitted_croatia2, fitted_croatia3, fitted_croatia4)) {
  cat("C19ProSo0", i, "\n", sep = "")
  collection <- model_evaluated(model)
  cat("R-squared:", collection[[1]], "\n")
  cat("R-squared Adjusted:", collection[[2]], "\n")
  cat("99.999% confidence interval significant predictors:\n")
  cat(collection[[3]], "\n")
  cat("Coefficients of predictors:\n")
  cat(collection[[4]], "\n")
  cat("\n")
  for (each in collection[[3]]) {
    each_model <- c(each_model, paste0("Croatia_C19ProSo0", i))
```

```
  }
  predictors <- c(predictors, collection[[3]])
  i <- i + 1
}
```

```
## C19ProSo01
## R-squared: 0.4686053
## R-squared Adjusted: 0.2519858
## 99.999% confidence interval significant predictors:
##
## Coefficients of predictors:
##
##
## C19ProSo02
## R-squared: 0.3355854
## R-squared Adjusted: 0.06474127
## 99.999% confidence interval significant predictors:
##
## Coefficients of predictors:
##
##
## C19ProSo03
## R-squared: 0.4872678
## R-squared Adjusted: 0.278256
## 99.999% confidence interval significant predictors:
## isoFriends_online
## Coefficients of predictors:
## 0.1795645
##
## C19ProSo04
## R-squared: 0.3617789
## R-squared Adjusted: 0.1016124
## 99.999% confidence interval significant predictors:
## rankOrdLife_4D
## Coefficients of predictors:
## 6.462742
```

`C19ProSo04` has the greatest adjusted R-squared at 0.1016124 out of all the models, indicating that the responses best predict it. `disc02`, `MLQ`, `c19NormShould`, and `c19IsPunish` are its best predictors. The `C19ProSo03` model has the least adjusted R-squared value at 0.08190663, and `PLRAC19`, `MLQ`, `c19NormShould`, `trustGovState`, and `edu` are its best predictors.

The fact that the majority of the survey items are deemed subjective makes the models' arguably low R-squared values predictable. For instance, different individuals interpret financial hardship differently and perceive various levels of serenity differently. Since Croatia is a large, populated nation with a wide range of living standards, its several regions are like independent nations with their own economies, healthcare systems, and general levels of satisfaction. Because of this, it is challenging to forecast the pro-social attitude reactions with consistency.

