FIT3152
Assignment 1

Daniel Nguyen
32471033

Focus Country: Brazil

Table of Contents

1. Descriptive Analysis and Pre-Processing

a)
The dataset has dimensions based on the contents of the survey, including:
- Employment Status
- Offline Isolation
- Online Isolation
- Loneliness
- Life Satisfaction
- Boredom
- Conspiracy
- What the Participant Values in Life
- Importance of Certain Behaviours to Minimize Coronavirus
- Proximity to Other People with Coronavirus
- Gender
- Age
- Education Level
- Country
- Pro Social Coronavirus Behaviours

By using the dim function, we can see that there are 52 columns in the dataset, which reference the questions asked in the survey. Most columns represent a question asked in the survey, however, some questions (specifically Employment Status, What the Participant Values in Life, and Proximity to Other People with Coronavirus) take up multiple columns, as the question allowed multiple answer to be selected, or asked the participant to click and drag items to rank them.

Each column in the dataset has a range of values corresponding to the information contained. For efficient visualisation of data, some of these columns can be combined to view the full picture of the meaning of the information contained.

The only text attributes in the dataset are country names, and the life order ranking, in which the letters A to F correspond to the different values to be ranked. There are 3 main things that the numerical attributes represent. Questions in which the participant was asked to select all answers that applied are stored as booleans (selected or not selected) for each potential answer, while questions in which the participant was asked to record their answer on a scale ahd responses recorded numerically based on their answer. Gender is also recorded numerically, where a value was assigned to each answer (female, male, or other). This is significant, as when analysing relationships in the data, while values on a scale can be analysed holistically, this cannot be done for the gender column, as there is no “gender scale” and each response must be analysed independently.

There are a large amount of missing (NA) values in the dataset, however many are justified. In the case of dummy variables, missing values simply correspond to answers that the participant did not select. For example, in the employment status question, the participant was asked to select all that apply, and the ones that were not selected are stored as NA in the dataset. However, various participants have not answered some questions, and their lack of an answer is stored as a missing value as well.

b)

Pre-processing the data is crucial for allowing easy analysis. To separate responses from the focus country and other responses, we can add another column indicating whether it came from a participant that declared their country of residence as the focus country and assign values of 0 or 1.

Additionally, some attributes can be combined to reveal more relevant data. For example, the attributes concerning the number of hours a participant has worked, which are currently stored as boolean values, can be combined to form a scale, with 0 being the option for least hours, and 3 the option for most hours. This also allows for easy extraction of other data, such as the proportion of employed to unemployed participants.

The removal of missing data can also be done, however in this case, I have decided to only remove responses in which the participant did not answer any question apart from their country of residence, as they have not provided a single piece of useful information. This is because the removal of all rows where some data is missing may cause bias in analysis if a significant portion of participant responses are removed.

Removal of missing data can also be done on a case by case basis, when looking at individual attributes, so that missing answers for that question only are removed.

When performing a linear regression to find attributes that may predict pro social coronavirus behaviour, I have changed the life order ranking columns from rankings 1-6, to the individuals ranking of each category, as this allows for the values to be numerical. I have also removed responses for that question specifically where participants have specified multiple categories for the same ranking, as their data for that question would not be usable.

For the gender attribute, normally one hot encoding would be used for regression analysis. However, for my focus country dataset and seed, there are no people identifying as other, so this is not needed for the initial regression analysis. I have done this when looking at the data for all countries, however.

2. Focus Country vs All Other Countries as a Group

a)
The focus country of this report is Brazil. All referenced graphs can be found in the appendix.

Fig1
When looking at the employstatus attributes, the first 3 relate to the number of hours per week that the participant works. From this, we can also extract the employment rate of the participants from our focus country and compare this with the rest of the participants. We can see in figure 1 that there is a higher proportion of Brazilian participants that are employed compared to those from other countries

Fig2
In figure 2, we can see that while Brazil has a higher rate of employment among participants, those that are employed work fewer hours. We can see that the other participants are more likely to work 40+ hours a week than Brazilian participants, while the Brazilian participants are more like to only work 1-24 hours a week.

Fig3
Looking at the rest of the employstatus attributes, we can see that they concern unemployed people who are or are not looking for work, and people that identify as disabled and unable to work, and/or as homemakers, retirees, students, or volunteers. From figure 3, we can see that a higher proportion of unemployed Brazilian participants are looking for work relative to those from other countries.

Fig4
Additionally, a much lower proportion of Brazilian participants are students, while a higher proportion are homemakers or retired. Participants from other countries are also more likely to be unable to work due to disability

Fig5
When looking at contact with people in the past week, Brazilian participants in the survey had much more online contact, but in person contact with people in general was quite similar between Brazil and the other countries.

Fig6
When analysing the loneliness question responses, it is evident that while overall, both Brazil and all other countries on average do not feel lonely, the Brazilian participants feel more loneliness than others, as seen by the skew of the graphs in figure 6.

Fig7
When analysing the data for the life satisfaction series of questions, we can also see that the Brazilian participants are, on average, slightly happier, and that the responses indicating very low happiness are outliers. Additionally, Brazilian participants have slightly less life satisfaction, but a slightly higher sense of purpose in life on average.

Fig8
When analysing boredom, Brazilian participants, on average, feel that time is moving slowly and wish that time would go by faster more than participants from other countries. They also feel less in control of their time.

Fig9
Brazilian participants also have responses to the conspiracy questions that are very slightly more in agreement than the rest of the participants, except for their view of government agencies closely monitoring citizens, where there is a larger difference, with more Brazilian participants agreeing. They also have a higher level of extreme views on both sides, as seen in the area of the extremes in figure 9.

Fig10
The valuation of different categories is remarkable similar between Brazilian participants and other participants, where the median for each category is mostly the same. However, the Brazilian participants have a higher median valuation of empathy, but for those that did not value empathy as highly, there is a very large range of ratings. The median rating for friendship was also lower for Brazilian participants. Additionally, there was, in general, a greater spread of rankings for the Brazilian participants relative to those from other countries.

Fig11
When answering questions regarding their personal behaviour around coronavirus, Most participants generally support a petition for a mandatory quarantine in the same proportions, while Brazil is more open to the other radical actions. In terms of personal behaviour, the distribution is very similar, except for participants view on self quarantining, where Brazilian participants are more in support than those from other countries.

Fig12
There was generally a similar distribution of answers for all participants regarding their knowledge of any other people with coronavirus

Fig13
A slightly higher proportion of Brazilian participants were male relative to other countries, with no participants identifying as neither male nor female, however this can also be caused by the difference in sample size.

Fig14
Brazilian participants were, on average, older than those from other countries. This can be see by the larger areas towards the bottom of the Brazil graph in figure 14. In particular, the age ranges of 35-44 and 45-54 are noticeably larger for Brazil, while the age range for 18-24 is noticeably smaller.

Fig15
In figure 15, we can see a clear difference in education levels, where Brazilian participants are much more likely to have a highest education level of general secondary education or higher education, while participants from other countries are more likely to have a bachelors, masters, or PhD degree.

Fig16
In figure 16, we can see that only about 2% of the sample is made up of Brazilian participants, while 98% are from other countries.

Fig17
In figure 17, which concerns the responses of participants to the coronavirus pro social behaviour questions, we can see that Brazilian participants are more likely to support these behaviours across the board. While participants from other countries generally also support the behaviours, the magnitude ofresponses in agreement with the statements is significantly higher for Brazilian participants.

b)
The relevant data takes up too much space, so it has been labelled in the appendix, which includes R-Squared values, P values, residual data, and tables of significant attributes. Additionally, the summary tables for each model are included in the appendix as well.

c19ProSo01: I am willing to help others who suffer from coronavirus.

The residual values for this regression model show that the median is close to 0, and there is a roughly even spread of residual values around the median, showing an approximately normal distribution.

Approximately 20% of the variation in responses to c19ProSo01 can by predicted by the question responses by Brazilian participants, as indicated by the R-Squared value. The P-Value represents the change that these results would have been caused if there were no relationship at all between our predictor attributes and our outcome attribute. The very low P-Value indicates that our results are statistically significant, and that the question responses have some prediction ability for the willingness of participants to help others who suffer from coronavirus.

c19ProSo02: I am willing to make donations to help others that suffer from coronavirus.

The residual data again shows a roughly equal distribution of residuals, as our median is close to 0. The 1st quartile is slightly further from the median compared to the 3rd quartile, which indicates a greater spread of residuals below the median value.

The R-Squared value indicates that roughly 25% of the variation in responses to c19ProSo02 can be predicted by the responses to the other questions. The P-Value is extremely low, indicating that our results are statistically significant, and that it is very unlikely that our results are due to chance.

c19ProSo03: I am willing to protect vulnerable groups from coronavirus even at my own expense.

The residual data again shows a roughly normal distribution of residual values, however our first and 3rd quartiles are quite far from the median compared to our previous regression models. We can also see that our R-Squared value is lower, indicating less predictive power, which explains this spread of residual values.

Our R-Squared Value indicates that 16% of the variation in responses to our outcome question can be explained by the responses to the other question by Brazilian participants. This is lower than for the previous 2 outcome attributes, which indicates that from our data, there is less predictability for the responses to this question, however our P-Value is still low, indicating statistical significance, and that our model results are not due to chance.

c19ProSo04: I am willing to make personal sacrifices to prevent the spread of coronavirus.

The residual data indicates a slight skew towards positive residual values, indicating that there are slightly more residuals that are higher than our predicted value than there are below. This may mean that the relationship for our model is not linear. However, looking at our R-squared value and our P-value, we can assume that this is not the case.

The R-squared value for this model indicates that around 19% of variation in responses to the output question can be predicted by the participants answers to the other questions. The P-Value indicates a high statistical significance, meaning that there is indeed some predictive power for this model, and the data was not created due to chance only.

General Observations Regarding Focus Country Regression Data

In general, for each c19ProSo attribute, there was only moderate predictability based on the rest of the attributes, as indicated by the relatively low R-Squared values (around 0.15-0.25). However, when working with real world data, especially when the data concerns human responses and behaviour, these R-Squared values are actually very good. Additionally, we can see that our models are all statistically significant, as seen by the very low P-Values. We can make the conclusion that there is in fact some predictability for all of our outcome variables based on the predictor attributes.

The most significant attributes for each coronavirus pro social behaviour are detailed in the appendix. I selected the most significant attributes based on their p values for the models in question, as only a few attributes for each linear regression model had low enough P-values to be statistically significant. This is mainly due to the low sample size for my focus country. Additionally, each attribute listed also has their corresponding coefficient value. Negative coefficient values indicate a negative relationship with the dependant variable, while positive coefficients indicate a positive relationship. The values of the coefficients can only be directly compared when the participant responses are on the same scale, where a coefficient further away from 0 indicates a stronger impact.

c19RCA01, bor01, and employstatus_10 were overall the most significant attributes when predicting the responses to the Corona Pro Social Behaviour prompts. In particular, c19RCA01, which concerns responses to the question:

> I would sign a petition that supports mandatory vaccination once a vaccine has been developed for coronavirus.

had significantly low P-values when used in the models for every outcome attribute. For data regarding the focus country, this attribute was the best predictor overall. The other attributes mentioned had low P-values for multiple outcome attribute models, but not for all.

c)
The dataset for all other countries except the focus country has a much larger sample size, which, when creating linear regression models, causes many attributes to have low P-values when compared to our focus country data. For this reason, I am not only basing my attribute choices on P-value, but also on correlation of the individual variables to the outcome variables. I have included a correlation heatmap in the appendix in addition to regression model data, significant attributes, and the regression summary table.

c19ProSo01: I am willing to help others who suffer from coronavirus.

This model shows an approximately normal distribution for the residual values, as indicated by the mean residual value being close to 0, and the 1st and 3rd quartile ranges being roughly the same distance from the median.

With an R-Squared of 0.11, we can predict around 10% of output data based on the predictors. This is lower than we have seen for the model based on the focus country data. However, the P-Value calculated for this model is as low as can be measured by a computer, which indicates extreme statistical significance for this model.

Additionally, the most significant attributes I have chosen for this model, which were based upon P-Value and correlation coefficient, are similar to the significant attributes for this model, but based on the focus country data. Both c19RCA01 and c19perBeh01 seem to be important predictors for this outcome variable in both models

c19ProSo02: I am willing to make donations to help others that suffer from coronavirus.

The residual data for this model shows an approximately normal distribution once again.

Again, the R-Squared value for this model is lower than when calculated using the focus country data, at around 0.16, and again the P-Value is as low as possible. This makes sense, as the sample size is much larger.

The common attributes between this dataset and the focus country dataset are c19RCA01 and lifeSat, and for this dataset, all chosen attributes have P-Values that are as low as can be measured before we reach integer overflow issues.

c19ProSo03: I am willing to protect vulnerable groups from coronavirus even at my own expense.

With an approximately normally distributed set of residual values, and an extremely low P-Value again, c19ProSo03 also had some predictability, specifically around 11%, as indicated by the R-Squared Value.

The most important attributes for this model were different from those for our focus country data, however c19RCA01 still made an appearance, with the 4th highest correlation coefficient, and the minimum P-Value possible.

c19ProSo04: I am willing to make personal sacrifices to prevent the spread of coronavirus.

Once again, the residual values are roughly normally distributed around the median value, which is close to 0. There is also a very low P-Value, and an R-Squared value of 0.154, indicating that around 15% of the outcome variable values can be predicted from the predictor variables.

The most important attributes for predicting the outcome variable were the c19PerBeh variables, which concern the participants agreement with certain statements about their personal actions to mitigate the spread of coranavirus, such as avoiding crowded spaces, washing hands, and putting themselves in self quarantine. However, the 4th most important attribute was c19RCA02, indicating that the more participants agreed with the measure for mandatory quarantine for those with coronavirus, the more likely they were to be willing to make personal sacrifices to prevent the spread of the disease.

General Observations Regarding Other Country Regression Data

In general, when compared to our focus country, we saw a decrease in R-Squared values across the board, however, with a much larger sample size, we were able to see that each model had an extremely low P-Value. This indicates that the models do infact have predictive power for the outcome variables. Additionally, many more variables had extremely low P-Values for each of the models. For this reason, I had to find some other criteria to choose the attributes that seemed the most important to each model. Despite this, there were some attributes that remained important across all the datasets. The main attribute that seemed to contribute to all models was c19RCA01, which had some of the lowest P-Values in our focus country models, and additionally had the highest correlation coefficients in our other country data as well. Another attribute that seemed important, although it was not the most significant for every single model, was c19perBeh01, which was a major factor across the whole dataset.

3. Focus Country vs Cluster of Similar Countries

a)
When creating a cluster of similar countries, I chose 7 aspects to compare countries on:

GHS Index Overall Score	The GHS Index measures the capacities of 195 countries to prepare for epidemics and pandemics.(https://ghsindex.org/)
GDP Per Capita 2021	Average economic output per person.
Unemployment Rate 2021	Percentage of people unemployed.
Happiness Score 2019	A measure of overall wellbeing for a country provided by the World Happiness Report
Birth Rate Per 1000 2021	Births per 1000 people
Press Freedom Score 2021	A ranking of countries based on the Reporters Without Borders assessment of press freedom in the previous year.
Corruption Perception Index (CPI) Score 2021	A ranking of countries based on perceived levels of corruption by experts and opinion surveys. It is published by Transparency International, a non governmental organisation.

The selection of these indicators was to cover a wide range of aspects about the countries in the dataset, including health, economic, social, and political aspects. Additionally, I aimed to choose indicators that had a lot of data, to exclude as few countries as possible for clustering. However, some countries still had to be excluded as they were missing from some of the datasets. The sources for the data used are contained in the references for this report.

Hierarchical mean clustering was used to create groups of similar countries, with 12 clusters chosen. I have included a coloured dendrogram in the appendix, in addition to the pre-normalised table of values used. The eventual cluster that contained my focus country was:

- Bulgaria
- Armenia
- Georgia
- Argentina
- Panama
- Turkey
- **Brazil (Focus Country)**
- Colombia
- Albania
- Ukraine
- Bosnia and Herzegovina
- Greece
- Montenegro

b)
Model data, significant attribute tables, and model summaries are included in the appendix.

For this set of models, I chose significant attributes based on P-Value, taking the most significant variables with the lowset P-Values

c19ProSo01: I am willing to help others who suffer from coronavirus.

This Model had an approximately normal distribution of residual values, with the quartile ranges spread evenly around the median. With a very low P-Value, and an R-Squared of roughly 0.11, this model was able to predict around 10% of the outcome variable with the results from the predictors, and was statistically significant.

The significant attributes for this model were similar to the same model for the other countries data, however there were a couple differences. The first was the appearance of the rank_A attribute, which had a positive coefficient, which indicated that the lower (closer to 6) that participants rated beauty, the more likely they were to be willing to help others who suffer from coronavirus. The other differing attribute was consp01, which also had a positive coefficient, which shows that the more participants were in agreement with the statement:

I think that many very important things happen in the world, which the public is never informed about.

The more they were willing to help others who suffer from coronavirus.

c19ProSo02: I am willing to make donations to help others that suffer from coronavirus.

The R-Squared value for this model is 0.135, indicating that around 14% of the output variable can be predicted from the input predictors. The P-Value is the minimum possible, indicating statistical significance. The residual values are very slightly skewed towards the first percentile, however it is not by much, and the distribution is still approximately normal.

The most significant attributes are partially the same for the model for all other countries, sharing the attributes MLQ and c19RCA01, however c19RCA03 and consp02 are also significant for this model, as they have some of the lowest p values.

c19ProSo03: I am willing to protect vulnerable groups from coronavirus even at my own expense.

With an approximately normal distribute of residual values, and an R-Squared value of 0.11 and minimum P-value, the overall model looks similar to the model for this outcome variable using the dataset for all other countries.

However, the only shared variable with that model is c19RCA01, which has been significant in many of the models we have seen. Some other significant attributes for this model are rank_A, and rank_B, which are the rankings of beauty and achievement. These attributes have so far only been some of the most significant for this dataset only.

c19ProSo04: I am willing to make personal sacrifices to prevent the spread of coronavirus.

This model had an approximately normally distributed set of residual values, in addition to an R-Squared value of 0.16, and an extremely low P-Value, indicating that the model was very statistically significant.

The significant attributes for this model were different to the other datasets, with a high significance on the participants ranking of beauty and achievement. The significance of the c19perBeh attributes was similar to the other country data, which also placed a high level of significance on these attributes, but was very different from our focus country, which showed these attributes as having lower statistical significance.

General Observations about the Cluster Countries Regression Data

In general, when looking at residual values, R-Squared values, and P-Values, this cluster of countries is most similar the other countries dataset, rather than our focus country. However, these values are also dependant on sample size, and as the cluster of countries has a much larger sample size than our focus country on its own, this result is to be expected.

Additionally, when looking at significant attributes, there is not a lot of similarity between the cluster of countries and our focus country. In particular, the ranking order attributes had a significantly increased presence in the models used with this dataset, but were not seen as the most significant in the other datasets. The main significant attributes that were the same between the cluster countries and our focus country were c19RCA01 and c19PerBeh01, both of which had a significant presence across many models and across all datasets.

The limited similarly between the cluster of countries and our focus country may indicate that the indicators used in our clustering table could use some improvement, and that they were not the best choice to find similarity between countries. Additionally, while the indicators used may lead to clusters of countries that are similar economically, politically, socially and healthwise, it may not indicate that the countries are similar in terms of culture, and that people from these countries could respond differently to the questions posed in the survey.

Overall, I would say that the group of other countries that were not the focus country had a slightly better match of significant attributes with my focus country, but both the other countries, and the cluster of similar countries were not very similar with the focus country overall, but instead were similar to each other, in terms of overall model values, in addition to having more significant attributes in common.

Appendix

Figures (Question 2a)

Fig 1.

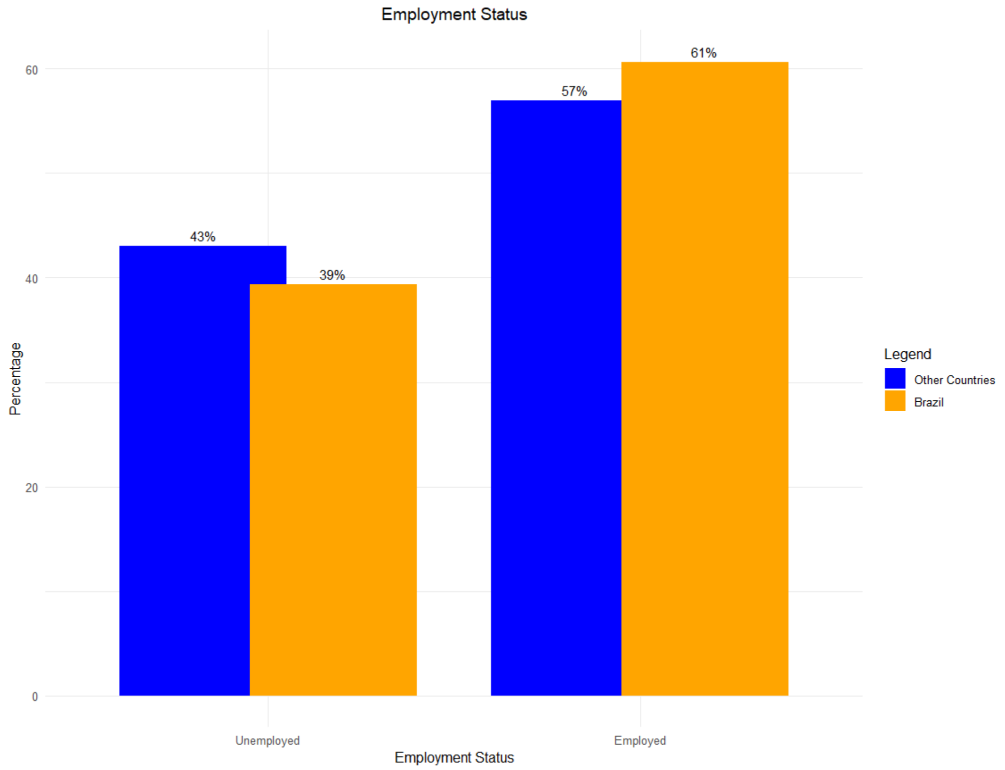


Fig 2.

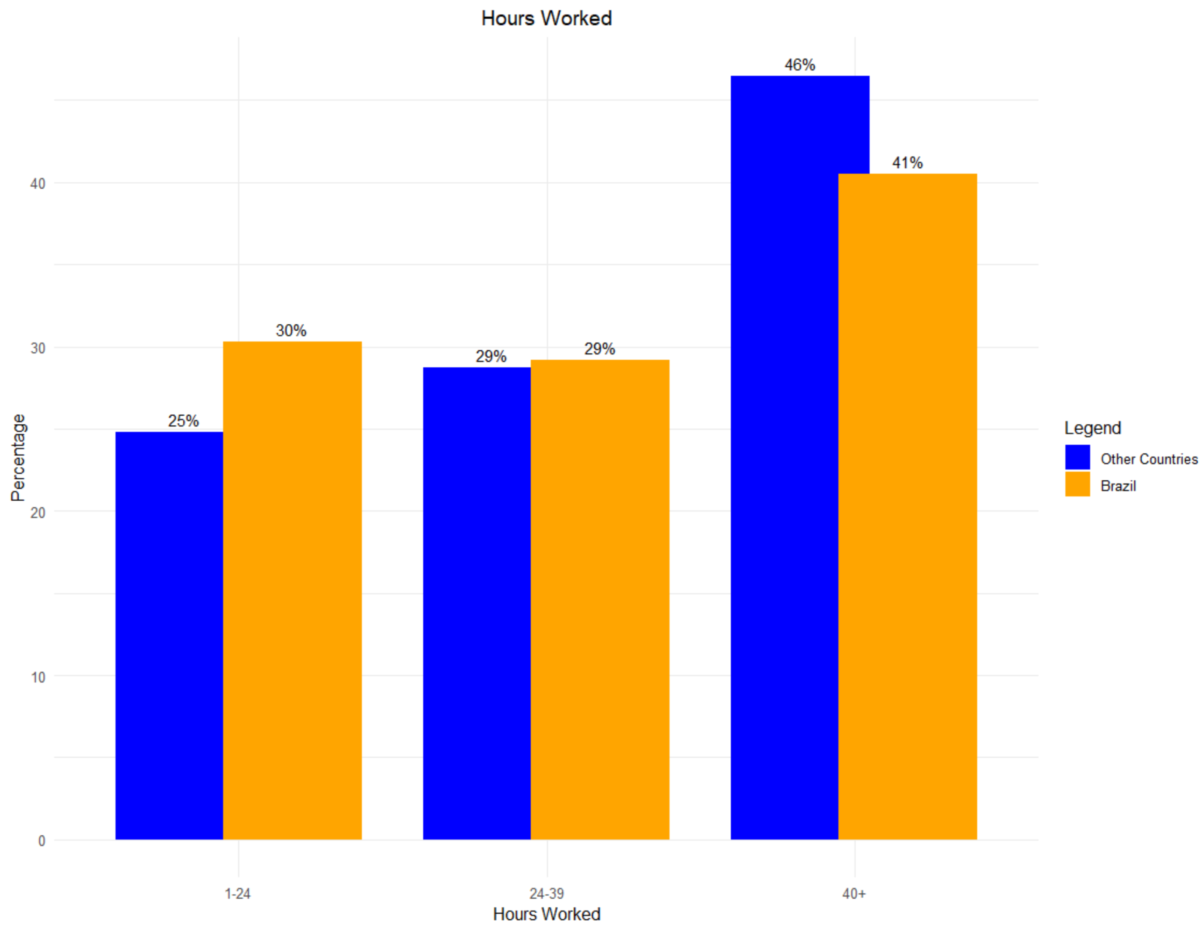


Fig 3.

