Analysis on Institutional Confidence: A Focus on the Czech Republic in Comparative Perspective

Mary Grace Paramitha Saraswati
(Student ID: 33410712)

Table of Contents

Part 1: Descriptive Analysis

1. **Introduction**
   This report explores the factors influencing trust in institutions, comparing various predictor variables. Towards the end a regression analysis is done to compare the trend in the Czech Republic to other countries and a cluster of socioeconomically similar nations. Generative AI is used in certain cases in this report, more details on Appendix 6.

2. **Data overall**
   The data describes the confidence level of the people from 66 countries world wide on social organisations, such as Parliament and Parties. The survey is conducted in 2017-2022. The individual sample created has 50000 rows and 40 variables. The target variables are the confidence level for each organisation and the predictors are the social factors such as trust, happiness, employment, age, and education shown in Appendix 1.1. The complete interpretation of values can be found in Appendix 1.1.

3. **Missing / negative values**
   The negative values (-1, -2, -4, -5) are considered as unique values. First, we decided that -4 (not asked) and -5 (missing) would not be used in the analysis as it would not contribute meaningfully. However, for the 'MF' column, we change value '-5' into 'Others' to represent the other gender categories and this value is not removed.

   Second, for the -2 (not answered) values, we do a check on whether it is random, or if there is a trend in refusal to answer. The result shows that all -2 answers are random occurrences as it occurs for less than 5% in all variables, in both the global and the focus country data (see Appendix 1.2). Hence, we deleted the -2 values.

   Third, for the -1 (not answered) values, we keep it in the dataset as this might show public uncertainty in certain aspects. We also delete the -1 values from the MF, Age, Education, and Employment columns. However, these values are not used in the regression.

4. **Special cases**
   - **Mutually exclusiveness:**
     For variables PIA and PIAB, we only use the responses which are mutually exclusive. We exclude the non-mutually exclusive results so that it will not contradict the ranking intent and ensures that each priority has a unique meaning.

   - **One-hot encoding:**
     The data in the variables PIA, PIAB, Employment, and MF are not categorical ordinal data, hence before the regression process, we did a one-hot encoding to those variables, creating new columns based on the responses and assigning binary numbers to them.

5. **Key findings**
   Generally, we use median to figure out the central tendency, and mode for the most frequent answer. Full descriptive diagrams and tables can be found in Appendix 1.3.
   - **Demographics:**
     It is found that there are 16628 females, 16046 males, and 6 categorised as Others. When compared with age, the respondents tend to be in the mid 20s to 30s with the average being 42, due to several numbers of old respondents up to 103 years old. While the age distribution by gender is approximately similar, however there are slightly more elderly males than females. Overall, respondents mostly have their Upper secondary education completed and work full-time.
   - **Confidence:**
     Overall, the data shows that the respondents have quite a lot of confidence in social organisations, while most people tend to have less trust in the Election, Union, Parties, and Parliament, shown by median and mode of 3.

- **Interpersonal Trust:**
  People tend to have different levels of trust in different kinds of people. They tend to entirely trust their family members and do not trust people they just met. However, there is a variability on how much people trust people they just met, where some fairly trust and some do not at all, shown by the interquartile range in the box plot.
- **Importance in life:**
  People tend to believe that family, leisure, and work are on average important to their life.
- **Well-being:**
  Most of the respondents are moderately happy with their overall condition. While most of the respondents are completely satisfied with the freedom of choice (mode=10), the tendency is that on average people are just moderately satisfied (median=7), which means some people are less satisfied than others.
- **Economic:**
  People tend to believe that the income they obtain needs to differ based on their hard work, and believe that success is due to hard work instead of luck. While the previous answers show the importance of individual hard work, most people believe that the government should take more responsibility to ensure that everyone is provided for. The central tendency for this aspect, however, stays in the middle, showing variations in people's opinion.
- **Priority:**
  Most people stated that a high level of economic growth is the important aspect a country needs to prioritise, followed by people having more say about how things are done at their jobs and in their communities.
- **Trust in Science:**
  People tend to agree that science and technology makes life easier with most of them completely agreeing.
- **Information Source:**
  On average, the internet and social media have become a popular daily source of information. On the other hand, people rarely receive information through email and radio.
- **Thoughts on Democracy:**
  People tend to agree that it is important to live in a democratically-governed country, however they are less satisfied with the current government's democracy.

Part 2: Focus Country vs Other Countries as a Group

1. Comparison of respondents answer in focus country vs all countries

In this analysis, the Czech Republic is the focus country. Results show that responses from the Czech Republic differ from those of the broader group of countries. This section will reveal unique insights into the country in terms of socio-economic and institutional trust, which will be further used in further analysis in other sections.

In the Czech Republic, the majority of respondents have attained either lower secondary or upper secondary education, while respondents from other countries are more likely to have attained upper secondary or bachelor's degrees as their highest level of education. More than 50% of respondents in the Czech Republic report being employed full-time. This is higher than the global average, where employment statuses are more distributed across part-time, self-employment, and students. There is stronger confidence in providing education for one's child without worry in the Czech Republic, which is completely the other way around in other countries as a group.

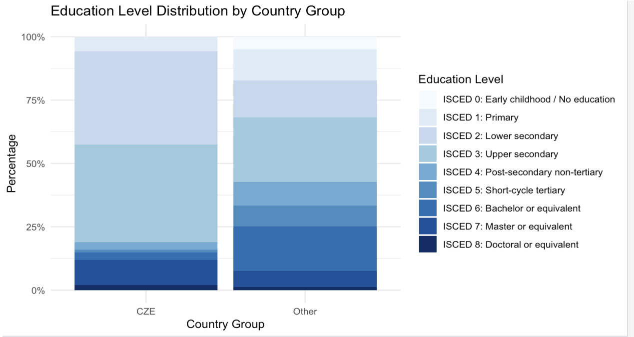


Image 2.1: Education Level Distribution Stacked Barplot

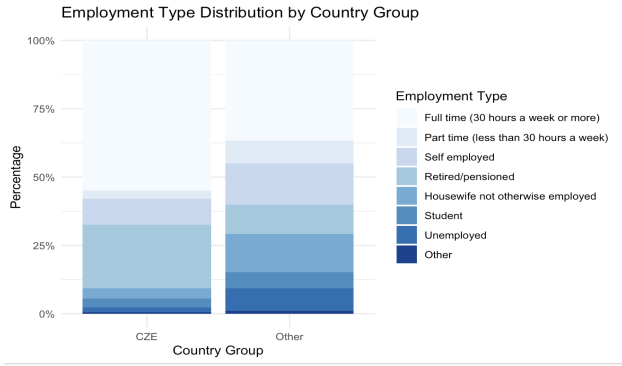


Image 2.2: Employment Type Distribution Stacked Barplot

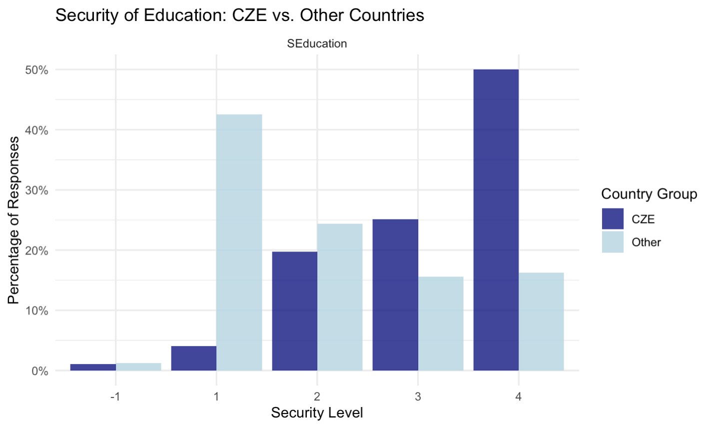


Image 2.3: Education Security Distribution Histogram

Compared to the respondents worldwide, the respondents in the Czech Republic are largely satisfied with their freedom of choice, although a very small amount of abstain is found (see Appendix 1.4). This positive trend correlates with a fair level of satisfaction with the current state of democracy, suggesting a feedback loop. When people experience more personal freedom, it may increase democratic trust, and a functioning democracy may preserve or promote that freedom. This is proved by fairly high satisfaction towards the current democracy of the country.

Czechs also show more trust in social organisations such as labor unions and election organisations than the global average as shown below. This might be perceived as higher transparency and trust in general in Czech, explaining the fact of higher trust on acquaintances and people they just met compared to other countries. Another interesting insight is, unlike global respondents who largely do not consume radio, Czech respondents still frequently use radio as a source of information alongside the internet (more details see Appendix 1.4).

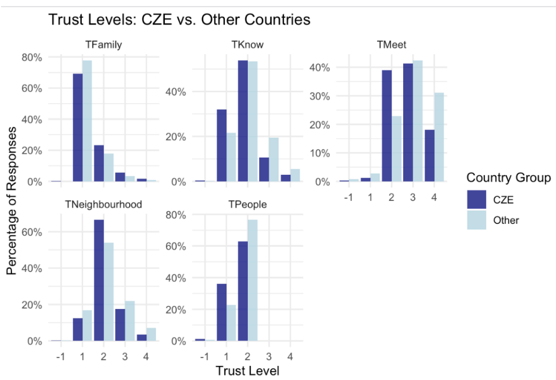


Image 2.4: Trust Level Distribution Histogram

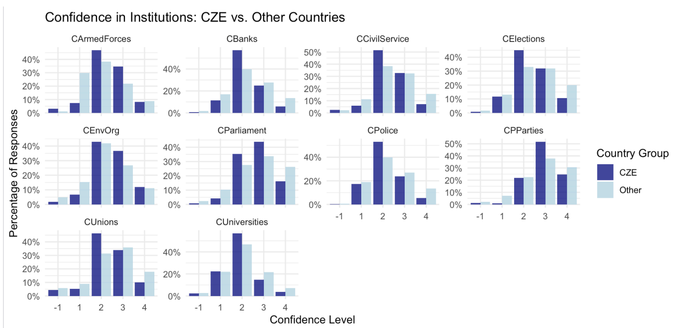


Image 2.5: Confidence Level Distribution Histogram

2. Evaluation on how well participant responses predict the confidence level in the focus country (CZE)

After discovering the key differences between the survey responses in the focus country and the other countries, the next step is to explore whether these distinguishing characteristics are better in predicting respondents' confidence in institutions within the focus country compared to the other countries. We are using regression analysis.

The R-squared values for all target variables (CArmedForces, CUnions, CPolice, CPParties, CParliament, CCivilService, CUniversities, CElections, CBanks, CEnvOrg) is low, with the value from 0.232 to 0.364. This indicates that the predictor variables combined only explain around 23.2% to 36.4% of the confidence level variables. This means that there are other variables that need to be taken into account for a better analysis of the response and the confidence level. The Q-Q plot shows that the fit is good, except some minor negligible deviations (see Appendix 2.1). The table below shows the main summary of the regression analysis.

	R^2	Adjusted R^2	Residual Std. Error (df = 484)	F Statistic (df = 12; 484)
CArmedForces	0.309	0.250	0.655	5.244^{***}
CUnions	0.284	0.223	0.669	4.652^{***}
CPolice	0.315	0.257	0.655	5.400^{***}
CPParties	0.233	0.167	0.662	3.564^{***}
CParliament	0.267	0.205	0.696	4.285^{***}
CCivilService	0.292	0.231	0.635	4.838^{***}
CUniversities	0.308	0.249	0.642	5.233^{***}
CElections	0.364	0.310	0.683	6.715^{***}
CBanks	0.245	0.181	0.655	3.813^{***}
CEnvOrg	0.232	0.167	0.731	3.557^{***}

Table 2.1: Regression result of focus country model

Next, I made a heatmap of p-values of the predictor variables towards the target variables. The lower the p-value means the higher the significance and less likely the occurrence happens randomly. Based on the stargazer table and heatmap, the best predictors overall are the TMeet and PRadio. These are the variables showing the most significant relationships in the p-value of 1% (shown by ***). This means, the more a person trusts other people they meet, also the more often they obtain information from the radio, the more trust they have in social organisations. This can be translated to the more trust a person can put into another person, the more trust they can put on the social organisations. However, the value differs for each confidence variable, indicating that a certain variable might be good for one confidence variable but not the other. Emp_student is notably the highest indicator for CUniversity with a coefficient of >1.000, much higher than average coefficients. Table 2.2 shows the complete list of best predictors for each institution in the confidence variable.

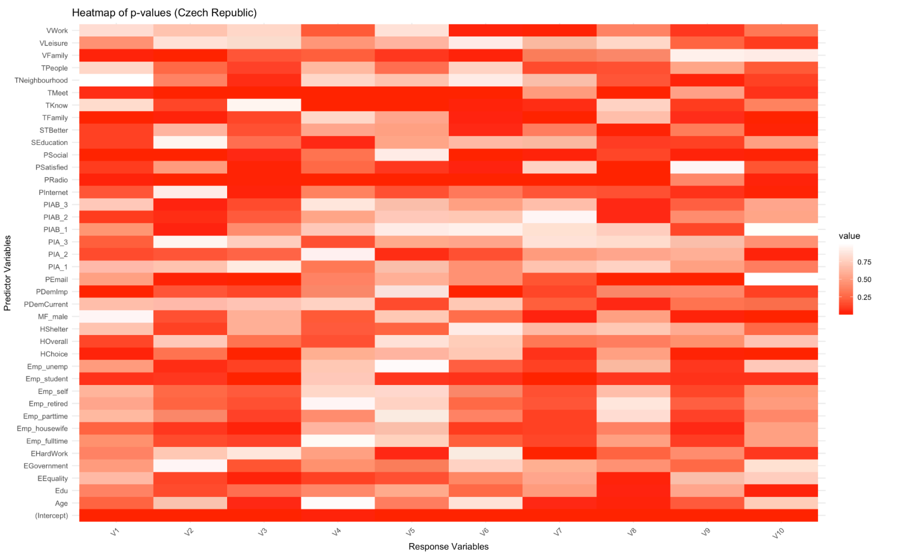


Image 2.5: Heatmap of focus country model

Institutions	Best predictors
CArmedForces	PDemImp, PSocial, TFamily
CUnions	PRadio, TMeet, PSocial
CPolice	PEmail, TMeet, HChoice
CPParties	TMeet, PRadio, TKnow
CParliament	PRadio, TKnow, TMeet
CCivilService	TMeet, PRadio, PSocial
CUniversities	PRadio, TFamily, Emp_student
CElections	PRadio, TMeet, PSatisfied
CBanks	PEmail, HChoice, TNeighbourhood
CEnvOrg	MF_male, TFamily, HChoice

Table 2.2: Best predictor per confidence variables of focus country model

We now do another regression on the focus country data, but with the best predictor variables from the previous analysis. Here is the main summary.

	R^2	Adjusted R^2	Residual Std. Error (df = 484)	F Statistic (df = 12; 484)
CArmedForces	0.194	0.169	0.689	7.758^{***}
CUnions	0.156	0.13	0.708	5.944^{***}
CPolice	0.197	0.172	0.691	7.897^{***}
CPParties	0.164	0.138	0.674	6.289^{***}
CParliament	0.187	0.161	0.715	7.377^{***}
CCivilService	0.213	0.189	0.653	8.700^{***}
CUniversities	0.217	0.192	0.666	8.885^{***}
CElections	0.244	0.22	0.726	10.343^{***}
CBanks	0.146	0.12	0.679	5.503^{***}
CEnvOrg	0.105	0.077	0.770	3.757^{***}

Table 2.3: Regression result of focus country model with the best predictors.

It is surprising that the new model has lower R^2 value, which means it does not do a better job in explaining a larger proportion of the variability in the dependent variable. However, the second model has greater F statistics value which shows that on average the model is statistically useful and better explains that the predictor variables have relation to the dependent variables.

3. Evaluation on how well participant responses predict the confidence level in the other countries

Similar to the focus country data, the model has a low R^2. However, the R^2 of the model with the other countries data is much lower than the focus country's, which means this model has much less explanatory power. The R^2 values for all target variables are low, ranging from 0.090 to 0.246, which means the predictor variables combined only can explain a maximum 24.6% of the confidence level variables. As the model did not do a great job in explaining the variability of the dependent variables, this means that there are other predictor variables that need to be taken into account for a better analysis of the response and the confidence level, similar to the focus country data. The residual std.error value is also significantly more than the focus country model, indicating that this model does not do well in predicting the values. However, the F statistics are much larger in this model, indicating that the model is overall statistically significant, in $p<0.001$, in a more diverse sample. The Q-Q plot also shows that the model has a good fit with the plots following the 45 degrees line, similar to the previous model, except some minor deviations. Table 2.4 shows the main summary of the regression analysis.

	R^2	Adjusted R^2	Residual Std. Error (df = 484)	F Statistic (df = 12; 484)
CArmedForces	0.133	0.132	0.865	123.892^{***}
CUnions	0.122	0.121	0.824	111.697^{***}
CPolice	0.160	0.159	0.854	153.927^{***}
CPParties	0.216	0.215	0.804	222.097^{***}
CParliament	0.246	0.245	0.830	262.414^{***}
CCivilService	0.183	0.182	0.807	180.636^{***}
CUniversities	0.105	0.104	0.794	94.613^{***}
CElections	0.207	0.206	0.843	210.581^{***}
CBanks	0.120	0.119	0.858	110.275^{***}
CEnvOrg	0.090	0.089	0.837	79.697^{***}

Table 2.4: Regression result of other countries' model

The heatmap shows that more variables are significant in the 1% levels. The best predictors overall are the PSatisfied, TPeople, TFamily, TKnow, TNeighborhood, and PDemCurrent. These are the variables showing the most significant relationships in the p-value of 1%. Compared to the focus country models, the best predictor variables are fairly different. First, the best predictors include all interpersonal variables, but TMeet, indicating trust to closer people plays more significance here, which is the other way around in the focus country model. Second, there are some significant variables, such as PDemCurrent and PDemSatisfied, which are not the top predictors in the focus country model. This can be translated as the more a person is happy about the current democratic state of their country, the more likely they have confidence in the social organisation. Besides the overall best predictors, there are some priority variables that have a high explanatory level to a certain confidence variable, which are not found in the focus country model, which are PIA_1, PIA_2, and PIA_3 (more info on the meaning of the variables in Appendix 1.1). Also, while Emp_student is notably the highest indicator for CUniversity in the focus country model, it is not included as the best predictor in the other countries' model. The table below shows the complete list of best predictors for each institution in the confidence variable.

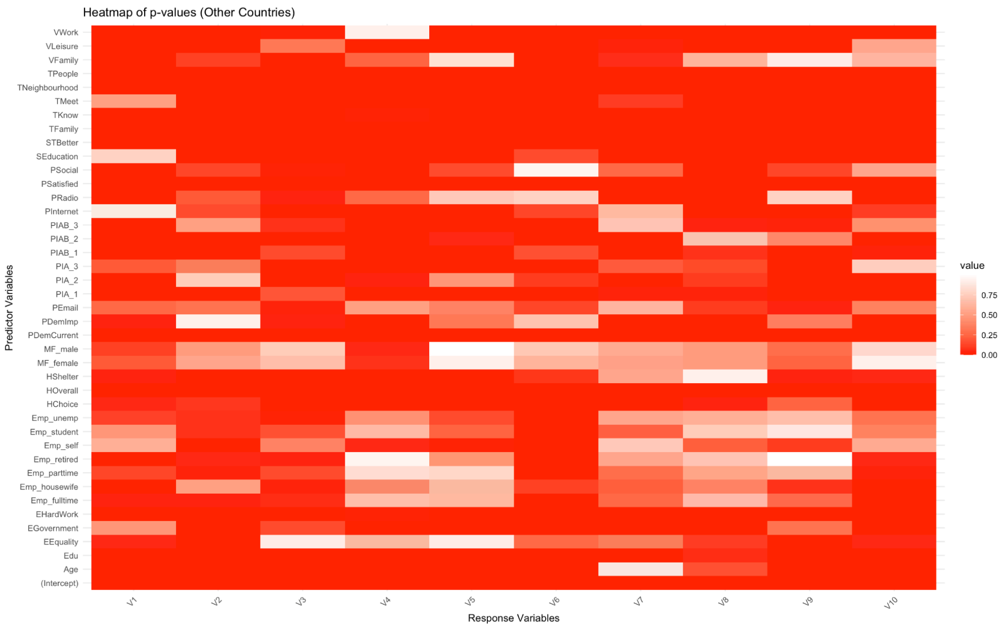


Image 2.6: Heatmap of other countries model

Institutions	Best predictors
CArmedForces	PIA_2, PIAB_2, Emp_housewife
CUnions	PSatisfied, TNeighbourhood, Emp_self
CPolice	TPeople, TFamily, TNeighbourhood
CPParties	PIA_3, TNeighbourhood, TPeople
CParliament	PSatisfied, PDemCurrent, TNeighbourhood
CCivilService	PSatisfied, TNeighbourhood, PDemCurrent
CUniversities	TKnow, TFamily, HOverall

CElections	PSatisfied, PDemCurrent, TNeighbourhood
CBanks	PIA_1, PIA_2, HOverall
CEnvOrg	TKnow, PIA_2, Emp_housewife

Table 2.5: Best predictor per confidence variables of the other countries model

Similar to the focus country model, we now do another regression with only the best predictor variables from the previous analysis. Here is the main summary.

	R^2	Adjusted R^2	Residual Std. Error (df = 484)	F Statistic (df = 12; 484)
CArmedForces	0.119	0.118	0.872	289.164***
CUnions	0.107	0.106	0.831	256.299***
CPolice	0.151	0.151	0.858	383.088***
CPParties	0.201	0.201	0.811	541.987***
CParliament	0.230	0.230	0.838	642.049***
CCivilService	0.171	0.171	0.813	444.585***
CUniversities	0.086	0.085	0.802	201.454***
CElections	0.200	0.200	0.846	538.077***
CBanks	0.099	0.099	0.868	236.646***
CEnvOrg	0.07	0.069	0.846	161.594***

Table 2.6: Regression result of other countries' model with the best predictors.

Similar to the focus country model, the R^2 value of the other country model with the extracted variables is also lower. This is because when we use fewer variables, we are capturing less total variation, so R^2 value drops. However the coefficients of some of the variables in this model are higher than in the previous complete model. This suggests that removing less influential variables allows the stronger predictors to stand out, due to reduced multicollinearity and more significance of variance to the key variables.

In conclusion, the key similarity between the two models is that the reduced models do not significantly improve the R^2 values, explaining the trade-off between model simplicity and explanatory power. The main differences between the models are the best predictors. This clearly shows societal or political dynamics between the Czech Republic and the broader international context. Although both models are limited in explaining the confidence level, they offer insights into which factors affect the most in national and international context. As for the best model, if we are looking at which one explains more variance in the dependent variable, the focus country model is better. However, the other countries' models offer stronger overall model significance and individual predictors, due to higher F statistics and more variables significant in 1% level of confidence. Hence, no model is strictly better.

Part 3: Focus Country vs Cluster of Similar Countries

1. Selection of cluster of similar countries

After comparing the model for the focus country and the other countries as a group, the next step is to examine whether these patterns hold in countries that are more socially, economically, and politically similar to the focus country.

Trust in institutions is often shaped by equality and economic conditions of the country (Palmisano & Sacchi, 2024). Therefore, we selected a range of economic indicators to assess the cross-country similarity. In addition, we included measures that reflect the quality of governance, such as corruption perception and democracy index score, as key criteria for clustering to best find the similar countries (Ma et al., 2024). Below is the complete list of criteria:

- **Economic indicators:** GDP, unemployment rate, average annual income
- **Political indicators:** Democracy score, corruption cleanliness, freedom of expression
- **Social indicators:** Number of people completed tertiary education
- **Demographic indicators:** Total population

Before proceeding to the clustering, NA values are removed and outliers are checked with the boxplot method and then removed. Since the variables are measured on different scales (e.g., GDP in billions USD and corruption scores on a fixed scale), we need to do scaling to standardise the data and all variables contribute equally to the clustering algorithm. I used the K-means algorithm to do the clustering.

The first step is to find out the right number of clusters and I use the Elbow method to determine this. Figure 3.1 shows the graph generated from the Elbow method, which depicts that the optimal number of centroids is 6. Then we do the clustering process where 8 countries are found similar to the Czech Republic, which are: **Hungary, Greece, Peru, Columbia, Malaysia, Romania, Portugal, and Chile** (more details on Appendix 4.3). They tend to have moderately high average annual income, high corruption cleanliness index, moderately high democracy, and average population. They also have low correlation, indicating that all data are distinct, meaningful, and well-justified for clustering.

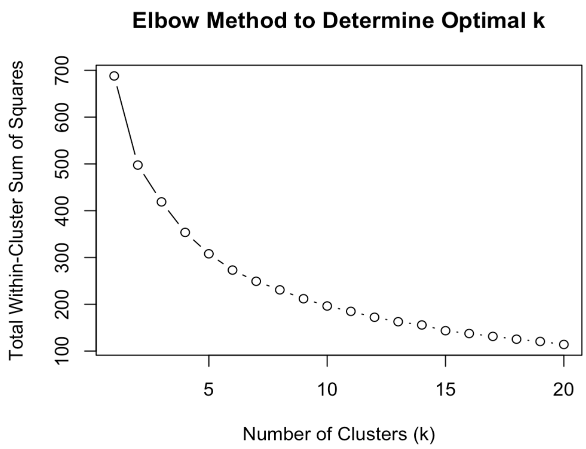


Figure 3.1: Elbow method for clustering

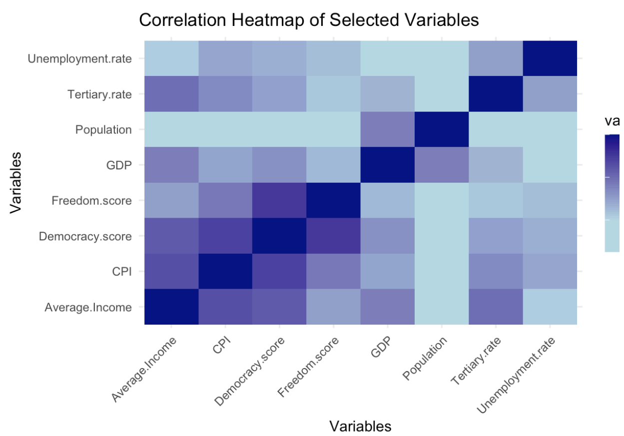


Figure 3.2: Correlation map of selected variables

To better visualise the clusters, we then plot the clusters using the Principal Component Analysis as shown in Figure 3.3.

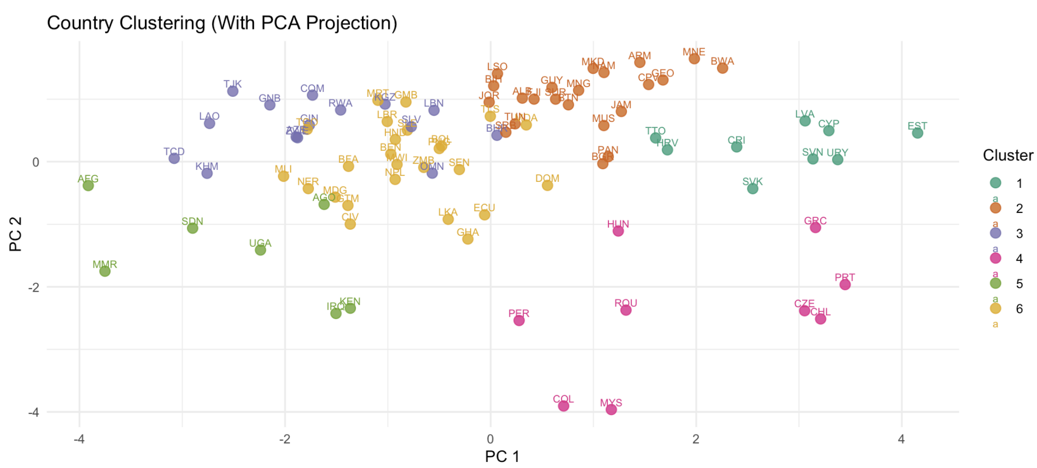


Figure 3.3: PCA scatter plot visualising the clusters

2. Evaluation on how well participant responses predict the confidence level in the other countries

We now do the regression analysis on the cluster of countries. The Q-Q plot shows that the fit is good. The R^2 values are lower than the focus country data model, however, it is higher than the other countries' model, except for the CArmedForce variable. The R^2 value ranges between 0.115 to 0.270, indicating that the predictor variables combined could only have an explanatory level of around 11.5% to 27.0% of the dependent variables. Although still lower than the focus country model, this model has a better explanatory power than the group of other countries' model on average. Although slightly higher than the focus country model, the F statistics are much lower than the other countries' model, stating that this model is less significant. There are also many more variables that have a high significance in $p<0.001$ in this model compared to the focus country model, but less than the other countries' model as seen in the stargazer table. The table below shows the main summary of the regression analysis.

