FIT3152 Assignment 1
Semester 1 - 2024

Ang Ke Er | 32581343

1. Descriptive analysis of the data

Generative AI was not used in this assignment.

The original given dataset has 57942 rows and 52 columns. After generating a random sample, the dataset has 40000 rows and 52 columns. For the data types, all the variables are integer types except coded_country and rankOrdLife which are character types.

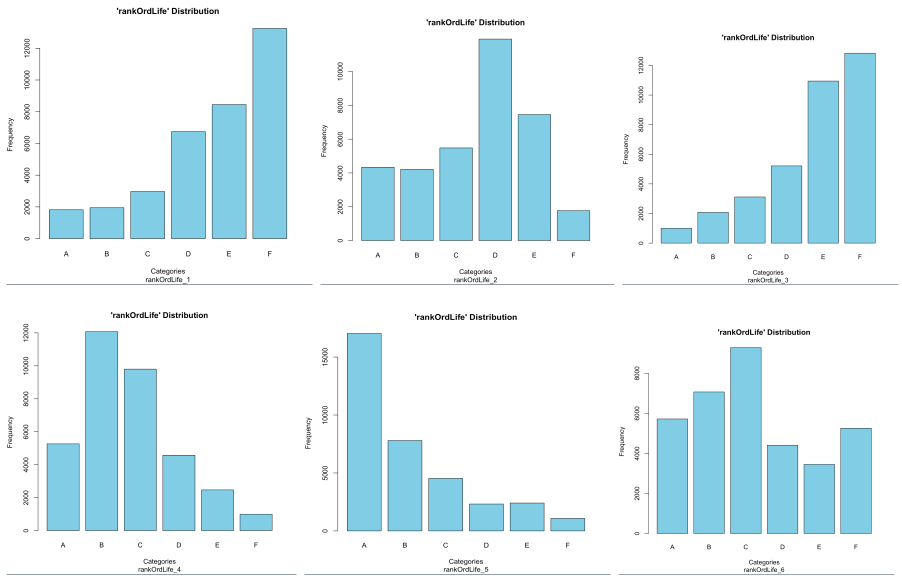
The column included are employstatus_1,employstatus_2, employstatus_3, employstatus_4, employstatus_5, employstatus_6, employstatus_7, employstatus_8, employstatus_9, employstatus_10, isoFriends_inPerson, isoOthPpl_inPerson, isoFriends_online, isoOthPpl_online, lone01,lone02, lone03, happy, lifeSat, MLQ, bor01, bor02, bor03, consp01, consp02, consp03, rankOrdLife_1, rankOrdLife_2,rankOrdLife_3 rankOrdLife_4 rankOrdLife_5 rankOrdLife_6, c19perBeh01, c19perBeh02, c19perBeh03, c19RCA01, c19RCA02, c19RCA03, coronaClose_1, coronaClose_2, coronaClose_3, coronaClose_4, coronaClose_5, coronaClose_6, gender age ,edu, coded_country, c19ProSo01, c19ProSo02, c19ProSo03, c19ProSo04.

After checking NA values in all columns, it is observed that all columns contain NA values except country. In employstatus and coronaClose columns, a large amount of NA values are observed, this is mostly due to the fact that the data collection is designed as a yes/no question. The frequency of "NA" values in other columns, which reflects the complexity of survey responses, may be the consequence of participants being uncertain, uncomfortable, or encountering other obstacles to responding.

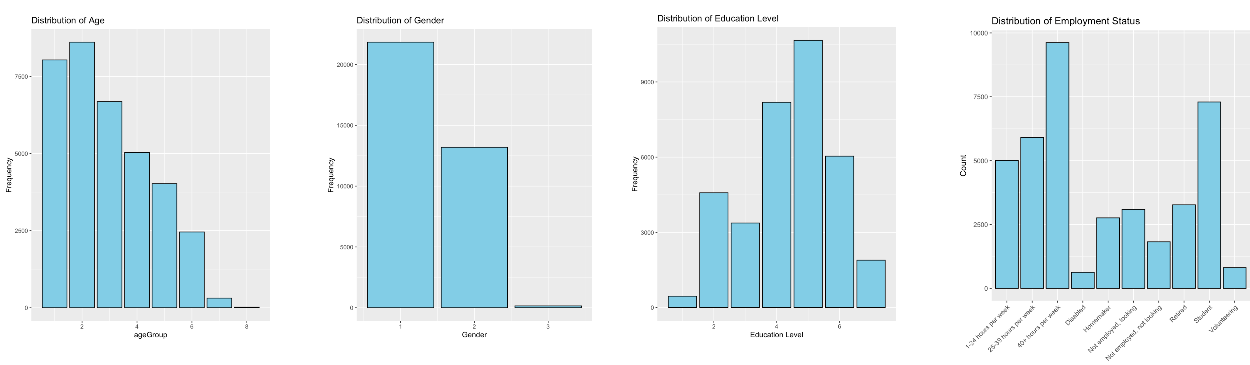
For coded_country, there are 108 unique values. However, one of them is “” which can be considered as missing value and there are 137 of them. More analysis of the dataset reveals an interesting trend in the coded_country, among the other 107 values, the United States of America has shown significantly higher frequency occurrences which is 7013 as it substantially exceeds the second highest occurrences which is Spain, totalling 2008. There are also countries where all of which record a single instance, those countries are Andorra, Azerbaijan, Botswana, Cameroon, Ethiopia, Laos, Mongolia, Myanmar, Nepal, Oman , Slovenia and Uzbekistan.

For rankOrdLife columns, all of them have 7 unique values and including NA as missing response. Ranging from rankOrdLife_1 to rankOrdLife_6, the characters in each cell, which range from A to F, represent different values that people value most in life: A is for Beauty; B is for Achievement; C is for Victory; D is for Friendship; E is for Love; and F is for Empathy. The purpose of these columns is to record the values' rank ordering based on how important they are to each participant, with rankOrdLife1 denoting the highest priority and rankOrdLife6 the lowest priority.

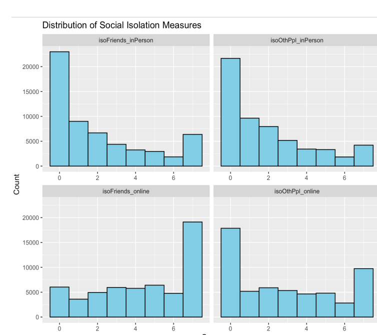
With a deeper analysis, we can see that in rankOrdLife_1, F has the highest occurrence which is 14268 times; in rankOrdLife_2, D has the highest occurrence, having 12617 times; in rankOrdLife_3, it is interesting to see that F also has the highest number which is 13512, which is same as rankOrdLife_1 but number of occurrence of E is close to F which is 11726; in rankOrdLife_4, B has the highest occurrence; in rankOrdLife_5, A has the highest. Last, in rankOrdLife_6, C has the highest occurrence. In general, people value empathy the most, then friendship, love, achievement, beauty, and victory the least.

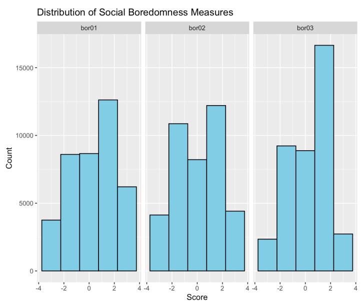


The data uses a numerical coding method in the "gender", “age” and “edu” columns to indicate the various categories. The values in gender are 1,2,3 and in age, the value is ranging from 1 to 8, each number represents a different age group. For education, the values range from 1 to 7 each represent a different education level.For the age group, the graph shows a right-skewed distribution, having group 2, which is people around 25-34 years old, has the most population. In terms of gender, group 1 has almost half more than group 2, which indicates that in the survey group, female is more than male. For education level, group 5 has the most occurrence, showing that in the survey group, most people are holding a bachelor degree. Regarding employment status, from the graph, we can see that most of the people who did the survey work 40+ per week, followed by students; also, having a small group of disabled persons and volunteers.

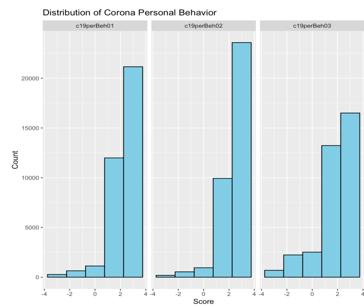


For isolation online and offline, from the figure below, we can see that most of the participants did not contact any friends or other people in person, as shown by the highest occurrence of 0 in both isoFriends_inPerson and isoOthppl_inPerson graphs. However, in isoFriends_online, with 7 being the highest occurrence, we can tell they contacted their friends more online. Also, comparing isoOthPpl_inPerson and isoOthPpl_online, we can observe that they contacted other people online more too.