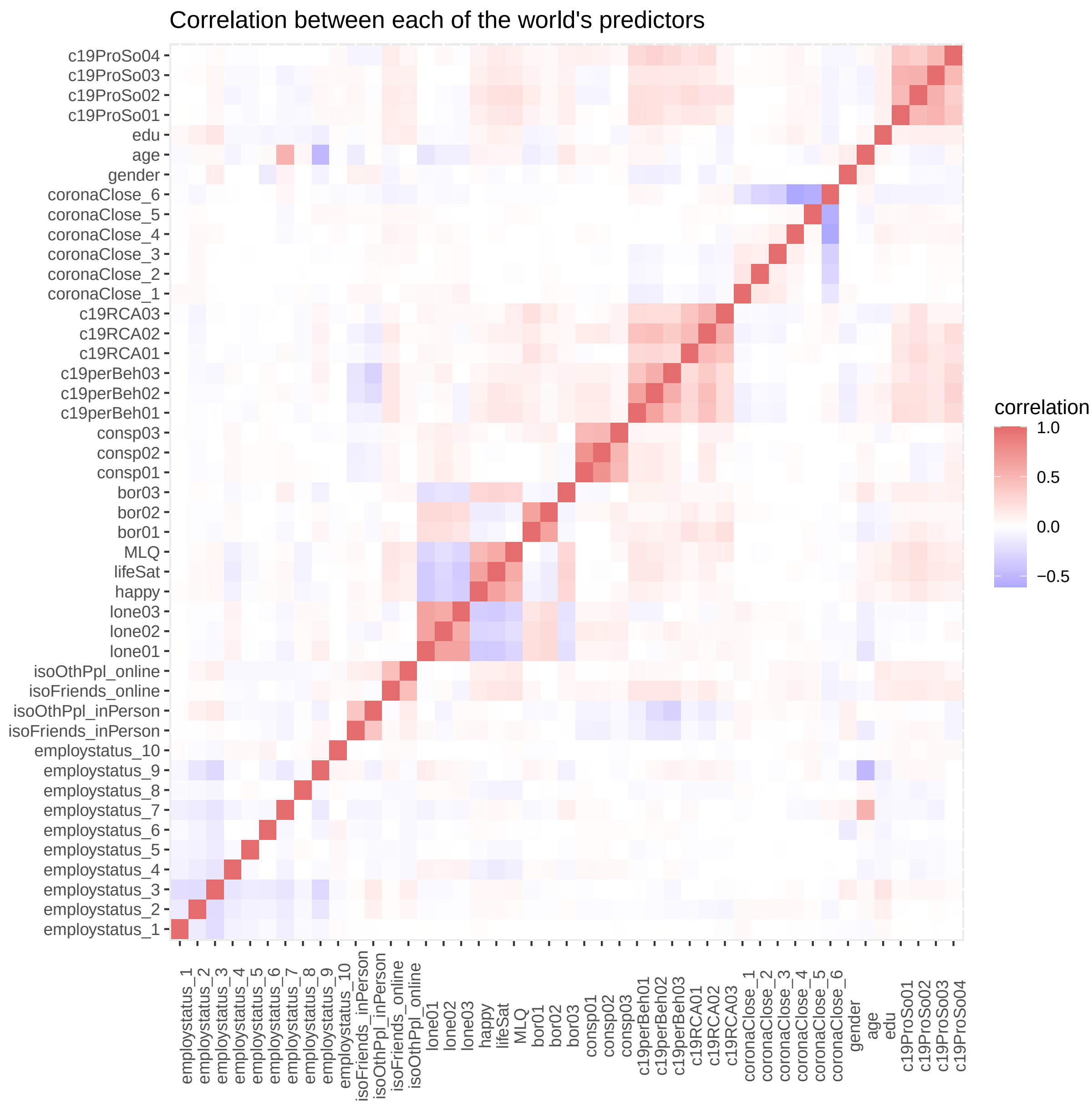
Although each model has a unique set of important predictors, some predictors can be regarded as generally more reliable because they are more frequently found in all of the models. The best illustration would be `c19NormShould`, a highly predictive variable for each of the models. During a pandemic, someone who is willing to assist society would want the best for it and would counsel individuals to isolate themselves and avoid social interactions.. The Centers for Disease Control and Prevention (CDC) in Croatia recommend these steps to stop the spread of viruses, and since Croatia is a developed country with a highly educated populace, people who aspire to be pro-social generally abide by these recommendations. Conversely, a person devoid of pro-social attitudes would not care about adhering to new rules or showing any interest in societal behaviors. Those who disagree with social distancing policies and believe that doing so benefits society as a whole may also have an impact on the predictive power of `c19NormShould`. During the epidemic, lock down protests were prevalent in Croatia, demonstrating the validity of this viewpoint.

`disc02`, `MLQ`, and `trustGovState` are additional variables that predict three of the models effectively. People are more likely to be pro-social if they care about the future of society, if they have a purpose in life, and if they think society can come to an agreement on how to handle the pandemic.

2(c)

The previous code blocks for 2(b) are modified, but with the `rem` data set instead of `croatia`, to produce a similar correlation matrix for the rest of the world. To keep the report concise, variants of code that have been modified further on, can be found in the **Appendix**.