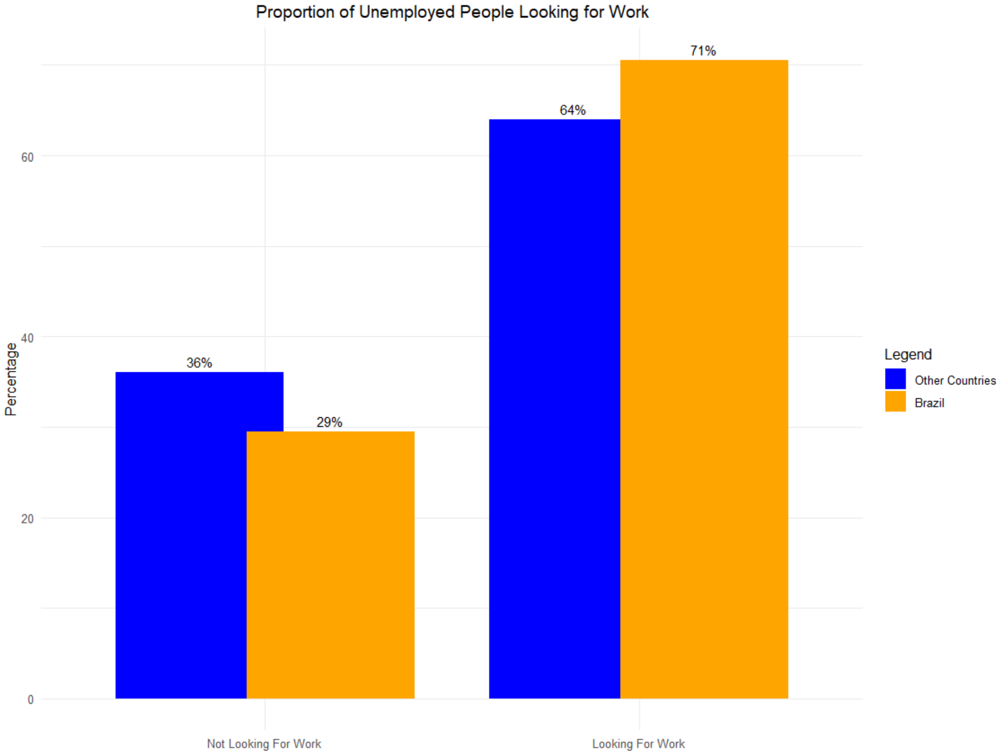

Proportion of Unemployed People Looking for Work
Percentage
0
20
40
60
36%
29%
64%
71%
Not Looking For Work
Looking For Work
Legend
Other Countries
Brazil

Fig 4.

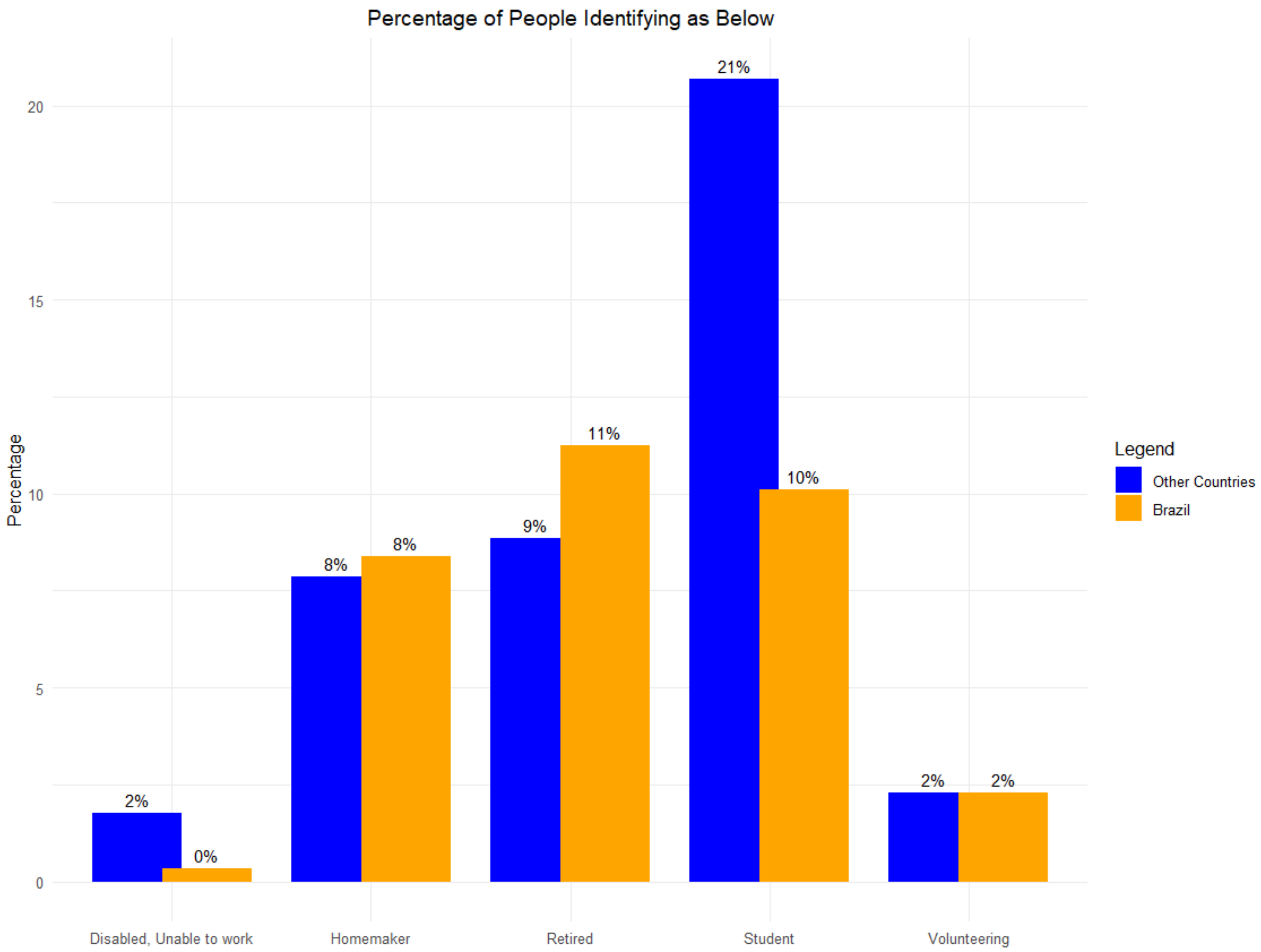
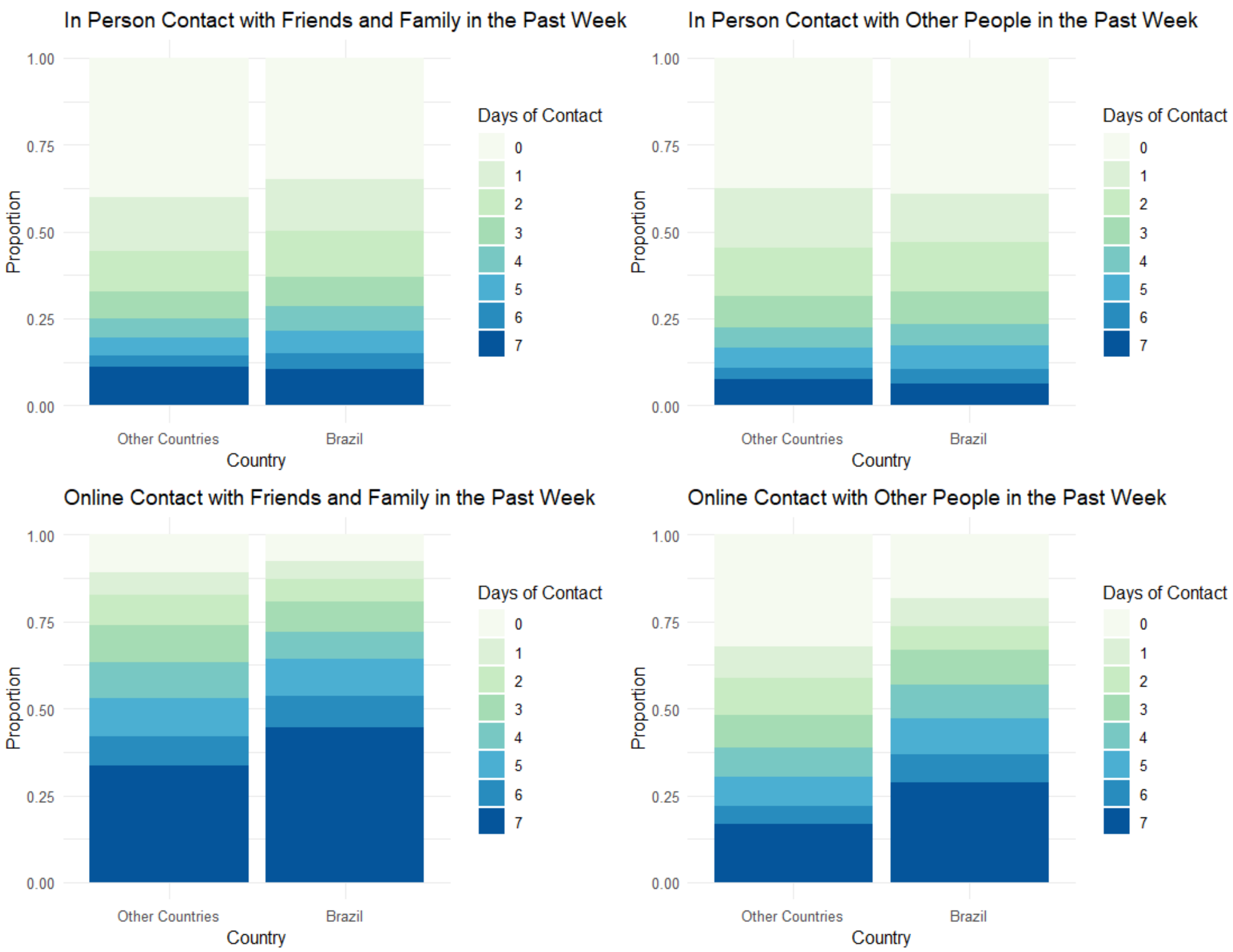
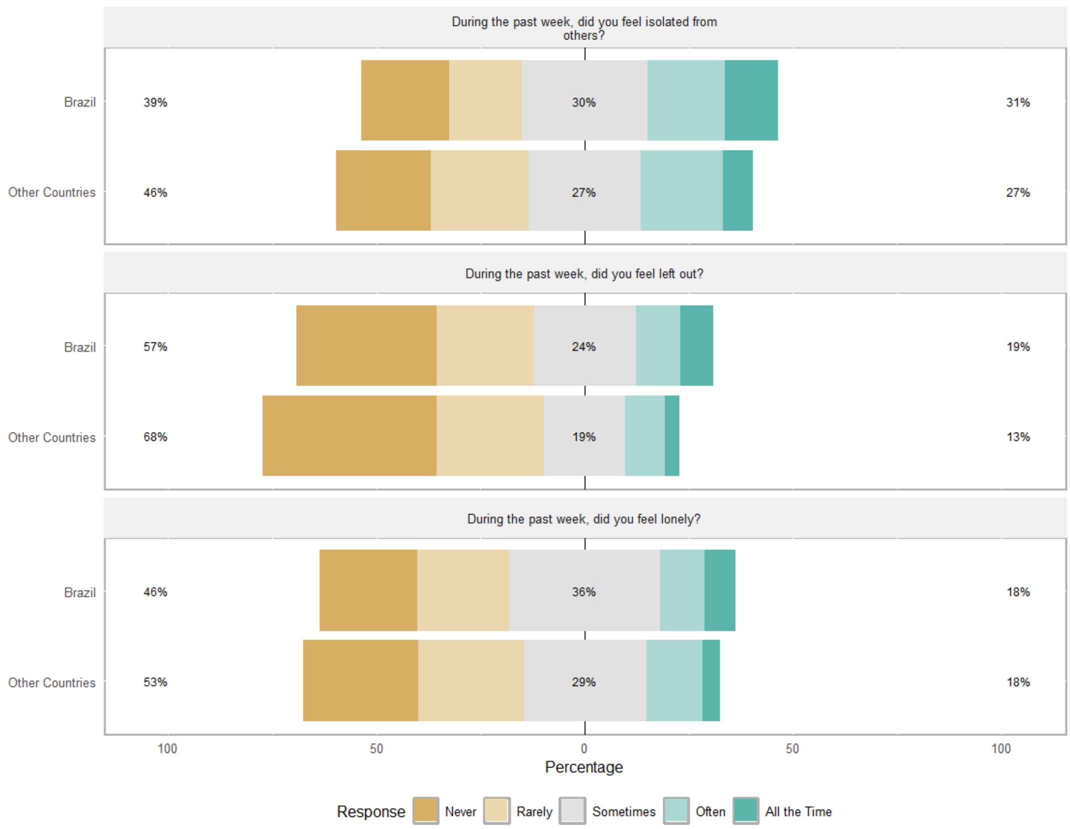

Percentage of People Identifying as Below
Percentage
0
5
10
15
20
2%
0%
8%
8%
9%
11%
21%
10%
2%
2%
Disabled, Unable to work
Homemaker
Retired
Student
Volunteering
Legend
Other Countries
Brazil

Fig 5.



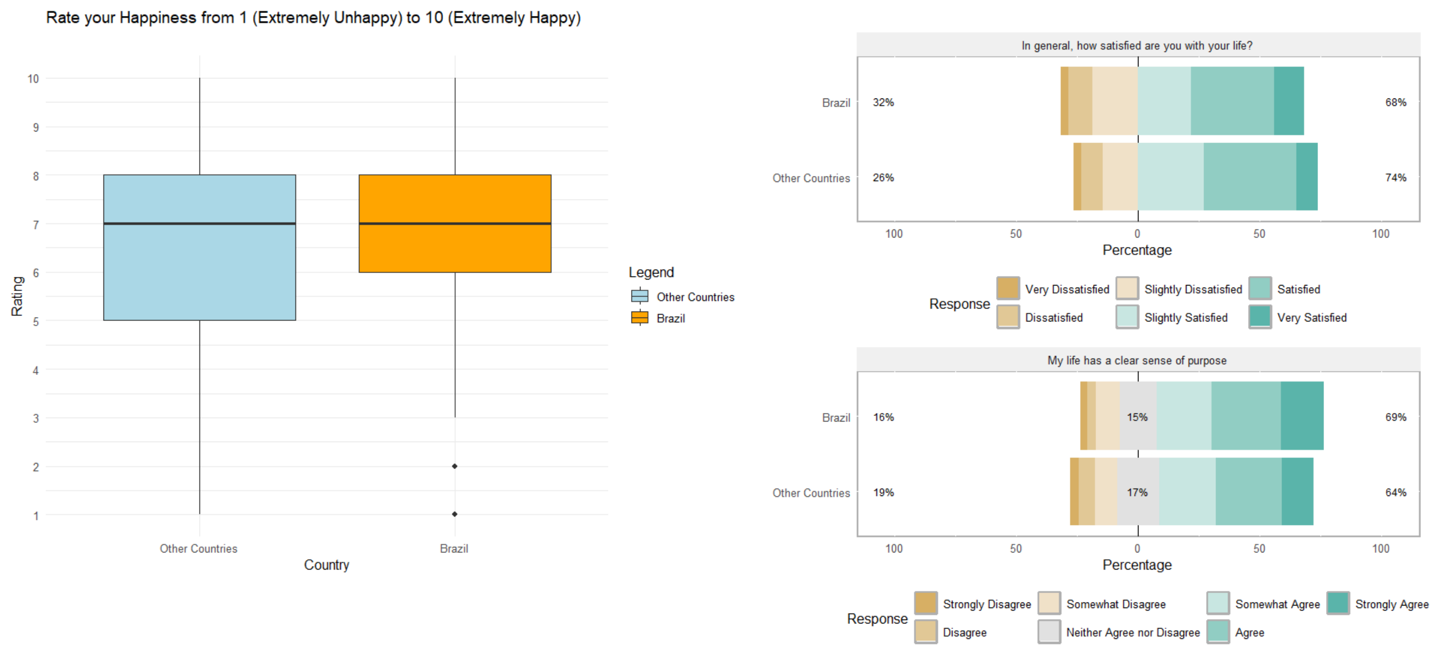
In Person Contact with Friends and Family in the Past Week
In Person Contact with Other People in the Past Week
Online Contact with Friends and Family in the Past Week
Online Contact with Other People in the Past Week
Proportion
0.00
0.25
0.50
0.75
1.00
Other Countries
Brazil
Country
Days of Contact
0
1
2
3
4
5
6
7

Fig 6.



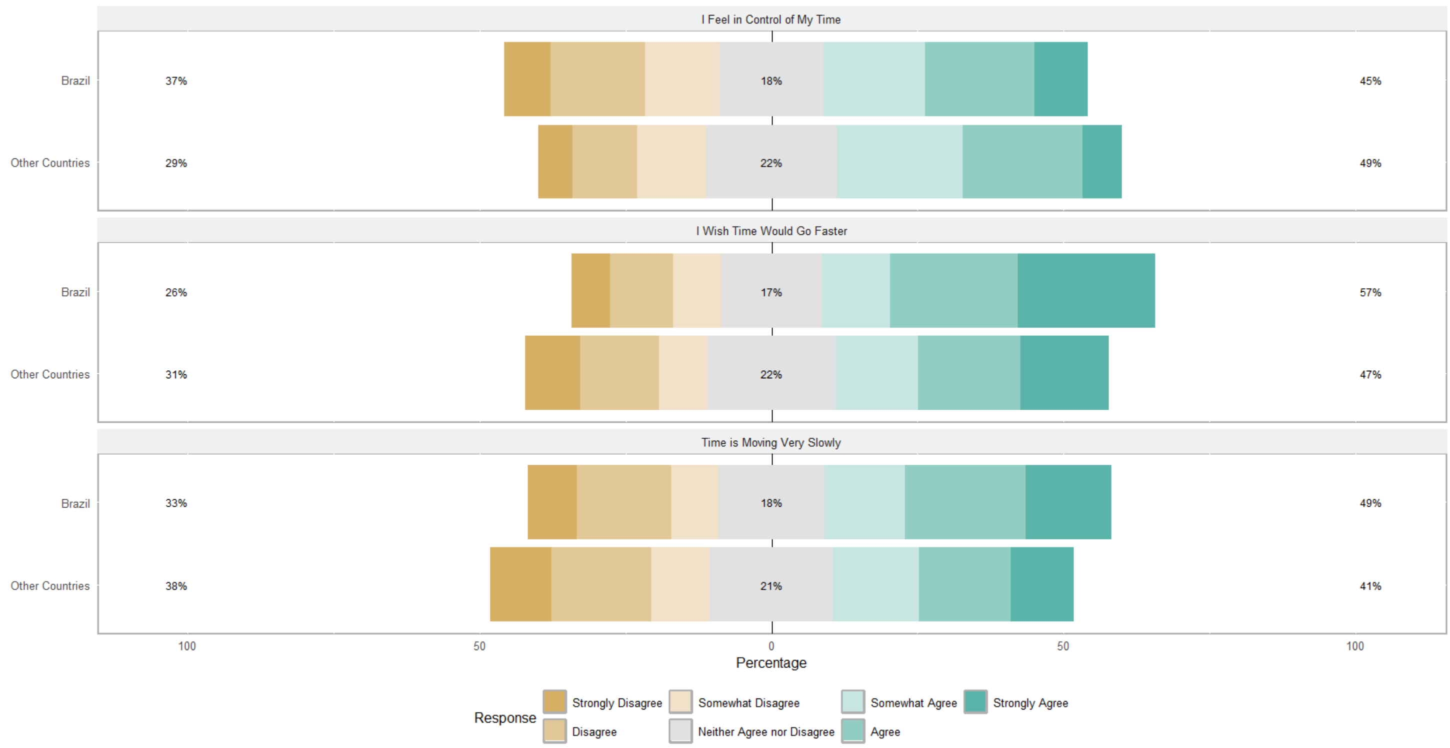
During the past week, did you feel isolated from others?
Brazil 39% 30% 31%
Other Countries 46% 27% 27%
During the past week, did you feel left out?
Brazil 57% 24% 19%
Other Countries 68% 19% 13%
During the past week, did you feel lonely?
Brazil 46% 36% 18%
Other Countries 53% 29% 18%
100
50
0
50
100
Percentage
Response
Never
Rarely
Sometimes
Often
All the Time

Fig 7.



Rate your Happiness from 1 (Extremely Unhappy) to 10 (Extremely Happy)
Rating
Country
Other Countries
Brazil
Legend
Other Countries
Brazil
In general, how satisfied are you with your life?
Brazil 32% 68%
Other Countries 26% 74%
Percentage
Response
Very Dissatisfied
Dissatisfied
Slightly Dissatisfied
Slightly Satisfied
Satisfied
Very Satisfied
My life has a clear sense of purpose
Brazil 16% 15% 69%
Other Countries 19% 17% 64%
Percentage
Response
Strongly Disagree
Disagree
Somewhat Disagree
Neither Agree nor Disagree
Somewhat Agree
Agree
Strongly Agree

Fig 8.



I Feel in Control of My Time
Brazil 37% 18% 45%
Other Countries 29% 22% 49%
I Wish Time Would Go Faster
Brazil 26% 17% 57%
Other Countries 31% 22% 47%
Time is Moving Very Slowly
Brazil 33% 18% 49%
Other Countries 38% 21% 41%
Percentage
Response
Strongly Disagree
Disagree
Somewhat Disagree
Neither Agree nor Disagree
Somewhat Agree
Agree
Strongly Agree

Fig 9.

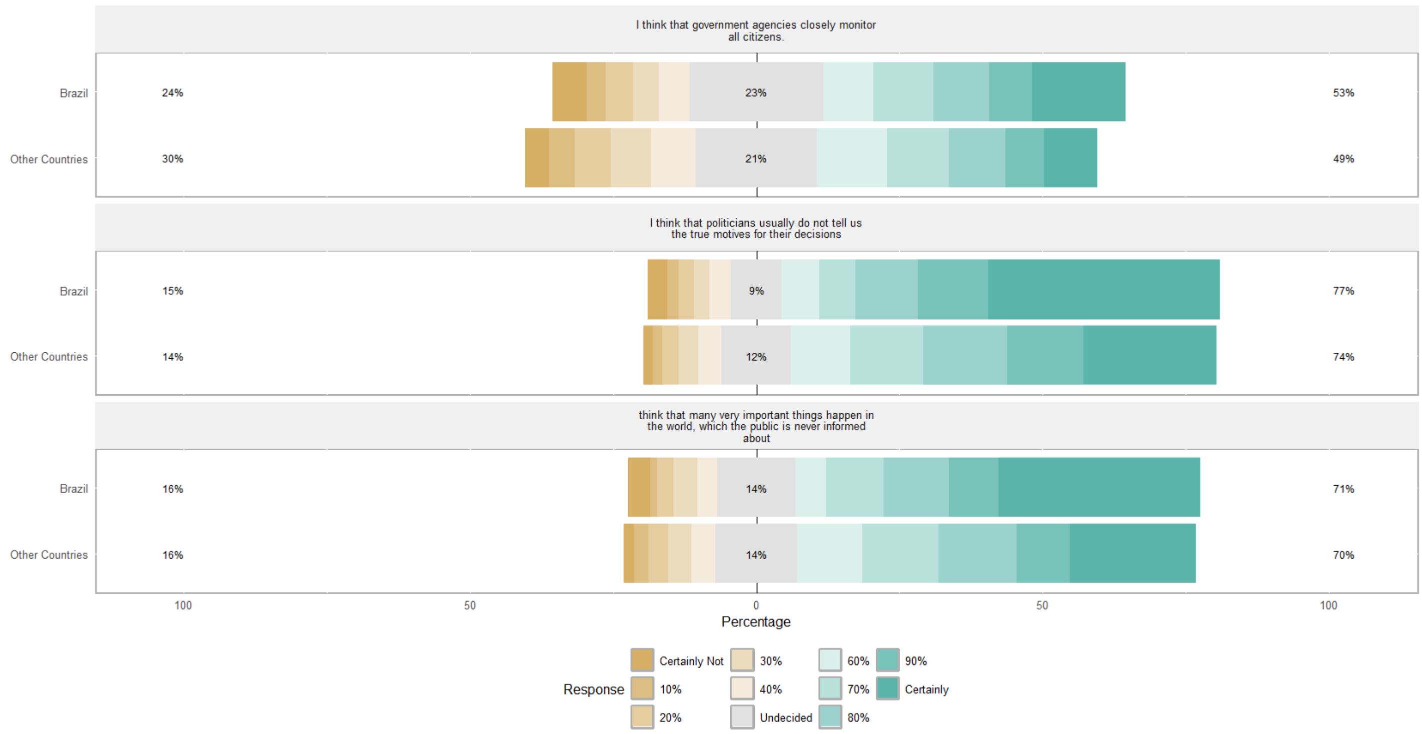

I think that government agencies closely monitor all citizens.
Brazil
24%
23%
53%
Other Countries
30%
21%
49%
I think that politicians usually do not tell us the true motives for their decisions
Brazil
15%
9%
77%
Other Countries
14%
12%
74%
think that many very important things happen in the world, which the public is never informed about
Brazil
16%
14%
71%
Other Countries
16%
14%
70%
100
50
0
50
100
Percentage
Response
Certainly Not
10%
20%
30%
40%
Undecided
60%
70%
80%
90%
Certainly

Fig 10.