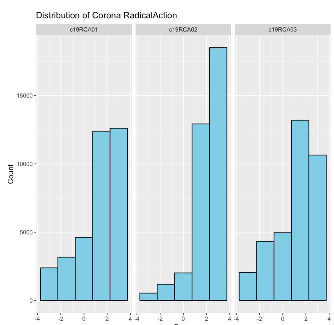




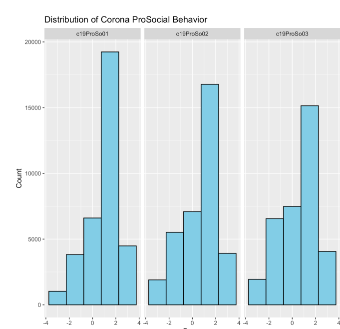
For boredomness attributes , in bor01 and bor02, the graphs show an almost symmetrical distribution. For bor01, the graph shows a slight lean towards the positive side, which shows that more respondents do wish time could go faster. While for bor2, a somewhat even distribution between negative and positive side, about the balanced amount of people think that time is moving very slow or the opposite. For bor03, the distribution clearing skewed to the left, which indicates most people felt they have control over their time.



Looking at the distribution of corona personal behaviour graphs. We can tell that the participants adopt correct measures in preventing spreads of coronavirus. As observation, c19proBeh01 and c19proBeh02 show a clear distribution of skewed towards high score, having score 3 as mode, suggesting that a high degree of adherence to advised procedures, which are wash hands frequently, avoid crowded place and following the self quarantine rules.



For Corona Radical Action, the respondents also show a strong agreement in the questionnaire, as all the distributions are skewed towards the highest score. In this case, c19RCA2 exhibits the strongest left-skewness, with the majority of answers scoring three. This suggests that most people agree in doing mandatory vaccination once it is launched, enforcing quarantine for those infected people and their close contacts and reporting for suspicious cases. These suggest that participants generally support vigorous public health interventions for coronavirus.



The results of the Corona Pro Social Behaviour survey show that people in society have a respectable attitude towards supporting one another during a pandemic. In this 3 graph, strong pro-social behaviours are indicated by a concentration of responses at score 3. This score reflects that people are willing in helping coronavirus victims, making donations for them, and taking personal risks to stop the virus's spread and safeguard vulnerable populations. This indicates there is a general inclination in the community to act in a selfless and helpful manner during the crisis.

To have a more detailed summary of the data, I use the skim function in the skimr library to obtain the information, and the result is shown in the appendix. Besides, I also create a heat map to visualise the relationship between the numerical variables.

2. Pre-processing or data manipulation

There is some preprocessing required for the analysis. As the data contains NA values, I have tackled this in 3 steps. First, the NA values in employstatus and coronaClose have been replaced with 0, this has adopted the binary coding system where 1 represents yes and 0 represents no. Second, I replaced the “” in coded_country with NA values as it represents nothing. Then, I dropped all the rows where there are NA values in order to maintain data integrity and prevent the following analysis impacted by the NA values. Besides, I have also factorised the chr type attributes before doing analysis.

3. Focus country vs all other countries as a group

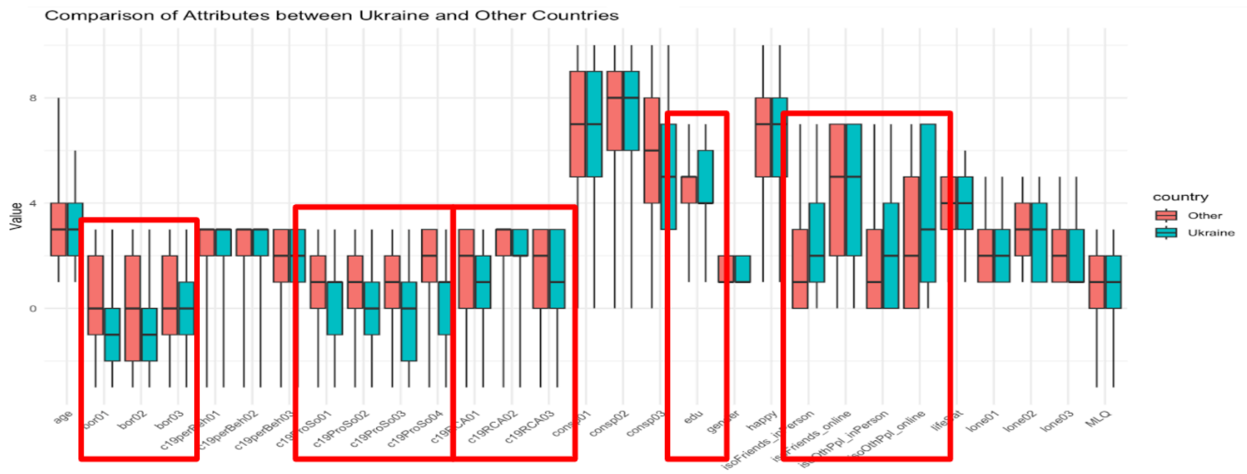
In order to make a comparison between Ukraine and other countries as a group, I have created a boxplot for the visualisation. So from the graph, we can see that there are some differences between the attributes.

First of all, in bor01 and bor02, Ukraine has lower median value than other countries as a whole, meaning the people in Ukraine felt less bored than other countries as a whole. And the variance in bor01, bor02, bor 03 in ukraine are lesser than other countries, as their IQR is smaller, meaning that people in Ukraine have more consistent feeling on the boredomness.

It is notable that in c19ProSo01 ,Ukraine shows a clear negative skewed distribution, showing that people in general have more agreement in helping people who suffer from coronavirus. However, in c19ProSo02, c19ProSo03 and c19ProSo04, Ukraine’s median approaches the lower boundary of the box representing other countries in the box plots , indicating a significant difference of opinion in both groups. More specifically, the lower medians show that people in Ukraine are generally less willing to make personal sacrifices to prevent spread of coronavirus. This discrepancy can be the result of certain social, cultural, or economic elements affecting the mindset of Ukrainians.

For Corona Radical Action, we can also see there are notable differences between both groups. In metrics of C19RCA01, C19RCA02 and C19RCA03, Ukraine shows a lower median score. This implies that, on the whole, Ukrainians are less inclined to agree or be prepared to take the preventive measures against COVID-19, such as mandatory vaccination, quarantine or report suspicious cases. In addition, Ukraine's median educational attainment is lower than that of other countries, suggesting that Ukrainians generally obtain less schooling.

We can also observe that in the attributes of isoFriends_inPerson, isoOthPpl_inPerson and isoOthPpl_online, Ukraine has a higher median score compared to the other group. This suggests that people in Ukraine communicate with friends and other people more regularly, both in person and online. Since more frequent social contacts may lessen emotions of loneliness and boredom, this may help to explain the observed lower levels of boredom among Ukrainians. However, a deeper analysis would be needed to explain this causal relationship.



4. Performance of attributes predicting pro-social attitudes for Ukraine

In order to see how well the pro-social attitudes has been predicted by the participants' responses, I have built a series of linear regression models. I use all the attributes besides coded_country as independent variables (predictors) and dependent variables are the pro-social attitudes (c19proSo01, c19proSo02, c19proSo03, c19proSo04) . After that, I try to identify the best predictors of each model by inspecting the p-values of each predictor and evaluate the performance of the model by observing the values of overall p-value, residuals and R-squared.

For the first model (c19proSo01), MLQ, consp03, c19RCA01 are the best predictors. Besides, the p value for the overall regression model is 1.14e-11(<0.05) Looking at the residuals distribution of the model, we can observe that it is approximately symmetrical. However, the R^2 for the model is low, which is only 0.1943.

For the second model (c19proSo02), c19RCA01 is the best predictor. For the overall model, p-value is 4.154e-14 (<0.05) for the significance and fairly symmetrical distributed residuals. Regarding R^2, this model still has a low R^2 but did slightly better than the previous one, which is 0.2113, still showing limited capability for explanation.

For the third model (c19proSo03) , the best predictor is consp03 with the smallest p-value. This model is similar to the previous 2 models, having a low p-value of 3.182e-11, nearly symmetrical residuals and low R^2 which is 0.1911.

For the fourth model (c19proSo04) ,rankOrdLife_3F is the best predictor. This model has a p-value of 1.099e-10 , approximately symmetrical residuals and R^2 which is 0.1871.

Overall, the low p-values of the regression models show statistical significance and the distribution of residuals show that there is a consistent performance for all of the models. Nevertheless, the consistently low R-squared values of all the models indicate that the models' ability to predict prosocial behaviour under study is restricted, as they are ranging from 0.18 to 0.21, showing that the model can only explain 18-21% of the data. This suggests that the model has to be improved or that other variables should be taken into account in order to fully describe the complexity of pro-social behaviours in the focus country. Another observation is that in all the models, rankOrdLife_6 shows a NA value, meaning that the variables are not used at all in the model, which might need some additional studies

5. Performance of attributes predicting pro-social attitudes for other countries as a group

I re-fitted the model again for all countries as a group.

For the 1st model (C19proSo01), It shows that MLQ, bor03, c19perBeh01, c19RCA01 are the strongest predictors as they have extremely low p-values (< 2e-16). The p-value in the overall model is low, having a value less than 2.2e-16 , showing statistical significance of the model. And also an approximate distribution of residuals and low R^2 which is 0.1178.

