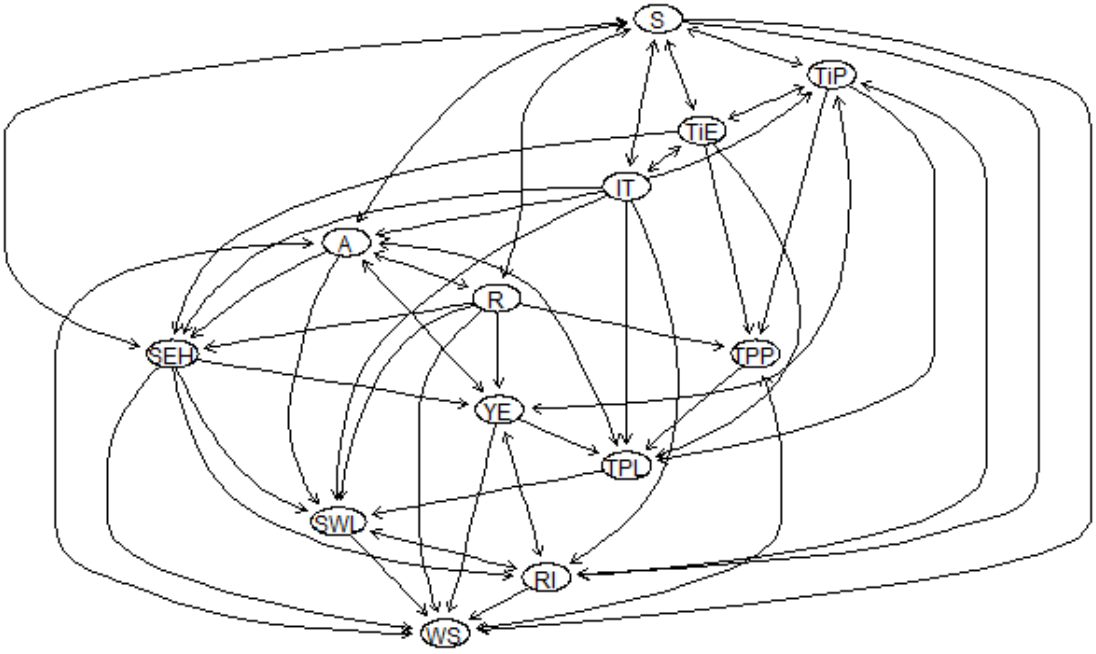
Causal Effects on Subjective Well-Being from Observational Data in Latin America

Michiel Bosma

June 2022

Supervisors: Yolanda Grift and Tina Dulam
Student Number: 5663067
Name of Programme: Applied Data Science, Utrecht University

Figure 1: CPDAG estimated for Sample 1



Contents

1 Abstract

This thesis aims to provide causal information for the study of happiness in Latin America. Using survey data from LAPOP from different countries in Latin America the causal structure surrounding Satisfaction with Life is investigated. First, the missing data is multiple imputed to solve the bias in the missingness. Then, building on the work of Pearl (2000) and others, causal discovery algorithms are performed to aim to learn a CPDAG from the observational data. The causal discovery algorithms work by searching through the conditional dependencies and creating a causal structure consistent with them. In the end, across samples, CPDAG's are obtained using the PC-algorithm. Then, using the IDA algorithm, the strength of the causal relationships is estimated. The most important causal effects that were estimated were that of Subjective Economic Hardship on Satisfaction with Life (negative 15 - 17 percent), that of Interpersonal Trust on Satisfaction with Life (positive 8 - 10 percent) and that of Interpersonal Trust on Subjective Economic Hardship (negative 9 - 11 percent). The results are important for happiness research and have implications for public policy.

2 Introduction

2.1 Motivation and context

This Thesis is based upon the work already done by Tina W. Dulam, Yolanda Grift and Annette van den Berg. Their provisional paper: 'Economic Hardship, institutions and subjective well-being in Latin America' forms the inspiration for this work. In this paper, the relationship between subjective well-being and economic hardship in Latin America and the mitigating effect of institutions is investigated. However, in this paper, many of the relationships discussed are from correlations and a causal relationship is not discerned. This is seen as a problem to make this paper published. Consequently, the first, direct motivation for this thesis is to provide extra information for this provisional paper by learning the causal structure from their observational cohort data.

The provisional paper attempts to fill a gap in the literature by studying the exact variables relating happiness, economic hardship and institutions in Latin America. Much study is already been done on the relationship between happiness or subjective-well being and economic factors, however, the role of institutions is hereby been neglected. The provisional paper is based upon the paper by Reeskens and Vandecasteele (2017) called: Economic Hardship and Well-Being: Examining the Relative Role of Individual Resources and Welfare State Effort in Resilience Against Economic Hardship' which did attempt to fill this gap by including the role of institutions. However, this paper is focusing on Europe and the provisional paper by Dulam, Grift and van den Berg attempts to extend their approach to Latin America to see whether the same results hold up or if different relations are present.

Further motivation of this research is naturally to contribute to the research of happiness in especially Latin America. Broadening our field of study in this less-studied area of the world is paramount in my opinion. When we discuss how to develop this part of the world or how to move forward, it is necessary to understand what causes individuals to be happy in Latin America. We should not extrapolate the results from Europe towards this continent, but attempt to study this continent in its own right. Increased understanding of this feature can be of crucial importance for policymaking in this area, therefore. The current study consequently attempts to play a minor role in moving this understanding forward, with as the ultimate goal naturally to make individuals in Latin America a little happier than they were before.

2.2 Contents

The remainder of this research follows the following structure. In section 2.3 a Literature overview will be provided, contextualizing this research in the field of the study of happiness. Then, in 2.4 the research question and the accompanying hypotheses are introduced. In Chapter 3, the preparation and exploration of the data are discussed. 3.1 discusses the preliminary data preparation including the selection of variables and countries in the data. Furthermore, 3.2 presents the initial results of the data exploration process, where initial distributions of variables and correlations between variables are shown. Then, in 3.3, the

advanced data preparation is analyzed. This includes the creation of samples and the handling of missing data through multiple imputation. Lastly, in 3.4, a short discussion of the ethical component of the data is provided.

In chapter 4, we turn to the subject of causality. The research question is turned into a data science question in section 4.1, where also the work on causality on which this research builds is introduced. 4.2 explains the specific methods used to perform causal discovery, specifically the PC-algorithm and the IDA-algorithm. Then, in 4.3, the settings for the algorithms are discussed to specify how the algorithms were used. Especially the context of causal discovery with multiple imputation is important here. Chapter 5 furthermore demonstrates the results following from the algorithms and describes which causal relationships are found. Furthermore, it provides estimations of the causal strengths between variables.

Chapter 6 reflects on the results in chapter 5 and answers the data and research question. The data question is answered in 6.1 and the research question in 6.2. Finally, in 6.3, implications for the domain, which is the research on happiness, are drawn from the results. Chapter 7, then, is the Appendix, where the code scripts can be found in 7.1, the full data exploration results in 7.2, the full analysis results in 7.3 and the list of tables and figures in 7.4. Lastly, Chapter 8, provides a List of References accompanying this research.

2.3 Literature overview

This project is concerned with the measurement of happiness [1]. Before continuing, therefore, it is helpful to look at the current literature to understand how in current research these concepts are understood and how they are measured. Rojas (2019a) locates the start of this research tradition in the seminal paper by Richard Easterlin (1974): *Does Economic Growth Improve the Human Lot? Some Empirical Evidence.*

Easterlin was not the first scholar or economist to focus on happiness, however, as Rojas (2019a) explains through the first decades of the twentieth century the study of happiness was gradually abandoned to focus on the study of choice under constraints. This abandonment of the study of happiness relied on a fundamental assumption: That the happiness of people increased whenever their choices reflected in their consumption possibilities increased. However, we do not see much interest in validating this fundamental assumption (Rojas, 2019a). In the paper by Easterlin, this dictum that there is a clear positive relationship between the concept of economic welfare and the concept of social welfare is investigated and challenged. What is exactly the relationship between economic growth and happiness (Easterlin, 1974)? Attempting to empirically answer this question, Easterlin started a research tradition in which this work can also be located. However, now not only the relationship between economic growth and happiness but the relationship between happiness and a wide variety of variables is studied broadening the economic research tradition.

The measurement of happiness can be traced back to the ideas of Jeremy Bentham (1780). He firstly conceptualized the idea of happiness as an experience of people. He narrowly

[1] Throughout this paper the concepts of happiness, satisfaction with life and subjective well-being will be used interchangeably.

thought of experience as consisting of pleasures and pains, however, this move is crucial in understanding how we currently measure happiness (Collard, 2006). Following Bentham, we do not attempt to objectively, from the outside, determine what should cause happiness in an individual, but are content with the subjective evaluations of those individuals that experience that happiness (Collard, 2006). Therefore, we can also describe this idea as the measurement of subjective well-being contrasted with the more objectively approached choice theory. The concept of happiness is therefore not objectively described but follows out of the subjective evaluation of an individual who declares that he is or is not experiencing happiness.

As Frey et al. (2008) state, this research on happiness or subjective well-being is important because, in the end, happiness is thought to be the ultimate goal of life. Most individuals strive for happiness and consequently economics, it is claimed, should not be about choice, but about happiness. The profound question is and should be: How do things like economic growth, inflation, inequality, environmental factors and institutional factors affect individual subjective well-being (Frey et al., 2008)? To promote happiness for the people on this planet, we should understand what makes people happy. Studying this question is called the economics of happiness, but it can include variables or factors which we would not in the normal or current sense of the world call economical.

For current economics, the neglect of the happiness question has caused some detrimental results. The difference in perspective can cause different economic policies to be proposed or let economics focus on factors such as the working of institutions which are neglected in the mainstream economic agenda (Rose, 2017). The research of happiness or subjective well-being can consequently inform our policymaking (Frey Stutzer, 2002). Moreover, happiness research can fill the gap within the economic tradition by connecting the research traditions of objective utility with that of subjective utilities.

Frey and Stutzer (2005) discussed the state of happiness research at the time of writing (2005). As they state, recent advances in the measurement of subjective well-being now make it quite convincing that we can inform ourselves of the subjective-well being of individuals through their self-reporting, where true well-being serves as the latent variable in which we are interested. Moreover, we can in this way possibly gain information about the determinants of happiness. The issue of causality is a great one, however (Powdthavee, 2007, 2010). Using a happiness function where happiness is the latent variable assumes that happiness is the dependent variable, however, reverse causation may also be possible. How do we know that the correlations that we find are not caused by a causal relationship going the other way around? Studying this question is what Frey and Stutzer (2005) see as the future of happiness research, and is also the question that this present project is concerned with.

As it is concluded in the provisional paper, economic hardship seems to be negatively related to happiness and institutional factors seem to have a positive correlation. However, what is unclear is whether economic hardship causes diminishing subjective well-being or whether decreasing subjective well-being can be thought of as also causing in a sense economic hardship? It seems logical that the first is true, and the latter is not, but clear results from the data are missing. The same is the case for the relationship between institutions and subjective well-being. It seems logical that successful institutions cause an increase in happiness instead

of the other way around, nevertheless, clear results from the data are missing.

When looking at research done on happiness in Latin America, it is often reported that the causal direction is unclear. In the seminal volume with work from different authors on happiness research in Latin America collected by Rojas, we can see how often this is reported. We can discern this from the following quotes in Rojas (2019b, p. 593): 'While the direction of causality is unclear, it is desirable to incorporate this variable', in Rojas (2019b, p. 515): 'This indicates that even though the wealth-health relationship is one of the most published topics in health economics and disciplines such as demographics, the direction of causality between the two variables remains an open debate' and in Rojas (2019b, p. 487): 'we cannot predicate causality, but we can report that there is a strong statistical correlation that appears in different studies'.

We conclude therefore that the study of happiness in Latin America is well underway, however, the study of the causal relationships in this field is lacking. This project will consequently aim to fill this gap in the literature.

2.4 Research Question

The research question guiding this project is, therefore: *Is it possible to gather information about the causal structure surrounding subjective well-being or happiness in Latin America?* This research question will in section 4.1 be transformed into two data science sub-questions to specify the research. Then, in 6.2. the answers to the data science question will be used to answer the research question.

Furthermore, this research will be guided by transforming two hypotheses in the provisional paper into hypotheses about the causal structure surrounding subjective well-being in Latin America. The hypothesis in the provisional paper H1: *Economic Hardship is negatively related to subjective well-being* can consequently be reformulated in the causal hypothesis H1B: *Economic hardship causes a decrease in Subjective Well-being.* And the hypothesis in the provisional paper *Institutional Quality mitigates Economic Hardship and has a net positive effect on Subjective Well-being* (H2) can be reformulated in H2B: *Institutional Quality has a different causal direction than Economic Hardship in their relationship to Subjective Well-Being, causes an increase in Subjective Well-Being and also has a negative causal influence on Economic Hardship.*

3 Data

The Data is collected from the 2016/2017 LAPOP barometer. The LAPOP barometer is the foremost survey institution in Latin America, collecting professional data from all Latin-American countries. The Data is collected from countries in North and South America. Not all countries have the same participants included, therefore a weight factor is present in the Dataset to be able to make population-wide analysis possible and create overall results.

3.1 Preliminary data preparation

From the complete Dataset that consisted of a data frame with 42451 rows (number of observations) and 535 variables, a selection for our analysis was made. First, participants from countries that were not analysed in the provisional paper were excluded. The included countries are Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Uruguay and Venezuela. These countries, therefore, include all countries in the Central and South American areas, which make up Latin America.

Then, a selection of variables was made, based on the variables analysed in the provisional paper, and only these variables were included for further analysis. The included variables in the questionnaire were weight1500, q1, q2, b21, b21a, b13, b47a, cp6, it1, q10d,ls3, ed, q10new and ocup4a. These variables translate to Country Weight, Gender, Age, Trust in Political Parties, Trust in Political Leader, Trust in Parliament, Trust in Elections, Religiosity, Interpersonal Trust, Subjective Economic Hardship, Satisfaction With Life, Years of Education, Relative Income and Work Status.

The variable Work Status was mutated to a binary variable, where all the non-working categories are grouped in one category. This is done to increase understanding since for our purposes it is I believe only necessary to know whether a participant is working or not. Moreover, the variables of Satisfaction with Life, Interpersonal Trust, Religiosity and Work Status were mutated to create the fact that a higher number in the data analysis would always signify a higher quantity of that variable instead of the other way around. The nature of the survey meant that this was not always the case, hence the necessary mutation.

3.2 Selected data exploration results

This selected data was explored to see if the distribution of the variables was logical or could present any problems. Furthermore, relationships between crucial variables were preliminarily explored. Finally, missing data were explored to achieve a full understanding of the scope of this problem.

For all the plots indicating the distribution of the data, see Appendix 7.3. Here I would like to discuss the distribution of the crucial variables for our understanding.

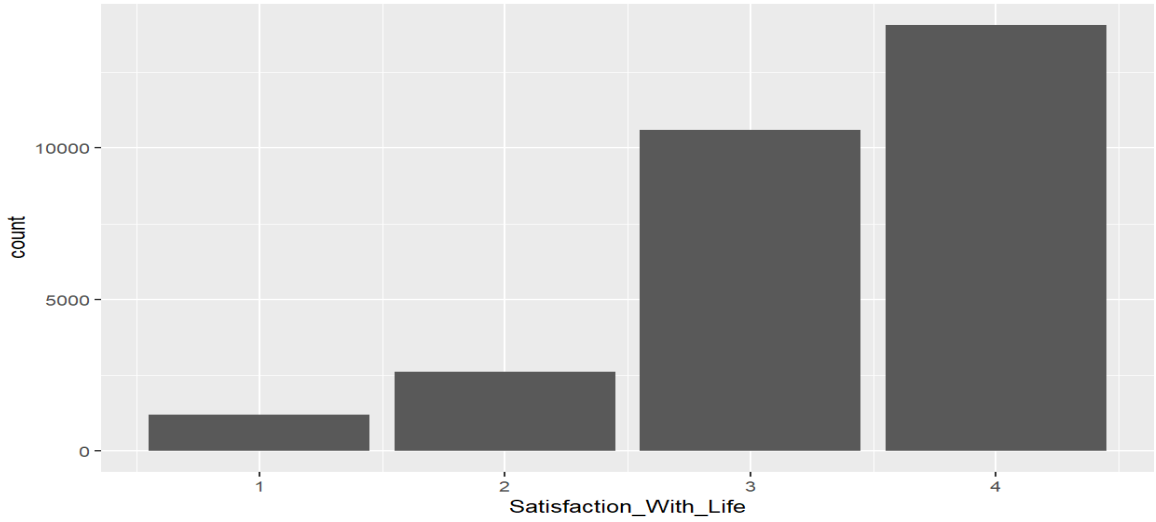
In Figure 2 [2] [3] the distribution of the variable Satisfaction with Life is plotted. This

[2]Source: LAPOP 2016-2017 Cohort

[3]footnote 2 is valid for all Figures, Tables and Images

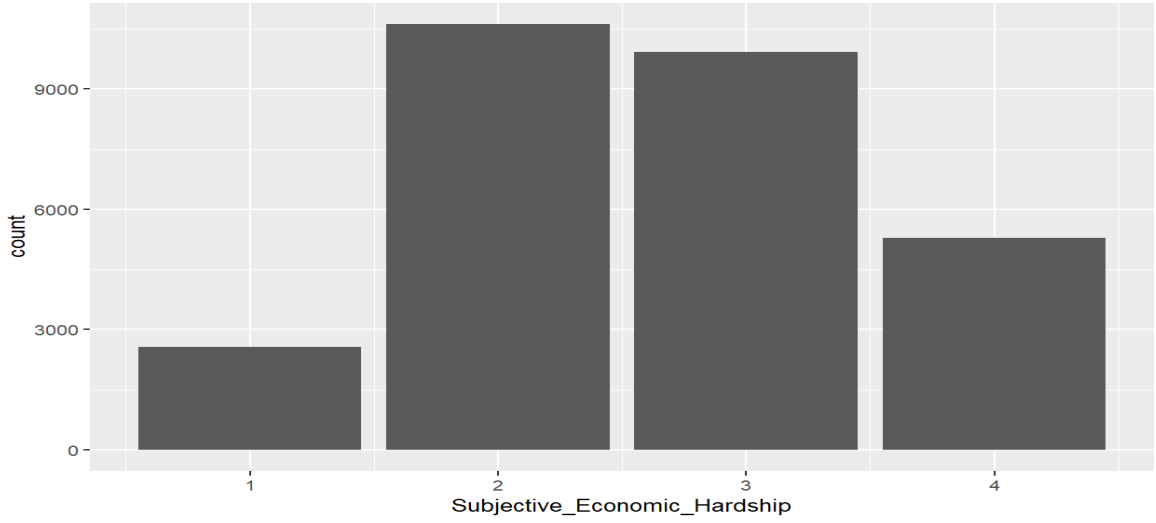
variable is our variable of interest since through this variable we measure the subjective well-being of our participants. Interestingly, we see that most participants self-report as being very satisfied and least participants self-report that they are very dissatisfied with their life. In countries where economic development sometimes trails that of the Western World, this is remarkable (Bértola and Ocampo, 2012).

Figure 2: Distribution of Satisfaction With Life



In Figure 3 the distribution of the Subjective Economic Hardship variable is shown. We see that most participants identify themselves around a 2 or a 3, meaning that they feel that their salary is just enough or just not enough, and fewer people locate themselves around the extremes.

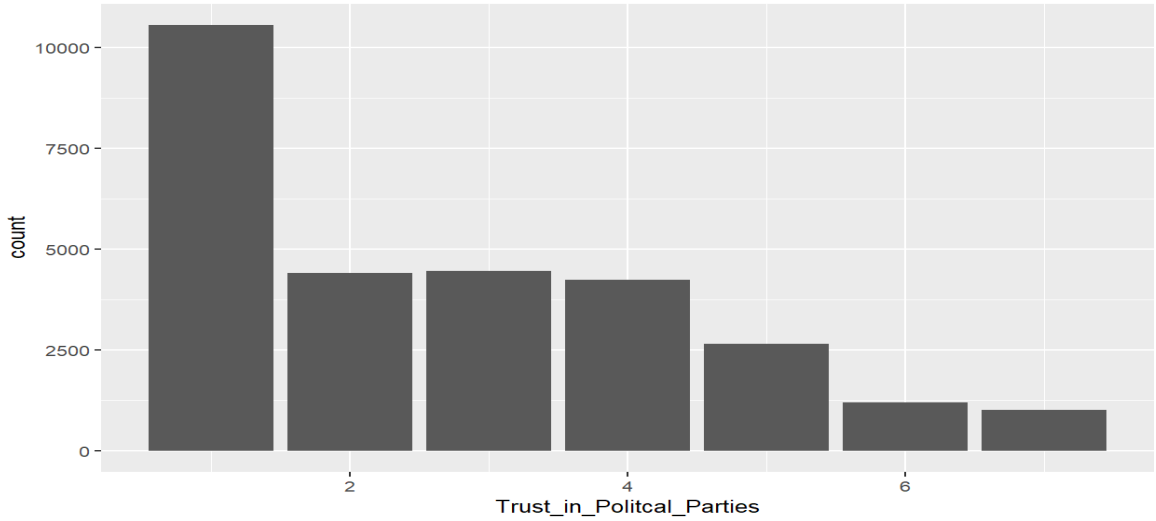
Figure 3: Distribution of Subjective Economic Hardship



Finally, in Figure 4, we see the distribution of the variable Trust in Political Parties. Importantly, most people report an astounding lack of trust in their political parties, whether

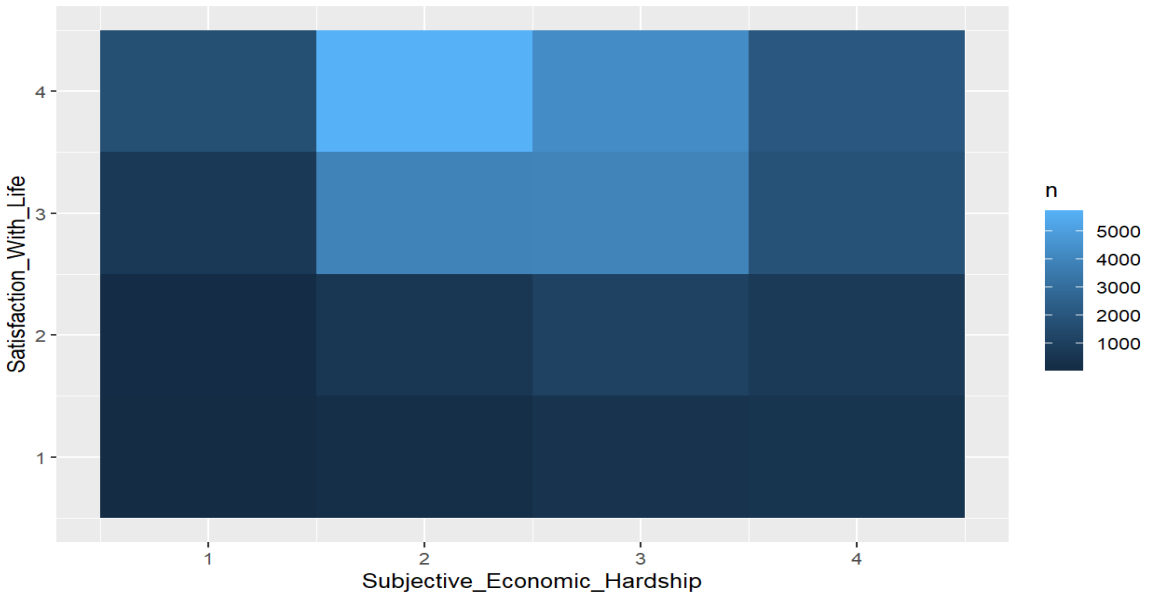
the amount of people choosing the lowest category trust is twice as high as the support for any other category. The high subjective well-being of the people of Latin America does not seem to come from this trust in the political parties that are supposed to represent them.

Figure 4: Distribution of Trust in Political Parties



To explore the relationship between Satisfaction with Life and Subjective Economic Hardship, the paired results are visualized. In Figure 5 a heatmap shows where most participants locate themselves in the paired matrix of these two variables. The combination of reasonably low Subjective Economic Hardship (2) and high Satisfaction with Life (4) is most prevalent.

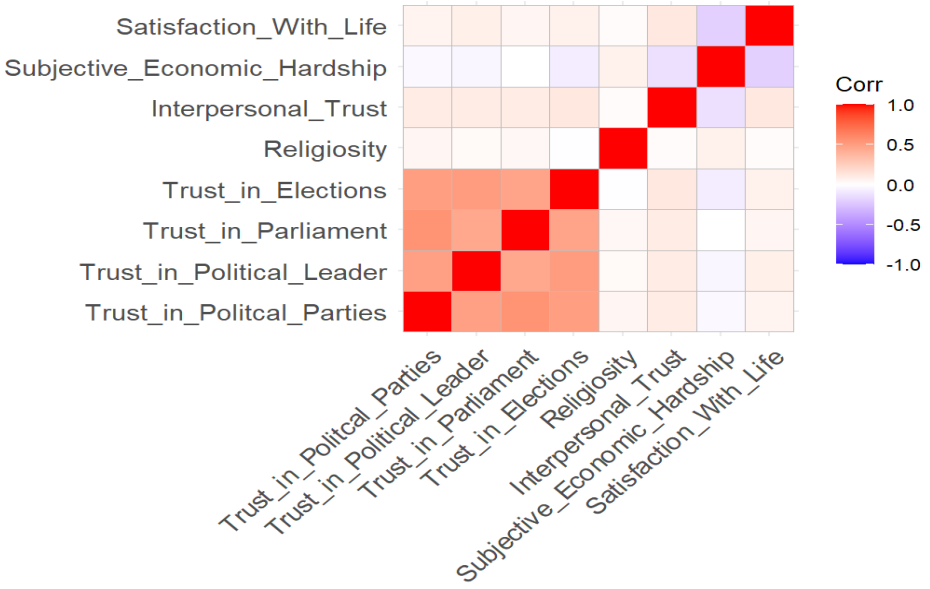
Figure 5: Heat Map of SWL and SEH



Furthermore, correlations between the variables are visualized in a correlation matrix. In Figure 6 we firstly see that the trust variables all exhibit large correlations across each other,

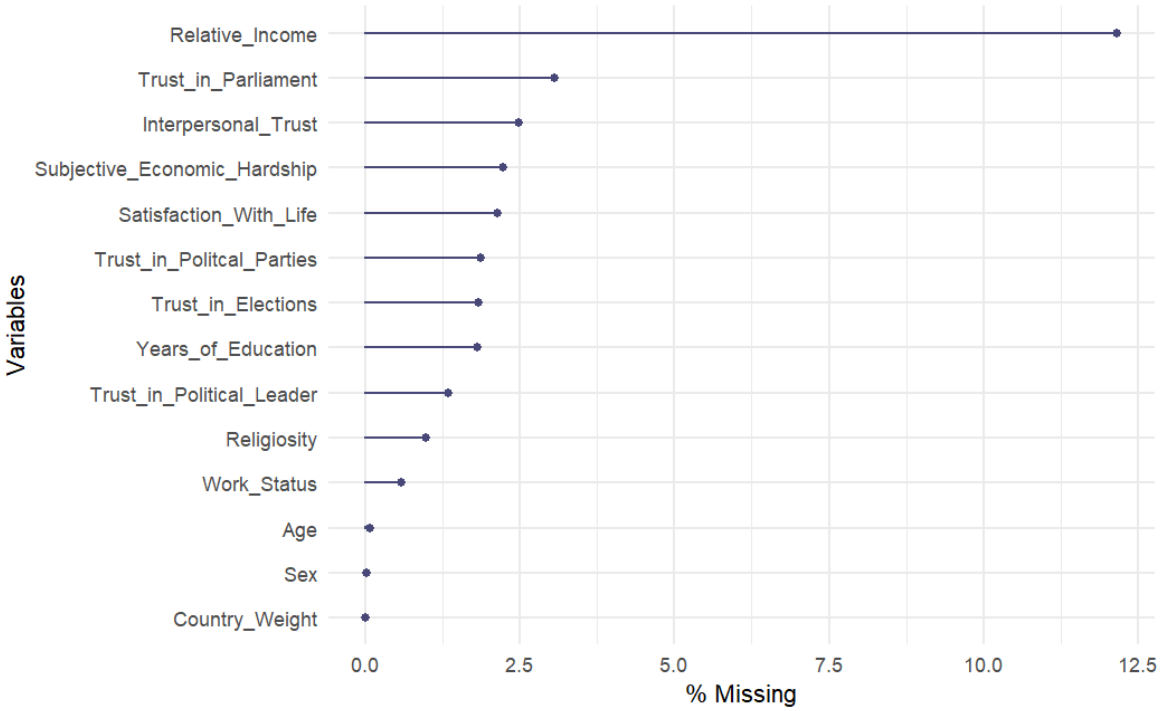
which is reasonable. Moreover, we see that most correlations are slightly positive, however, the Subjective Economic Hardship variable possesses slightly negative correlations with the other variables. So, some more economic hardship is correlated with for example a lower satisfaction with life, lower religiosity and lower trust in political leaders and parties.