```
global_correlation_plot
```



When we compare the two heatmaps we currently have, we see that `cro_cor_plot` has tiles coloured in a deeper shade of red, which suggests a larger overall correlation between the predictors. Apart from having lighter tiles, `rem_cor_plot` appears "cleaner" due to a reduced dispersion of coloured tiles. However, since the subsections in both plots appear fairly similar, focusing on the heatmap subsections that illustrate the association between pro-social attitudes and all other features, we can speculate that the characteristics for both sets of data should predict pro-social attitudes with somewhat comparable performance.

```
## Pro-social attitudes in the world predictors model summary

## C19ProSo01
## R-squared value: 0.1280411
## Adjusted R-squared value: 0.12646
## Significant predictors with p-value < 0.001:
## employstatus_6 employstatus_7 employstatus_8 employstatus_10 isoFriends_inPerson isoOthPpl_inPerson
```

```
## Coefficients of predictors:
## -0.09356577 -0.2272962 -0.1897203 0.3730428 0.01719855 0.02667376 0.01609172 0.01721033 0.0606518 -0
##
## C19ProSo02
## R-squared value: 0.1766909
## Adjusted R-squared value: 0.1751979
## Significant predictors with p-value < 0.001:
## employstatus_3 employstatus_4 employstatus_5 employstatus_8 employstatus_10 isoFriends_inPerson isoF
## Coefficients of predictors:
## 0.1072481 -0.1977689 -0.1317748 -0.3161566 0.2091906 0.03416122 0.0152765 0.02820841 0.05811839 -0.0
##
## C19ProSo03
## R-squared value: 0.1199772
## Adjusted R-squared value: 0.1183814
## Significant predictors with p-value < 0.001:
## employstatus_3 employstatus_7 employstatus_10 isoFriends_inPerson isoOthPpl_inPerson isoOthPpl_onlin
## Coefficients of predictors:
## 0.1555728 -0.2152103 0.3369801 0.02274039 0.02160123 0.02621581 0.066345 0.09423445 0.0588311 0.0282
##
## C19ProSo04
## R-squared value: 0.17178
## Adjusted R-squared value: 0.1702782
## Significant predictors with p-value < 0.001:
## employstatus_2 employstatus_3 employstatus_10 isoFriends_online lone02 lone03 lifeSat MLQ bor02 bor0
## Coefficients of predictors:
## 0.09869336 0.1272541 0.2820308 0.01248981 0.03724033 0.03827886 0.07587075 0.02828067 0.02557799 0.0
```

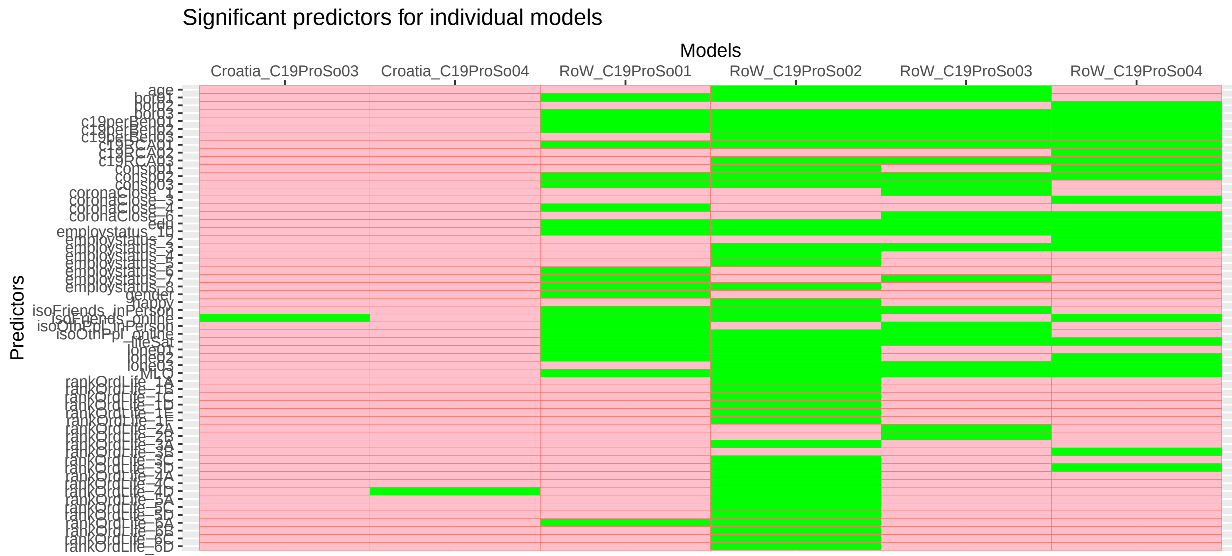
All four models had adjusted R-squared somewhere between 0.12 and 0.17, pointing to a smaller range than the comparable one for the Croatia data set (0.06 - 0.49), according to the summary for the rest of the world. The models' predictors are significantly more important than those of the `croatia` models. The four models are well predicted by `disc02`, `lifeSat`, `c19NormShould`, `c19NormDo`, and `trustGovState`. The majority of the predictors that performed well in all four of the `croatia` models, `c19NormShould`, `disc02`, and `trustGovState` are included in this set. As was already said, Croatia's sheer vastness and diversity make it seem like a collection of independent nations. It follows that strong predictors for Croatia would also apply to other nations collectively.

The table below displays the results of the strongest predictors for each pro-social behaviour for Croatia and the rest of the globe.

TODO LATER!!!! fix columnds missing

```r
tables_combined <- table(predictors = predictors, models = each_model)

tables_combined <- tables_combined[, c("Croatia_C19ProSo03",
  "Croatia_C19ProSo04", "RoW_C19ProSo01", "RoW_C19ProSo02", "RoW_C19ProSo03",
  "RoW_C19ProSo04")]

tables_combined_plotted <- ggplot(data = as.data.frame(tables_combined)) +
  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 = "Significant predictors for individual models")
```

```
tables_combined_plotted
```



Task 3: Focus country vs cluster of similar countries

3(a)

Some additional data is collected from external sources to supplement the indicators available in the sources mentioned in the references. Eight indicators make up the final data table I have compiled (in **Appendix**) for use in clustering: `HDI`, `GHS`, `freedom`, `political_stability`, `happiness`, `total_vax_per_hundred`, `total_cases_per_mil`, and `total_deaths_per_mil`. The **Appendix** contains information and explanations regarding each indicator as well as its sources.

Using k-means clustering, nations that are comparable to Croatia are found. For the `kmeans()` function to function, countries with NA values must first be eliminated. This has no bearing on our findings because the majority of these nations—such as Afghanistan and Syria—do not initially appear in the baseline data and differ greatly from Croatia in terms of development and data transparency. After scaling the data, we use a few random beginnings to carry out K-means clustering.