For model 2 (C19proSo02), the strongest predictors are isoFriends_inPerson, lifeSat, MLQ ,bor01, edu, c19perBeh01 , c19RCA01, c19RCA03. Besides, the p-value in the overall model is also low, having a value less than 2.2e-16. The distribution of the residuals is fairly symmetrical and R^2 is 0.1672 which is better than model 1 but still has limited effectiveness .

For model 3 (C19proSo03), the strongest predictors are lifeSat , consp02, edu, c19perBeh01, c19RCA01. The p-value in is less than 2.2e-16. Continuing the trend of the previous model of having an approximate distribution of residuals and low R^2 which is 0.1131.

For model 4 (C19proSo04), the strongest predictors are c19perBeh01, c19perBeh02, c19perBeh03, c19RCA01, c19RCA02 , c19RCA03. The pattern of p-value and residuals in the overall model are similar to others, having a low value of 2.2e-16 and symmetrical distribution. The R^2 is 0.1581 suggesting the limited power in the explanation of the model.

From the statistics above, the low p-values show that there is significant correlation between the independent and the dependent variables, but the models' R-squared values, which range from 0.10 to 0.16, are low. This suggests that only roughly 10–16% of the variability in the dependent variables can be explained by the models.

Comparing these attributes to those of my focus country, the focus country's models outperformed all of these by only a slight amount, as they are having slightly higher R^2 values. However, in the models for the other countries, I found that more variables are regarded as strong predictors when looking at the summary of each model, indicating a more complicated interaction of variables impacting pro-social behaviours throughout the larger dataset.

6. Focus country vs cluster of similar countries.

The table below shows the indicators I have used. These indicators covered the aspects of social, economic, health, political, using these offers a comprehensive perspective that may be used to find countries with comparable profiles.

No.	Indicators	Description
1	Literacy rate	The index that measures the literacy skill of the population.
2	Gender Equality Index	The index evaluates differences between genders in terms of economic status, empowerment, and reproductive health.
3	Gini Coefficient	The index quantifies the inequality in income in a country.
4	NCD mortality rate	The index measures the number of deaths from non-communicable diseases per 100,000 population.
5	Natural disaster risk	The index evaluates the possible harm caused by natural disasters.
6	Public Health vulnerabilities	The index examines the limitations of health systems that impact how quickly medical emergencies are handled.
7	Government effectiveness	The index evaluates the government's development and execution of public policies as well as the quality of public services.
8	Political and security risk	The index evaluates a country's level of safety, stability, and war risk.
9	Corruption	The index measures the extent of abuse of authority for personal benefit.
10	Unemployment rate	The index shows the proportion of the unemployed workforce.
11	GDP	The index measures the total market value of all completed products and services generated inside a country.
12	Urban population	The index shows the portion of people who live in urban areas.
13	Coverage of insurance	The index shows the portion of people who have insurance in health, life or other areas.

Next, I built a k-means clustering model based on the indicators and data collected. I did some preprocessing and cleaning of my data as there are some missing values. Then, I fit the model with my data after scaling it. After that, the 5 countries that are similar to Ukraine I have obtained include Algeria, Belarus, China, Thailand, Greece. In order to see what is the best number of clusters for the data, I find the silhouette_score for each number of clusters, ranging from 2 to 14. And the diagram shows that the score is the best at 2 clusters, then it plummets from 2 to 5, however, the score from 6 to 8 is about the same, it continues to drop after 8 clusters, thus, I decided to choose 8 clusters as my cut off since clusters beyond of this range don't offer any more meaningful groupings.

(b) Performance of attributes predicting pro-social attitudes for similar countries. Discuss similarities and differences between 2(c) & 3(b).

I subset the data of the chosen countries (Algeria, Belarus, China, Thailand, Greece) from the cvbase dataset, then I fit the model on the countries as a cluster. Next, I investigate the p-value of the attributes, residuals, R-squared and p-value of the model.

The first model (c19ProSo01), c19perBeh01, rankOrdLife_2D, rankOrdLife_2E, rankOrdLife_2F, rankOrdLife_3D, rankOrdLife_3E, rankOrdLife_3F, rankOrdLife_5C are the strongest predictors. With an overall p-value of less than 2.2e-16, the first model exhibits excellent statistical significance overall. There appears to be a reasonable fit as the residuals are fairly dispersed. But with an R-squared of 0.1125, it appears that just 11.25 % of the variability in the dependent variable is explained by the model.

For the second model (c19ProSo02), the strongest attribute is c19perBeh01. The significance of the model is shown by the exceptionally low (<2.2e-16) overall model p-value. Even with the minor improvement to the R-squared value of 0.1375 compared to the previous one, there is still little explanatory power.

For the third model (c19ProSo03), like the strongest predictors of the second model, the p-value for the best predictor, isoOthPpl_online, age indicates that they are not as strong of a predictor. Again, the R^2 of the model is low, having a value of 0.1203, and overall p-value of the model is <2.2e-16.

For the fourth model (c19ProSo04), the strongest attribute is c19perBeh01, c19perBeh02, c19perBeh03, c19RCA02. Despite its statistical significance(the overall p-value < 2.2e-16), the model's ability to explain the variance in the dependent variable is quite restricted as indicated by the R-squared of 0.1614.

Comparing models for all countries , similar countries and Ukraine, the overall p-values of the models are low, indicating the statistical significance of the models. Besides, the overall R^2 is also low for all models. All models have this constraint, however it's noteworthy that models for Ukraine perform marginally better than those for comparable or other nations (higher R-squared values). This could imply that, as opposed to more general models, localised models that are adapted to particular country contexts could capture more meaningful variance. With the exception of c19ProSo04 where c19perBeh01, c19perBeh02, c19perBeh03, c19RCA02 and, in c19ProSo01 and c19ProSo02, c19perBeh01 are a common significant predictor between similar countries and other countries, the fact that different models have wildly divergent predictors suggests that pro-social behaviours are driven differently in various circumstances.

In conclusion, these results and insights suggest that the models for similar or other countries as a group may not perfectly fit the pro-social behaviour in the focused country. These can be explained by several factors, for example, the differences in the social, religion, economic, and cultural disparities across the countries in each model group. Also, the timing, methods of data collection and how the participants interpret the question might affect the accuracy of data. Hence, the data might not reflect the underlying prosocial behaviours well. In order to improve the performance of the model, more variables that are distinctive to a given culture or region added could help to improve the R-squared values. Also, comparative research across cultures in different countries might need to be conducted. Last is to create more uniform data collection practices across various regions which may help to improve the models with higher levels of accuracy.

c19ProSo01			
	Strongest predictors	**p-value**	**R^2**
Ukraine	MLQ, consp03, c19RCA01	1.14e-11	0.1943
Other Countries	MLQ, bor03, c19perBeh01, c19RCA01	<2.2e-16	0.1178
Similar countries	c19perBeh01, rankOrdLife_2D, rankOrdLife_2E, rankOrdLife_2F, rankOrdLife_3D, rankOrdLife_3E, rankOrdLife_3F, rankOrdLife_5C	< 2.2e-16	0.1125

c19ProSo02			
	Strongest predictors	**p-value**	**R^2**
Ukraine	c19RCA01	4.154e-14	0.2113
Other Countries	isoFriends_inPerson, lifeSat, MLQ ,bor01, edu, c19perBeh01 , c19RCA01, c19RCA03	< 2.2e-16	0.1672
Similar countries	c19perBeh01	< 2.2e-16	0.1375

c19ProSo03			
	Strongest predictors	**p-value**	**R^2**
Ukraine	consp03	3.182e-11	0.1911
Other Countries	lifeSat , consp02, edu, c19perBeh01, c19RCA01	<2.2e-16	0.1131
Similar countries	isoOthPpl_online, age	< 2.2e-16	0.1203

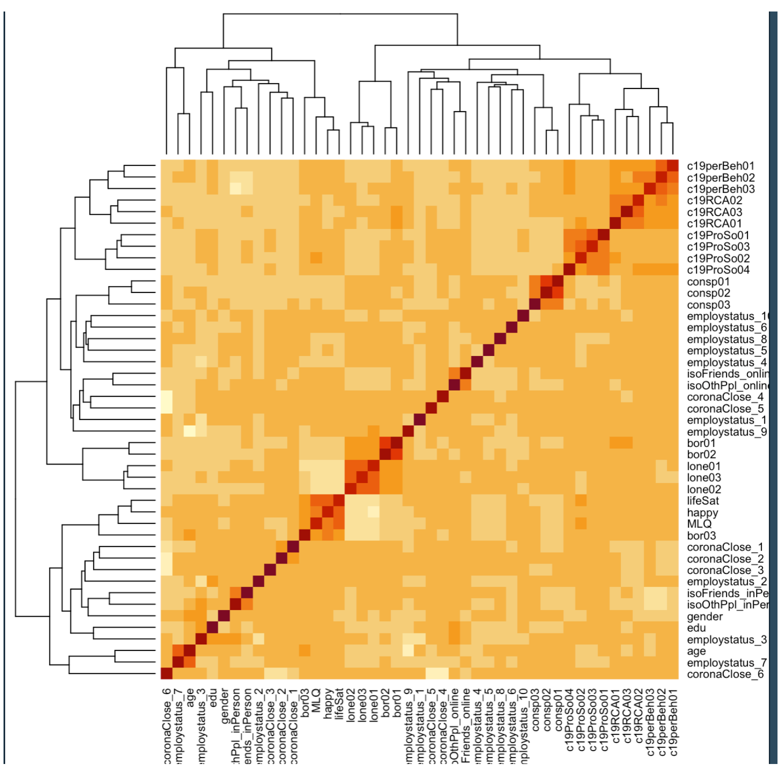
c19ProSo04			
	Strongest predictors	**p-value**	**R^2**
Ukraine	rankOrdLife_3F	1.099e-10	0.1871
Other Countries	c19perBeh01, c19perBeh02, c19perBeh03, c19RCA01, c19RCA02 , c19RCA03	< 2.2e-16	0.1581
Similar countries	c19perBeh01, c19perBeh02, c19perBeh03, c19RCA02	< 2.2e-16	0.1614