	R^2	Adjusted R^2	Residual Std. Error (df = 484)	F Statistic (df = 12; 484)
CArmedForces	0.132	0.122	0.839	12.776^{***}
CUnions	0.186	0.177	0.754	19.234^{***}
CPolice	0.184	0.174	0.808	18.910^{***}
CPParties	0.195	0.186	0.684	20.369^{***}
CParliament	0.257	0.249	0.713	29.116^{***}
CCivilService	0.270	0.261	0.733	30.994^{***}
CUniversities	0.122	0.111	0.786	11.649^{***}
CElections	0.238	0.229	0.782	26.200^{***}
CBanks	0.154	0.144	0.814	15.331^{***}
CEnvOrg	0.115	0.104	0.832	10.862^{***}

Table 3.1: Regression result of cluster countries model

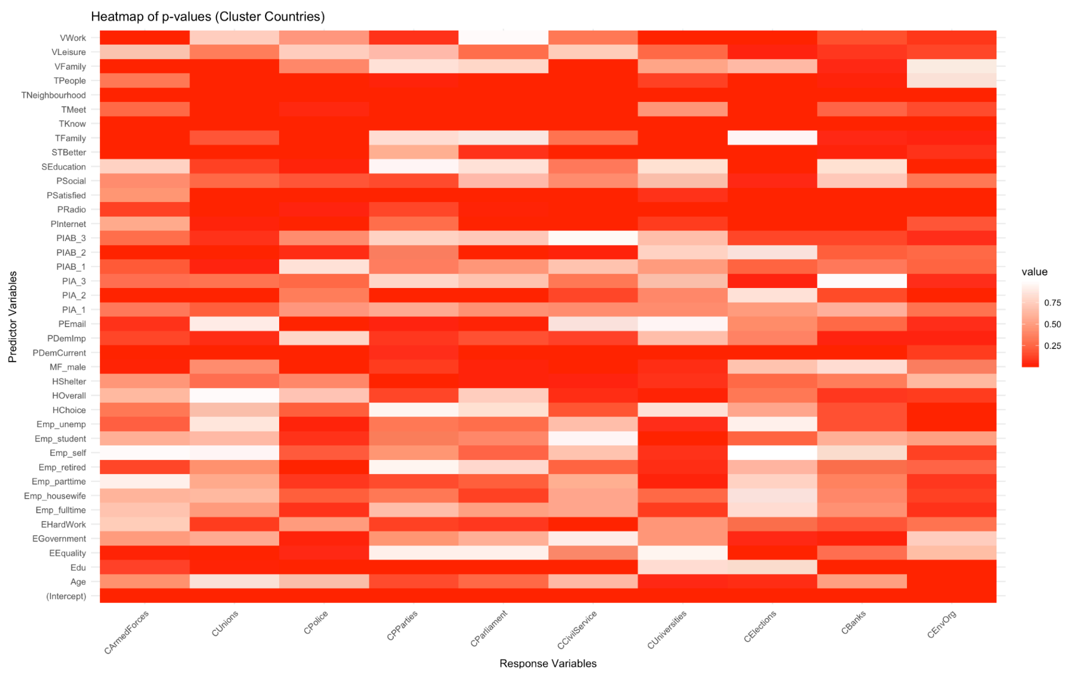


Figure 3.3: Heatmap of cluster of countries model

Based on the regression result, the best predictors overall are TNeighborhood, PSatisfied, PDemCurrent, and TKnow, shown by a fairly high coefficient in $p<0.001$ in all confidence variables. The list of best predictors variables resembles the top predictors in the other countries' model, instead of the focus country model. This might suggest that although the selected cluster of countries shares similar socio-economic and political situations with the focus country, the pattern of their institutional trust may be more influenced by broader, possibly global factors rather than country-specific factors. This shows that while clustering countries based on macro-level indicators may improve overall model fit and explanatory power, as well as provide a more contextual comparison than analysing all countries together, it does not guarantee similarity in specific-level predictors. However, there are some variables that are best predictors in the focus country model, and are present in the best predictors in the cluster country model, such as TMeet (see Table 3.2), indicating some overlaps in the pattern, possibly due to cultural and social similarities. For example, TMeet reflects how much individuals trust people they have just met, which can also represent a broader sense of interpersonal trust, which resembles the condition in the focus country model, but not other countries' model. This type of trust may translate into greater confidence in institutional trust as well. Moreover, the variables where TMeet is significant in the cluster model (CPParties and CParliament) also appear among the key predictors in the similar institutions for the focus country model. Interestingly, similar to the focus country model, Emp_student is one of the best predictors for CUniversity, which is not the case with the other countries' model. This suggests a consistent role of the strong predictors in the focus country across both contexts.

Institutions	Best predictors
CArmedForces	PIA_2, PIAB_2, TFamily
CUnions	TNeighbourhood, PSatisfied, PDemCurrent
CPolice	TNeighbourhood, PSatisfied, TKnow
CPParties	PSatisfied, TMeet, TNeighbourhood
CParliament	PIA_2, TMeet, TPeople
CCivilService	TPeople, TKnow, TNeighbourhood
CUniversities	TNeighbourhood, Emp_student, Emp_unemp
CElections	TNeighbourhood, TPeople, TMeet
CBanks	TNeighbourhood, TKnow, PEmail
CEnvOrg	TKnow, SEducation, Emp_unemp

Table 3.2: Best predictor per confidence variables of the cluster countries model

Part 4: Conclusion

The regression results discovered broad insights into the predictors of institutional trust. The cluster country model is the middle ground as it offers better performance than the other countries' model although underperforming compared to the focus country model in terms of R^2 value. However, it captures more significant variables at $p<0.001$ than the focus country model, indicating specific variables have stronger relationships with the dependent variables.

Between the other countries' model and the cluster countries model, the cluster model is more effective. While the other countries' model shows higher F-statistics and a greater number of highly significant predictor variables, the cluster model gives a better overall balance. In the other countries' model, the results come from a very mixed group of countries that may not share much in common with the Czech Republic. Hence, the insights might be too general or less useful for understanding institutional trust in the micro context. On the other hand, the cluster model has lower F-statistics value and fewer significant predictors, but higher R^2 than the other countries' model. This means it is doing a better job at explaining the variables that affect institutional trust. Also, strong predictor variables in the focus country model are also seen here, like TMeet and Emp_student, which shows that the model captures some shared social or cultural factors as well. The residual std. error value is also slightly less, showing that the model has better fit.

Overall, each model has strengths and limitations. The focus country model has the best explanatory power with high contextual relevance but it identifies fewer highly significant predictors, while other countries' model provides more significant predictors with a high F-statistics, but lacks contextual relevance. The cluster model provides a strong middle ground. It has a solid number of significant predictors and better explanatory power than other countries' models, also incorporating similarities with the Czech Republic.

REFERENCES

Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano J., M. Lagos, P. Norris, E. Ponarin & B. Puranen (eds.). 2022. World Values Survey: Round Seven – Country-Pooled Datafile Version 6.0. Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat. doi:10.14281/18241.24

Palmisano, F., & Sacchi, A. (2024). Trust in public institutions, inequality, and digital interaction: empirical evidence from European Union countries. *Journal of Macroeconomics*, *79*, 103582–103582. https://doi.org/10.1016/j.jmacro.2023.103582

Ma, Y., Ma, B., Yu, L., Ma, M., & Dong, Y. (2024). Perceived social fairness and trust in government serially mediate the effect of governance quality on subjective well-being. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-67124-4

APPENDIX

Appendix 1

1.1 Descriptive statistics:

1.1.a Below is the table which shows the data types of all variables:

Variable	Data_Type
Country	character
TPeople	integer
TFamily	integer
TNeighbourhood	integer
TKnow	integer
TMeet	integer
VFamily	integer
VLeisure	integer
VWork	integer
HOverall	integer
HChoice	integer
HShelter	integer
EEquality	integer
EGovernment	integer
EHardWork	integer
SEducation	integer
PIA	integer
PIAB	integer
STBetter	integer
PRadio	integer
PEmail	integer
PInternet	integer
PSocial	integer
PDemImp	integer
PDemCurrent	integer
PSatisfied	integer
MF	integer
Age	integer
Edu	integer
Employment	integer
CArmedForces	integer
CUnions	integer
CPolice	integer
CPParties	integer
CParliament	integer
CCivilService	integer
CUniversities	integer
CElections	integer
CBanks	integer
CEnvOrg	integer

1.1.b: Below is the table showing the target variables

Category	Variables	Values	Data Type
Confidence Level	CArmedForces CUnions CPolice CPParties CParliament CCivilService CUniversities CElections CBanks CEnvOrg	1.- A great deal 2.- Quite a lot 3.- Not very much 4.- None at all -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Not available	Integer

1.1.c: Below is the table showing the predictor variables:

Category	Variables	Values	Data Type
Interpersonal Trust	TPeople TFamily TNeighbourhood TKnow TMeet	1.- Trust completely 2.- Trust somewhat 3.- Do not trust very much 4.- Do not trust at all -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Not available	Integer
Importance in Life	VFamily VLeisure VWork	1.- Very important 2.- Rather important 3.- Not very important 4.- Not at all important -1-.- Don´t know -2-.- No answer -4-.- Not asked in this country -5-.- Missing; Not available	
Well-being	HOverall	1.- Very happy 2.- Quite happy 3.- Not very happy 4.- Not at all happy -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Not available.	
	HChoice	10.- A great deal 9.- 9 8.- 8 7.- 7 6.- 6 5.- 5 4.- 4 3.- 3 2.- 2 1.- None at all -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Unknown.	
	HShelter	1.- Often 2.- Sometimes 3.- Rarely 4.- Never -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Not available.	

Economic	EEquality	1.- Incomes more equal 2.- 2 3.- 3 4.- 4 5.- 5 6.- 6 7.- 7 8.- 8 9.- 9 10.- Larger income differences -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Unknown	
	EGovernment	1.- The government should take more responsibility to ensure that everyone is provided for 2.- 2 3.- 3 4.- 4 5.- 5 6.- 6 7.- 7 8.- 8 9.- 9 10.- People should take more responsibility to provide for themselves -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Not available.	
	EHardWork	1.- In the long run, hard work usually brings a better life 2.- 2 3.- 3 4.- 4 5.- 5 6.- 6 7.- 7 8.- 8 9.- 9 10.- Hard work doesn´t generally bring success - it´s more a matter of luck and connection -1-.- Don´t know -2-.- No answer -4-.- Not asked	

		-5-.- Missing; Not available.	
Education security	SEducation	1.- Very much 2.- A great deal 3.- Not much 4.- Not at all -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Not available	
Priority	PIA PIAB	1. A high level of economic growth. 2. Making sure this country has strong defence forces. 3. People have more say about how things are done at their jobs and in their communities. 4. Trying to make our cities and countryside more beautiful.	
Trust in Science & Technology	STBetter	1.- Completely disagree 2.- 2 3.- 3 4.- 4 5.- 5 6.- 6 7.- 7 8.- 8 9.- 9 10.- Completely agree -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Not available	
Information Source	PRadio PEmail PInternet PSocial	1.- Daily 2.- Weekly 3.- Monthly 4.- Less than monthly 5.- Never -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Not	

		available	
International organizations being effective or democratic	PDemImp PDemCurrent PSatisfied	1.- Not at all democratic 2.- 2 3.- 3 4.- 4 5.- 5 6.- 6 7.- 7 8.- 8 9.- 9 10.- Completely democratic -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Not available	
Gender	MF	1.- Male 2.- Female -2-.- No answer -4-.- Not asked -5-.- Missing; Not available; AU: Other	
Age	Age	-	
Education	Edu	0.- Early childhood education (ISCED 0) / no education 1.- Primary education (ISCED 1) 2.- Lower secondary education (ISCED 2) 3.- Upper secondary education (ISCED 3) 4.- Post-secondary non-tertiary education (ISCED 4) 5.- Short-cycle tertiary education (ISCED 5) 6.- Bachelor or equivalent (ISCED 6) 7.- Master or equivalent (ISCED 7) 8.- Doctoral or equivalent (ISCED 8) -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Not available	
Employment	Employment	1.- Full time (30 hours a week or more) 2.- Part time (less than 30 hours a week) 3.- Self	

		employed 4.- Retired/pensioned 5.- Housewife not otherwise employed 6.- Student 7.- Unemployed 8.- Other -1-.- Don´t know -2-.- No answer -4-.- Not asked -5-.- Missing; Not available	

1.2 Data cleaning:

1.2.a Below is the proportion of 'no answer' responses (value = -2) in the data overall.

```
> no_answer_proportion
      Country       TPeople       TFamily TNeighbourhood         TKnow         TMeet       VFamily      VLeisure         VWork
      0.00000       0.00290       0.00114        0.00196       0.00206       0.00246       0.00060       0.00192       0.00298
     HOverall       HChoice      HShelter      EEquality   EGovernment     EHardWork    SEducation           PIA          PIAB
      0.00228       0.00244       0.00278        0.00300       0.00308       0.00272       0.01388       0.00552       0.01536
     STBetter        PRadio        PEmail      PInternet       PSocial       PDemImp   PDemCurrent    PSatisfied            MF
      0.00254       0.00344       0.00452        0.00380       0.00314       0.00486       0.00622       0.00580       0.00044
          Age           Edu    Employment   CArmedForces       CUnions       CPolice      CPParties   CParliament CCivilService
      0.00202       0.00504       0.00804        0.00242       0.00626       0.00246       0.00490       0.00492       0.00432
 CUniversities    CElections        CBanks        CEnvOrg
      0.00420       0.00454       0.00388        0.00534
```

1.2.b Below is the proportion of 'no answer' responses (value = -2) in the focus country data.

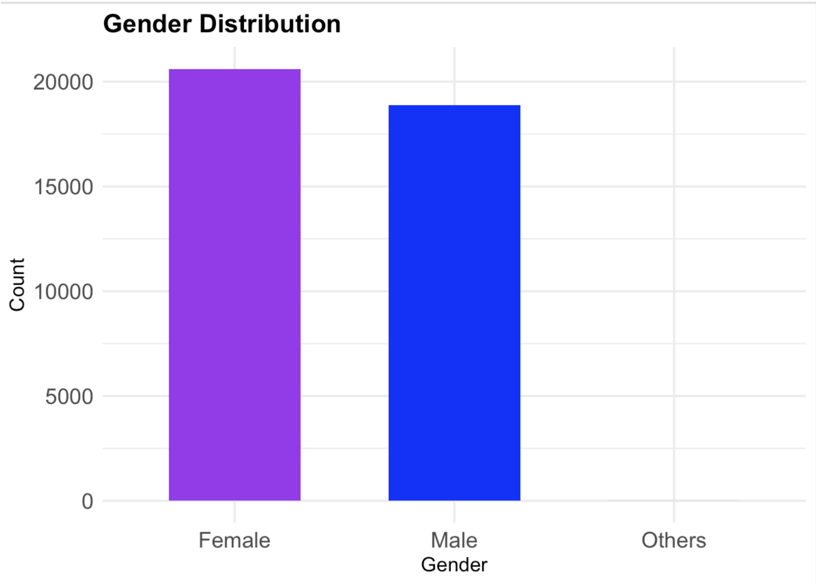
```
> no_answer_proportion
      Country       TPeople       TFamily TNeighbourhood        TKnow         TMeet       VFamily      VLeisure         VWork
  0.000000000   0.000000000   0.001677852    0.000000000  0.000000000   0.000000000   0.001677852   0.000000000   0.006711409
     HOverall       HChoice      HShelter      EEquality  EGovernment     EHardWork    SEducation           PIA          PIAB
  0.001677852   0.000000000   0.000000000    0.000000000  0.001677852   0.000000000   0.028523490   0.001677852   0.006711409
     STBetter        PRadio        PEmail      PInternet      PSocial       PDemImp   PDemCurrent    PSatisfied            MF
  0.000000000   0.000000000   0.000000000    0.001677852  0.000000000   0.000000000   0.001677852   0.000000000   0.000000000
          Age           Edu    Employment   CArmedForces      CUnions       CPolice     CPParties   CParliament CCivilService
  0.000000000   0.000000000   0.000000000    0.000000000  0.003355705   0.000000000   0.000000000   0.000000000   0.001677852
CUniversities    CElections        CBanks        CEnvOrg
  0.001677852   0.000000000   0.000000000    0.000000000
```

1.2.c Below is the number of 'don't know' responses (value = -1) in the data overall.

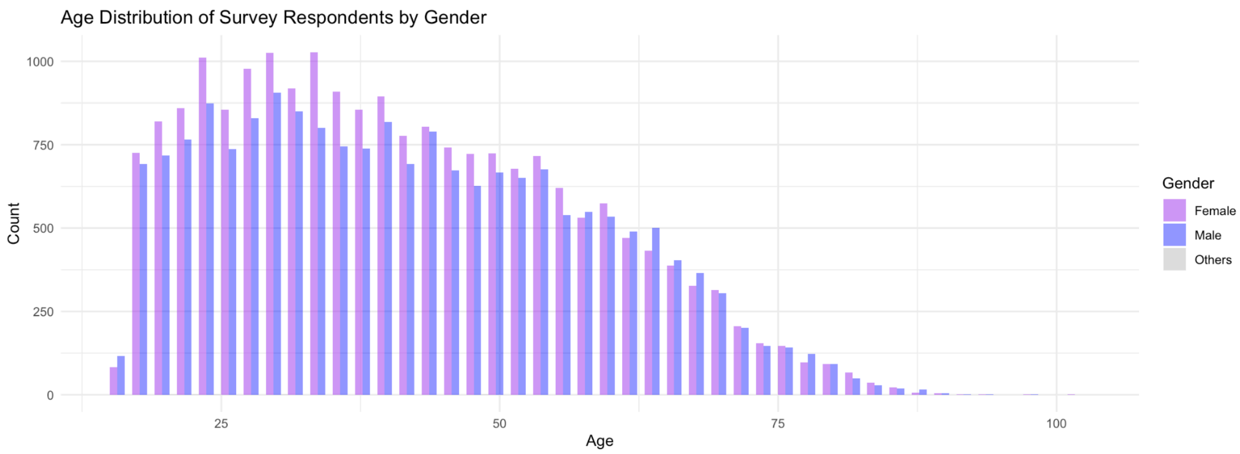
```
      Country       TPeople       TFamily TNeighbourhood         TKnow         TMeet
            0           233            18            128            74           286
      VFamily      VLeisure         VWork       HOverall       HChoice      HShelter
            6            60           120             93           155           103
    EEquality   EGovernment     EHardWork     SEducation           PIA          PIAB
          245           182             0            491             4           307
     STBetter        PRadio        PEmail      PInternet       PSocial       PDemImp
          219           170           331            276           280           312
  PDemCurrent    PSatisfied            MF            Age           Edu    Employment
          558           611             0              0             0             0
 CArmedForces       CUnions       CPolice      CPParties   CParliament CCivilService
          501          2232           256            827           893           855
CUniversities    CElections        CBanks        CEnvOrg
         1016           668           712           1902
```

1.3 Distribution of data after data cleansing:

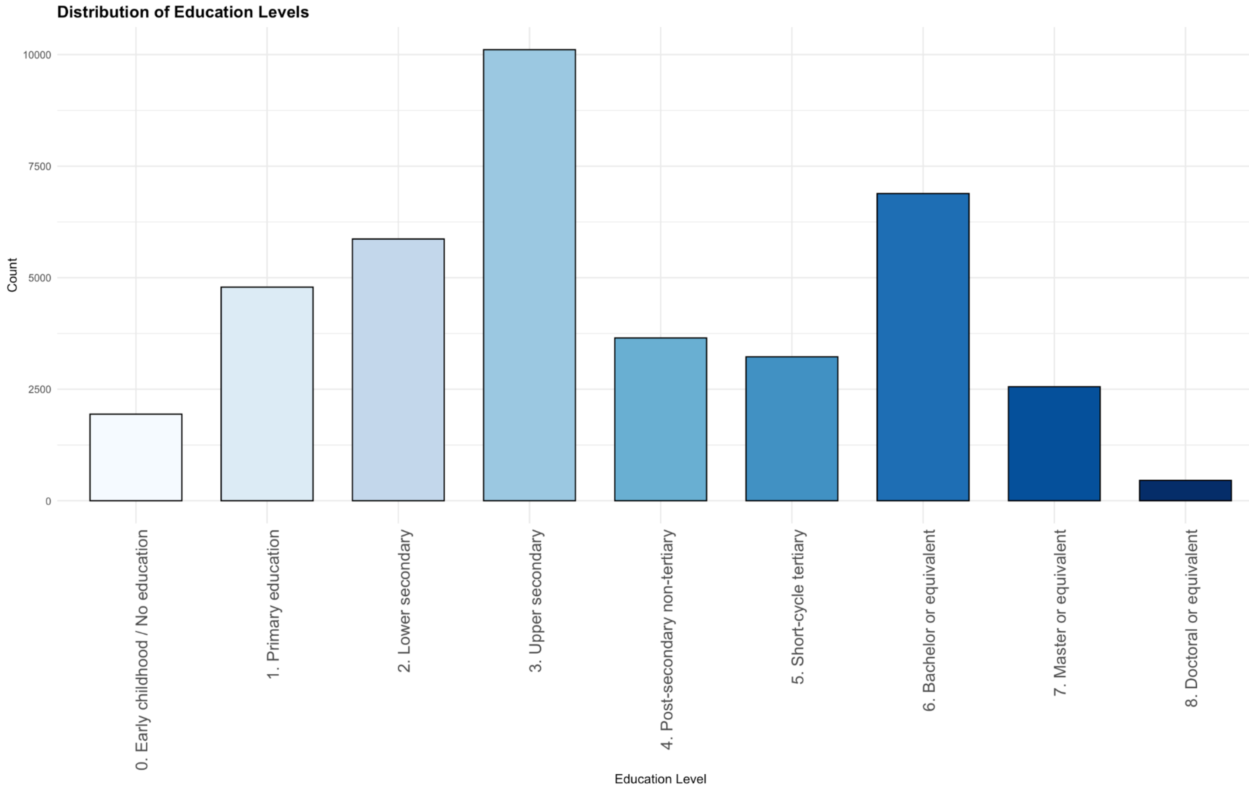
1.3.a Below is the comparison between each gender group (male, female, others).



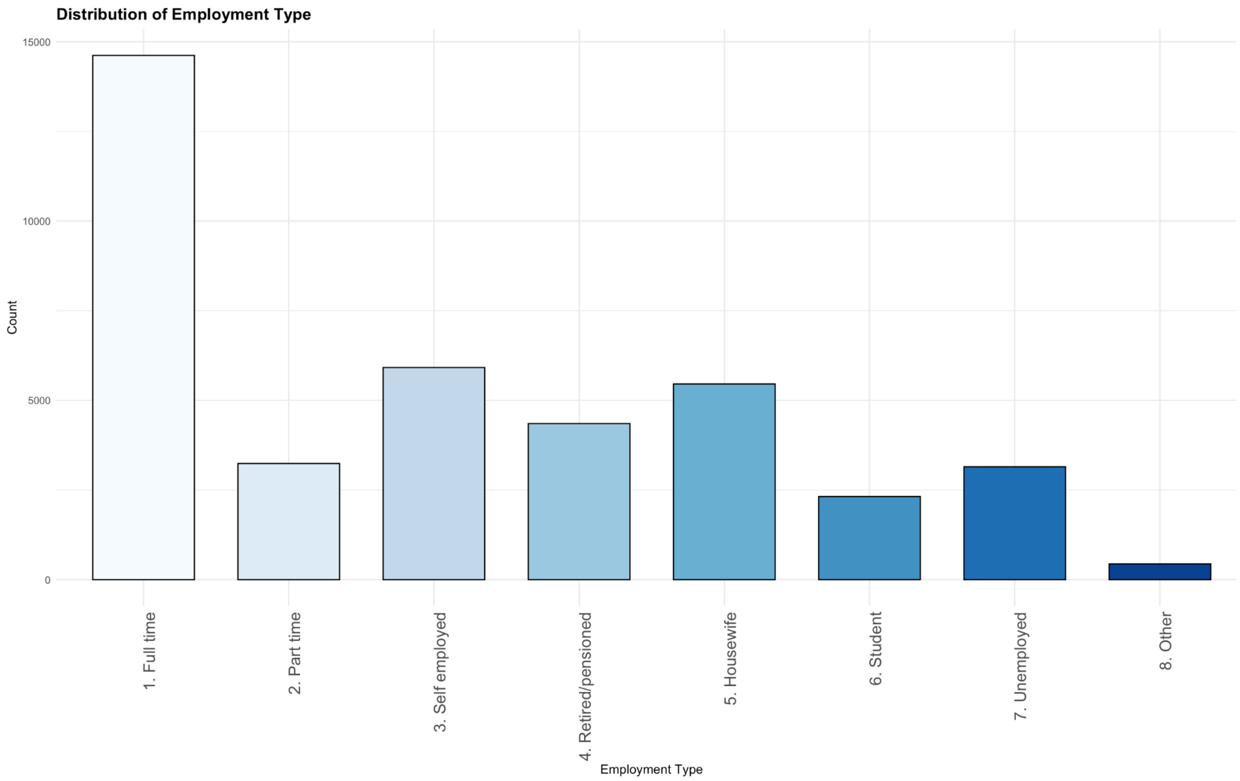
1.3.b Below is the distribution of the age, grouped by gender.