Figure 6: Correlation Matrix of Important Variables



Finally, the amount of missingness in the Dataset is investigated. First analysis showed that out of 377.832 possible observations in the data frame there were 8843 missings. This results in a missingness percentage of 2.3 percent. In figure 7 the missingness percentage per variable is shown. We see that almost all the variables have a reasonably low amount of missingness around 2.5 percent. However, Relative Income has a Missingness Percentage of 12.5 percent. Any method dealing with this missingness should incorporate this 12.5 percent to be certain that precision is reached. In this project, this is done by using the 12.5 percent in the calculation of the number of imputations that are necessary to ensure reproducible standard deviations in the imputed data.

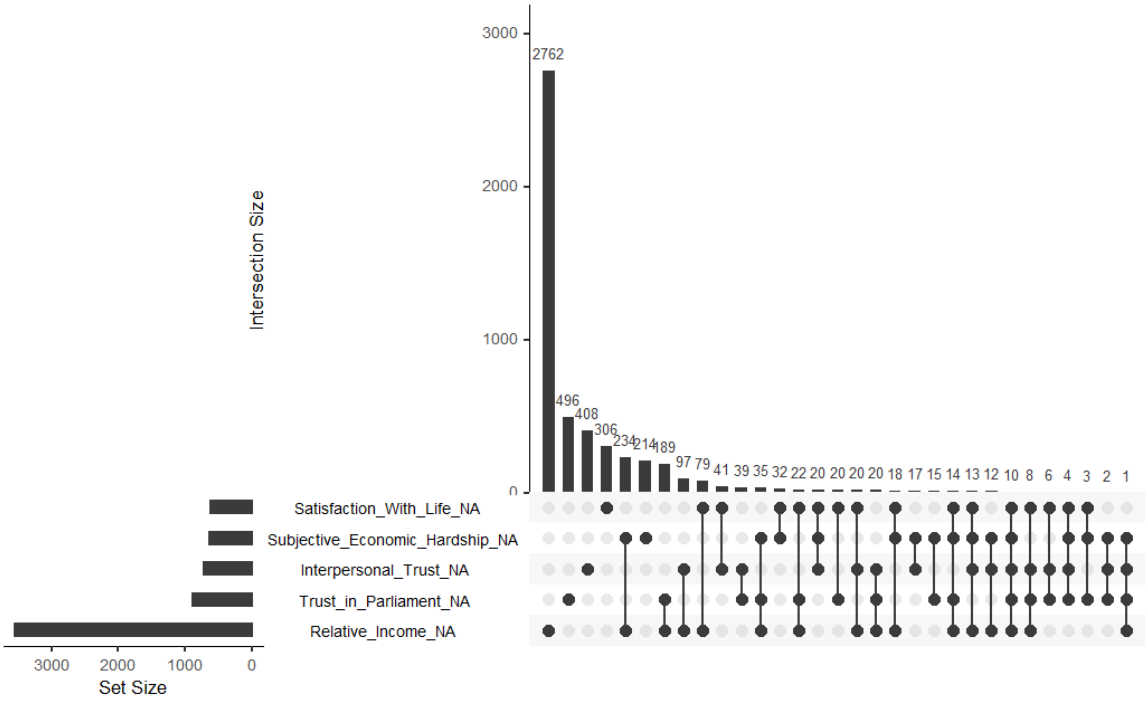
Figure 7: Missingness Percentage per Variable



Lastly, in Figure 8 the relationship in missingness between the variables is visualized. Most participants only miss one variable, however, there are also some more missingness relationships present in the data. The fact that most participants only have one value missing makes the hereafter proposed method of multiple imputation more credible since it is not the case that for a participant often almost all variables are imputed.

Moreover, this figure provides evidence for the idea that the missingness in the data is not MCAR, but MAR. That is, the missingness is probably not Missing Completely at Random, but only Missing at Random. The difference is that under MCAR the missing is caused by a full random process, while under MAR the missingness is related to variables in the dataset (Van Buuren, 2018). This is finally also been tested with the MCAR test (Li, 2013; Little, 1988). The result of this test is that the data is not MCAR with a certainty of almost 100 percent. This means that we are almost certain that there are missingness patterns in the data and that the missingness in the data is not random. Moreover, 297 missingness patterns are detected.

Figure 8: Relationships in Missingness



3.3 Advanced preparation for analysis

Hereafter, the fact that the data consisted of samples of different sizes from different countries had to be solved. The weight factor supplied by LAPOP was used in this preparation. A country out of which fewer observations were collected had a higher weight factor to remedy this problem. In traditional statistical methods, the weight-factor is used to achieve representative population statistics. However, the data methods used hereafter were too complicated to be able to incorporate immediately this weight factor. Therefore, samples of the original data were created. These samples consisted of 80 percent of the observations of the original dataset, where the probability for each observation to be included in the dataset was equal to the weight factor. In this way, representative samples were created and the use of samples can make a comparison in later data analysis possible. Moreover, any results that are consistent across samples can be believed to be quite robust.

The missingness in the data was dealt with per sample using multiple imputation methods. As Van Buuren (2018) indicates, other methods, such as pairwise deletion (available case analysis) or single imputation methods suffer from problems if the data is not MCAR. Wulff and Jeppesen (2017) moreover state that multiple imputation also outperforms pairwise deletion and similar techniques under different missing mechanisms and sample sizes. Pairwise deletion introduces bias in the final result, since the probability of missing is related to the data, and single imputation methods are unable to incorporate the uncertainty that is present in missing data. Multiple imputation consists basically of the idea that multiple complete data sets are created based upon the original data, where missing data is imputed

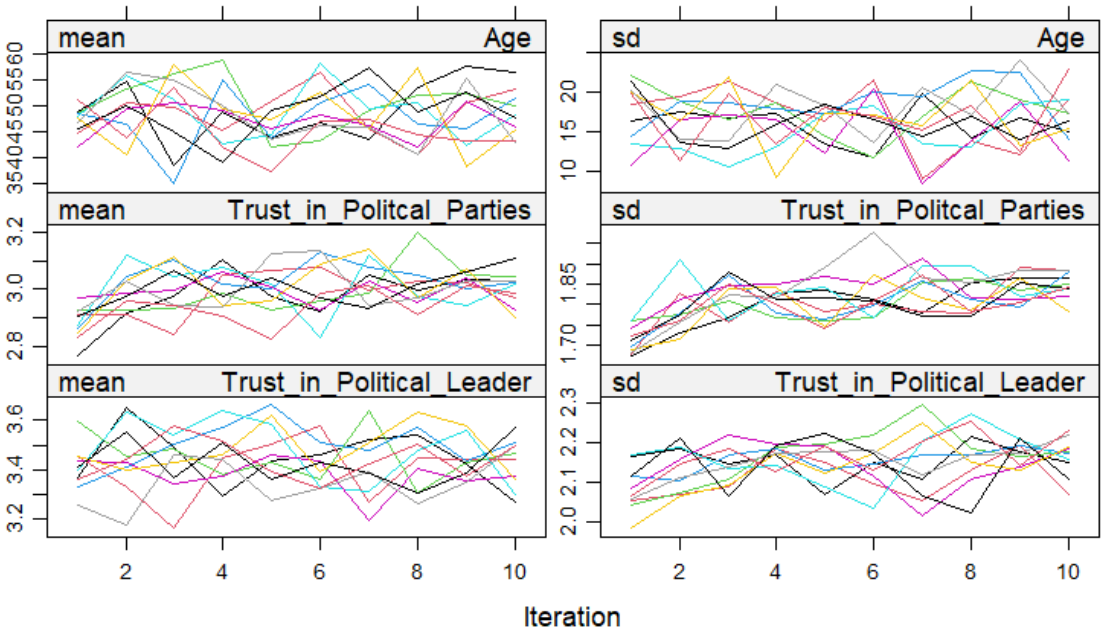
stochastically, and plausible values are imputed into a data set. Consequently, a final result is reached where multiple full datasets are created, which allows for uncertainty in the missing values, since those values are different across data sets. This uncertainty can be incorporated into further analysis.

The mice package by Van Buuren and Groothuis-Oudshoorn (2011) was used to perform the multiple imputation, where the appropriate method is immediately chosen based on the type of data. Since most of our data can be seen as numeric, the predictive mean matching method was used to create the multiple imputed data. The data unfortunately had to be numeric and could not be ordinary factors, since that would be the most appropriate interpretation, based upon the fact that the data is ranked survey data. However, as will be explained in the next section, further data methods were unable to handle this data type. Consequently, the data has to be interpreted as numeric (continuous and ranked) data.

Finally, the work of Von Hippel (2020) was used to calculate how many imputations were needed to be able to achieve replicable standard errors and uncertainty intervals in our variables. For only the point estimates few imputations are needed, however, to also compute reliable uncertainty intervals the amount of imputations is estimated using a 2-step quadratic method. This resulted in the answer that for the Relative Income variable, which was the variable with the highest missingness percentage, 10 imputations were needed. For the other variables, consequently, fewer imputations are needed, but to achieve replicability of the uncertainty intervals across all variables 10 imputations were taken as the standard. The replicability of uncertainty intervals is crucial in the next phase since the conditional independence tests used in the causal discovery methods are dependent on p-values and consequently on those uncertainty intervals.

The results of the Multiple Imputation process are checked in two ways. First, it is checked whether the imputation has converged towards a value. If this is not the case, then possibly the multiple imputation length was not enough or something went wrong in the process. The result of this process can be seen for Sample 1 in Figure 9. The X-axis in Figure 9 displays the number of iterations and every line signifies a different imputated data set. The Multiple Imputation process is an iterative process and we hope that the iterations become stable after a number of iterations. This means that they do not deviate much across iterations anymore. Furthermore, we hope that across imputed data sets there is no considerable difference across imputations, which can be checked by looking at the difference in the coloured lines.

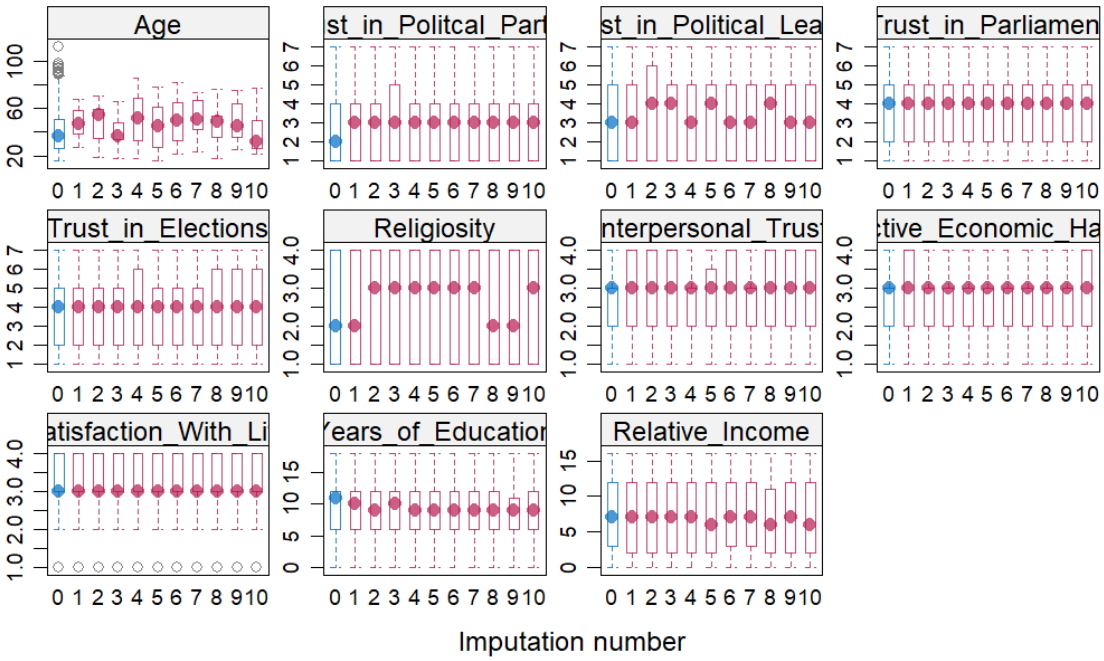
Figure 9: Convergence Plots of Multiple Imputation Process: Sample 1



Hereafter, the results of the multiple imputation are compared with the distribution of the observable data. When a considerable difference is detected, this can be further investigated. Nevertheless, it needs to be highlighted that the success of the multiple imputation method can not be finally demonstrated, but has to be made plausible. Differences between observable and imputated values can come because of a wrong imputation process, but can also be warranted if the missing data significantly differs from the observed data. This consequently has to be judged by the domain and data science experts after careful consideration.

The results of the imputation process are displayed in Figure 10. In blue are the observed data and displayed at numbers 1 through 10 on the X-axis are the mean results of the imputed data sets. It is to be believed that the process has not resulted in significant deviations from the observable results, since the differences are minor. Only for the religiosity variable, some significant difference exists, however, this is I believe due to the low missingness of this variable which makes the missing variables more prone to deviations. For full results, check Appendix 7.4.

Figure 10: Observed and Imputed Values of Numerical variables: Sample 1



3.4 Ethical and legal considerations of the data

The Data that are used are public and published on the LAPOP website. LAPOP themselves has an extensive ethical guideline, which explicates how the data is collected and how to ensure that this is done ethically and in accordance with legal guidelines. Important issues that are discussed are the conduct of the interviewer, the informed consent of the participant, the importance of confidentiality and the duty to not alter the data collection process in any way. Any participant is asked to fill in a letter of informed consent to establish the voluntary process and state that they have answered the survey truthfully. In this way, the published data already reflects an ethical process in which utmost care has been used to ensure that the rights of the participants have not been violated. Consequently, using this data does not seem problematic. If LAPOP, as an institution with a high standard of ethical survey analysis, has published this open data, it can be believed that any ethical guidelines have been followed. Any more ethical considerations are therefore deemed to be unnecessary by the author.

4 Methods

4.1 Translation of the research question to a data science question

The methods used in this project are inspired by the work of Judea Pearl, which culminated in his book about Causality (Pearl, 2009). The task of causal discovery in this framework is seen as an inductive process, which scientists are playing to understand Nature. Nature is assumed to consist of stable causal mechanisms which can be described through functional causal relationships between variables. The task of scientists consequently is to learn about these causal mechanisms and the culminating causal structure through observational data in an inductive process (Pearl, 2009).

Causality can be understood through the idea of counterfactuals. If a variable would take on a different state, leaving all else equivalent, which variables would correspondingly inhibit which changes? (Pearl, 2009) That is the question that causal discovery algorithms attempt to answer. The assumptions behind these approaches are that first, we are attempting to search for a minimal description of our structure. Moreover, we assume that the conditional independencies that we find are due to stable relationships and not due to accidental cancelling out dynamics. Finally, we assume that no hidden variables are present in the data.

The causal structure is displayed in a Directed Acyclical Graph (DAG) and the set of DAGs which is estimated to be consistent with the observational data is the target of our causal inference (Glymour et al., 2019). DAGs are directed since the existence of arrows between the variables signifies the direction of the causal relationship. Moreover, the DAG is acyclic, because the structure does not allow for a cyclical arrow structure where variables become the cause of themselves (Rohrer, 2018). This set of DAGs can be displayed as a Complete Partial Directed Acyclical Graph (CPDAG), where the equivalence sets of conditional dependencies are captured (Pearl, 2009).

The basic idea in causal discovery learning is that different causal relationships inhibit different conditional dependencies in our observational data. Between three variables, three different structures are possible. We can imagine these variables as a chain structure, a fork structure or a collider (invented fork) structure. A chain structure is a structure A $\rightarrow$ B $\rightarrow$ C, through which genuine causal effects are transmitted. Furthermore, a fork structure is a structure A $\leftarrow$ B $\rightarrow$ C. In a fork, associations are possible through common outcomes, but these associations inhibit no causal information. In colliders (invented forks) a structure A $\rightarrow$ B $\leftarrow$ C is present. This structure formalizes the idea of common causes, and it possesses no statistical associations (Rohrer, 2018). The difference in the associations between these structures is paramount in our causal discovery process (Spirtes et al., 2000).

Through algorithms based upon these basic ideas and the estimation of conditional dependencies and independencies, a DAG or CPDAG is estimated. When we have achieved the task of creating a reliable DAG or CPDAG, we can estimate the strength of the causal relationships between variables by conditionalizing on backdoor paths (Spirtes et al., 2000). Using the backdoor criterion (Pearl, 2009) and the DAG, we can discern on which variables we should condition to estimate the direct or total causal effect. Generally speaking, the

backdoor criterion describes the idea that when we attempt to estimate a direct effect of X on Y we condition on all variables in a fork structure, but not on collider structures since this opens up a spurious association. Paths in the DAG which satisfy this criterion are described as backdoor paths. (Rohrer, 2018, Pearl, 2009).

The identification of the causal structure through the formalization of a CPDAG in our variables is crucial consequently to estimating any causal effect. Otherwise, we may condition on colliders, which opens up a spurious association, or fail to condition on a fork structure and again include non-causal associations in our causal analysis. The original research question can consequently be reformulated into a data science question, which consists of two parts: *What is the CPDAG that is consistent with the observational data of the LAPOP 2016/2017 survey?* And: *Based upon this CPDAG what can we estimate the strength and direction of the causal relationship between our variables in our hypotheses to be?*

4.2 Motivated selection of methods for analysis

Different groups of methods perform causal discovery, such as constraint-based methods, score-based methods and non-linear methods based upon functional causal models (Glymour et al., 2019). The approach that is taken in this paper is the constraint-based method. The choice for this is practical, these methods are researched the most, and it is the method for which most functions have been developed through the R package pcalg. Constraint-based methods exploit different relationships in the conditional independencies in the data to discover and learn the causal structure that supposedly generated the data. Constraint-based methods search through the possible space of possible causal structures to find structures that are compatible with the observed statistical dependencies (Spirtes and Zhang, 2016). The output is not always a single causal structure, but a set of possible structures (Glymour et al., 2019).

The PC algorithm is the specific algorithm that is used to discover the causal structure. Under the assumptions that there are no hidden variables and that there are stable causal relations, this algorithm provides a search which outputs a CPDAG. The PC algorithm works through the following steps. First, a complete graph is formed, where every variable is assumed to possess a causal relationship with any other variable. This is indicated by a CPDAG with edges, or undirected lines between the variables, between all the variables. Then it is checked whether variables can be made conditionally independent using any subset of other variables. When this is possible, it is concluded, that therefore these variables do not directly cause each other and the edge is eliminated (Kalisch et al., 2012). This is because if two variables share a direct causal path, then their causal influence can never be blocked by conditioning on another path. Consequently, whenever this is possible, we must conclude that these variables do not share a direct causal relationship and the edge must be eliminated (Glymour et al., 2019). The result of this first step is called the skeleton of the graph, since it consists of the structure of variables, but does not have any orienting edges.

Then, we attempt to learn the orientation of the remaining edges. The basis of this idea is that chains and forks have a different pattern of dependencies than colliders (invented

forks). If there is a collider, we can learn the direction of the causal path. A triple A, B, C is oriented A $\rightarrow$ B $\leftarrow$ C if B is not in any separating set, since we can then conclude that this structure possesses a collider structure. Separating sets consist of those variables which made A and C conditionally independent. Then, after this step, it is investigated whenever we can orient more edges if another direction will introduce another collider, which is not allowed, or make the graph cyclical, which is also forbidden (Kalisch et al., 2012). Important is that the PC algorithm will, if the assumptions hold, converge to the true equivalence class of possible causal structures under large samples (Spirtes et al., 2000). The PC algorithm is implemented in the package pcalg created by Kalisch et al. (2012) and Kalisch et al. (2020). The PC algorithm is performed with help from the practical guide to causal discovery by Andrews et al. (2021).

After the CPDAG is estimated, we can use this output to estimate the causal strength between variables. The crucial idea behind estimating causal effects is that adjusting for more variables is not always better. As explained above, we need to adjust for different variables based on our estimated CPDAG. By using valid adjustment sets that open up causal paths, but close non-causal associations, we can estimate a causal effect. However, a valid CPDAG is necessary, therefore. Using valid adjustment sets, we can use linear regression to estimate causal effects (Kalisch et al., 2020). The IDA algorithm developed by Maathuis et al. (2009) provides estimates of the causal effects based upon the idea of an intervention of raising a variable by one and calculating the response in the target variable. The IDA algorithm works by firstly extracting a collection of adjustment sets of the intervention variables from the CPDAG. Then, these sets are used in linear regression using the adjustment sets as covariates. Important to understand here is that in an incomplete or complex estimated causal structure more adjustment sets may be valid, and consequently, the algorithm may output more than one estimation of the causal effect. Therefore, a range of the estimation or a summary of the estimations can be provided to allow insight into the distribution of the estimations.

4.3 Motivated settings for selected methods

A problem with the original PC algorithm is that it is order-dependent. That is, the algorithm may in each of the 3 steps be dependent on the order in which the variables are presented (Colombo and Maathuis, 2014). After experimenting with the original algorithm it is concluded that also in this case the original algorithm suffers from the order-dependence, especially with the orientation of the edges. The methods PC-stable, PC-conservative and Solve Conflict are implemented to solve the order-dependency problems. They are based upon the work of Colombo and Maathuis (2014), who partly based their work on the original conservative algorithm by Ramsey et al. (2012). The PC stable setting, which is the default in the pcalg package, solved the order-dependency in the creation of the skeleton. Furthermore, the conservative algorithm solves the order-dependency in the detection of collider structures. Lastly, the method Solve Conflict allows for bidirectional edges whenever we have information that conflicting collider structures have been detected. We could for example

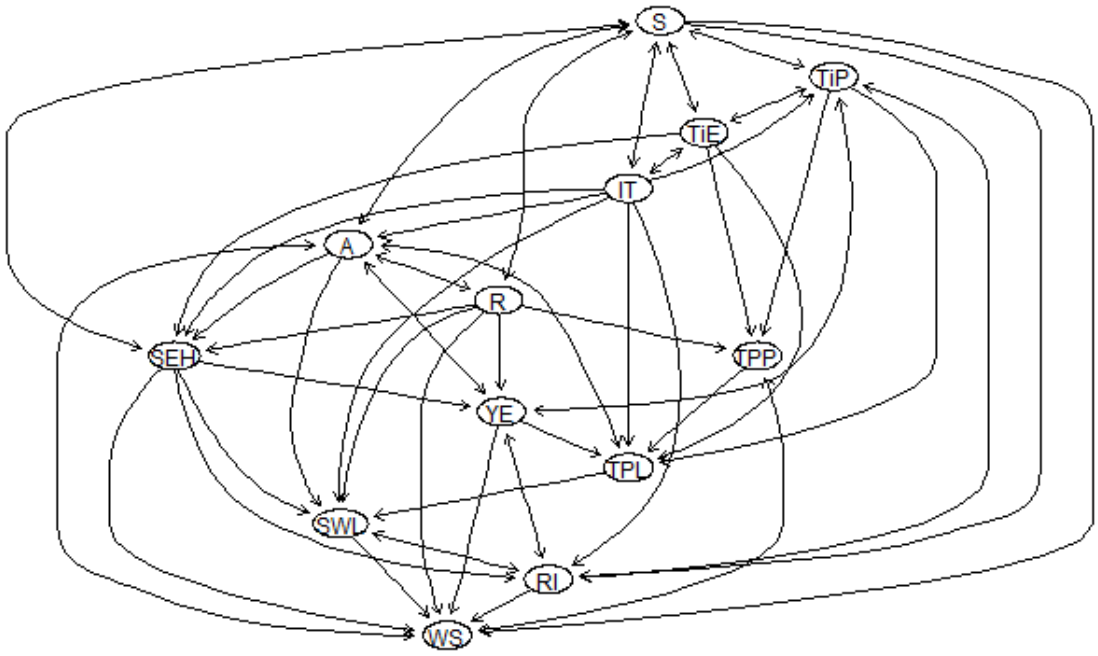
have detected a collider structure with an arrow going from A to B, but also a collider with an arrow going from B to A. The original PC algorithm took a preference for the orientation which was detected first. The Solve Conflict method however allows then for bidirectional edges. These edges should not be interpreted causally but are a sign that the assumption that there were no hidden variables is problematic for these variables (Colombo and Maathuis, 2014; Kalisch et al., 2020).

Moreover, settings to incorporate missing and mixed data were implemented by using the package micd based upon the work of Witte et al. (2021). The setting mixMItest allowed the implementation of multiple imputed mixed data. The multiple imputated data is incorporated by setting in the pc algorithm. Moreover, this setting allows for a combination of numerical (continuous) and categorical data, which is necessary, since the used data is of a mixed data type. The used data actually arrived in an ordinal form, which is of ordered factor data in R. However, the micd package and the setting mixMitest can not handle this type of data, and it is advised to turn the ordinal data into numeric data instead by the authors themselves. One sample is used for the testwise deletion method for handling missing data as a comparison. The method mixCItwd from the package micd employs testwise deletion in the pc algorithm. Finally, the setting with alpha = 0,05 is used to create 95 percent confidence in our conditional independence tests used in the pc algorithm. Lastly, the local method in the IDA algorithm is selected to allow for faster convergence. The global method is only advised to use for maximal 10 variables. The unique estimations between the global and local methods will be equal, however, the distributions of those variables may differ. Therefore, the lower and upper bounds of the local method are sound estimates of the uncertainty in the causal effect (Kalisch et al., 2012; Maathuis et al., 2009).