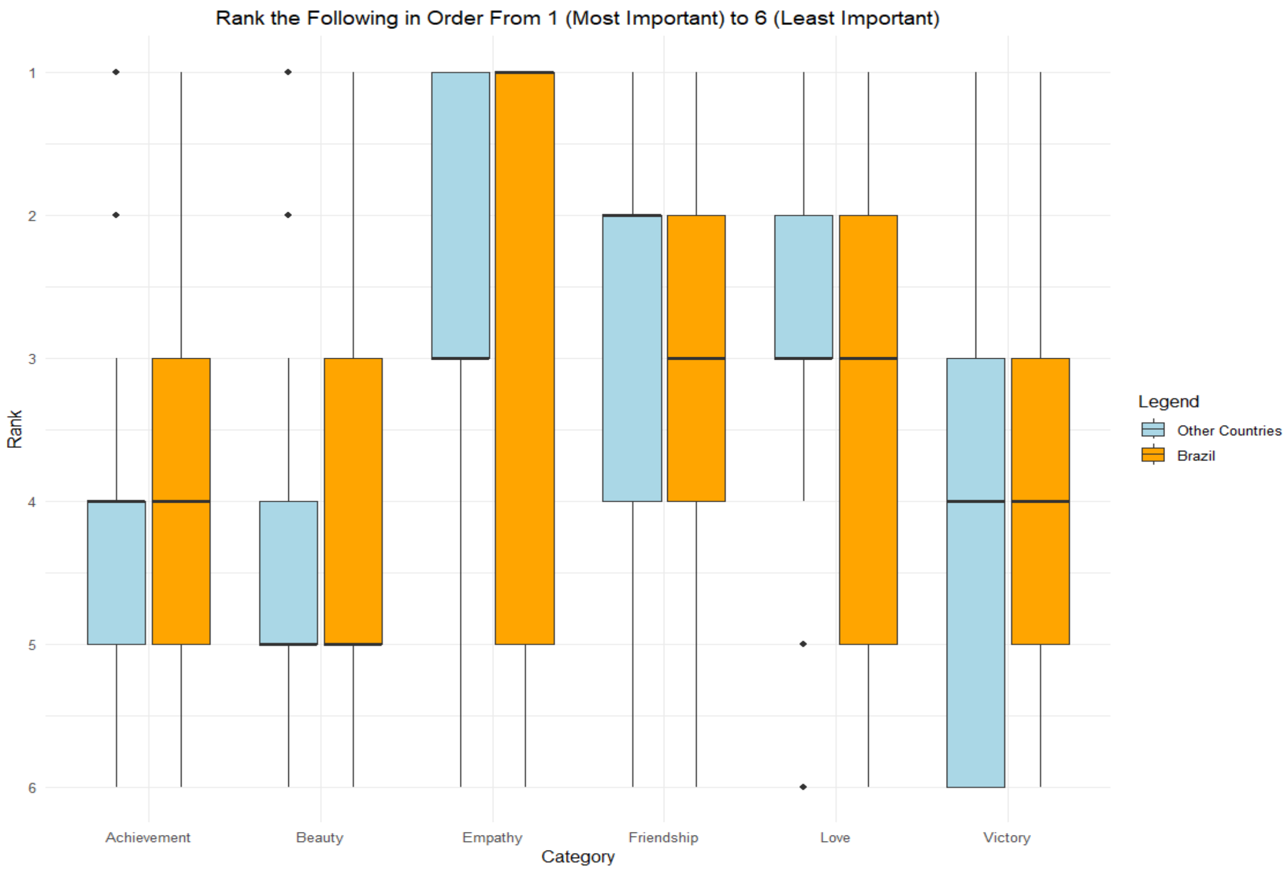
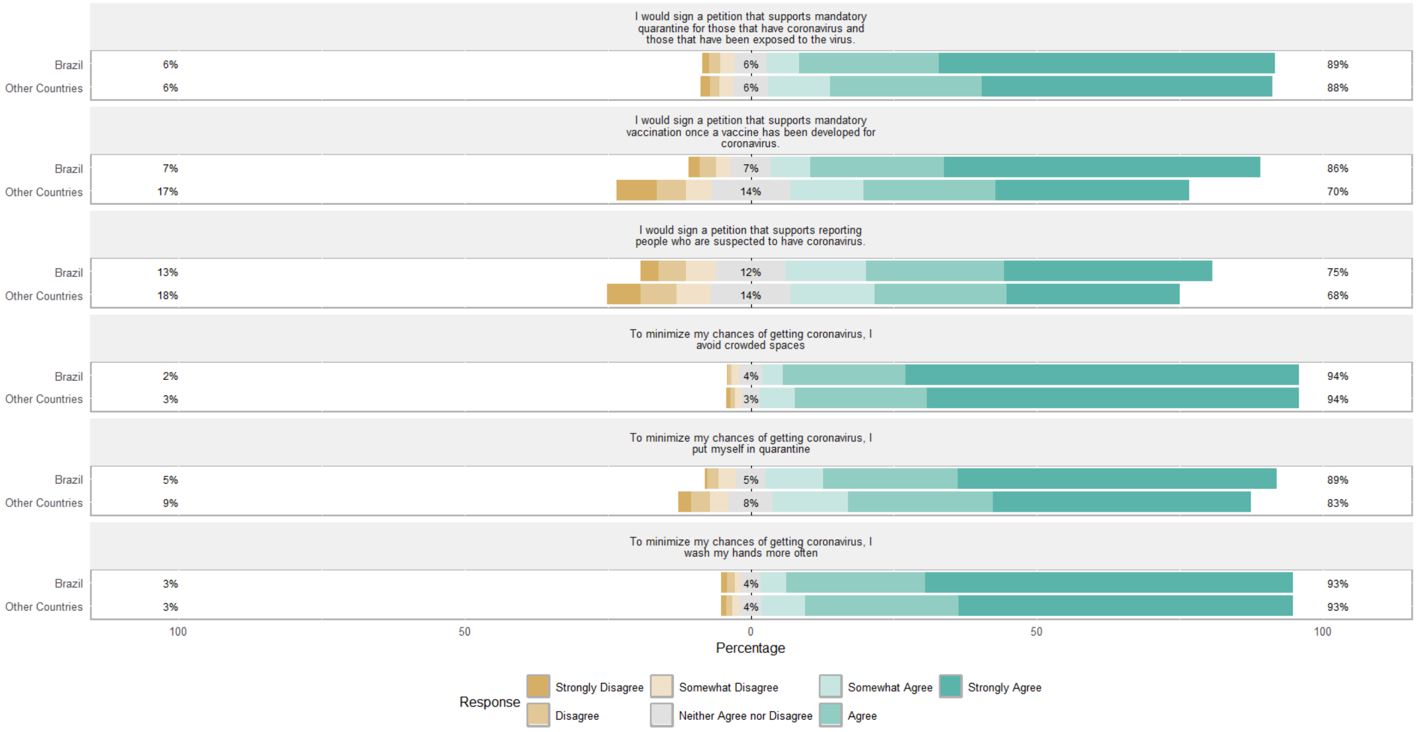
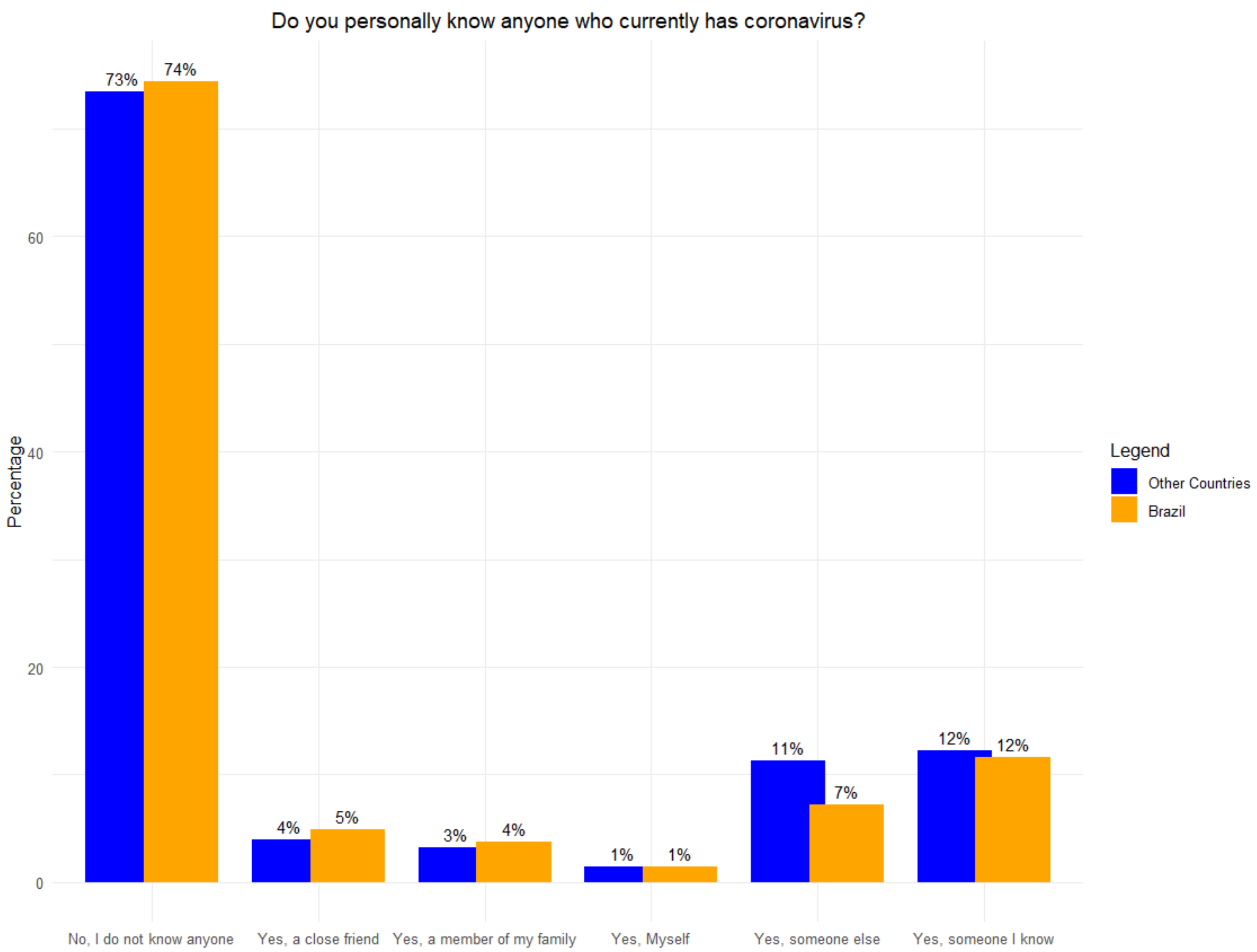

Rank the Following in Order From 1 (Most Important) to 6 (Least Important)
Rank
1
2
3
4
5
6
Achievement
Beauty
Empathy
Friendship
Love
Victory
Category
Legend
Other Countries
Brazil

Fig 11.



I would sign a petition that supports mandatory quarantine for those that have coronavirus and those that have been exposed to the virus.
Brazil 6% 6% 89%
Other Countries 6% 6% 88%
I would sign a petition that supports mandatory vaccination once a vaccine has been developed for coronavirus.
Brazil 7% 7% 86%
Other Countries 17% 14% 70%
I would sign a petition that supports reporting people who are suspected to have coronavirus.
Brazil 13% 12% 75%
Other Countries 18% 14% 68%
To minimize my chances of getting coronavirus, I avoid crowded spaces
Brazil 2% 4% 94%
Other Countries 3% 3% 94%
To minimize my chances of getting coronavirus, I put myself in quarantine
Brazil 5% 5% 89%
Other Countries 9% 8% 83%
To minimize my chances of getting coronavirus, I wash my hands more often
Brazil 3% 4% 93%
Other Countries 3% 4% 93%
100
50
0
50
100
Percentage
Response
Strongly Disagree
Disagree
Somewhat Disagree
Neither Agree nor Disagree
Somewhat Agree
Agree
Strongly Agree

Fig 12.



Do you personally know anyone who currently has coronavirus?
Percentage
60
40
20
0
73%
74%
4%
5%
3%
4%
1%
1%
11%
7%
12%
12%
No, I do not know anyone
Yes, a close friend
Yes, a member of my family
Yes, Myself
Yes, someone else
Yes, someone I know
Legend
Other Countries
Brazil

Fig 13.

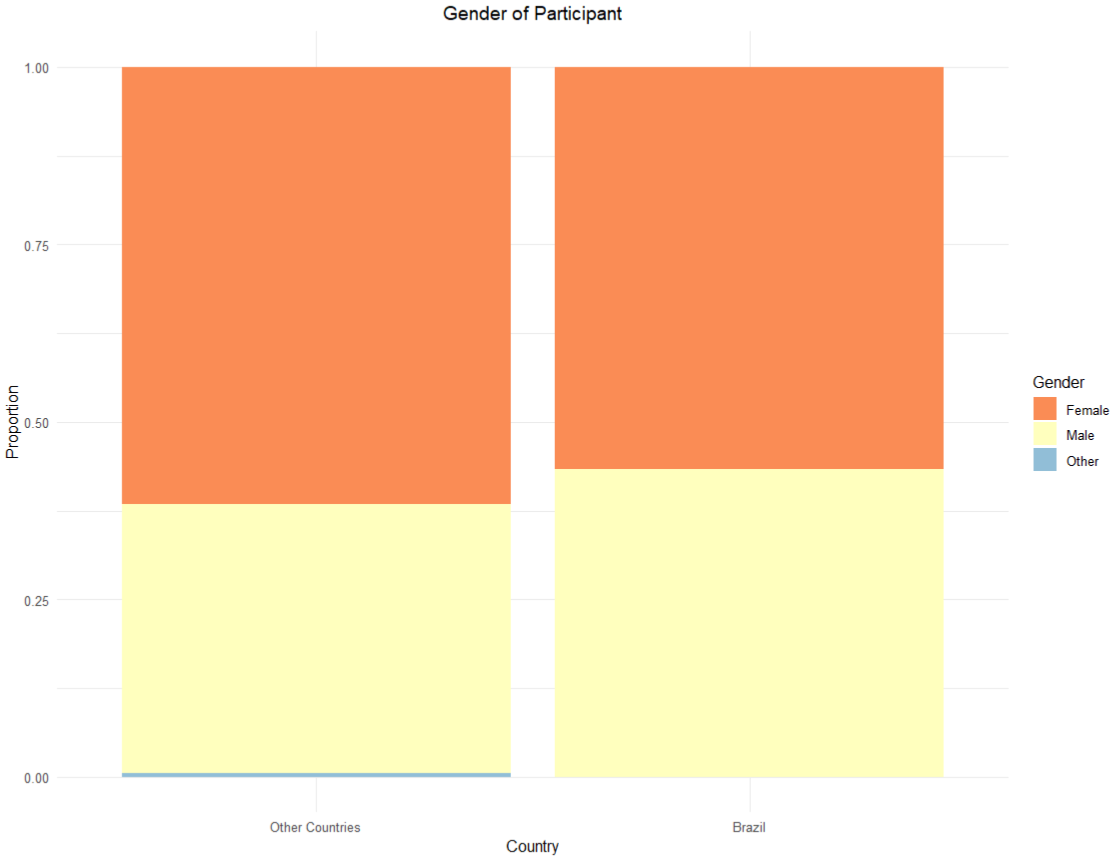

Gender of Participant
1.00
0.75
0.50
0.25
0.00
Proportion
Other Countries
Brazil
Country
Gender
Female
Male
Other

Fig 14.

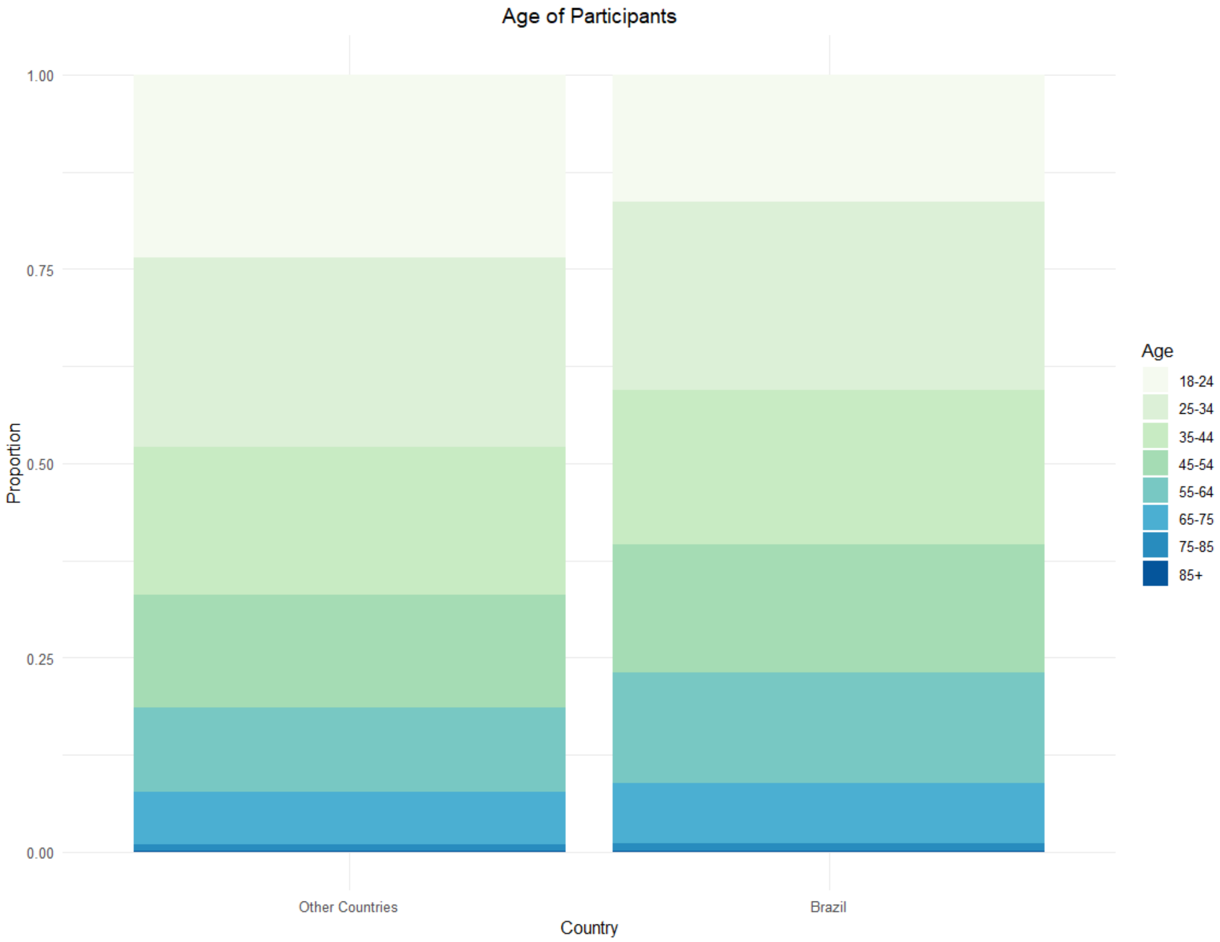

Age of Participants
1.00
0.75
0.50
0.25
0.00
Proportion
Other Countries
Brazil
Country
Age
18-24
25-34
35-44
45-54
55-64
65-75
75-85
85+

Fig 15.

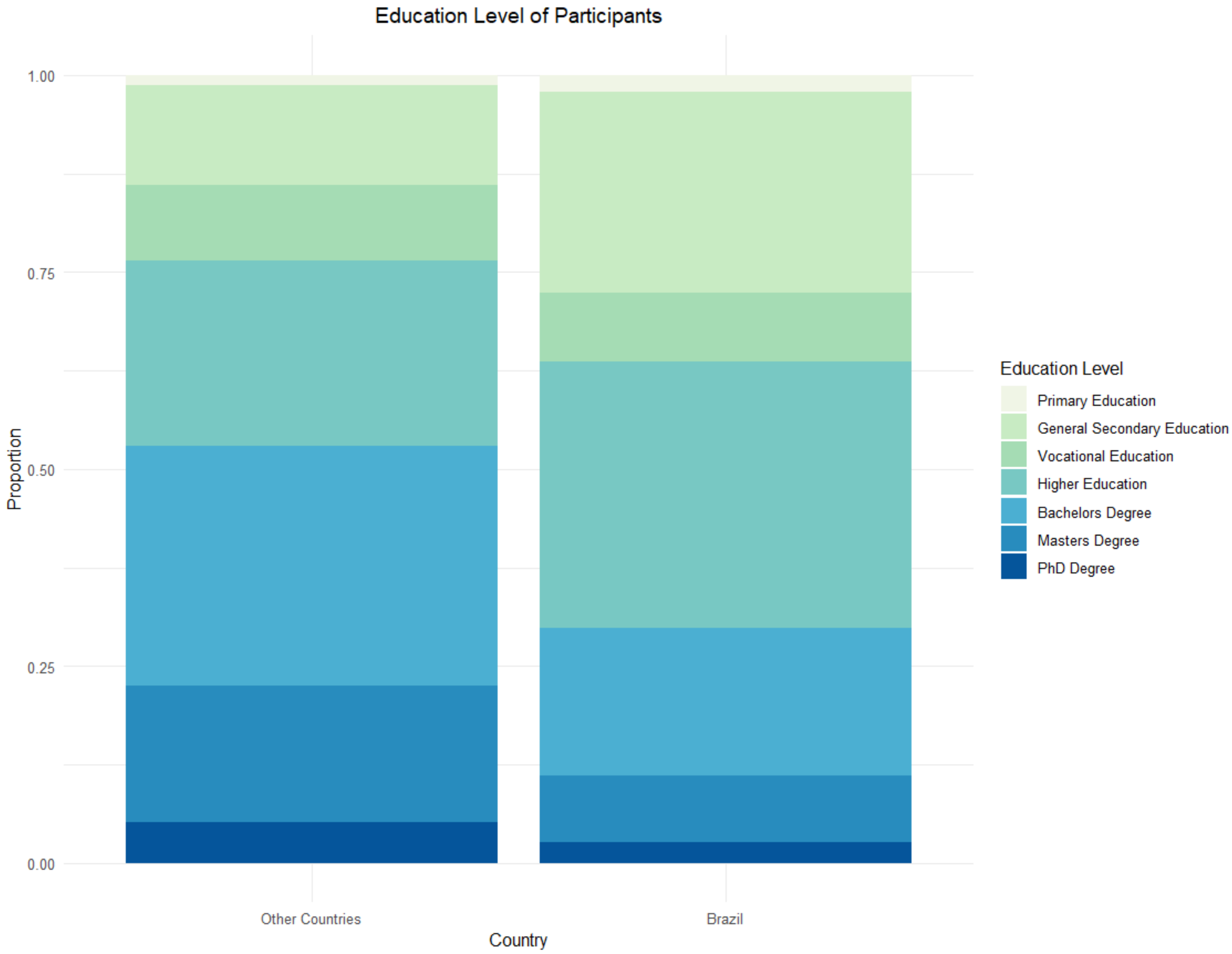

Education Level of Participants
1.00
0.75
0.50
0.25
0.00
Proportion
Other Countries
Brazil
Country
Education Level
Primary Education
General Secondary Education
Vocational Education
Higher Education
Bachelors Degree
Masters Degree
PhD Degree

Fig 16.

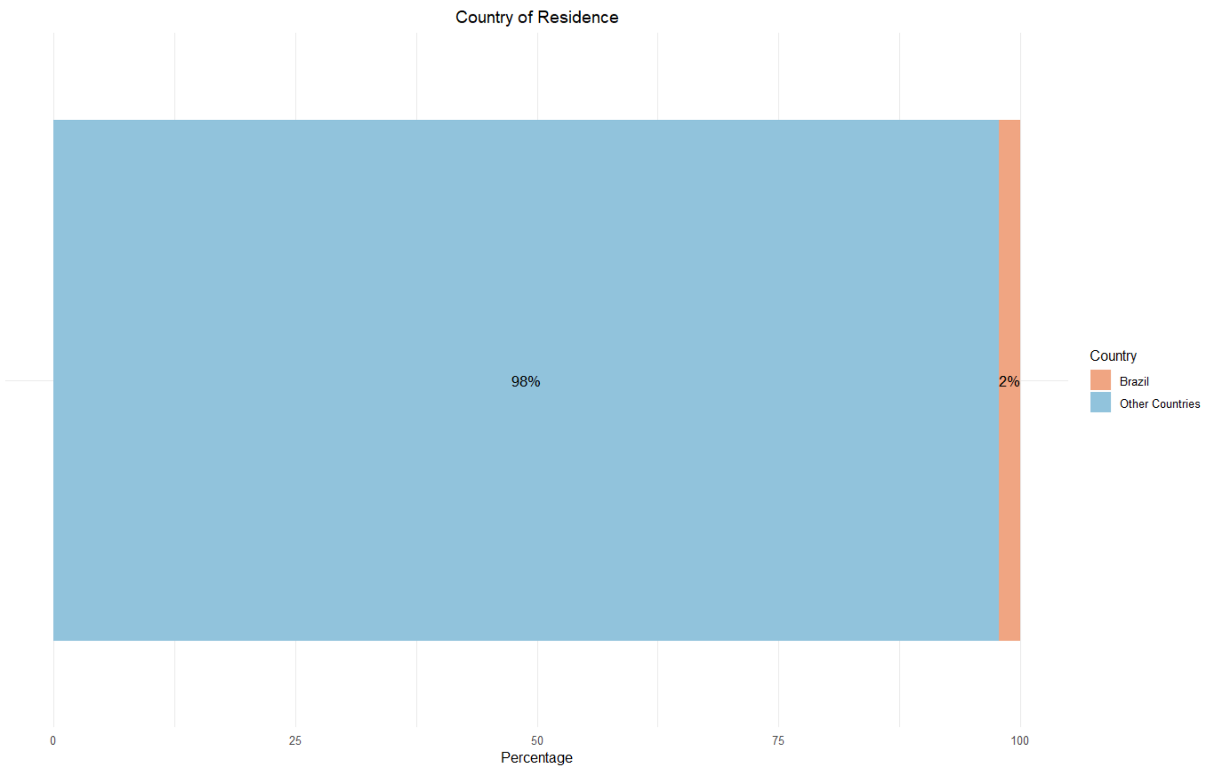
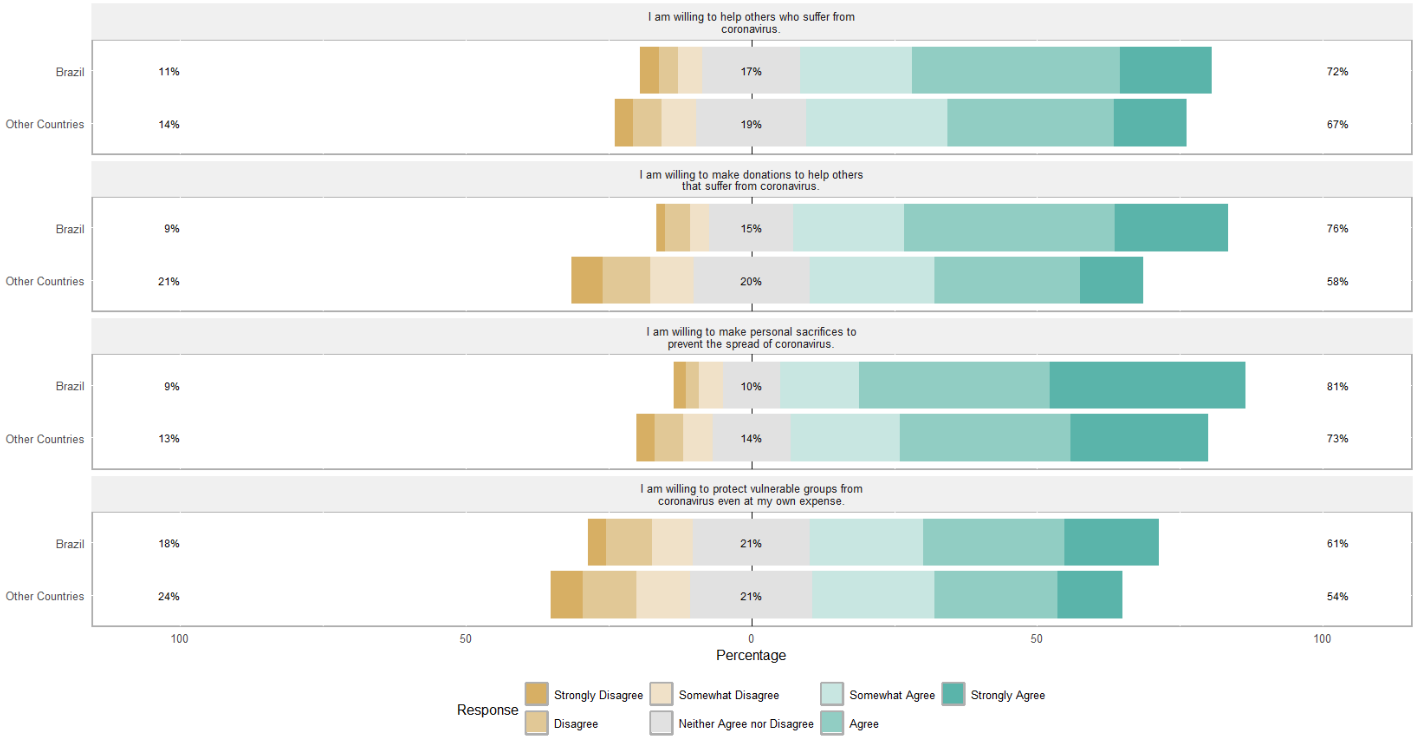

Country of Residence
98%
2%
Country
Brazil
Other Countries
0
25
50
75
100
Percentage

Fig 17.