```r
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 == "Croatia")$cluster
similar <- filter(clusters, cluster == target)
similar
```

```
##      country cluster
## 42   Croatia      24
## 76   Hungary      24
## 137   Poland      24
```

```
## 140  Romania      24
## 155 Slovakia      24
```

TODO: Emit croatia from above

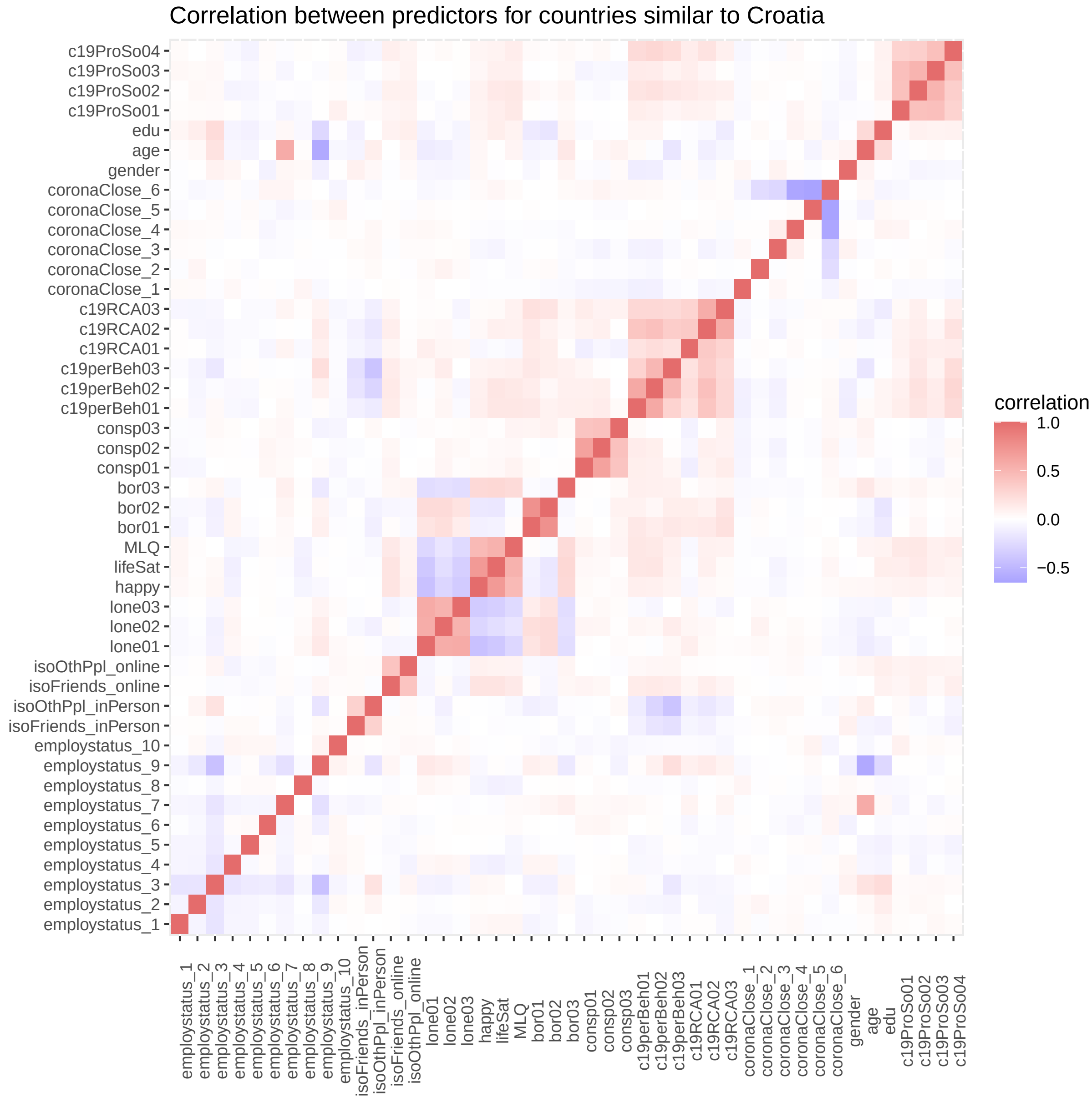
The clustering indicates that Hungary, Poland, Romania and Slovakia are comparable to Croatia.

3(b)

After removing the data from Croatia, the baseline data of the cluster's member nations are initially extracted by inner joining `cvbase` and `similar`. Finally, we visualize the given correlation matrix, in line with `croatia` and the rest of the globe.