5 Results

The results from the analysis come in two parts. First, the CPDAG representing the causal structure in the variables is estimated. The CPDAG for Sample 1 is shown in Figure 11. The legend for this CPDAG is found in Table 1. For the CPDAG for samples 2-5 obtained through the multiple imputation setting and sample 6 through the testwise deletion setting, see Appendix 7.4. Due to the long runtimes of the algorithm (c.a. 8 hours per sample), it was unfeasible to obtain the CPDAG for even more samples. However, results seem to be quite robust over the samples, so that provides credibility to the idea that the results are quite reliable. To recapture, a CPDAG describes two aspects of the causal structure. First, which variables possess a direct and which only have an indirect relationship with each other indicated by the existence or non-existence of arrows between the variables. Second, the direction of the causal relationships is indicated by the direction of the arrow.

Figure 11: CPDAG estimated for Sample 1



Variable	Abbreviation	Variable	Abbreviation
Satisfaction With Life	SWL	Age	A
Economic Hardship	SEH	Relative Income	RI
Religiosity	R	Trust in Elections	TiE
Interpersonal Trust	IT	Gender	S
Trust in Political Parties	TPP	Years of Education	YE
Trust in Political Leader	TPL	Work Status	WS
Trust in Parliament	TiP		

Table 1: Legend of CPDAG

When looking at the estimated CPDAG's the first thing to note is that not all edges are one-directional, but also bidirectional edges are shown. These bidirectional edges do not imply any causal information but are merely a sign that the PC algorithm has detected multiple collider structures, due to the fact that in these relations probably some variables are still missing. Further research is therefore still necessary to become even more clear about the causal structure surrounding the Satisfaction with Life variable.

Hereafter, we observe that across the samples, the skeleton seems to be quite similar. To remind ourselves, the skeleton of the CPDAG is the way variables are directly or indirectly related but does not take the direction of the causal relationships into account. The variables Trust in Political Parties, Trust in Political Leader, Trust in Elections, Trust in Parliament and Interpersonal Trust form a cluster in most CPDAG's. Only in the CPDAG from Sample 4 are these variables not interrelated. Moreover, the variables Subjective Economic Hardship and Satisfaction with Life share a direct causal relationship in all estimated CPDAG's. The number of times that other variables share a direct causal link with the SWL variable can be seen in Table 2, where the number indicates in how many samples a direct relationship with SWL was found.

We see that apart from the Subjective Economic Hardship variable, also the variables Religiosity, Interpersonal Trust, Trust in Political Leader, Work Status, Relative Income and Age are estimated to share a direct causal relationship with the Satisfaction with Life variable in all or almost all samples. On the other hand, the variables Trust in Political Parties, Trust in Parliament, Years of Education, Gender and Trust in Elections are estimated to have no direct causal relationship with the Satisfaction with Life variable for none or almost none samples. The usefulness of this information can be seen in the following: If we attempt to influence or improve the Satisfaction with Life of people in Latin America, doing this through the improvement of variables that share a direct causal relationship with this variable is much more effective than intervening on variables that only influence this variable through other variables.

Variable	0	1	2	3	4	5
Economic Hardship						X
Religiosity					X	
Interpersonal Trust						X
Trust in Political Parties		X				
Trust in Political Leader						X
Trust in Parliament	X					
Work Status					X	
Years of Education	X					
Gender		X				
Trust in Elections	X					
Relative Income						X
Age						X

Table 2: Do variables share a direct causal link with SWL?

Hereafter, the causal effects of the different variables that were of interest in the preliminary paper on Satisfaction with Life were calculated using the IDA algorithm. Table 3 displays the results of this. The first column states the lower bound of the estimated causal effect, which is the lowest causal effect that was estimated. Oppositely, the second column states the higher bound of the estimated causal effect, which is naturally the highest causal effect that was estimated. It is not possible to report one causal effect, since we have results from 5 samples, and it is also possible that one sample outputs multiple estimations of the causal effect, since multiple adjustment sets may be valid per CPDAG.

In the third column, it is visible for how many samples the IDA algorithm was unable to estimate the causal effect and reported a NA. These NA can have two origins. Firstly, it is possible that the CPDAG in that sample estimated the causal direction to be the other way around. This fortunately only happened twice. Once in the estimation of the causal effect of economic hardship on Satisfaction with Life in samples 3 and 6 and once with the causal effect of religiosity on economic hardship in sample 5. The other reported NA were caused by the bidirectional edges in our CPDAG. If a causal effect has to be estimated across a causal path in which the direction is unclear, then the IDA algorithm will not be able, and rightly so, to estimate the causal effect. The fewer NA are present in a column, the more samples were able to output a causal effect, and the more trust we can place in our estimations of the causal effect. Lastly, the final column states as means of comparing the information provided by sample 6 in which testwise deletion instead of multiple imputation was used.

Variable	Lower Bound	Higher Bound	Number of NA	Test Deletion
Economic Hardship	-0.167	-0.157	1	NA
Religiosity	0.014	0.022	0	0.007
Interpersonal Trust	0.086	0.103	0	0.092
Trust in Political Parties	0.014	0.021	2	0.018
Trust in Political Leader	0.015	0.020	2	NA
Trust in Parliament	0.008	0.020	3	NA
Trust in Elections	0.019	0.030	0	0.029

Table 3: Causal Effects on Satisfaction with Life

The causal effect of the economic hardship variable on satisfaction with life is estimated to be between -0.167 and -0.157, which is between 15 and 17 percent. The interpretation of this is as follows: if we intervene on our economic hardship variable and increase that variable by 1, we expect the satisfaction with life variable for that specific person to decrease with that amount. If we can decrease the economic hardship variable by 1 the expectation is that this causes an increase in satisfaction with life between 15 and 17 percent.

The other variables are estimated to have a positive causal effect on satisfaction with life, on the other hand. The variable interpersonal trust is estimated to have the greatest causal impact of between 8 and 10 percent, while the trust in politics variables and the religiosity variable have an effect that is between 1 and 3 percent. The test deletion sample did not significantly differ in these outputs, except for the economic hardship variable, where

it estimated the causal direction to be the other way around. It was more often possible finally for the variables that are closer to the satisfaction with life variable in the CPDAG to estimate a causal effect. This is to be expected, since the shorter the causal path, the lower the chance of encountering bidirectional edges in that path. A path can be shorter or longer indicating how many variables are causally speaking between two variables.

In Table 4 the causal effects of the other variables on subjective economic hardship are shown to help answer hypothesis 2. The variable interpersonal trust is estimated to have a large negative effect on subjective economic hardship, which might explain as well its relatively high effect on satisfaction with life. This effect is estimated to be between -0.118 and -0.090. Surprisingly, the religiosity variable is estimated to have a positive causal effect on subjective economic hardship, with an effect of between 0.028 and 0.040. The trust in elections variable is also estimated to have a significant impact on subjective economic hardship, with a causal effect of between -0.048 and -0.030. Finally, the effect of the other variables is estimated to be around 0, so these variables have no significant causal impact.

Variable	Lower Bound	Higher Bound	Number of NA	Test Deletion
Religiosity	0.028	0.040	1	0.044
Interpersonal Trust	-0.118	-0.090	0	-0.104
Trust in Political Parties	-0.001	0.001	2	-0.001
Trust in Political Leader	-0.001	0.002	2	NA
Trust in Parliament	0.001	0.003	3	NA
Trust in Elections	-0.048	-0.030	0	-0.036

Table 4: Causal Effects on Subjective Economic Hardship

6 Conclusion and Discussion

6.1 Answering the data science question

The data science questions that were asked were the following: *What is the CPDAG that is consistent with the observational data of the LAPOP 2016/2017 survey?* And: *Based upon this CPDAG what can we estimate the strength and direction of the causal relationship between our variables in our hypotheses to be?* It was possible to estimate a CPDAG from the observational data, however, due to the use of samples somewhat different CPDAG's were estimated between samples. They were largely similar but still differed in the details. Important to state here is that the CPDAG's may not be entirely valid if the assumptions are not fully met. Especially, the assumption that there are no hidden variables may be problematic, as is shown by the fact that there were some bidirectional edges in each CPDAG. Nevertheless, I still believe that this research provides important insights into the causal structure related to satisfaction with life in Latin America since the results are quite robust under different samples and by using somewhat differing methods.

The second data science question was also able to be answered and presented in Table 2 and Table 3. The most important causal effects that were estimated were that of Subjective Economic Hardship on Satisfaction with Life (negative 15 - 17 percent), that of Interpersonal Trust on Satisfaction with Life (positive 8 - 10 percent) and that of Interpersonal Trust on Subjective Economic Hardship (negative 9 - 11 percent).

6.2 Answering the research question

The research question that guided this project was *Is it possible to gather information about the causal structure surrounding subjective well-being or happiness in Latin America?*. Moreover, the goal was to gather evidence for or against two hypotheses: *Economic hardship causes a decrease in Subjective Well-being.* And: *Institutional Quality has a different causal direction than Economic Hardship in their relationship to Subjective Well-Being, causes an increase in Subjective Well-Being and also has a negative causal influence on Economic Hardship.*

It can firstly be concluded that it was possible to gather information about the causal structure surrounding satisfaction with life. The estimated CPDAG's provided this information. Satisfaction with Life shares a direct causal relationship with Subjective Economic Hardship, Religiosity, Interpersonal Trust, Trust in Political Leader, Work Status, Relative Income and Age in almost all of these CPDAG's. Moreover, the estimated causal effect can help us argue in favour of the two hypotheses. We have estimated that Subjective Economic Hardship causes a decrease in Subjective Well Being of 15 - 17 percent. Moreover, we have estimated that the Institutional Quality variables (Interpersonal Trust, Religiosity, Trust in Political Leader, Trust in Elections, Trust in Parliament and Trust in Political Parties) have a positive causal effect on Satisfaction with Life.

Consequently, we can conclude that we have evidence to believe that Institutional Quality has a different causal direction than Economic Hardship and has a positive causal effect. The mitigating effect of these variables on Subjective Economic Hardship directly is mostly

present in the Interpersonal Trust variable, which has a large negative causal effect (9 - 11 percent) on Subjective Economic Hardship. Interpersonal Trust therefore directly influences the subjective experience of Economic Hardship and consequently can mitigate the negative causal effect of Subjective Economic Hardship on Satisfaction with Life.

6.3 Describing implications for domain

This research has multiple implications for the domain of happiness research and happiness research in Latin America. First, this research has, I believe, been one of the first that explicitly investigated the causal relationships surrounding happiness instead of assuming that happiness has been the dependent variable. The positive answer to the question of whether happiness is in most cases the dependent variable provides evidence in favour of this assumption and therefore helps in making other research which has assumed this causal direction more credible.

Moreover, the estimated CPDAG's are a first step in becoming clearer what exactly causes subjective well-being, satisfaction with life or happiness, which is crucial for policymaking and development theories. From this research, an important finding is that subjective economic hardship does indeed seem to have the largest impact on satisfaction with life. However, interpersonal trust also has a large impact and interpersonal trust does moreover have a large impact on subjective economic hardship as well. The implications of this are that economic development which destroys interpersonal trust across people may even increase the experience of economic hardship of individuals. In this way, economic development may not help to increase satisfaction with life. The picture that arises from this research is that we should aim for a decrease in subjective economic hardship combined with an increase in interpersonal trust and institutional quality at large. This is an important conclusion to keep in mind in creating policies for development. At least, our policies to enhance economic development should not decrease interpersonal trust.

This is a finding which may be only present in Latin America, but it should be investigated whether this finding holds for other parts of the world. Even when this is not the case, and especially when this is not the case, we should adopt policies that are tuned to the situation and mentality of the people in Latin America. In the western world for example interpersonal trust may be less important, and it may be possible that overlooking this crucial component is the reason why exported European policies have not worked in the Latin American context. Finally, when creating personal interventions that can be used by for example NGOs, this information is crucially important. NGOs attempt to devise interventions which improve the lives of individuals. Causal information is paramount, therefore, since NGOs should know and understand which interventions are more likely to improve the experience of people in Latin America.

An important limitation of this research however is that the data came from a cohort study, and individuals or countries have consequently not been followed across time. A next step would be to extend this research to time series data, which can answer hopefully the question of what causes individuals to be satisfied across their entire lives. Moreover, it may

be possible that even more information about the causal structures and relationships can be identified due to the time component. Causality runs from earlier to later, and therefore the inclusion of a time component is an important improvement of the current research. Nevertheless, this research provides an important first step and without further research, we should aim to adopt its conclusions in our policymaking, I believe.

7 Appendix

7.1 Annotated scripts of analyses and method settings

```
library(dplyr)
library(tidyverse)
library(mice)
library(haven)
library(dagitty)
library(qgraph)
library(pcalg)
library(micd)
library(RBGL)
library(naniar)
library(ggplot2)
library(hexbin)
library(ggcorrplot)
library(UpSetR)
library(lattice)
library(howManyImputations)
library(mifa)
library(bmem)
```

```
#Prepare Data
#Select Correct Countries From DataSet
Countries <- Merged_2016 %>%
  filter(pais == 1 | pais == 2 | pais == 3 | pais == 4 | pais == 5 | pais == 6 |
           pais == 7 | pais == 8 | pais == 9 | pais == 10 | pais == 11 |pais == 12 |
           pais == 13 |  pais == 14 | pais == 15| pais == 16| pais == 17 | pais == 21)
#Select Correct Variables From DataSet
Variable_Data <- Countries %>%
  select(weight1500, q1, q2, b21, b21a, b13, b47a, cp6, it1, q10d,ls3, ed, q10new,ocup4a)
```

```
#Rename Variables
names<-c('Country_Weight', 'Sex', 'Age', 'Trust_in_Politcal_Parties',
         'Trust_in_Political_Leader', 'Trust_in_Parliament', 'Trust_in_Elections',
         'Religiosity',   'Interpersonal_Trust', 'Subjective_Economic_Hardship',
         'Satisfaction_With_Life' ,  'Years_of_Education' , 'Relative_Income' ,
         'Work_Status')
colnames(Variable_Data) <- names
#Short names for graph
names_short<-c('S','A','TPP','TPL','TiP','TiE','R','IT','SEH','SWL','YE','RI', 'WS' )
```

```
#Work Status is a categorical variable, which for our purposes only informative
#if people work or not. So new variable created with working or not working
Better_Data<-mutate(Variable_Data,Work_Status=case_when
                    (Work_Status > 1 ~ "2", Work_Status  == 1  ~ "1"))
Better_Data$Work_Status<- as.numeric(Better_Data$Work_Status)

#Variables are recorded differently. The data is mutated such that a higher value
#signifies a higher number in real life for that variable
Prepared_Data<-mutate(Better_Data,Satisfaction_With_Life = 5 - Satisfaction_With_Life,
                  Interpersonal_Trust= 5 - Interpersonal_Trust,
                  Religiosity= 5 - Religiosity,
                  Work_Status = 3 - Work_Status)
```

```r
#Basic Exploration
summary(Prepared_Data)
#Bar Plots of Data
ggplot(data = Prepared_Data) +
  geom_bar(mapping = aes(x = Sex))
ggplot(data = Prepared_Data) +
  geom_bar(mapping = aes(x = Trust_in_Politcal_Parties))
ggplot(data = Prepared_Data) +
  geom_bar(mapping = aes(x = Trust_in_Political_Leader))
ggplot(data = Prepared_Data) +
  geom_bar(mapping = aes(x = Trust_in_Parliament))
ggplot(data = Prepared_Data) +
  geom_bar(mapping = aes(x = Trust_in_Elections))
ggplot(data = Prepared_Data) +
  geom_bar(mapping = aes(x = Religiosity))
ggplot(data = Prepared_Data) +
  geom_bar(mapping = aes(x = Interpersonal_Trust))
ggplot(data = Prepared_Data) +
  geom_bar(mapping = aes(x = Subjective_Economic_Hardship))
ggplot(data = Prepared_Data) +
  geom_bar(mapping = aes(x = Satisfaction_With_Life))
ggplot(data = Prepared_Data) +
  geom_bar(mapping = aes(x = Relative_Income))
ggplot(data = Prepared_Data) +
  geom_bar(mapping = aes(x = Work_Status))
#Histograms of Data
ggplot(data = Prepared_Data) +
  geom_histogram(mapping = aes(x = Age))
ggplot(data = Prepared_Data) +
  geom_histogram(mapping = aes(x = Years_of_Education))
```

```r
#Relations in Data
#Heatmaps
Prepared_Data %>%
  count(Subjective_Economic_Hardship, Satisfaction_With_Life) %>%
  ggplot(mapping = aes(x = Subjective_Economic_Hardship, y = Satisfaction_With_Life)) +
    geom_tile(mapping = aes(fill = n))
Prepared_Data %>%
  count(Interpersonal_Trust, Satisfaction_With_Life) %>%
  ggplot(mapping = aes(x = Interpersonal_Trust, y = Satisfaction_With_Life)) +
    geom_tile(mapping = aes(fill = n))
Prepared_Data %>%
  count(Religiosity, Satisfaction_With_Life) %>%
  ggplot(mapping = aes(x = Religiosity, y = Satisfaction_With_Life)) +
    geom_tile(mapping = aes(fill = n))
#Correlation Matrix
Data_For_Correlation_Matrix<-  Prepared_Data %>%
  select( Trust_in_Politcal_Parties, Trust_in_Political_Leader, Trust_in_Parliament,
          Trust_in_Elections, Religiosity,   Interpersonal_Trust,
          Subjective_Economic_Hardship,  Satisfaction_With_Life)
res <- cor(na.omit(Data_For_Correlation_Matrix))
Correlation_matrix<-round(res, 2)
ggcorrplot(Correlation_matrix)
```

```r
#Missing Values
totalcells = 29064*13
missingcells = sum(is.na(Prepared_Data))
percentage = (missingcells/totalcells)
print(totalcells)
print(missingcells)
print(percentage)
#Visualize Missing Values
gg_miss_var(Prepared_Data)
gg_miss_var(Prepared_Data, show_pct = TRUE)
#Find out which variable has missing in first question Sex
which(is.na(Prepared_Data$Sex))
#Inspect that Row
Prepared_Data[26413,]
#Empty Row, so Remove Row
Correct_Data = Prepared_Data[-c(26413),]
gg_miss_upset(Correct_Data)
#Test whether Missingness is missing Complete at Random(MCAR)
mcar_test(Correct_Data)
#Data is not MCAR, high Chi-square Statistic, p-value of 0, 297 Missing patterns detected
```

```r
#Making sure that the variables are the right variable type for the
#Imputation and Causal Learning
Correct_Data$Country_Weight<- as.numeric(Correct_Data$Country_Weight)
Correct_Data$Sex<- as.factor(Correct_Data$Sex)
Correct_Data$Age<- as.numeric(Correct_Data$Age)
Correct_Data$Trust_in_Politcal_Parties<-as.numeric(Correct_Data$Trust_in_Politcal_Parties)
Correct_Data$Trust_in_Political_Leader<-as.numeric(Correct_Data$Trust_in_Political_Leader)
Correct_Data$Trust_in_Parliament<- as.numeric(Correct_Data$Trust_in_Parliament)
Correct_Data$Trust_in_Elections<- as.numeric(Correct_Data$Trust_in_Elections)
Correct_Data$Religiosity<- as.numeric(Correct_Data$Religiosity)
Correct_Data$Interpersonal_Trust<- as.numeric(Correct_Data$Interpersonal_Trust)
Correct_Data$Subjective_Economic_Hardship<-as.numeric(
  Correct_Data$Subjective_Economic_Hardship)
Correct_Data$Satisfaction_With_Life<- as.numeric(Correct_Data$Satisfaction_With_Life)
Correct_Data$Years_of_Education<- as.numeric(Correct_Data$Years_of_Education)
Correct_Data$Relative_Income<- as.numeric(Correct_Data$Relative_Income)
Correct_Data$Work_Status<- as.factor(Correct_Data$Work_Status)
```

```r
#Check whether this is done successfully
str(Correct_Data)
```

```r
#The dataset is weighted with variables for country weight. Samples are taken to get a
#representative analysis with as probability that an entry is included in the
#sample the Country Weight
set.seed(8)
Sample_Data_1<-sample_frac(Correct_Data,0.80, prob=Correct_Data$Country_Weight)
set.seed(27)
Sample_Data_2<-sample_frac(Correct_Data,0.80, prob=Correct_Data$Country_Weight)
set.seed(100)
Sample_Data_3<-sample_frac(Correct_Data,0.80, prob=Correct_Data$Country_Weight)
set.seed(19)
Sample_Data_4<-sample_frac(Correct_Data,0.80, prob=Correct_Data$Country_Weight)
```

```r
set.seed(4)
Sample_Data_5<-sample_frac(Correct_Data,0.80, prob=Correct_Data$Country_Weight)
set.seed(111)
Sample_Data_6<-sample_frac(Correct_Data,0.80, prob=Correct_Data$Country_Weight)
set.seed(88)
Sample_Data_7<-sample_frac(Correct_Data,0.80, prob=Correct_Data$Country_Weight)
```

```r
#Removing the Country Weight Variable from the Sample Data
Sample_Data_1<-subset(Sample_Data_1, select= -c(Country_Weight))
Sample_Data_2<-subset(Sample_Data_2, select= -c(Country_Weight))
Sample_Data_3<-subset(Sample_Data_3, select= -c(Country_Weight))
Sample_Data_4<-subset(Sample_Data_4, select= -c(Country_Weight))
Sample_Data_5<-subset(Sample_Data_5, select= -c(Country_Weight))
Sample_Data_6<-subset(Sample_Data_6, select= -c(Country_Weight))
Sample_Data_7<-subset(Sample_Data_7, select= -c(Country_Weight))
```

```r
#estimate How Many Imputations Necessary

impdata1 <- mice(Sample_Data_1, m = 100, maxit = 10)
modelFit1 <- with(impdata1, lm(Relative_Income ~ Sex + Age + Trust_in_Politcal_Parties +
                                 Trust_in_Political_Leader + Trust_in_Parliament +
                                 Trust_in_Elections + Religiosity +    Interpersonal_Trust
                               + Subjective_Economic_Hardship + Satisfaction_With_Life
                               + Years_of_Education  + Work_Status))
```

```r
how_many_imputations(modelFit1, c = 0.05)
```

```r
#Multiple Imputation on Missing Data, automatically mice
#chooses the right imputation method based upon the variable type
mi_object_1 <- mice(Sample_Data_1, m=10, maxit=10)
```

```r
#Combining the different imputations into a single object
mi_dat_1 <- complete(mi_object_1, action = "all")
```

```r
#Check Imputed Data
plot(mi_object_1)
#Check whether there is healthy convergence
```

```r
#Check whether our imputation results match the observed data
bwplot(mi_object_1)
propplot(mi_object_1)
```

```r
#Perform the PC algorithm, with an independence test for Mixed Data and alpha=0,05
#(95% confidence)
pc_outcome_sample_1 <- pc(suffStat = mi_dat_1, indepTest = mixMItest , u2pd= 'relaxed',
                          conservative  = TRUE, solve.confl = TRUE,alpha = 0.05,
                          labels = colnames(Sample_Data_1))
```

```r
#Plot the outcome of the PC Algorithm as a CPDAG
plot(pc_outcome_sample_1,labels=names_short, main= 'Causality Sample 1')
```

```r
#Sample 2 Full Code
mi_object_2 <- mice(Sample_Data_2, m=10, maxit=10)
mi_dat_2 <- complete(mi_object_2, action = "all")
plot(mi_object_2)
bwplot(mi_object_2)
propplot(mi_object_2)
pc_outcome_sample_2 <- pc(suffStat = mi_dat_2, indepTest = mixMItest ,u2pd= 'relaxed',
                          conservative = TRUE, solve.confl = TRUE,alpha = 0.05,
                          labels = colnames(Sample_Data_2))
plot(pc_outcome_sample_2,labels=names_short, main= 'Causality Sample 2')
```

```r
#Sample 3 Full Code
mi_object_3 <- mice(Sample_Data_3, m=10, maxit=10)
mi_dat_3 <- complete(mi_object_3, action = "all")
plot(mi_object_3)
bwplot(mi_object_3)
propplot(mi_object_3)
pc_outcome_sample_3 <- pc(suffStat = mi_dat_3, indepTest = mixMItest , u2pd= 'relaxed',
                          conservative = TRUE, solve.confl = TRUE,alpha = 0.05,
                          labels = colnames(Sample_Data_3))
plot(pc_outcome_sample_3,labels=names_short, main= 'Causality Sample 3')
```

```r
#Sample 4 Full Code
mi_object_4 <- mice(Sample_Data_4, m=10, maxit=10)
mi_dat_4 <- complete(mi_object_4, action = "all")
plot(mi_object_4)
bwplot(mi_object_4)
propplot(mi_object_4)
pc_outcome_sample_4 <- pc(suffStat = mi_dat_4, indepTest = mixMItest ,u2pd= 'relaxed',
                          conservative = TRUE, solve.confl = TRUE,alpha = 0.05,
                          labels = colnames(Sample_Data_4))
plot(pc_outcome_sample_4,labels=names_short, main= 'Causality Sample 4')
```

```r
#Sample 5 Full Code
mi_object_5 <- mice(Sample_Data_5, m=10, maxit=10)
mi_dat_5 <- complete(mi_object_5, action = "all")
plot(mi_object_5)
bwplot(mi_object_5)
propplot(mi_object_5)
pc_outcome_sample_5 <- pc(suffStat = mi_dat_5, indepTest = mixMItest , u2pd= 'relaxed',
                          conservative = TRUE, solve.confl = TRUE,alpha = 0.05,
                          labels = colnames(Sample_Data_5))
plot(pc_outcome_sample_5,labels=names_short, main= 'Causality Sample 5')
```

```r
#Sample 6 Full Code Test Wise Deletion
pc_outcome_sample_6_Testdeletion <- pc(suffStat = Sample_Data_6, indepTest = mixCItwd ,
                                       u2pd= 'relaxed',conservative = TRUE,
                                       solve.confl = TRUE, alpha = 0.05,
                                       labels = colnames(Sample_Data_6))
plot(pc_outcome_sample_6_Testdeletion,labels=names_short, main= 'Causality Sample 6')
```

```
#Calculation of Causation in Correlation Matrix is unfortunately only possible
# with numeric values. #Therefore, calculations of causation with the variables
#Sex and Work Status are not possible,
# but these variables have to be converted to numeric.
Sample_Data_1$Work_Status<- as.numeric(Sample_Data_1$Work_Status)
Sample_Data_1$Sex<- as.numeric(Sample_Data_1$Sex)
Sample_Data_2$Work_Status<- as.numeric(Sample_Data_2$Work_Status)
Sample_Data_2$Sex<- as.numeric(Sample_Data_2$Sex)
Sample_Data_3$Work_Status<- as.numeric(Sample_Data_3$Work_Status)
Sample_Data_3$Sex<- as.numeric(Sample_Data_3$Sex)
Sample_Data_4$Work_Status<- as.numeric(Sample_Data_4$Work_Status)
Sample_Data_4$Sex<- as.numeric(Sample_Data_4$Sex)
Sample_Data_5$Work_Status<- as.numeric(Sample_Data_5$Work_Status)
Sample_Data_5$Sex<- as.numeric(Sample_Data_5$Sex)
Sample_Data_6$Work_Status<- as.numeric(Sample_Data_6$Work_Status)
Sample_Data_6$Sex<- as.numeric(Sample_Data_6$Sex)
```

```
#Test Whether a longer imputation makes a difference
cov_var_sample_1_test<-mifa(Sample_Data_1,m=20, maxit=100)
cov_combined_1_test<-cov_var_Sample_1_test[["cov_combined"]]
ida(9,10,cov_combined_1_test,pc_outcome_sample_1@graph,verbose = TRUE,method = "local")
```

```
#Calculate Causation comparable with previous but with fewer iterations and imputations
cov_var_sample_1_test_2<-mifa(Sample_Data_1,m=10, maxit=10)
cov_combined_1_test_2<-cov_var_sample_1_test_2[["cov_combined"]]
ida(9,10,cov_combined_1_test_2, pc_outcome_sample_1@graph,verbose = TRUE,method = "local")
#Answer, it only makes a difference in the 4the decimal places. An imputation method with
#m=10 and maxit=10 as used in the observation of CPDAG is deemed suitable therefore.
```

```
#Calculate Covariance Matrices for all Samples
cov_var_sample_1<-mifa(Sample_Data_1,m=10, maxit=10)
cov_combined_1<-cov_var_sample_1[["cov_combined"]]
cov_var_sample_2<-mifa(Sample_Data_2,m=10, maxit=10)
cov_combined_2<-cov_var_sample_2[["cov_combined"]]
cov_var_sample_3<-mifa(Sample_Data_3,m=10, maxit=10)
cov_combined_3<-cov_var_sample_3[["cov_combined"]]
cov_var_sample_4<-mifa(Sample_Data_4,m=10, maxit=10)
cov_combined_4<-cov_var_sample_4[["cov_combined"]]
cov_var_sample_5<-mifa(Sample_Data_5,m=10, maxit=10)
cov_combined_5<-cov_var_sample_5[["cov_combined"]]
cov_combined_6<-bmem.list.cov(Sample_Data_6)
```

```
#Calculate Effects Of SEH on SWL (9 on 10)
ida(9,10, cov_combined_1, pc_outcome_sample_1@graph, verbose = TRUE, method = "local")
ida(9,10, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(9,10, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(9,10, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(9,10, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(9,10, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
```

```
# Institutions (Religiosity and Interpersonal Trust) on Satisfaction with Life
#Religiosity on SWL
ida(7,10, cov_combined_1, pc_outcome_sample_1@graph, verbose = TRUE, method = "local")
ida(7,10, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(7,10, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(7,10, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(7,10, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(7,10, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
```

```
#Interpersonal Trust on SSWL
ida(8,10, cov_combined_1, pc_outcome_sample_1@graph, verbose = TRUE, method = "local")
ida(8,10, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(8,10, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(8,10, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(8,10, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(8,10, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
```

```
#Institutional Quality (Trust in Political Parties, Political Leader, Parliament and Elections)
#on Satisfaction with Life
#Trust in Political Parties on SWL
ida(3,10, cov_combined_1, pc_outcome_sample_1@graph, verbose = TRUE, method = "local")
ida(3,10, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(3,10, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(3,10, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(3,10, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(3,10, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
```

```
#Trust in Political Leader on SWL
ida(4,10, cov_combined_1, pc_outcome_sample_1@graph, verbose = TRUE, method = "local")
ida(4,10, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(4,10, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(4,10, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(10,4, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(4,10, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
```

```
#Trust in  Parliament on SWL
ida(5,10, cov_combined_1, pc_outcome_sample_1@graph,verbose = TRUE, method = "local")
ida(5,10, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(5,10, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(5,10, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(5,10, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(5,10, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
```

```
#Trust in Elections on SWL
ida(6,10, cov_combined_1, pc_outcome_sample_1@graph, verbose = TRUE, method = "local")
ida(6,10, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(6,10, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
```

```
ida(6,10, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(6,10, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(6,10, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
```

```
#Trust in Political Parties, Political Leader, Parliament and Elections)
#on Subjective Economic Hardship

#Trust in Political Parties on SEH
ida(3,9, cov_combined_1, pc_outcome_sample_1@graph, verbose = TRUE, method = "local")
ida(3,9, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(3,9, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(3,9, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(3,9, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(3,9, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
#Trust in Political Leader on SEH
ida(4,9, cov_combined_1, pc_outcome_sample_1@graph, verbose = TRUE, method = "local")
ida(4,9, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(4,9, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(4,9, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(4,9, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(4,9, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
#Trust in  Parliament on SEH
ida(5,9, cov_combined_1, pc_outcome_sample_1@graph,verbose = TRUE, method = "local")
ida(5,9, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(5,9, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(5,9, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(5,9, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(5,9, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
#Trust in Elections on SEH
ida(6,9, cov_combined_1, pc_outcome_sample_1@graph, verbose = TRUE, method = "local")
ida(6,9, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(6,9, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(6,9, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(6,9, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(6,9, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
```