I am willing to help others who suffer from coronavirus.
Brazil
11%
17%
72%
Other Countries
14%
19%
67%
I am willing to make donations to help others that suffer from coronavirus.
Brazil
9%
15%
76%
Other Countries
21%
20%
58%
I am willing to make personal sacrifices to prevent the spread of coronavirus.
Brazil
9%
10%
81%
Other Countries
13%
14%
73%
I am willing to protect vulnerable groups from coronavirus even at my own expense.
Brazil
18%
21%
61%
Other Countries
24%
21%
54%
100
50
0
50
100
Percentage
Response
Strongly Disagree
Disagree
Somewhat Disagree
Neither Agree nor Disagree
Somewhat Agree
Agree
Strongly Agree

Brazil Regression Data (2b)

Brazil c19ProSo01
R-Squared = 0.1954
P-Value = 6.416e-15
Residual Data:

1Q	Median	3Q
-0.7084	0.2260	0.9187

Attribute	c19perBeh01	bor01	c19RCA01	employstatus_10
Meaning	To minimize my chances of getting coronavirus, I wash my hands more often	I wish time would go by faster	I would sign a petition that Supports mandatory vaccination once a vaccine has been developed for coronavirus.	I am a volunteer
P-value	9.1e-05	0.00145	0.00305	0.00499
Est. Coefficient	0.268787	0.113238	0.127370	0.915019

```
Residuals:
    Min      1Q  Median      3Q     Max 
-4.2202 -0.7084  0.2260  0.9187  3.2779 

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)          1.027648   0.775762   1.325  0.18569    
employstatus_1       0.263493   0.161530   1.631  0.10328    
employstatus_2       0.230316   0.171364   1.344  0.17936    
employstatus_3       0.271696   0.166616   1.631  0.10339    
employstatus_4       0.340280   0.196318   1.733  0.08347 .  
employstatus_5      -0.287392   0.248595  -1.156  0.24803    
employstatus_6       0.019290   0.211161   0.091  0.92724    
employstatus_7      -0.073150   0.208744  -0.350  0.72612    
employstatus_8      -0.349303   0.785307  -0.445  0.65660    
employstatus_9      -0.119180   0.190578  -0.625  0.53193    
employstatus_10      0.915019   0.324902   2.816  0.00499 ** 
isoFriends_inPerson  0.043176   0.022615   1.909  0.05663 .  
isoOthPpl_inPerson  -0.023372   0.026295  -0.889  0.37438    
isoFriends_online    0.046268   0.026524   1.744  0.08152 .  
isoOthPpl_online    -0.001602   0.023274  -0.069  0.94514    
lone01               0.011354   0.059677   0.190  0.84916    
lone02               0.080262   0.052928   1.516  0.12984    
lone03              -0.052826   0.055978  -0.944  0.34564    
happy                0.037524   0.030649   1.224  0.22123    
lifeSat              0.070330   0.053123   1.324  0.18595    
MLQ                  0.083301   0.040779   2.043  0.04144 *  
bor01                0.113238   0.035412   3.198  0.00145 ** 
bor02               -0.058301   0.034579  -1.686  0.09222 .  
bor03                0.010757   0.028551   0.377  0.70647    
consp01             -0.036394   0.022800  -1.596  0.11087    
consp02              0.019609   0.023394   0.838  0.40220    
consp03              0.024733   0.018489   1.338  0.18140    
rank_A               0.062813   0.037088   1.694  0.09076 .  
rank_B               0.010314   0.036635   0.282  0.77838    
rank_C               0.053834   0.036942   1.457  0.14547    
rank_D              -0.040461   0.038456  -1.052  0.29309    
rank_E               0.013177   0.034848   0.378  0.70545    
rank_F                     NA         NA      NA       NA    
c19perBeh01          0.268787   0.068298   3.935  9.1e-05 ***
c19perBeh02         -0.073911   0.090415  -0.817  0.41393    
c19perBeh03         -0.048859   0.054118  -0.903  0.36693    
c19RCA01             0.127370   0.042848   2.973  0.00305 ** 
c19RCA02            -0.027231   0.055925  -0.487  0.62647    
c19RCA03             0.068657   0.034232   2.006  0.04527 *  
coronaClose_1        0.771512   0.435449   1.772  0.07685 .  
coronaClose_2       -0.141697   0.354359  -0.400  0.68937    
coronaClose_3       -0.079037   0.314859  -0.251  0.80187    
coronaClose_4       -0.093658   0.303295  -0.309  0.75756    
coronaClose_5       -0.023083   0.322547  -0.072  0.94297    
coronaClose_6       -0.115375   0.310539  -0.372  0.71035    
gender               0.154669   0.102871   1.504  0.13314    
age                 -0.033869   0.040007  -0.847  0.39752    
edu                 -0.021447   0.035437  -0.605  0.54523    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.295 on 726 degrees of freedom
  (98 observations deleted due to missingness)
Multiple R-squared:  0.1954,	Adjusted R-squared:  0.1445 
F-statistic: 3.834 on 46 and 726 DF,  p-value: 6.416e-15
```

Brazil c19ProSo02
R-Squared = 0.2553
P-Value < 2.2e-16
Residual Data:

1Q	Median	3Q
-0.6236	0.1735	0.8230

Attribute	c19RCA01	lifeSat	c19perBeh03	employstatus_10
Meaning	I would sign a petition that Supports mandatory vaccination once a vaccine has been developed for coronavirus.	In general, how satisfied are you with your life?	To minimize my chances of getting coronavirus, put myself in quarantine.	I am a volunteer
P-Value	0.00124	0.00339	0.01159	0.01991
Est. Coefficient	0.127789	0.143611	0.125953	0.697136

```
Residuals:
    Min      1Q  Median      3Q     Max 
-4.3156 -0.6236  0.1735  0.8230  2.6728 

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)   
(Intercept)          1.670808   0.713423   2.342  0.01945 * 
employstatus_1       0.011165   0.148550   0.075  0.94011   
employstatus_2      -0.045065   0.157594  -0.286  0.77499   
employstatus_3       0.161085   0.153227   1.051  0.29348   
employstatus_4      -0.199893   0.180542  -1.107  0.26858   
employstatus_5       0.080125   0.228618   0.350  0.72608   
employstatus_6      -0.274725   0.194192  -1.415  0.15758   
employstatus_7       0.095659   0.191970   0.498  0.61842   
employstatus_8       0.998532   0.722201   1.383  0.16721   
employstatus_9      -0.121033   0.175264  -0.691  0.49005   
employstatus_10      0.697136   0.298793   2.333  0.01991 * 
isoFriends_inPerson  0.024682   0.020798   1.187  0.23571   
isoOthPpl_inPerson  -0.018222   0.024182  -0.754  0.45138   
isoFriends_online    0.005755   0.024393   0.236  0.81354   
isoOthPpl_online     0.042170   0.021404   1.970  0.04919 * 
lone01              -0.019528   0.054882  -0.356  0.72208   
lone02               0.100749   0.048675   2.070  0.03882 * 
lone03              -0.020733   0.051479  -0.403  0.68726   
happy                0.058195   0.028186   2.065  0.03931 * 
lifeSat              0.143611   0.048854   2.940  0.00339 **
MLQ                  0.012099   0.037502   0.323  0.74708   
bor01                0.066722   0.032566   2.049  0.04084 * 
bor02               -0.015403   0.031801  -0.484  0.62828   
bor03                0.019142   0.026257   0.729  0.46622   
consp01             -0.002034   0.020967  -0.097  0.92274   
consp02              0.015281   0.021514   0.710  0.47777   
consp03             -0.016183   0.017003  -0.952  0.34154   
rank_A               0.011387   0.034107   0.334  0.73858   
rank_B               0.057070   0.033691   1.694  0.09071 . 
rank_C               0.007494   0.033973   0.221  0.82547   
rank_D              -0.060389   0.035366  -1.708  0.08815 . 
rank_E              -0.024894   0.032047  -0.777  0.43754   
rank_F                     NA         NA      NA       NA   
c19perBeh01          0.111697   0.062810   1.778  0.07577 . 
c19perBeh02         -0.018788   0.083149  -0.226  0.82130   
c19perBeh03          0.125953   0.049769   2.531  0.01159 * 
c19RCA01             0.127789   0.039405   3.243  0.00124 **
c19RCA02             0.048456   0.051431   0.942  0.34642   
c19RCA03             0.052867   0.031482   1.679  0.09352 . 
coronaClose_1        0.283990   0.400457   0.709  0.47845   
coronaClose_2        0.236835   0.325883   0.727  0.46762   
coronaClose_3       -0.154801   0.289557  -0.535  0.59308   
coronaClose_4       -0.056437   0.278922  -0.202  0.83971   
coronaClose_5        0.002255   0.296628   0.008  0.99394   
coronaClose_6       -0.130565   0.285584  -0.457  0.64768   
gender              -0.176988   0.094604  -1.871  0.06177 . 
age                 -0.076370   0.036792  -2.076  0.03827 * 
edu                 -0.070252   0.032590  -2.156  0.03144 * 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.191 on 726 degrees of freedom
  (98 observations deleted due to missingness)
Multiple R-squared:  0.2553,	Adjusted R-squared:  0.2081 
F-statistic: 5.409 on 46 and 726 DF,  p-value: < 2.2e-16
```

Brazil c19ProSo03

R-Squared = 0.16

P-Value = 5.383e-10

Residual Data:

1Q	Median	3Q
-0.9242	0.2141	1.0945

Attribute	bor01	c19RCA01	Age	employstatus_10
Meaning	I wish time would go by faster	I would sign a petition that Supports mandatory vaccination once a vaccine has been developed for coronavirus.	Age range of the participant	I am a volunteer
P-Value	0.00119	0.01492	0.02046	0.02774
Est Coefficient	0.134965	0.122448	-0.108830	0.839129

```
Residuals:
    Min      1Q  Median      3Q     Max 
-4.6792 -0.9242  0.2141  1.0945  3.4769 

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)   
(Intercept)          2.430460   0.908476   2.675  0.00763 **
employstatus_1       0.090490   0.189164   0.478  0.63253   
employstatus_2      -0.022153   0.200681  -0.110  0.91213   
employstatus_3       0.355649   0.195120   1.823  0.06876 . 
employstatus_4      -0.104329   0.229903  -0.454  0.65011   
employstatus_5      -0.465195   0.291124  -1.598  0.11049   
employstatus_6      -0.407070   0.247285  -1.646  0.10016   
employstatus_7      -0.138237   0.244455  -0.565  0.57191   
employstatus_8       1.191698   0.919654   1.296  0.19545   
employstatus_9      -0.268513   0.223182  -1.203  0.22932   
employstatus_10      0.839129   0.380485   2.205  0.02774 * 
isoFriends_inPerson  0.001448   0.026484   0.055  0.95640   
isoOthPpl_inPerson   0.006000   0.030793   0.195  0.84556   
isoFriends_online   -0.015948   0.031062  -0.513  0.60782   
isoOthPpl_online     0.023825   0.027256   0.874  0.38234   
lone01               0.052119   0.069886   0.746  0.45605   
lone02              -0.017933   0.061983  -0.289  0.77242   
lone03              -0.044430   0.065554  -0.678  0.49814   
happy                0.051016   0.035892   1.421  0.15564   
lifeSat              0.072471   0.062211   1.165  0.24443   
MLQ                 -0.002210   0.047756  -0.046  0.96311   
bor01                0.134965   0.041470   3.255  0.00119 **
bor02               -0.048549   0.040495  -1.199  0.23096   
bor03                0.006588   0.033436   0.197  0.84386   
consp01             -0.009626   0.026700  -0.361  0.71856   
consp02             -0.040639   0.027396  -1.483  0.13841   
consp03              0.036862   0.021652   1.703  0.08909 . 
rank_A              -0.012032   0.043432  -0.277  0.78183   
rank_B               0.031185   0.042902   0.727  0.46753   
rank_C               0.028300   0.043262   0.654  0.51322   
rank_D              -0.068832   0.045035  -1.528  0.12685   
rank_E               0.020534   0.040809   0.503  0.61500   
rank_F                     NA         NA      NA       NA   
c19perBeh01          0.117567   0.079982   1.470  0.14202   
c19perBeh02         -0.137439   0.105882  -1.298  0.19469   
c19perBeh03          0.121422   0.063376   1.916  0.05577 . 
c19RCA01             0.122448   0.050178   2.440  0.01492 * 
c19RCA02             0.087163   0.065492   1.331  0.18364   
c19RCA03             0.058055   0.040089   1.448  0.14800   
coronaClose_1        0.403860   0.509944   0.792  0.42864   
coronaClose_2        0.212182   0.414981   0.511  0.60929   
coronaClose_3       -0.077462   0.368724  -0.210  0.83366   
coronaClose_4       -0.220577   0.355181  -0.621  0.53478   
coronaClose_5       -0.346819   0.377727  -0.918  0.35883   
coronaClose_6       -0.422310   0.363664  -1.161  0.24592   
gender               0.115122   0.120469   0.956  0.33959   
age                 -0.108830   0.046852  -2.323  0.02046 * 
edu                 -0.063071   0.041500  -1.520  0.12900   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.517 on 726 degrees of freedom
  (98 observations deleted due to missingness)
Multiple R-squared:   0.16,	Adjusted R-squared:  0.1068 
F-statistic: 3.006 on 46 and 726 DF,  p-value: 5.383e-10
```

Brazil c19ProSo04

R-Squared = 0.1901

P-Value = 3.858e-14

Residual Data:

1Q	Median	3Q
-0.4133	0.3121	0.8257

Attribute	c19RCA01	c19RCA03	isoFriends_online
Meaning	I would sign a petition that Supports mandatory vaccination once a vaccine has been developed for coronavirus.	I would sign a petition that supports reporting people who are suspected to have coronavirus.	I wish time would go by faster
P-Value	0.01492	0.00854	0.03176
Est Coefficient	0.1358647	0.0901534	0.0569937

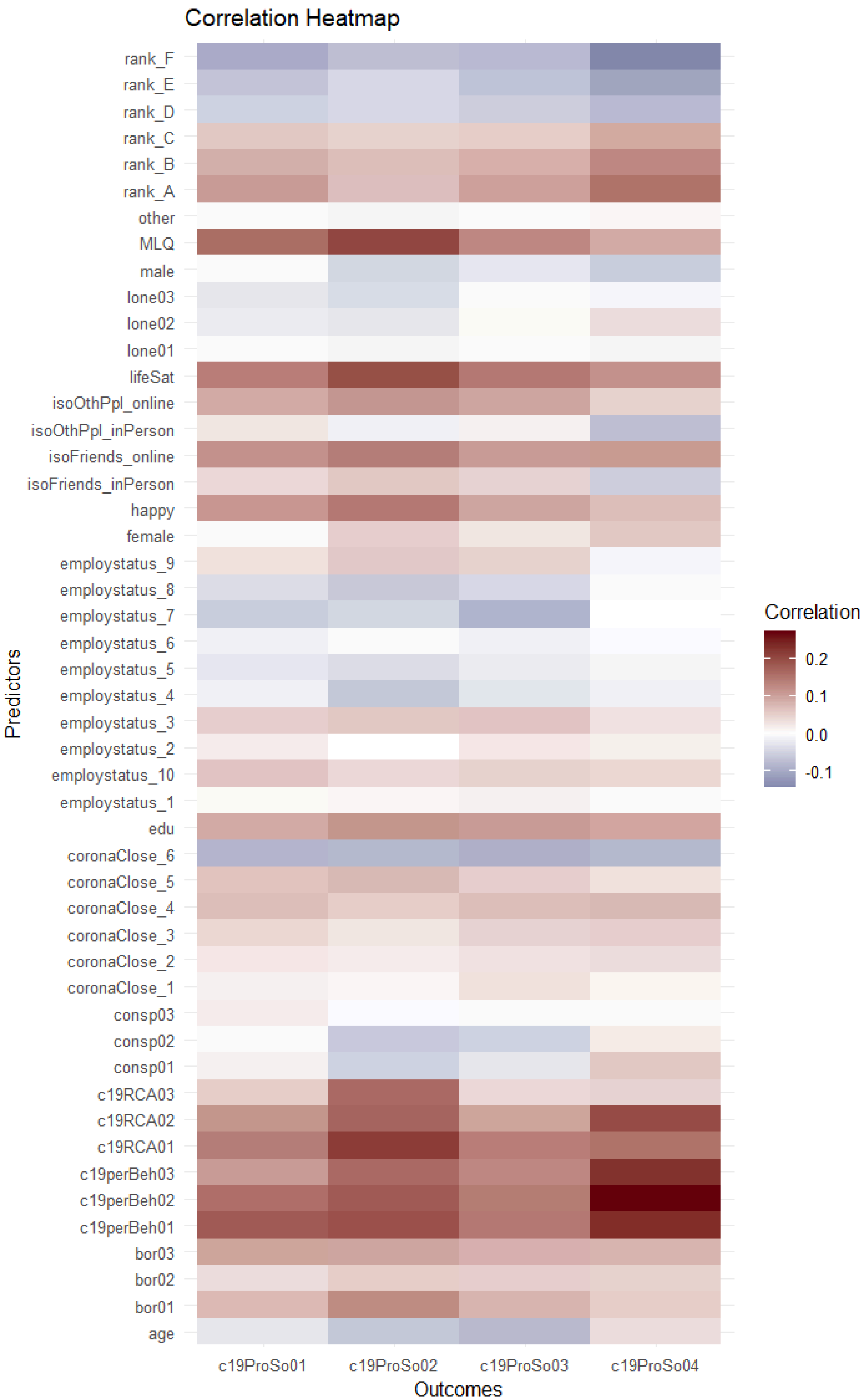
```
Residuals:
    Min      1Q  Median      3Q     Max 
-5.5642 -0.4133  0.3121  0.8257  2.7743 

Coefficients: (1 not defined because of singularities)
                      Estimate Std. Error t value Pr(>|t|)   
(Intercept)          2.2751588  0.7747280   2.937  0.00342 **
employstatus_1      -0.1183917  0.1613148  -0.734  0.46324   
employstatus_2      -0.0250093  0.1711359  -0.146  0.88385   
employstatus_3       0.0038777  0.1663937   0.023  0.98141   
employstatus_4      -0.2515086  0.1960562  -1.283  0.19996   
employstatus_5      -0.4768038  0.2482636  -1.921  0.05518 . 
employstatus_6      -0.2047836  0.2108793  -0.971  0.33183   
employstatus_7      -0.0592808  0.2084662  -0.284  0.77621   
employstatus_8      -1.0398693  0.7842605  -1.326  0.18528   
employstatus_9      -0.1179101  0.1903244  -0.620  0.53577   
employstatus_10      0.3161200  0.3244691   0.974  0.33025   
isoFriends_inPerson -0.0040385  0.0225848  -0.179  0.85813   
isoOthPpl_inPerson  -0.0288592  0.0262599  -1.099  0.27214   
isoFriends_online    0.0569937  0.0264890   2.152  0.03176 * 
isoOthPpl_online    -0.0092966  0.0232429  -0.400  0.68929   
lone01               0.0727040  0.0595976   1.220  0.22289   
lone02               0.0506576  0.0528578   0.958  0.33819   
lone03              -0.0874885  0.0559029  -1.565  0.11802   
happy                0.0218903  0.0306079   0.715  0.47472   
lifeSat             -0.0323361  0.0530518  -0.610  0.54237   
MLQ                  0.0677020  0.0407250   1.662  0.09686 . 
bor01               -0.0130650  0.0353646  -0.369  0.71191   
bor02               -0.0274963  0.0345333  -0.796  0.42616   
bor03               -0.0112197  0.0285133  -0.393  0.69407   
consp01             -0.0035295  0.0227691  -0.155  0.87685   
consp02             -0.0172329  0.0233631  -0.738  0.46099   
consp03              0.0056164  0.0184641   0.304  0.76108   
rank_A               0.0079735  0.0370381   0.215  0.82961   
rank_B               0.0232812  0.0365858   0.636  0.52475   
rank_C               0.0501839  0.0368928   1.360  0.17417   
rank_D              -0.0525643  0.0384048  -1.369  0.17152   
rank_E              -0.0079695  0.0348013  -0.229  0.81893   
rank_F                      NA         NA      NA       NA   
c19perBeh01          0.1188910  0.0682069   1.743  0.08174 . 
c19perBeh02          0.0246880  0.0902940   0.273  0.78461   
c19perBeh03          0.0844274  0.0540460   1.562  0.11869   
c19RCA01             0.1358647  0.0427910   3.175  0.00156 **
c19RCA02             0.0841600  0.0558501   1.507  0.13227   
c19RCA03             0.0901534  0.0341869   2.637  0.00854 **
coronaClose_1        0.5105984  0.4348688   1.174  0.24072   
coronaClose_2        0.2122314  0.3538868   0.600  0.54888   
coronaClose_3        0.1326609  0.3144393   0.422  0.67323   
coronaClose_4       -0.1024818  0.3028903  -0.338  0.73520   
coronaClose_5        0.0012178  0.3221172   0.004  0.99698   
coronaClose_6       -0.1159476  0.3101247  -0.374  0.70861   
gender               0.0004325  0.1027335   0.004  0.99664   
age                 -0.0353432  0.0399541  -0.885  0.37667   
edu                  0.0046347  0.0353900   0.131  0.89584   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.294 on 726 degrees of freedom
  (98 observations deleted due to missingness)
Multiple R-squared:  0.1901,	Adjusted R-squared:  0.1388 
F-statistic: 3.705 on 46 and 726 DF,  p-value: 3.858e-14
```

Other Country Regression Data (2c)

Correlation Heatmap



Other Countries c19ProSo01
R-Squared = 0.1132
P-Value < 2.2e-16
Residual Data:

1Q	Median	3Q
-0.7466	0.1934	0.9671

Attribute	c19perBeh01	MLQ	c19perBeh02	c19RCA01
Meaning	To minimize my chances of getting coronavirus, I wash my hands more often.	My life has a clear sense of purpose.	To minimize my chances of getting coronavirus, I avoid crowded spaces.	I would sign a petition that Supports mandatory vaccination once a vaccine has been developed for coronavirus.
P-Value	< 2e-16	< 2e-16	5.14e-14	< 2e-16
Correlation Coeff.	0.180313299	0.159279332	0.155676679	0.139918773

```
Residuals:
    Min      1Q  Median      3Q     Max 
-5.1093 -0.7466  0.1934  0.9671  4.3675 

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)          0.387515   0.120251   3.223 0.001272 ** 
employstatus_1       0.018206   0.027030   0.674 0.500603    
employstatus_2       0.054737   0.027973   1.957 0.050385 .  
employstatus_3       0.113453   0.026812   4.231 2.33e-05 ***
employstatus_4       0.082883   0.031430   2.637 0.008366 ** 
employstatus_5      -0.035424   0.036781  -0.963 0.335498    
employstatus_6      -0.055208   0.030598  -1.804 0.071193 .  
employstatus_7      -0.242475   0.037339  -6.494 8.48e-11 ***
employstatus_8      -0.152375   0.059498  -2.561 0.010441 *  
employstatus_9       0.120152   0.026411   4.549 5.40e-06 ***
employstatus_10      0.484736   0.049517   9.789  < 2e-16 ***
isoFriends_inPerson  0.019756   0.003479   5.679 1.36e-08 ***
isoOthPpl_inPerson   0.025332   0.003991   6.347 2.22e-10 ***
isoFriends_online    0.018236   0.003475   5.248 1.55e-07 ***
isoOthPpl_online     0.013079   0.003210   4.075 4.62e-05 ***
lone01               0.071885   0.009545   7.531 5.15e-14 ***
lone02              -0.046929   0.008385  -5.597 2.20e-08 ***
lone03               0.033941   0.009086   3.736 0.000187 ***
happy                0.020489   0.005412   3.786 0.000153 ***
lifeSat              0.041795   0.009332   4.479 7.53e-06 ***
MLQ                  0.078555   0.006132  12.811  < 2e-16 ***
bor01                0.037546   0.005357   7.009 2.45e-12 ***
bor02                0.003549   0.005365   0.661 0.508311    
bor03                0.049355   0.004979   9.912  < 2e-16 ***
consp01              0.011888   0.003930   3.025 0.002488 ** 
consp02             -0.020811   0.004129  -5.040 4.67e-07 ***
consp03              0.012056   0.003110   3.877 0.000106 ***
rank_A               0.097377   0.006530  14.912  < 2e-16 ***
rank_B               0.077142   0.006360  12.130  < 2e-16 ***
rank_C               0.052390   0.005835   8.978  < 2e-16 ***
rank_D               0.028256   0.006267   4.508 6.55e-06 ***
rank_E               0.012327   0.006190   1.992 0.046428 *  
rank_F                     NA         NA      NA       NA    
c19perBeh01          0.123015   0.009042  13.604  < 2e-16 ***
c19perBeh02          0.082324   0.010931   7.531 5.14e-14 ***
c19perBeh03          0.008987   0.006329   1.420 0.155572    
c19RCA01             0.072492   0.004753  15.253  < 2e-16 ***
c19RCA02             0.013186   0.007793   1.692 0.090665 .  
c19RCA03            -0.026169   0.005115  -5.116 3.14e-07 ***
coronaClose_1        0.190557   0.072163   2.641 0.008278 ** 
coronaClose_2        0.096500   0.047706   2.023 0.043102 *  
coronaClose_3        0.197590   0.043595   4.532 5.85e-06 ***
coronaClose_4        0.147931   0.035623   4.153 3.29e-05 ***
coronaClose_5        0.124053   0.036077   3.439 0.000585 ***
coronaClose_6       -0.030178   0.037060  -0.814 0.415471    
gender               0.092804   0.015893   5.839 5.28e-09 ***
age                  0.001838   0.006754   0.272 0.785533    
edu                  0.046091   0.005644   8.166 3.29e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.384 on 34359 degrees of freedom
  (4684 observations deleted due to missingness)
Multiple R-squared:  0.1132,	Adjusted R-squared:  0.112 
F-statistic: 95.33 on 46 and 34359 DF,  p-value: < 2.2e-16
```