```r
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 != "Croatia")

non_numeric_columns <- sapply(clus, function(x) !is.numeric(x))
clus_numeric <- subset(clus, select = !non_numeric_columns)

clus_cor <- cor(clus_numeric, use = "complete.obs")
clus_melted <- reshape2::melt(clus_cor)
```

```r
clus_cor_plot
```



This heat map's scatter of colored tiles is identical to the Croatia heat map's, showing how similar these two nations are. In comparison to the previous plots, the portion of tiles displaying the correlation between predictors and pro-social views is generally darker, suggesting the predictors related to the cluster in question may perform positively in terms of prediction than the data from the previous two groups.

Finally, repeating the code fragments from Task 2, b and c, we can output a structured summary for all models in order to determine how respondent answers can be a significant predictor of pro-social views pertaining to the similar countries cluster in question.

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

```
## C19ProSo01
## R-squared value: 0.1019308
## Adjusted R-squared value: 0.07645638
```

```
## Significant predictors with p-value < 0.001:
## employstatus_10 MLQ rankOrdLife_3B c19RCA01
## Coefficients of predictors:
## 0.6937677 0.1364176 -0.6291484 0.05595581
##
## C19ProSo02
## R-squared value: 0.1275736
## Adjusted R-squared value: 0.1028265
## Significant predictors with p-value < 0.001:
## MLQ c19RCA01
## Coefficients of predictors:
## 0.1080792 0.07625105
##
## C19ProSo03
## R-squared value: 0.1031383
## Adjusted R-squared value: 0.07769807
## Significant predictors with p-value < 0.001:
## MLQ rankOrdLife_3C c19RCA01
## Coefficients of predictors:
## 0.09845237 -0.5655349 0.08508227
##
## C19ProSo04
## R-squared value: 0.1595349
## Adjusted R-squared value: 0.1356944
## Significant predictors with p-value < 0.001:
## MLQ rankOrdLife_1B rankOrdLife_1C rankOrdLife_1D rankOrdLife_1F rankOrdLife_3B c19perBeh01 c19perBeh0
## Coefficients of predictors:
## 0.0850501 1.498947 1.385919 1.365358 1.5909 -0.68414 0.1338745 0.1736953 0.1314914
```

Also check last sentence of next paragraph

Based on the results, the models for these comparable nations typically have adjusted R-squared values that are comparable to those of Croatia and all other nations combined. Similar to the Croatia models, `C19ProSo04` seems to have the greatest adjusted R-squared value (0.1356944). Except for `C19ProSo04`, which has significant predictors including `disc02`, `MLQ`, `c19NormShould` and `c19IsPunish`, none of these models, in contrast to the preceding eight models, had significant predictors with p-values less than 0.001. In the Croatia model, `disc02` is also a very good predictor for `C19ProSo04`, but not for `PFS02`. However, `disc02`, `MLQ`, `c19NormShould` and `c19IsPunish` are all significant predictors in the rest-of-the-world model for `C19ProSo04`.

Consequently, the predictive performance of qualities for this cluster of countries is not significantly better than that of Croatia or the rest of the globe, with comparable R-squared values and predictors with generally greater p-values. The previously reported large correlation may have been due to chance or small sample size rather than representing genuine statistically significant links between attribute and pro-social attitude.

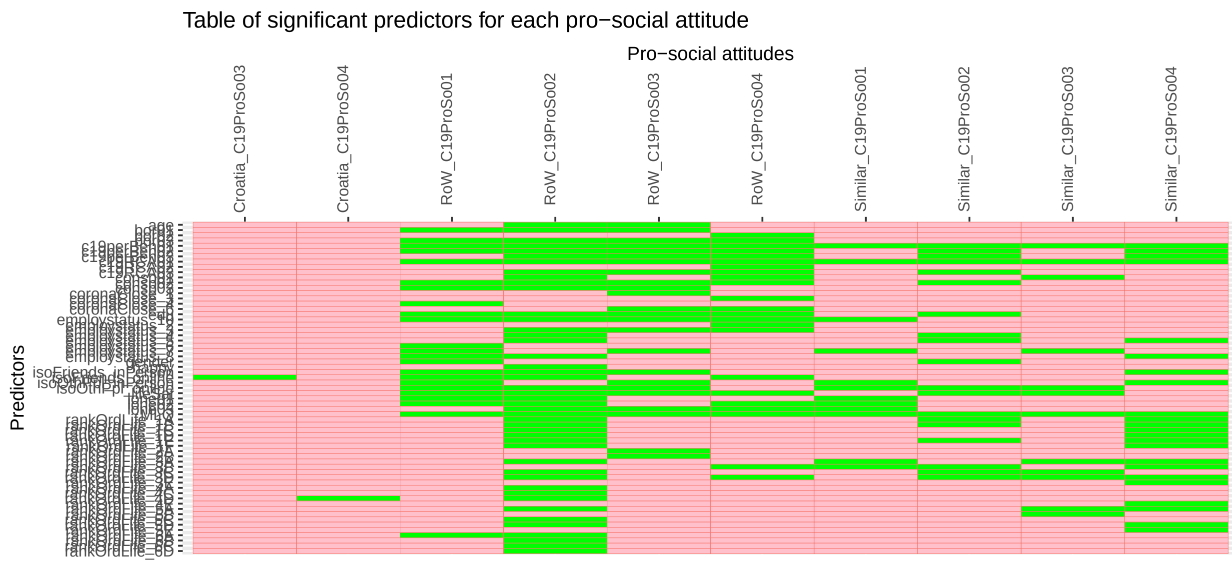