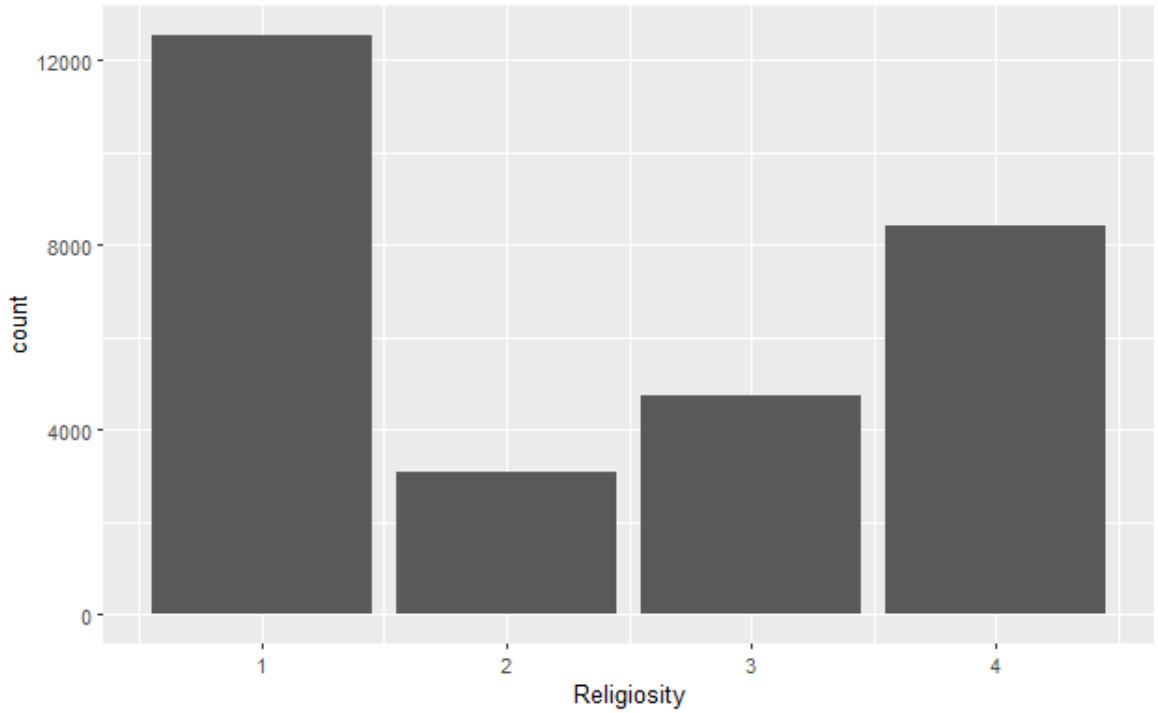
```
# Institutions (Religiosity and Interpersonal Trust) on SEH
#Religiosity on SEH
ida(7,9, cov_combined_1, pc_outcome_sample_1@graph, verbose = TRUE, method = "local")
ida(7,9, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(7,9, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(7,9, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(7,9, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(7,9, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
```

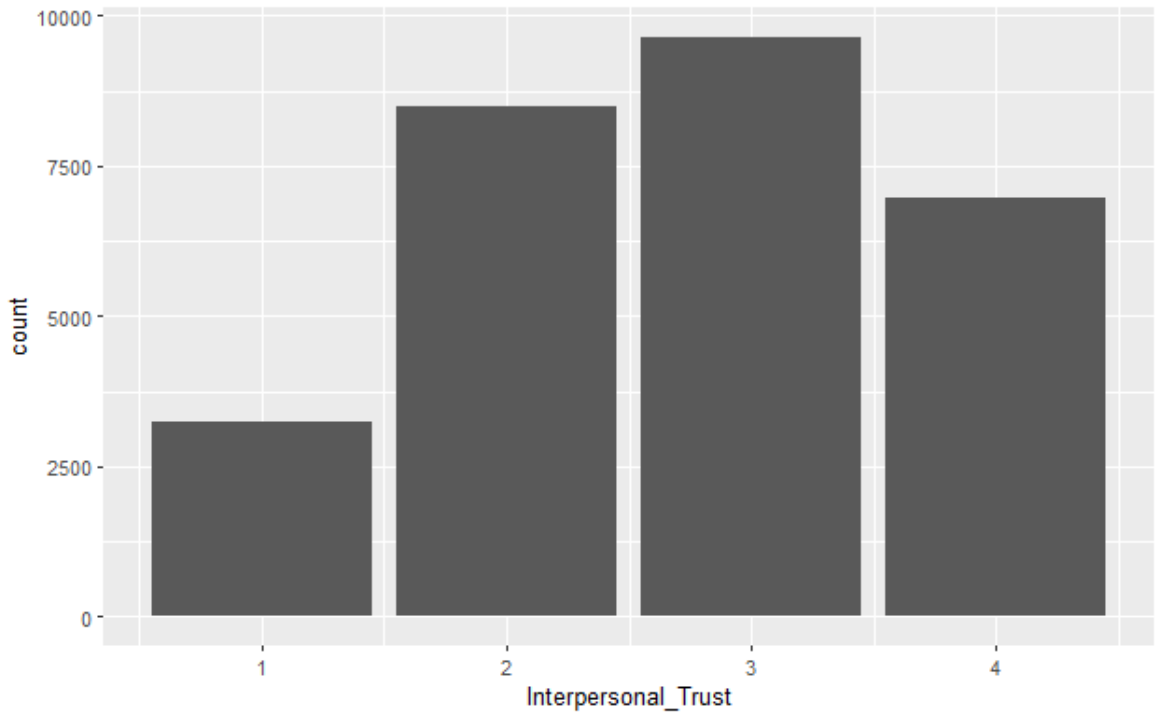
```
#Interpersonal Trust on SEH
ida(8,9, cov_combined_1, pc_outcome_sample_1@graph, verbose = TRUE, method = "local")
ida(8,9, cov_combined_2, pc_outcome_sample_2@graph, verbose = TRUE, method = "local")
ida(8,9, cov_combined_3, pc_outcome_sample_3@graph, verbose = TRUE, method = "local")
ida(8,9, cov_combined_4, pc_outcome_sample_4@graph, verbose = TRUE, method = "local")
ida(8,9, cov_combined_5, pc_outcome_sample_5@graph, verbose = TRUE, method = "local")
ida(8,9, cov_combined_6, pc_outcome_sample_6_Testdeletion@graph,
    verbose = TRUE, method = "local")
```

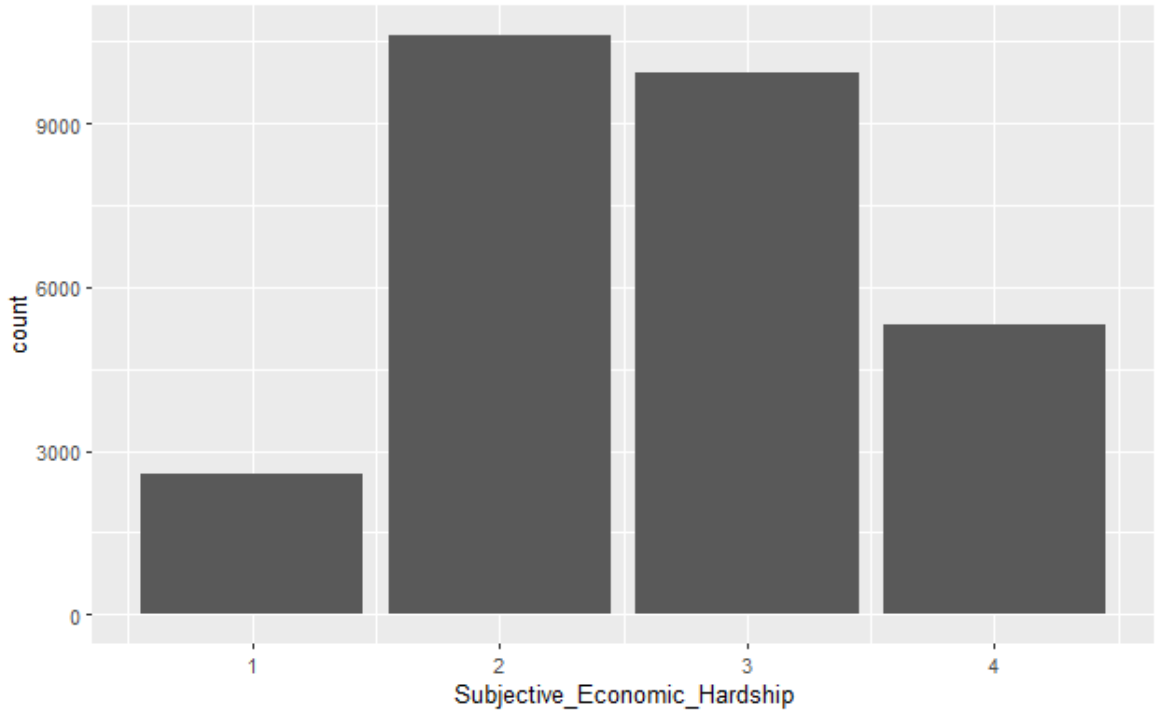
```
#Again Create Plots without Titles for Thesis
Sample1_plot<-plot(pc_outcome_sample_1,labels=names_short, main= '')
Sample2_plot<-plot(pc_outcome_sample_2,labels=names_short, main= '')
Sample3_plot<-plot(pc_outcome_sample_3,labels=names_short, main= '')
Sample4_plot<-plot(pc_outcome_sample_4,labels=names_short, main= '')
Sample5_plot<-plot(pc_outcome_sample_5,labels=names_short, main= '')
Sample6_plot<-plot(pc_outcome_sample_6_Testdeletion,labels=names_short, main= '')
```