Appendix:

1. Summary of numerical variable

	skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
1	isoFriends_inPerson	0	1	1.96774010	2.3802715	0	0	1	3	7	
2	isoOthPpl_inPerson	0	1	1.85186027	2.1571212	0	0	1	3	7	
3	isoFriends_online	0	1	4.42617173	2.4682279	0	2	5	7	7	
4	isoOthPpl_online	0	1	2.83989427	2.6430702	0	0	2	5	7	
5	lone01	0	1	2.42173777	1.1494569	1	1	2	3	5	
6	lone02	0	1	2.69399994	1.2317868	1	2	3	4	5	
7	lone03	0	1	2.06940852	1.1471827	1	1	2	3	5	
8	happy	0	1	6.34582042	1.9883651	1	5	7	8	10	
9	lifeSat	0	1	4.16033312	1.2178042	1	3	4	5	6	
10	MLQ	0	1	0.86305886	1.5688781	-3	0	1	2	3	
11	bor01	0	1	0.31299207	1.9054399	-3	-1	0	2	3	
12	bor02	0	1	0.04675554	1.8775402	-3	-2	0	2	3	
13	bor03	0	1	0.29767217	1.6275955	-3	-1	0	2	3	
14	consp01	0	1	6.93877725	2.5801909	0	5	7	9	10	
15	consp02	0	1	7.26379217	2.4718801	0	6	8	9	10	
16	consp03	0	1	5.63687008	2.6798698	0	4	6	8	10	
17	c19perBeh01	0	1	2.36185089	1.0417125	-3	2	3	3	3	
18	c19perBeh02	0	1	2.48045931	0.9584975	-3	2	3	3	3	
19	c19perBeh03	0	1	1.87755450	1.4693010	-3	1	2	3	3	
20	c19RCA01	0	1	1.31097405	1.8415476	-3	0	2	3	3	
21	c19RCA02	0	1	2.11164483	1.2831191	-3	2	3	3	3	
22	c19RCA03	0	1	1.14663332	1.8165739	-3	0	2	3	3	
23	gender	0	1	1.38370804	0.4951571	1	1	1	2	3	
24	age	0	1	2.91802859	1.5926969	1	2	3	4	8	
25	edu	0	1	4.41258562	1.4266890	1	4	5	5	7	
26	c19ProSo01	0	1	0.97802916	1.4630540	-3	0	1	2	3	
27	c19ProSo02	0	1	0.66358753	1.6433132	-3	0	1	2	3	
28	c19ProSo03	0	1	0.54520649	1.6697853	-3	0	1	2	3	
29	c19ProSo04	0	1	1.35193133	1.5449087	-3	1	2	3	3	

2. Heat Map for the numerical variable



Data Source:

https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?end=2021&start=1991

https://ghsindex.org/report-model/

https://www.imf.org/external/datamapper/profile

https://www.globaldata.com/data-insights/macroeconomic/the-unemployment-rate-of-antigua-and-barbuda-220119/

https://www.globaldata.com/data-insights/macroeconomic/the-unemployment-rate-of-andorra-220118/