We can define a strong predictor in relation to a model's total p-values for comparative purposes. For these novel cluster models, we characterize a strong predictor as one with a p-value of less than 0.05, which is a widely accepted threshold. A new visualization table is constructed and the `model_eval` function is changed to reflect this (see **Appendix**).

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

## C19ProSo01
## R-squared value: 0.1019308
## Adjusted R-squared value: 0.07645638
```

```
## Significant predictors with p-value < 0.05:
## employstatus_7 employstatus_10 isoOthPpl_inPerson isoOthPpl_online lone01 lone02 lone03 MLQ rankOrdLi
## Coefficients of predictors:
## -0.4443775 0.6937677 0.03384447 0.033907 0.06722975 -0.0930139 0.09173 0.1364176 -0.5174174 -0.62914
##
## C19ProSo02
## R-squared value: 0.1275736
## Adjusted R-squared value: 0.1028265
## Significant predictors with p-value < 0.05:
## employstatus_4 employstatus_5 isoOthPpl_online lifeSat MLQ consp02 rankOrdLife_1A rankOrdLife_1B ran
## Coefficients of predictors:
## -0.2919775 -0.3600833 0.03217282 0.1372649 0.1080792 -0.04087629 1.069375 1.125655 0.8700702 -0.4339
##
## C19ProSo03
## R-squared value: 0.1031383
## Adjusted R-squared value: 0.07769807
## Significant predictors with p-value < 0.05:
## employstatus_7 isoOthPpl_online lifeSat MLQ consp01 rankOrdLife_3A rankOrdLife_3C rankOrdLife_3D ran
## Coefficients of predictors:
## -0.5393024 0.03719592 0.09921346 0.09845237 -0.04105437 -0.5538729 -0.5655349 -0.3214164 -0.4687776
##
## C19ProSo04
## R-squared value: 0.1595349
## Adjusted R-squared value: 0.1356944
## Significant predictors with p-value < 0.05:
## employstatus_5 employstatus_8 isoFriends_inPerson isoOthPpl_inPerson MLQ rankOrdLife_1A rankOrdLife_
## Coefficients of predictors:
## -0.4926586 0.5835967 -0.02775915 0.03682043 0.0850501 1.325203 1.498947 1.385919 1.365358 1.238381 1
```

Omitted “Croatia_C19ProSo01”, “Croatia_C19ProSo02” and similar

```
summ_table_vis_2
```



We see that the distribution of strong predictors between the models of similar countries is more similar to the Croatian models, with a few shared significant predictors (i.e., appearing quite as “sparse” as the Croatia models). Given more comparable p-values, the models from the group of all other nations have a great deal more significant predictors in common with the Croatian models. These models do, however, also contain a large number of powerful predictors that are absent from the Croatian models. Consequently, the group of comparable nations may provide a better fit to the critical characteristics needed to predict pro-social sentiments. The higher p-values and fewer shared common strong predictors included in their models may become insignificant with additional study or a larger sample size.

One explanation could be that, despite their similarities to Croatia, each country in the cluster varies slightly from the others in terms of socioeconomic factors that are not taken into account by the clustering indicators.. Their collective performance in forecasting pro-social attitudes differs substantially from Croatia alone when these small variations are taken into account. However, because Croatia's politics, culture, and other aspects of society are complicated, much like those of a group of many countries, its models have many strong predictors in common with the models of the group of all countries. The models may report multiple significant predictors that are not important in reality because the group of all other countries may be too large and complex.