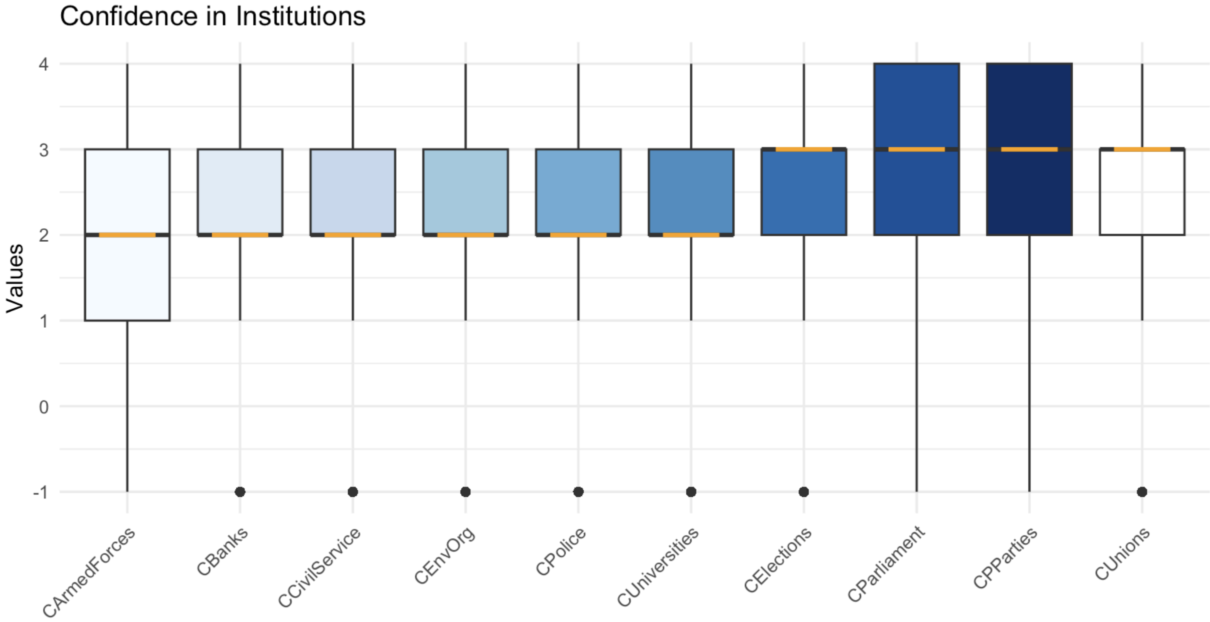
1.3.c Below is the distribution of the education levels



1.3.c Below is the distribution of the employment levels

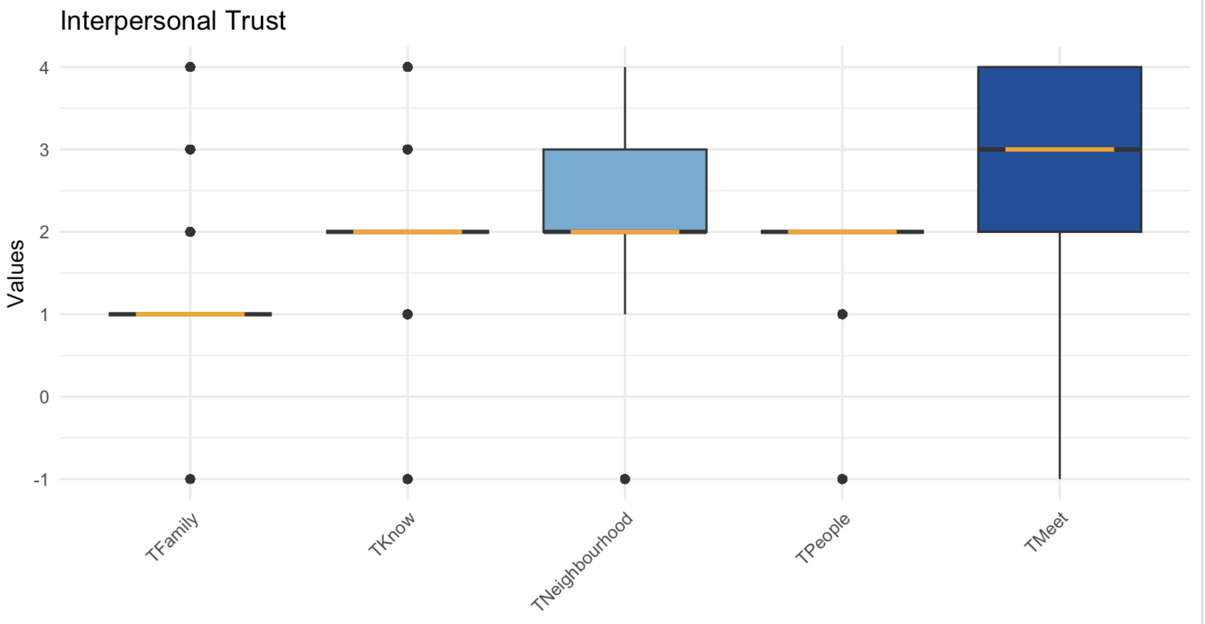


1.3.d Below is the distribution and the descriptive statistics of the variables related to the Confidence in Institutions.



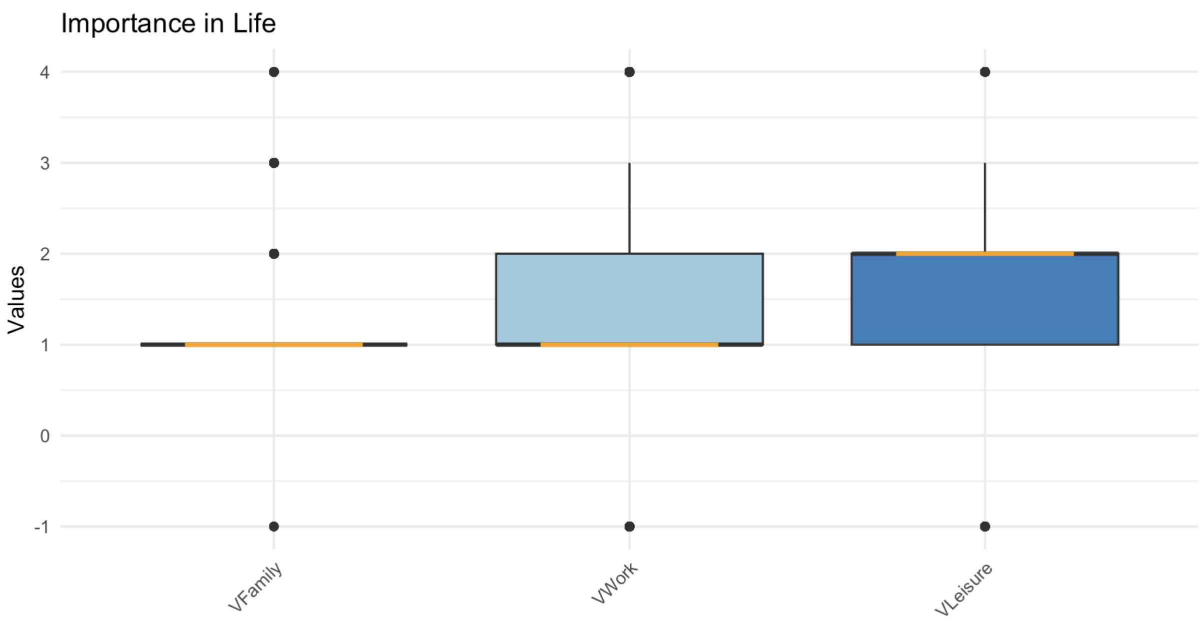
	Variable	Mean	Median	IQR	Min	Max	Mode	Category
1	CArmedForces	2.06	2	2	−1	4	2	Confidence
2	CUnions	2.46	3	1	−1	4	3	Confidence
3	CPolice	2.32	2	1	−1	4	2	Confidence
4	CPParties	2.86	3	2	−1	4	3	Confidence
5	CParliament	2.69	3	2	−1	4	3	Confidence
6	CCivilService	2.46	2	1	−1	4	2	Confidence
7	CUniversities	2.06	2	1	−1	4	2	Confidence
8	CElections	2.54	3	1	−1	4	2	Confidence
9	CBanks	2.33	2	1	−1	4	2	Confidence
10	CEnvOrg	2.20	2	1	−1	4	2	Confidence

1.3.e Below is the distribution and the descriptive statistics of the variables related to the Interpersonal Trust.



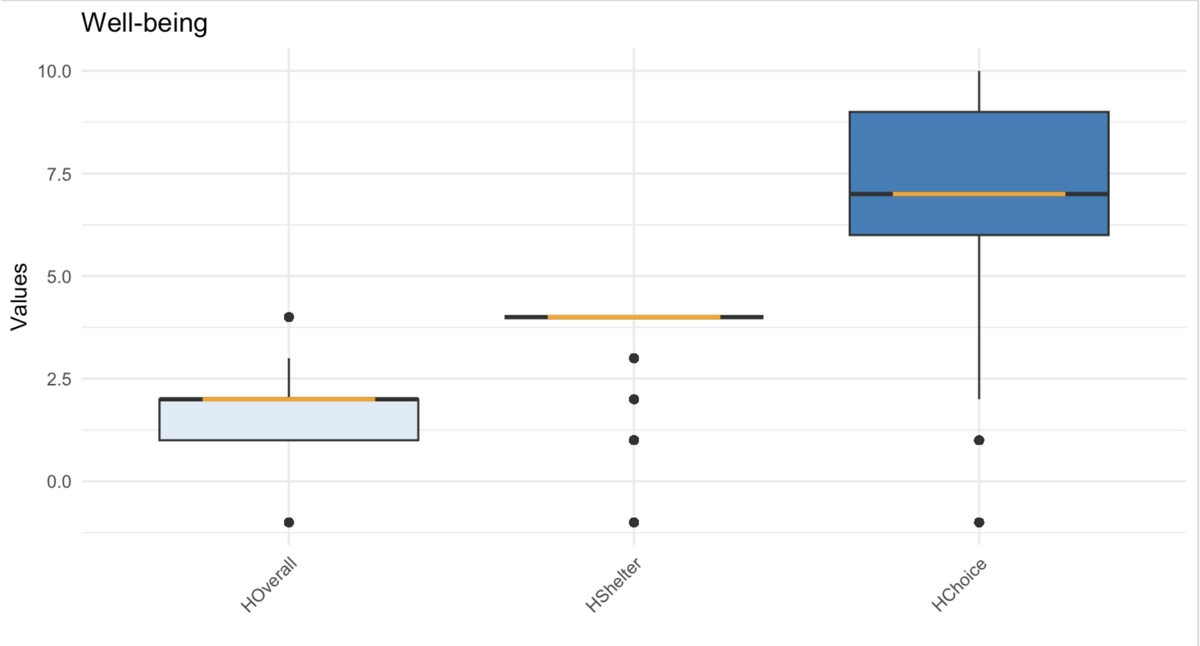
	Variable	Mean	Median	IQR	Min	Max	Mode	Category
1	TPeople	1.75	2	0	−1	2	2	Interpersonal Trust
2	TFamily	1.27	1	0	−1	4	1	Interpersonal Trust
3	TNeighbourhood	2.18	2	1	−1	4	2	Interpersonal Trust
4	TKnow	2.08	2	0	−1	4	2	Interpersonal Trust
5	TMeet	2.99	3	2	−1	4	3	Interpersonal Trust

1.3.f Below is the distribution and the descriptive statistics of the variables related to the Importance in Life.



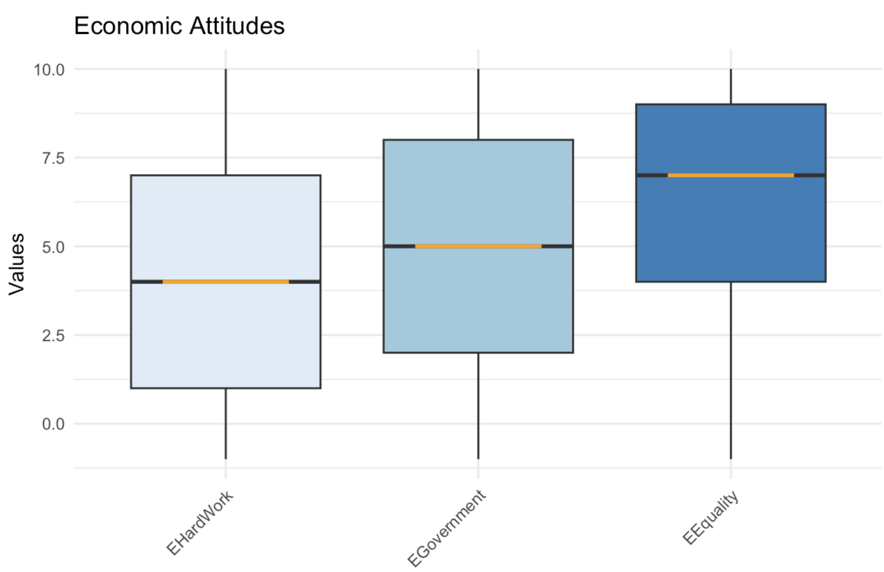
	Variable	Mean	Median	IQR	Min	Max	Mode	Category
1	VFamily	1.11	1	0	−1	4	1	Importance in Life
2	VLeisure	1.79	2	1	−1	4	2	Importance in Life
3	VWork	1.52	1	1	−1	4	1	Importance in Life

1.3.g Below is the distribution and the descriptive statistics of the variables related to Wellbeing.



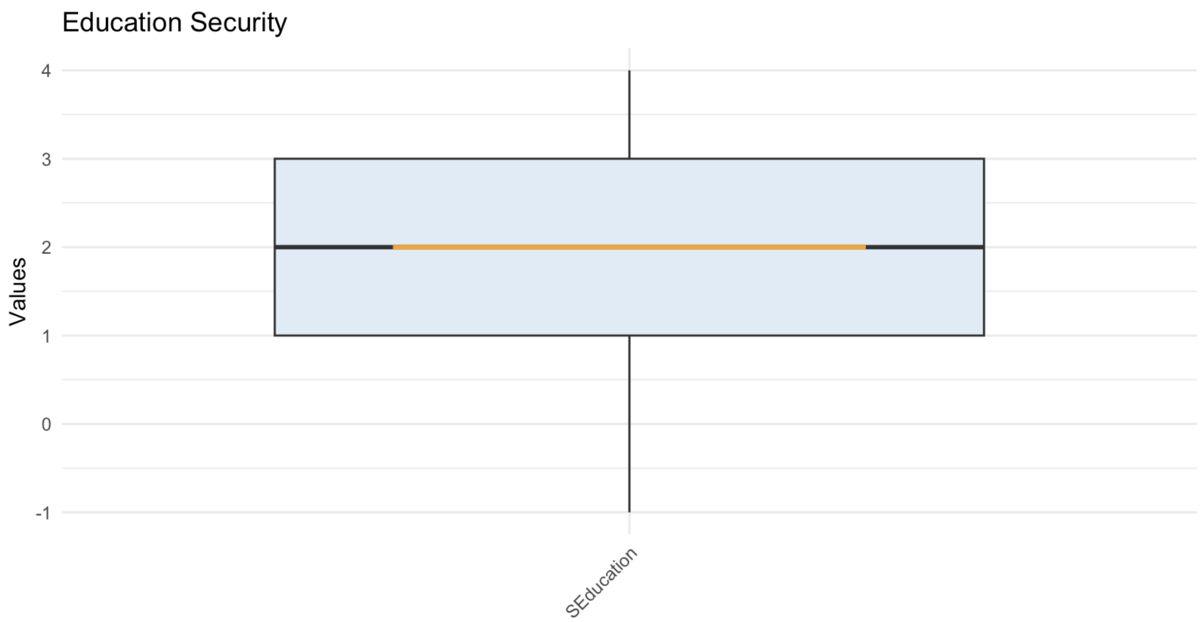
	Variable	Mean	Median	IQR	Min	Max	Mode	Category
1	HOverall	1.84	2	1	−1	4	2	Well−being
2	HChoice	7.19	7	3	−1	10	10	Well−being
3	HShelter	3.71	4	0	−1	4	4	Well−being

1.3.h Below is the distribution and the descriptive statistics of the variables related to the aspects that influence someone's economy.



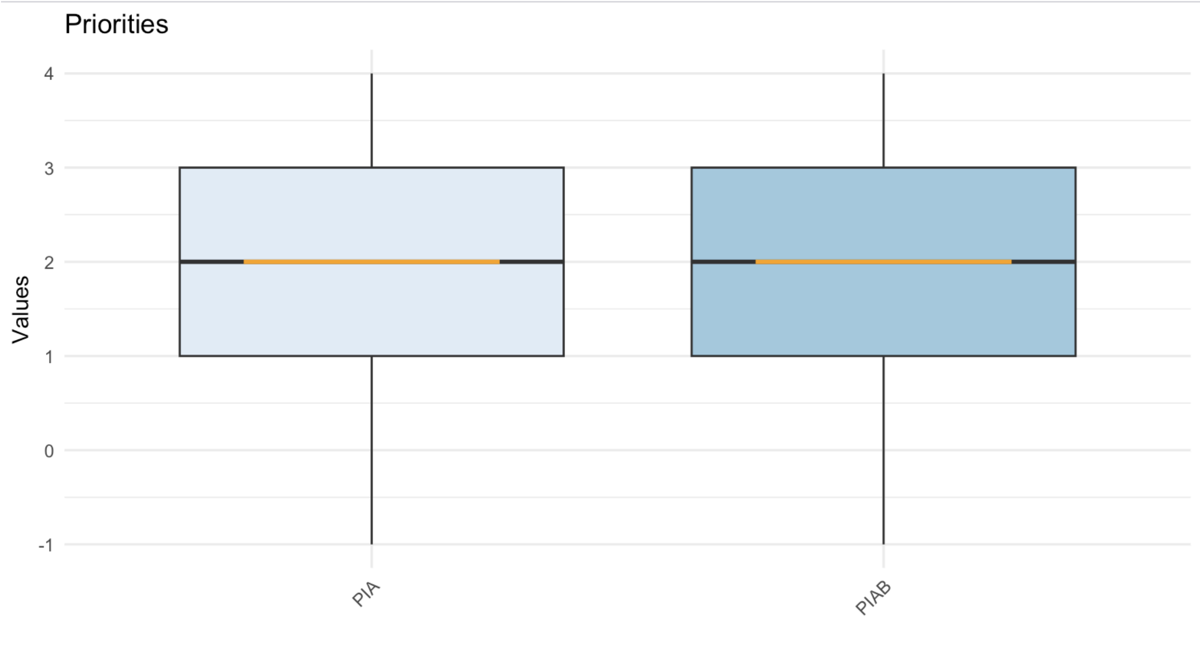
	Variable	Mean	Median	IQR	Min	Max	Mode	Category
1	EEquality	6.29	7	5	-1	10	10	Economic Attitudes
2	EGovernment	5.05	5	6	-1	10	1	Economic Attitudes
3	EHardWork	4.38	4	6	-1	10	1	Economic Attitudes

1.3.i Below is the distribution and the descriptive statistics of how secure someone is to continue pursuing education.



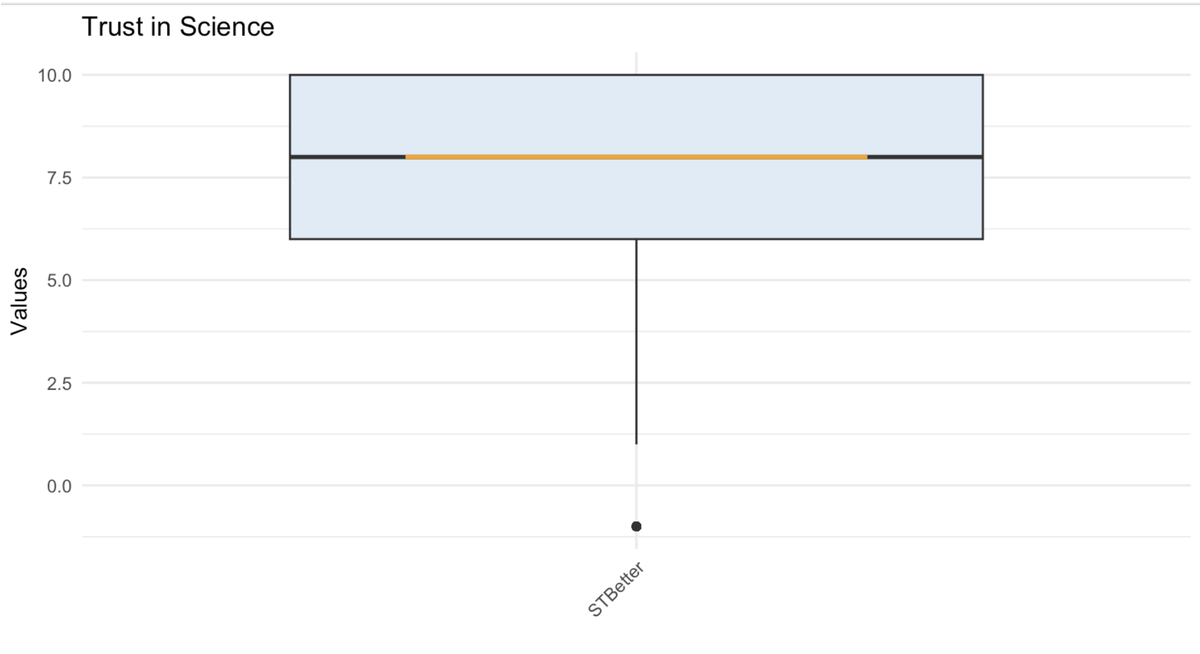
	Variable	Mean	Median	IQR	Min	Max	Mode	Category
1	SEducation	2.03	2	2	-1	4	1	Education Security

1.3.j Below is the distribution and the descriptive statistics of the variables related to the people's priorities in life.



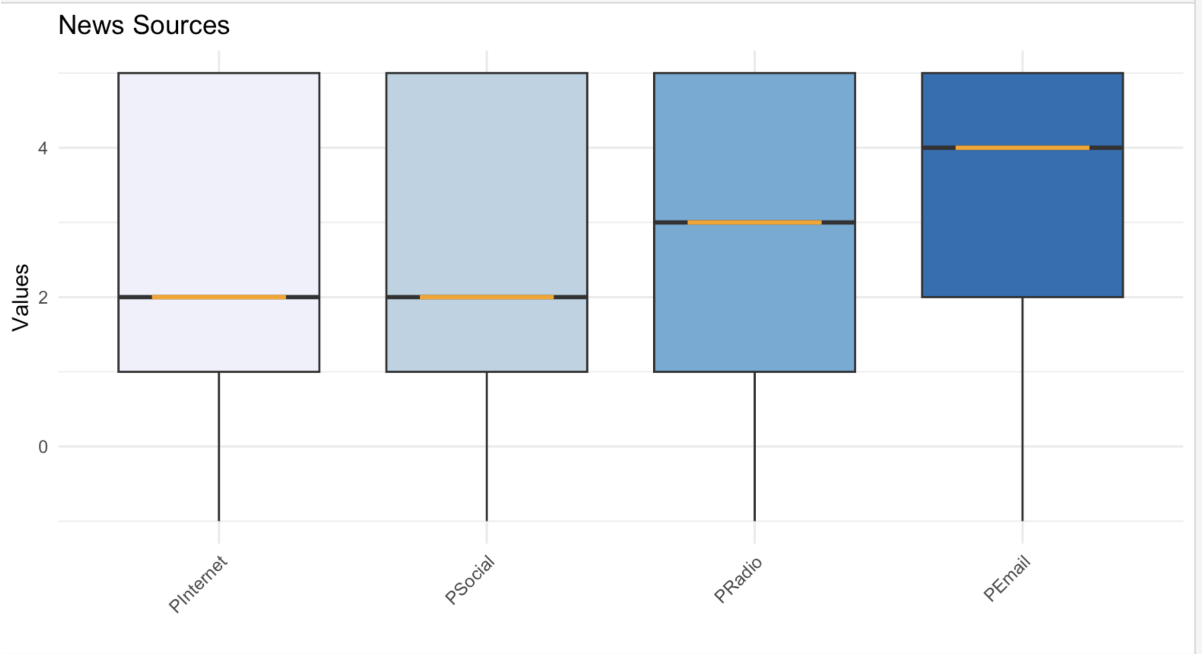
	Variable	Mean	Median	IQR	Min	Max	Mode	Category
1	PIA	1.92	2	2	−1	4	1	Priorities
2	PIAB	2.38	2	2	−1	4	3	Priorities

1.3.k Below is the distribution and the descriptive statistics of how much people trust Science & Technology.



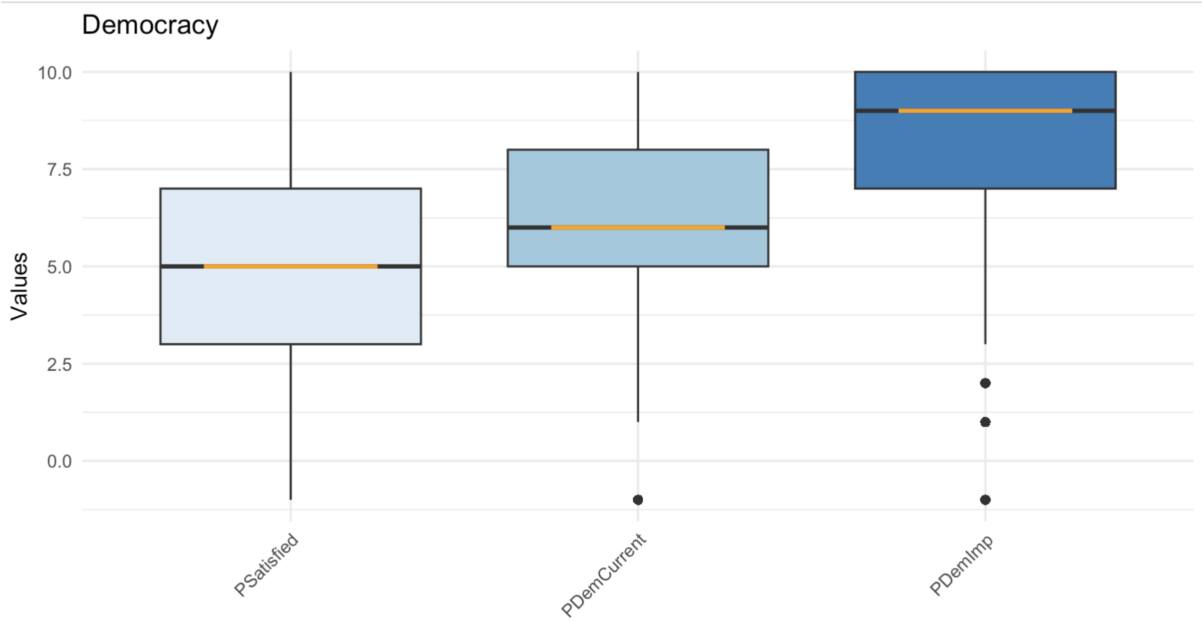
	Variable	Mean	Median	IQR	Min	Max	Mode	Category
1	STBetter	7.43	8	4	−1	10	10	Trust in Science

1.3.l Below is the distribution and the descriptive statistics of the variables related to the News Sources.



	Variable	Mean	Median	IQR	Min	Max	Mode	Category
1	PRadio	3.11	3	4	−1	5	5	News Sources
2	PEmail	3.50	4	3	−1	5	5	News Sources
3	PInternet	2.51	2	4	−1	5	1	News Sources
4	PSocial	2.62	2	4	−1	5	1	News Sources

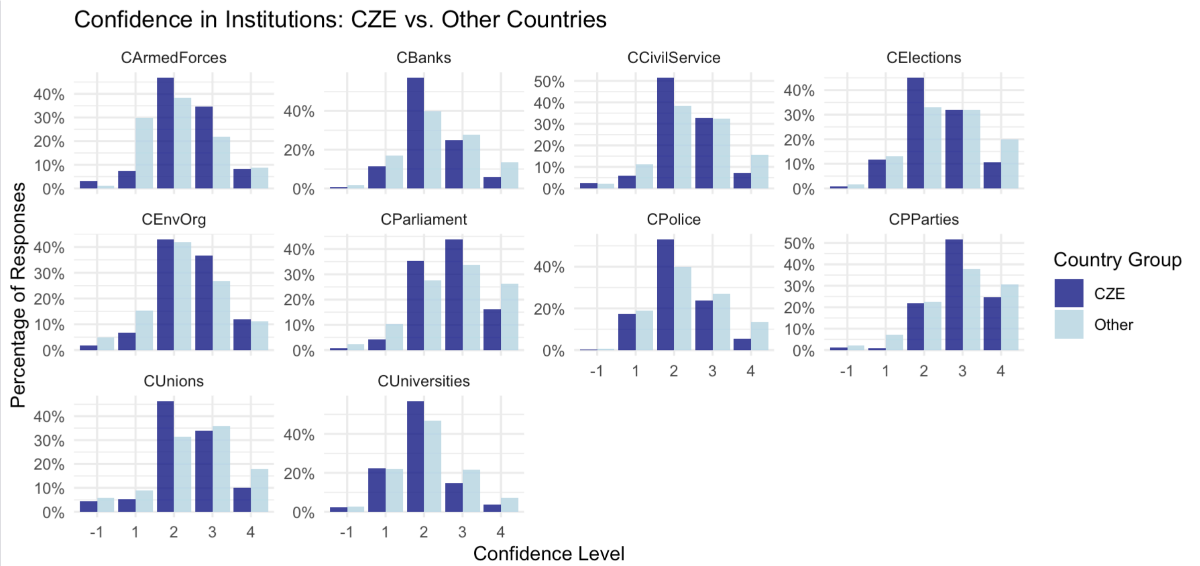
1.3.m Below is the distribution and the descriptive statistics of the variables related to Democracy.