Other Countries c19ProSo02
R-Squared = 0.1648
P-Value < 2.2e-16
Residual Data:

1Q	Median	3Q
-0.8799	0.2287	1.0734

Attribute	c19RCA01	MLQ	lifeSat	c19perBeh01
Meaning	I would sign a petition that Supports mandatory vaccination once a vaccine has been developed for coronavirus.	My life has a clear sense of purpose.	Participant level of life satisfaction	To minimize my chances of getting coronavirus, I wash my hands more often.
P-Value	< 2e-16	< 2e-16	< 2e-16	< 2e-16
Correlation Coeff	0.2153153605	0.2048230515	0.1940621471	0.1900348719

```
Residuals:
    Min      1Q  Median      3Q     Max 
-5.3886 -0.8799  0.2287  1.0734  5.1448 

Coefficients: (1 not defined because of singularities)
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -0.2044875  0.1307070  -1.564  0.11772    
employstatus_1      -0.0006562  0.0293832  -0.022  0.98218    
employstatus_2      -0.0154220  0.0304095  -0.507  0.61206    
employstatus_3       0.1351896  0.0291472   4.638 3.53e-06 ***
employstatus_4      -0.1868611  0.0341667  -5.469 4.55e-08 ***
employstatus_5      -0.1192877  0.0399798  -2.984  0.00285 ** 
employstatus_6       0.0265635  0.0332582   0.799  0.42447    
employstatus_7      -0.0798194  0.0405811  -1.967  0.04920 *  
employstatus_8      -0.3354086  0.0646721  -5.186 2.16e-07 ***
employstatus_9       0.0832932  0.0287113   2.901  0.00372 ** 
employstatus_10      0.3590305  0.0537925   6.674 2.52e-11 ***
isoFriends_inPerson  0.0399927  0.0037815  10.576  < 2e-16 ***
isoOthPpl_inPerson  -0.0009605  0.0043386  -0.221  0.82480    
isoFriends_online    0.0192032  0.0037778   5.083 3.73e-07 ***
isoOthPpl_online     0.0249117  0.0034892   7.140 9.54e-13 ***
lone01               0.0729763  0.0103760   7.033 2.06e-12 ***
lone02              -0.0515241  0.0091152  -5.653 1.59e-08 ***
lone03               0.0357931  0.0098760   3.624  0.00029 ***
happy                0.0255530  0.0058826   4.344 1.40e-05 ***
lifeSat              0.1025233  0.0101443  10.107  < 2e-16 ***
MLQ                  0.0935499  0.0066652  14.035  < 2e-16 ***
bor01                0.0706815  0.0058243  12.136  < 2e-16 ***
bor02               -0.0064454  0.0058328  -1.105  0.26916    
bor03                0.0291595  0.0054130   5.387 7.21e-08 ***
consp01             -0.0167712  0.0042725  -3.925 8.68e-05 ***
consp02             -0.0343117  0.0044885  -7.644 2.15e-14 ***
consp03              0.0124897  0.0033800   3.695  0.00022 ***
rank_A               0.0627755  0.0070972   8.845  < 2e-16 ***
rank_B               0.0731485  0.0069131  10.581  < 2e-16 ***
rank_C               0.0478222  0.0063423   7.540 4.81e-14 ***
rank_D               0.0171703  0.0068121   2.521  0.01172 *  
rank_E               0.0146447  0.0067276   2.177  0.02950 *  
rank_F                      NA         NA      NA       NA    
c19perBeh01          0.0999531  0.0098298  10.168  < 2e-16 ***
c19perBeh02          0.0755121  0.0118837   6.354 2.12e-10 ***
c19perBeh03          0.0546291  0.0068809   7.939 2.10e-15 ***
c19RCA01             0.1120986  0.0051667  21.696  < 2e-16 ***
c19RCA02            -0.0064143  0.0084715  -0.757  0.44896    
c19RCA03             0.0536485  0.0055605   9.648  < 2e-16 ***
coronaClose_1        0.0604747  0.0784385   0.771  0.44072    
coronaClose_2        0.1218388  0.0518552   2.350  0.01880 *  
coronaClose_3        0.1041132  0.0473860   2.197  0.02802 *  
coronaClose_4        0.0903662  0.0387220   2.334  0.01962 *  
coronaClose_5        0.1678458  0.0392149   4.280 1.87e-05 ***
coronaClose_6       -0.0499935  0.0402832  -1.241  0.21459    
gender              -0.0440540  0.0172762  -2.550  0.01078 *  
age                 -0.0409899  0.0073402  -5.584 2.36e-08 ***
edu                  0.0843874  0.0061349  13.755  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.504 on 34357 degrees of freedom
  (4686 observations deleted due to missingness)
Multiple R-squared:  0.1648,	Adjusted R-squared:  0.1637 
F-statistic: 147.4 on 46 and 34357 DF,  p-value: < 2.2e-16
```

Other Countries c19ProSo03
R-Squared = 0.1127
P-Value < 2.2e-16
Residual Data:

1Q	Median	3Q
-0.9851	0.1877	1.1803

Attribute	c19perBeh01	lifeSat	c19perBeh02	c19RCA01
Meaning	To minimize my chances of getting coronavirus, I wash my hands more often.	Participant level of life satisfaction	To minimize my chances of getting coronavirus, I avoid crowded spaces.	I would sign a petition that Supports mandatory vaccination once a vaccine has been developed for coronavirus.
P-Value	< 2e-16	< 2e-16	3.09e-11	< 2e-16
Correlation Coeff	0.144347155	0.143268372	0.141135792	0.137849017

```
Residuals:
    Min      1Q  Median      3Q     Max 
-5.2460 -0.9851  0.1877  1.1803  5.0641 

Coefficients: (1 not defined because of singularities)
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)          0.2622534  0.1367490   1.918 0.055148 .  
employstatus_1       0.0719239  0.0307396   2.340 0.019301 *  
employstatus_2       0.0952044  0.0318163   2.992 0.002771 ** 
employstatus_3       0.1899040  0.0304966   6.227 4.81e-10 ***
employstatus_4       0.0074101  0.0357548   0.207 0.835817    
employstatus_5      -0.0126318  0.0418183  -0.302 0.762605    
employstatus_6      -0.0322816  0.0347991  -0.928 0.353592    
employstatus_7      -0.2451900  0.0424601  -5.775 7.78e-09 ***
employstatus_8      -0.1631371  0.0677150  -2.409 0.015994 *  
employstatus_9       0.1065238  0.0300432   3.546 0.000392 ***
employstatus_10      0.4195813  0.0563150   7.451 9.51e-14 ***
isoFriends_inPerson  0.0288377  0.0039562   7.289 3.18e-13 ***
isoOthPpl_inPerson   0.0194965  0.0045392   4.295 1.75e-05 ***
isoFriends_online    0.0111174  0.0039519   2.813 0.004909 ** 
isoOthPpl_online     0.0235657  0.0036505   6.455 1.09e-10 ***
lone01               0.0257323  0.0108554   2.370 0.017772 *  
lone02              -0.0166177  0.0095362  -1.743 0.081415 .  
lone03               0.0783423  0.0103308   7.583 3.45e-14 ***
happy                0.0085016  0.0061547   1.381 0.167193    
lifeSat              0.1096175  0.0106134  10.328  < 2e-16 ***
MLQ                  0.0419493  0.0069719   6.017 1.80e-09 ***
bor01                0.0345127  0.0060917   5.666 1.48e-08 ***
bor02                0.0182705  0.0061009   2.995 0.002749 ** 
bor03                0.0509905  0.0056622   9.005  < 2e-16 ***
consp01              0.0008307  0.0044697   0.186 0.852558    
consp02             -0.0457165  0.0046959  -9.735  < 2e-16 ***
consp03              0.0123473  0.0035366   3.491 0.000481 ***
rank_A               0.1026117  0.0074262  13.818  < 2e-16 ***
rank_B               0.0815709  0.0072333  11.277  < 2e-16 ***
rank_C               0.0492398  0.0066357   7.420 1.19e-13 ***
rank_D               0.0182195  0.0071273   2.556 0.010583 *  
rank_E               0.0011651  0.0070394   0.166 0.868543    
rank_F                      NA         NA      NA       NA    
c19perBeh01          0.0889968  0.0102819   8.656  < 2e-16 ***
c19perBeh02          0.0825991  0.0124312   6.644 3.09e-11 ***
c19perBeh03          0.0566608  0.0071978   7.872 3.59e-15 ***
c19RCA01             0.0847317  0.0054055  15.675  < 2e-16 ***
c19RCA02             0.0072678  0.0088639   0.820 0.412264    
c19RCA03            -0.0358425  0.0058180  -6.161 7.32e-10 ***
coronaClose_1        0.3480063  0.0819844   4.245 2.19e-05 ***
coronaClose_2        0.0578040  0.0542499   1.066 0.286651    
coronaClose_3        0.1572460  0.0495861   3.171 0.001520 ** 
coronaClose_4        0.0948932  0.0405265   2.342 0.019211 *  
coronaClose_5       -0.0017412  0.0410323  -0.042 0.966153    
coronaClose_6       -0.1251235  0.0421534  -2.968 0.002997 ** 
gender               0.0426353  0.0180731   2.359 0.018327 *  
age                 -0.0572341  0.0076794  -7.453 9.34e-14 ***
edu                  0.0670659  0.0064180  10.450  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.574 on 34358 degrees of freedom
  (4685 observations deleted due to missingness)
Multiple R-squared:  0.1127,	Adjusted R-squared:  0.1116 
F-statistic: 94.91 on 46 and 34358 DF,  p-value: < 2.2e-16
```

Other Countries c19ProSo04
R-Squared = 0.1542
P-Value < 2.2e-16
Residual Data:

1Q	Median	3Q
-0.7002	0.3045	1.0251

Attribute	c19perBeh02	c19perBeh01	c19perBeh03	c19RCA02
Meaning	To minimize my chances of getting coronavirus, I avoid crowded spaces.	To minimize my chances of getting coronavirus, I wash my hands more often.	To minimize my chances of getting coronavirus, I put myself in quarantine.	"I would sign a petition that Supports mandatory quarantine for those that have coronavirus and those that have been exposed to the virus.
P-Value	< 2e-16	< 2e-16	< 2e-16	< 2e-16
Correlation Coeff	0.2723188175	0.2304197578	0.2274017933	0.1976518559

```
Residuals:
    Min      1Q  Median      3Q     Max 
-5.5919 -0.7002  0.3045  1.0251  4.5343 

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -0.260354   0.123984  -2.100 0.035746 *  
employstatus_1       0.072248   0.027877   2.592 0.009554 ** 
employstatus_2       0.131332   0.028852   4.552 5.33e-06 ***
employstatus_3       0.182691   0.027654   6.606 4.00e-11 ***
employstatus_4       0.085308   0.032411   2.632 0.008491 ** 
employstatus_5       0.063117   0.037923   1.664 0.096050 .  
employstatus_6      -0.004522   0.031548  -0.143 0.886027    
employstatus_7      -0.024691   0.038497  -0.641 0.521282    
employstatus_8       0.180367   0.061398   2.938 0.003309 ** 
employstatus_9       0.057576   0.027234   2.114 0.034512 *  
employstatus_10      0.321918   0.051026   6.309 2.84e-10 ***
isoFriends_inPerson  0.001262   0.003586   0.352 0.724936    
isoOthPpl_inPerson   0.003124   0.004115   0.759 0.447711    
isoFriends_online    0.014678   0.003583   4.097 4.20e-05 ***
isoOthPpl_online     0.001854   0.003310   0.560 0.575349    
lone01              -0.014092   0.009841  -1.432 0.152177    
lone02               0.030378   0.008647   3.513 0.000443 ***
lone03               0.057512   0.009367   6.140 8.35e-10 ***
happy               -0.003120   0.005580  -0.559 0.576042    
lifeSat              0.085665   0.009622   8.903  < 2e-16 ***
MLQ                  0.012623   0.006321   1.997 0.045826 *  
bor01                0.004310   0.005523   0.780 0.435192    
bor02                0.028930   0.005532   5.230 1.70e-07 ***
bor03                0.042559   0.005133   8.290  < 2e-16 ***
consp01              0.031578   0.004052   7.794 6.67e-15 ***
consp02             -0.028629   0.004257  -6.725 1.78e-11 ***
consp03             -0.003217   0.003206  -1.003 0.315628    
rank_A               0.120205   0.006733  17.854  < 2e-16 ***
rank_B               0.102794   0.006558  15.675  < 2e-16 ***
rank_C               0.073416   0.006016  12.204  < 2e-16 ***
rank_D               0.037915   0.006462   5.868 4.46e-09 ***
rank_E               0.012020   0.006382   1.883 0.059665 .  
rank_F                     NA         NA      NA       NA    
c19perBeh01          0.092067   0.009322   9.876  < 2e-16 ***
c19perBeh02          0.173799   0.011270  15.421  < 2e-16 ***
c19perBeh03          0.095715   0.006525  14.668  < 2e-16 ***
c19RCA01             0.055457   0.004900  11.317  < 2e-16 ***
c19RCA02             0.108286   0.008036  13.475  < 2e-16 ***
c19RCA03            -0.059966   0.005274 -11.369  < 2e-16 ***
coronaClose_1        0.275338   0.074327   3.704 0.000212 ***
coronaClose_2        0.140499   0.049179   2.857 0.004281 ** 
coronaClose_3        0.233126   0.044937   5.188 2.14e-07 ***
coronaClose_4        0.127819   0.036724   3.481 0.000501 ***
coronaClose_5       -0.019730   0.037192  -0.531 0.595768    
coronaClose_6       -0.105598   0.038203  -2.764 0.005710 ** 
gender              -0.003037   0.016385  -0.185 0.852941    
age                  0.027287   0.006962   3.919 8.89e-05 ***
edu                  0.041789   0.005818   7.182 7.00e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.427 on 34359 degrees of freedom
  (4684 observations deleted due to missingness)
Multiple R-squared:  0.1542,    Adjusted R-squared:  0.1531 
F-statistic: 136.2 on 46 and 34359 DF,  p-value: < 2.2e-16
```

otherCountry correlation table

	c19ProSo01	c19ProSo 02	c19ProSo03	c19ProSo04
employstatus_1	0.005690684	0.007876043	0.01504904	0.002280419
employstatus_2	0.017322067	-0.00026973 4	0.022111273	0.016030365
employstatus_3	0.050193795	0.056889835	0.059104804	0.029292891
employstatus_4	-0.013119962	-0.06508296 6	-0.029949354	-0.01629027
employstatus_5	-0.026957254	-0.03820997 1	-0.023262907	-0.01076053
employstatus_6	-0.015644074	0.002157015	-0.016109347	-0.000594386
employstatus_7	-0.058498234	-0.04713675 1	-0.088743128	0.000408755
employstatus_8	-0.039129835	-0.06408276 3	-0.04206523	-0.001825513
employstatus_9	0.030750078	0.054431321	0.045695057	-0.007448108
employstatus_1 0	0.060195894	0.039673773	0.04686414	0.040741807
isoFriends_inP erson	0.038657381	0.05580354	0.044051257	-0.057985553
isoOthPpl_inPe rson	0.025911129	-0.01544790 5	0.01341513	-0.073410306
isoFriends_onli ne	0.116414262	0.139636591	0.107626858	0.108620455
isoOthPpl_onlin e	0.088804622	0.112604007	0.096319095	0.047321767
lone01	-0.003570964	-0.00984855 3	0.004660972	-0.011564971
lone02	-0.022345259	-0.02856968 8	0.005663251	0.033240857
lone03	-0.028343179	-0.04133845	0.002034925	-0.007884432
happy	0.110052729	0.14323159	0.094796566	0.06674562
lifeSat	0.139136372	0.194062147	0.143268372	0.118217472

MLQ	0.159279332	0.204823052	0.125799591	0.090918061
bor01	0.073531489	0.124882548	0.079812534	0.050729181
bor02	0.034224735	0.053613819	0.050176638	0.047939675
bor03	0.097304395	0.094082769	0.081074099	0.078697082
consp01	0.014469735	-0.053717878	-0.028814464	0.055463471
consp02	-0.004152228	-0.060655347	-0.053276161	0.020745296
consp03	0.017638715	-0.000755384	-0.004483439	-0.001936446
c19perBeh01	0.180313299	0.190034872	0.144347155	0.230419758
c19perBeh02	0.155676679	0.179383269	0.141135792	0.272318818
c19perBeh03	0.103307652	0.162838533	0.127008955	0.227401793
c19RCA01	0.139918773	0.215315361	0.137849017	0.153732618
c19RCA02	0.114335122	0.167444743	0.097442399	0.197651856
c19RCA03	0.050509761	0.160004685	0.037672341	0.044588693
coronaClose_1	0.014752024	0.006558037	0.031443933	0.010476016
coronaClose_2	0.021955326	0.018866324	0.027362029	0.033195318
coronaClose_3	0.041821901	0.025328569	0.04443174	0.049005077
coronaClose_4	0.067199603	0.052328971	0.069771644	0.074871853
coronaClose_5	0.06090847	0.074140151	0.048743045	0.030790429
coronaClose_6	-0.087131918	-0.082702171	-0.091978763	-0.083115798
female	0.002757872	0.049007517	0.025245693	0.05826052
male	-0.003624739	-0.047760424	-0.026079779	-0.059616209
other	0.004566622	-0.010605923	0.005109789	0.008863094
age	-0.027991852	-0.064709905	-0.081602734	0.034417664
edu	0.089990182	0.113952473	0.105555285	0.092243774

rank_E	-0.068648646	-0.044402233	-0.070981355	-0.109582083
rank_C	0.055906945	0.04716458	0.051959649	0.091741654
rank_D	-0.053791387	-0.043371073	-0.056936039	-0.080836199
rank_B	0.084410729	0.069274541	0.081565217	0.127264336
rank_A	0.105536254	0.065467606	0.098196463	0.153914696
rank_F	-0.097638482	-0.075457151	-0.080415957	-0.144836061

Clustering (3a)

Clustering Table of Values

coded_country	ghs_overall_score	GDP_per_capita_2021	unemployment.rate.2021	happiness_score_2019	birth_rate_per_1000_2021	press_freedom_score_2021	CPI_score_2021
Albania	45	6377.2031	12.59	4.719	10.24	30.59	35
Algeria	26.2	3700.3147	13.729	5.211	21.524	47.26	33
Argentina	54.4	10650.8605	8.74	6.086	13.902	28.99	38
Armenia	61.8	4972.7832	10.01	4.559	12.049	28.83	49
Australia	71.1	60697.2454	5.12	7.228	12.1	19.79	73
Austria	56.9	53517.8904	6.46	7.246	9.6	16.34	74
Azerbaijan	34.7	5408.0454	6.04	5.208	11.1	58.77	30
Bahrain	36.3	26850.0034	1.548	6.199	11.926	61.1	42
Bangladesh	35.5	2457.924	5.246	4.456	17.821	49.71	26
Belarus	43.9	7489.7189	3.9	5.323	9.283	50.82	41
Belgium	59.3	51850.3972	6.26	6.923	10.2	11.69	73
Benin	25.4	1360.9115	1.784	4.883	36.608	38.18	42
Bosnia and Herzegovina	35.4	7230.1988	14.9	5.386	8.42	28.34	35

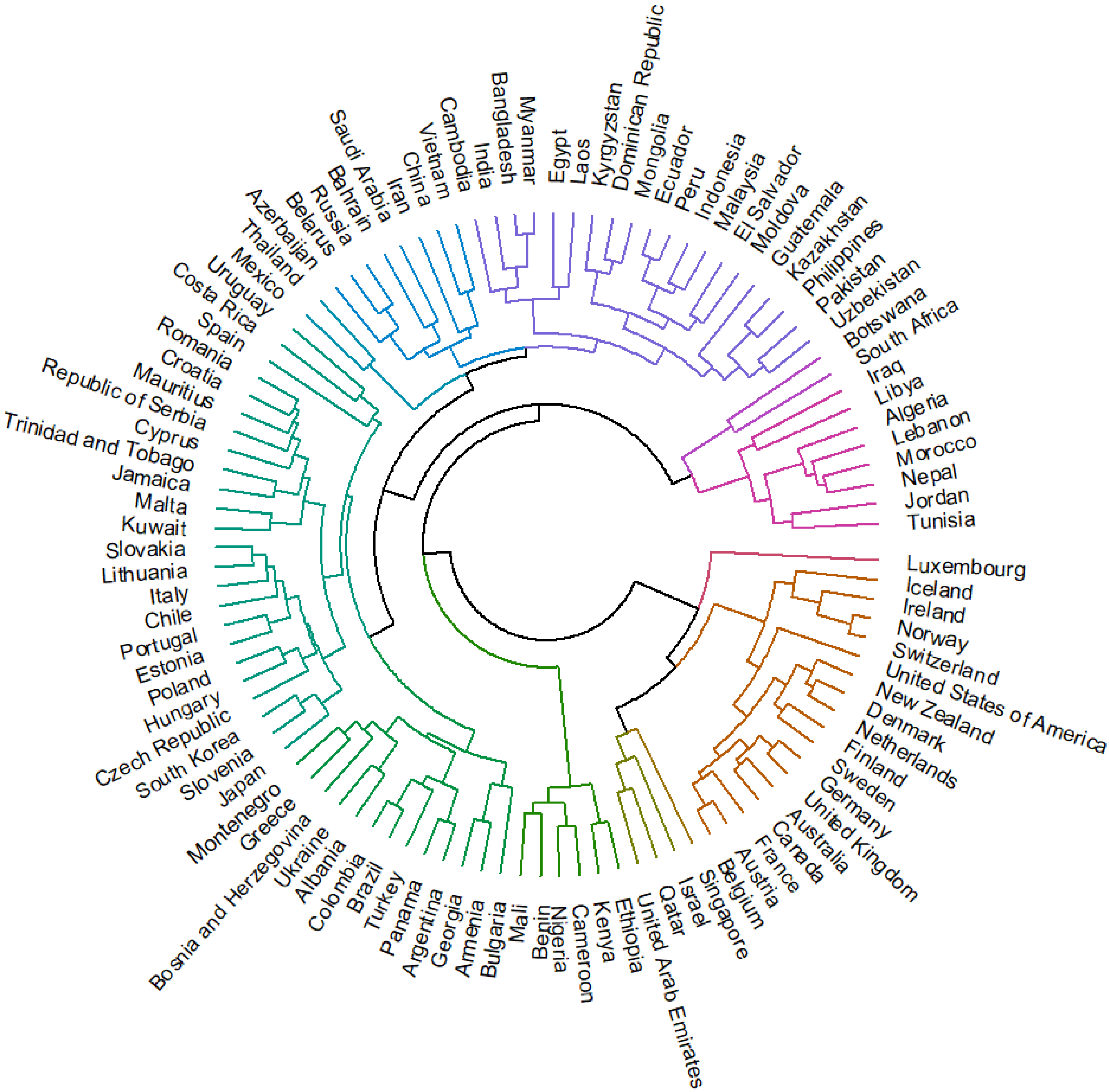
Botswana	33.6	7238.7961	23.11	3.488	23.576	23.25	55
Brazil	51.2	7696.7848	13.16	6.3	12.883	36.25	38
Bulgaria	59.9	12219.3419	5.27	5.011	8.5	37.29	42
Cambodia	31.1	1625.235	0.4	4.7	19.334	46.84	23
Cameroon	28.6	1654.257	4.145	5.044	34.938	43.78	27
Canada	69.8	52515.1998	7.53	7.278	9.6	15.25	74
Chile	56.2	16240.6078	9.28	6.444	11.788	27.89	67
China	47.5	12617.5051	4.55	5.191	7.52	78.72	45
Colombia	53.2	6182.7071	13.9	6.125	14.204	43.74	39
Costa Rica	40.8	12604.0488	15.14	7.167	11.873	8.76	58
Croatia	48.8	17809.0324	7.61	5.432	9.4	27.95	47
Cyprus	41.9	32745.8438	7.51	6.046	10.313	19.85	53
Czech Republic	52.8	26822.5142	2.8	6.852	10.6	23.38	54
Denmark	64.4	69268.6518	5.04	7.6	10.8	8.57	88
Dominican Republic	34.5	8476.7496	7.7	5.425	18.432	25.6	30
Ecuador	50.8	5965.1329	4.55	6.028	16.817	32.83	36
Egypt	28	3886.7225	7.44	4.166	22.558	56.17	33
El Salvador	40.8	4664.3112	4.33	6.253	16.025	30.49	34
Estonia	55.5	27943.7012	6.18	5.893	10	15.25	74
Ethiopia	37.8	925.0007	3.93	4.286	32.383	33.63	39
Finland	70.9	53504.6937	7.61	7.769	9	6.99	88
France	61.9	43671.3084	7.86	6.592	10.9	22.6	71
Georgia	52.6	5023.2744	11.851	4.519	13.412	28.64	55
Germany	65.5	51426.7504	3.64	6.985	9.6	15.24	80
Greece	51.5	20310.6825	14.66	5.287	8	29.01	49
Guatemala	29.1	5029.4776	2.17	6.436	21.12	38.45	25
Hungary	54.4	18753.0469	4.05	5.758	9.7	31.76	43

Iceland	48.5	68710.2442	6.03	7.494	13.1	15.37	74
India	42.8	2238.1271	6.38	4.015	16.419	46.56	40
Indonesia	50.4	4334.216	3.83	5.192	16.425	37.4	38
Iran	36.5	4084.2003	9.28	4.548	13.699	72.7	25
Iraq	24	4770.8353	16.17	4.437	27.367	55.57	23
Ireland	55.3	102001.7982	6.19	7.021	11.6	11.91	74
Israel	47.2	52129.516	4.81	7.139	19.7	30.9	59
Italy	51.9	36449.2583	9.5	6.223	6.8	23.39	56
Jamaica	31.8	5183.581	6.156	5.89	11.712	9.96	44
Japan	60.5	40058.5373	2.83	5.886	6.6	28.88	73
Jordan	42.8	4152.758	19.84	4.906	21.95	42.89	49
Kazakhstan	46.1	10373.7898	5.572	5.809	23.5	50.28	37
Kenya	38.8	2069.6611	5.69	4.509	27.685	33.65	30
Kuwait	36.8	32324.8409	2.943	6.021	10.41	34.36	43
Kyrgyzstan	42.4	1365.5083	4.1	5.261	22.4	30.37	27
Laos	34.8	2535.6234	4.15	4.796	21.973	70.56	30
Lebanon	33.4	4136.1466	12.777	5.197	14.948	34.93	24
Libya	25.3	5908.9513	19.71	5.525	17.828	55.73	17
Lithuania	59.5	23849.6157	7.11	6.149	8.3	20.15	61
Luxembourg	48.4	133711.7944	5.25	7.09	10.5	17.56	81
Malaysia	56.4	11134.623	4.083	5.339	15.24	39.47	48
Mali	29	881.5101	2.44	4.39	41.643	33.5	29
Malta	40.2	34881.2913	3.39	6.726	8.5	30.46	54
Mauritius	39.7	9068.9802	7.72	5.888	10.3	28.74	54
Mexico	57	10359.1499	4.09	6.595	14.857	46.71	31
Moldova	41	5274.7448	0.79	5.529	12.432	31.61	36
Mongolia	41	4566.1401	7.75	5.285	21.413	28.97	35
Montenegro	44.1	9465.9615	16.54	5.523	11.4	34.33	46