References

ChatGPT was used to help model some code fragments in this assignment, including 2a, 2b, 2c, 3a and 3b. Grammarly was used to assist with spell checking and grammar use in all parts of this assignment.

Prompts used, include Fix any grammar or spelling mistakes in this paragraph Write R code to use ggplot to plot a correlation matrix for the entire world from a data set

However, the output from generative AI was not the final text that appears in any case in this assignment, as it has been checked, amended and modified by the submitting student.

Appendix

`cvbase` head, 1(b)

```r
head(cvbase)
```

```
##       employstatus_1 employstatus_2 employstatus_3 employstatus_4
## 21475              0              1              0              0
## 3823               0              0              0              1
## 57606              0              1              0              0
## 49874              0              0              0              0
## 16726              1              1              1              0
## 23113              0              0              0              0
##       employstatus_5 employstatus_6 employstatus_7 employstatus_8
## 21475              0              0              0              0
## 3823               0              0              0              0
## 57606              0              0              0              0
## 49874              0              0              1              0
## 16726              0              0              0              0
## 23113              0              0              1              0
##       employstatus_9 employstatus_10 isoFriends_inPerson isoOthPpl_inPerson
## 21475              0               0                   0                  2
## 3823               0               0                   1                  0
## 57606              0               0                   1                  2
## 49874              0               0                   1                  1
## 16726              0               0                   3                  2
```

```
## 23113                0               0                  0               0
##        isoFriends_online isoOthPpl_online lone01 lone02 lone03 happy lifeSat MLQ
## 21475                  6                1      1      1      1     9       5   2
## 3823                   6                0      3      3      2     6       2   1
## 57606                  4                4      2      2      1     7       4   1
## 49874                  7                3      2      3      2     8       5   0
## 16726                  7                4      3      3      4     9       5   3
## 23113                  7                0      4      5      5     4       2   0
##        bor01 bor02 bor03 consp01 consp02 consp03 rankOrdLife_1 rankOrdLife_2
## 21475      1     1    -2       7       8       0             F             D
## 3823      -2    -1     1       6       1       2             C             E
## 57606      0     0     0       8       8       2             D             E
## 49874     -2    -2     3      10       8       0             F             D
## 16726      2     1     3       6       8       6             0             0
## 23113      0     0    -3      10      10       7             E             D
##        rankOrdLife_3 rankOrdLife_4 rankOrdLife_5 rankOrdLife_6 c19perBeh01
## 21475              E             C             A             B           3
## 3823               F             D             A             B           3
## 57606              F             C             A             B           3
## 49874              E             C             A             B           3
## 16726              0             0             0             0           3
## 23113              F             B             A             C           3
##        c19perBeh02 c19perBeh03 c19RCA01 c19RCA02 c19RCA03 coronaClose_1
## 21475            3           3        0        0       -2             0
## 3823             2           2        0        3        2             0
## 57606            3           1        1        2       -1             0
## 49874            3           3        0        3        1             0
## 16726            3           3        3        3        3             0
## 23113            3           3       -3       -3       -3             0
##        coronaClose_2 coronaClose_3 coronaClose_4 coronaClose_5 coronaClose_6
## 21475              1             0             1             1             0
## 3823               0             0             0             0             1
## 57606              0             0             0             0             1
## 49874              0             0             0             0             1
## 16726              1             0             0             0             0
## 23113              0             0             0             0             1
##        gender age edu coded_country c19ProSo01 c19ProSo02 c19ProSo03 c19ProSo04
## 21475       1   3   5         Spain          2          2         -1          3
## 3823        1   3   6        Greece          1         -1          2          2
## 57606       2   4   4       Germany          2          2          1          1
## 49874       1   5   4        Canada         -2          1          1          3
## 16726       1   2   6     Indonesia          3          3          3          3
## 23113       2   6   5        France          0         -3         -3          0
```

rem Correlation matrix, 2(c)

```
others <- anti_join(cvbase, croatia)

numeric_others <- others[sapply(others, is.numeric)]

global_correlation <- cor(numeric_others, use = "complete.obs")

global_melted <- reshape2::melt(global_correlation)
```

```
global_correlation_plot <- ggplot(data = global_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 each of the world's predictors", x = "", y = "",
    fill = "correlation") +
  theme(axis.text.x = element_text(angle = 90))
```

`rem` models, 2(c)