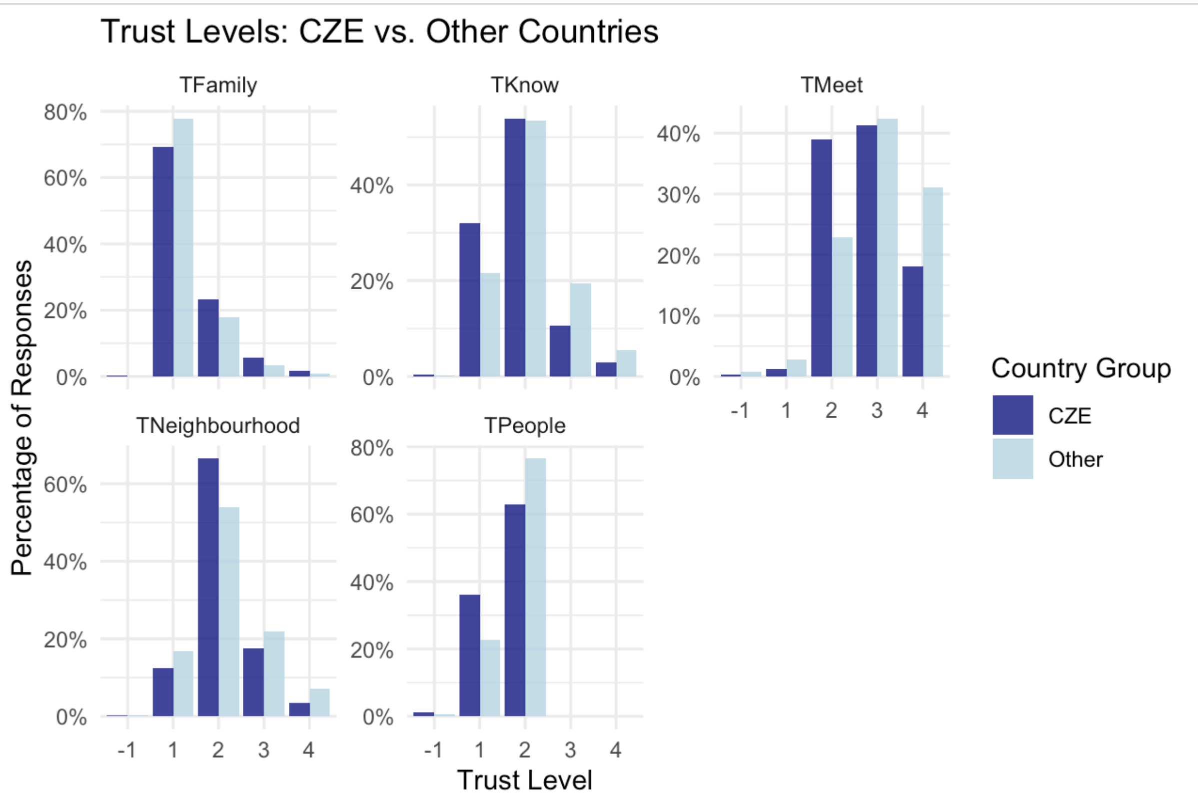
	Variable	Mean	Median	IQR	Min	Max	Mode	Category
1	PDemImp	8.27	9	3	-1	10	10	Democracy
2	PDemCurrent	6.08	6	3	-1	10	8	Democracy
3	PSatisfied	5.28	5	4	-1	10	5	Democracy

1.4 Response comparison Focus Country VS Other Countries

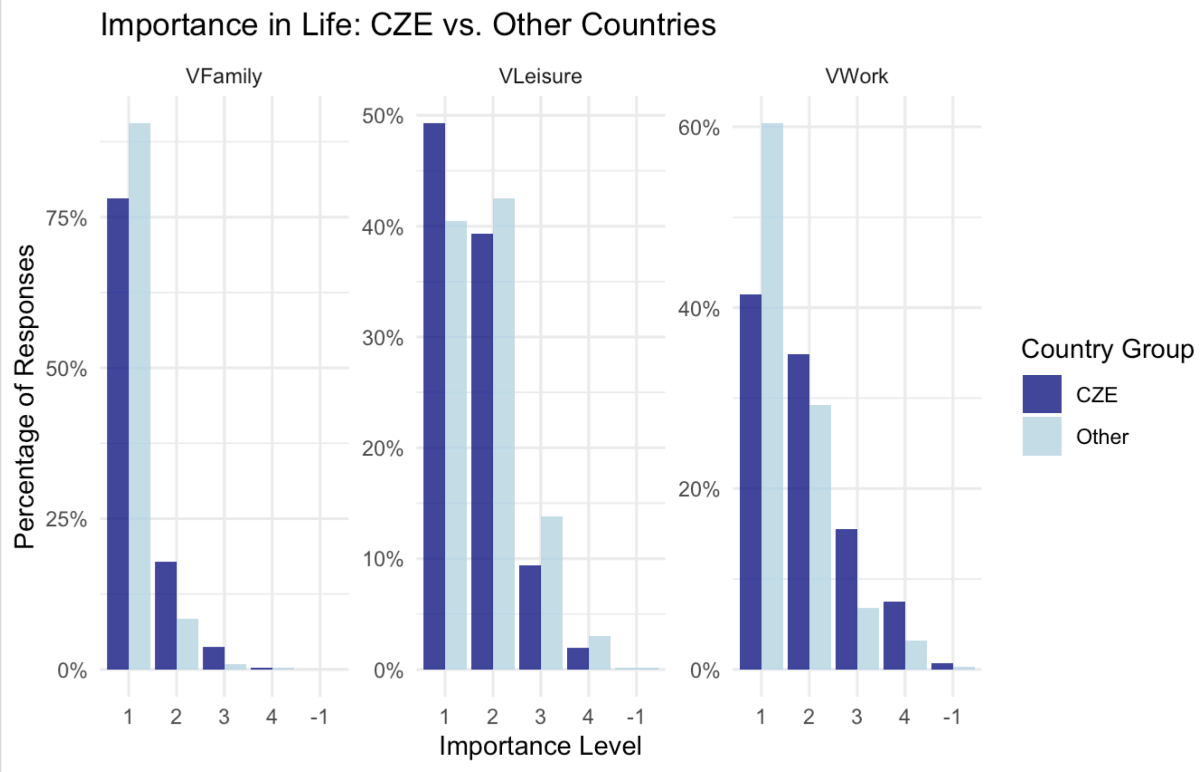
1.4.a Below is the comparison of the variables related to Confidence Level of the Czech Republic vs other countries.



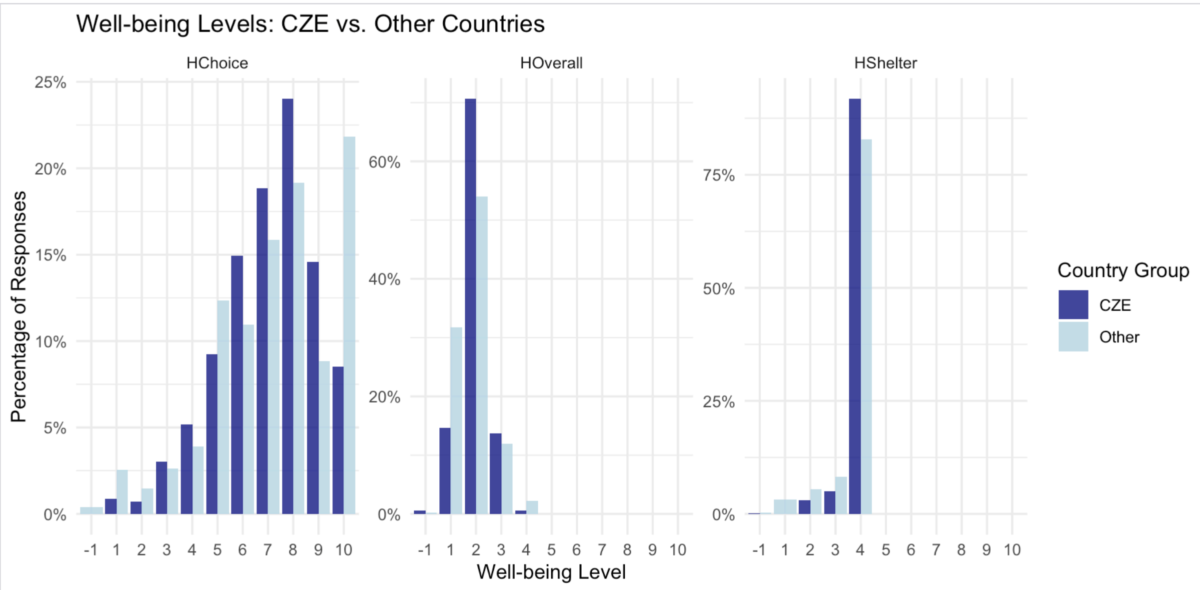
1.4.b Below is the comparison of the variables related to the Interpersonal Trust Level of the Czech Republic vs other countries.



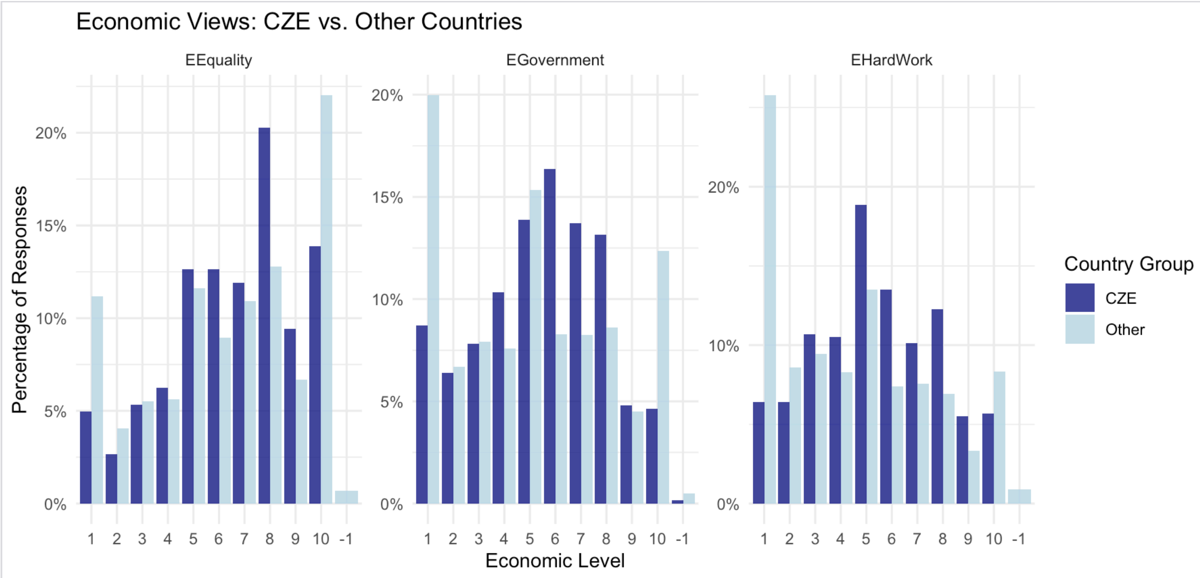
1.4.c Below is the comparison of the variables related to Importance in Life of the Czech Republic vs other countries.



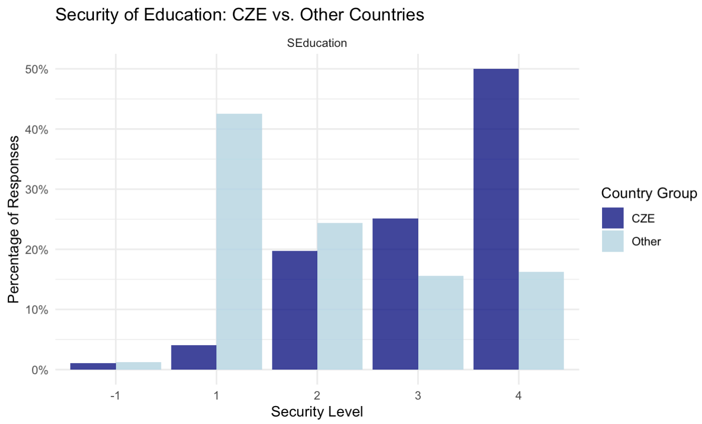
1.4.d Below is the comparison of the variables related to the Wellbeing Level of the Czech Republic vs other countries.



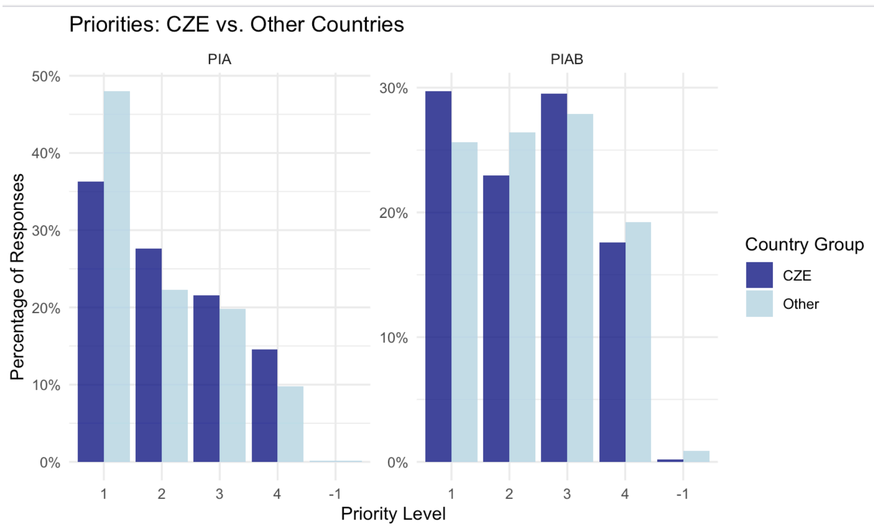
1.4.e Below is the comparison of the variables related to aspects which influences someone's economy of the Czech Republic vs other countries.



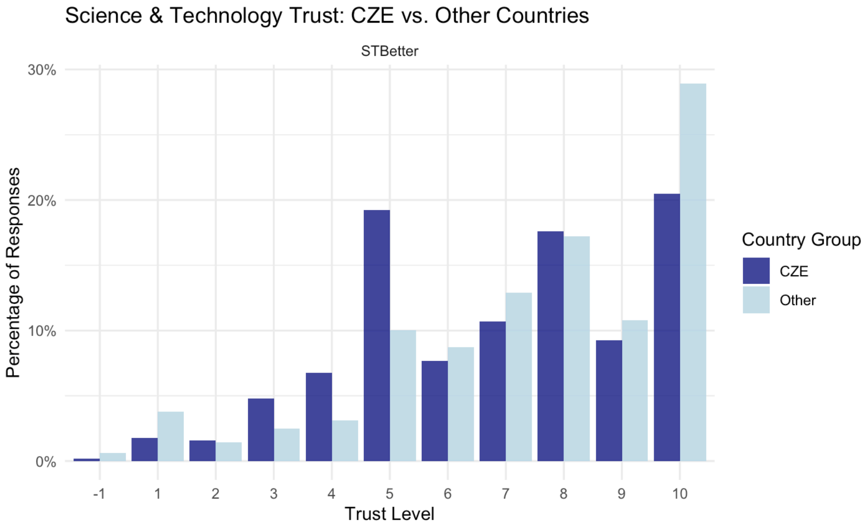
1.4.f Below is the comparison of the variables related to security of continuing pursuing education in the Czech Republic vs other countries.



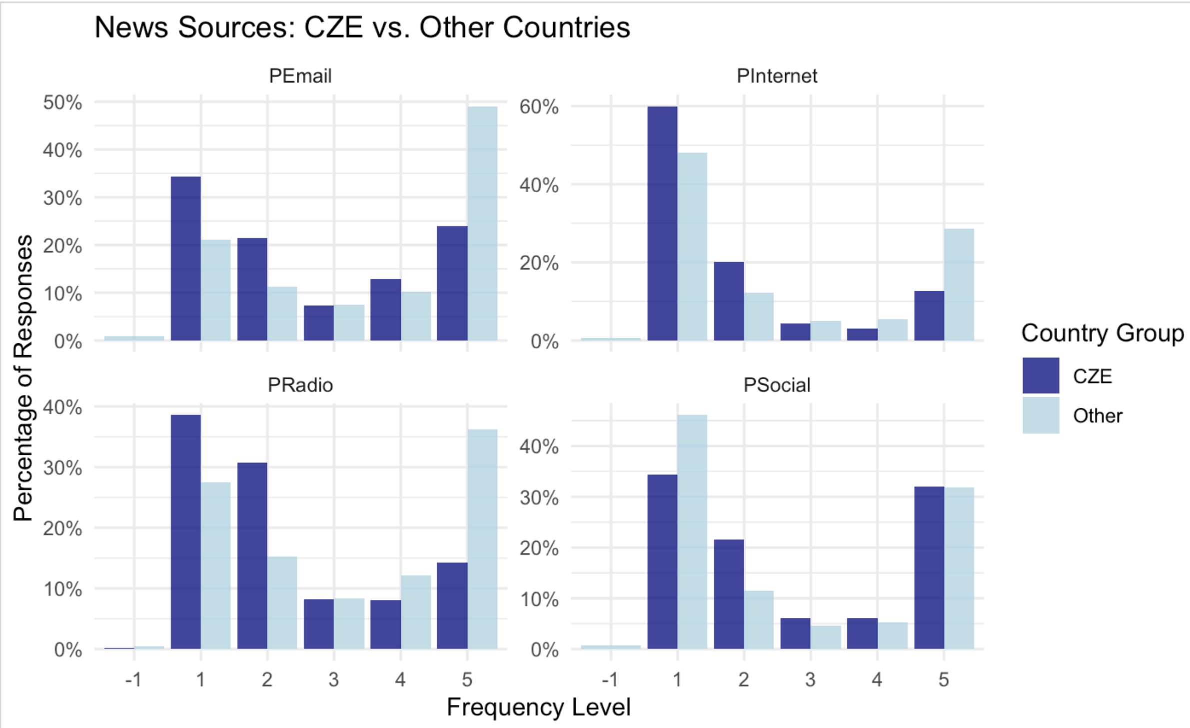
1.4.g Below is the comparison of the variables related to Priorities in Life in the Czech Republic vs other countries.



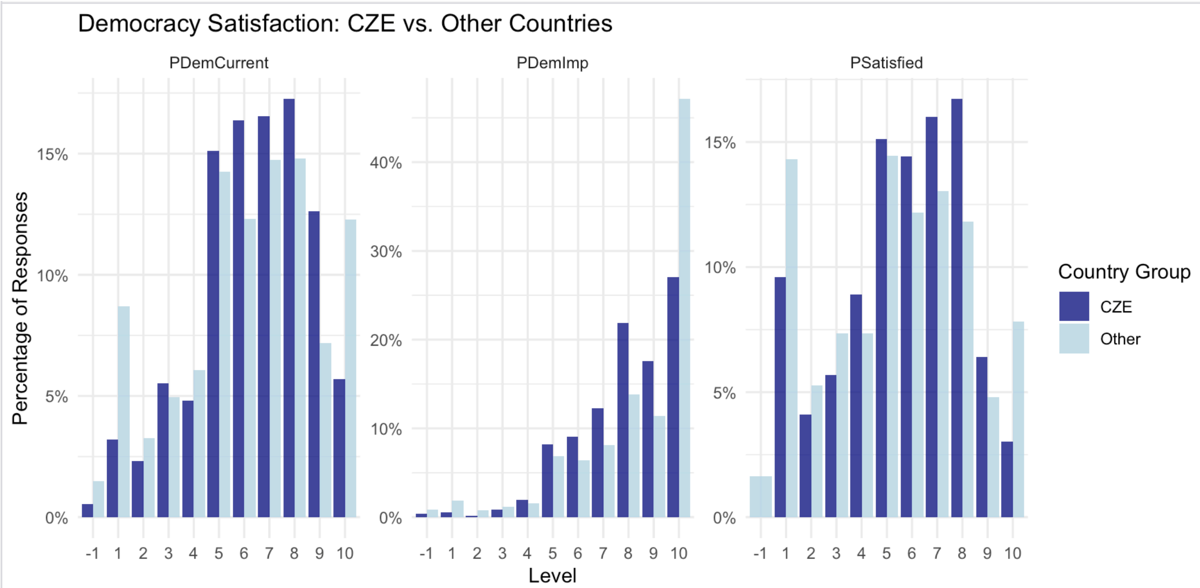
1.4.h Below is the comparison of the variables related to Trust in Science and Technology in the Czech Republic vs other countries.



1.4.i Below is the comparison of the variables related to News Source in the Czech Republic vs other countries.



1.4.l Below is the comparison of the variables related to Democratic Level in the Czech Republic vs other countries.



Appendix 2

2.1 Regression results for focus country (2b)

2.1.a Stargazer correlation table

	Dependent variable:									
	CArmedForces	CUnions	CPolice	CPParties	CParliament	CCivilService	CUniversities	CElections	CBanks	CEnvOrg
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
TPeople	0.018	0.080	0.110	-0.033	0.079	-0.020	0.102	0.103	-0.044	0.100
	(0.071)	(0.072)	(0.071)	(0.071)	(0.075)	(0.069)	(0.069)	(0.074)	(0.071)	(0.079)
TFamily	0.162***	0.129**	0.084	0.016	0.036	0.112**	0.179***	-0.025	0.105*	0.161**
	(0.057)	(0.058)	(0.057)	(0.057)	(0.060)	(0.055)	(0.056)	(0.059)	(0.057)	(0.063)
TNeighbourhood	-0.00003	-0.054	0.115*	0.016	-0.025	-0.006	0.025	0.085	0.130**	-0.108
	(0.061)	(0.063)	(0.061)	(0.062)	(0.065)	(0.059)	(0.060)	(0.064)	(0.061)	(0.068)
TMeet	0.091*	0.209***	0.155***	0.167***	0.128**	0.204***	-0.032	0.159***	0.031	0.095*
	(0.048)	(0.049)	(0.048)	(0.049)	(0.051)	(0.047)	(0.047)	(0.050)	(0.048)	(0.054)
TKnow	0.012	0.085	-0.003	0.163***	0.189***	0.116**	0.101*	0.016	0.091*	0.055
	(0.054)	(0.055)	(0.054)	(0.055)	(0.058)	(0.053)	(0.053)	(0.057)	(0.054)	(0.061)
VFamily	0.154**	0.176**	0.090	0.085	0.124*	0.147**	-0.062	0.059	-0.006	0.009
	(0.068)	(0.070)	(0.068)	(0.069)	(0.073)	(0.066)	(0.067)	(0.071)	(0.068)	(0.076)
VLeisure	0.038	0.010	0.011	-0.037	-0.026	-0.006	-0.024	0.013	0.059	0.094*
	(0.051)	(0.052)	(0.051)	(0.051)	(0.054)	(0.049)	(0.050)	(0.053)	(0.051)	(0.056)
VWork	0.008	-0.016	-0.010	0.051	-0.008	0.082**	0.091**	-0.036	0.069*	-0.043
	(0.040)	(0.040)	(0.040)	(0.040)	(0.042)	(0.038)	(0.039)	(0.041)	(0.040)	(0.044)
HOverall	0.097	-0.023	0.067	0.088	0.013	0.021	0.061	0.061	-0.051	0.027
	(0.065)	(0.066)	(0.065)	(0.065)	(0.069)	(0.063)	(0.063)	(0.067)	(0.065)	(0.072)
HChoice	-0.047**	-0.023	-0.053**	-0.012	-0.011	-0.007	-0.038*	-0.015	-0.050**	-0.086***
	(0.021)	(0.022)	(0.021)	(0.021)	(0.023)	(0.021)	(0.021)	(0.022)	(0.021)	(0.024)
HShelter	-0.033	0.135	-0.044	0.104	0.103	-0.009	0.036	0.031	0.046	0.101
	(0.082)	(0.084)	(0.082)	(0.083)	(0.087)	(0.079)	(0.080)	(0.085)	(0.082)	(0.091)
EEquality	0.006	-0.021	-0.031**	-0.023	-0.021	-0.012	-0.008	-0.033**	-0.006	0.005
	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)	(0.015)	(0.014)	(0.016)
EGovernment	0.010	-0.001	0.020	-0.012	-0.015	0.004	0.008	0.012	-0.016	-0.003
	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)	(0.014)	(0.014)	(0.015)	(0.015)	(0.016)
EHardWork	-0.012	0.005	0.002	-0.009	-0.029**	-0.002	0.037***	0.017	-0.011	-0.027*
	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)	(0.015)	(0.014)	(0.016)
SEducation	-0.058	-0.002	-0.039	-0.072*	-0.024	-0.016	0.017	-0.062	-0.058	0.027
	(0.037)	(0.038)	(0.037)	(0.037)	(0.039)	(0.036)	(0.036)	(0.038)	(0.037)	(0.041)
STBetter	0.024	-0.007	0.020	0.009	-0.010	0.028*	-0.013	-0.038**	-0.014	0.039**
	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)	(0.015)	(0.015)	(0.016)	(0.015)	(0.017)
PRadio	0.053**	0.118***	0.057**	0.079***	0.085***	0.066***	0.079***	0.125***	0.020	0.057**
	(0.024)	(0.025)	(0.024)	(0.024)	(0.026)	(0.023)	(0.024)	(0.025)	(0.024)	(0.027)