Morocco	33.6	3767.5249	11.219	5.208	17.545	43.94	39
Myanmar	38.3	1231.6947	4.34	4.36	17.103	46.14	28
Nepal	34	1229.3942	12.581	4.913	20.402	34.62	33
Netherlands	64.7	58727.8705	4.21	7.488	10.2	9.67	82
New Zealand	62.5	49996.4207	3.78	7.307	11.48	10.04	88
Nigeria	38	2065.7744	5.264	5.265	37.117	39.69	24
Norway	60.2	93072.8925	4.37	7.554	10.4	6.72	85
Pakistan	30.4	1506.1083	6.34	5.653	27.519	46.86	28
Panama	53.5	15491.2898	10.451	6.321	17.692	29.94	36
Peru	54.9	6635.4641	5.1	5.697	17.622	31.71	36
Philippines	45.7	3460.5394	3.4	5.631	21.813	45.64	33
Poland	55.7	18050.2794	3.36	6.182	8.8	28.84	56
Portugal	54.7	24661.1665	6.58	5.693	7.7	10.11	62
Qatar	48.7	66858.7417	0.14	6.374	9.816	42.6	63
Republic of Serbia	45	9232.9616	10.06	5.603	9.1	32.03	38
Romania	45.7	14946.625	5.59	6.07	9.3	24.91	45
Russia	49.1	12532.0508	4.72	5.648	9.6	48.71	29
Saudi Arabia	44.9	24315.6185	6.62	6.375	17.473	62.73	53
Singapore	57.4	77710.0892	4.64	6.262	8.6	55.2	85
Slovakia	54.4	21768.1487	6.89	6.198	10.4	23.02	52
Slovenia	67.8	29331.0647	4.74	6.118	9	23.1	57
South Africa	45.8	7073.6128	28.77	4.722	19.821	21.59	44
South Korea	65.4	35142.2643	3.64	5.895	5.1	23.43	62
Spain	60.9	30488.821	14.78	6.354	7.1	20.44	61
Sweden	64.9	61417.6809	8.72	7.343	11	7.24	85
Switzerland	58.8	93446.4345	5.1	7.48	10.3	10.55	84

Thailand	68.2	7060.8977	1.21	6.008	8.999	45.22	35
Trinidad and Tobago	36.8	16056.302	4.45	6.192	11.687	21.55	41
Tunisia	31.5	3807.1841	18.524	4.461	16.093	29.53	44
Turkey	50	9743.2131	11.98	5.373	14.678	49.79	38
Ukraine	38.9	4827.8457	9.83	4.332	7.3	32.96	32
United Arab Emirates	39.6	44332.34	3.11	6.825	10.307	43.13	69
United Kingdom	67.2	46869.7591	4.826	7.054	10.1	21.59	78
United States of America	75.9	70219.4725	5.35	6.892	11	23.93	67
Uruguay	40.3	17923.9953	9.29	6.293	10.473	16.38	73
Uzbekistan	39	1993.4245	5.419	6.174	25.9	50.74	28
Vietnam	42.9	3756.4889	2.38	5.175	15.008	78.46	39

Coloured Circular Dendrogram with 12 Clusters



Cluster Countries Regression Models Data

Cluster Countries c19ProSo01
R-Squared = 0.1089
P-Value < 2.2e-16
Residual Data:

1Q	Median	3Q
-0.7907	0.2045	0.9902

Attribute	MLQ	rank_A	c19RCA01	consp01
Meaning	My life has a clear sense of purpose.	Ranking 1-6 of Beauty	I would sign a petition that Supports mandatory vaccination once a vaccine has been developed for coronavirus.	I think that many very important things happen in the world, which the public is never informed about.
P-Value	1.27e-08	8.92e-07	6.71e-06	1.53e-05
Est. Coefficient	0.0926819	0.0916628	0.0565728	0.0493682

```
Residuals:
    Min      1Q  Median      3Q     Max 
-4.8447 -0.7907  0.2045  0.9902  3.8340 

Coefficients: (1 not defined because of singularities)
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)          0.7723929  0.3501975   2.206 0.027464 *  
employstatus_1       0.0157241  0.0759820   0.207 0.836062    
employstatus_2      -0.0526234  0.0776278  -0.678 0.497875    
employstatus_3       0.0822537  0.0760700   1.081 0.279628    
employstatus_4       0.0391649  0.0814781   0.481 0.630768    
employstatus_5      -0.0314161  0.0957488  -0.328 0.742844    
employstatus_6      -0.0740902  0.0857313  -0.864 0.387518    
employstatus_7      -0.1605385  0.0989431  -1.623 0.104761    
employstatus_8      -0.2233712  0.2251089  -0.992 0.321116    
employstatus_9       0.1106871  0.0764727   1.447 0.147855    
employstatus_10      0.5298172  0.1459711   3.630 0.000287 ***
isoFriends_inPerson  0.0074621  0.0104703   0.713 0.476078    
isoOthPpl_inPerson   0.0106515  0.0116037   0.918 0.358703    
isoFriends_online    0.0212334  0.0104352   2.035 0.041932 *  
isoOthPpl_online     0.0097742  0.0086414   1.131 0.258076    
lone01               0.1007008  0.0253238   3.977 7.11e-05 ***
lone02              -0.0336653  0.0219418  -1.534 0.125028    
lone03              -0.0531078  0.0249833  -2.126 0.033582 *  
happy                0.0152495  0.0145959   1.045 0.296183    
lifeSat              0.0085534  0.0238339   0.359 0.719707    
MLQ                  0.0926819  0.0162573   5.701 1.27e-08 ***
bor01                0.0221403  0.0161920   1.367 0.171585    
bor02                0.0414155  0.0167545   2.472 0.013478 *  
bor03                0.0292034  0.0136983   2.132 0.033070 *  
consp01              0.0493682  0.0114022   4.330 1.53e-05 ***
consp02             -0.0422189  0.0120617  -3.500 0.000469 ***
consp03              0.0152717  0.0083626   1.826 0.067890 .  
rank_A               0.0916628  0.0186260   4.921 8.92e-07 ***
rank_B               0.0441067  0.0178023   2.478 0.013265 *  
rank_C               0.0447921  0.0159993   2.800 0.005139 ** 
rank_D              -0.0011365  0.0171143  -0.066 0.947058    
rank_E               0.0009513  0.0172477   0.055 0.956019    
rank_F                      NA         NA      NA       NA    
c19perBeh01          0.1025110  0.0271468   3.776 0.000161 ***
c19perBeh02          0.0977793  0.0330194   2.961 0.003080 ** 
c19perBeh03         -0.0074680  0.0229199  -0.326 0.744569    
c19RCA01             0.0565728  0.0125489   4.508 6.71e-06 ***
c19RCA02            -0.0090499  0.0232178  -0.390 0.696715    
c19RCA03             0.0505328  0.0152336   3.317 0.000917 ***
coronaClose_1        0.1840976  0.2385769   0.772 0.440364    
coronaClose_2        0.4580518  0.2028342   2.258 0.023979 *  
coronaClose_3        0.2748079  0.1691882   1.624 0.104390    
coronaClose_4        0.3143930  0.1285745   2.445 0.014516 *  
coronaClose_5        0.2649074  0.1288655   2.056 0.039872 *  
coronaClose_6        0.0799271  0.1329117   0.601 0.547635    
gender               0.0689548  0.0454608   1.517 0.129390    
age                 -0.0209137  0.0186694  -1.120 0.262685    
edu                  0.0186264  0.0176966   1.053 0.292610    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.382 on 4367 degrees of freedom
  (413 observations deleted due to missingness)
Multiple R-squared:  0.1089,	Adjusted R-squared:  0.09953 
F-statistic:  11.6 on 46 and 4367 DF,  p-value: < 2.2e-16
```

Cluster Countries c19ProSo02
R-Squared = 0.135
P-Value < 2.2e-16
Residual Data:

1Q	Median	3Q
-0.9669	0.1969	0.9902

Attribute	c19RCA01	consp02	c19RCA03	MLQ
Meaning	I would sign a petition that Supports mandatory vaccination once a vaccine has been developed for coronavirus.	I think that politicians usually do not tell us the true motives for their decisions.	I would sign a petition that supports reporting people who are suspected to have coronavirus.	My life has a clear sense of purpose.
P-Value	1.65e-15	1.19e-07	4.19e-07	7.09e-07
Est. Coefficient	0.1114517	-0.0711024	0.0858048	0.0896536

```
Residuals:
    Min      1Q  Median      3Q     Max 
-4.5184 -0.9669  0.1969  1.1420  4.6669 

Coefficients: (1 not defined because of singularities)
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)          0.9902513  0.3889172   2.546  0.01093 *  
employstatus_1       0.0688647  0.0843775   0.816  0.41446    
employstatus_2      -0.0976046  0.0862203  -1.132  0.25768    
employstatus_3       0.1448959  0.0844831   1.715  0.08640 .  
employstatus_4      -0.0468258  0.0904674  -0.518  0.60476    
employstatus_5       0.0164012  0.1063224   0.154  0.87741    
employstatus_6       0.0369527  0.0951752   0.388  0.69784    
employstatus_7      -0.0581479  0.1098897  -0.529  0.59673    
employstatus_8      -0.2552901  0.2500028  -1.021  0.30724    
employstatus_9      -0.0813326  0.0849055  -0.958  0.33816    
employstatus_10      0.4203892  0.1612603   2.607  0.00917 ** 
isoFriends_inPerson  0.0094257  0.0116291   0.811  0.41768    
isoOthPpl_inPerson  -0.0103097  0.0128890  -0.800  0.42382    
isoFriends_online    0.0238760  0.0115899   2.060  0.03945 *  
isoOthPpl_online     0.0114264  0.0095989   1.190  0.23396    
lone01               0.0648160  0.0281384   2.303  0.02130 *  
lone02               0.0124755  0.0243711   0.512  0.60875    
lone03              -0.0597593  0.0277492  -2.154  0.03133 *  
happy                0.0329642  0.0162148   2.033  0.04212 *  
lifeSat              0.0668246  0.0264741   2.524  0.01163 *  
MLQ                  0.0896536  0.0180533   4.966 7.09e-07 ***
bor01                0.0150094  0.0180074   0.834  0.40460    
bor02                0.0377822  0.0186318   2.028  0.04264 *  
bor03                0.0387597  0.0152124   2.548  0.01087 *  
consp01              0.0207205  0.0126804   1.634  0.10232    
consp02             -0.0711024  0.0134056  -5.304 1.19e-07 ***
consp03              0.0291618  0.0092880   3.140  0.00170 ** 
rank_A               0.0069838  0.0206851   0.338  0.73566    
rank_B               0.0337896  0.0197718   1.709  0.08753 .  
rank_C              -0.0168595  0.0177711  -0.949  0.34282    
rank_D              -0.0003811  0.0190127  -0.020  0.98401    
rank_E              -0.0109159  0.0191566  -0.570  0.56883    
rank_F                      NA         NA      NA       NA    
c19perBeh01          0.1261559  0.0301505   4.184 2.92e-05 ***
c19perBeh02          0.0655612  0.0366934   1.787  0.07405 .  
c19perBeh03          0.0694169  0.0255118   2.721  0.00653 ** 
c19RCA01             0.1114517  0.0139407   7.995 1.65e-15 ***
c19RCA02            -0.0564964  0.0257905  -2.191  0.02853 *  
c19RCA03             0.0858048  0.0169320   5.068 4.19e-07 ***
coronaClose_1        0.2796232  0.2649702   1.055  0.29135    
coronaClose_2        0.2230252  0.2252798   0.990  0.32223    
coronaClose_3        0.1463917  0.1879100   0.779  0.43599    
coronaClose_4        0.0303251  0.1428111   0.212  0.83185    
coronaClose_5        0.0934284  0.1431264   0.653  0.51394    
coronaClose_6       -0.1843869  0.1476287  -1.249  0.21174    
gender              -0.0104764  0.0505034  -0.207  0.83568    
age                 -0.0885717  0.0207387  -4.271 1.99e-05 ***
edu                  0.0123471  0.0196543   0.628  0.52990    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.535 on 4366 degrees of freedom
  (414 observations deleted due to missingness)
Multiple R-squared:  0.135,	Adjusted R-squared:  0.1259 
F-statistic: 14.81 on 46 and 4366 DF,  p-value: < 2.2e-16
```

Cluster Countries c19ProSo03
R-Squared = 0.1075
P-Value < 2.2e-16
Residual Data:

1Q	Median	3Q
-1.1176	0.1274	1.2131

Attribute	bor03	rank_A	c19RCA01	rank_B
Meaning	I feel in control of my time.	1-6 ranking of beauty	I would sign a petition that supports mandatory vaccination once a vaccine has been developed for coronavirus.	1-6 ranking of achievement
P-Value	1.72e-07	2.10e-06	1.30e-05	6.48e-05
Est. Coefficient	0.082811	0.102227	0.063305	0.082274

```
Residuals:
    Min      1Q  Median      3Q     Max 
-4.3267 -1.1176  0.1274  1.2131  4.2020 

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -0.011991   0.404627  -0.030 0.976359    
employstatus_1       0.209534   0.087758   2.388 0.017000 *  
employstatus_2       0.137522   0.089724   1.533 0.125417    
employstatus_3       0.201134   0.087921   2.288 0.022205 *  
employstatus_4       0.100094   0.094165   1.063 0.287860    
employstatus_5       0.126670   0.110492   1.146 0.251686    
employstatus_6      -0.054326   0.099091  -0.548 0.583557    
employstatus_7      -0.369173   0.114352  -3.228 0.001254 ** 
employstatus_8      -0.056088   0.260195  -0.216 0.829340    
employstatus_9       0.294886   0.088391   3.336 0.000857 ***
employstatus_10      0.557775   0.168723   3.306 0.000954 ***
isoFriends_inPerson  0.011340   0.012096   0.938 0.348537    
isoOthPpl_inPerson   0.032533   0.013410   2.426 0.015306 *  
isoFriends_online    0.015793   0.012060   1.310 0.190428    
isoOthPpl_online     0.020473   0.009988   2.050 0.040454 *  
lone01               0.002858   0.029271   0.098 0.922212    
lone02              -0.013511   0.025348  -0.533 0.594065    
lone03               0.067853   0.028843   2.353 0.018689 *  
happy                0.049817   0.016861   2.954 0.003149 ** 
lifeSat              0.096792   0.027530   3.516 0.000443 ***
MLQ                 -0.015466   0.018767  -0.824 0.409897    
bor01                0.046552   0.018713   2.488 0.012897 *  
bor02                0.019882   0.019360   1.027 0.304496    
bor03                0.082811   0.015818   5.235 1.72e-07 ***
consp01              0.017343   0.013173   1.317 0.188072    
consp02             -0.062561   0.013941  -4.488 7.39e-06 ***
consp03              0.033631   0.009664   3.480 0.000506 ***
rank_A               0.102227   0.021522   4.750 2.10e-06 ***
rank_B               0.082274   0.020576   3.998 6.48e-05 ***
rank_C               0.031439   0.018492   1.700 0.089170 .  
rank_D               0.039175   0.019781   1.980 0.047712 *  
rank_E               0.008570   0.019936   0.430 0.667312    
rank_F                     NA         NA      NA       NA    
c19perBeh01          0.071234   0.031319   2.274 0.022987 *  
c19perBeh02          0.086841   0.038138   2.277 0.022834 *  
c19perBeh03          0.081073   0.026488   3.061 0.002221 ** 
c19RCA01             0.063305   0.014502   4.365 1.30e-05 ***
c19RCA02             0.025758   0.026837   0.960 0.337215    
c19RCA03            -0.045543   0.017601  -2.587 0.009700 ** 
coronaClose_1       -0.080699   0.272665  -0.296 0.767271    
coronaClose_2        0.169967   0.234093   0.726 0.467837    
coronaClose_3        0.315138   0.195393   1.613 0.106851    
coronaClose_4        0.198901   0.148356   1.341 0.180089    
coronaClose_5        0.016222   0.148702   0.109 0.913136    
coronaClose_6       -0.032219   0.153285  -0.210 0.833527    
gender               0.083495   0.052539   1.589 0.112092    
age                 -0.079134   0.021579  -3.667 0.000248 ***
edu                  0.004417   0.020448   0.216 0.828983    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.597 on 4368 degrees of freedom
  (412 observations deleted due to missingness)
Multiple R-squared:  0.1075,	Adjusted R-squared:  0.09808 
F-statistic: 11.44 on 46 and 4368 DF,  p-value: < 2.2e-16
```

Cluster Countries c19ProSo04
R-Squared = 0.1633
P-Value < 2.2e-16
Residual Data:

1Q	Median	3Q
-0.7730	0.3061	1.0591

Attribute	c19perBeh02	c19perBeh03	rank_A	rank_B
Meaning	To minimize my chances of getting coronavirus, I avoid crowded spaces.	To minimize my chances of getting coronavirus, I put myself in quarantine.	1-6 ranking of beauty	1-6 ranking of achievement
P-Value	1.58e-08	5.63e-07	1.53e-06	4.61e-06
Est. Coefficient	0.197263	0.121215	0.094614	0.086211

```
Residuals:
    Min      1Q  Median      3Q     Max 
-5.3105 -0.7730  0.3061  1.0591  4.2463 

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -0.393842   0.369511  -1.066 0.286552    
employstatus_1       0.052403   0.080137   0.654 0.513200    
employstatus_2       0.098132   0.081935   1.198 0.231107    
employstatus_3       0.208001   0.080288   2.591 0.009610 ** 
employstatus_4       0.017812   0.085977   0.207 0.835881    
employstatus_5       0.135756   0.100893   1.346 0.178519    
employstatus_6      -0.109791   0.090459  -1.214 0.224928    
employstatus_7      -0.087642   0.104428  -0.839 0.401372    
employstatus_8       0.003047   0.237624   0.013 0.989769    
employstatus_9       0.267357   0.080697   3.313 0.000930 ***
employstatus_10      0.397750   0.153273   2.595 0.009490 ** 
isoFriends_inPerson -0.034476   0.011041  -3.122 0.001805 ** 
isoOthPpl_inPerson  -0.007383   0.012245  -0.603 0.546569    
isoFriends_online    0.031634   0.011012   2.873 0.004090 ** 
isoOthPpl_online     0.016374   0.009122   1.795 0.072723 .  
lone01               0.005107   0.026732   0.191 0.848507    
lone02              -0.008923   0.023150  -0.385 0.699933    
lone03               0.040973   0.026337   1.556 0.119850    
happy               -0.009453   0.015399  -0.614 0.539339    
lifeSat              0.096692   0.025142   3.846 0.000122 ***
MLQ                  0.022306   0.017136   1.302 0.193091    
bor01                0.030430   0.017090   1.781 0.075044 .  
bor02                0.033865   0.017681   1.915 0.055518 .  
bor03                0.039317   0.014445   2.722 0.006516 ** 
consp01              0.054177   0.012031   4.503 6.87e-06 ***
consp02             -0.048047   0.012732  -3.774 0.000163 ***
consp03              0.012592   0.008825   1.427 0.153665    
rank_A               0.094614   0.019653   4.814 1.53e-06 ***
rank_B               0.086211   0.018791   4.588 4.61e-06 ***
rank_C               0.057144   0.016888   3.384 0.000721 ***
rank_D               0.047990   0.018065   2.656 0.007926 ** 
rank_E               0.002136   0.018206   0.117 0.906601    
rank_F                     NA         NA      NA       NA    
c19perBeh01          0.037615   0.028603   1.315 0.188555    
c19perBeh02          0.197263   0.034831   5.664 1.58e-08 ***
c19perBeh03          0.121215   0.024190   5.011 5.63e-07 ***
c19RCA01             0.037012   0.013244   2.795 0.005218 ** 
c19RCA02             0.098897   0.024509   4.035 5.55e-05 ***
c19RCA03            -0.011030   0.016074  -0.686 0.492645    
coronaClose_1       -0.077576   0.249021  -0.312 0.755417    
coronaClose_2        0.262715   0.213793   1.229 0.219203    
coronaClose_3        0.172493   0.178446   0.967 0.333777    
coronaClose_4        0.092912   0.135489   0.686 0.492905    
coronaClose_5       -0.038214   0.135805  -0.281 0.778428    
coronaClose_6       -0.131288   0.139992  -0.938 0.348387    
gender              -0.028888   0.047983  -0.602 0.547184    
age                  0.012654   0.019707   0.642 0.520829    
edu                  0.047641   0.018671   2.552 0.010758 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.459 on 4369 degrees of freedom
  (411 observations deleted due to missingness)
Multiple R-squared:  0.1633,	Adjusted R-squared:  0.1545 
F-statistic: 18.54 on 46 and 4369 DF,  p-value: < 2.2e-16
```

Libraries Used

Ggplot2:
    Graphing and plotting functions
Dplyr:
    For data wrangling and dataframe manipulation
TidyR:
    Pivoting tables to long and wide formats
Patchwork:
    Easy combination of plots
Likert:
    Easy processing and plotting of likert scale (or similar) data
RColorBrewer:
    Graph customisation
Dendextend:
    Dendrogram customisation
Circlize:
    Circular graph creation.

R Code

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(patchwork)
library(likert)
library(RColorBrewer)


#########################################################
#data setup
#########################################################

rm(list = ls())
set.seed(32471033) # XXXXXXXX = your student ID
cvbase = read.csv("PsyCoronaBaselineExtract.csv")
cvbase <- cvbase[sample(nrow(cvbase), 40000), ] # 40000 rows

#########################################################
#1a
#########################################################

#get num of rows and columns
dim(cvbase)

#########################################################
#1b
#########################################################

#filter out responses where all relevant questions are left blank
cvbase <- cvbase %>%
```

```
  filter(rowSums(!is.na(select(., -coded_country))) > 0)

#convert all hours worked into 1 column
cvbase <- cvbase %>%
  transform(employstatus_hours = case_when
    (
      employstatus_1 == 1 ~ 1,
      employstatus_2 == 1 ~ 2,
      employstatus_3 == 1 ~ 3,
      TRUE ~ NA
    )
  )

cvbase <- cvbase %>%
  transform(employstatus_lookingForWork = case_when
    (
      employstatus_4 == 1 ~ 1,
      employstatus_5 == 1 ~ 0,
      TRUE ~ NA
    )
  )

#create new column for employed or unemployed
cvbase <- cvbase %>%
  transform(employstatus_employed = ifelse(is.na(employstatus_hours), 0, 1))

#add focus country indicator
cvbase <- cvbase %>%
  mutate(focusCountry = ifelse(coded_country == "Brazil", 1, 0))

########################################################
#2a
########################################################

#################
#Fig1
#################

#create new dataframe for employment percentage values
percentage_data_employment <- cvbase %>%
  group_by(focusCountry, employstatus_employed) %>%
  summarise(count = n()) %>%
  group_by(focusCountry) %>%
  mutate(total_count = sum(count)) %>%
  mutate(percentage = count / total_count * 100)

#Factor Variables for graph labelling
percentage_data_employment$employstatus_employed <-
factor(percentage_data_employment$employstatus_employed, levels = c(0, 1), labels = c("Unemployed",
"Employed"))
```

```
percentage_data_employment$focusCountry <- factor(percentage_data_employment$focusCountry, levels =
c(0, 1), labels = c("Other Countries", "Brazil"))

# Create the dodged bar chart
ggplot(percentage_data_employment, aes(x = employstatus_employed, y = percentage, fill = focusCountry)) +

  geom_bar(position = position_dodge(width = 0.7), stat = "identity") +

  geom_text(aes(label = paste0(round(percentage), "%")),
            position = position_dodge(width = 0.7), vjust = -0.5, size = 3.5, color = "black") +

  scale_fill_manual(values = c("Other Countries" = "blue", "Brazil" = "orange")) +

  labs(title = "Employment Status", x = "Employment Status", y = "Percentage", fill = "Legend") +

  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

#################
#Fig2
#################

#create new dataframe for hours worked percentage values
percentage_data_hours <- cvbase %>%
  group_by(focusCountry, employstatus_hours) %>%
  summarise(count = n()) %>%
  group_by(focusCountry)
  percentage_data_hours <- na.omit(percentage_data_hours) %>%
  mutate(total_count = sum(count)) %>%
  mutate(percentage = count / total_count * 100)

#Factor Variables for graph labelling
percentage_data_hours$employstatus_hours <- factor(percentage_data_hours$employstatus_hours, levels =
c(1, 2, 3), labels = c("1-24", "24-39", "40+"))

percentage_data_hours$focusCountry <- factor(percentage_data_hours$focusCountry, levels = c(0, 1), labels
= c("Other Countries", "Brazil"))

# Create dodged bar chart
ggplot(percentage_data_hours, aes(x = employstatus_hours, y = percentage, fill = focusCountry)) +

  geom_bar(position = position_dodge(width = 0.7), stat = "identity") +

  geom_text(aes(label = paste0(round(percentage), "%")),
            position = position_dodge(width = 0.7), vjust = -0.5, size = 3.5, color = "black") +

  scale_fill_manual(values = c("Other Countries" = "blue", "Brazil" = "orange")) +

  labs(title = "Hours Worked", x = "Hours Worked", y = "Percentage", fill = "Legend") +

  theme_minimal() +
```

```
  theme(plot.title = element_text(hjust = 0.5))

#################
#Fig3
#################

#create new dataframe for looking for work percentage values
percentage_lookingForWork <- cvbase %>%
  group_by(focusCountry, employstatus_lookingForWork) %>%
  summarise(count = n()) %>%
  group_by(focusCountry)
percentage_lookingForWork <- na.omit(percentage_lookingForWork) %>%
  mutate(total_count = sum(count)) %>%
  mutate(percentage = count / total_count * 100)

#Factor Variables for graph labelling
percentage_lookingForWork$employstatus_lookingForWork <-
factor(percentage_lookingForWork$employstatus_lookingForWork, levels = c(0, 1), labels = c("Not Looking
For Work", "Looking For Work"))

percentage_lookingForWork$focusCountry <- factor(percentage_lookingForWork$focusCountry, levels = c(0,
1), labels = c("Other Countries", "Brazil"))