https://tradingeconomics.com/palestine/unemployment-rate

https://www.statista.com/statistics/1423918/unemployment-rate-in-palestine/

https://documents1.worldbank.org/curated/en/099455304272328937/pdf/IDU00e7074dc01c850437a0b9770f1fcae9a85c8.pdf

```r
install.packages("ggplot2")
install.packages("skimr")
install.packages("dplyr")
install.packages("treemapify")
library(ggplot2)
library(skimr)
library(dplyr)
library(cluster)
library(treemapify)

#--------------------------------------------
# Assignment 1
#--------------------------------------------
rm(list = ls())
set.seed(32581343) # XXXXXXXX = your student ID
cvbase = read.csv("PsyCoronaBaselineExtract.csv")
cvbase = cvbase[sample(nrow(cvbase), 40000), ] # 40000 rows
View(cvbase)
attach(cvbase)

#--------------------------------------------
# Question 1a : Description of data
#--------------------------------------------
# Check the datatypes of the columns
str(cvbase)

# Check occurrences of NA values in all columns
colSums(is.na(cvbase))

# Get a summary of the data
skim(cvbase)

# Get the unique values and number of unique values of the data
unique_value = function(col){
  uniq = unique(col)
  num = length(uniq)
  return(list(UniqueValues = uniq, NumberOfUniqueValues = num))
}
col_uniq_value = lapply(cvbase, unique_value)
col_uniq_value


#--------------------------------------------
# Question 1b : Pre-processing & cleaning data
#--------------------------------------------
# Step 1 : Replace employStatus and coronaClose with 0
columns_to_replace = c("employstatus_1", "employstatus_2", "employstatus_3","employstatus_4",
                       "employstatus_5", "employstatus_6","employstatus_7","employstatus_8",
                       "employstatus_9","employstatus_10","coronaClose_1","coronaClose_2",
                       "coronaClose_3","coronaClose_4","coronaClose_5","coronaClose_6")

for (col in columns_to_replace) {
  cvbase[is.na(cvbase[[col]]), col] = 0
}

# Step 2 : Replace "" in coded_country as NA value
coded_country[coded_country == ""] = NA
unique(cvbase$coded_country)

# Step 3 : Drop rows with NA values
cvbase = na.omit(cvbase)

# Make sure no more NA values
colSums(is.na(cvbase))

# Take out character type attribute
cvbase_subset = cvbase %>% select(-rankOrdLife_1,-rankOrdLife_2,-rankOrdLife_3,-rankOrdLife_4,
                                  -rankOrdLife_5,-rankOrdLife_6,-coded_country)
matrix = cor(cvbase_subset)

# Create heat map of numerical attribute
heatmap(matrix)
```

```
# Check Occurrence of country
country_occurrence = as.data.frame(table(cvbase$coded_country))
View(country_occurrence)

# Factorize rankOrdLife
# levels_description = c("A" = "Beauty", "B" = "Achievement", "C" = "Victory", "D" =
"Friendship", "E" = "Love", "F" = "Empathy")
cvbase$rankOrdLife_1 = factor(cvbase$rankOrdLife_1)
cvbase$rankOrdLife_2 = factor(cvbase$rankOrdLife_2)
cvbase$rankOrdLife_3 = factor(cvbase$rankOrdLife_3)
cvbase$rankOrdLife_4 = factor(cvbase$rankOrdLife_4)
cvbase$rankOrdLife_5 = factor(cvbase$rankOrdLife_5)
cvbase$rankOrdLife_6 = factor(cvbase$rankOrdLife_6)

contrasts(cvbase$rankOrdLife_1) = contr.treatment(levels(cvbase$rankOrdLife_1))
contrasts(cvbase$rankOrdLife_2) = contr.treatment(levels(cvbase$rankOrdLife_2))
contrasts(cvbase$rankOrdLife_3) = contr.treatment(levels(cvbase$rankOrdLife_3))
contrasts(cvbase$rankOrdLife_4) = contr.treatment(levels(cvbase$rankOrdLife_4))
contrasts(cvbase$rankOrdLife_5) = contr.treatment(levels(cvbase$rankOrdLife_5))
contrasts(cvbase$rankOrdLife_6) = contr.treatment(levels(cvbase$rankOrdLife_6))

# Rank Order Life
rankOrdLife = c("rankOrdLife_1","rankOrdLife_2","rankOrdLife_3",
                "rankOrdLife_4","rankOrdLife_5","rankOrdLife_6")
? barplot
? table
for(i in rankOrdLife){
  occurrence = table(cvbase[[i]])
  print(occurrence)
  barplot(occurrence, xlab = "Categories", ylab = "Frequency",main = "'rankOrdLife'
Distribution",sub = i, col= 'skyblue')
}

# Employment Status
Count = c(sum(cvbase$employstatus_1), sum(cvbase$employstatus_2),sum(cvbase$employstatus_3),
sum(cvbase$employstatus_4),
          sum(cvbase$employstatus_5), sum(cvbase$employstatus_6),sum(cvbase$employstatus_7),
sum(cvbase$employstatus_8),
          sum(cvbase$employstatus_9), sum(cvbase$employstatus_10)
)
employmentStatus = c("1-24 hours per week", "25-39 hours per week", "40+ hours per week",
                     "Not employed, looking", "Not employed, not looking", "Homemaker",
                     "Retired", "Disabled", "Student", "Volunteering")

employment_status_summary = data.frame(employmentStatus,Count)

ggplot(employment_status_summary, aes(x = employmentStatus, y = Count)) +
  geom_bar(stat = "identity",color = 'black', fill = 'skyblue') +
  labs(title = "Distribution of Employment Status", y = "Count", x= "Employment Status") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Education level
ggplot(cvbase, aes(x = edu)) + geom_bar(fill = "skyblue", color = "black") +
  labs(title = "Distribution of Education Level", x = "Education Level", y = "Frequency")

# Gender
ggplot(cvbase, aes(x = gender)) + geom_bar(fill = "skyblue", color = "black") +
  labs(title = "Distribution of Gender", x = "Gender", y = "Frequency")

# Age
ggplot(cvbase, aes(x = age)) + geom_bar(fill = "skyblue", color = "black") +
  labs(title = "Distribution of Age", x = "ageGroup", y = "Frequency")


# Isolation online & offline
iso_df = cvbase[,c("isoFriends_inPerson","isoOthPpl_inPerson",
                   "isoFriends_online", "isoOthPpl_online")]
iso_long = reshape2::melt(iso_df)
ggplot(iso_long, aes(x = value)) +
  geom_histogram(bins = 8, fill = 'skyblue', color = "black") +
  facet_wrap(~ variable, scales = "free_x") +
  labs(title = "Distribution of Social Isolation Measures",x = "Score",y = "Count")

# Loneliness
```

```
lone_df = cvbase[,c("lone01","lone02","lone03")]
lone_long = reshape2::melt(lone_df)
ggplot(lone_long, aes(x = value)) +
  geom_histogram(bins = 5, fill = 'skyblue', color = "black") +
  facet_wrap(~ variable, scales = "free_x") +
  labs(title = "Distribution of Social Loneliness Measures",x = "Score",y = "Count")

# Boredomness
bor_df = cvbase[,c("bor01","bor02","bor03")]
bor_long = reshape2::melt(bor_df)
ggplot(bor_long, aes(x = value)) +
  geom_histogram(bins = 5, fill = 'skyblue', color = "black") +
  facet_wrap(~ variable, scales = "free_x") +
  labs(title = "Distribution of Social Boredomness Measures",x = "Score",y = "Count")

# Corona Personal Behavior
c19perBeh_df = cvbase[,c("c19perBeh01","c19perBeh02","c19perBeh03")]
c19perBeh_long = reshape2::melt(c19perBeh_df)
ggplot(c19perBeh_long, aes(x = value)) +
  geom_histogram(bins = 5, fill = 'skyblue', color = "black") +
  facet_wrap(~ variable, scales = "free_x") +
  labs(title = "Distribution of Corona Personal Behavior",x = "Score",y = "Count")

# Corona Radical Action
c19RCA_df = cvbase[,c("c19RCA01","c19RCA02","c19RCA03")]
c19RCA_long = reshape2::melt(c19RCA_df)
ggplot(c19RCA_long, aes(x = value)) +
  geom_histogram(bins = 5, fill = 'skyblue', color = "black") +
  facet_wrap(~ variable, scales = "free_x") +
  labs(title = "Distribution of Corona Radical Action",x = "Score",y = "Count")

# Corona ProSocial Behavior
c19ProSo_df = cvbase[,c("c19ProSo01","c19ProSo02","c19ProSo03")]
c19ProSo_long = reshape2::melt(c19ProSo_df)
ggplot(c19ProSo_long, aes(x = value)) +
  geom_histogram(bins = 5, fill = 'skyblue', color = "black") +
  facet_wrap(~ variable, scales = "free_x") +
  labs(title = "Distribution of Corona ProSocial Behavior",x = "Score",y = "Count")

# Comprehensive distribution of numerical attributes
cvbase_subset = cvbase %>%
  select(-employstatus_1, -employstatus_2, -employstatus_3, -employstatus_4,
         -employstatus_5, -employstatus_6, -employstatus_7, -employstatus_8,
         -employstatus_9,-employstatus_10, -coronaClose_1,-coronaClose_2,
         -coronaClose_3,-coronaClose_4,-coronaClose_5,-coronaClose_6,
         -rankOrdLife_1,-rankOrdLife_2,-rankOrdLife_3,-rankOrdLife_4,
         -rankOrdLife_5,-rankOrdLife_6,-coded_country)
skim_df = skim(cvbase_subset)
skim_num_df = yank(skim_df, "numeric")
View(skim_num_df)

#---------------------------------------------
# Question 2 : Ukraine vs Other Countries
#---------------------------------------------
# Create other countries data set
other = cvbase %>% filter(coded_country != "Ukraine")
other_subset = other %>% select(-employstatus_1, -employstatus_2, -employstatus_3, -
employstatus_4,
                                    -employstatus_5, -employstatus_6, -employstatus_7, -
employstatus_8,
                                    -employstatus_9,-employstatus_10, -coronaClose_1,-
coronaClose_2,
                                    -coronaClose_3,-coronaClose_4,-coronaClose_5,-coronaClose_6,
                                    -rankOrdLife_1,-rankOrdLife_2,-rankOrdLife_3,-rankOrdLife_4,
                                    -rankOrdLife_5,-rankOrdLife_6,-coded_country)
other_df = skim(other_subset)
other_df = yank(other_df, "numeric")
View(other_df)

# Create Ukraine data set
ukraine = cvbase %>% filter(coded_country == "Ukraine")
ukraine_subset = ukraine %>% select(-employstatus_1, -employstatus_2, -employstatus_3, -
employstatus_4,
                               -employstatus_5, -employstatus_6, -employstatus_7, -
```

```
employstatus_8,
                               -employstatus_9,-employstatus_10, -coronaClose_1,-coronaClose_2,
                               -coronaClose_3,-coronaClose_4,-coronaClose_5,-coronaClose_6,
                               -rankOrdLife_1,-rankOrdLife_2,-rankOrdLife_3,-rankOrdLife_4,
                               -rankOrdLife_5,-rankOrdLife_6,-coded_country)
ukraine_df = skim(ukraine_subset)
ukraine_df = yank(ukraine_df, "numeric")
View(ukraine_df)

# Retrieve numeric data only from both skim data frames
subset_other = other_df[, c("skim_variable", "p0","p25","p50","p75","p100")]
subset_other = subset_other %>% mutate(country = "Other")

subset_Ukraine = ukraine_df[, c("skim_variable", "p0","p25","p50","p75","p100")]
subset_Ukraine = subset_Ukraine %>% mutate(country = "Ukraine")

# Combine both Ukraine and other countries data
whole = rbind(subset_other,subset_Ukraine)
View(whole)

# plot a box plot to do comparison
ggplot(whole, aes(x = skim_variable, lower = p25, upper = p75, middle = p50, ymin = p0, ymax =
p100, fill = country)) +
  geom_boxplot(stat = "identity") +
  theme_minimal() +
  labs(y = "Value", title = "Comparison of Attributes between Ukraine and Other Countries") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#----------------------------------------------
# Question 2b : Focus Group Model
#-----------------------------------------------
model = function(target,data) {
  formula_str <- paste(target, "~ isoFriends_inPerson + isoOthPpl_inPerson + isoFriends_online +
isoOthPpl_online +
            lone01 + lone02 + lone03 + happy + lifeSat + MLQ + bor01 + bor02 + bor03 + consp01 +
consp02 + consp03 +
           coronaClose_1 + coronaClose_2 + coronaClose_3 + coronaClose_4 + coronaClose_5 +
coronaClose_6 +
           gender + age + edu + c19perBeh01 + c19perBeh02 + c19perBeh03 + c19RCA01 + c19RCA02 +
c19RCA03 +
           rankOrdLife_1 + rankOrdLife_2 + rankOrdLife_3 + rankOrdLife_4 + rankOrdLife_5 +
rankOrdLife_6 +
           employstatus_1 + employstatus_2 + employstatus_3 +employstatus_4 +employstatus_5 +
employstatus_6 +
           employstatus_7 + employstatus_8 + employstatus_9 + employstatus_10")

  formula = as.formula(formula_str)
  fit = lm(formula, data = data)
  return(summary(fit))
}
summary_model_c19ProSo01 = model("c19ProSo01", ukraine)
summary_model_c19ProSo01
summary_model_c19ProSo02 = model("c19ProSo02", ukraine)
summary_model_c19ProSo02
summary_model_c19ProSo03 = model("c19ProSo03", ukraine)
summary_model_c19ProSo03
summary_model_c19ProSo04 = model("c19ProSo04", ukraine)
summary_model_c19ProSo04


#----------------------------------------------
# Question 2c : Other group model
#----------------------------------------------
# Create linear regression model for other countries as a group, added coded_country as an
attribute
model = function(target,data) {
  formula_str = paste(target, "~ isoFriends_inPerson + isoOthPpl_inPerson + isoFriends_online +
isoOthPpl_online +
           lone01 + lone02 + lone03 + happy + lifeSat + MLQ + bor01 + bor02 + bor03 + consp01 +
consp02 + consp03 +
           coronaClose_1 + coronaClose_2 + coronaClose_3 + coronaClose_4 + coronaClose_5 +
coronaClose_6 +
           gender + age + edu + c19perBeh01 + c19perBeh02 + c19perBeh03 + c19RCA01 + c19RCA02 +
c19RCA03 +
```

```
           rankOrdLife_1 + rankOrdLife_2 + rankOrdLife_3 + rankOrdLife_4 + rankOrdLife_5 +
rankOrdLife_6 +
           employstatus_1 + employstatus_2 + employstatus_3 +employstatus_4 +employstatus_5 +
employstatus_6 + employstatus_7 +
           employstatus_8 +employstatus_9 + employstatus_10 + coded_country")

  formula = as.formula(formula_str)
  data$coded_country = as.numeric(as.factor(data$coded_country))
  fit = lm(formula, data = data)
  return(summary(fit))
}

summary_model_c19ProSo01 = model("c19ProSo01",other)
summary_model_c19ProSo01
summary_model_c19ProSo02 = model("c19ProSo02",other)
summary_model_c19ProSo02
summary_model_c19ProSo03 = model("c19ProSo03",other)
summary_model_c19ProSo03
summary_model_c19ProSo04 = model("c19ProSo04",other)
summary_model_c19ProSo04



#----------------------------------------------
# Question 3 : Other group model
#----------------------------------------------
data = read.csv('data1.csv')
data = subset(data, select = -Year)
View(data)
str(data)

# Make gdp as numeric
data[,14] = as.numeric(as.character(data[,14]))

# Check NA data
colSums(is.na(data))

# Make NA as mean
for(i in 2:ncol(data)) {
    column_mean = mean(data[[i]], na.rm = TRUE)
    data[[i]][is.na(data[[i]])] = column_mean

}

# Factorise country and scale the data
data$Country = factor(data$Country)
data_scaled = data
data_scaled[,2:14]  = scale(data_scaled[,2:14])
View(data_scaled)

?kmeans
# K- means
dkfit = kmeans(data_scaled[,2:14],8, nstart = 50)
dkfit$cluster
T1 = table(actual = data$Country, fitted = dkfit$cluster)
T1 = as.data.frame.matrix(T1)
View(T1)

# Find similar countries
focus_country_index = which(data$Country == "Ukraine") #
focus_cluster = dkfit$cluster[focus_country_index]
similar_countries = row.names(T1)[T1[, focus_cluster] == 1]
similar_countries

# Check how many clusters is suitable
silhouette_score = function(k){
  km = kmeans(data_scaled[,2:14], k, nstart = 50)
  # calculate the silhouette score
  ss = silhouette(km$cluster, dist(data_scaled[,2:14]))
  mean(ss[,3])
}

k = 2:15
```

```
# Run silhouette function for all value of k and plot a graph
avg_sil = sapply(k,silhouette_score)
plot(k, type = 'b', avg_sil, xlab  = "Number of Clusters",ylab  = "Average Silhouette Score")

#-------------------------------------
# Question 3b :
#-------------------------------------
# Five countries
specific_countries = c("Algeria", "Belarus", "China", "Thailand", "Greece")

# Retrieve data of 5 countries
df_kmeans = cvbase[cvbase$coded_country %in% specific_countries,]
View(df_kmeans)

# Fit linear model of the 5 countries as a group
model = function(target,data) {
  formula_str = paste(target, "~ isoFriends_inPerson + isoOthPpl_inPerson + isoFriends_online +
isoOthPpl_online +
           lone01 + lone02 + lone03 + happy + lifeSat + MLQ + bor01 + bor02 + bor03 + consp01 +
consp02 + consp03 +
           coronaClose_1 + coronaClose_2 + coronaClose_3 + coronaClose_4 + coronaClose_5 +
coronaClose_6 +
           gender + age + edu + c19perBeh01 + c19perBeh02 + c19perBeh03 + c19RCA01 + c19RCA02 +
c19RCA03 +
           rankOrdLife_1 + rankOrdLife_2 + rankOrdLife_3 + rankOrdLife_4 + rankOrdLife_5 +
rankOrdLife_6 +
           employstatus_1 + employstatus_2 + employstatus_3 +employstatus_4 +employstatus_5 +
employstatus_6 + employstatus_7 +
           employstatus_8 +employstatus_9 + employstatus_10 + coded_country")

  formula = as.formula(formula_str)
  data$coded_country = as.numeric(as.factor(data$coded_country))
  fit = lm(formula, data = data)
  return(summary(fit))
}
summary_kmodel_c19ProSo01 = model("c19ProSo01",df_kmeans)
summary_kmodel_c19ProSo01
summary_kmodel_c19ProSo02 = model("c19ProSo02",df_kmeans)
summary_kmodel_c19ProSo02
summary_kmodel_c19ProSo03 = model("c19ProSo03",df_kmeans)
summary_kmodel_c19ProSo03
summary_kmodel_c19ProSo04 = model("c19ProSo04",df_kmeans)
summary_kmodel_c19ProSo04
```