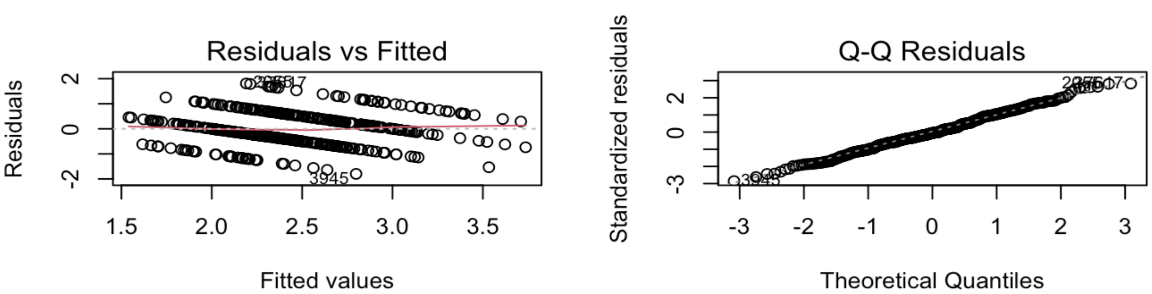
PEmail	-0.017	-0.065**	-0.090***	-0.023	0.015	-0.019	-0.034	-0.071**	-0.065**	0.00004
	(0.026)	(0.027)	(0.026)	(0.027)	(0.028)	(0.025)	(0.026)	(0.027)	(0.026)	(0.029)
PInternet	0.043	0.004	0.070**	0.030	0.049	0.031	0.043	0.048	0.060*	0.086**
	(0.033)	(0.034)	(0.033)	(0.033)	(0.035)	(0.032)	(0.032)	(0.034)	(0.033)	(0.037)
PSocial	-0.074***	-0.094***	-0.050*	0.027	-0.003	-0.065***	-0.058**	-0.040	-0.059**	-0.067**
	(0.026)	(0.027)	(0.026)	(0.026)	(0.028)	(0.025)	(0.025)	(0.027)	(0.026)	(0.029)
PDemImp	-0.070***	0.028	-0.030	0.017	-0.004	-0.055***	-0.030	-0.018	-0.017	-0.036
	(0.020)	(0.021)	(0.020)	(0.021)	(0.022)	(0.020)	(0.020)	(0.021)	(0.020)	(0.023)
PDemCurrent	-0.009	-0.009	0.008	0.006	-0.031	-0.007	-0.024	-0.041*	-0.021	0.024
	(0.020)	(0.021)	(0.020)	(0.020)	(0.022)	(0.020)	(0.020)	(0.021)	(0.020)	(0.023)
PSatisfied	-0.029*	-0.012	-0.042**	-0.020	-0.032*	-0.034**	-0.005	-0.049***	-0.001	-0.026
	(0.017)	(0.018)	(0.017)	(0.018)	(0.018)	(0.017)	(0.017)	(0.018)	(0.017)	(0.019)
MF_male	0.003	0.087	0.034	0.082	0.025	0.065	0.126**	-0.043	0.141**	0.202***
	(0.063)	(0.065)	(0.063)	(0.064)	(0.067)	(0.061)	(0.062)	(0.066)	(0.063)	(0.071)
MF_female										
MF_others										
Age	-0.004	0.001	-0.006*	-0.0001	-0.003	0.001	-0.006*	-0.007**	-0.004	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Edu	-0.017	0.028	0.020	0.016	-0.012	0.021	-0.013	-0.041**	0.011	0.047**
	(0.019)	(0.019)	(0.019)	(0.019)	(0.020)	(0.018)	(0.019)	(0.020)	(0.019)	(0.021)
Emp_fulltime	-0.387	-0.738	-0.769	-0.015	-0.159	-0.596	-0.735	-0.340	-0.807	-0.365
	(0.495)	(0.506)	(0.495)	(0.501)	(0.526)	(0.480)	(0.485)	(0.516)	(0.495)	(0.553)
Emp_parttime	-0.239	-0.438	-0.803	0.418	-0.071	-0.441	-0.800	-0.115	-0.801	-0.486
	(0.517)	(0.528)	(0.517)	(0.522)	(0.549)	(0.501)	(0.506)	(0.539)	(0.517)	(0.577)
Emp_self	-0.254	-0.568	-0.642	0.135	-0.127	-0.406	-0.671	-0.224	-0.759	-0.416
	(0.506)	(0.517)	(0.506)	(0.511)	(0.537)	(0.490)	(0.496)	(0.527)	(0.506)	(0.565)
Emp_retired	-0.291	-0.605	-0.669	0.021	-0.152	-0.544	-0.630	-0.077	-0.594	-0.371
	(0.493)	(0.504)	(0.493)	(0.499)	(0.524)	(0.478)	(0.483)	(0.514)	(0.493)	(0.550)
Emp_housewife	-0.608	-0.925*	-1.097**	0.256	-0.229	-0.838*	-0.818	-0.476	-1.032**	-0.391
	(0.523)	(0.534)	(0.523)	(0.529)	(0.555)	(0.507)	(0.512)	(0.545)	(0.523)	(0.584)
Emp_student	-0.947*	-0.935*	-1.319**	-0.189	-0.974*	-0.923*	-1.384***	-0.953*	-0.940*	-1.065*
	(0.536)	(0.548)	(0.536)	(0.542)	(0.570)	(0.520)	(0.525)	(0.559)	(0.536)	(0.599)
Emp_unemp	-0.376	-1.063*	-0.868	0.203	0.014	-0.640	-0.851	-0.269	-0.945*	0.208
	(0.549)	(0.560)	(0.549)	(0.555)	(0.583)	(0.532)	(0.537)	(0.572)	(0.548)	(0.612)
PIA_1	-0.045	-0.042	0.013	0.101	0.045	0.075	-0.040	0.032	0.069	-0.106
	(0.103)	(0.106)	(0.103)	(0.105)	(0.110)	(0.100)	(0.101)	(0.108)	(0.103)	(0.115)
PIA_2	-0.146	-0.135	0.116	-0.007	-0.202*	0.134	0.068	0.065	0.054	-0.238**
PIAB_3	-0.032	-0.198**	0.139	-0.015	0.042	0.060	0.040	0.196*	0.116	-0.064
	(0.096)	(0.098)	(0.096)	(0.097)	(0.102)	(0.093)	(0.094)	(0.100)	(0.096)	(0.107)
Constant	3.483***	2.009***	2.952***	1.502**	2.450***	2.208***	2.656***	3.237***	3.333***	2.292***
	(0.673)	(0.687)	(0.673)	(0.680)	(0.715)	(0.652)	(0.659)	(0.702)	(0.673)	(0.751)
Observations	498	498	498	498	498	498	498	498	498	498
R^2	0.309	0.284	0.315	0.233	0.267	0.292	0.308	0.364	0.245	0.232
Adjusted R^2	0.250	0.223	0.257	0.167	0.205	0.231	0.249	0.310	0.181	0.167
Residual Std. Error (df = 458)	0.655	0.669	0.655	0.662	0.696	0.635	0.642	0.683	0.655	0.731
F Statistic (df = 39; 458)	5.244***	4.652***	5.400***	3.564***	4.285***	4.838***	5.233***	6.715***	3.813***	3.557***
Note:										*p<0.1; **p<0.05; ***p<0.01

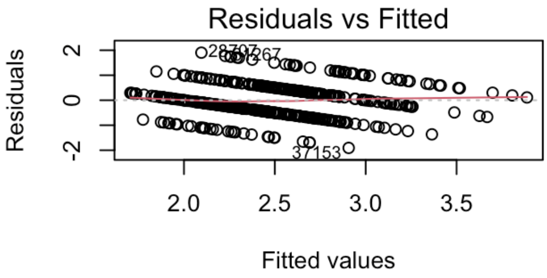
2.1.b Stargazer correlation table with selected variables

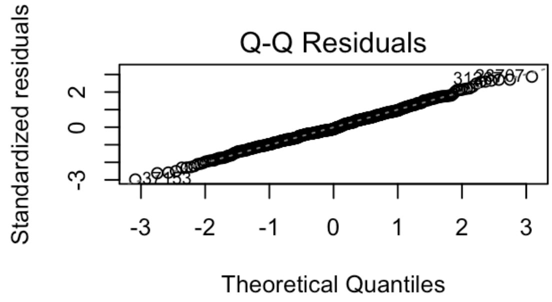
	CArmedForces	CUnions	CPolice	CPParties	CParliament	CCivilService	CUniversities	CElections	CBanks	CEnvOrg
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
TNeighbourhood	0.047	-0.014	0.163***	0.046	-0.012	0.037	0.063	0.116*	0.177***	-0.037
	(0.062)	(0.064)	(0.062)	(0.061)	(0.064)	(0.059)	(0.060)	(0.065)	(0.061)	(0.069)
TFamily	0.261***	0.193***	0.174***	0.012	0.042	0.176***	0.190***	0.029	0.152***	0.167***
	(0.054)	(0.055)	(0.054)	(0.053)	(0.056)	(0.051)	(0.052)	(0.057)	(0.053)	(0.060)
PDemCurrent	-0.030	0.008	0.002	0.019	-0.040**	-0.011	-0.039**	-0.064***	-0.033*	0.021
	(0.019)	(0.020)	(0.019)	(0.019)	(0.020)	(0.018)	(0.019)	(0.020)	(0.019)	(0.022)
PSatisfied	-0.025	-0.013	-0.036**	-0.035**	-0.045**	-0.034**	-0.008	-0.051***	-0.003	-0.035*
	(0.018)	(0.018)	(0.018)	(0.017)	(0.018)	(0.017)	(0.017)	(0.018)	(0.017)	(0.020)
TMeet	0.062	0.208***	0.114**	0.217***	0.182***	0.198***	-0.020	0.170***	0.020	0.121**
	(0.048)	(0.050)	(0.049)	(0.047)	(0.050)	(0.046)	(0.047)	(0.051)	(0.048)	(0.054)
TPeople	0.012	0.052	0.073	-0.046	0.076	-0.035	0.060	0.083	-0.071	0.080
	(0.072)	(0.074)	(0.073)	(0.071)	(0.075)	(0.069)	(0.070)	(0.076)	(0.071)	(0.081)
TKnow	0.079	0.113**	0.038	0.152***	0.207***	0.148***	0.145***	0.066	0.103*	0.091
	(0.055)	(0.057)	(0.055)	(0.054)	(0.057)	(0.052)	(0.053)	(0.058)	(0.054)	(0.062)
HOverall	0.163***	0.029	0.123**	0.162***	0.086	0.075	0.117**	0.120*	0.074	0.150**
	(0.060)	(0.062)	(0.060)	(0.059)	(0.062)	(0.057)	(0.058)	(0.063)	(0.059)	(0.067)
EHardWork	0.0002	0.005	0.013	0.003	-0.011	0.008	0.041***	0.031**	-0.014	-0.015
	(0.014)	(0.014)	(0.014)	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)	(0.015)
PIA_2	-0.152	-0.089	0.110	0.011	-0.210**	0.119	0.065	0.024	0.037	-0.206*
	(0.103)	(0.105)	(0.103)	(0.100)	(0.106)	(0.097)	(0.099)	(0.108)	(0.101)	(0.115)
PIAB_2	-0.109	-0.037	0.053	0.099	-0.009	0.041	0.017	0.129	0.001	0.016
	(0.080)	(0.082)	(0.080)	(0.078)	(0.083)	(0.076)	(0.077)	(0.084)	(0.079)	(0.089)
Emp_housewife	-0.036	-0.178	-0.148	0.342**	0.148	-0.180	0.139	0.172	-0.177	0.005
	(0.165)	(0.170)	(0.166)	(0.162)	(0.172)	(0.157)	(0.160)	(0.174)	(0.163)	(0.185)
Emp_self	0.108	0.188*	0.173	0.151	0.049	0.152	0.076	0.087	-0.026	0.016
	(0.109)	(0.112)	(0.110)	(0.107)	(0.113)	(0.104)	(0.106)	(0.115)	(0.108)	(0.122)
PIA_3	-0.087	0.121	0.063	0.155	-0.083	0.046	0.004	-0.041	-0.030	-0.024
	(0.106)	(0.109)	(0.106)	(0.104)	(0.110)	(0.100)	(0.103)	(0.112)	(0.105)	(0.119)
PIA_1	-0.046	0.036	-0.039	0.053	-0.017	0.061	-0.097	-0.086	-0.047	-0.098
	(0.097)	(0.099)	(0.097)	(0.095)	(0.100)	(0.092)	(0.093)	(0.102)	(0.095)	(0.108)
Constant	1.735***	1.312***	0.879***	1.699***	2.148***	1.320***	1.135***	1.684***	1.689***	1.704***
	(0.238)	(0.245)	(0.239)	(0.233)	(0.247)	(0.225)	(0.230)	(0.251)	(0.235)	(0.266)
Observations	498	498	498	498	498	498	498	498	498	498
R^2	0.194	0.156	0.197	0.164	0.187	0.213	0.217	0.244	0.146	0.105
Adjusted R^2	0.169	0.130	0.172	0.138	0.161	0.189	0.192	0.220	0.120	0.077
Residual Std. Error (df = 482)	0.689	0.708	0.691	0.674	0.715	0.653	0.666	0.726	0.679	0.770
F Statistic (df = 15; 482)	7.758***	5.944***	7.897***	6.289***	7.377***	8.700***	8.885***	10.343***	5.503***	3.757***

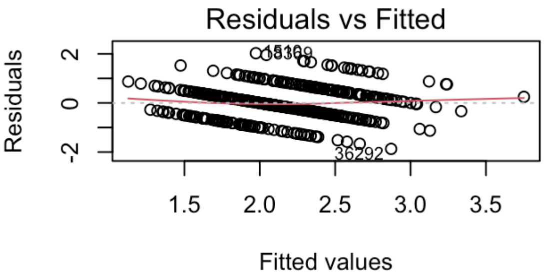
Note: *p<0.1; **p<0.05; ***p<0.01

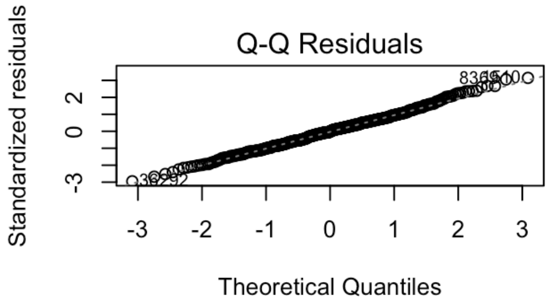
2.1.c Residuals vs Fitted and Q-Q Residual plots for focus country for each confidence variable

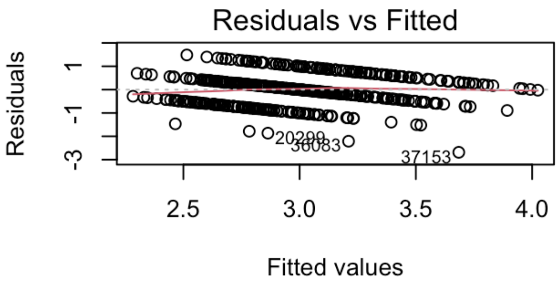


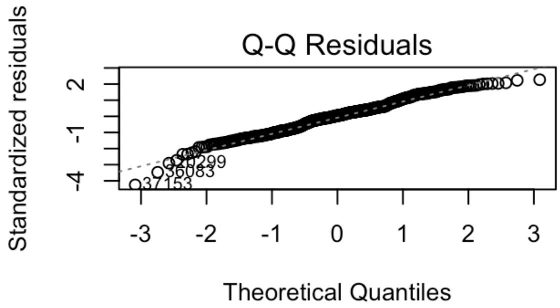

Residuals vs Fitted
Residuals
Fitted values

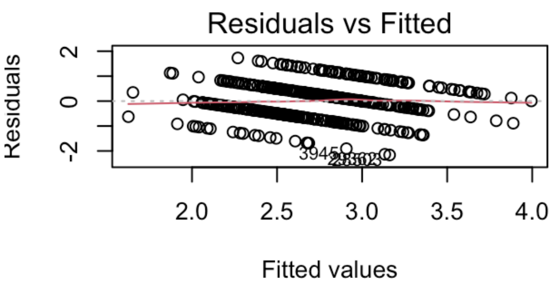

Q-Q Residuals
Standardized residuals
Theoretical Quantiles

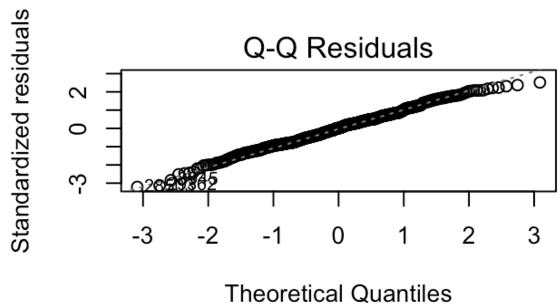

Residuals vs Fitted
Residuals
Fitted values

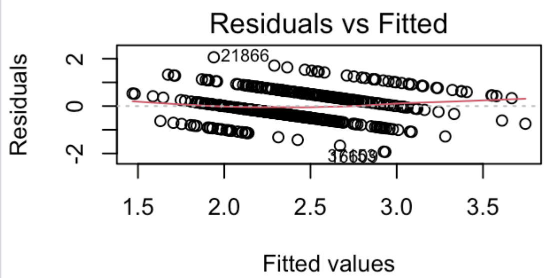

Q-Q Residuals
Standardized residuals
Theoretical Quantiles

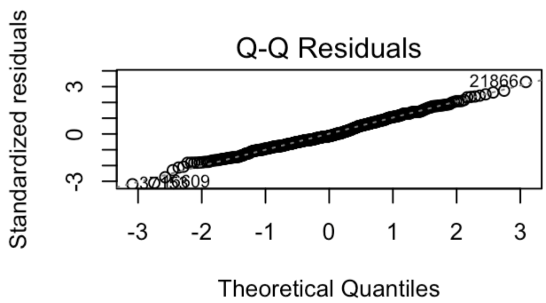

Residuals vs Fitted
Residuals
Fitted values

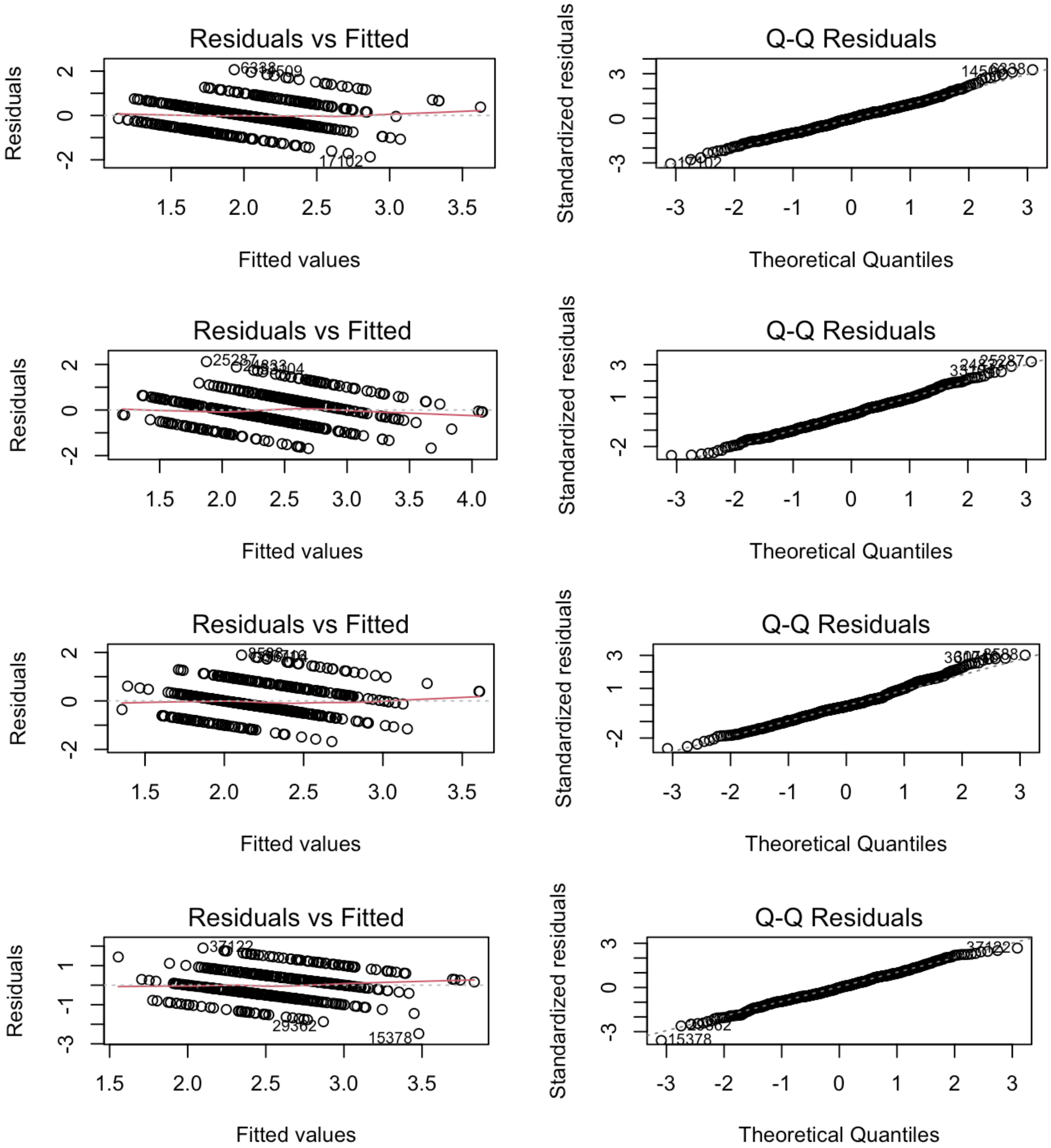

Q-Q Residuals
Standardized residuals
Theoretical Quantiles


Residuals vs Fitted
Residuals
Fitted values


Q-Q Residuals
Standardized residuals
Theoretical Quantiles


Residuals vs Fitted
Residuals
Fitted values


Q-Q Residuals
Standardized residuals
Theoretical Quantiles



2.2 Regression results for other countries (2c)

2.2.a Stargazer correlation table

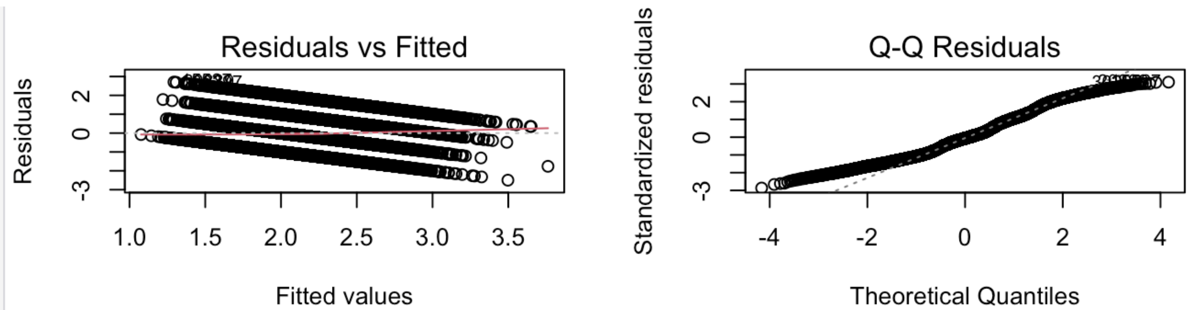
	Dependent variable:									
	CArmedForces	CUnions	CPolice	CPParties	CParliament	CCivilService	CUniversities	CElections	CBanks	CEnvOrg
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
TPeople	0.065***	0.088***	0.130***	0.187***	0.164***	0.131***	0.061***	0.115***	0.075***	0.057***
	(0.013)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)
TFamily	0.140***	0.064***	0.138***	0.069***	0.082***	0.078***	0.090***	0.059***	0.065***	0.050***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
TNeighbourhood	0.133***	0.115***	0.116***	0.118***	0.119***	0.127***	0.072***	0.108***	0.080***	0.057***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
TMeet	0.005	0.099***	0.040***	0.094***	0.080***	0.072***	0.010	0.075***	0.040***	0.055***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
TKnow	0.067***	0.072***	0.092***	0.016**	0.047***	0.087***	0.098***	0.056***	0.090***	0.105***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
VFamily	0.111***	0.022	0.059***	-0.016	-0.003	0.036***	0.024*	-0.007	0.001	-0.007
	(0.014)	(0.014)	(0.014)	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
VLeisure	-0.019***	-0.032***	-0.006	-0.048***	-0.068***	-0.054***	-0.013**	-0.027***	-0.037***	-0.004
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
VWork	0.051***	0.033***	0.017**	-0.001	0.031***	0.023***	0.077***	0.035***	0.060***	0.047***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
HOverall	0.041***	0.041***	0.059***	0.067***	0.070***	0.046***	0.089***	0.068***	0.107***	0.064***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
HChoice	-0.004*	0.004*	0.007***	0.016***	0.016***	0.009***	-0.012***	0.005**	-0.003	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
HShelter	-0.014**	0.041***	-0.034***	0.018***	0.024***	-0.012*	-0.004	0.001	0.014**	0.013*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
EEquality	-0.003*	0.006***	0.0002	-0.001	0.0002	-0.002	0.001	0.003	-0.013***	0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
EGovernment	0.001	0.011***	-0.002	0.004***	0.006***	0.009***	0.011***	0.008***	-0.002	0.009***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
EHardWork	0.013***	0.005***	0.012***	0.004**	0.011***	0.009***	0.010***	0.007***	0.025***	0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
SEducation	-0.001	0.029***	-0.026***	0.032***	0.022***	0.006	0.023***	-0.014***	0.043***	0.033***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
STBetter	-0.018***	-0.016***	-0.008***	-0.010***	-0.016***	-0.025***	-0.015***	-0.009***	-0.011***	-0.009***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PRadio	-0.018***	0.004	0.006**	-0.003	-0.001	0.001	0.011***	0.018***	-0.001	0.015***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
PEmail	-0.004	0.004	-0.008**	0.002	0.003	-0.005	0.002	-0.006	0.008**	-0.003
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
PInternet	-0.001	-0.006	0.024***	-0.025***	-0.026***	0.006	-0.002	0.015***	-0.013***	0.007
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
PSocial	0.014***	0.006	0.010**	0.015***	0.006	-0.0003	-0.004	-0.032***	0.006	0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)