#create dodged bar chart
ggplot(percentage_lookingForWork, aes(x = employstatus_lookingForWork, y = percentage, fill =
focusCountry)) +

  geom_bar(position = position_dodge(width = 0.7), stat = "identity") +

  geom_text(aes(label = paste0(round(percentage), "%")),
            position = position_dodge(width = 0.7), vjust = -0.5, size = 3.5, color = "black") +

  scale_fill_manual(values = c("Other Countries" = "blue", "Brazil" = "orange")) +

  labs(title = "Proportion of Unemployed People Looking for Work", x = "", y = "Percentage", fill = "Legend") +

  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

#################
#Fig4
#################

#Create new dataframe for the rest of employstatus attributes
percentage_employstatusRest <- cvbase %>%
  group_by(focusCountry) %>%
  summarise(
    employstatus_6 = sum(employstatus_6 == 1, na.rm = TRUE),
    employstatus_7 = sum(employstatus_7 == 1, na.rm = TRUE),
    employstatus_8 = sum(employstatus_8 == 1, na.rm = TRUE),
    employstatus_9 = sum(employstatus_9 == 1, na.rm = TRUE),
```

```
    employstatus_10 = sum(employstatus_10 == 1, na.rm = TRUE),
    total_count = sum(focusCountry %in% c(0, 1), na.rm = TRUE),  # Calculate total count for country 0 or 1
  )

#convert to long format
percentage_employstatusRest <- percentage_employstatusRest %>%
  pivot_longer(cols = starts_with("employstatus_"),
               names_to = "employment_status",
               values_to = "count")

#calculate percentages
percentage_employstatusRest <- percentage_employstatusRest %>%
  mutate(percentage = count / total_count * 100)

#Rename column factors for labelling purposes
percentage_employstatusRest <- percentage_employstatusRest %>%
  mutate(employment_status = case_when(
    employment_status == "employstatus_6" ~ "Homemaker",
    employment_status == "employstatus_7" ~ "Retired",
    employment_status == "employstatus_8" ~ "Disabled, Unable to work",
    employment_status == "employstatus_9" ~ "Student",
    employment_status == "employstatus_10" ~ "Volunteering"
  ))

#factor data for labelling purposes
percentage_employstatusRest$focusCountry <- factor(percentage_employstatusRest$focusCountry, levels =
c(0, 1), labels = c("Other Countries", "Brazil"))

# Plot the dodged bar chart
ggplot(percentage_employstatusRest, aes(x = employment_status, y = percentage, fill = focusCountry)) +

  geom_bar(position = position_dodge(width = 0.7), stat = "identity") +

  geom_text(aes(label = paste0(round(percentage), "%")),
            position = position_dodge(width = 0.7), vjust = -0.5, size = 3.5, color = "black") +

  scale_fill_manual(values = c("Other Countries" = "blue", "Brazil" = "orange")) +

  labs(title = "Percentage of People Identifying as Below", x = "", y = "Percentage", fill = "Legend") +

  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

#################
#Fig5
#################

#create new dataframe for isoFriends_inPerson
isoFriends_inPerson <- cvbase[, c("isoFriends_inPerson", "focusCountry")]

#remove NA values
```

```
isoFriends_inPerson <- na.omit(isoFriends_inPerson)

#factor data for labelling
isoFriends_inPerson$focusCountry <- factor(isoFriends_inPerson$focusCountry, levels = c(0, 1), labels =
c("Other Countries", "Brazil"))

#draw graph 1
isoPlot1 <- ggplot(isoFriends_inPerson, aes(x = factor(focusCountry), fill = factor(isoFriends_inPerson))) +
  geom_bar(position = "fill") +
  scale_fill_brewer(palette = "GnBu") +
  labs(title = "In Person Contact with Friends and Family in the Past Week", x = "Country", y = "Proportion", fill
= "Days of Contact") +
  theme_minimal()

#create new dataframe for isoOthPpl_inPerson
isoOthPpl_inPerson <- cvbase[, c("isoOthPpl_inPerson", "focusCountry")]

#remove NA values
isoOthPpl_inPerson <- na.omit(isoOthPpl_inPerson)

#factor data for labelling
isoOthPpl_inPerson$focusCountry <- factor(isoOthPpl_inPerson$focusCountry, levels = c(0, 1), labels =
c("Other Countries", "Brazil"))

#draw graph 2
isoPlot2 <- ggplot(isoOthPpl_inPerson, aes(x = factor(focusCountry), fill = factor(isoOthPpl_inPerson))) +
  geom_bar(position = "fill") +
  scale_fill_brewer(palette = "GnBu") +
  labs(title = "In Person Contact with Other People in the Past Week", x = "Country", y = "Proportion", fill =
"Days of Contact") +
  theme_minimal()

#create new dataframe for isoFriends_Online
isoFriends_online <- cvbase[, c("isoFriends_online", "focusCountry")]

#remove NA values
isoFriends_online <- na.omit(isoFriends_online)

#factor data for labelling
isoFriends_online$focusCountry <- factor(isoFriends_online$focusCountry, levels = c(0, 1), labels = c("Other
Countries", "Brazil"))

#draw graph 3
isoPlot3 <- ggplot(isoFriends_online, aes(x = factor(focusCountry), fill = factor(isoFriends_online))) +
  geom_bar(position = "fill") +
  scale_fill_brewer(palette = "GnBu") +
  labs(title = "Online Contact with Friends and Family in the Past Week", x = "Country", y = "Proportion", fill =
"Days of Contact") +
  theme_minimal()

#create new dataframe for isoOthPpl_Online
```

```
isoOthPpl_online <- cvbase[, c("isoOthPpl_online", "focusCountry")]

#remove NA values
isoOthPpl_online <- na.omit(isoOthPpl_online)

#factor data for labelling
isoOthPpl_online$focusCountry <- factor(isoOthPpl_online$focusCountry, levels = c(0, 1), labels = c("Other
Countries", "Brazil"))

#draw graph 4
isoPlot4 <- ggplot(isoOthPpl_online, aes(x = factor(focusCountry), fill = factor(isoOthPpl_online))) +
  geom_bar(position = "fill") +
  scale_fill_brewer(palette = "GnBu") +
  labs(title = "Online Contact with Other People in the Past Week", x = "Country", y = "Proportion", fill = "Days
of Contact") +
  theme_minimal()

#combine plots using patchwork library

isoPlot1 +
isoPlot2 +
isoPlot3 +
isoPlot4

#################
#Fig6
#################

#NOTES
#Refer to https://stackoverflow.com/questions/27061286/likert-grouping-with-different-levels-in-r
#for future likert plotting
#extremely helpful

#create dataframe for loneliness
lone_data <- cvbase[, c("focusCountry", "lone01", "lone02", "lone03")]

#apply factor function to all question responses to give response levels
lone_data[2:4] <- lapply(lone_data[2:4], factor, levels = 1:5, labels = c("Never", "Rarely", "Sometimes", "Often",
"All the Time"))

#factor focusCountry to get label names
lone_data$focusCountry <- factor(lone_data$focusCountry, levels = c(0, 1), labels = c("Other Countries",
"Brazil"))

#rename question columns to represent the question text
lone_data <- lone_data %>%
  rename("During the past week, did you feel lonely?" = lone01,
      "During the past week, did you feel isolated from others?" = lone02,
      "During the past week, did you feel left out?" = lone03
   )
```

```
#convert responses to a likert object, and group by focus country
lone_likert <- likert(lone_data[, c(2:4)], grouping = lone_data$focusCountry)

#plot the graph
plot(lone_likert)

#################
#Fig7
#################

#create dataframe for happy

happy_data <- cvbase[, c("focusCountry", "happy")]
happy_data <- na.omit(happy_data)

#factor for labelling
happy_data$focusCountry <- factor(happy_data$focusCountry, levels = c(0, 1), labels = c("Other Countries",
"Brazil"))

#create and store boxplot
happy_plot <- ggplot(happy_data, aes(x = focusCountry, y = happy, fill = focusCountry)) +

  geom_boxplot() +

  scale_y_continuous(breaks = seq(0, 10, by = 1)) +

  scale_fill_manual(values = c("Other Countries" = "lightblue", "Brazil" = "orange")) +

  labs(title = "Rate your Happiness from 1 (Extremely Unhappy) to 10 (Extremely Happy)", x = "Country", y =
"Rating", fill = "Legend") +

  theme_minimal()

#create dataframe for lifeSat
lifeSat_data <- cvbase[, c("focusCountry", "lifeSat")]
lifeSat_data <- na.omit(lifeSat_data)

#assign levels to response values
lifeSat_data[2] <- lapply(lifeSat_data[2], factor, levels = 1:6, labels = c("Very Dissatisfied", "Dissatisfied",
"Slightly Dissatisfied", "Slightly Satisfied", "Satisfied", "Very Satisfied"))

#factor to assign labels to focusCountry
lifeSat_data$focusCountry <- factor(lifeSat_data$focusCountry, levels = c(0, 1), labels = c("Other Countries",
"Brazil"))

#rename question column to represent the actual question
lifeSat_data <- lifeSat_data %>%
  rename("In general, how satisfied are you with your life?" = lifeSat
    )

#convert to likert object
```

```
lifeSat_likert <- likert(items=lifeSat_data[,2, drop=FALSE], grouping = lifeSat_data$focusCountry)

#Store the plot
lifeSat_plot <- plot(lifeSat_likert)

#create dataframe for MLQ
mlq_data <- cvbase[, c("focusCountry", "MLQ")]
mlq_data <- na.omit(mlq_data)

#convert values from -3:3 to 1:7 because idk how to assign levels to negative numbers
mlq_data <- mlq_data %>%
  mutate(
    MLQ = MLQ + 4
    )

#assign levels to response values
mlq_data[2] <- lapply(mlq_data[2], factor, levels = 1:7,
            labels = c("Strongly Disagree", "Disagree", "Somewhat Disagree", "Neither Agree nor Disagree",
"Somewhat Agree", "Agree", "Strongly Agree")
                    )

#assign labels to focusCountry
mlq_data$focusCountry <- factor(mlq_data$focusCountry, levels = c(0, 1), labels = c("Other Countries",
"Brazil"))

#rename question name
mlq_data <- mlq_data %>%
  rename("My life has a clear sense of purpose" = MLQ
  )

#convert to likert object
mlq_likert <- likert(items=mlq_data[,2, drop=FALSE], grouping = mlq_data$focusCountry)

mlq_plot <- plot(mlq_likert)

#combine plots
happy_plot + lifeSat_plot / mlq_plot

#################
#Fig8
#################

#create dataframe for boredom
bor_data <- cvbase[, c("focusCountry", "bor01", "bor02", "bor03")]

#convert from -3:3 to 1:7
bor_data <- bor_data %>%
  mutate(
    bor01 = bor01 + 4,
    bor02 = bor02 + 4,
    bor03 = bor03 + 4
```

```
  )

#assign levels to response values
bor_data[2:4] <- lapply(bor_data[2:4], factor, levels = 1:7,
                    labels = c("Strongly Disagree", "Disagree", "Somewhat Disagree", "Neither Agree nor Disagree",
"Somewhat Agree", "Agree", "Strongly Agree")
)

#factor focusCountry to get label names
bor_data$focusCountry <- factor(bor_data$focusCountry, levels = c(0, 1), labels = c("Other Countries",
"Brazil"))

#rename question columns to represent the question text
bor_data <- bor_data %>%
  rename("I Wish Time Would Go Faster" = bor01,
        "Time is Moving Very Slowly" = bor02,
        "I Feel in Control of My Time" = bor03
  )

#convert responses to a likert object, and group by focus country
bor_likert <- likert(bor_data[, c(2:4)], grouping = bor_data$focusCountry)

#plot the graph
plot(bor_likert)

#################
#Fig9
#################

#create dataframe for Conspiracy
consp_data <- cvbase[, c("focusCountry", "consp01", "consp02", "consp03")]

#assign levels to response values
consp_data[2:4] <- lapply(consp_data[2:4], factor, levels = 0:10,
  labels = c("Certainly Not", "10%", "20%", "30%", "40%", "Undecided", "60%", "70%", "80%", "90%",
"Certainly")
)

#factor focusCountry to get label names
consp_data$focusCountry <- factor(consp_data$focusCountry, levels = c(0, 1), labels = c("Other Countries",
"Brazil"))

#rename question columns to represent the question text
consp_data <- consp_data %>%
  rename("think that many very important things happen in the world, which the public is never informed about"
= consp01,
        "I think that politicians usually do not tell us the true motives for their decisions" = consp02,
        "I think that government agencies closely monitor all citizens." = consp03
  )

#convert responses to a likert object, and group by focus country
```

```
consp_likert <- likert(consp_data[, c(2:4)], grouping = consp_data$focusCountry)

#plot the graph
plot(consp_likert)

#################
#Fig10
#################

#create dataframe for rankOrdLife
rankOrd_data <- cvbase[, c("focusCountry", "rankOrdLife_1", "rankOrdLife_2", "rankOrdLife_3",
"rankOrdLife_4", "rankOrdLife_5", "rankOrdLife_6")]

#pivot table
rankOrd_data <- rankOrd_data %>%
  pivot_longer(cols = starts_with("rankOrdLife"), names_to = "rank", values_to = "category")

#omit NA vals
rankOrd_data <- na.omit(rankOrd_data)

#rename row values to numerical ranking
rankOrd_data <- rankOrd_data %>%
  mutate(rank = case_when(
    rank == "rankOrdLife_1" ~ 1,
    rank == "rankOrdLife_2" ~ 2,
    rank == "rankOrdLife_3" ~ 3,
    rank == "rankOrdLife_4" ~ 4,
    rank == "rankOrdLife_5" ~ 5,
    rank == "rankOrdLife_6" ~ 6
  ))

#rename row values to categories
rankOrd_data <- rankOrd_data %>%
  mutate(category = case_when(
    category == "A" ~ "Beauty",
    category == "B" ~ "Achievement",
    category == "C" ~ "Victory",
    category == "D" ~ "Friendship",
    category == "E" ~ "Love",
    category == "F" ~ "Empathy"
  ))

#label country
rankOrd_data$focusCountry <- factor(rankOrd_data$focusCountry, levels = c(0, 1), labels = c("Other
Countries", "Brazil"))

#draw boxplot
ggplot(rankOrd_data, aes(x=category, y=rank, fill=focusCountry)) +

  geom_boxplot() +
  scale_fill_manual(values = c("Other Countries" = "lightblue", "Brazil" = "orange")) +
```

```
  scale_y_continuous(trans = "reverse", breaks = unique(rankOrd_data$rank)) +
  labs(title = "Rank the Following in Order From 1 (Most Important) to 6 (Least Important)", x = "Category", y =
"Rank", fill = "Legend") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

#################
#Fig11
#################

#create dataframe for c19perBeh and c19RCA
c19Resp_data <- cvbase[, c("focusCountry", "c19perBeh01", "c19perBeh02", "c19perBeh03", "c19RCA01",
"c19RCA02", "c19RCA03")]

#convert from -3:3 to 1:7
c19Resp_data <- c19Resp_data %>%
  mutate(
    c19perBeh01 = c19perBeh01 + 4,
    c19perBeh02 = c19perBeh02 + 4,
    c19perBeh03 = c19perBeh03 + 4,
    c19RCA01 = c19RCA01 + 4,
    c19RCA02 = c19RCA02 + 4,
    c19RCA03 = c19RCA03 + 4
  )

#assign levels to response values
c19Resp_data[2:7] <- lapply(c19Resp_data[2:7], factor, levels = 1:7,
                        labels = c("Strongly Disagree", "Disagree", "Somewhat Disagree", "Neither Agree nor
Disagree", "Somewhat Agree", "Agree", "Strongly Agree")
)

#factor focusCountry to get label names
c19Resp_data$focusCountry <- factor(c19Resp_data$focusCountry, levels = c(0, 1), labels = c("Other
Countries", "Brazil"))

#rename question columns to represent the question text
c19Resp_data <- c19Resp_data %>%
  rename("To minimize my chances of getting coronavirus, I wash my hands more often" = c19perBeh01,
        "To minimize my chances of getting coronavirus, I avoid crowded spaces" = c19perBeh02,
        "To minimize my chances of getting coronavirus, I put myself in quarantine" = c19perBeh03,
        "I would sign a petition that supports mandatory vaccination once a vaccine has been developed for
coronavirus." = c19RCA01,
        "I would sign a petition that supports mandatory quarantine for those that have coronavirus and those
that have been exposed to the virus." = c19RCA02,
        "I would sign a petition that supports reporting people who are suspected to have coronavirus." =
c19RCA03
  )

#convert responses to a likert object, and group by focus country
c19Resp_likert <- likert(c19Resp_data[, c(2:7)], grouping = c19Resp_data$focusCountry)
```

```
#plot the graph
plot(c19Resp_likert)

#################
#Fig12
#################

#Create new dataframe for coronaClose
percentage_coronaClose <- cvbase %>%
  group_by(focusCountry) %>%
  summarise(
    coronaClose_1 = sum(coronaClose_1 == 1, na.rm = TRUE),
    coronaClose_2 = sum(coronaClose_2 == 1, na.rm = TRUE),
    coronaClose_3 = sum(coronaClose_3 == 1, na.rm = TRUE),
    coronaClose_4 = sum(coronaClose_4 == 1, na.rm = TRUE),
    coronaClose_5 = sum(coronaClose_5 == 1, na.rm = TRUE),
    coronaClose_6 = sum(coronaClose_6 == 1, na.rm = TRUE),
    total_count = sum(focusCountry %in% c(0, 1), na.rm = TRUE),  # Calculate total count for country 0 or 1
  )

#convert to long format
percentage_coronaClose <- percentage_coronaClose %>%
  pivot_longer(cols = starts_with("coronaClose_"),
               names_to = "coronaClose",
               values_to = "count")

#calculate percentages
percentage_coronaClose <- percentage_coronaClose %>%
  mutate(percentage = count / total_count * 100)

#Rename column factors for labelling purposes
percentage_coronaClose <- percentage_coronaClose %>%
  mutate(coronaClose = case_when(
    coronaClose == "coronaClose_1" ~ "Yes, Myself",
    coronaClose == "coronaClose_2" ~ "Yes, a member of my family",
    coronaClose == "coronaClose_3" ~ "Yes, a close friend",
    coronaClose == "coronaClose_4" ~ "Yes, someone I know",
    coronaClose == "coronaClose_5" ~ "Yes, someone else",
    coronaClose == "coronaClose_6" ~ "No, I do not know anyone"
  ))

percentage_coronaClose$focusCountry <- factor(percentage_coronaClose$focusCountry, levels = c(0, 1),
labels = c("Other Countries", "Brazil"))

# Plot the dodged bar chart
ggplot(percentage_coronaClose, aes(x = coronaClose, y = percentage, fill = focusCountry)) +

  geom_bar(position = position_dodge(width = 0.7), stat = "identity") +
  geom_text(aes(label = paste0(round(percentage), "%")),
          position = position_dodge(width = 0.7), vjust = -0.5, size = 3.5, color = "black") +
```

```
  scale_fill_manual(values = c("Other Countries" = "blue", "Brazil" = "orange")) +
  labs(title = "Do you personally know anyone who currently has coronavirus?", x = "", y = "Percentage", fill =
"Legend") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

#################
#Fig13
#################

#create dataframe for gender
gender_data <- cvbase[, c("focusCountry", "gender")]

#remove NA values
gender_data <- na.omit(gender_data)

#factor data for labelling
gender_data$focusCountry <- factor(gender_data$focusCountry, levels = c(0, 1), labels = c("Other Countries",
"Brazil"))

gender_data$gender <- factor(gender_data$gender, levels = (1:3), labels = c("Female", "Male", "Other"))

#draw stacked bar chart
ggplot(gender_data, aes(x = factor(focusCountry), fill = factor(gender))) +
  geom_bar(position = "fill") +
  scale_fill_brewer(palette = "RdYlBu") +
  labs(title = "Gender of Participant", x = "Country", y = "Proportion", fill = "Gender") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

#################
#Fig14
#################

#create dataframe for age
age_data <- cvbase[, c("focusCountry", "age")]

#remove NA values
age_data <- na.omit(age_data)

#factor focusCountry for labelling
age_data$focusCountry <- factor(age_data$focusCountry, levels = c(0, 1), labels = c("Other Countries",
"Brazil"))

#factor age values for labelling
age_data$age <- factor(age_data$age, levels = (1:8), labels = c("18-24", "25-34", "35-44", "45-54", "55-64",
"65-75", "75-85", "85+"))

#draw graph 4
ggplot(age_data, aes(x = factor(focusCountry), fill = factor(age))) +
  geom_bar(position = "fill") +
```

```
  scale_fill_brewer(palette = "GnBu") +
  labs(title = "Age of Participants", x = "Country", y = "Proportion", fill = "Age") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

#################
#Fig15
#################

#create dataframe for education level
edu_data <- cvbase[, c("focusCountry", "edu")]

#remove NA values
edu_data <- na.omit(edu_data)

#factor focusCountry for labelling
edu_data$focusCountry <- factor(edu_data$focusCountry, levels = c(0, 1), labels = c("Other Countries",
"Brazil"))

#factor age values for labelling
edu_data$edu <- factor(edu_data$edu, levels = (1:7),
    labels = c("Primary Education", "General Secondary Education", "Vocational Education", "Higher
Education", "Bachelors Degree", "Masters Degree", "PhD Degree"))

#draw graph 4
ggplot(edu_data, aes(x = factor(focusCountry), fill = factor(edu))) +
  geom_bar(position = "fill") +
  scale_fill_brewer(palette = "GnBu") +
  labs(title = "Education Level of Participants", x = "Country", y = "Proportion", fill = "Education Level") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

#################
#Fig16
#################

#create dataframe for country data
country_data <- cvbase %>%
  summarise(
    "Other Countries" = sum(focusCountry == 0, na.rm = TRUE),
    "Brazil" = sum(focusCountry == 1, na.rm = TRUE),
  )

#pivot to longer, and add percentage column
country_data <- country_data %>%
  pivot_longer(everything(), names_to = "focusCountry", values_to = "count") %>%
  mutate(
    percentage = count / sum(count) * 100
  )

#create plot
```

```
ggplot(country_data, aes(x = "", y = percentage, fill = focusCountry)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  geom_text(aes(label = paste0(round(percentage), "%")),
           position = position_stack(vjust = 0.5), size = 4, color = "black") +
  scale_fill_manual(values = c("Brazil" = "#f4a582", "Other Countries" = "#92c5de")) +
  labs(title = "Country of Residence",
       x = NULL, y = "Percentage", fill = "Country") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

#################
#Fig17
#################

#create dataframe for c19ProSo
c19ProSo_data <- cvbase[, c("focusCountry", "c19ProSo01", "c19ProSo02", "c19ProSo03", "c19ProSo04")]

#convert from -3:3 to 1:7
c19ProSo_data <- c19ProSo_data %>%
  mutate(
    c19ProSo01 = c19ProSo01 + 4,
    c19ProSo02 = c19ProSo02 + 4,
    c19ProSo03 = c19ProSo03 + 4,
    c19ProSo04 = c19ProSo04 + 4
  )

#assign levels to response values
c19ProSo_data[2:5] <- lapply(c19ProSo_data[2:5], factor, levels = 1:7,
                        labels = c("Strongly Disagree", "Disagree", "Somewhat Disagree", "Neither Agree nor
Disagree", "Somewhat Agree", "Agree", "Strongly Agree")
)

#factor focusCountry to get label names
c19ProSo_data$focusCountry <- factor(c19ProSo_data$focusCountry, levels = c(0, 1), labels = c("Other
Countries", "Brazil"))

#rename question columns to represent the question text
c19ProSo_data <- c19ProSo_data %>%
  rename("I am willing to help others who suffer from coronavirus." = c19ProSo01,
         "I am willing to make donations to help others that suffer from coronavirus." = c19ProSo02,
         "I am willing to protect vulnerable groups from coronavirus even at my own expense." = c19ProSo03,
         "I am willing to make personal sacrifices to prevent the spread of coronavirus." = c19ProSo04,
  )

#convert responses to a likert object, and group by focus country
c19ProSo_likert <- likert(c19ProSo_data[, c(2:5)], grouping = c19ProSo_data$focusCountry)

#plot the graph
plot(c19ProSo_likert)
```

```
############################################################
#2b and 2c
############################################################

#assign dataframe
regression_data <- cvbase

#change na vals to 0 for dummy vars
regression_data[, c(1:10, 39:44)][is.na(regression_data[, c(1:10, 39:44)])] <- 0

#convert scales that include neg vals to positive only
regression_data <- regression_data %>%
  mutate(
    c19ProSo01 = c19ProSo01 + 4,
    c19ProSo02 = c19ProSo02 + 4,
    c19ProSo03 = c19ProSo03 + 4,
    c19ProSo04 = c19ProSo04 + 4,
    c19perBeh01 = c19perBeh01 + 4,
    c19perBeh02 = c19perBeh02 + 4,
    c19perBeh03 = c19perBeh03 + 4,
    c19RCA01 = c19RCA01 + 4,
    c19RCA02 = c19RCA02 + 4,
    c19RCA03 = c19RCA03 + 4,
    bor01 = bor01 + 4,
    bor02 = bor02 + 4,
    bor03 = bor03 + 4,
    MLQ = MLQ + 4
  )

#convert rankOrd to be in terms of category, rather than rank
regression_processing <- regression_data %>%
  mutate(rowNum = row_number()) %>%
  pivot_longer(
    cols = starts_with("rankOrdLife"),
    names_to = "rank",
    values_to = "category"
  )

#NA out rankOrd responses where multiple categories have the same rank (idk how this would even occur
given the nature of the question)
#look at rowNum == 141 for example
regression_processing <- regression_processing %>%
  group_by(rowNum) %>%
  mutate(category = if (!any(is.na(category)) && length(unique(category)) < 6) NA else category = category)
%>%
  ungroup()

#rename column text values to be numerical
regression_processing <- regression_processing %>%
  mutate(rank = case_when(
    rank == "rankOrdLife_1" ~ 1,
```

```
    rank == "rankOrdLife_2" ~ 2,
    rank == "rankOrdLife_3" ~ 3,
    rank == "rankOrdLife_4" ~ 4,
    rank == "rankOrdLife_5" ~ 5,
    rank == "rankOrdLife_6" ~ 6
  ))

#pivot back to wide format to return to correct num of rows
regression_processing <- regression_processing %>%
  pivot_wider(
    id_cols = rowNum,
    names_from = category,
    values_from = rank,
    names_prefix = "rank_"
  )

#remove NA ranking column and rowNum column
regression_processing <- regression_processing %>%
  select(-c(rowNum, rank_NA))