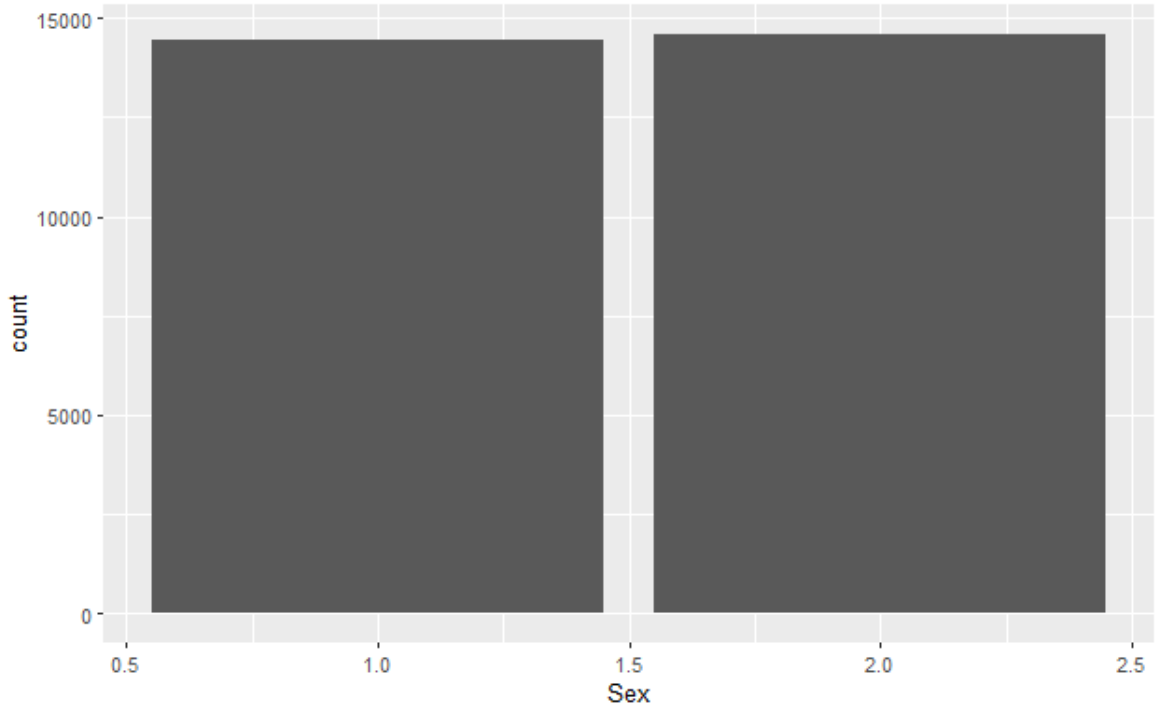
7.2 Full data exploration results

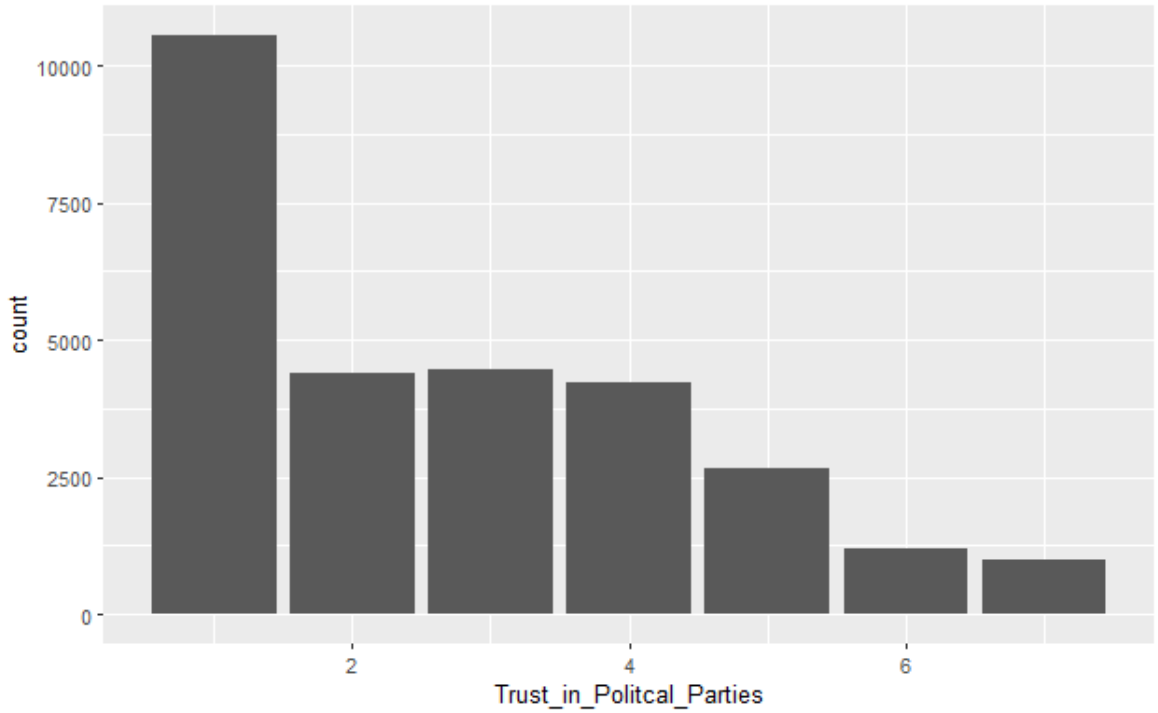
Histograms of Variables

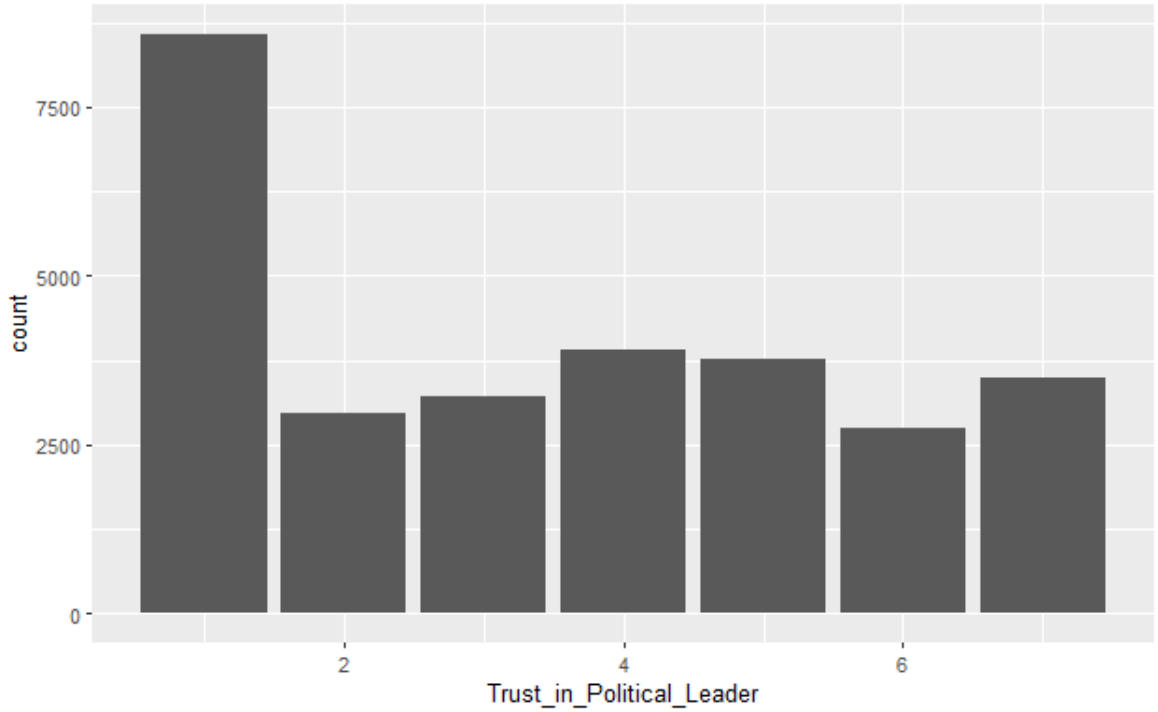


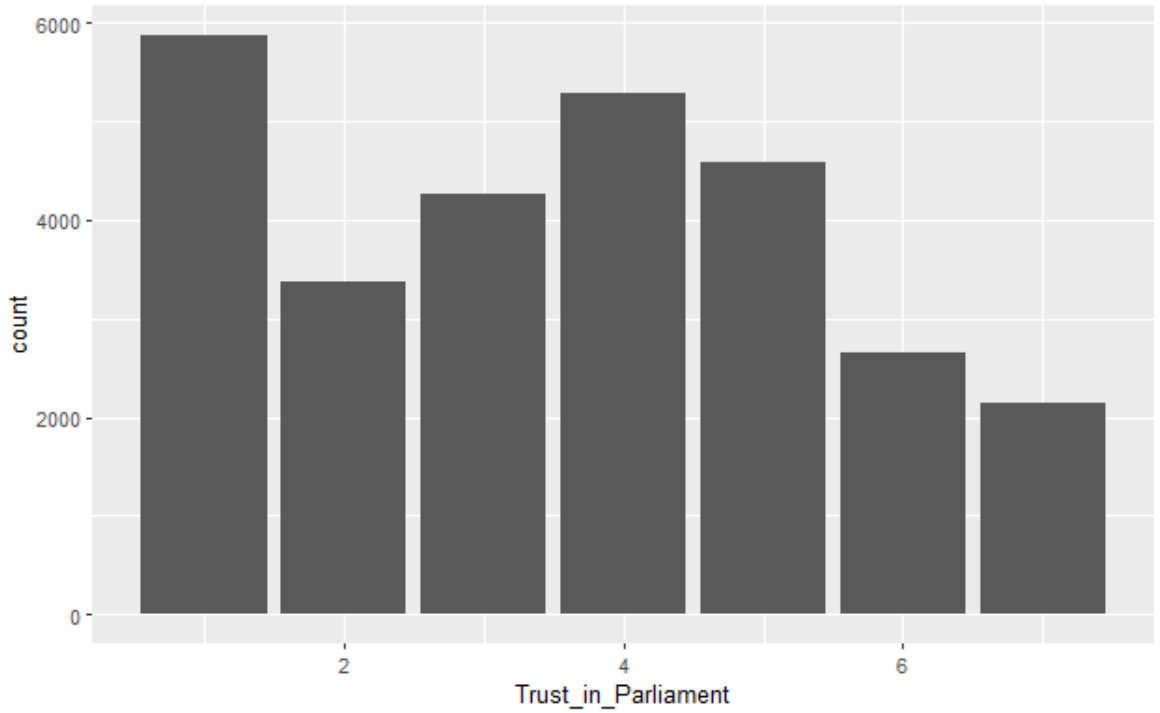


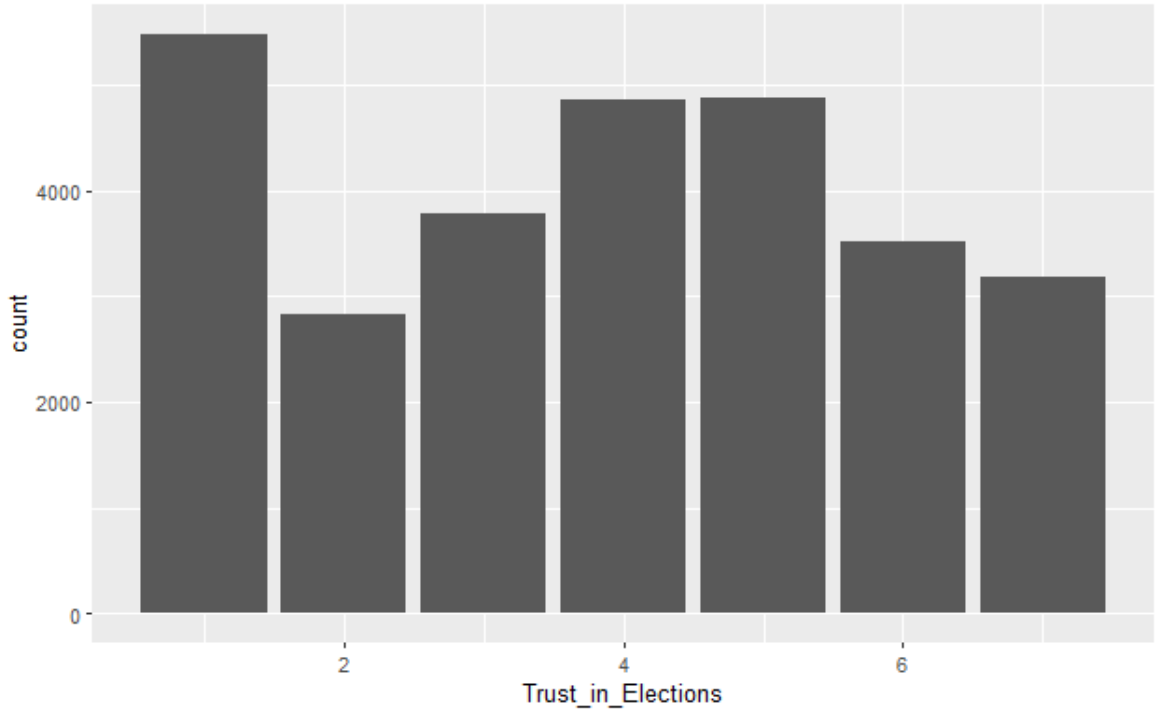

9000
6000
3000
0
count
1
2
3
4
Subjective_Economic_Hardship

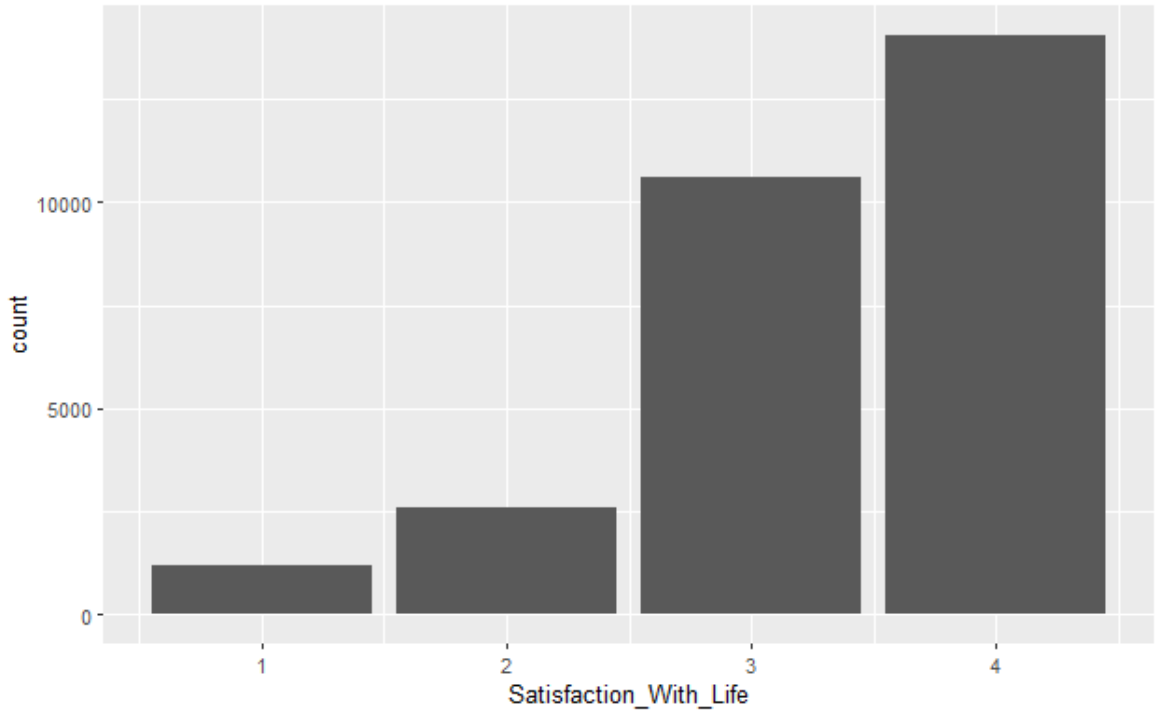

15000
10000
5000
0
count
0.5
1.0
1.5
2.0
2.5
Sex

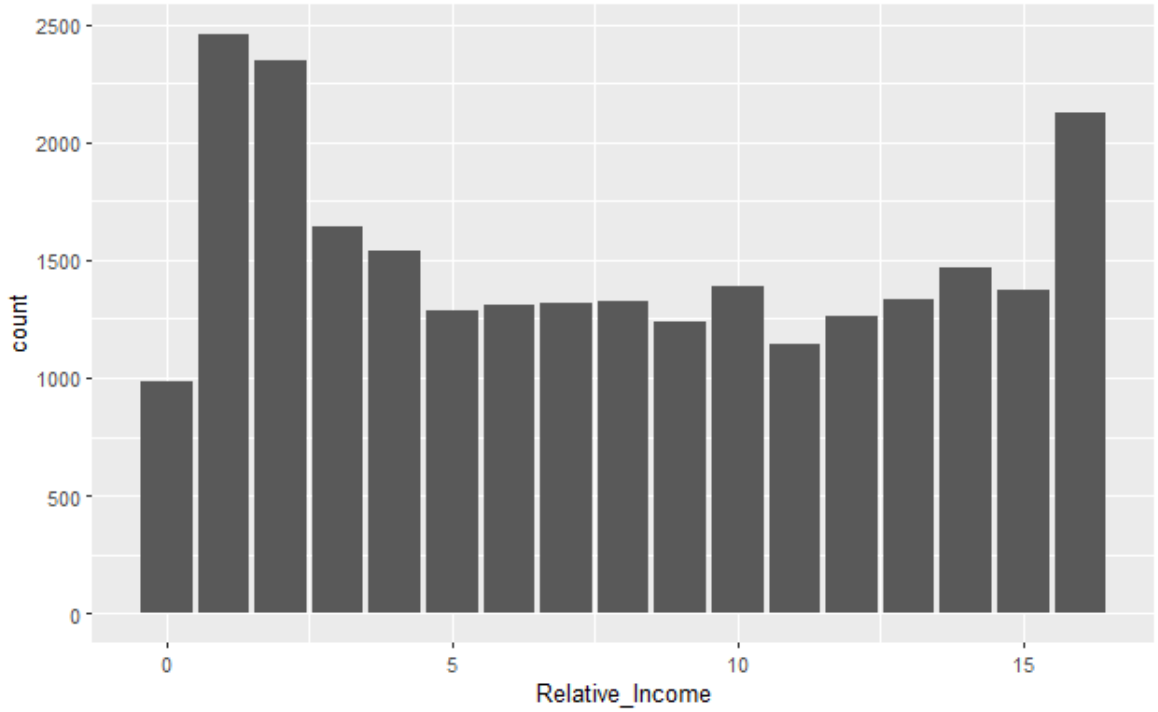

10000
7500
5000
2500
0
count
2
4
6
Trust_in_Politcal_Parties

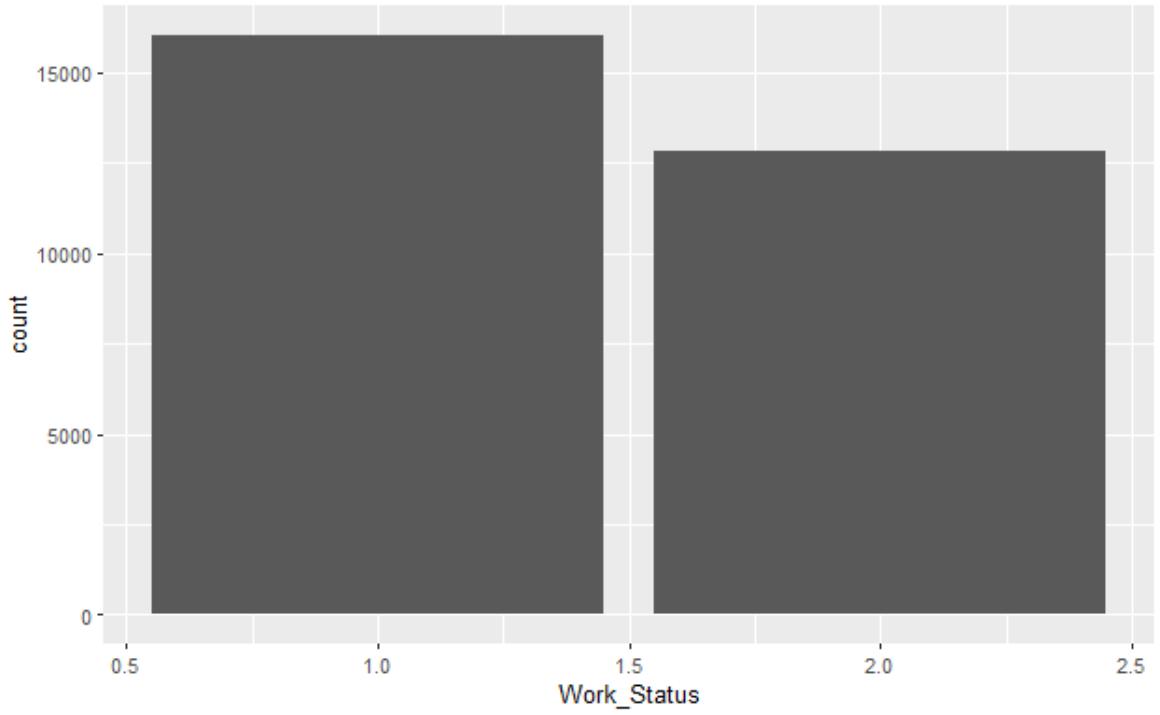

7500
5000
2500
0
count
2
4
6
Trust_in_Political_Leader

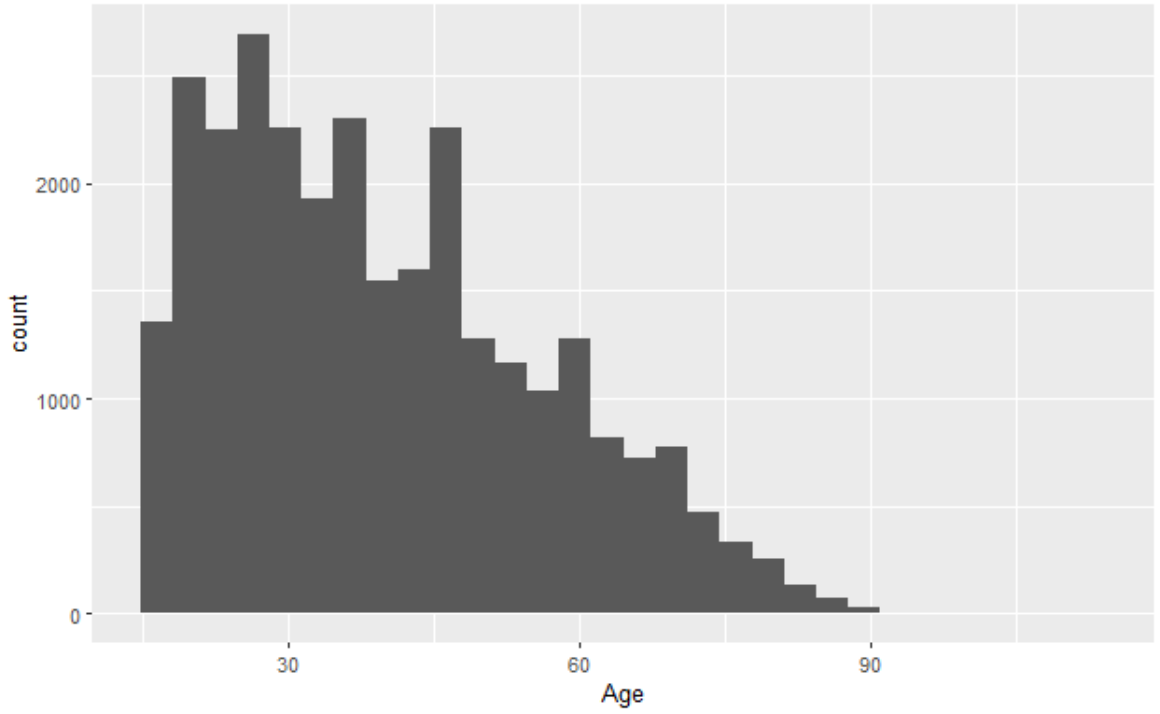

6000
4000
2000
0
count
2
4
6
Trust_in_Parliament

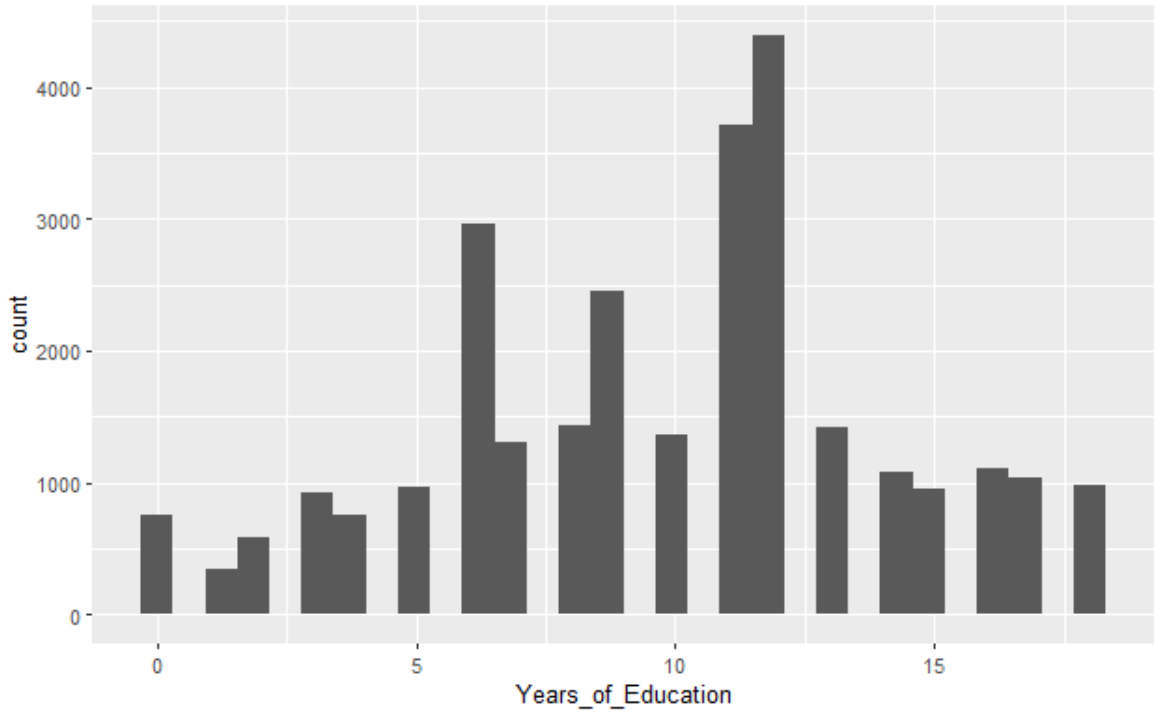

4000
2000
0
count
2
4
6
Trust_in_Elections


count
10000
5000
0
1
2
3
4
Satisfaction_With_Life

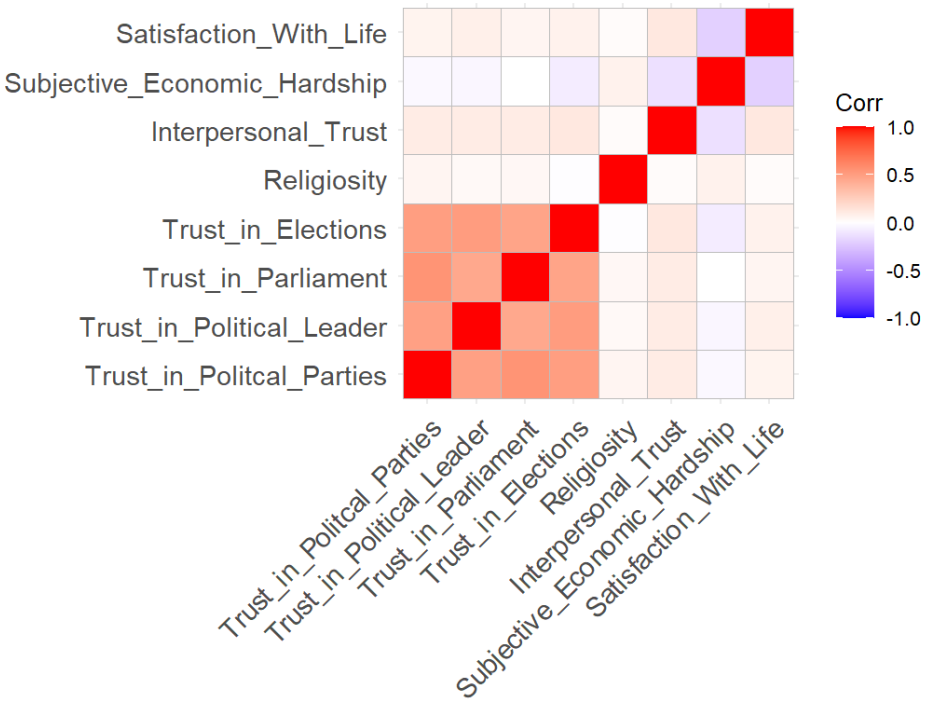

count
2500
2000
1500
1000
500
0
0
5
10
15
Relative_Income


15000
10000
5000
0
count
0.5
1.0
1.5
2.0
2.5
Work_Status

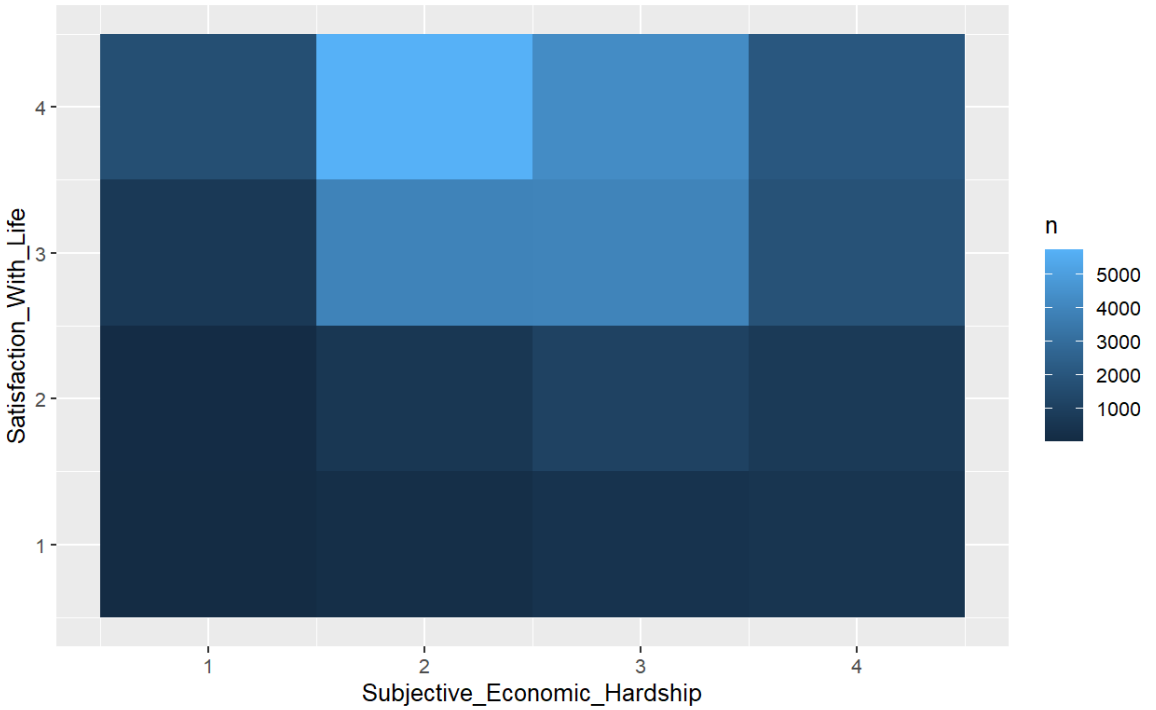

2000
1000
0
count
30
60
90
Age



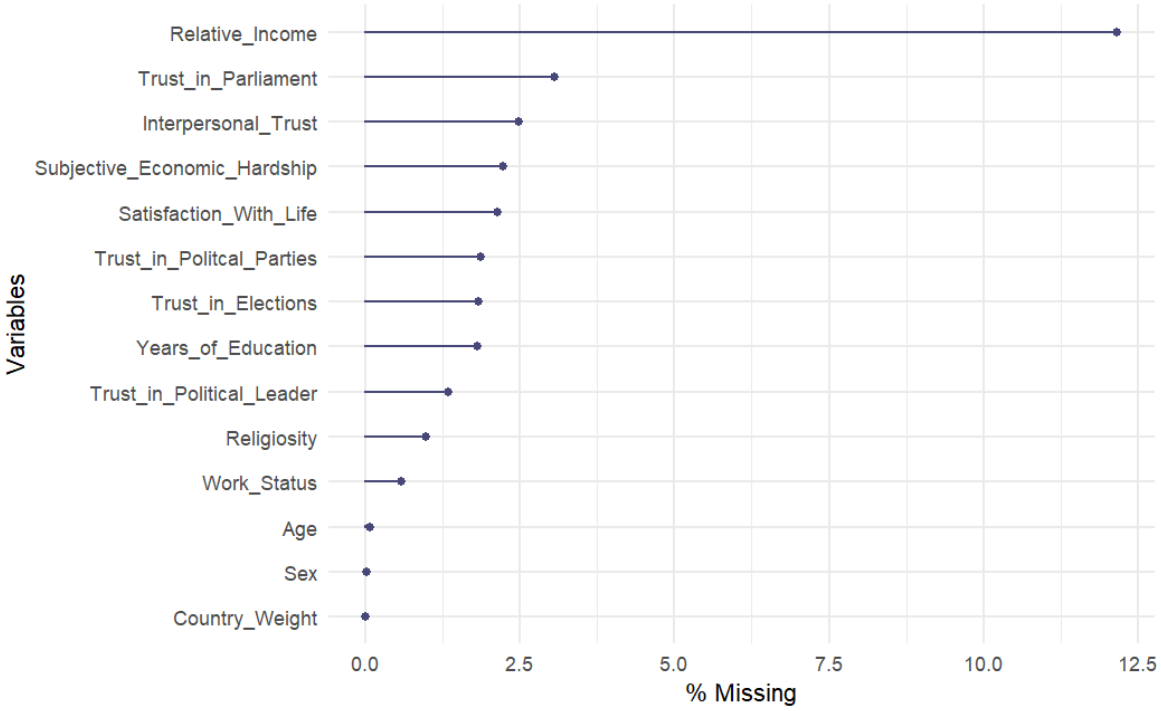
Correlation Matrix of Important Variables