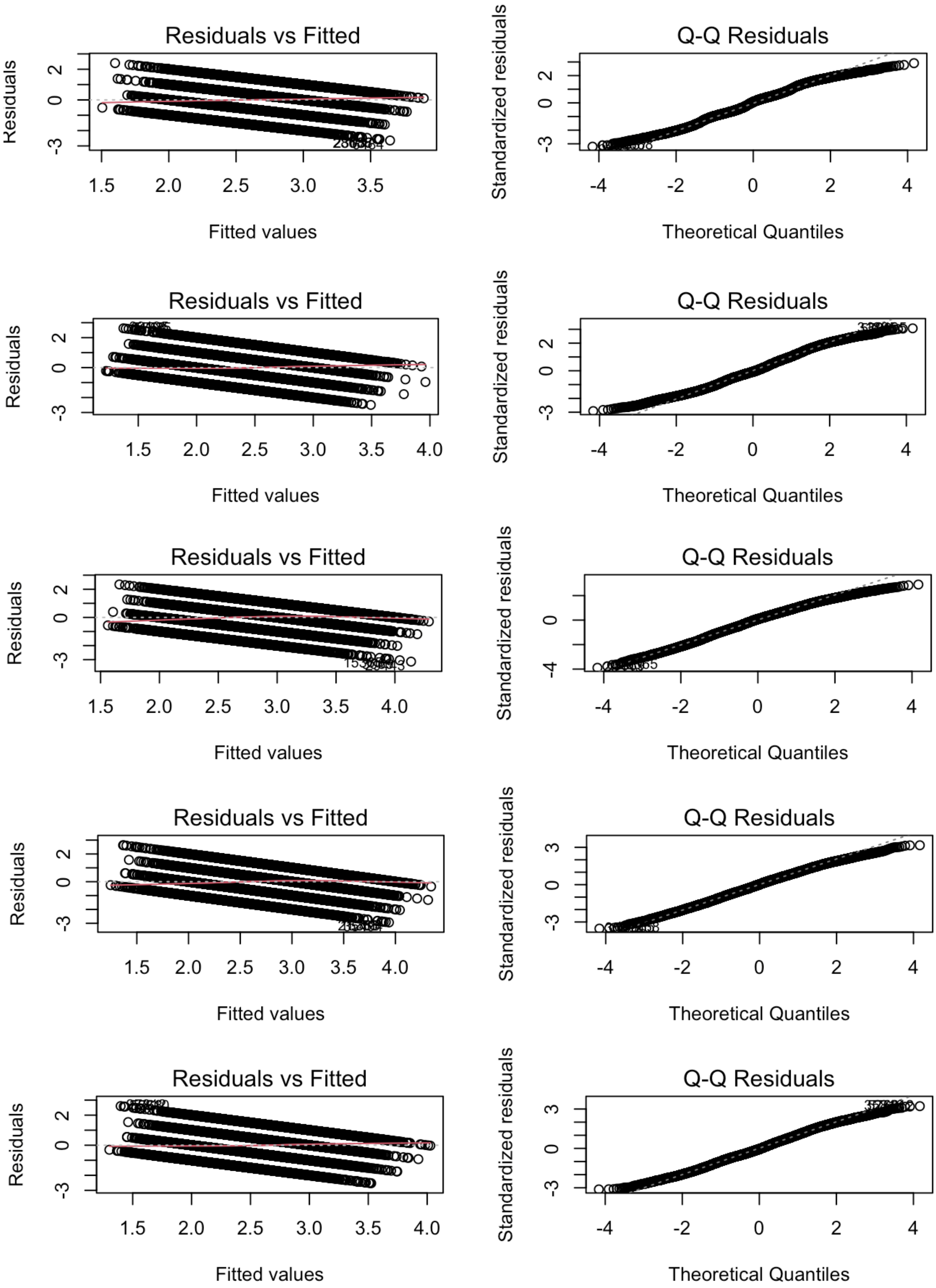
PDemImp	0.005**	-0.0002	0.005**	0.008***	0.002	-0.001	-0.024***	-0.006***	0.002	-0.022***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PDemCurrent	-0.036***	-0.024***	-0.041***	-0.027***	-0.042***	-0.032***	-0.023***	-0.056***	-0.029***	-0.025***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PSatisfied	-0.028***	-0.045***	-0.047***	-0.094***	-0.102***	-0.065***	-0.027***	-0.080***	-0.040***	-0.028***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
MF_male	-0.558	0.235	-0.116	0.630*	0.0001	-0.121	-0.185	-0.239	-0.372	0.083
	(0.354)	(0.337)	(0.349)	(0.329)	(0.339)	(0.330)	(0.324)	(0.345)	(0.351)	(0.342)
MF_female	-0.452	0.198	-0.150	0.589*	-0.028	-0.162	-0.203	-0.235	-0.420	0.026
	(0.354)	(0.337)	(0.349)	(0.329)	(0.339)	(0.330)	(0.324)	(0.344)	(0.351)	(0.342)
MF_others										
Age	-0.002***	0.002***	-0.003***	0.001***	0.002***	0.002***	-0.0001	-0.001	0.003***	0.003***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Edu	0.018***	0.024***	0.022***	0.025***	0.019***	0.019***	0.013***	0.005*	0.033***	0.021***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Emp_fulltime	-0.096**	-0.089**	-0.087*	-0.018	-0.021	-0.153***	0.048	0.020	0.050	0.130***
	(0.047)	(0.044)	(0.046)	(0.043)	(0.045)	(0.043)	(0.043)	(0.045)	(0.046)	(0.045)
Emp_parttime	-0.075	-0.102**	-0.071	-0.009	-0.012	-0.137***	0.048	0.029	0.023	0.100**
	(0.049)	(0.047)	(0.048)	(0.045)	(0.047)	(0.046)	(0.045)	(0.048)	(0.048)	(0.047)
Emp_self	-0.026	-0.141***	-0.041	-0.086*	-0.110**	-0.196***	-0.015	-0.056	-0.078*	0.026
	(0.047)	(0.045)	(0.047)	(0.044)	(0.046)	(0.044)	(0.044)	(0.046)	(0.047)	(0.046)
Emp_retired	-0.136***	-0.091*	-0.098**	-0.002	-0.035	-0.185***	-0.027	0.018	-0.0001	0.091*
	(0.049)	(0.047)	(0.048)	(0.046)	(0.047)	(0.046)	(0.045)	(0.048)	(0.049)	(0.047)
Emp_housewife	-0.183***	-0.030	-0.104**	0.037	0.022	-0.070	0.055	0.041	0.085*	0.140***
	(0.048)	(0.046)	(0.047)	(0.045)	(0.046)	(0.045)	(0.044)	(0.047)	(0.048)	(0.047)
Emp_student	-0.037	-0.086*	-0.069	-0.021	-0.058	-0.134***	-0.056	-0.017	0.007	0.044
	(0.051)	(0.048)	(0.050)	(0.047)	(0.049)	(0.047)	(0.046)	(0.049)	(0.050)	(0.049)
Emp_unemp	-0.076	-0.083*	-0.115**	-0.034	-0.066	-0.136***	-0.027	-0.025	0.021	0.049
	(0.049)	(0.047)	(0.048)	(0.045)	(0.047)	(0.046)	(0.045)	(0.048)	(0.049)	(0.047)
PIA_1	-0.051***	0.067***	0.024	0.091***	0.092***	0.072***	0.038**	0.039**	0.095***	0.081***
	(0.018)	(0.017)	(0.018)	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)
PIA_2	-0.262***	-0.006	-0.090***	-0.036**	-0.013	-0.029	0.114***	0.031*	0.052***	0.110***
	(0.019)	(0.018)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.018)
PIA_3	0.025	0.018	0.078***	0.105***	0.101***	0.067***	0.023	0.028	0.107***	-0.006
	(0.020)	(0.019)	(0.019)	(0.018)	(0.019)	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)
PIAB_1	-0.071***	0.040***	-0.023	0.040***	0.061***	0.020	0.038***	0.027*	0.046***	0.032**
	(0.016)	(0.015)	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)	(0.015)
PIAB_2	-0.218***	-0.046***	-0.103***	-0.061***	-0.028**	-0.051***	0.047***	-0.006	-0.012	0.049***
	(0.015)	(0.014)	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.015)	(0.014)
PIAB_3	-0.044***	0.009	0.026*	0.049***	0.063***	0.049***	0.005	0.029**	0.032**	-0.011
	(0.015)	(0.014)	(0.015)	(0.014)	(0.014)	(0.014)	(0.013)	(0.014)	(0.015)	(0.014)
Constant	2.457***	1.412***	2.004***	1.522***	2.120***	2.201***	1.815***	2.603***	1.776***	1.380***
	(0.361)	(0.344)	(0.356)	(0.336)	(0.346)	(0.337)	(0.331)	(0.352)	(0.358)	(0.349)
Observations	32,254	32,254	32,254	32,254	32,254	32,254	32,254	32,254	32,254	32,254
R^2	0.133	0.122	0.160	0.216	0.246	0.183	0.105	0.207	0.120	0.090
Adjusted R^2	0.132	0.121	0.159	0.215	0.245	0.182	0.104	0.206	0.119	0.089
Residual Std. Error (df = 32213)	0.865	0.824	0.854	0.804	0.830	0.807	0.794	0.843	0.858	0.837
F Statistic (df = 40; 32213)	123.892***	111.697***	153.927***	222.097***	262.414***	180.636***	94.613***	210.581***	110.275***	79.697***
Note:										*p<0.1; **p<0.05; ***p<0.01

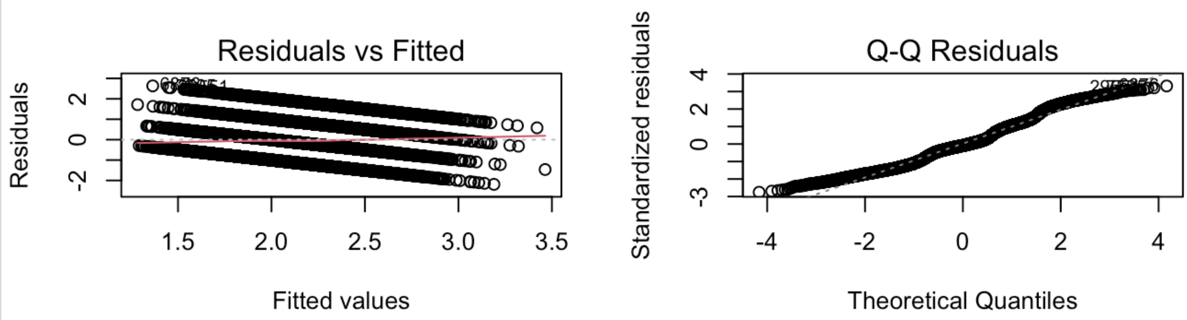
2.1.b Stargazer correlation table with selected variables

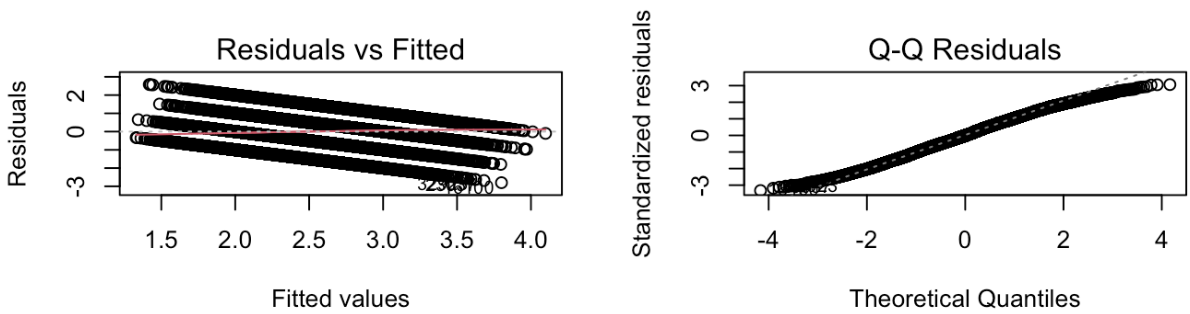
	Dependent variable:									
	CArmedForces	CUnions	CPolice	CPParties	CParliament	CCivilService	CUniversities	CElections	CBanks	CEnvOrg
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
TNeighbourhood	0.146***	0.117***	0.123***	0.122***	0.124***	0.131***	0.077***	0.115***	0.080***	0.051***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
TFamily	0.165***	0.071***	0.147***	0.067***	0.084***	0.095***	0.109***	0.059***	0.073***	0.062***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)
PDemCurrent	-0.035***	-0.021***	-0.042***	-0.022***	-0.038***	-0.031***	-0.026***	-0.057***	-0.026***	-0.026***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PSatisfied	-0.029***	-0.047***	-0.048***	-0.098***	-0.107***	-0.068***	-0.027***	-0.081***	-0.043***	-0.028***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
TMeet	-0.006	0.094***	0.038***	0.088***	0.072***	0.061***	-0.001	0.075***	0.028***	0.045***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
TPeople	0.042***	0.055***	0.132***	0.158***	0.130***	0.113***	0.023**	0.106***	0.022*	0.018
	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)
TKnow	0.066***	0.060***	0.097***	-0.003	0.029***	0.082***	0.097***	0.054***	0.073***	0.102***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
HOverall	0.048***	0.036***	0.054***	0.053***	0.056***	0.040***	0.100***	0.060***	0.118***	0.076***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
EHardWork	0.016***	0.007***	0.012***	0.004***	0.013***	0.012***	0.015***	0.010***	0.026***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PIA_2	-0.262***	-0.008	-0.101***	-0.041**	-0.017	-0.039**	0.124***	0.035*	0.050***	0.117***
	(0.019)	(0.018)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)
PIAB_2	-0.195***	-0.066***	-0.113***	-0.107***	-0.084***	-0.088***	0.041***	-0.023**	-0.052***	0.053***
	(0.012)	(0.011)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)	(0.011)
Emp_housewife	-0.064***	-0.005	-0.035**	-0.025*	-0.019	0.028**	0.008	0.021	-0.026*	-0.016
	(0.015)	(0.014)	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.014)
Emp_self	0.033**	-0.099***	0.045***	-0.118***	-0.134***	-0.076***	-0.076***	-0.086***	-0.176***	-0.116***
	(0.014)	(0.013)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)
PIA_3	0.063***	0.037**	0.066***	0.124***	0.120***	0.069***	0.036**	0.025	0.137***	0.005
	(0.019)	(0.018)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)
PIA_1	-0.026	0.061***	0.032*	0.106***	0.099***	0.076***	0.019	0.035**	0.097***	0.061***
	(0.017)	(0.016)	(0.017)	(0.016)	(0.017)	(0.016)	(0.016)	(0.017)	(0.017)	(0.017)
Constant	1.724***	2.087***	1.711***	2.548***	2.489***	1.994***	1.596***	2.360***	1.851***	1.830***
	(0.037)	(0.036)	(0.037)	(0.035)	(0.036)	(0.035)	(0.034)	(0.036)	(0.037)	(0.036)
Observations	32,254	32,254	32,254	32,254	32,254	32,254	32,254	32,254	32,254	32,254
R^2	0.119	0.107	0.151	0.201	0.230	0.171	0.086	0.200	0.099	0.070
Adjusted R^2	0.118	0.106	0.151	0.201	0.230	0.171	0.085	0.200	0.099	0.069
Residual Std. Error (df = 32238)	0.872	0.831	0.858	0.811	0.838	0.813	0.802	0.846	0.868	0.846
F Statistic (df = 15; 32238)	289.164***	256.299***	383.088***	541.987***	642.049***	444.585***	201.454***	538.077***	236.646***	161.594***
Note:										*p<0.1; **p<0.05; ***p<0.01

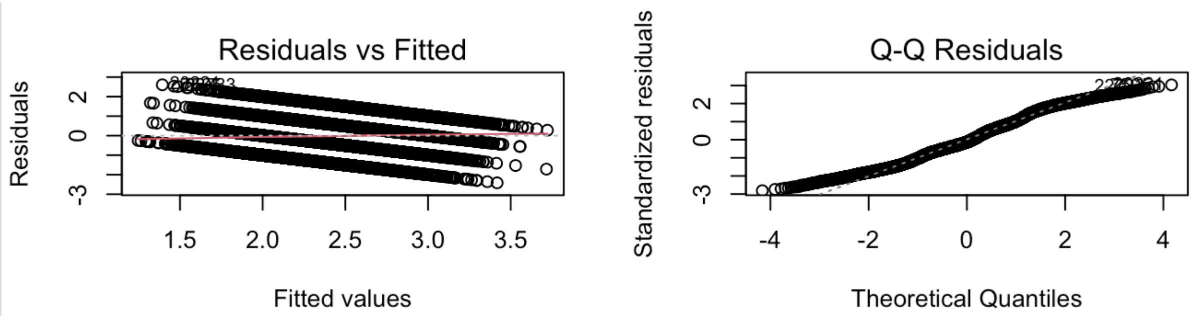
2.2.c Residuals vs Fitted and Q-Q Residual plots for other countries for each confidence variable

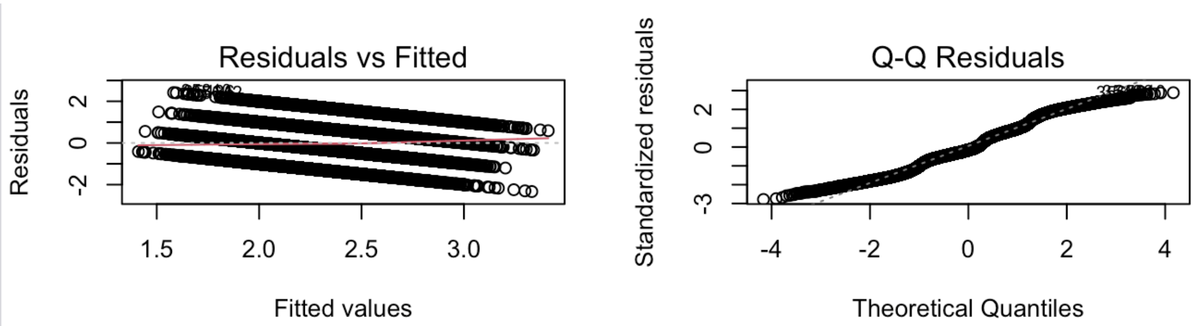



Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles









Appendix 3

3.1 External data sources used for clustering

Data	Source	Link
Democracy Score	Our World in Data	https://ourworldindata.org/grapher/democracy-index-eiu?tab=table&time=lates
GDP	World Bank	https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

Unemployment Rate	World Bank	https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?end=2023&name_desc=false&start=2023
Tertiary Education Completed	World Bank	https://data.worldbank.org/indicator/SE.PRM.UNER?end=2023&start=1970&view=chart
Total Population	World Bank	http://data.worldbank.org/indicator/SP.POP.TOTL?end=2022&start=2017
Corruption Cleanliness	Transparency	https://www.transparency.org/en/cpi/202
Freedom of Expression	Our World in Data	https://ourworldindata.org/grapher/freedom-of-expression-index
Average Annual Income	World Bank	https://api.worldbank.org/v2/en/indicator/NY.ADJ.NNTY.PC.CD?downloadformat=csv

3.2 Cleaned dataset used for clustering

Country	Average.Income	CPI	Democracy.score	Freedom.score	GDP	Population	Tertiary.rate	Unemployment.rate
AFG	449.7242928	24	0.32	0.071	17386548673	40578842	4.7	11.895
AGO	1333.401438	33	3.96	0.492	73907540671	35635029	4.676	16.132
ALB	4403.670127	36	6.41	0.713	16045842188	2777689	17.689	11.781
ARM	3717.990223	46	5.63	0.817	13939800618	2969200	53.163	17
AZE	3558.640406	23	2.87	0.152	52079869175	10141756	32.53	5.638
BEN	1038.97714	43	4.28	0.73	15349854212	13759501	5.153	1.541
BFA	659.5170533	42	3.08	0.749	17045430141	22509038	4.868	4.948
BGR	8263.037754	43	6.53	0.811	73334727523	6643324	26.729	5.044
BHR	18119.6316	44	2.52	0.132	40096307624	1524693	40.64	1.378
BIH	5135.590416	34	5	0.664	21287741406	3204802	19.239	16.34
BOL	2784.303502	31	4.51	0.756	39955975220	12077154	30.754	4.567
BTN	2762.577478	68	5.54	0.694	2672502056	780914	11	4.144
BWA	5167.505147	60	7.73	0.799	17316033958	2439892	30.653	21.196
CHL	11208.19543	67	8.22	0.917	2.86995E+11	19553036	30.141	8.473

CIV	1959.213949	37	4.22	0.661	62902224340	30395002	7.382	2.674
COL	5240.365314	39	6.72	0.84	3.17217E+11	51737944	27.01	11.526
COM	1395.10442	19	3.2	0.625	1204150804	834188	16.495	4.098
CPV	3125.997381	60	7.65	0.854	2095738294	519741	11.87	12.908
CRI	11176.83894	54	8.29	0.929	63992307115	5081765	19.594	11.74
CYP	23830.81992	52	7.38	0.831	26966897359	1331370	43.477	8.148
CZE	17385.22522	56	7.97	0.949	2.62377E+11	10672118	22.845	2.455
DOM	6957.795123	32	6.39	0.931	90186709903	11230734	18.756	6.23
ECU	4570.402316	36	5.69	0.822	1.06571E+11	17823897	14.29	4.27
EST	19222.70179	74	7.96	0.969	32994537358	1348840	49.068	5.74
FJI	4403.849611	53	5.55	0.513	5014383856	919422	27.77	4.471
GEO	3751.617543	56	5.2	0.879	18643760189	3712502	53.776	12.223
GHA	1749.058243	43	6.43	0.893	69978948922	33149152	10.073	3.26
GIN	779.6735109	25	2.32	0.561	14448803693	14055137	31.139	5.29
GMB	559.5699689	34	4.47	0.774	1836709124	2636470	10.512	5.343
GNB	629.4877993	21	2.56	0.628	1651592732	2105529	6.11	2.866
GRC	16043.20911	52	7.97	0.804	2.08255E+11	10436882	36.575	16.77
GTM	3871.083288	24	4.68	0.643	80330543643	17847877	4.51	2.52
GUY	5712.517834	40	6.34	0.792	7156763171	821637	11.331	13.694
HND	2196.188458	23	5.15	0.884	25834805048	10463872	4.131	7.523
HRV	12893.09578	50	6.5	0.804	62915118507	3855641	22.746	8.055
HUN	13231.67466	42	6.64	0.513	1.64017E+11	9644377	31.514	3.864
IRQ	3886.403482	23	3.13	0.563	2.20909E+11	44070551	16.594	14.68
JAM	4745.612096	44	7.13	0.923	15323052205	2839144	33.728	5.605
JOR	3894.208432	47	3.17	0.545	44707291596	11256263	31.54	18.405
KEN	1609.384118	32	5.05	0.874	99904678313	54252461	6.704	4.976
KGZ	949.723762	27	3.62	0.571	9166642243	6975219.5	41.828	4.051
KHM	1306.596106	24	3.18	0.2	35131614172	17201724	5.87	0.198
LAO	1903.483772	31	1.77	0.027	17871849511	7559007	12.12	2.39
LBN	5735.655936	24	3.64	0.68	39228608455	5744489	29.036	11.642
LBR	352.2913502	26	5.43	0.839	3470546250	5373294	5.376	3.353
LKA	3735.939127	36	6.47	0.707	87490596840	22181000	4.655	4.651
LSO	1025.902301	37	6.19	0.808	2345700713	2286110	6.107	17.203

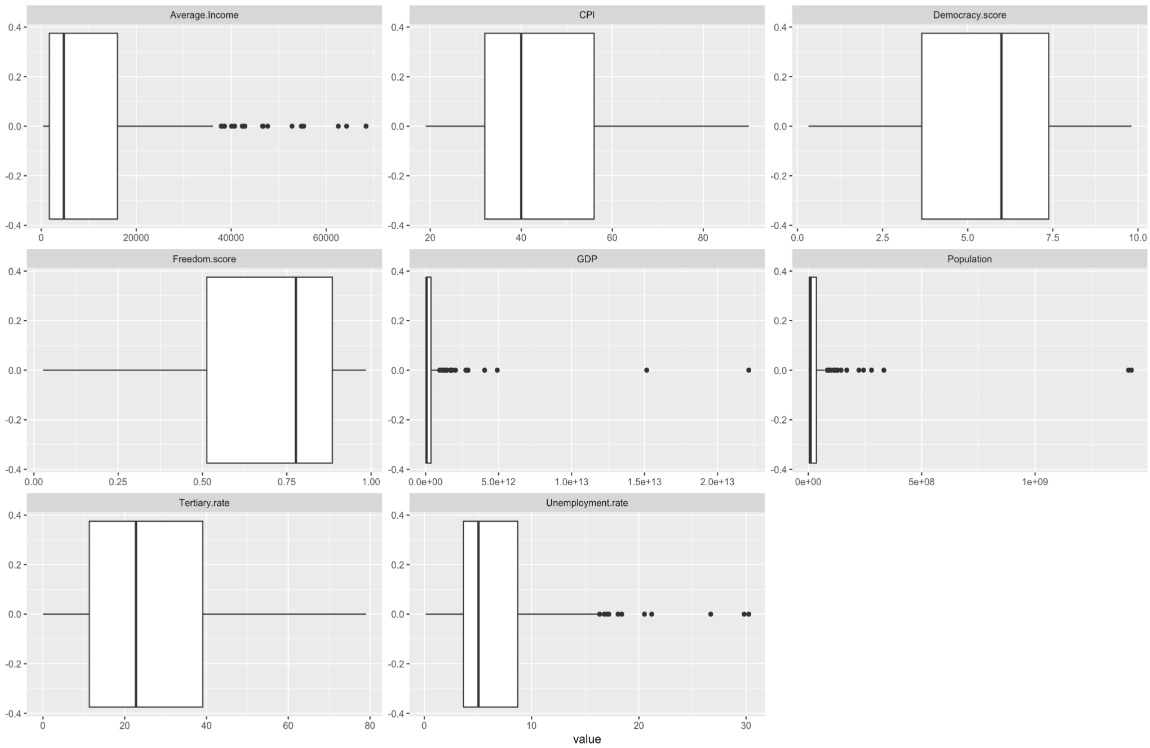
LVA	14028.62362	59	7.37	0.938	34220409691	1879383	49.459	7.477
MDA	4064.214226	39	6.23	0.884	12039443442	2528654	20.388	1.186
MDG	412.4708023	26	5.7	0.725	13963651180	30437261	8.361	2.807
MKD	5339.945193	40	6.1	0.736	12815038487	1831712	33.382	18.055
MLI	679.9127821	28	3.23	0.72	17545334987	23072640	2.92	2.714
MMR	1134.582963	23	0.74	0.044	69430519566	53756787	28.906	1.901
MNE	7610.668581	45	6.45	0.777	5461136975	617213	59.259	15.945
MNG	2817.767634	33	6.35	0.772	14101866060	3433748	59.721	6.926
MRT	1607.55507	30	4.03	0.726	8201617798	4875637	3.29	10.648
MUS	9291.276086	50	8.14	0.676	13117663939	1262523	0	7.029
MWI	503.9070218	34	5.91	0.796	11075190744	20568728	2.979	5.241
MYS	7524.235121	47	7.3	0.619	3.6032E+11	34695493	25.64	3.847
NAM	3972.660391	49	6.52	0.817	12445434413	2889662	14.094	20.525
NER	559.6024121	32	3.73	0.802	13500912594	25311973	1.384	0.569
NPL	1090.969971	34	4.49	0.837	34635122529	29715436	11.71	11.269
OMN	12494.08012	44	3.12	0.148	89266753359	4730226	18.424	2.417
PAN	12503.25079	36	6.91	0.791	67025917817	4400773	16.771	8.28
PER	5564.275738	36	5.92	0.804	2.22701E+11	33475438	26.333	4.448
PNG	1973.475566	30	5.97	0.822	25534707504	10203169	14.212	2.736
PRT	18294.44251	62	7.95	0.935	2.40941E+11	10434332	25.093	6.953
ROU	10441.97478	46	6.45	0.76	2.5671E+11	19048502	19.56	4.878
RWA	636.9021902	51	3.1	0.19	10633311382	13651030	6.983	12.76
SDN	999.2997863	22	2.47	0.419	36481076735	49383346	10.855	11.241
SEN	1268.835202	43	5.72	0.914	24531339453	17651103	5.717	3.352
SLE	434.540484	34	5.03	0.864	6617586513	8276807	5.693	3.411
SLV	3448.650071	33	5.06	0.418	27305738333	6280319	24.714	4.149
SRB	6123.748715	36	6.33	0.427	56909419188	6664449	25.671	10.65
SUR	3861.116642	40	6.95	0.885	3569208794	623164	13.837	8.091
SVK	15810.96776	53	7.07	0.942	1.08781E+11	5431752	27.914	6.701
SVN	20782.25078	56	7.75	0.872	55098988975	2112076	34.66	4.974
TCD	474.2697318	19	1.67	0.473	11241116389	18455316	3.5	1.281
TGO	771.1001411	30	2.99	0.648	7386904494	9089738	6.765	2.534
TJK	885.4924987	24	1.94	0.07	8564585092	10182222	24.558	11.772

TLS	1058.135134	42	7.06	0.801	2361586717	1369295	2.634	2.964
TTO	13865.76766	42	7.16	0.898	24623781226	1365805	20.758	3.908
TUN	3067.274063	40	5.51	0.8	43439918259	12119334	16.14	16.252
UGA	623.4410485	26	4.55	0.533	37119190793	47312719	12.553	3.313
URY	15547.81678	74	8.91	0.869	62861176767	3390913	14.788	8.78
ZMB	914.1018272	33	5.8	0.856	24148623386	20152938	10.804	5.388
ZWE	1354.80384	23	2.92	0.433	32974390594	16069056	24.637	8.126

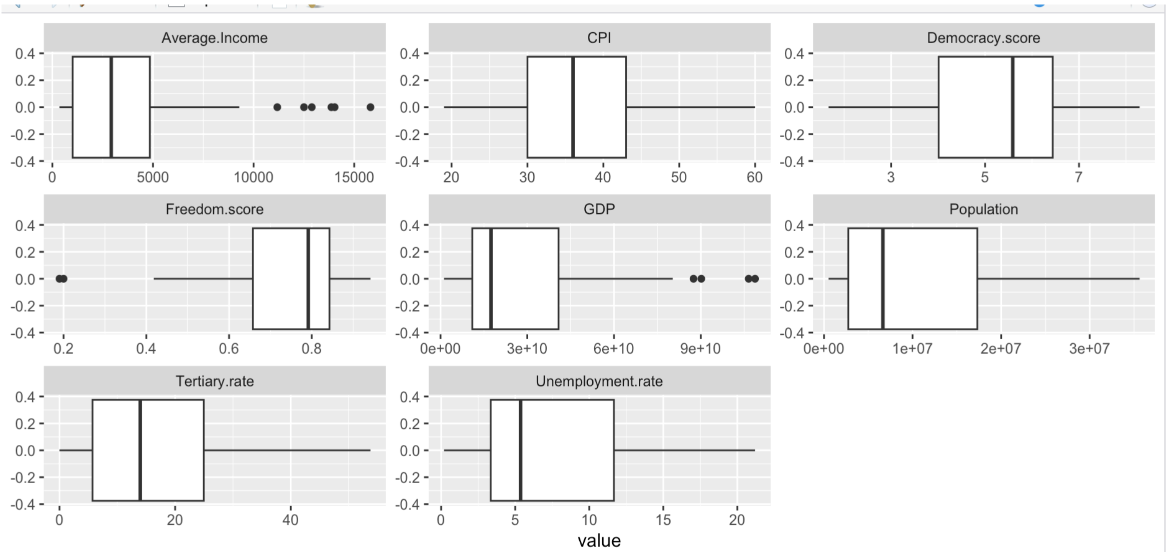
Appendix 4

4.1 Boxplots

4.1.a Below are the boxplots before data cleaning and removal of outliers.