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

cat("Pro-social attitudes in the world predictors model summary\n\n")
res_counter <- 1
for (model in list(fitted_rem1, fitted_rem2, fitted_rem3, fitted_rem4)) {
  cat("C19ProSo0", res_counter, "\n", sep = "")
  res <- model_evaluated(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]]) {
    each_model <- c(each_model, paste0("RoW_C19ProSo0", res_counter))
  }
  predictors <- c(predictors, res[[3]])
  res_counter <- res_counter + 1
}
```

Clustering data, 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
```

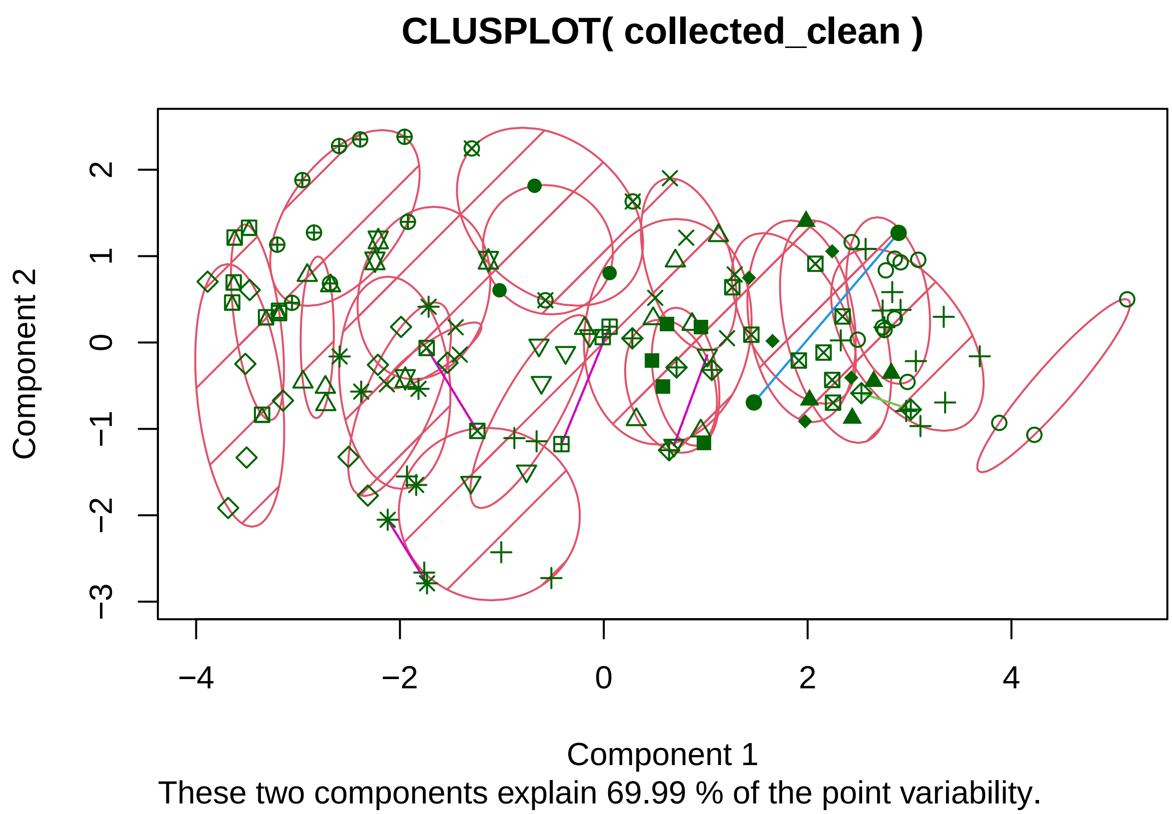
Clustering indicators, 3(a) ##### TODO: WARNING!!!!!!!!########## - `HDI`: Human Development Index (2021); a value between 0 and 1 that measures average achievement in human development based on three dimensions - life expectancy, education and standard of living. (Source: Human Development Reports) - `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.

K-means clustering, 3(a)

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



`rem` correlation matrix, 3(b)

```r
# 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 Croatia",
```

```
#     x = "", y = "", fill = "correlation") +
#   theme(axis.text.x = element_text(angle = 90))
```

`clus` model, 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
# }
```

`clus` models, updated `model_eval` function with p-value less than 0.05, 3(b)

```
# 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]]) {
#     each_model <- c(each_model, paste0("Similar C19ProSo0", counter))
#   }
#   predictors <- c(predictors, res[[3]])
#   counter <- counter + 1
# }
```

`croatia`, `rem` and `clus` models strongest predictors, 3(b)

```
# summ_table_2 <- table(predictors = predictors, models = each_model)
# summ_table_2 <- summ_table_2[, c("Croatia_C19ProSo01", "Croatia_C19ProSo02", "Croatia_C19ProSo03",
#   "Croatia_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))
```