data

Country	Year	Literacy	Gender equality	Gini coefficient	NCD mortality rate	Natural disaster risk	Public health vulnerabilities	Government effectiveness	Political and security risk	Coverage of social insurance programs (% of population)	Urban population (% of total population)	Corruption
Afghanistan	2021	26.6	32.5	89.5	38.2	25	45.1	9.9	5	0	85.6	19
Albania	2021	97.6	75.3	78.9	64.9	25	46.5	37.3	62.5	66.7	44.8	36
Algeria	2021	76.1	49.1	92.1	79.8	75	49	15.9	41.6	66.7	30.9	36
Andorra	2021	99.9	82.7	73.7	77.8	100	63.5	95.3	92.2	100	13.8	67
Angola	2021	56.2	32.1	31.6	63.1	50	39.6	21.7	60.2	33.3	39	27
Antigua & Barbuda	2021	98.7	63.1	26.3	74.6	0	57.1	51.1	83.5	33.3	87.1	58
Argentina	2021	98.8	60.2	52.6	80.9	75	64.7	45.3	62.4	66.7	9.2	42
Armenia	2021	99.6	72.1	86.8	72.2	25	53.5	39.1	38.9	100	42.4	49
Australia	2021	99.9	91.7	76.3	95.8	50	83	86	80.1	100	16	77
Austria	2021	99.9	95.5	84.2	91	100	85.2	75.1	88.1	100	47.9	76
Azerbaijan	2021	99.7	64.4	94.7	42.2	75	64.2	25.7	26.3	100	50.7	30
Bahamas	2021	99.9	60.4	44.7	71.6	0	58.5	59	79.9	33.3	19.5	63
Bahrain	2021	96.8	78.7	0	60.9	100	59	38.1	41.2	66.7	12.2	42
Bangladesh	2021	66.4	37.4	81.6	77.2	0	59.9	18	50.2	0	72.2	26
Barbados	2021	99.5	72.5	63.2	77	0	57.4	66.3	86.9	33.3	79.4	64
Belarus	2021	99.7	89.7	100	64.3	75	55	10.3	28.9	100	24.2	47
Belgium	2021	99.9	99	94.7	91.9	100	78.3	75.1	75	100	2.3	76
Belize	2021	91.9	55.6	26.3	79.3	0	55.6	45.1	62.4	66.7	62.4	41
Benin	2021	25.9	27.7	39.5	61.8	100	25	34.4	60.9	33.3	60.1	41
Bhutan	2021	57	49.9	68.4	70.3	0	51.6	52.6	73	0	67.4	68
Bolivia	2021	90.3	48.7	55.3	66.7	50	49.8	25.9	58.5	33.3	34.8	31
Bosnia and Hercegovina	2021	96.1	84.3	78.9	69.8	25	48.6	22.9	43.7	66.7	59.3	35
Botswana	2021	84.2	46.4	26.3	52.8	100	52.3	55	81.7	0	34.4	60
Brazil	2021	91.2	56.2	26.3	81.9	100	56.6	41.1	65.4	66.7	15.2	38
Brunei	2021	96.4	75.3	0	62.3	100	63.7	55	75.7	33.3	25.5	60
Bulgaria	2021	97.9	77.3	57.9	63.5	75	46.2	42	69.1	100	28.5	44
Burkina Faso	2021	24.3	27.9	73.7	60.5	75	33.7	23.6	22.4	0	80.7	40
Burundi	2021	59.3	39.4	63.2	60.2	25	52.7	13.4	26.9	0	99.9	19
Cabo Verde	2021	83	58	55.3	73.4	25	51.6	54.7	79.2	33.3	39	58
Cambodia	2021	74.9	45.2	65.8	60.1	0	45.8	17.3	47.7	0	87.9	21
Cameroon	2021	70.5	33.6	42.1	59.3	25	38.5	25	21.4	0	49.6	25
Canada	2021	99.9	94.2	78.9	93.6	100	77.7	89.6	87.8	100	21.3	77
Central African Republic	2021	19.4	19.1	18.4	35.4	0	34.4	10.9	13.5	33.3	67.1	26
Chad	2021	0	16.7	52.6	62.6	0	17.1	10.1	24.1	0	88.5	21
Chile	2021	96	68.5	50	91	50	57.8	66.7	73.8	100	14.3	67
China	2021	95.9	84.2	63.2	75.5	25	60.8	38.1	62.6	66.7	45.8	42
Colombia	2021	93.7	53.1	31.6	91.3	25	59	52	46.7	33.3	21.8	39
Comoros	2021	47	30.6	47.4	67.2	25	36.3	3	51.6	0	81.7	21
Congo (Brazzaville)	2021	74.6	32	36.8	61	50	26.6	13.4	32.9	33.3	37.6	19
Congo (Democratic Republic)	2021	70.4	22.5	55.3	60.1	75	32.5	13.3	22.1	0	63.4	18
Cook Islands	2021	84.7	47.9	52.6	58.8	25	55	32	62.9	66.7	28.3	74
Costa Rica	2021	97.3	68.9	39.5	92.9	100	67.2	58.1	72.6	33.3	23	57
Côte d'Ivoire	2021	32	30.6	55.3	63.4	75	27.6	30.1	53.1	33.3	56.3	36
Croatia	2021	98.8	89.3	86.8	80	25	58.1	46	63.7	100	49.4	47
Cuba	2021	99.7	65.4	65.8	81.2	25	61.7	35.3	61	33.3	26.4	47
Cyprus	2021	98.3	93.9	78.9	95.3	100	52.7	61.7	48.1	66.7	38.3	57
Czech Republic	2021	99.7	87.5	100	82.6	75	75.9	61.3	86.1	100	30.1	54