4.1.b Below are the boxplots after outlier removal.



4.2 List of countries similar to Czech Republic (CZE)

4.2.a Below is the table consisting of a cluster of countries similar to the Czech Republic.

Country Code	Country Name
CHL	Chile
COL	Columbia
GRC	Greece
HUN	Hungary
MYS	Malaysia
PER	Peru
PRT	Portugal
ROU	Romania

4.2.b Below is the table of the clustering results

	Country	Average.Income	CPI	Democracy.score	Freedom.score	GDP	Population	Tertiary.rate	Unemployment.rate
1	CHL	1.01038142	2.15789824	1.5023907	0.9620009	2.932204	0.4405523	0.72189236	0.1966846
2	COL	−0.03287304	−0.01606376	0.7301283	0.6398593	3.318467	2.7359986	0.50357982	0.7852959
3	CZE	2.09020672	1.30384174	1.3736803	1.0958780	2.617561	−0.1928399	0.21317044	−0.9635716
4	GRC	1.85560481	0.99327574	1.3736803	0.4892476	1.925826	−0.2096171	1.17051035	1.7963268
5	HUN	1.36411196	0.21686074	0.6889410	−0.7281970	1.360413	−0.2661390	0.81762635	−0.6919198
6	MYS	0.36637716	0.60506824	1.0287364	−0.2847292	3.869381	1.5205211	0.40805500	−0.6951973
7	PER	0.02375072	−0.24898826	0.3182550	0.4892476	2.110451	1.4335061	0.45637522	−0.5793259
8	PRT	2.24914975	1.76969074	1.3633835	1.0373068	2.343581	−0.2097989	0.36991481	−0.0963678
9	ROU	0.87643607	0.52742674	0.5911210	0.3051666	2.545126	0.4045687	−0.01587993	−0.4964229

Appendix 5

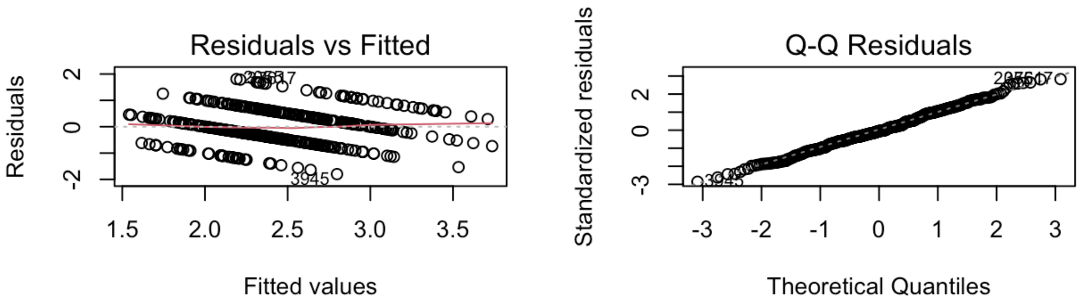
5.1 Regression results for cluster country

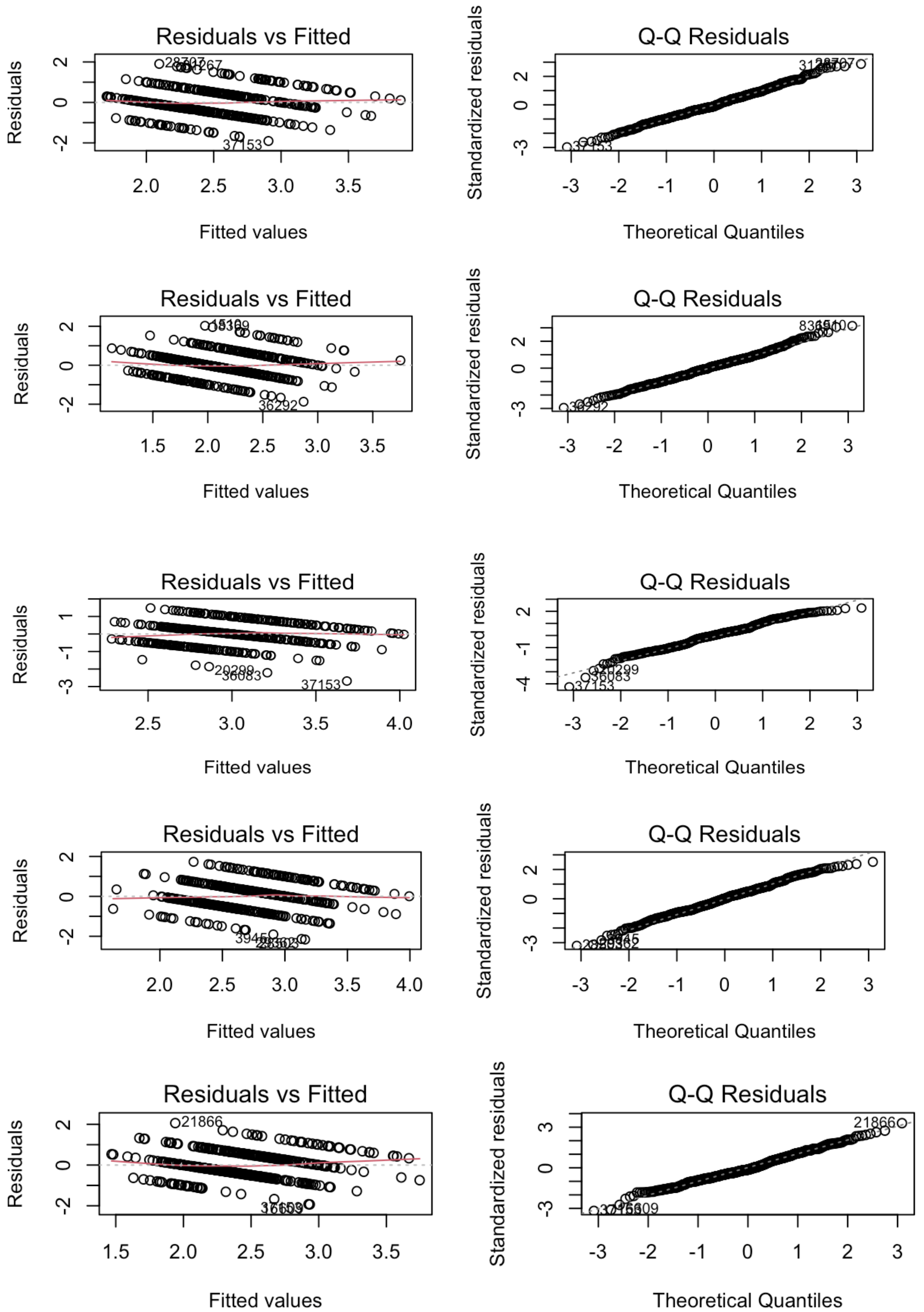
5.1.a Stargazer correlation table

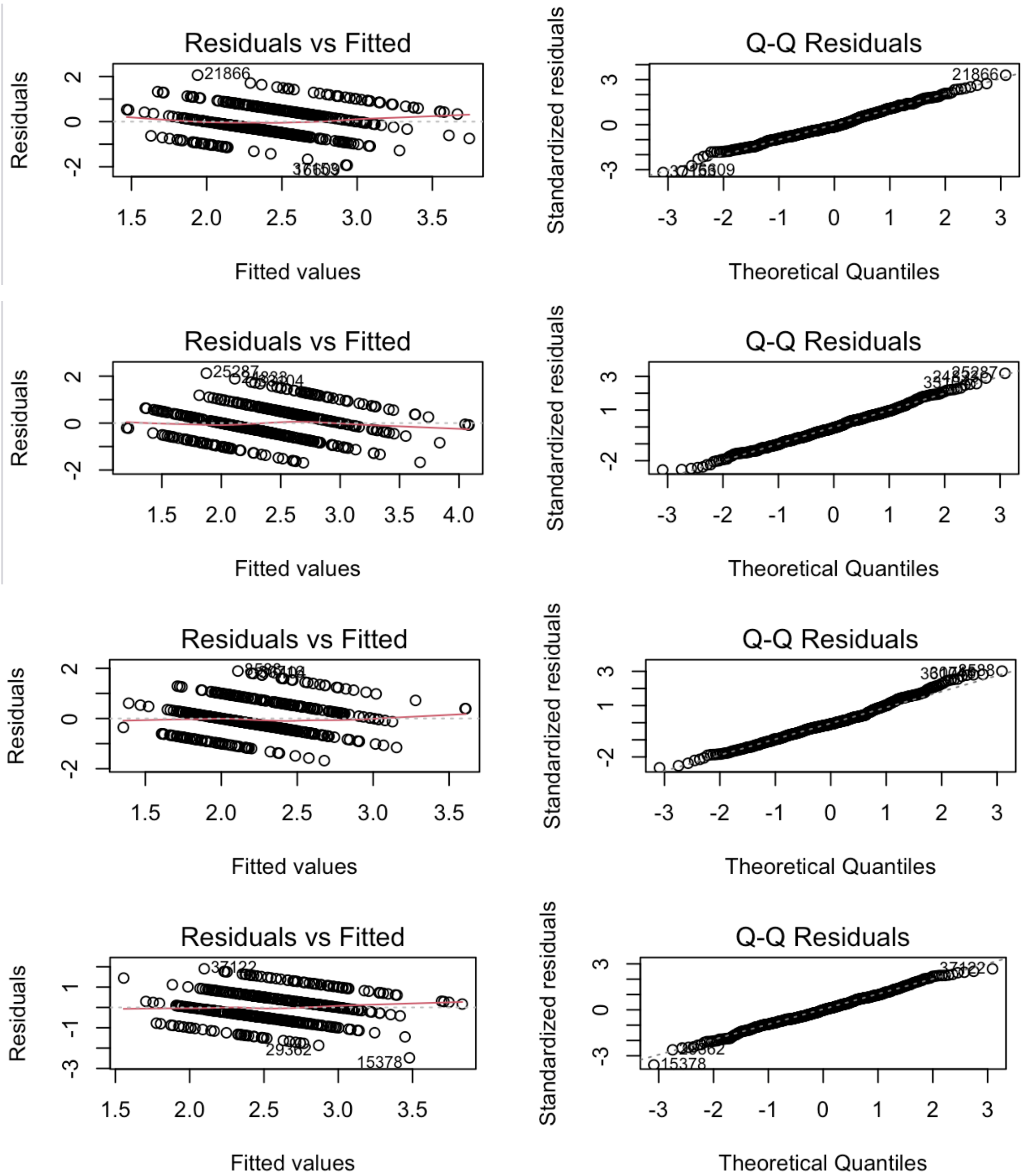
	Dependent variable:									
	CArmedForces	CUnions	CPolice	CPParties	CParliament	CCivilService	CUniversities	CElections	CBanks	CEnvOrg
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
TPeople	-0.044	0.126***	0.134***	0.082**	0.161***	0.105***	0.067	0.102**	0.095**	-0.008
	(0.045)	(0.040)	(0.043)	(0.036)	(0.038)	(0.039)	(0.042)	(0.042)	(0.043)	(0.044)
TFamily	0.194***	0.030	0.093***	-0.005	0.003	0.023	0.100***	0.001	0.048*	0.053**
	(0.026)	(0.023)	(0.025)	(0.021)	(0.022)	(0.023)	(0.024)	(0.024)	(0.025)	(0.026)
TNeighbourhood	0.073***	0.155***	0.155***	0.101***	0.114***	0.128***	0.121***	0.135***	0.151***	0.058**
	(0.023)	(0.021)	(0.022)	(0.019)	(0.020)	(0.020)	(0.022)	(0.022)	(0.023)	(0.023)
TMeet	0.026	0.091***	0.045**	0.141***	0.134***	0.081***	-0.016	0.114***	0.026	0.034
	(0.023)	(0.021)	(0.022)	(0.019)	(0.020)	(0.020)	(0.022)	(0.022)	(0.023)	(0.023)
TKnow	0.102***	0.078***	0.109***	0.067***	0.095***	0.152***	0.107***	0.056***	0.086***	0.124***
	(0.022)	(0.020)	(0.022)	(0.018)	(0.019)	(0.020)	(0.021)	(0.021)	(0.022)	(0.022)
VFamily	0.117***	0.102***	0.036	0.007	0.010	0.095**	0.026	-0.019	-0.084**	0.006
	(0.044)	(0.039)	(0.042)	(0.035)	(0.037)	(0.038)	(0.041)	(0.041)	(0.042)	(0.043)
VLeisure	-0.009	0.018	-0.007	0.008	0.020	-0.006	-0.022	0.040**	0.035*	0.032
	(0.021)	(0.019)	(0.020)	(0.017)	(0.018)	(0.019)	(0.020)	(0.020)	(0.021)	(0.021)
VWork	0.082***	-0.007	0.016	0.033*	-0.001	0.019	0.067***	0.052**	-0.031	0.039*
	(0.023)	(0.020)	(0.022)	(0.019)	(0.019)	(0.020)	(0.021)	(0.021)	(0.022)	(0.023)
HOverall	-0.011	-0.0005	0.009	0.029	0.006	-0.038*	0.052**	0.021	0.039*	0.039*
	(0.023)	(0.021)	(0.022)	(0.019)	(0.020)	(0.020)	(0.022)	(0.022)	(0.023)	(0.023)
HChoice	0.008	-0.003	0.009	-0.0004	0.001	0.009	-0.001	0.004	-0.010	-0.026***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
HShelter	-0.015	0.020	-0.017	0.048***	0.040**	-0.038**	-0.035*	-0.022	-0.019	-0.010
	(0.021)	(0.019)	(0.020)	(0.017)	(0.018)	(0.018)	(0.020)	(0.020)	(0.020)	(0.021)
EEquality	-0.013**	0.014***	-0.010*	0.0005	-0.001	-0.004	-0.0003	0.015***	-0.006	0.002
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
EGovernment	-0.004	-0.003	0.011**	-0.003	-0.003	0.001	0.004	0.010*	-0.012**	0.002
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
EHardWork	-0.002	-0.008*	-0.004	-0.007	-0.008*	-0.014***	0.004	-0.005	-0.007	0.006
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
SEducation	0.004	0.020	-0.031**	-0.001	-0.002	-0.012	0.003	-0.043***	0.003	0.063***
	(0.014)	(0.013)	(0.014)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)
STBetter	-0.027***	-0.019***	-0.015**	-0.003	-0.009*	-0.025***	-0.032***	-0.022***	-0.013**	-0.011*
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
PRadio	0.016	0.031***	0.020**	0.012	0.020**	0.023***	0.024***	0.045***	0.044***	0.029***
	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)
PEmail	-0.022*	-0.001	-0.029**	0.021**	0.025**	0.002	0.001	0.009	0.013	0.023*
	(0.012)	(0.011)	(0.012)	(0.010)	(0.010)	(0.011)	(0.012)	(0.011)	(0.012)	(0.012)
PInternet	0.009	0.031**	0.041***	0.014	0.042***	0.073***	0.024	0.046***	0.047***	0.021
	(0.016)	(0.014)	(0.015)	(0.013)	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)	(0.016)
PSocial	0.012	0.015	-0.019	0.017	-0.006	-0.010	-0.006	-0.027*	0.005	0.014
	(0.015)	(0.013)	(0.014)	(0.012)	(0.013)	(0.013)	(0.014)	(0.014)	(0.015)	(0.015)

PDemImp	-0.012	0.013*	-0.002	0.011*	0.009	0.010	0.003	0.006	0.015**	-0.015**
	(0.008)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
PDemCurrent	-0.036***	-0.023***	-0.035***	-0.010*	-0.024***	-0.022***	-0.025***	-0.046***	-0.025***	-0.012*
	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)
PSatisfied	-0.005	-0.030***	-0.037***	-0.053***	-0.061***	-0.058***	-0.011*	-0.061***	-0.030***	-0.022***
	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
MF_male	-0.074**	0.023	0.075**	0.043*	0.056**	0.075***	0.055*	0.011	0.007	0.029
	(0.031)	(0.028)	(0.030)	(0.026)	(0.027)	(0.027)	(0.029)	(0.029)	(0.030)	(0.031)
MF_female										
MF_others										
Age	-0.001	0.0002	0.001	0.002	0.001	0.001	-0.002*	-0.002*	0.001	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Edu	0.014	0.020**	0.028***	0.028***	0.035***	0.028***	-0.002	-0.002	0.036***	0.023**
	(0.009)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
Emp_fulltime	-0.058	-0.093	-0.253*	0.052	0.083	-0.079	-0.230*	-0.031	0.109	0.259*
	(0.147)	(0.132)	(0.142)	(0.120)	(0.125)	(0.129)	(0.138)	(0.137)	(0.143)	(0.146)
Emp_parttime	0.017	0.085	-0.263*	0.184	0.166	0.077	-0.318**	0.043	0.139	0.273*
	(0.160)	(0.144)	(0.154)	(0.130)	(0.136)	(0.140)	(0.150)	(0.149)	(0.155)	(0.159)
Emp_self	-0.004	-0.007	-0.182	0.092	0.147	0.053	-0.253*	-0.001	0.034	0.243
	(0.151)	(0.136)	(0.145)	(0.123)	(0.128)	(0.132)	(0.141)	(0.141)	(0.146)	(0.150)
Emp_retired	-0.237	-0.110	-0.384**	0.008	0.033	-0.158	-0.275*	-0.071	0.162	0.178
	(0.155)	(0.139)	(0.149)	(0.127)	(0.132)	(0.136)	(0.145)	(0.145)	(0.151)	(0.154)
Emp_housewife	-0.077	0.065	-0.179	0.124	0.205	0.080	-0.159	0.026	0.126	0.242
	(0.152)	(0.137)	(0.147)	(0.124)	(0.130)	(0.133)	(0.143)	(0.142)	(0.148)	(0.151)
Emp_student	-0.086	-0.067	-0.276*	0.125	0.116	0.008	-0.467***	-0.175	0.085	0.102
	(0.161)	(0.144)	(0.155)	(0.131)	(0.136)	(0.140)	(0.150)	(0.150)	(0.156)	(0.159)
Emp_unemp	-0.191	-0.024	-0.339**	0.128	0.144	0.060	-0.274*	-0.015	0.210	0.414***
	(0.158)	(0.142)	(0.152)	(0.129)	(0.134)	(0.138)	(0.148)	(0.147)	(0.153)	(0.157)
PIA_1	0.054	-0.061	0.038	0.026	-0.037	0.040	-0.042	-0.036	0.028	-0.057
	(0.056)	(0.050)	(0.054)	(0.046)	(0.047)	(0.049)	(0.052)	(0.052)	(0.054)	(0.055)
PIA_2	-0.180***	-0.178***	-0.054	-0.132***	-0.213***	-0.078	-0.048	-0.012	-0.086	-0.147**
	(0.060)	(0.054)	(0.058)	(0.049)	(0.051)	(0.053)	(0.056)	(0.056)	(0.058)	(0.060)
PIA_3	0.063	-0.055	0.065	0.013	-0.020	0.050	0.023	0.112**	-0.001	-0.107*
	(0.059)	(0.053)	(0.057)	(0.048)	(0.050)	(0.052)	(0.055)	(0.055)	(0.057)	(0.059)
PIAB_1	0.066	-0.097**	-0.010	0.038	-0.032	-0.018	-0.034	-0.056	-0.048	-0.061
	(0.052)	(0.046)	(0.050)	(0.042)	(0.044)	(0.045)	(0.048)	(0.048)	(0.050)	(0.051)
PIAB_2	-0.120***	-0.104**	-0.083*	-0.035	-0.099**	-0.092**	-0.013	-0.008	-0.055	-0.052
	(0.047)	(0.042)	(0.045)	(0.038)	(0.040)	(0.041)	(0.044)	(0.043)	(0.045)	(0.046)
PIAB_3	0.047	-0.071*	0.034	0.010	-0.013	0.002	-0.018	0.062	-0.063	-0.080*
	(0.044)	(0.039)	(0.042)	(0.036)	(0.037)	(0.038)	(0.041)	(0.041)	(0.043)	(0.043)
Constant	2.159***	1.681***	1.992***	1.798***	1.752***	2.011***	1.992***	2.235***	1.746***	1.531***
	(0.240)	(0.215)	(0.231)	(0.196)	(0.204)	(0.210)	(0.225)	(0.224)	(0.233)	(0.238)
Observations	3,316	3,316	3,316	3,316	3,316	3,316	3,316	3,316	3,316	3,316
R^2	0.132	0.186	0.184	0.195	0.257	0.270	0.122	0.238	0.154	0.115
Adjusted R^2	0.122	0.177	0.174	0.186	0.249	0.261	0.111	0.229	0.144	0.104
Residual Std. Error (df = 3276)	0.839	0.754	0.808	0.684	0.713	0.733	0.786	0.782	0.814	0.832
F Statistic (df = 39; 3276)	12.776***	19.234***	18.910***	20.369***	29.116***	30.994***	11.649***	26.200***	15.331***	10.862***
Note:										*p<0.1; **p<0.05; ***p<0.01

5.1.b Residuals vs Fitted and Q-Q Residual plots for focus country for each confidence variable




Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles



Appendix 6: AI Acknowledgement

Tool	Information Generated	Prompt	How is it used?
ChatGPT 4.0	Understanding Institutional Confidence: Cross-National Insights with a Focus on the Czech Republic Institutional Trust Across Borders: A Comparative Model-Based Approach	(copy pasted important parts of the report) This is my current title: “Analysis on different models related to institutional confidence”. Please make a nicer title.	I used generative AI to make a more interesting title. I did a minor modification from the suggestions generated.

	Predicting Institutional Confidence: A Clustering and Regression Analysis Centered on the Czech Republic Exploring Predictors of Institutional Trust: A Regional and Global Comparison Mapping Institutional Confidence: How Social, Political, and Economic Factors Influence Trust Institutional Confidence in Context: A Comparative Study Using Country Clustering Modeling Institutional Trust: National Insights and International Parallels From Local to Global: A Data-Driven Study of Institutional Confidence Unpacking Institutional Confidence: Comparing the Czech Republic to Its Global Peers Bridging Contexts: How Similar Countries Explain Trust in Institutions		
	🧠 **What's going wrong:** You're **converting `-2` to `NA`**, but then immediately trying to **filter rows with `NA` using `if_all()`**, expecting it to drop all rows containing any NA. That part should work. But the issue is **probably that some columns like `MF` or `Country` were left out of the conversion**, or not numeric, so they might still contain `"NA"` as a **string**, or `-2` as a character,	(copy and pasted the relevant code fragment). I have checked that my code follows the correct logic, why is it still not deleting the -2 data.	I used generative AI to help me in the debugging process, especially when the root cause of the error is unforeseen. I ended up not using the code fragment generated as I did a minor modification from the suggestion generated.

<table>
<tr>
<td></td>
<td>depending on how the earlier transformation worked.
Let me walk you through a clear and safe version of your code to make sure all <code>-2</code> are properly removed:
<pre>
library(dplyr)

# Convert all -2 to NA (including MF column if needed)
VC = VC %>%
  mutate(across(everything(), ~ ifelse(. == -2, NA, .)))