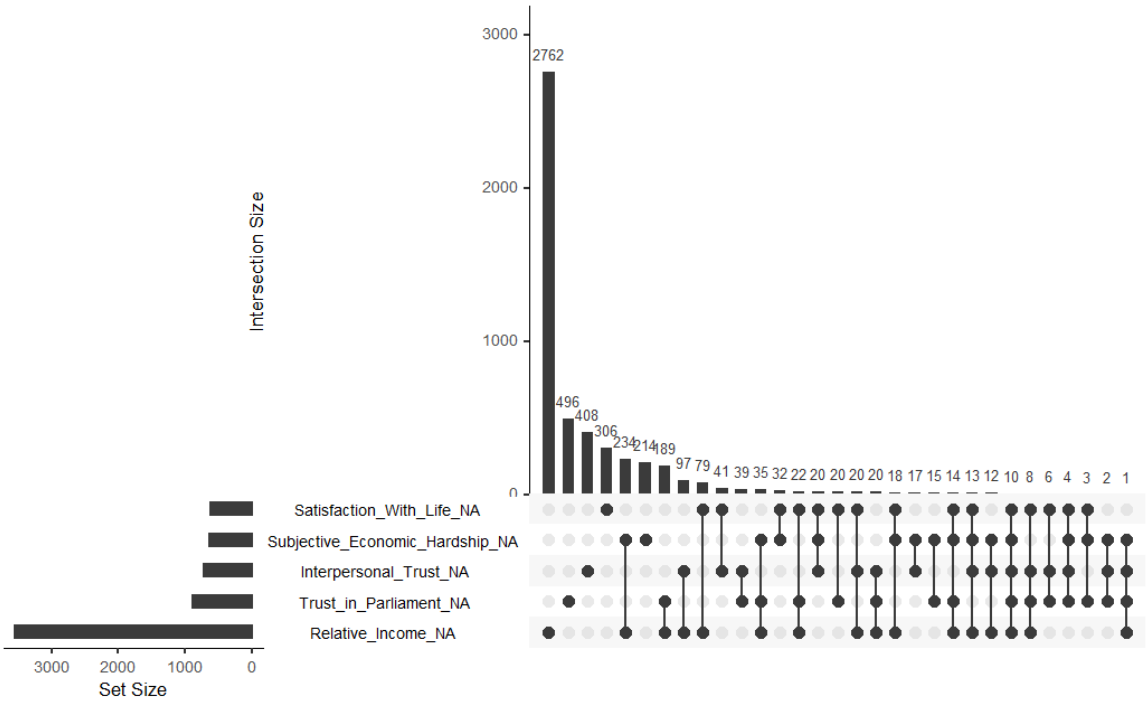
Heat Map of SWL and SEH



Missingness Percentage per Variable

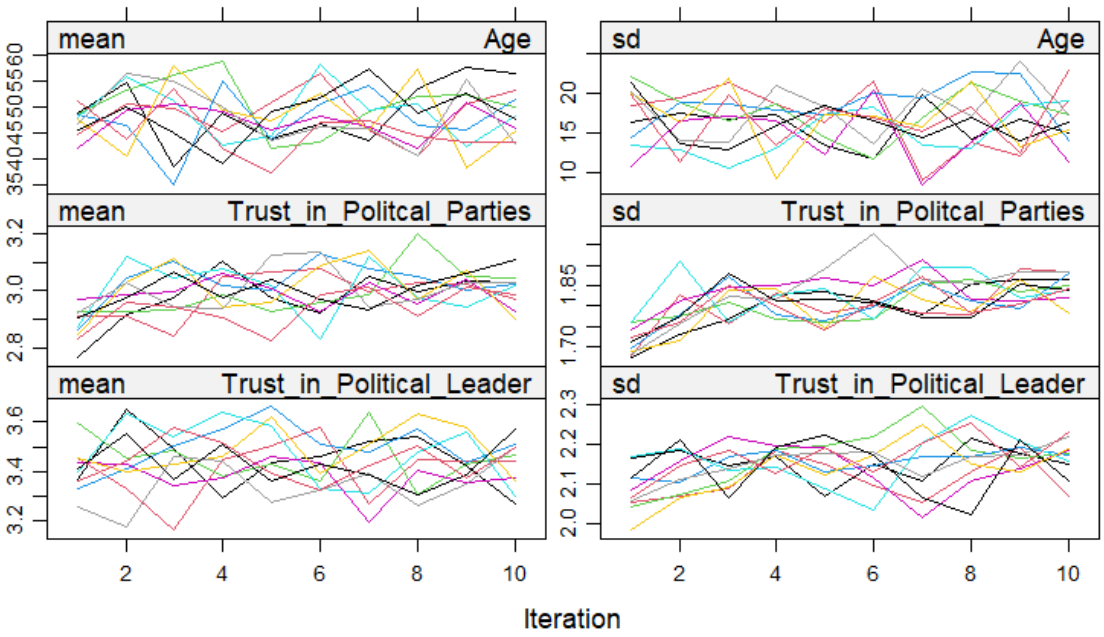


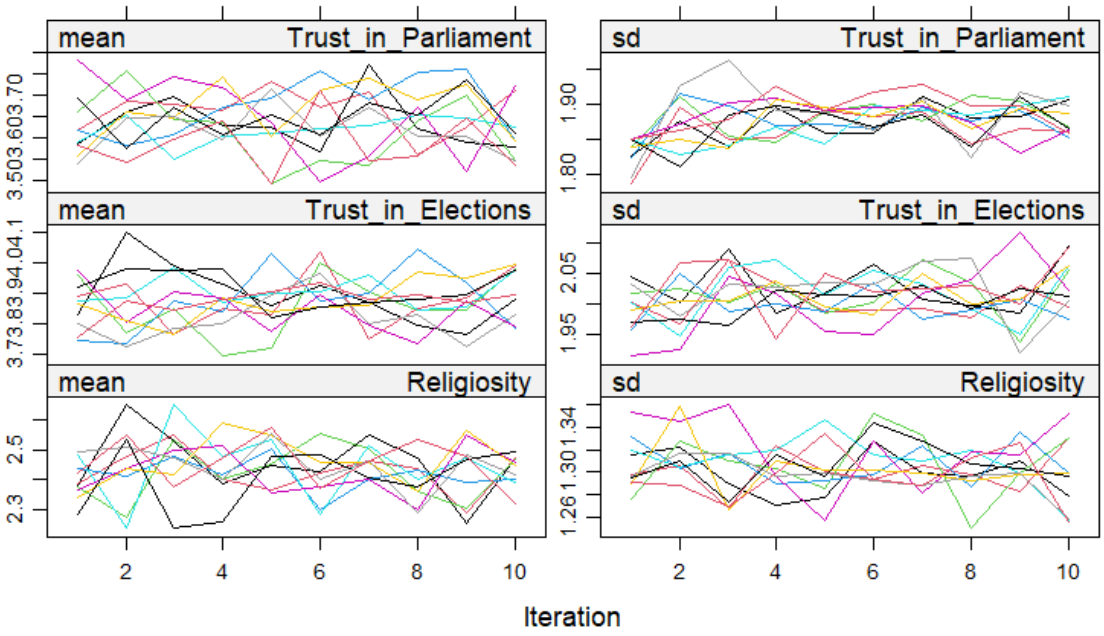
Relationships in Missingness

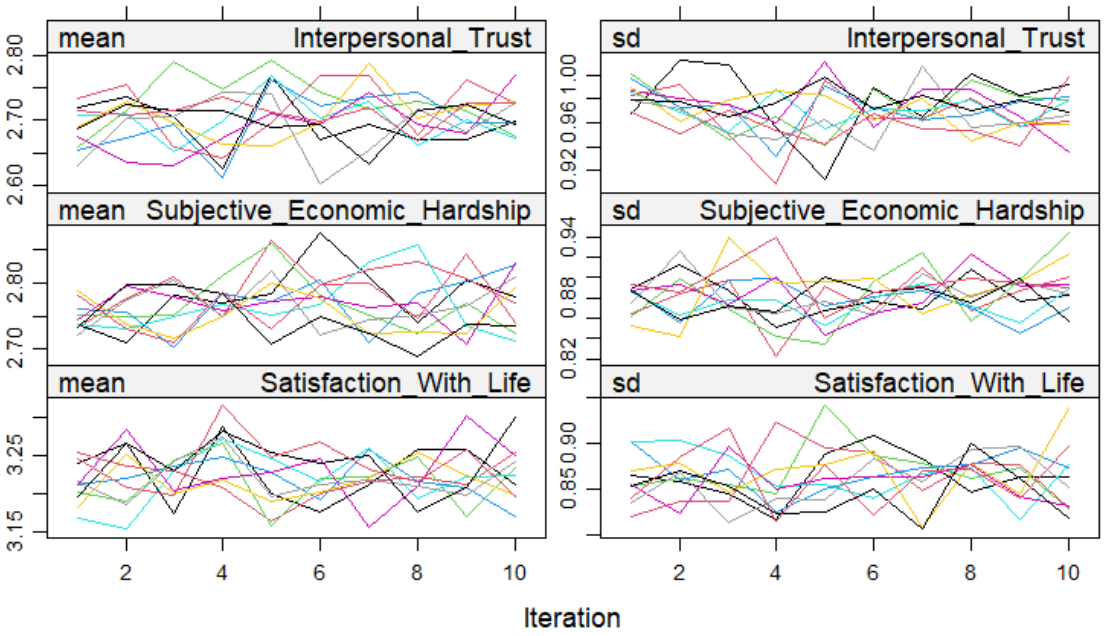


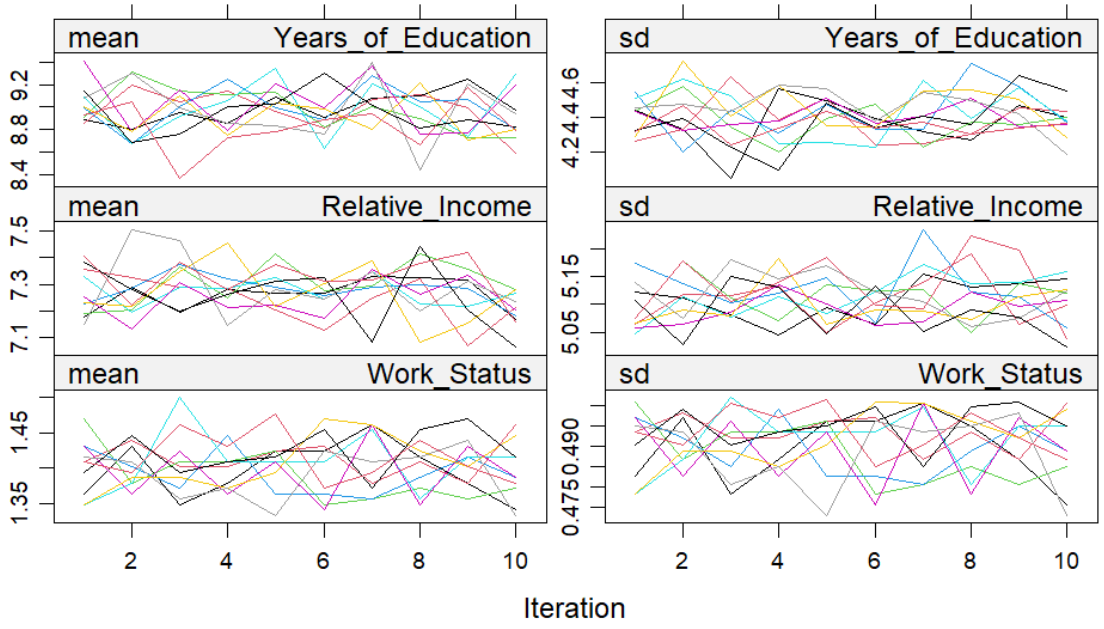
7.3 Full analysis results

Convergence Plots of Multiple Imputation Process: Sample 1

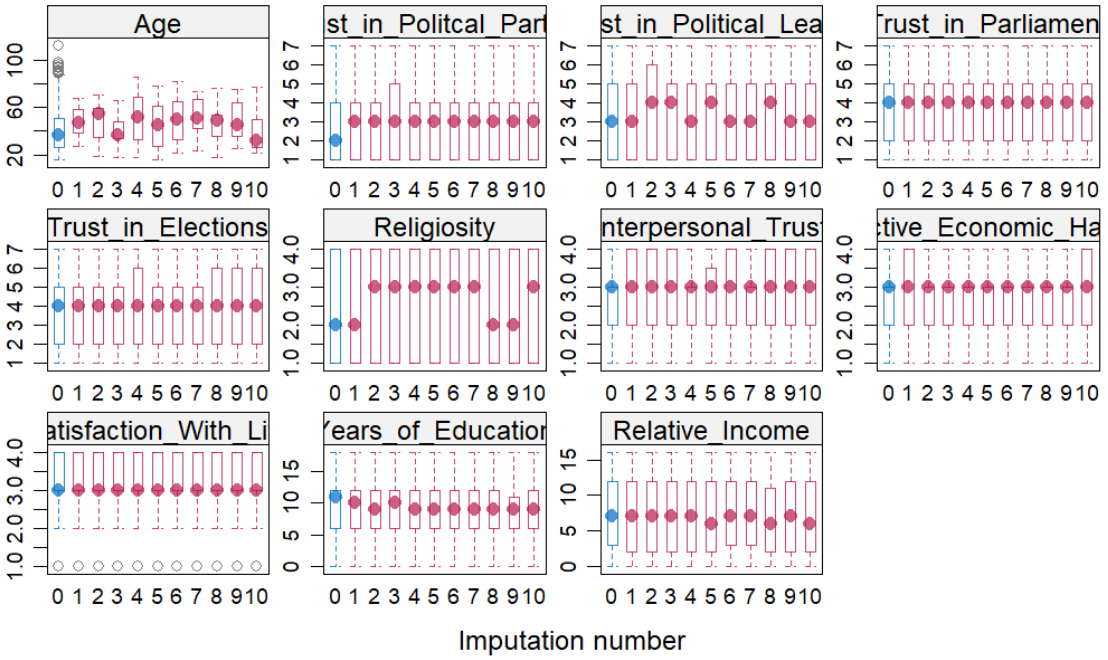


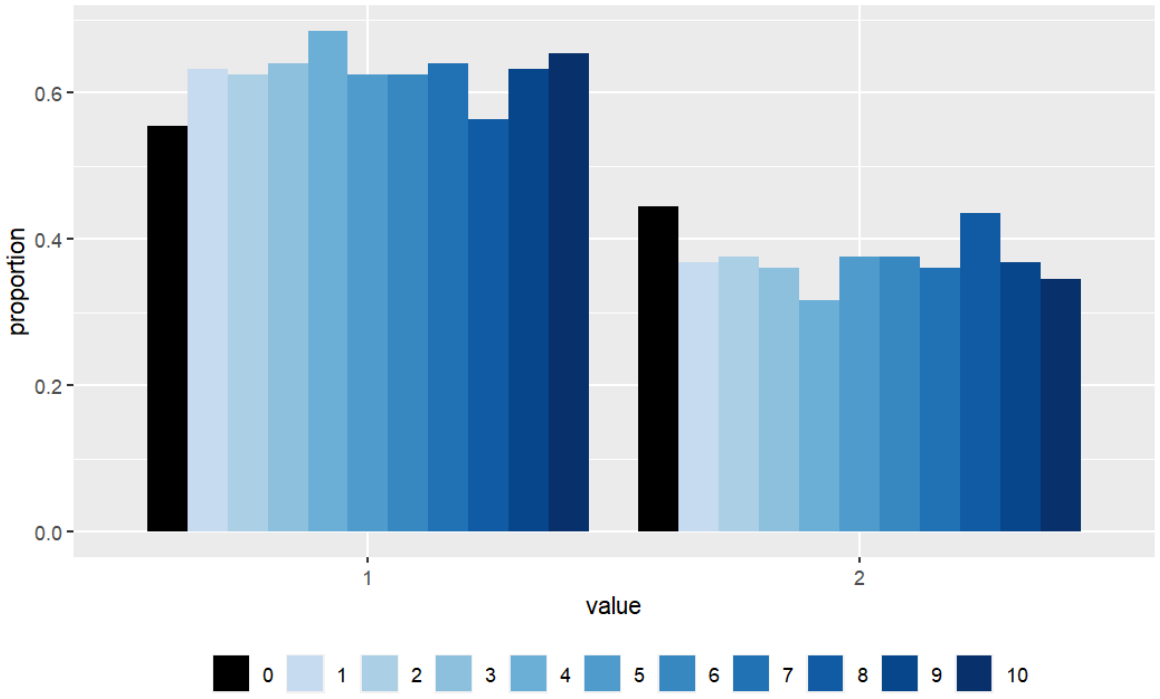



mean Interpersonal_Trust
sd Interpersonal_Trust
mean Subjective_Economic_Hardship
sd Subjective_Economic_Hardship
mean Satisfaction_With_Life
sd Satisfaction_With_Life
Iteration

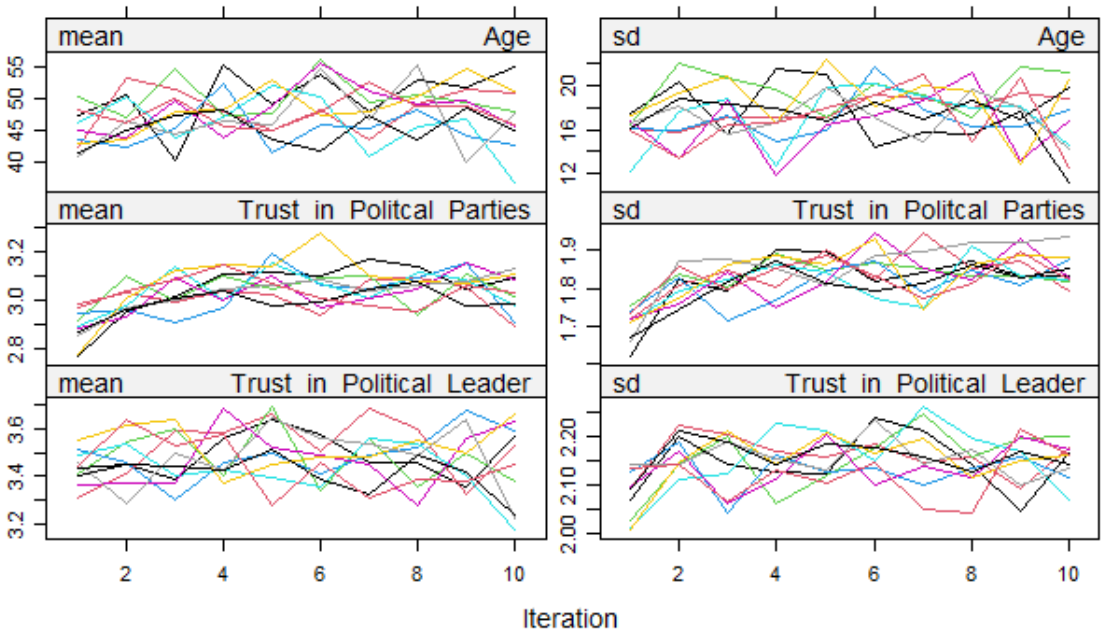

mean Years_of_Education
sd Years_of_Education
mean Relative_Income
sd Relative_Income
mean Work_Status
sd Work_Status
Iteration

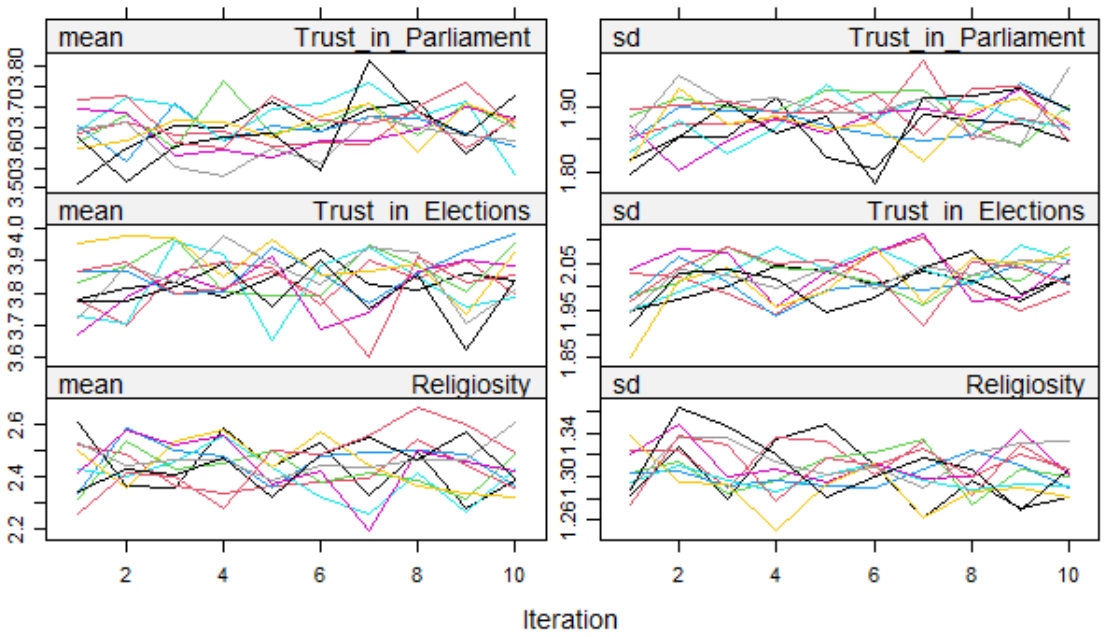
Result of the Multiple Imputation Process: Sample 1

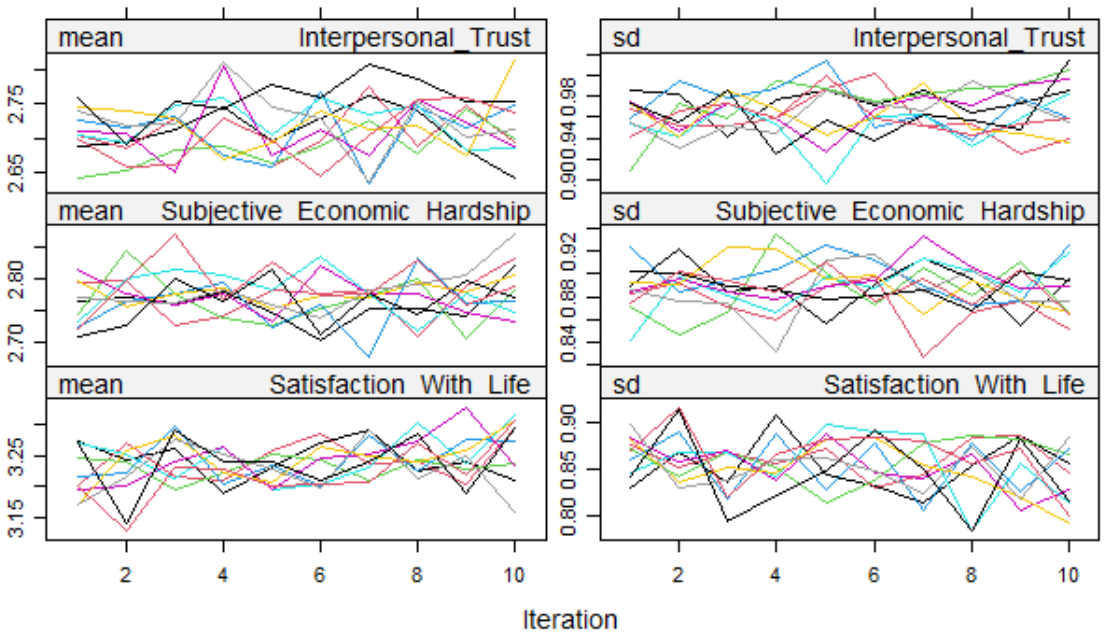


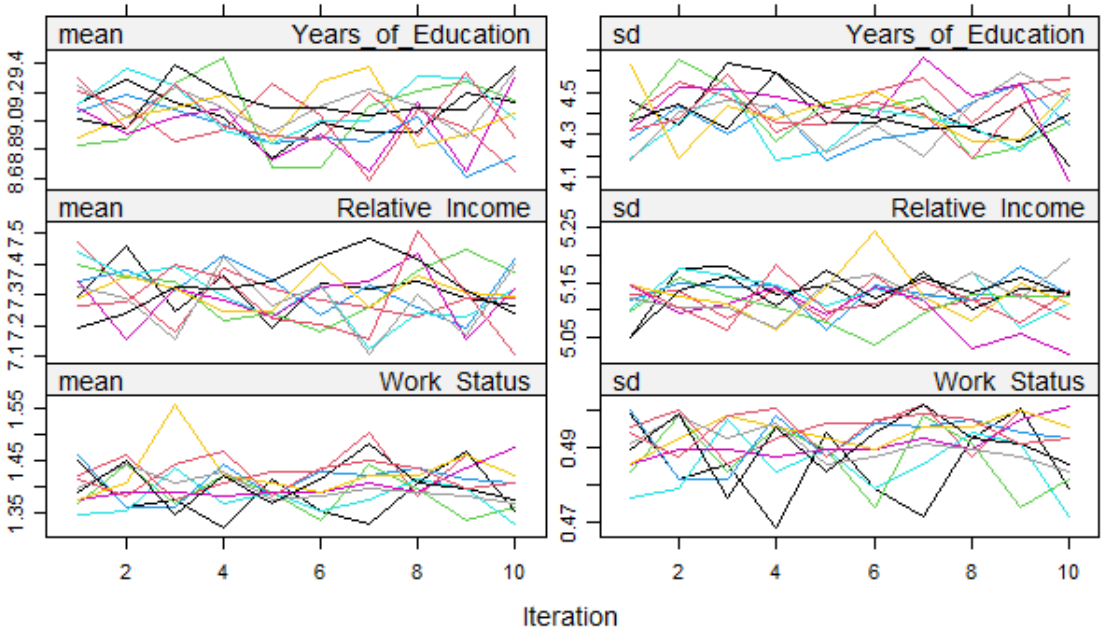


Convergence Plots of Multiple Imputation Process: Sample 2

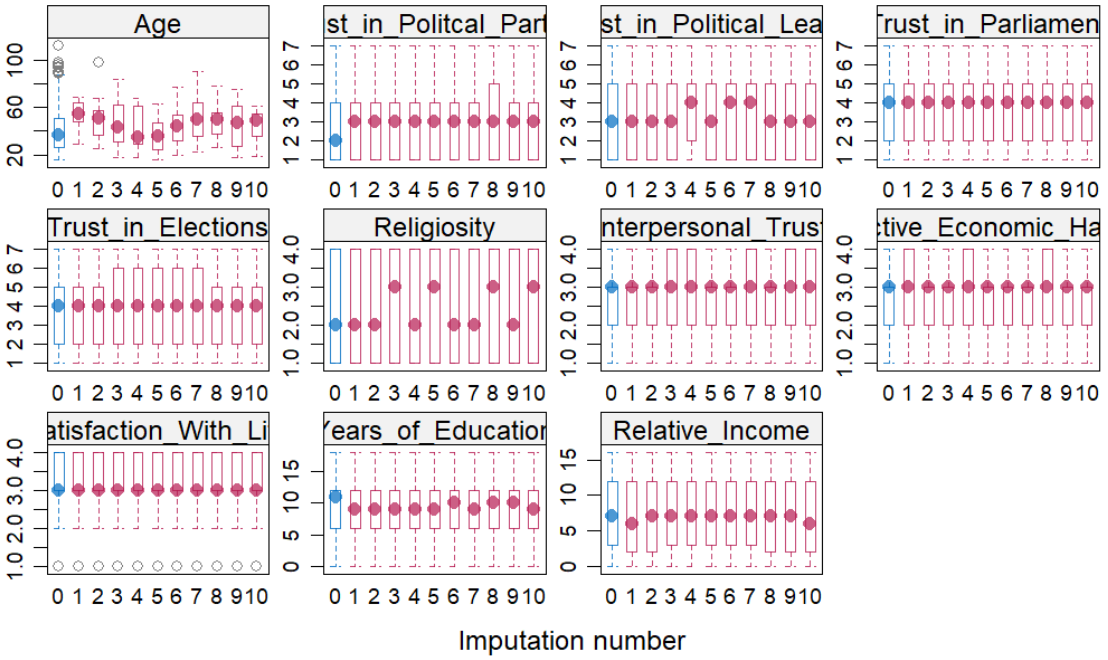


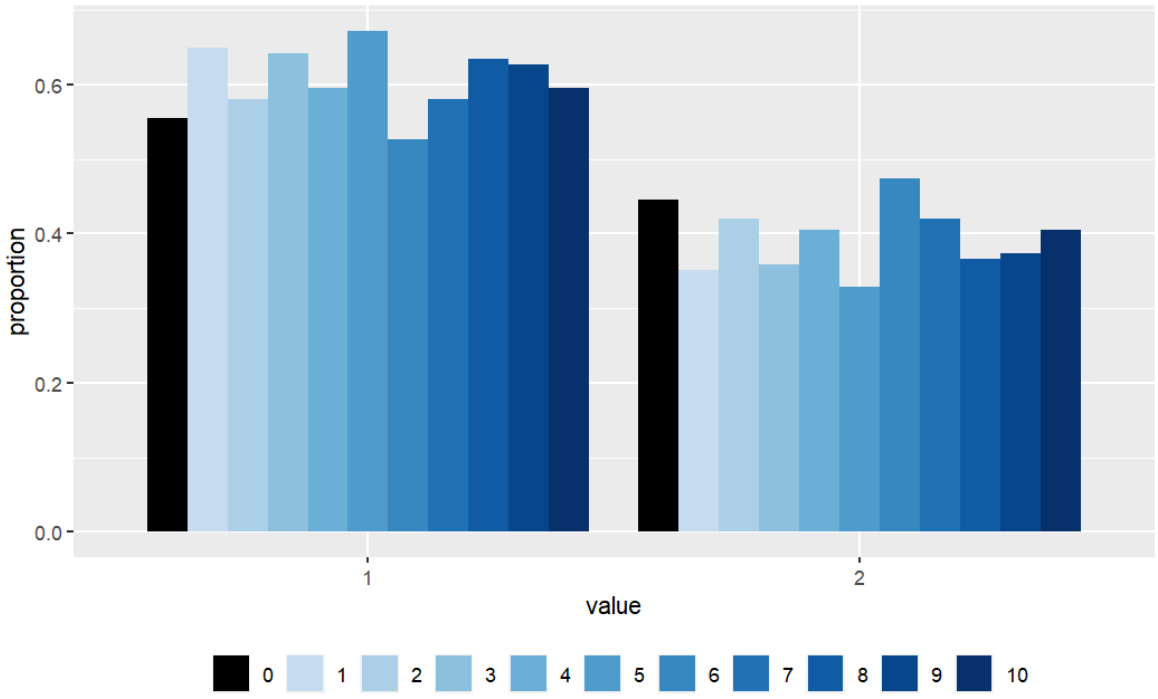



mean
Interpersonal_Trust
sd
Interpersonal_Trust
mean
Subjective Economic Hardship
sd
Subjective Economic Hardship
mean
Satisfaction With Life
sd
Satisfaction With Life
Iteration

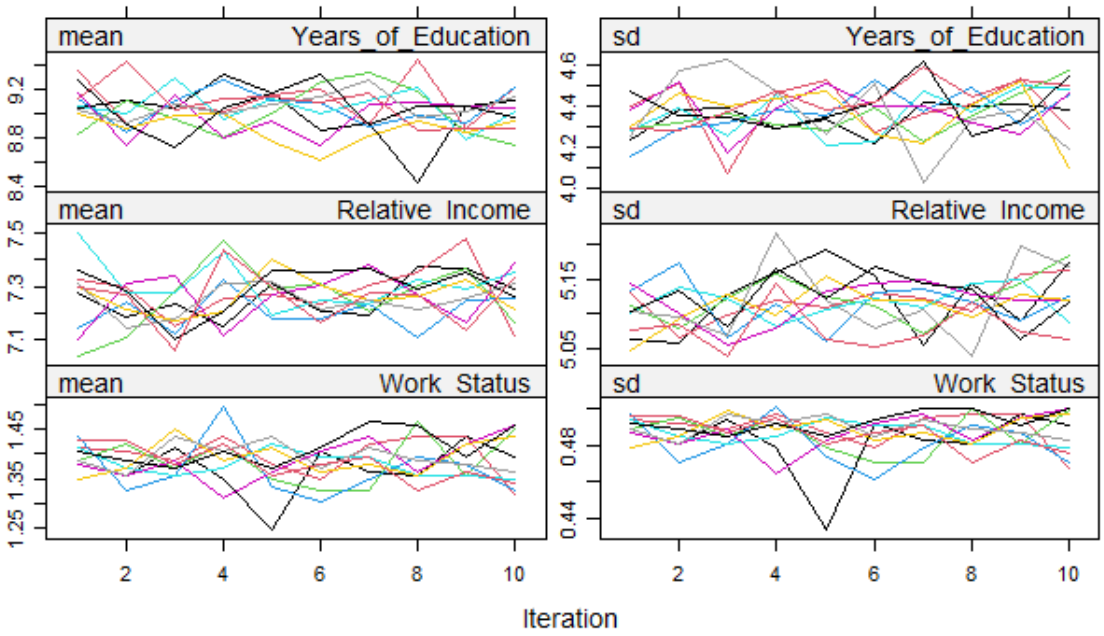

mean
Years_of_Education
sd
Years_of_Education
mean
Relative Income
sd
Relative Income
mean
Work Status
sd
Work Status
Iteration