Denmark	2021	99.9	99.6	92.1	89.7	100	82.2	84	85.8	100	13.8	88
Djibouti	2021	41.8	30.6	55.3	64.1	50	44.9	18.1	35.9	33.3	25.5	27
Dominica	2021	91.9	63.1	36.8	77	0	54.5	50.7	69.1	66.7	33.7	55
Dominican Republic	2021	92	47.8	55.3	74	25	61.1	39.7	64	33.3	21	28
Ecuador	2021	92.7	55.8	44.7	87.6	0	56.2	30.6	49.6	33.3	41.5	39
Egypt	2021	62.9	48.2	81.6	52.3	25	65.3	29.7	56.6	66.7	66.1	33
El Salvador	2021	85.8	54.8	63.2	84.9	25	55.3	40.9	53.5	33.3	31.5	36
Equatorial Guinea	2021	93.6	41.8	0	61.6	75	43.6	5.9	49.6	33.3	31.6	16
Eritrea	2021	69.9	30.6	0	54.3	75	37.9	10.1	39.5	0	67.7	21
Estonia	2021	99.9	93.2	86.8	81.7	75	66.8	82.1	82	100	35.6	75
eSwatini	2021	85.1	32	21.1	34.8	50	48.2	26.1	54.9	0	87.7	33
Ethiopia	2021	38	40.9	73.7	73.6	25	45.4	30.4	16.3	0	90.9	38
Fiji	2021	98.8	59.8	68.4	34.2	25	50.2	49	76	33.3	49.8	43
Finland	2021	99.9	98.4	94.7	90.5	100	75.8	87.1	82.7	100	16.8	85
France	2021	99.9	98.2	81.6	94.7	100	76.4	77.7	81.3	100	22.3	69
Gabon	2021	80.3	37.6	65.8	63.8	50	34	32.9	61.8	33.3	11.9	30
Gambia	2021	36.7	26.9	71.1	65.5	100	55.1	30.3	62.7	0	43.9	37
Georgia	2021	99.5	60.6	71.1	56	25	58	58	36.9	66.7	47.3	56
Germany	2021	99.9	94.1	81.6	91	100	86.5	79.3	87.5	100	26.1	80
Ghana	2021	73	36.8	50	63.4	75	54.3	41.9	71.5	100	49.9	43
Greece	2021	97.3	89.3	78.9	91	0	57.6	57.1	67.7	100	23.8	50
Grenada	2021	98.2	63.1	47.4	63.4	0	53.7	50.4	83.4	33.3	73.4	53
Guatemala	2021	75.9	42.9	39.5	75	0	50.9	21.4	31.6	33.3	56.1	25
Guinea	2021	12.5	30.6	76.3	59.4	75	30.5	18.3	50.2	0	73.2	28
Guinea-Bissau	2021	30	30.6	31.6	58.2	0	36.8	2.7	25.4	33.3	64.8	19
Guyana	2021	81.5	42.9	47.4	45.9	0	52.7	38	60.2	66.7	84.5	41
Haiti	2021	50.7	26.9	57.9	42.4	0	37.6	16.9	51.2	0	50.5	18
Honduras	2021	85.8	44.5	39.5	63.8	0	52.7	24.9	35.7	0	48.8	24
Hungary	2021	98.8	72.3	86.8	69.5	50	64	56.3	74.7	100	32.8	44
Iceland	2021	99.9	97.5	97.4	93.3	75	80.8	82.1	92.7	100	7	75
India	2021	67.1	41.8	71.1	69.1	25	61	41.4	58.3	66.7	75.5	40
Indonesia	2021	94.5	48.1	65.8	59.3	50	48.6	41	61.8	33.3	50.7	37
Iran	2021	81.3	42.9	55.3	79.8	0	63.2	17.9	32.3	66.7	28.4	25
Iraq	2021	81.5	36.9	86.8	64.1	0	53.7	13.7	9.1	66.7	33.8	21
Ireland	2021	99.9	93	84.2	91.2	100	85	71	78	100	42.2	72
Israel	2021	99.9	92.1	63.2	95.2	0	76.8	58.6	60.7	66.7	8.7	60
Italy	2021	99	96	71.1	94.8	75	72.4	54	66	100	33.8	53
Jamaica	2021	84.7	53.8	44.7	79	25	53.8	49.1	71.3	33.3	50.7	44
Japan	2021	99.9	92.2	78.9	99.9	50	78.3	71.3	80.4	66.7	9.6	74
Jordan	2021	97.7	45.8	76.3	80.3	75	62	39.1	47.3	66.7	10.1	49
Kazakhstan	2021	99.7	79.2	92.1	63.1	75	54.9	30.4	53.2	66.7	49	38
Kenya	2021	76.2	36.3	57.9	66.3	50	38.7	29.4	50.6	0	83.6	31
Kiribati	2021	61.8	41.8	68.4	0	0	33.7	26.6	77.6	33.3	52.1	36
Kuwait	2021	95	73.9	0	91.4	100	81.4	38.1	62.6	66.7	0	42
Kyrgyz Republic	2021	99.5	56.8	86.8	66.2	75	64.3	25.9	54.9	66.7	73.1	31
Laos	2021	80.3	46.5	63.2	55.6	0	46.5	18.4	56.2	0	74.3	29
Latvia	2021	99.9	83.4	73.7	69.1	75	54.7	68.9	78.9	100	36.7	57
Lebanon	2021	93.7	59.2	81.6	72.1	0	59.4	10.7	13.4	66.7	12.9	25
Lesotho	2021	69.9	36.1	47.4	13.8	25	35.6	34.4	57.3	0	82.4	41