# Drop all rows with any NA
VC = VC %>%
  filter(if_all(everything(), ~ !is.na(.)))

Double check
sum(is.na(VC)) # Should return 0
</pre></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><pre>
extract_p_values_and_plot <- function(models, title) {
  p_matrix = sapply(models, function(model) coef(summary(model))[, "Pr(>|t|)"])
  p_matrix = as.data.frame(p_matrix)
  p_matrix$Variable = rownames(p_matrix)
  p_long = melt(p_matrix, id.vars = "Variable")
</pre></td>
<td>Please generate a function to create a heatmap that takes the p-value from a regression model result.</td>
<td>I know how to generate a heatmap but I do not know how to generate one that takes the p-value. So, I use it to generate the heatmap that takes the p-value to better visualise the regression result.</td>
</tr>
</table>

	ggplot(p_long, aes(x = variable, y = Variable, fill = value)) +
	    geom_tile() +
	
	scale_fill_gradient2(low = "red", mid = "red", high = "white", midpoint = 0) +
	    theme_minimal() +
	    labs(title = title, x = "Response Variables", y = "Predictor Variables") +
	
	theme(axis.text.x = element_text(angle = 45, hjust = 1))
	}		

Appendix 7: R Codes

Question 1

```
#import libraries
library(ggplot2)
library(dplyr)
library(tidyr)
library(stargazer)
library(reshape2)
library(scales)

#Q1
#Setup
rm(list = ls())
set.seed(33410712)
VCData =
read.csv("/Users/paramithaaa/Documents/Monash/Academics/FIT3152/Assi
gnment 1/WVSExtract.csv")
VC = VCData[sample(1:nrow(VCData),50000, replace=FALSE),]
VC = VC[,c(1:6, sort(sample(7:46,17, replace = FALSE)), 47:53,
sort(sample(54:69,10, replace = FALSE)))]

#Overview
nrow(VC) #50000 rows
ncol(VC) #40 columns
str(VC)
```

```
#Print Variable Types
variable_types = data.frame(
  Variable = names(VC),
  Data_Type = sapply(VC, class)
)
print(variable_types)

#Cleaning
#Check if not answered (-2) is random
no_answer = sapply(VC, function(x) sum(x == -2, na.rm = TRUE))
no_answer_proportion = no_answer / nrow(VC)
no_answer_proportion

subset = VC[VC$Country == "CZE", ]
no_answer = sapply(subset, function(x) sum(x == -2, na.rm = TRUE))
no_answer_proportion = no_answer / nrow(subset)
no_answer_proportion

#missing values
missing_codes = sapply(VC, function(x) sum(x %in% c(-2, -3, -4,
-5)))
missing_codes
VC = VC %>% mutate(MF = ifelse(MF == -5, "Others", MF))

# Delete N/A
VC = VC %>% filter(if_all(everything(), ~ !.x %in% c(-2, -3, -4,
-5)))
VC = VC %>% mutate(across(-c(MF, Country), ~
as.integer(as.character(.))))
str(VC)
missing_codes = sapply(VC, function(x) sum(x %in% c(-2, -3, -4,
-5)))
missing_codes

#Calculate "don't know" or "-1" values
unique_values = sapply(VC, function(x) sum(x == -1, na.rm = TRUE))
unique_values
cols = c("MF", "Age", "Edu", "Employment")
VC = VC %>% filter(if_all(all_of(cols), ~ . != -1))

# Remove rows where PIA == PIAB (mutually exclusive)
VC = VC %>% filter(PIA != PIAB)

#Gender
VC = VC %>% mutate(MF = ifelse(MF == 1, "Male", ifelse(MF == 2,
"Female", ifelse(MF == -5, "Others", MF))))
VC_gender = data.frame(table(VC$MF))
VC_gender

ggplot(VC_gender, aes(x = Var1, y = Freq, fill = Var1)) +
  geom_bar(stat = "identity", width = 0.6) +
  scale_fill_manual(values = c("Others" = "grey", "Male" = "blue",
"Female" = "purple")) +
  theme_minimal() +
  labs(title = "Gender Distribution", x = "Gender", y = "Count") +
  theme(legend.position = "none",
        axis.text = element_text(size = 12),
```

```
        plot.title = element_text(size = 14, face = "bold"))

#Age
VC$Age = as.numeric(VC$Age)
summary(VC$Age)
ggplot(VC, aes(x = Age, fill = factor(MF))) +
geom_histogram(binwidth = 2, position = "dodge", alpha = 0.5) +
  labs(title = "Age Distribution of Survey Respondents by Gender", x
= "Age", y = "Count", fill = "Gender") +
  scale_fill_manual(values = c("Others" = "grey", "Male" = "blue",
"Female" = "purple")) +
  theme_minimal()

#Education
edu_labels = c(
  "0. Early childhood / No education",
  "1. Primary education",
  "2. Lower secondary",
  "3. Upper secondary",
  "4. Post-secondary non-tertiary",
  "5. Short-cycle tertiary",
  "6. Bachelor or equivalent",
  "7. Master or equivalent",
  "8. Doctoral or equivalent"
)

VC_Edu = data.frame(table(VC$Edu))
VC_Edu$Edu_Label = factor(VC_Edu$Var1, levels = 0:8, labels =
edu_labels)

ggplot(VC_Edu, aes(x = Edu_Label, y = Freq, fill = Edu_Label)) +
  geom_bar(stat = "identity", width = 0.7, color = "black") +
  scale_fill_brewer(palette = "Blues") +
  theme_minimal() +
  labs(title = "Distribution of Education Levels", x = "Education
Level", y = "Count") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size =
14), legend.position = "none", plot.title = element_text(size = 14,
face = "bold"))

#Employment
employment_labels = c(
  "1. Full time",
  "2. Part time",
  "3. Self employed",
  "4. Retired/pensioned",
  "5. Housewife",
  "6. Student",
  "7. Unemployed",
  "8. Other"
)

VC_Employment = data.frame(table(VC$Employment))
VC_Employment$Employment_Label = factor(VC_Employment$Var1, levels =
1:8, labels = employment_labels)

ggplot(VC_Employment, aes(x = Employment_Label, y = Freq, fill =
```

```
Employment_Label)) +
  geom_bar(stat = "identity", width = 0.7, color = "black") +
  scale_fill_brewer(palette = "Blues") +
  theme_minimal() +
  labs(title = "Distribution of Employment Type", x = "Employment
Type", y = "Count") +
  theme(
    axis.text.x = element_text(angle = 90, hjust = 1, size = 14),
    legend.position = "none",
    plot.title = element_text(size = 14, face = "bold")
  )

#Get mode
get_mode <- function(x) {
  uniq_v <- unique(x)
  uniq_v[which.max(tabulate(match(x, uniq_v)))]
}

summarize <- function(df, vars, category_name) {
  cat(paste0("\n--- ", category_name, " ---\n"))

  stats <- lapply(vars, function(var) {
    values <- df[[var]]
    data.frame(
      Variable = var,
      Mean = round(mean(values, na.rm = TRUE), 2),
      Median = median(values, na.rm = TRUE),
      IQR = IQR(values, na.rm = TRUE),
      Min = min(values, na.rm = TRUE),
      Max = max(values, na.rm = TRUE),
      Mode = get_mode(values),
      Category = category_name
    )
  })

  summary_table = do.call(rbind, stats)
  return(summary_table)

  modes = sapply(df[vars], get_mode)
  return(data.frame(Variable = vars, Mode = modes, Category =
category_name))
}

# Boxplots
plot_boxplots <- function(df, vars, title) {
  df_long <- df %>%
    pivot_longer(cols = all_of(vars), names_to = "Variable",
values_to = "Value") %>%
    mutate(Variable = reorder(Variable, Value, FUN = median))

  ggplot(df_long, aes(x = Variable, y = Value, fill = Variable)) +
    geom_boxplot() +
    stat_summary(fun = median, geom = "crossbar", width = 0.5, color
= "orange", size = 0.4) +
    scale_fill_brewer(palette = "Blues") +
    theme_minimal() +
    labs(title = title, x = "", y = "Values") +
```

```
    theme(axis.text.x = element_text(angle = 45, hjust = 1),
legend.position = "none")
}

# Variables per category
confidence_vars = c("CArmedForces", "CUnions", "CPolice",
"CPParties", "CParliament",
                     "CCivilService", "CUniversities", "CElections",
"CBanks", "CEnvOrg")
interpersonal_vars = c("TPeople", "TFamily", "TNeighbourhood",
"TKnow", "TMeet")
importance_vars = c("VFamily", "VLeisure", "VWork")
wellbeing_vars = c("HOverall", "HChoice", "HShelter")
economic_vars = c("EEquality", "EGovernment", "EHardWork")
education_security_vars = c("SEducation")
priority_vars = c("PIA", "PIAB")
science_vars = c("STBetter")
news_source_vars = c("PRadio", "PEmail", "PInternet", "PSocial")
democracy_vars = c("PDemImp", "PDemCurrent", "PSatisfied")

# Run summaries + get modes + plot for each category

summary = summarize(VC, confidence_vars, "Confidence")
plot_boxplots(VC, confidence_vars, "Confidence in Institutions")

summary = summarize(VC, interpersonal_vars, "Interpersonal Trust")
plot_boxplots(VC, interpersonal_vars, "Interpersonal Trust")

summary = summarize(VC, importance_vars, "Importance in Life")
plot_boxplots(VC, importance_vars, "Importance in Life")

summary = summarize(VC, wellbeing_vars, "Well-being")
plot_boxplots(VC, wellbeing_vars, "Well-being")

summary = summarize(VC, economic_vars, "Economic Attitudes")
plot_boxplots(VC, economic_vars, "Economic Attitudes")

summary = summarize(VC, education_security_vars, "Education
Security")
plot_boxplots(VC, education_security_vars, "Education Security")

summary = summarize(VC, priority_vars, "Priorities")
plot_boxplots(VC, priority_vars, "Priorities")

summary = summarize(VC, science_vars, "Trust in Science")
plot_boxplots(VC, science_vars, "Trust in Science")

summary = summarize(VC, news_source_vars, "News Sources")
plot_boxplots(VC, news_source_vars, "News Sources")

summary = summarize(VC, democracy_vars, "Democracy")
plot_boxplots(VC, democracy_vars, "Democracy")
```

Question 2

```
#import libraries
library(ggplot2)
library(dplyr)
library(tidyr)
library(stargazer)
library(reshape2)
library(scales)

#Q2a
focus_country_data = VC %>% filter(Country == "CZE")
other_countries_data = VC %>% filter(Country != "CZE")

#Functions to compute and plot distributions
compute_distribution <- function(var_names, df_focus, df_other,
group_label) {
  bind_rows(
    bind_rows(lapply(var_names, function(var) {
      dist = prop.table(table(df_focus[[var]])) * 100
      df = as.data.frame(dist)
      df$Country_Group = "CZE"
      df$Variable = var
      colnames(df) = c("Level", "Percentage", "Country_Group",
group_label)
      df
    })),
    bind_rows(lapply(var_names, function(var) {
      dist = prop.table(table(df_other[[var]])) * 100
      df = as.data.frame(dist)
      df$Country_Group = "Other"
      df$Variable = var
      colnames(df) = c("Level", "Percentage", "Country_Group",
group_label)
      df
    }))
  )
}

plot_distribution <- function(data, level_col, group_col, facet_col,
title, xlab) {
  data[[level_col]] = as.factor(data[[level_col]])

  ggplot(data, aes_string(x = level_col, y = "Percentage", fill =
"Country_Group")) +
    geom_col(position = "dodge", alpha = 0.8) +
    scale_y_continuous(labels = percent_format(scale = 1)) +
    scale_fill_manual(values = c("CZE" = "darkblue", "Other" =
"lightblue")) +
    labs(title = title,
         x = xlab,
         y = "Percentage of Responses",
         fill = "Country Group") +
    theme_minimal() +
    facet_wrap(as.formula(paste("~", facet_col)), scales = "free_y")
}

# Education
edu_labels = c(
```

```
  "0" = "ISCED 0: Early childhood / No education",
  "1" = "ISCED 1: Primary",
  "2" = "ISCED 2: Lower secondary",
  "3" = "ISCED 3: Upper secondary",
  "4" = "ISCED 4: Post-secondary non-tertiary",
  "5" = "ISCED 5: Short-cycle tertiary",
  "6" = "ISCED 6: Bachelor or equivalent",
  "7" = "ISCED 7: Master or equivalent",
  "8" = "ISCED 8: Doctoral or equivalent"
)
edu_distribution = compute_distribution("Edu", focus_country_data,
other_countries_data, "Variable")
edu_distribution$Country_Group =
factor(edu_distribution$Country_Group)
edu_distribution$Level = as.character(edu_distribution$Level)
edu_distribution$Level = factor(edu_distribution$Level, levels =
names(edu_labels), labels = edu_labels)

ggplot(edu_distribution, aes(x = Country_Group, y = Percentage, fill
= as.factor(Level))) +
  geom_bar(stat = "identity", position = "fill") +
  scale_y_continuous(labels = percent_format()) +
  scale_fill_brewer(palette = "Blues") +
  labs(title = "Education Level Distribution by Country Group",
       x = "Country Group", y = "Percentage", fill = "Education
Level") +
  theme_minimal()

# Employment
emp_labels = c(
  "1" = "Full time (30 hours a week or more)",
  "2" = "Part time (less than 30 hours a week)",
  "3" = "Self employed",
  "4" = "Retired/pensioned",
  "5" = "Housewife not otherwise employed",
  "6" = "Student",
  "7" = "Unemployed",
  "8" = "Other"
)
employment_distribution = compute_distribution("Employment",
focus_country_data, other_countries_data, "Variable")
employment_distribution$Level =
as.character(employment_distribution$Level)
employment_distribution$Level =
factor(employment_distribution$Level, levels = names(emp_labels),
labels = emp_labels)
ggplot(employment_distribution, aes(x = Country_Group, y =
Percentage, fill = as.factor(Level))) +
  geom_bar(stat = "identity", position = "fill") +
  scale_y_continuous(labels = percent_format()) +
  scale_fill_brewer(palette = "Blues") +
  labs(title = "Employment Type Distribution by Country Group",
       x = "Country Group", y = "Percentage", fill = "Employment
Type") +
  theme_minimal()

# Trust
```

```
trust_vars = c("TPeople", "TFamily", "TNeighbourhood", "TKnow",
"TMeet")
trust_distribution = compute_distribution(trust_vars,
focus_country_data, other_countries_data, "Trust_Variable")
plot_distribution(trust_distribution, "Level", "Country_Group",
"Trust_Variable",
                  "Trust Levels: CZE vs. Other Countries", "Trust
Level")

# Importance
importance_vars = c("VFamily", "VLeisure", "VWork")
importance_distribution = compute_distribution(importance_vars,
focus_country_data, other_countries_data, "Importance_Variable")
plot_distribution(importance_distribution, "Level", "Country_Group",
"Importance_Variable",
                  "Importance in Life: CZE vs. Other Countries",
"Importance Level")

# Well-being
wellbeing_vars = c("HOverall", "HChoice", "HShelter")
wellbeing_distribution = compute_distribution(wellbeing_vars,
focus_country_data, other_countries_data, "Wellbeing_Variable")
plot_distribution(wellbeing_distribution, "Level", "Country_Group",
"Wellbeing_Variable",
                  "Well-being Levels: CZE vs. Other Countries",
"Well-being Level")

# Economic
economic_vars = c("EEquality", "EGovernment", "EHardWork")
economic_distribution = compute_distribution(economic_vars,
focus_country_data, other_countries_data, "Economic_Variable")
plot_distribution(economic_distribution, "Level", "Country_Group",
"Economic_Variable",
                  "Economic Views: CZE vs. Other Countries",
"Economic Level")

# Secure Education
secure_edu_vars = "SEducation"
secure_edu_distribution = compute_distribution(secure_edu_vars,
focus_country_data, other_countries_data, "Secure_Edu_Variable")
plot_distribution(secure_edu_distribution, "Level", "Country_Group",
"Secure_Edu_Variable",
                  "Security of Education: CZE vs. Other Countries",
"Security Level")

# Priority
priority_vars = c("PIA", "PIAB")
priority_distribution <- compute_distribution(priority_vars,
focus_country_data, other_countries_data, "Priority_Variable")
plot_distribution(priority_distribution, "Level", "Country_Group",
"Priority_Variable",
                  "Priorities: CZE vs. Other Countries", "Priority
Level")

# Science & Technology
st_vars = "STBetter"
st_distribution = compute_distribution(st_vars, focus_country_data,
```

```
other_countries_data, "St_Variable")
plot_distribution(st_distribution, "Level", "Country_Group",
"St_Variable",
                  "Science & Technology Trust: CZE vs. Other
Countries", "Trust Level")

# News
news_vars = c("PRadio", "PEmail", "PInternet", "PSocial")
news_distribution = compute_distribution(news_vars,
focus_country_data, other_countries_data, "News_Variable")
plot_distribution(news_distribution, "Level", "Country_Group",
"News_Variable",
                  "News Sources: CZE vs. Other Countries",
"Frequency Level")

# Democratic Institutions
democratic_vars = c("PDemImp", "PDemCurrent", "PSatisfied")
democratic_distribution <- compute_distribution(democratic_vars,
focus_country_data, other_countries_data, "Democratic_Variable")
plot_distribution(democratic_distribution, "Level", "Country_Group",
"Democratic_Variable",
                  "Democracy Satisfaction: CZE vs. Other Countries",
"Level")

# Confidence in Institutions
confidence_vars = c("CArmedForces", "CUnions", "CPolice",
"CPParties", "CParliament", "CCivilService", "CUniversities",
"CElections", "CBanks", "CEnvOrg")
confidence_distribution = compute_distribution(confidence_vars,
focus_country_data, other_countries_data, "Confidence_Variable")
plot_distribution(confidence_distribution, "Level", "Country_Group",
"Confidence_Variable",
                  "Confidence in Institutions: CZE vs. Other
Countries", "Confidence Level")

#Q2b and c

#One-hot encoding
VC = VC %>%
  mutate(
    MF_male = ifelse(MF == "Male", 1, 0),
    MF_female = ifelse(MF == "Female", 1, 0),
    MF_others = ifelse(MF == "Others", 1, 0),

    Emp_fulltime = ifelse(Employment == 1, 1, 0),  # Full time
    Emp_parttime = ifelse(Employment == 2, 1, 0),  # Part time
    Emp_self = ifelse(Employment == 3, 1, 0),  # Self employed
    Emp_retired = ifelse(Employment == 4, 1, 0),  #
Retired/pensioned
    Emp_housewife = ifelse(Employment == 5, 1, 0),  # Housewife
    Emp_student = ifelse(Employment == 6, 1, 0),  # Student
    Emp_unemp = ifelse(Employment == 7, 1, 0),  # Unemployed
    Emp_other = ifelse(Employment == 8, 1, 0),   # Other

    PIA_1 = ifelse(PIA == 1, 1, 0),   # Economic growth
    PIA_2 = ifelse(PIA == 2, 1, 0),   # Strong defence
    PIA_3 = ifelse(PIA == 3, 1, 0),   # Public say in
```

```
jobs/communities
    PIA_4 = ifelse(PIA == 4, 1, 0),   # Beautiful country

    PIAB_1 = ifelse(PIAB == 1, 1, 0), # Economic growth
    PIAB_2 = ifelse(PIAB == 2, 1, 0), # Strong defence
    PIAB_3 = ifelse(PIAB == 3, 1, 0), # Public say in
jobs/communities
    PIAB_4 = ifelse(PIAB == 4, 1, 0)  # Beautiful country
  )

variables = c("TPeople", "TFamily", "TNeighbourhood", "TMeet",
"TKnow",
              "VFamily", "VLeisure", "VWork", "HOverall", "HChoice",
"HShelter",
              "EEquality", "EGovernment", "EHardWork", "SEducation",
"STBetter",
              "PRadio", "PEmail", "PInternet", "PSocial", "PDemImp",
              "PDemCurrent", "PSatisfied",
              "MF_male", "MF_female", "MF_others",
              "Age", "Edu",
              "Emp_fulltime", "Emp_parttime", "Emp_self",
"Emp_retired",
              "Emp_housewife", "Emp_student", "Emp_unemp",
              "PIA_1", "PIA_2", "PIA_3",
              "PIAB_1", "PIAB_2", "PIAB_3")

response_vars = c("CArmedForces", "CUnions", "CPolice", "CPParties",
"CParliament",
                  "CCivilService", "CUniversities", "CElections",
"CBanks", "CEnvOrg")

focus_country_data = VC[VC$Country == "CZE" & rowSums(VC[,
c(variables, response_vars)] == -1, na.rm = TRUE) == 0, ]
other_countries_data = VC[VC$Country != "CZE" & rowSums(VC[,
c(variables, response_vars)] == -1, na.rm = TRUE) == 0, ]

focus_models = list()
other_models = list()

for (response in response_vars) {
  formula_model <- as.formula(paste(response, "~", paste(variables,
collapse=" + ")))
  focus_models[[response]] <- lm(formula_model,
data=focus_country_data)
  other_models[[response]] <- lm(formula_model,
data=other_countries_data)
}

par(mfrow = c(2,2))
plot(focus_models[[response_vars[1]]])
plot(focus_models[[response_vars[2]]])
plot(focus_models[[response_vars[3]]])
plot(focus_models[[response_vars[4]]])
plot(focus_models[[response_vars[5]]])
plot(focus_models[[response_vars[6]]])
plot(focus_models[[response_vars[7]]])
plot(focus_models[[response_vars[8]]])
```

```
plot(focus_models[[response_vars[9]]])
plot(focus_models[[response_vars[10]]])

#Other country model plots
plot(other_models[[response_vars[1]]])
plot(other_models[[response_vars[2]]])
plot(other_models[[response_vars[3]]])
plot(other_models[[response_vars[4]]])
plot(other_models[[response_vars[5]]])
plot(other_models[[response_vars[6]]])
plot(other_models[[response_vars[7]]])
plot(other_models[[response_vars[8]]])
plot(other_models[[response_vars[9]]])
plot(other_models[[response_vars[10]]])

focus_country_models = lapply(response_vars[1:10], function(r)
focus_models[[r]])
other_country_models = lapply(response_vars[1:10], function(r)
other_models[[r]])

#Print stargazer table
#Czech Republic
sink("focus_models_output.html")
stargazer(focus_country_models,type = "html")
sink()

#Other Country
sink("other_country_output.html")
stargazer(other_country_models,type = "html")
sink()

#Heatmap
extract_p_values_and_plot <- function(models, title) {
  p_matrix = sapply(models, function(model) coef(summary(model))[,
"Pr(>|t|)"])
  p_matrix = as.data.frame(p_matrix)
  p_matrix$Variable = rownames(p_matrix)
  p_long = melt(p_matrix, id.vars = "Variable")

  ggplot(p_long, aes(x = variable, y = Variable, fill = value)) +
    geom_tile() +
    scale_fill_gradient2(low = "red", mid = "red", high = "white",
midpoint = 0) +
    theme_minimal() +
    labs(title = title, x = "Response Variables", y = "Predictor
Variables") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
}

# Plot heatmap for Czech Republic
extract_p_values_and_plot(focus_country_models, "Heatmap of p-values
(Czech Republic)")

# Plot heatmap for other countries
extract_p_values_and_plot(other_country_models, "Heatmap of p-values
(Other Countries)")
```

```
#Q2b and c but with chosen variables

variables_extract = c("PDemImp", "PSocial", "TFamily", "PRadio",
"TMeet", "PEmail", "HChoice", "TKnow",
              "Emp_student", "PSatisfied", "TNeighbourhood",
"MF_male")
variables_other_extract = c("TNeighbourhood", "TFamily",
"PDemCurrent", "PSatisfied", "TMeet", "TPeople",
                            "TKnow", "HOverall", "EHardWork",
"PIA_2", "PIAB_2", "Emp_housewife",
                            "Emp_self", "PIA_3", "PIA_1")

focus_country_data_extract = VC[VC$Country == "CZE" & rowSums(VC[,
c(variables_extract, response_vars)] == -1, na.rm = TRUE) == 0, ]
other_countries_data_extract = VC[VC$Country != "CZE" & rowSums(VC[,
c(variables_other_extract, response_vars)] == -1, na.rm = TRUE) ==
0, ]
focus_models_extract = list()
other_models_extract = list()

for (response in response_vars) {
  formula_model <- as.formula(paste(response, "~",
paste(variables_other_extract, collapse=" + ")))
  focus_models_extract[[response]] <- lm(formula_model,
data=focus_country_data)
  other_models_extract[[response]] <- lm(formula_model,
data=other_countries_data)
}

focus_country_models_extract = lapply(response_vars[1:10],
function(r) focus_models_extract[[r]])
other_country_models_extract = lapply(response_vars[1:10],
function(r) other_models_extract[[r]])

sink("focus_models_extract_output.html")
stargazer(focus_country_models_extract,type = "html")
sink()

sink("other_models_extract_output.html")
stargazer(other_country_models_extract,type = "html")
sink()
```

Question 3

```
#import libraries
library(ggplot2)
library(dplyr)
library(tidyr)
library(stargazer)
library(reshape2)
library(scales)

#Q3
comb_external_data = read.csv("combined_data.csv")
vars = c("Average.Income", "CPI", "Democracy.score",
"Freedom.score", "GDP", "Population", "Tertiary.rate",
```

```
"Unemployment.rate")

comb_external_data = comb_external_data %>%
  select(Country, all_of(vars)) %>%
  drop_na()

#boxplot to identify outliers
data_cols = as_tibble(comb_external_data%>% select(., all_of(vars)))
data_long = data_cols %>% pivot_longer(vars, names_to = "column")
ggplot(data_long, aes(x = value)) + geom_boxplot() + facet_wrap(~
column, scales = "free")

#delete outliers
clean_extdata = comb_external_data %>% select(Country, all_of(vars))

remove_outliers <- function(x) {
  Q1 = quantile(x, 0.25, na.rm = TRUE)
  Q3 = quantile(x, 0.75, na.rm = TRUE)
  IQR_val = Q3 - Q1
  lower_bound = Q1 - 1.5 * IQR_val
  upper_bound = Q3 + 1.5 * IQR_val
  x[x >= lower_bound & x <= upper_bound]
}

for (var in vars) {
  clean_extdata = clean_extdata %>%
    filter(
      .data[[var]] >= quantile(.data[[var]], 0.25, na.rm = TRUE) -
1.5 * IQR(.data[[var]], na.rm = TRUE) &
        .data[[var]] <= quantile(.data[[var]], 0.75, na.rm = TRUE) +
1.5 * IQR(.data[[var]], na.rm = TRUE)
    )
}

data_cols = as_tibble(clean_extdata%>% select(., all_of(vars)))
data_long = data_cols %>% pivot_longer(vars, names_to = "column")
ggplot(data_long, aes(x = value)) + geom_boxplot() + facet_wrap(~
column, scales = "free")

write.csv(clean_extdata, "clean_extdata.csv", row.names = FALSE)
clean_extdata[,vars] = scale( clean_extdata[,vars])

#Elbow Method
elbowdata = data.frame()

for (k in 1:20) {
  kfit = kmeans(clean_extdata[, vars], centers = k, nstart = 6)
  print(kfit$tot.withinss)

  elbowdata = rbind(elbowdata, t(c(k, kfit$tot.withinss)))
}

colnames(elbowdata) = c("k", "tot.within.ss")

plot(elbowdata$k, elbowdata$tot.within.ss, type = "b",
     xlab = "Number of Clusters (k)", ylab = "Total Within-Cluster
Sum of Squares",
```

```
     main = "Elbow Method to Determine Optimal k")

#Clustering
kfit = kmeans(clean_extdata[, vars], centers = 6, nstart = 6)

kcentroids = aggregate(clean_extdata[, vars], by = list(cluster =
kfit$cluster), mean)
print(kcentroids)

clean_extdata$cluster = kfit$cluster

country_cluster = clean_extdata %>%
  filter(Country == "CZE") %>%
  select(Country, cluster)
print(country_cluster)

select_country = clean_extdata%>%
  filter(cluster == country_cluster$cluster[1])

print(select_country)

# Correlation matrix
correlation_matrix = cor(clean_extdata[, vars], use =
"complete.obs")
cor_melted = melt(correlation_matrix)
ggplot(cor_melted, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "lightblue", high = "darkblue", mid =
"lightblue", midpoint = 0) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Correlation Heatmap of Selected Variables", x =
"Variables", y = "Variables")


#Clustering plot using Principal Component Analysis
pca_result = prcomp(clean_extdata[, vars], center = TRUE, scale. =
TRUE)
pca_data = data.frame(pca_result$x[, 1:2])
pca_data$Country = clean_extdata$Country
pca_data$Cluster = as.factor(clean_extdata$cluster)

ggplot(pca_data, aes(x = PC1, y = PC2, color = Cluster, label =
Country)) +
  geom_point(size = 3, alpha = 0.8) +
  geom_text(size = 2.5, vjust = -0.5) +
  labs(title = "Country Clustering (With PCA Projection)",
       x = "PC 1", y = "PC 2") +
  theme_minimal() +
  scale_color_brewer(palette = "Dark2")

#Q3c
selected_countries = select_country$Country
selected_countries = selected_countries[selected_countries != "CZE"]
cluster_country_data = VC %>%
  filter(Country %in% selected_countries) %>%
  filter(rowSums(across(all_of(c(variables, response_vars)), ~ . ==
```

```
-1), na.rm = TRUE) == 0)

variables = c("TPeople", "TFamily", "TNeighbourhood", "TMeet",
"TKnow",
              "VFamily", "VLeisure", "VWork", "HOverall", "HChoice",
"HShelter",
              "EEquality", "EGovernment", "EHardWork", "SEducation",
"STBetter",
              "PRadio", "PEmail", "PInternet", "PSocial", "PDemImp",
              "PDemCurrent", "PSatisfied",
              "MF_male", "MF_female", "MF_others",
              "Age", "Edu",
              "Emp_fulltime", "Emp_parttime", "Emp_self",
"Emp_retired",
              "Emp_housewife", "Emp_student", "Emp_unemp",
              "PIA_1", "PIA_2", "PIA_3",
              "PIAB_1", "PIAB_2", "PIAB_3")

response_vars = c("CArmedForces", "CUnions", "CPolice", "CPParties",
"CParliament",
                  "CCivilService", "CUniversities", "CElections",
"CBanks", "CEnvOrg")

select_country_models = list()

for (response in response_vars) {
  formula_model <- as.formula(paste(response, "~", paste(variables,
collapse=" + ")))
  select_country_models[[response]] <- lm(formula_model,
data=cluster_country_data)
}

par(mfrow = c(2,2))
plot(select_country_models[[response_vars[1]]])
plot(select_country_models[[response_vars[2]]])
plot(select_country_models[[response_vars[3]]])
plot(select_country_models[[response_vars[4]]])
plot(select_country_models[[response_vars[5]]])
plot(select_country_models[[response_vars[6]]])
plot(select_country_models[[response_vars[7]]])
plot(select_country_models[[response_vars[8]]])
plot(select_country_models[[response_vars[9]]])
plot(select_country_models[[response_vars[10]]])

select_cluster_country_models = lapply(response_vars[1:10],
function(r) select_country_models[[r]])

#Print stargazer table
sink("select_models_output.html")
stargazer(select_cluster_country_models,type = "html")
sink()

#Heatmap
extract_p_values_and_plot(select_country_models, "Heatmap of
p-values (Cluster Countries)")
```