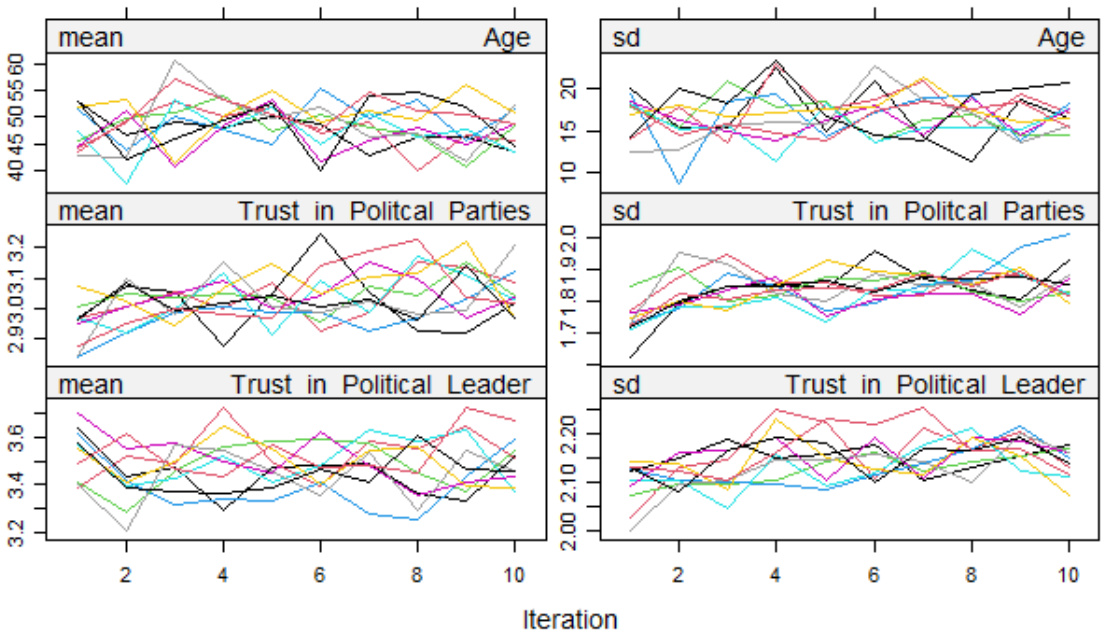
Result of the Multiple Imputation Process: Sample 2

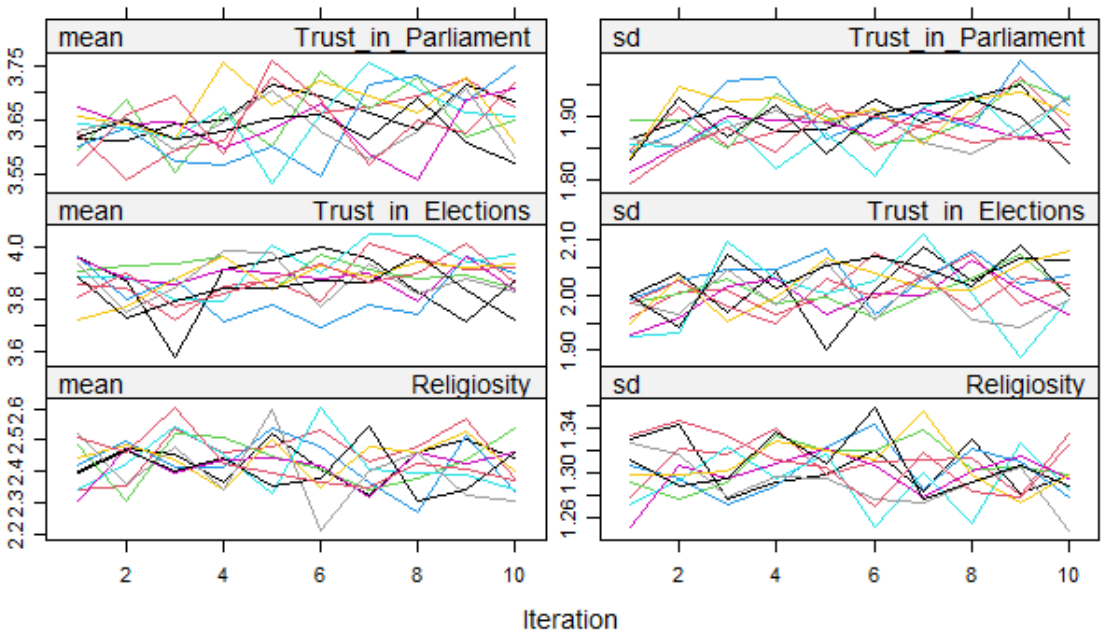


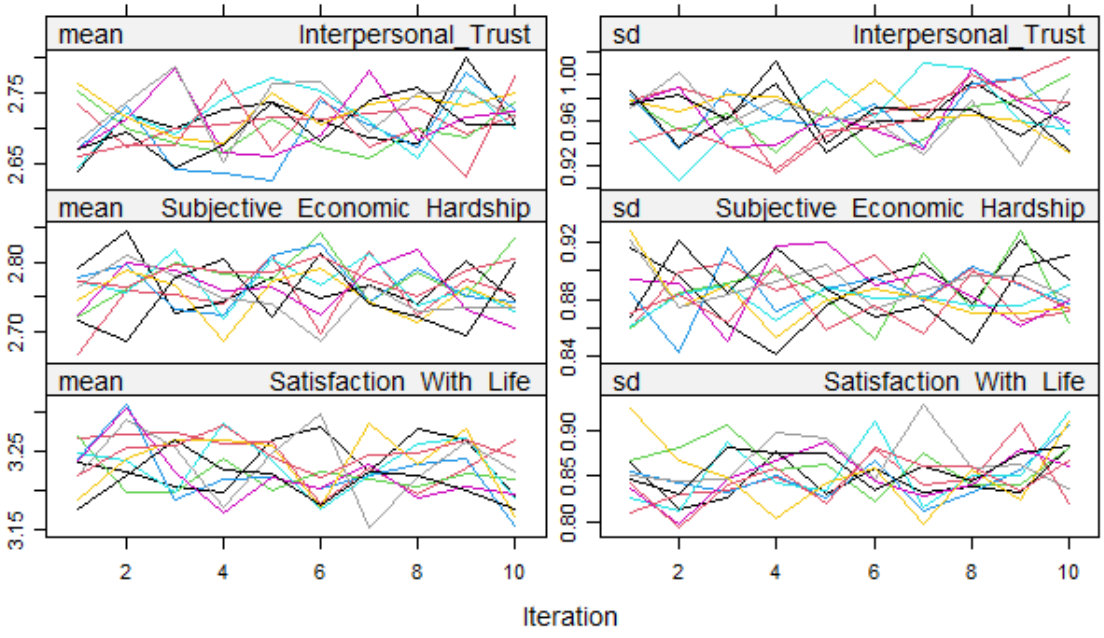


Convergence Plots of Multiple Imputation Process: Sample 3

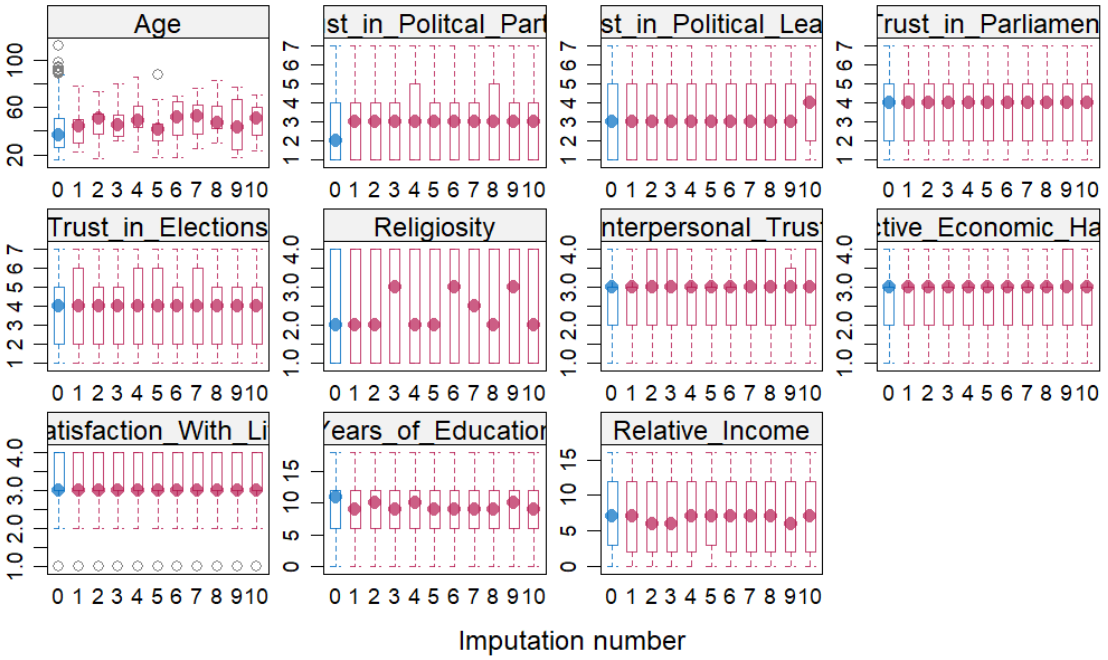


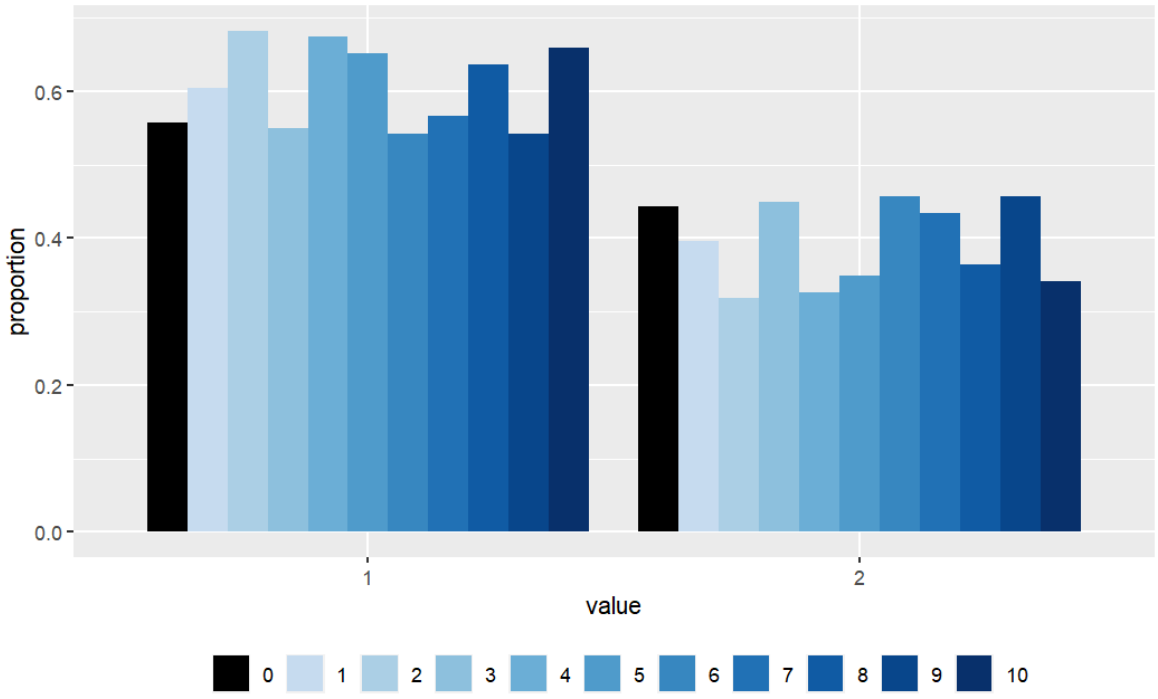



mean
Trust_in_Parliament
sd
Trust_in_Parliament
mean
Trust_in_Elections
sd
Trust_in_Elections
mean
Religiosity
sd
Religiosity
Iteration

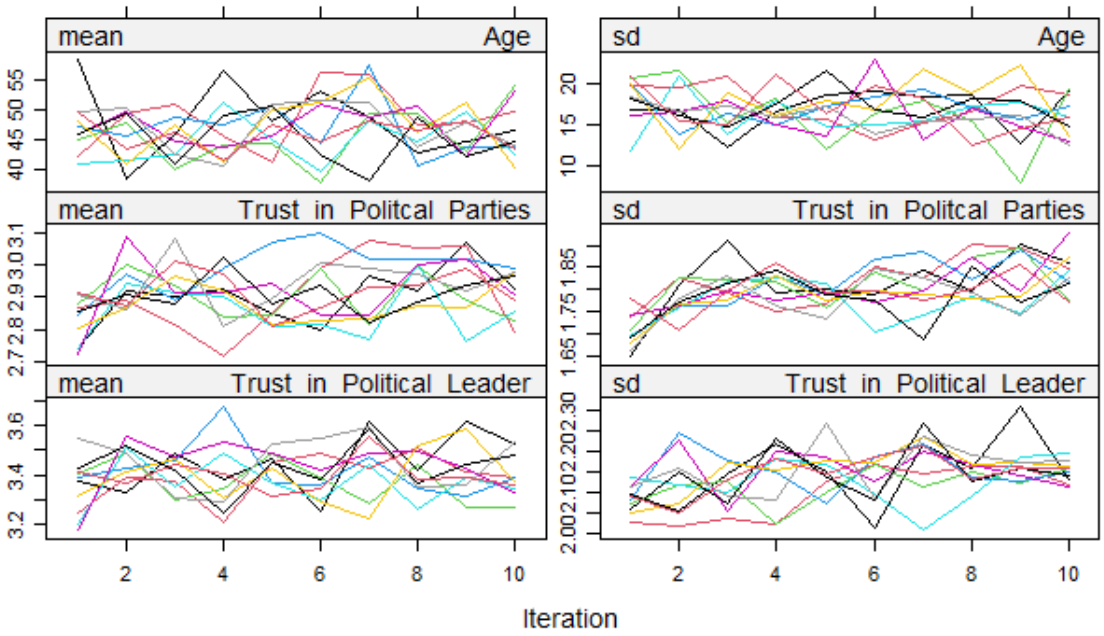

mean
Interpersonal_Trust
sd
Interpersonal_Trust
mean
Subjective_Economic_Hardship
sd
Subjective_Economic_Hardship
mean
Satisfaction_With_Life
sd
Satisfaction_With_Life
Iteration

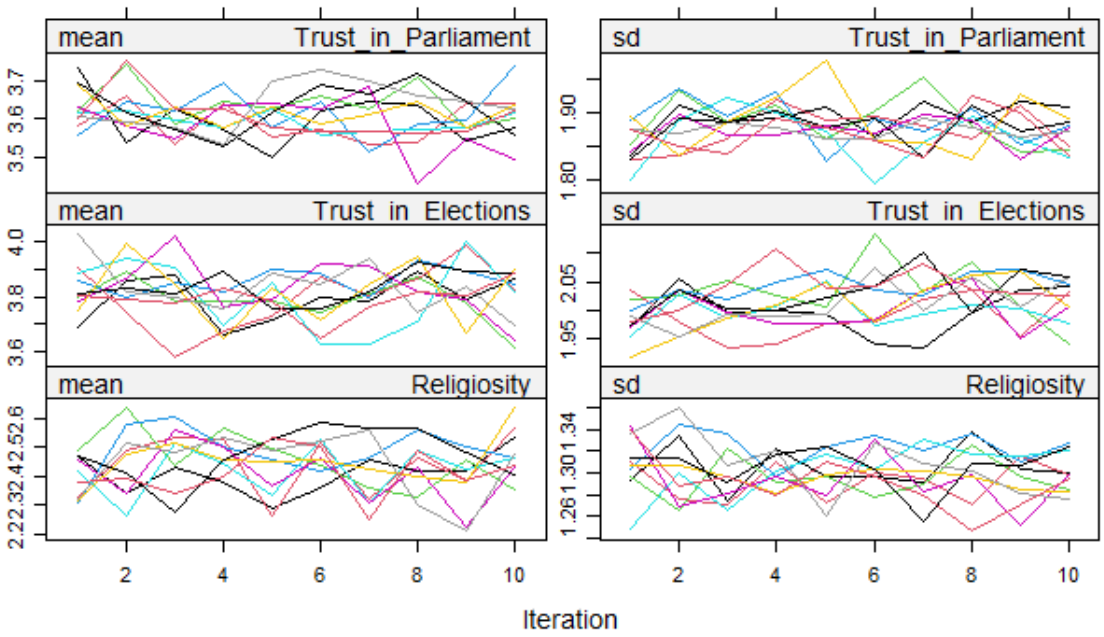
Result of the Multiple Imputation Process: Sample 3



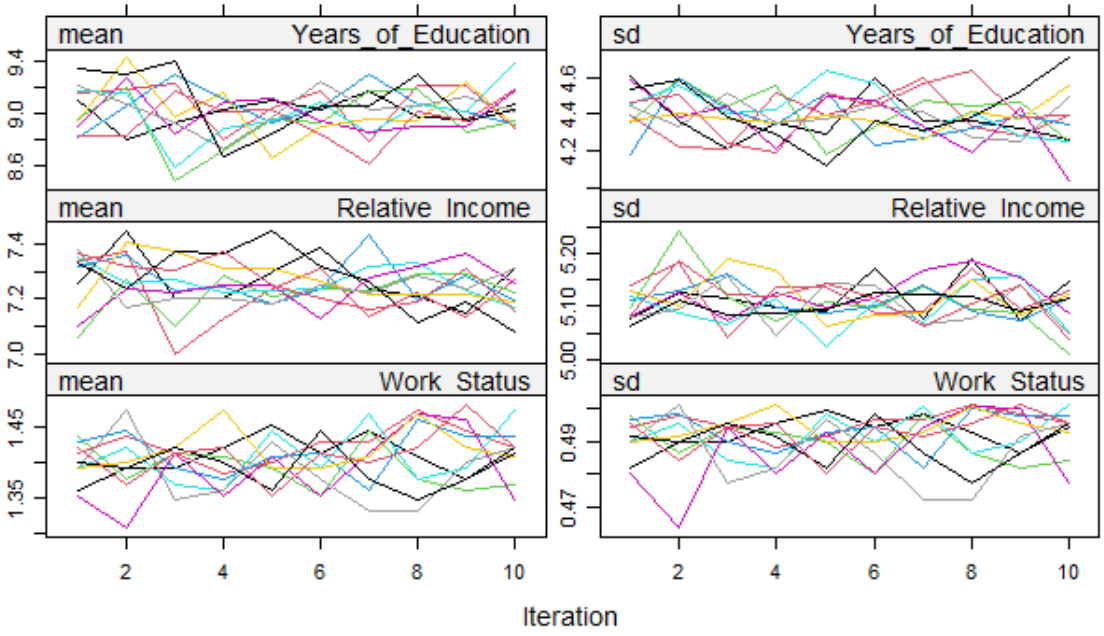


Convergence Plots of Multiple Imputation Process: Sample 4

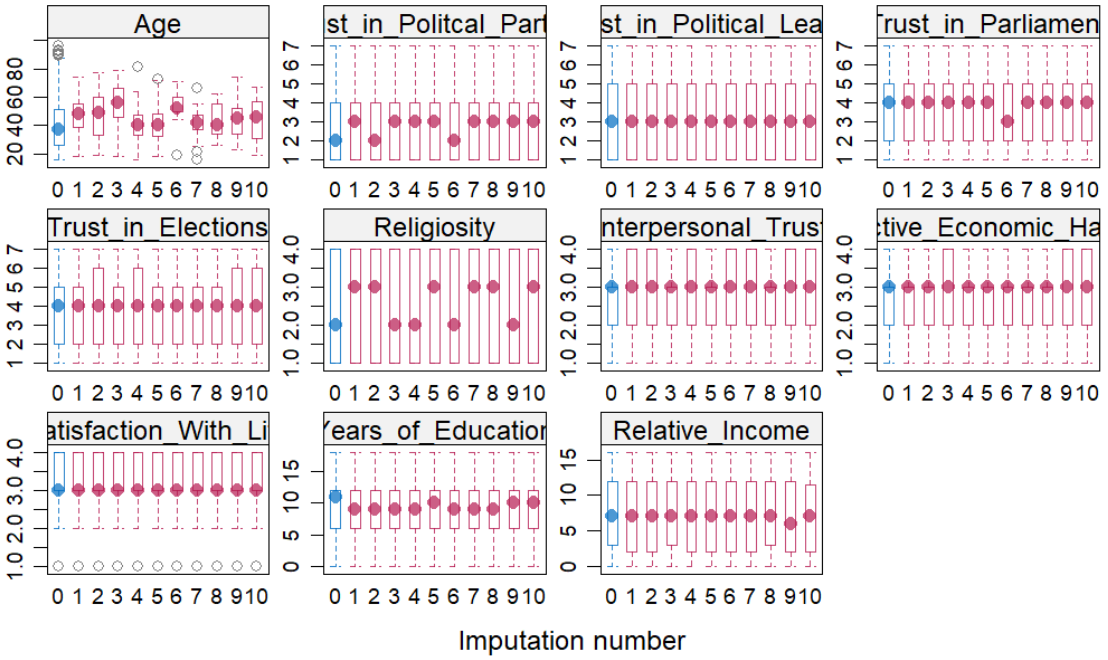


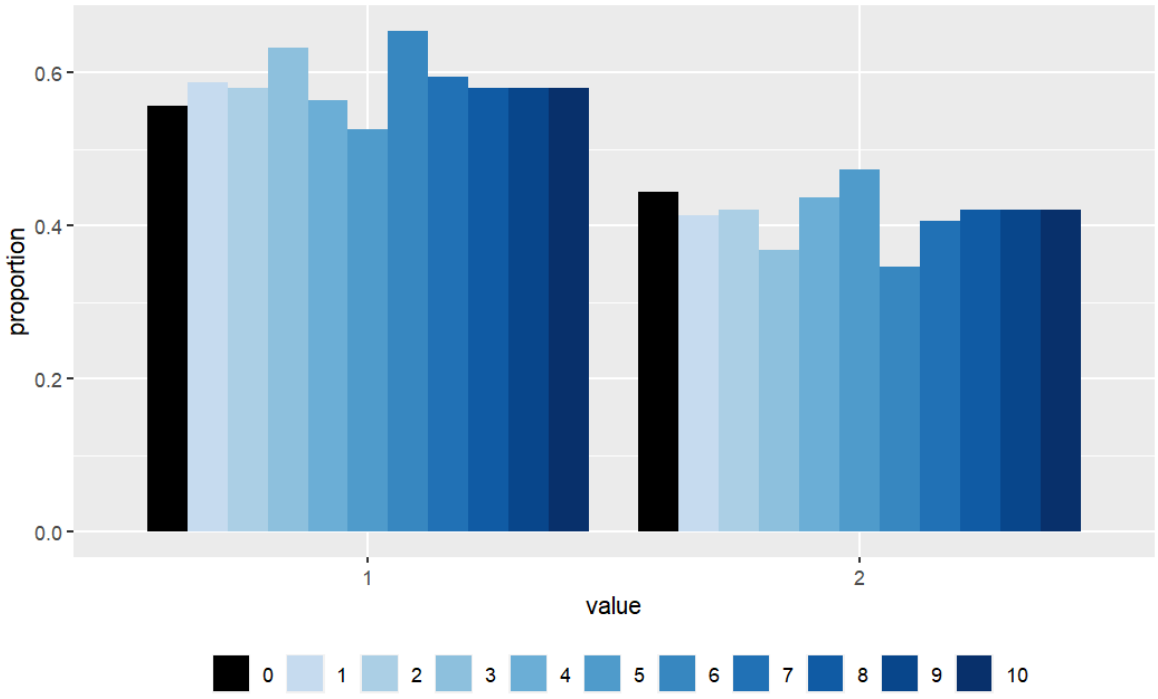



mean
Interpersonal_Trust
sd
Interpersonal_Trust
mean
Subjective_Economic_Hardship
sd
Subjective_Economic_Hardship
mean
Satisfaction_With_Life
sd
Satisfaction_With_Life
Iteration

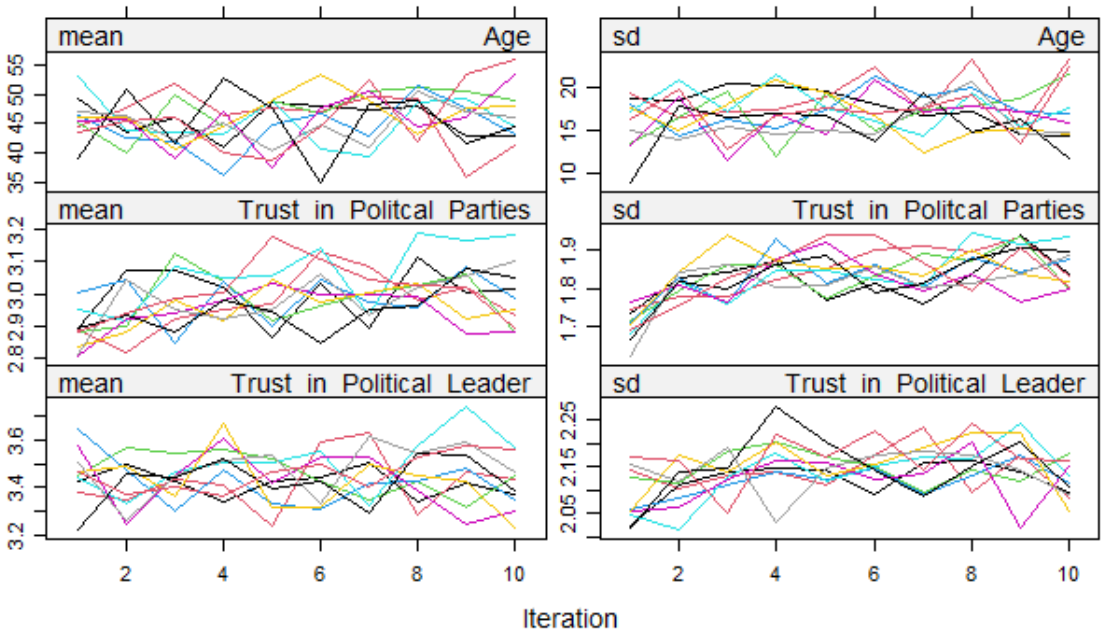

mean
Years_of_Education
sd
Years_of_Education
mean
Relative_Income
sd
Relative_Income
mean
Work_Status
sd
Work_Status
Iteration