Liberia	2021	33.5	23	73.7	74	75	59.6	21.9	61.5	0	55.8	28
Libya	2021	91.9	83.1	0	76.7	100	54.1	13.1	9	66.7	22.6	17
Liechtenstein	2021	99.9	82.7	0	64.5	100	66.2	83.6	92.9	100	98.7	85
Lithuania	2021	99.7	89.1	71.1	72.3	75	57.4	72.9	77.1	100	37	60
Luxembourg	2021	99.9	94.9	73.7	93.5	100	90.1	86.4	86.2	100	10.1	80
Madagascar	2021	67.6	30.6	52.6	58.9	0	26.7	32.1	58.2	0	71.6	25
Malawi	2021	51.2	27.5	47.4	66.1	0	52.3	22.1	65.1	0	95.5	30
Malaysia	2021	93.4	70.3	57.9	74.6	100	70.2	57.3	73.7	33.3	27	51
Maldives	2021	98.2	58.6	84.2	84	50	61.1	49	60.6	0	69	43
Mali	2021	17	19.8	78.9	64	0	42.6	29.3	26.8	0	65.6	30
Malta	2021	92.9	80.2	89.5	92.4	100	76.9	57.6	79.7	100	6.1	53
Marshall Islands	2021	97.8	63.1	0	58.8	0	47.8	27.6	74.2	33.3	26.1	43
Mauritania	2021	40.2	26.9	78.9	76.9	75	30.7	25.6	52.5	33.3	52.5	29
Mauritius	2021	91.2	58.3	68.4	66.9	50	56.1	54	80.3	66.7	68.3	53
Mexico	2021	94.1	62.7	47.4	78	100	62.6	36.6	36.2	66.7	22.6	31
Micronesia, Federated States (	2021	61.8	41.8	60.5	10.5	0	43.1	26.6	77.6	33.3	89	36
Moldova	2021	99.2	76	97.4	61.5	50	42.7	29.9	35.2	100	66.1	34
Monaco	2021	99.9	70	0	77.8	100	59.9	95.3	92.2	100	0	67
Mongolia	2021	97.9	64.2	78.9	32.7	100	50.7	37.1	79.1	100	36.3	35
Montenegro	2021	98.5	89.7	63.2	64.8	50	50.2	42.1	57.2	100	37.8	45
Morocco	2021	66.3	42.9	60.5	60.9	25	51.5	34.3	51.3	66.7	42.7	40
Mozambique	2021	49.4	33.2	23.7	48	50	35.7	25	35.7	33.3	73.2	25
Myanmar	2021	68.6	47.1	84.2	58.1	0	56.3	14.7	12.8	33.3	79.7	28
Namibia	2021	89.1	46.9	10.5	62.4	0	43.5	50.1	71.4	0	56.5	51
Nauru	2021	84.7	47.9	73.7	58.8	25	47	6.1	55.6	0	0	43
Nepal	2021	58.7	44.9	78.9	64.8	25	53.7	19	57.5	33.3	92	33
Netherlands	2021	99.9	99.5	92.1	91.9	75	76.5	79.6	81.6	100	9.3	82
New Zealand	2021	99.9	88	71.1	92.9	25	80.8	87.6	93.5	100	15.5	88
Nicaragua	2021	77.6	47.6	44.7	70.2	25	49.5	13.9	37.7	33.3	47.5	22
Niger	2021	16.5	23.5	76.3	65.1	0	34.4	29.6	20.9	0	96.3	32
Nigeria	2021	37.1	30.6	73.7	71.8	75	47.9	17.9	29.9	0	56.3	25
Niue	2021	84.7	47.9	76.3	58.8	0	52.5	49	82	66.7	62.1	43
North Korea	2021	70.4	0	0	65.4	0	52.2	6.1	35.4	66.7	43.7	18
North Macedonia	2021	97.2	63.1	78.9	52.3	50	47.3	40.7	59.4	100	48.2	35
Norway	2021	99.9	99.1	92.1	94.6	100	93.6	90.6	96.3	100	20.1	84
Oman	2021	95	66.5	84.2	57.1	75	60.9	43.4	57.4	66.7	16.8	54
Pakistan	2021	47.4	36	81.6	50.4	0	53.9	22.3	37.7	33.3	72.8	31
Palau	2021	95.6	63.1	97.4	58.8	25	55.6	32	71.2	66.7	22.5	74
Panama	2021	94.1	46.9	34.2	90.6	75	59.8	51.4	65.7	33.3	36.8	35
Papua New Guinea	2021	50.6	11.8	55.3	42.4	25	28.2	32.4	53.4	0	100	27
Paraguay	2021	93.2	44.2	44.7	79.7	25	56.6	39.7	67.6	33.3	43.9	28
Peru	2021	92.8	56.8	55.3	93.5	25	54.3	41.1	57.1	33.3	25.3	38
Philippines	2021	97.7	51.3	55.3	54.3	0	63.5	40.6	43.9	33.3	61	34
Poland	2021	98.3	89.6	86.8	80.2	100	65.7	54.4	73.3	100	46.1	56
Portugal	2021	95	94.5	76.3	92.1	50	68.6	69.4	77.8	100	39.4	61
Qatar	2021	91.6	79.3	57.9	60.3	100	65.5	48.3	73.6	66.7	0.9	63
Romania	2021	98.5	65	71.1	69.4	50	49.6	45.6	76.7	100	52.9	44
Russia	2021	99.6	72.6	65.8	63.2	50	48	22.1	22.2	100	29.3	30
Rwanda	2021	65.5	52.9	50	67.1	50	56.2	43.4	55	33.3	95.4	54

Samoa	2021	98.8	59	63.2	49.3	25	51.4	70.4	76.7	66.7	94.5	43
San Marino	2021	99.9	70	0	77.8	100	65.9	95.3	95.8	100	3	67
São Tomé and Príncipe	2021	90.7	36	18.4	63.8	50	46.5	38.9	62.7	33.3	30.4	47
Saudi Arabia	2021	94	76.5	44.7	69.1	50	70.2	39.7	54.5	66.7	18.3	53
Senegal	2021	38.1	39	60.5	69.7	25	46.5	38.6	65	33.3	60.3	45
Serbia	2021	98.5	84.3	71.1	65	25	57.9	44.7	42.1	100	50.4	38
Seychelles	2021	94.7	63.1	81.6	68	0	59.2	55.9	77	0	49.5	66
Sierra Leone	2021	26.9	23.8	71.1	61.6	25	34.1	26.1	58.5	33.3	66.3	33
Singapore	2021	96.5	96.5	44.7	100	100	80.4	83.6	86.9	33.3	0	85
Slovakia	2021	99.9	80.8	100	80.2	75	66.6	67.7	71.6	100	53.4	49
Slovenia	2021	99.6	96	100	90	75	64.4	65.7	78.4	100	52.1	60
Solomon Islands	2021	84.7	30.6	68.4	35.6	0	36.3	27.4	68.2	0	87.4	42
Somalia	2021	41.8	32.9	68.4	48.9	0	35.7	5.3	4.3	0	62.7	12
South Africa	2021	83.3	51.7	0	63.3	75	49.6	56.3	73.5	0	38.2	44
South Korea	2021	99.9	97.4	84.2	99.8	50	78	73	66.4	66.7	21.5	61
South Sudan	2021	5.8	30.6	50	76.5	0	26.6	1.7	9.8	0	92.4	12
Spain	2021	97.9	95.4	73.7	95.5	100	71.5	66	73.7	100	22.4	62
Sri Lanka	2021	89.6	57	63.2	77.8	50	67	37.6	61.3	33.3	93.9	38
St Kitts & Nevis	2021	91.9	63.1	60.5	77	0	57.4	51.1	79.9	33.3	79.8	58
St Lucia	2021	91.9	62.9	31.6	73.2	0	55	50.9	78.7	33.3	93.7	56
St Vincent & The Grenadines	2021	91.9	63.1	18.4	71.1	0	53.7	51.3	76.4	33.3	54.7	59
Sudan	2021	49.4	34.4	76.3	62.7	25	34.6	13	9	66.7	75.1	16
Suriname	2021	92.8	46.3	13.2	58.9	0	54.5	37.6	66.1	66.7	39.1	38
Sweden	2021	99.9	99.6	86.8	93.7	75	81.8	87.1	87.4	100	14.2	85
Switzerland	2021	99.9	100	78.9	96.7	100	78.6	90.7	86.8	100	30.2	85
Syria	2021	41.8	36	65.8	61.9	0	52.3	2	0.3	33.3	52.1	14
Tajikistan	2021	99.7	57.3	76.3	38.7	0	62	7.1	41.5	66.7	83.9	25
Tanzania	2021	71.6	37	57.9	75	75	50.7	26.9	55	0	75.5	38
Thailand	2021	92	57.3	73.7	87.8	50	70.7	40.9	41.6	66.7	56.9	36
Timor-Leste	2021	58.9	41.8	89.5	67.9	25	44.4	20	60	0	79.7	40
Togo	2021	53.3	33.6	52.6	60.4	100	24.6	32.7	49.9	0	66.7	29
Tonga	2021	99.2	52.2	65.8	62.7	25	50.9	27.6	63.5	0	88.7	43
Trinidad and Tobago	2021	98.3	64	60.5	80.5	50	58.8	41.4	66.6	33.3	54	40
Tunisia	2021	73	67	78.9	78.2	75	54.7	45.6	51.7	66.7	35.4	44
Turkey	2021	95.1	66.4	55.3	80.7	50	63.5	34.3	38.2	66.7	28.1	40
Turkmenistan	2021	99.6	63.1	57.9	58.6	25	66.5	2.7	52.8	66.7	55.4	19
Tuvalu	2021	84.7	47.9	63.2	58.8	0	48.3	49	82	66.7	42.4	43
Uganda	2021	69.8	38	52.6	67.8	75	35	28.9	30.3	0	87.2	27
Ukraine	2021	100	69	94.7	60.4	50	45.4	36.9	24.3	100	35.2	33
United Arab Emirates	2021	91.2	90.5	97.4	71.9	75	74.8	53	71.9	66.7	15.2	71
United Kingdom	2021	99.9	89.7	73.7	90.9	50	82.7	78.9	81.5	100	18.8	77
United States of America	2021	99.9	81.8	57.9	83.5	75	75.9	66.7	69.1	100	20.2	67
Uruguay	2021	98.3	70.1	60.5	79.9	100	66.5	70.9	76.8	66.7	5.3	71
Uzbekistan	2021	100	66.6	73.7	58.1	75	68.3	18	49	66.7	57.2	26
Vanuatu	2021	83.9	41.8	65.8	28.2	25	41.7	70.4	80.3	0	86	43
Venezuela	2021	96.3	47.2	47.4	83.1	75	48.3	2.1	25.3	33.3	13.6	15
Vietnam	2021	93.6	65.2	71.1	68	25	53.3	33.7	63.1	33.3	73.1	36
Yemen	2021	41.8	0	68.4	53.9	0	40.9	9.3	1.3	33.3	72.3	15
Zambia	2021	82.9	36.9	15.8	57.5	25	37.4	26.1	64.4	0	64.5	33

Zimbabwe	2021	85.5	38.8	34.2	52.2	25	50.7	7	36.7	0	78.2	24