#when pivoting back to wide format, list columns were created, due to the NA values
#we will have to convert the list columns back into regular columns to make any more changes
regression_processing <- unnest_longer(regression_processing, col = c(rank_C, rank_E, rank_F, rank_D,
rank_A, rank_B), keep_empty = TRUE)

#Replace NULL (No Value) with NA(Missing Value)
regression_processing <- replace(regression_processing, is.null(regression_processing), NA)

#Bind table to retrieve all other data
regression_processing <- regression_data %>%
  select(-starts_with("Rank")) %>%
  bind_cols(regression_processing)

#assign clean and tidy data to regression_data
regression_data <- regression_processing

#Extract data from focus country (Brazil) only
brazil_data <- regression_data[regression_data$focusCountry == 1, ]

#Extract data from other countries only
otherCountry_data <- regression_data[regression_data$focusCountry == 0, ]

#one-hot encode gender data
otherCountry_data <- otherCountry_data %>%
  transform(female = case_when
          (
            gender == 1 ~ 1,
            TRUE ~ 0
          )
  ) %>%
  transform(male = case_when
```

```
          (
            gender == 2 ~ 1,
            TRUE ~ 0
          )
  ) %>%
  transform(other = case_when
          (
            gender == 3 ~ 1,
            TRUE ~ 0
          )
  )
##########################
#Regressions
##########################

#regression for Brazil proso01
brazil_c19ProSo01 = lm(c19ProSo01 ~
                    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 +
                    rank_A +
                    rank_B +
                    rank_C +
                    rank_D +
                    rank_E +
                    rank_F +
                    c19perBeh01 +
                    c19perBeh02 +
```

```
                  c19perBeh03 +
                  c19RCA01 +
                  c19RCA02 +
                  c19RCA03 +
                  coronaClose_1 +
                  coronaClose_2 +
                  coronaClose_3 +
                  coronaClose_4 +
                  coronaClose_5 +
                  coronaClose_6 +
                  gender +
                  age +
                  edu,
                  data = brazil_data)

#regression for Brazil proso02
brazil_c19ProSo02 = lm(c19ProSo02 ~
                  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 +
                  rank_A +
                  rank_B +
                  rank_C +
                  rank_D +
                  rank_E +
                  rank_F +
                  c19perBeh01 +
                  c19perBeh02 +
```

```
                c19perBeh03 +
                c19RCA01 +
                c19RCA02 +
                c19RCA03 +
                coronaClose_1 +
                coronaClose_2 +
                coronaClose_3 +
                coronaClose_4 +
                coronaClose_5 +
                coronaClose_6 +
                gender +
                age +
                edu,
               data = brazil_data)

#regression for Brazil proso03
brazil_c19ProSo03 = lm(c19ProSo03 ~
                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 +
                rank_A +
                rank_B +
                rank_C +
                rank_D +
                rank_E +
                rank_F +
                c19perBeh01 +
                c19perBeh02 +
```

```
                  c19perBeh03 +
                  c19RCA01 +
                  c19RCA02 +
                  c19RCA03 +
                  coronaClose_1 +
                  coronaClose_2 +
                  coronaClose_3 +
                  coronaClose_4 +
                  coronaClose_5 +
                  coronaClose_6 +
                  gender +
                  age +
                  edu,
                data = brazil_data)

#regression for Brazil proso04
brazil_c19ProSo04 = lm(c19ProSo04 ~
                  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 +
                  rank_A +
                  rank_B +
                  rank_C +
                  rank_D +
                  rank_E +
                  rank_F +
                  c19perBeh01 +
                  c19perBeh02 +
```

```
                c19perBeh03 +
                c19RCA01 +
                c19RCA02 +
                c19RCA03 +
                coronaClose_1 +
                coronaClose_2 +
                coronaClose_3 +
                coronaClose_4 +
                coronaClose_5 +
                coronaClose_6 +
                gender +
                age +
                edu,
               data = brazil_data)

#create brazil summaries
brazil_proSo1_summary <- summary(brazil_c19ProSo01)
brazil_proSo2_summary <- summary(brazil_c19ProSo02)
brazil_proSo3_summary <- summary(brazil_c19ProSo03)
brazil_proSo4_summary <- summary(brazil_c19ProSo04)

#create otherCountry regressions
#regression for otherCountry proso01
otherCountry_c19ProSo01 = lm(c19ProSo01 ~
                  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 +
                  rank_A +
```

```
                  rank_B +
                  rank_C +
                  rank_D +
                  rank_E +
                  rank_F +
                  c19perBeh01 +
                  c19perBeh02 +
                  c19perBeh03 +
                  c19RCA01 +
                  c19RCA02 +
                  c19RCA03 +
                  coronaClose_1 +
                  coronaClose_2 +
                  coronaClose_3 +
                  coronaClose_4 +
                  coronaClose_5 +
                  coronaClose_6 +
                  gender +
                  age +
                  edu,
                 data = otherCountry_data)

#regression for otherCountry proso02
otherCountry_c19ProSo02 = lm(c19ProSo02 ~
                   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 +
                   rank_A +
```

```
                rank_B +
                rank_C +
                rank_D +
                rank_E +
                rank_F +
                c19perBeh01 +
                c19perBeh02 +
                c19perBeh03 +
                c19RCA01 +
                c19RCA02 +
                c19RCA03 +
                coronaClose_1 +
                coronaClose_2 +
                coronaClose_3 +
                coronaClose_4 +
                coronaClose_5 +
                coronaClose_6 +
                gender +
                age +
                edu,
               data = otherCountry_data)

#regression for otherCountry proso03
otherCountry_c19ProSo03 = lm(c19ProSo03 ~
                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 +
                rank_A +
```

```
                    rank_B +
                    rank_C +
                    rank_D +
                    rank_E +
                    rank_F +
                    c19perBeh01 +
                    c19perBeh02 +
                    c19perBeh03 +
                    c19RCA01 +
                    c19RCA02 +
                    c19RCA03 +
                    coronaClose_1 +
                    coronaClose_2 +
                    coronaClose_3 +
                    coronaClose_4 +
                    coronaClose_5 +
                    coronaClose_6 +
                    gender +
                    age +
                    edu,
                   data = otherCountry_data)

#regression for otherCountry proso04
otherCountry_c19ProSo04 = lm(c19ProSo04 ~
                    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 +
                    rank_A +
```

```
                   rank_B +
                   rank_C +
                   rank_D +
                   rank_E +
                   rank_F +
                   c19perBeh01 +
                   c19perBeh02 +
                   c19perBeh03 +
                   c19RCA01 +
                   c19RCA02 +
                   c19RCA03 +
                   coronaClose_1 +
                   coronaClose_2 +
                   coronaClose_3 +
                   coronaClose_4 +
                   coronaClose_5 +
                   coronaClose_6 +
                   gender +
                   age +
                   edu,
                  data = otherCountry_data)

#create otherCountry summaries
otherCountry_proSo1_summary <- summary(otherCountry_c19ProSo01)
otherCountry_proSo2_summary <- summary(otherCountry_c19ProSo02)
otherCountry_proSo3_summary <- summary(otherCountry_c19ProSo03)
otherCountry_proSo4_summary <- summary(otherCountry_c19ProSo04)

#view summaries (highlight to use)
brazil_proSo1_summary
brazil_proSo2_summary
brazil_proSo3_summary
brazil_proSo4_summary
otherCountry_proSo1_summary
otherCountry_proSo2_summary
otherCountry_proSo3_summary
otherCountry_proSo4_summary

#create correlation table for otherCountries
otherCountry_corTable <- cor(cbind(otherCountry_data[, c(1:38, 57:59, 40:41, 51:56, 43:46)]), use =
"complete.obs")

#take only relevant columns
otherCountry_corTable <- otherCountry_corTable[, 50:53]

#convert matrix to a dataframe
otherCountry_corTable <- as.data.frame(otherCountry_corTable)

#assign rows to remove (rows should be predictors, columns should be outcomes)
rows_to_remove <- c("c19ProSo01",
          "c19ProSo02",
```

```
                "c19ProSo03",
                "c19ProSo04")

#remove outcome rows, keeping only outcome columns
otherCountry_corTable <- subset(otherCountry_corTable, subset = !(rownames(otherCountry_corTable) %in%
rows_to_remove))

#create correlation heatmap
#add rownames (predictors) as a column
otherCountry_corTable$predictor <- rownames(otherCountry_corTable)

#pivot the table to get correlation of variable pairs
otherCountry_corHeatmap <- pivot_longer(otherCountry_corTable, cols = -predictor, names_to = "outcome",
values_to = "Correlation")

#plot the heatmap
ggplot(otherCountry_corHeatmap, aes(x = outcome, y = predictor, fill = Correlation)) +
  geom_tile() +
  scale_fill_gradient2(low = "#08306b", mid = "white", high = "#67000d", midpoint = 0) +
  theme_minimal() +
  labs(
    title = "Correlation Heatmap",
    x = "Outcomes",
    y = "Predictors",
    fill = "Correlation"
  )

View(otherCountry_corTable)

##########################################################
#3b
##########################################################

#label cluster countries
regression_data <- regression_data %>%
  mutate(clusterCountry = ifelse(coded_country %in% c("Bulgaria",
                                           "Armenia",
                                           "Georgia",
                                           "Argentina",
                                           "Panama",
                                           "Turkey",
                                           "Colombia",
                                           "Albania",
                                           "Ukraine",
                                           "Bosnia and Herzegovina",
                                           "Greece",
                                           "Montenegro"), 1, 0))

#extract cluster country data
clusterCountry_data <- regression_data[regression_data$clusterCountry == 1, ]
```

```
clusterCountry_c19ProSo01 = lm(c19ProSo01 ~
                          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 +
                          rank_A +
                          rank_B +
                          rank_C +
                          rank_D +
                          rank_E +
                          rank_F +
                          c19perBeh01 +
                          c19perBeh02 +
                          c19perBeh03 +
                          c19RCA01 +
                          c19RCA02 +
                          c19RCA03 +
                          coronaClose_1 +
                          coronaClose_2 +
                          coronaClose_3 +
                          coronaClose_4 +
                          coronaClose_5 +
                          coronaClose_6 +
                          gender +
                          age +
                          edu,
                        data = clusterCountry_data)

#regression for otherCountry proso02
```

```
clusterCountry_c19ProSo02 = lm(c19ProSo02 ~
                        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 +
                        rank_A +
                        rank_B +
                        rank_C +
                        rank_D +
                        rank_E +
                        rank_F +
                        c19perBeh01 +
                        c19perBeh02 +
                        c19perBeh03 +
                        c19RCA01 +
                        c19RCA02 +
                        c19RCA03 +
                        coronaClose_1 +
                        coronaClose_2 +
                        coronaClose_3 +
                        coronaClose_4 +
                        coronaClose_5 +
                        coronaClose_6 +
                        gender +
                        age +
                        edu,
                      data = clusterCountry_data)

#regression for otherCountry proso03
```

```
clusterCountry_c19ProSo03 = lm(c19ProSo03 ~
                         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 +
                         rank_A +
                         rank_B +
                         rank_C +
                         rank_D +
                         rank_E +
                         rank_F +
                         c19perBeh01 +
                         c19perBeh02 +
                         c19perBeh03 +
                         c19RCA01 +
                         c19RCA02 +
                         c19RCA03 +
                         coronaClose_1 +
                         coronaClose_2 +
                         coronaClose_3 +
                         coronaClose_4 +
                         coronaClose_5 +
                         coronaClose_6 +
                         gender +
                         age +
                         edu,
                       data = clusterCountry_data)

#regression for otherCountry proso04
```

```
clusterCountry_c19ProSo04 = lm(c19ProSo04 ~
                          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 +
                          rank_A +
                          rank_B +
                          rank_C +
                          rank_D +
                          rank_E +
                          rank_F +
                          c19perBeh01 +
                          c19perBeh02 +
                          c19perBeh03 +
                          c19RCA01 +
                          c19RCA02 +
                          c19RCA03 +
                          coronaClose_1 +
                          coronaClose_2 +
                          coronaClose_3 +
                          coronaClose_4 +
                          coronaClose_5 +
                          coronaClose_6 +
                          gender +
                          age +
                          edu,
                        data = clusterCountry_data)

#create otherCountry summaries
```

```
clusterCountry_proSo1_summary <- summary(clusterCountry_c19ProSo01)
clusterCountry_proSo2_summary <- summary(clusterCountry_c19ProSo02)
clusterCountry_proSo3_summary <- summary(clusterCountry_c19ProSo03)
clusterCountry_proSo4_summary <- summary(clusterCountry_c19ProSo04)

#show summaries
clusterCountry_proSo1_summary
clusterCountry_proSo2_summary
clusterCountry_proSo3_summary
clusterCountry_proSo4_summary
```

Clustering R Code:

```
library(dplyr)
library(dendextend)
library(circlize)

################################
#Data Gathering and Cleaning
################################

#load cvbase
cvbase = read.csv("PsyCoronaBaselineExtract.csv")

#extract countries
countryList <- data.frame("coded_country" = cvbase$coded_country)

#remove duplicates
clustering_data <- unique(countryList)

#remove NA values and blanks
clustering_data <- clustering_data %>%
  filter_all(any_vars(!is.na(.) & . != ""))

#load ghs data
ghs_data <- read.csv("2021-GHS-Index-April-2022.csv")

#rename columns
ghs_data <- data.frame(
    "coded_country" = ghs_data$"Country",
    "ghs_overall_score" = ghs_data$"OVERALL.SCORE")

#rename countries to match original data country names
ghs_data <- ghs_data %>%
  transform(coded_country = case_when
          (
            coded_country == "Bosnia and Hercegovina" ~ "Bosnia and Herzegovina",
            coded_country == "Kyrgyz Republic" ~ "Kyrgyzstan",
```

```
            coded_country == "Serbia" ~ "Republic of Serbia",
            TRUE ~ coded_country
          )
  )

#merge by country
clustering_data <- merge(clustering_data, ghs_data, by = "coded_country")

#load gdp per capita data (worldbank)
gdpPC2021_data <-read.csv("API_NY.GDP.PCAP.CD_DS2_en_csv_v2_133.csv")

#rename columns
gdpPC2021_data <- data.frame(
  "coded_country" = gdpPC2021_data$"Country.Name",
  "GDP_per_capita_2021" = gdpPC2021_data$"X2021")

#rename countries to match original dataframe country names
gdpPC2021_data <- gdpPC2021_data %>%
  transform(coded_country = case_when
          (
            coded_country == "Egypt, Arab Rep." ~ "Egypt",
            coded_country == "Venezuela, RB" ~ "Venezuela",
            coded_country == "Lao PDR" ~ "Laos",
            coded_country == "Iran, Islamic Rep." ~ "Iran",
            coded_country == "Kyrgyz Republic" ~ "Kyrgyzstan",
            coded_country == "Viet Nam" ~ "Vietnam",
            coded_country == "Brunei Darussalam" ~ "Brunei",
            coded_country == "Serbia" ~ "Republic of Serbia",
            coded_country == "Russian Federation" ~ "Russia",
            coded_country == "Turkiye" ~ "Turkey",
            coded_country == "Czechia" ~ "Czech Republic",
            coded_country == "Slovak Republic" ~ "Slovakia",
            coded_country == "Korea, Rep." ~ "South Korea",
            coded_country == "United States" ~ "United States of America",
            TRUE ~ coded_country
          )
  )

#add data to the clustering table
clustering_data <- merge(clustering_data, gdpPC2021_data, by = "coded_country")

#load unemployment rate data (worldbank)
unemployment2021_data <-read.csv("world bank unemployment.csv")

#rename columns
unemployment2021_data <- data.frame(
  "coded_country" = unemployment2021_data$"Country.Name",
  "unemployment rate 2021" = unemployment2021_data$"X2021")

#rename countries to match original dataframe country names
unemployment2021_data <- unemployment2021_data %>%
```

```
  transform(coded_country = case_when
          (
            coded_country == "Egypt, Arab Rep." ~ "Egypt",
            coded_country == "Venezuela, RB" ~ "Venezuela",
            coded_country == "Lao PDR" ~ "Laos",
            coded_country == "Iran, Islamic Rep." ~ "Iran",
            coded_country == "Kyrgyz Republic" ~ "Kyrgyzstan",
            coded_country == "Viet Nam" ~ "Vietnam",
            coded_country == "Brunei Darussalam" ~ "Brunei",
            coded_country == "Serbia" ~ "Republic of Serbia",
            coded_country == "Russian Federation" ~ "Russia",
            coded_country == "Turkiye" ~ "Turkey",
            coded_country == "Czechia" ~ "Czech Republic",
            coded_country == "Slovak Republic" ~ "Slovakia",
            coded_country == "Korea, Rep." ~ "South Korea",
            coded_country == "United States" ~ "United States of America",
            TRUE ~ coded_country
          )
  )

clustering_data <- merge(clustering_data, unemployment2021_data, by = "coded_country")

#load world happiness index data
happiness_data_2019 <- read.csv("World Happiness Index 2019.csv")

#rename columns
happiness_data_2019 <- data.frame(
  "coded_country" = happiness_data_2019$"Country.or.region",
  "happiness_score_2019" = happiness_data_2019$"Score")

#rename countries to match existing ones in data
happiness_data_2019 <- happiness_data_2019 %>%
  transform(coded_country = case_when
          (
            coded_country == "United States" ~ "United States of America",
            coded_country == "Trinidad & Tobago" ~ "Trinidad and Tobago",
            coded_country == "Serbia" ~ "Republic of Serbia",
            TRUE ~ coded_country
          )
  )

#add to clustering table
clustering_data <- merge(clustering_data, happiness_data_2019, by = "coded_country")

#load birth rate per 1000 data (worldbank)
birthRate_per_1000 <- read.csv("birthRate per 1000.csv")

#rename columns
birthRate_per_1000 <- data.frame(
  "coded_country" = birthRate_per_1000$"Country.Name",
  "birth_rate_per_1000_2021" = birthRate_per_1000$"X2021")
```

```
#rename countries to match original dataframe country names
birthRate_per_1000 <- birthRate_per_1000 %>%
  transform(coded_country = case_when
          (
            coded_country == "Egypt, Arab Rep." ~ "Egypt",
            coded_country == "Venezuela, RB" ~ "Venezuela",
            coded_country == "Lao PDR" ~ "Laos",
            coded_country == "Iran, Islamic Rep." ~ "Iran",
            coded_country == "Kyrgyz Republic" ~ "Kyrgyzstan",
            coded_country == "Viet Nam" ~ "Vietnam",
            coded_country == "Brunei Darussalam" ~ "Brunei",
            coded_country == "Serbia" ~ "Republic of Serbia",
            coded_country == "Russian Federation" ~ "Russia",
            coded_country == "Turkiye" ~ "Turkey",
            coded_country == "Czechia" ~ "Czech Republic",
            coded_country == "Slovak Republic" ~ "Slovakia",
            coded_country == "Korea, Rep." ~ "South Korea",
            coded_country == "United States" ~ "United States of America",
            TRUE ~ coded_country
          )
  )

#add to clustering table
clustering_data <- merge(clustering_data, birthRate_per_1000, by = "coded_country")

#load birth rate per 1000 data (worldbank)
press_freedom_2021_data <- read.csv("press-freedom-index-rsf.csv")

#take only 2021 data
press_freedom_2021_data <- press_freedom_2021_data %>%
  filter(Year == 2021)

#remove year and code columns
press_freedom_2021_data$Year <- NULL
press_freedom_2021_data$Code <- NULL

#rename columns
press_freedom_2021_data <- data.frame(
  "coded_country" = press_freedom_2021_data$"Entity",
  "press_freedom_score_2021" = press_freedom_2021_data$"Press.Freedom.Score")

#rename countries to match original dataframe country names
press_freedom_2021_data <- press_freedom_2021_data %>%
  transform(coded_country = case_when
          (
            coded_country == "Serbia" ~ "Republic of Serbia",
            coded_country == "Czechia" ~ "Czech Republic",
            coded_country == "United States" ~ "United States of America",
            TRUE ~ coded_country
          )
```

```
  )

#add to clustering table
clustering_data <- merge(clustering_data, press_freedom_2021_data, by = "coded_country")

#load corruption perception data, and trim whitespace (why is this dataset like this lol)
corruption_perception_2021_data <- read.csv("corruption_data.csv", strip.white=TRUE)

#rename columns
corruption_perception_2021_data <- data.frame(
  "coded_country" = corruption_perception_2021_data$"region_name",
  "CPI_score_2021" = corruption_perception_2021_data$"X2021")

#rename countries
corruption_perception_2021_data <- corruption_perception_2021_data %>%
  transform(coded_country = case_when
          (
            coded_country == "Serbia" ~ "Republic of Serbia",
            coded_country == "United States" ~ "United States of America",
            TRUE ~ coded_country
          )
  )

#add to clustering table
clustering_data <- merge(clustering_data, corruption_perception_2021_data, by = "coded_country")

##############################################################
#Hierarchical Clustering
#for the marker:
#If it is possible I will include all relevant csv files in a zip in the submission,
#otherwise please save the table of values and load it as clustering_data
##############################################################

#

#omit NA Values
clustering_data <- na.omit(clustering_data)

#convert country column into rownames
rownames(clustering_data) <- clustering_data[,1]

#remove country column
clustering_data <- clustering_data[, -1]

#normalise values
clustering_data <- as.data.frame(scale(clustering_data))

#create distance matrix
distance_matrix <- dist(clustering_data, method = "euclidean")

#create cluster
```

```r
cluster <- hclust(distance_matrix, method = "average")

#create dendrogram object
dendrogram <- as.dendrogram(cluster)

#customize the dendrogram
dendrogram <- dendrogram %>%
  color_branches(k = 12) %>%
  set("branches_lwd", 2)

#adjust margins so labels are not cutoff
par(mar = c(5, 4, 4, 2))

#plot the dendrogram
circlize_dendrogram(dendrogram, labels_track_height = 0.3)
```

References

Birth rate, crude (per 1,000 people). World Bank Open Data. (n.d.-a). https://data.worldbank.org/indicator/SP.DYN.CBRT.IN

Circular dendrogram in R with circlize and Dendexten. R CHARTS | A collection of charts and graphs made with the R programming language. (2021, March 27). https://r-charts.com/part-whole/circular-dendrogram/

GDP per capita (current US$). World Bank Open Data. (n.d.-b). https://data.worldbank.org/indicator/ny.gdp.pcap.cd

GHS index report and Data. GHS Index. (2023, December 19). https://ghsindex.org/report-model/

Hierarchical Cluster Analysis. Hierarchical Cluster Analysis · UC Business Analytics R Programming Guide. (n.d.). https://uc-r.github.io/hc_clustering

How can I trim leading and trailing white space?. Stack Overflow. (1955, November 1). https://stackoverflow.com/questions/2261079/how-can-i-trim-leading-and-trailing-white-space

Network, S. D. S. (2019, November 27). World happiness report. Kaggle. https://www.kaggle.com/datasets/unsdsn/world-happiness?resource=download

Press freedom index. Our World in Data. (n.d.). https://ourworldindata.org/grapher/press-freedom-index-rsf

R is plotting labels off the page. Stack Overflow. (1956, February 1). https://stackoverflow.com/questions/2807060/r-is-plotting-labels-off-the-page

Tyagi, M. (2022, September 4). World corruption index dataset. Kaggle. https://www.kaggle.com/datasets/tr1gg3rtrash/global-corruption-index

Unemployment, total (% of total labor force) (national estimate). World Bank Open Data. (n.d.-c). https://data.worldbank.org/indicator/SL.UEM.TOTL.NE.ZS?most_recent_year_desc=true

[R] hclust - Error in if (is.na(n) || n > 65536L | Kaggle. (n.d.). Www.kaggle.com. Retrieved April 16, 2024, from https://www.kaggle.com/discussions/questions-and-answers/43142

11.7 Bar Chart Dodge with Labels | Artificial Intelligence Desktop Survival Guide. (n.d.). In onepager.togaware.com. Retrieved April 16, 2024, from https://onepager.togaware.com/bar-chart-dodge-with-labels.html

A general vectorised if — case_when. (n.d.). Dplyr.tidyverse.org. https://dplyr.tidyverse.org/reference/case_when.html

Chanenson, G. J., Chanenson, G. J., & Chanenson, G. J. (2021, June 23). On Likert Scales In R. Jake Chanenson. https://jakec007.github.io/2021-06-23-R-likert/

Controlling Layouts. (n.d.). Patchwork.data-Imaginist.com. Retrieved April 16, 2024, from https://patchwork.data-imaginist.com/articles/guides/layout.html

Convert the values in a column into row names in an existing data frame. (n.d.). Stack Overflow. Retrieved April 16, 2024, from https://stackoverflow.com/questions/5555408/convert-the-values-in-a-column-into-row-names-in-an-existing-data-frame

Copy an R data.frame to an Excel spreadsheet. (n.d.). Stack Overflow. Retrieved April 16, 2024, from https://stackoverflow.com/questions/24704344/copy-an-r-data-frame-to-an-excel-spreadsheet

CRAN: Manuals. (2019). R-Project.org. https://cran.r-project.org/manuals.html

Differences between the base R and magrittr pipes. (n.d.). Www.tidyverse.org. Retrieved April 16, 2024, from https://www.tidyverse.org/blog/2023/04/base-vs-magrittr-pipe/

factor function - RDocumentation. (n.d.). Www.rdocumentation.org. https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/factor

Labelling circular dendextend dendrogram. (n.d.). Stack Overflow. Retrieved April 16, 2024, from https://stackoverflow.com/questions/46238364/labelling-circular-dendextend-dendrogram

Likert grouping with different levels in R. (n.d.). Stack Overflow. Retrieved April 16, 2024, from https://stackoverflow.com/questions/27061286/likert-grouping-with-different-levels-in-r

Muenchen, B. (2013, January 22). Comparing Transformation Styles: attach, transform, mutate and within | R-bloggers. https://www.r-bloggers.com/2013/01/comparing-transformation-styles-attach-transform-mutate-and-within/

par function | R Documentation. (n.d.). Www.rdocumentation.org. https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/par

Pivoting. (n.d.). Tidyr.tidyverse.org. https://tidyr.tidyverse.org/articles/pivot.html

r - What do hjust and vjust do when making a plot using ggplot? (n.d.). Stack Overflow. https://stackoverflow.com/questions/7263849/what-do-hjust-and-vjust-do-when-making-a-plot-using-ggplot

R colour Cheatsheet. (n.d.). https://www.nceas.ucsb.edu/sites/default/files/2020-04/colorPaletteCheatsheet.pdf

Return rows with matching conditions — filter. (2020). Tidyverse.org. https://dplyr.tidyverse.org/reference/filter.html

Summarise each group to fewer rows — summarise. (n.d.). Dplyr.tidyverse.org. https://dplyr.tidyverse.org/reference/summarise.html

Wikipedia Contributors. (2019, May 15). Corruption Perceptions Index. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Corruption_Perceptions_Index