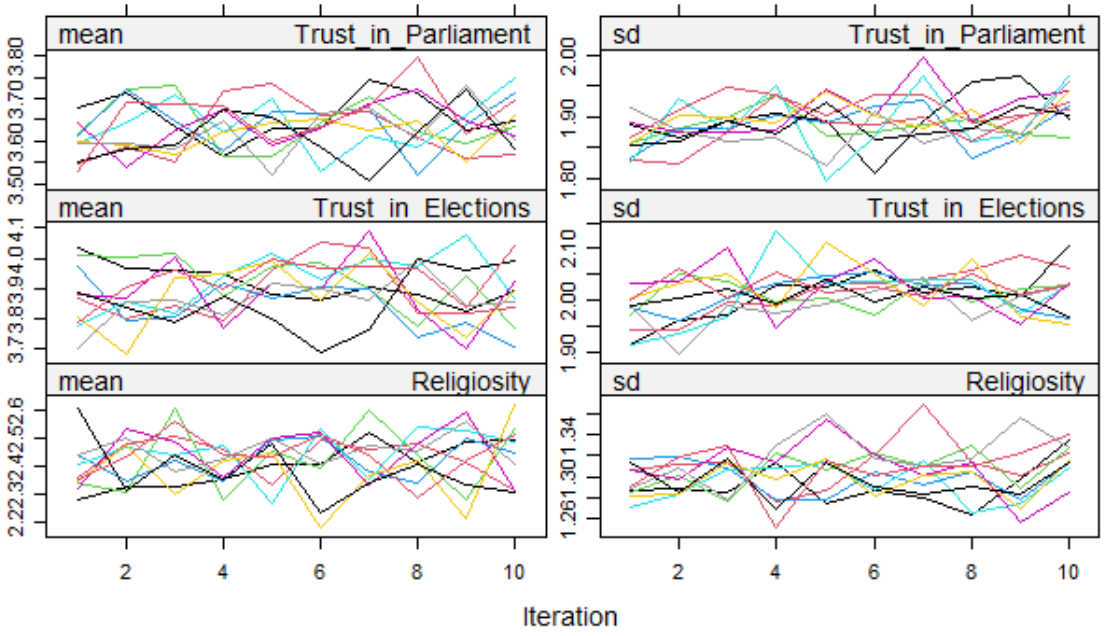
Result of the Multiple Imputation Process: Sample 4

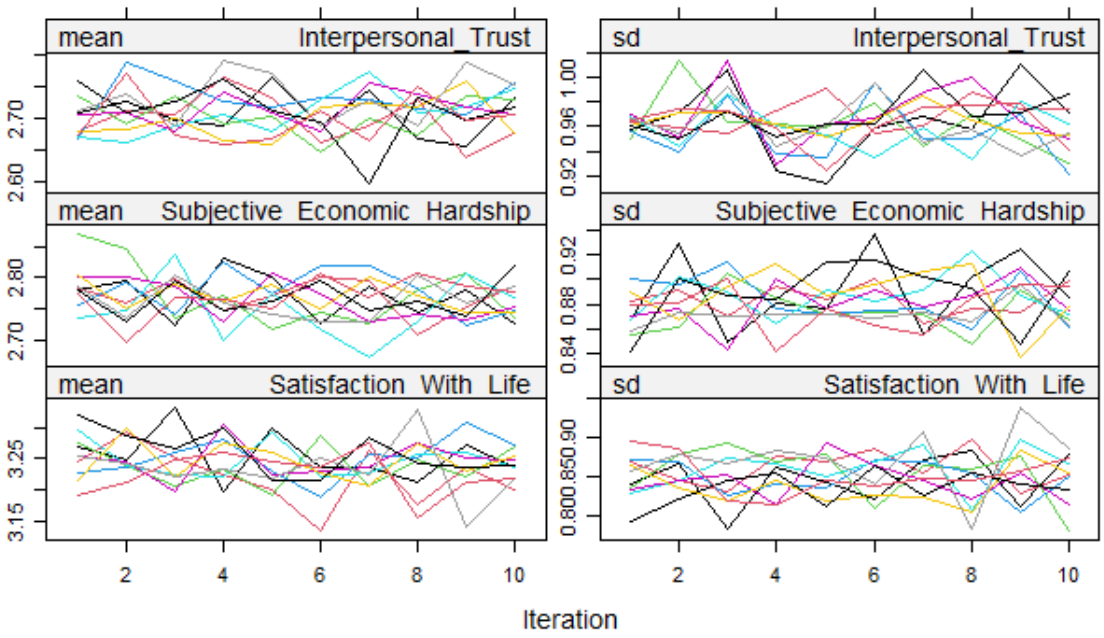


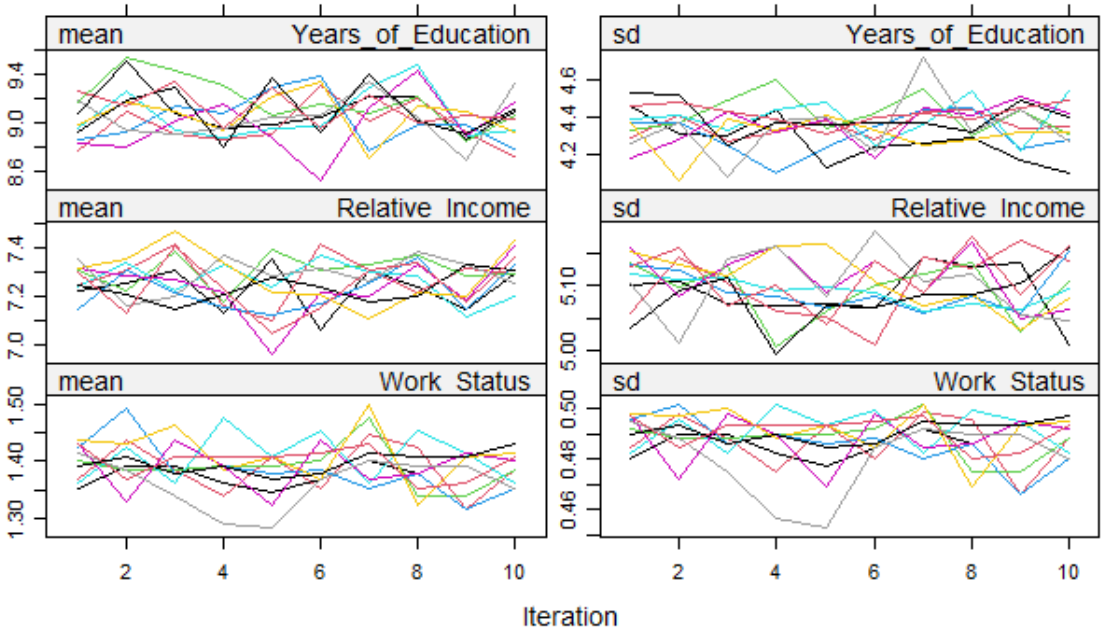


Convergence Plots of Multiple Imputation Process: Sample 5

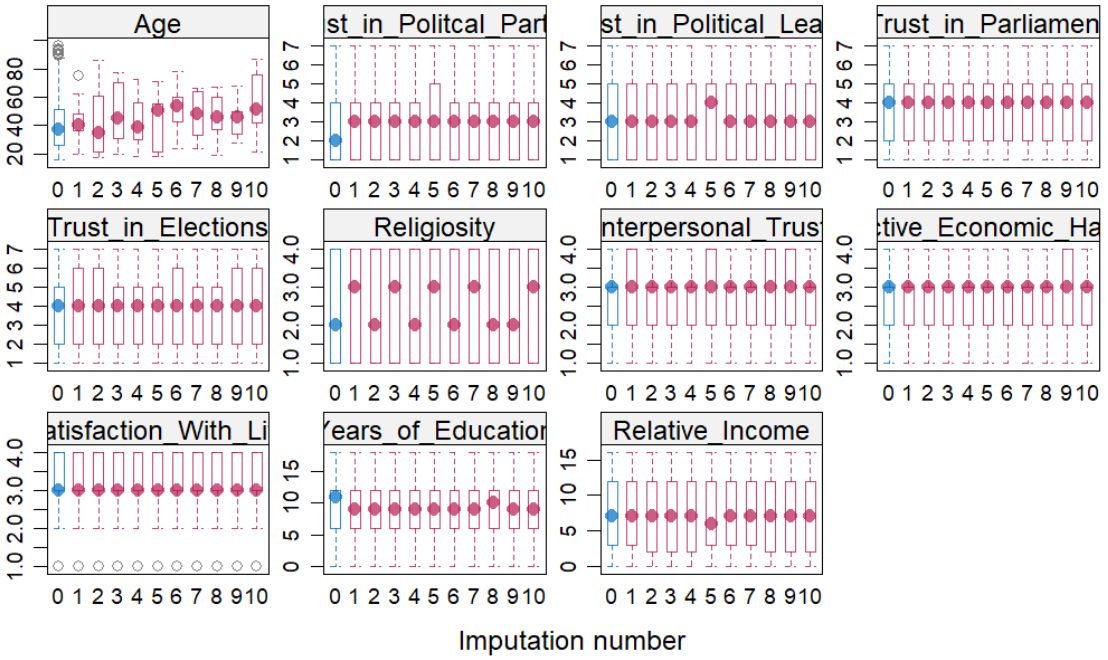





mean
Interpersonal_Trust
sd
Interpersonal_Trust
mean
Subjective_Economic_Hardship
sd
Subjective_Economic_Hardship
mean
Satisfaction_With_Life
sd
Satisfaction_With_Life
Iteration


mean
Years_of_Education
sd
Years_of_Education
mean
Relative_Income
sd
Relative_Income
mean
Work_Status
sd
Work_Status
Iteration

Result of the Multiple Imputation Process: Sample 5



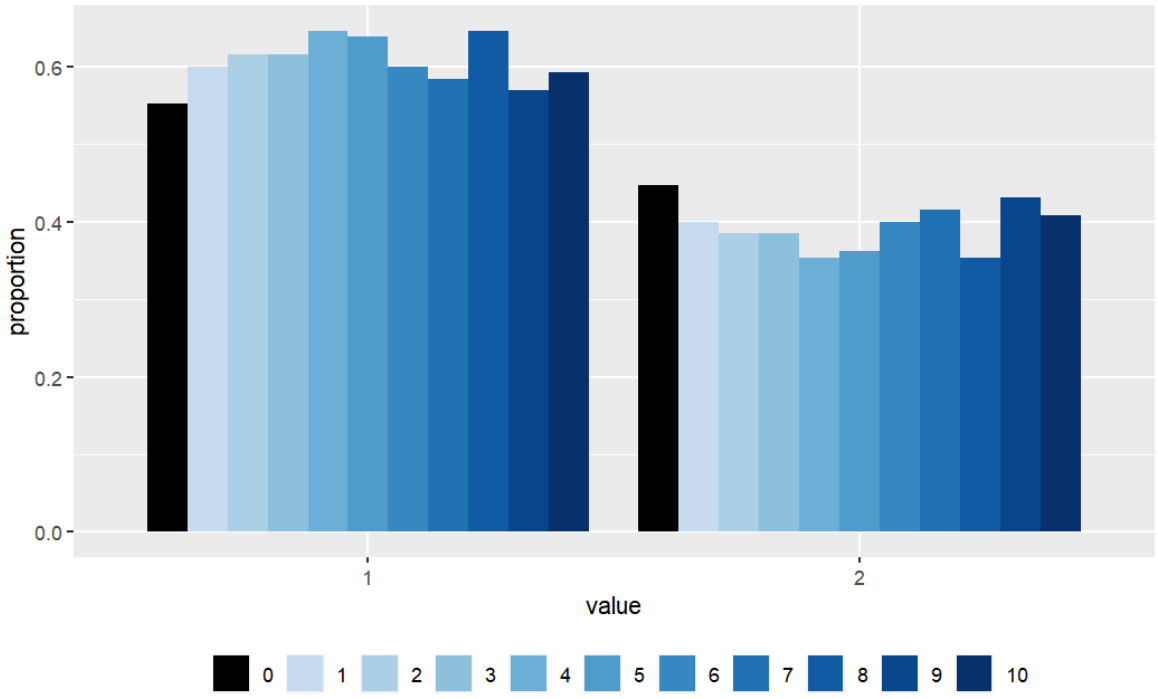


Figure 12: CPDAG estimated for Sample 1

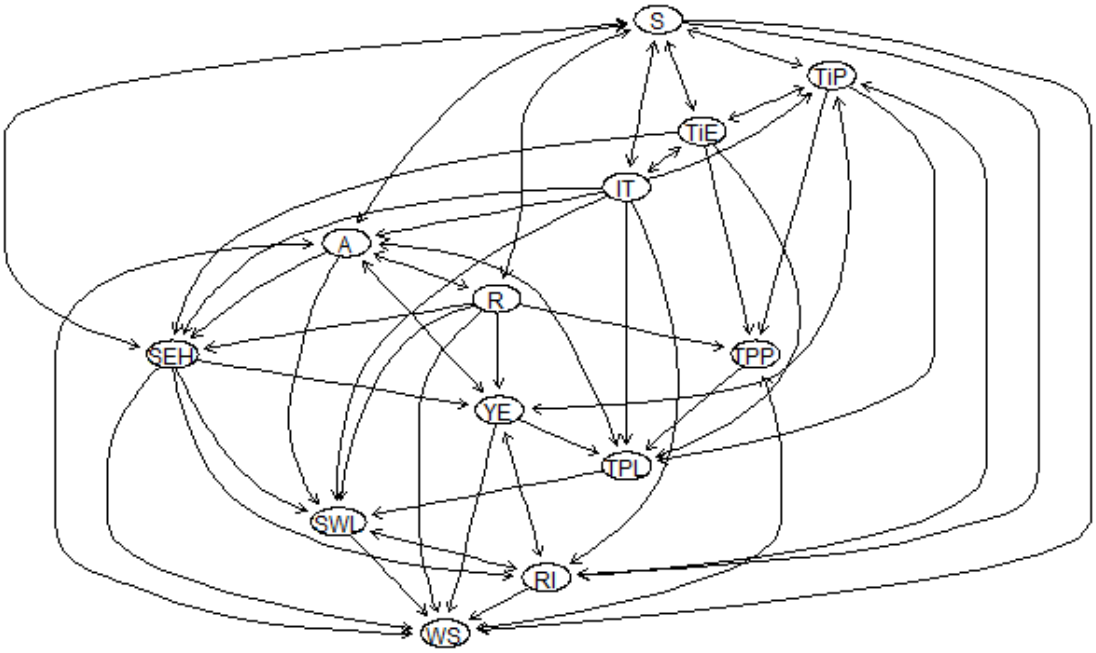


Figure 13: CPDAG estimated for Sample 2

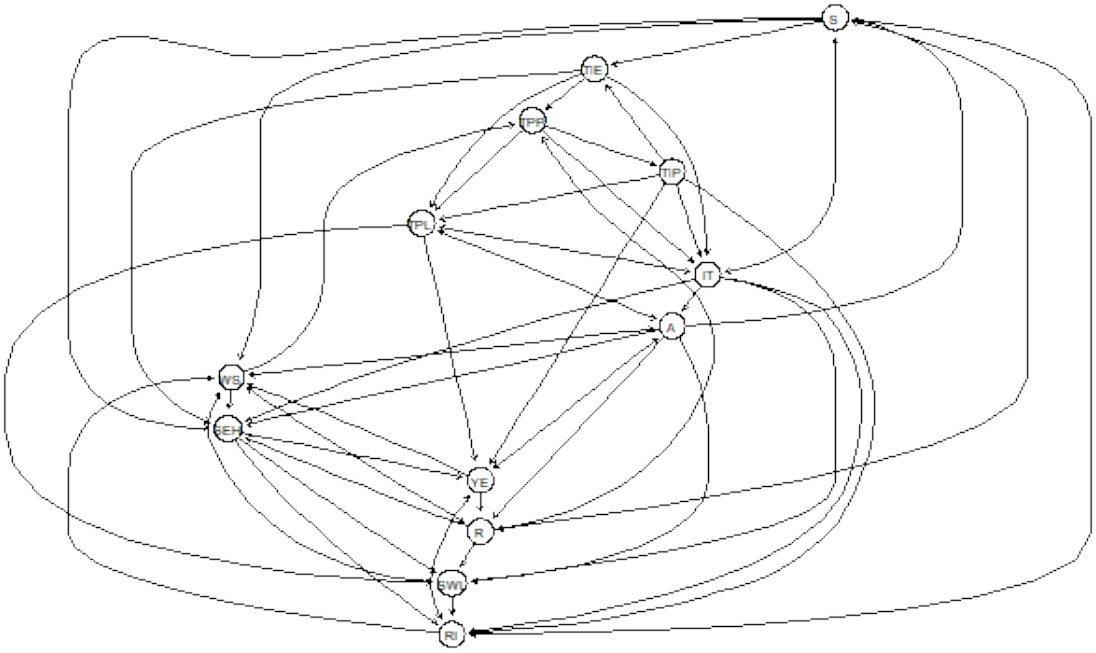


Figure 14: CPDAG estimated for Sample 3

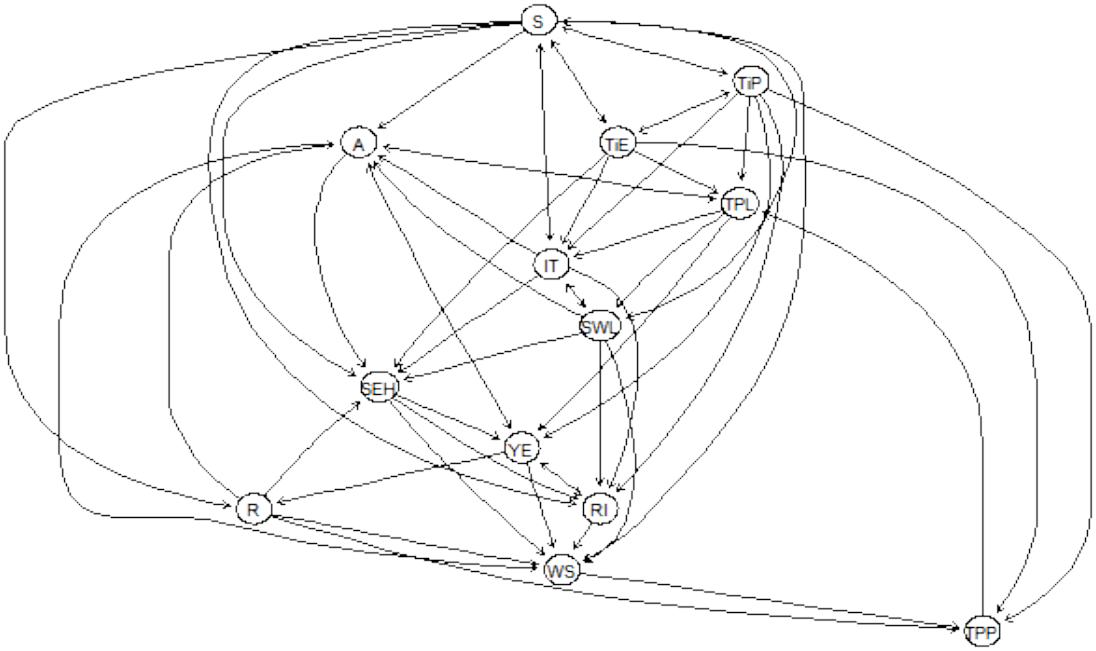


Figure 15: CPDAG estimated for Sample 4

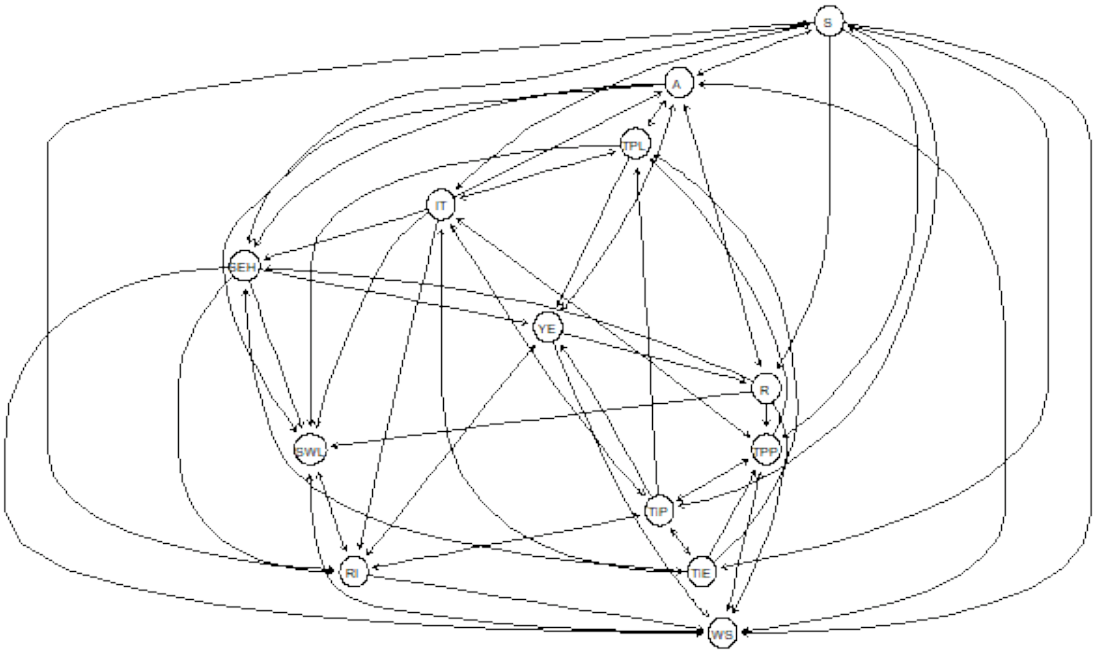


Figure 16: CPDAG estimated for Sample 5

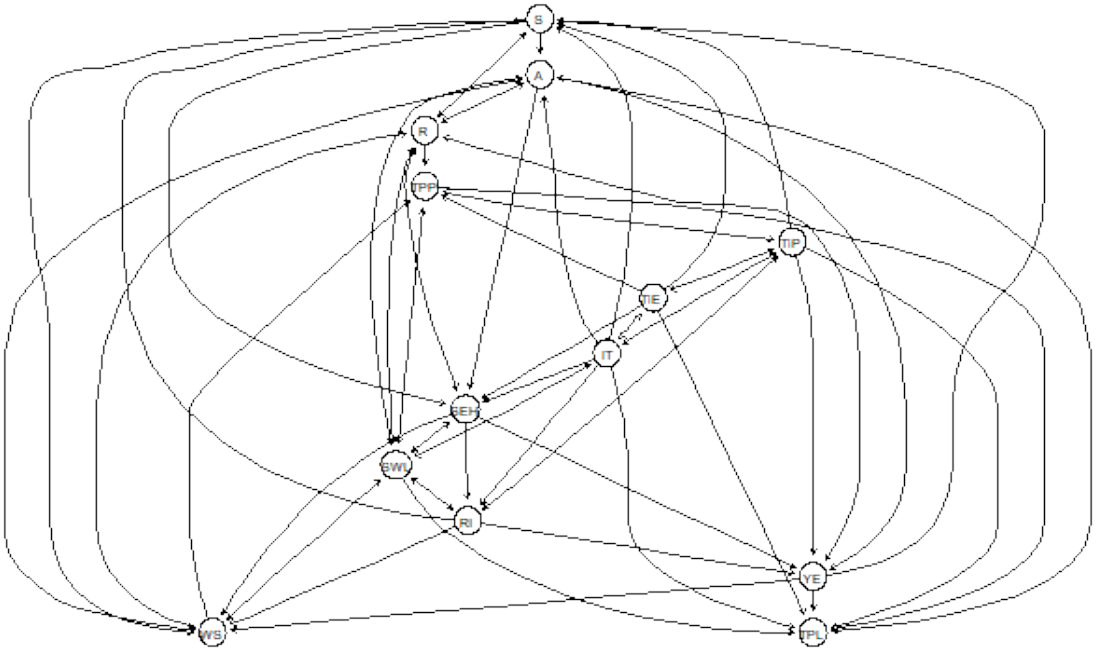
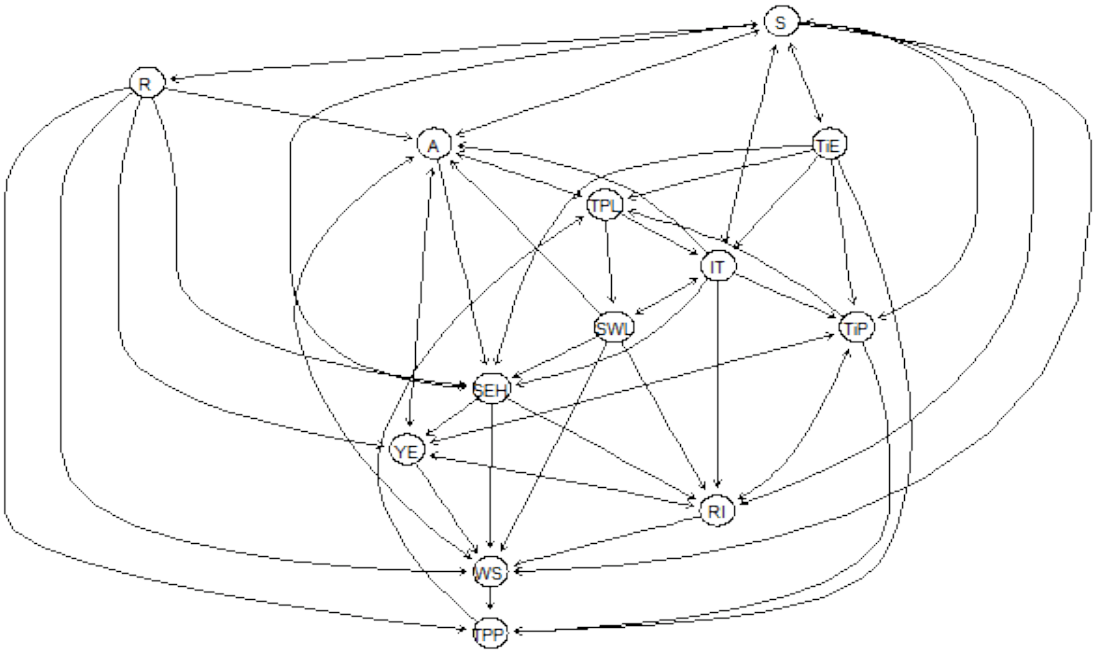


Figure 17: CPDAG estimated for Sample 6 - Testwise Deletion



Legend of CPDAG

Variable	Abbreviation
Satisfaction With Life	SWL
Economic Hardship	SEH
Religiosity	R
Interpersonal Trust	IT
Trust in Political Parties	TPP
Trust in Political Leader	TPL
Trust in Parliament	TiP
Work Status	WS
Years of Education	YE
Gender	S
Trust in Elections	TiE
Relative Income	RI
Age	A

Do variables share a direct causal link with SWL?

Variable	0	1	2	3	4	5
Economic Hardship						X
Religiosity					X	
Interpersonal Trust						X
Trust in Political Parties		X				
Trust in Political Leader						X
Trust in Parliament	X					
Work Status					X	
Years of Education	X					
Gender		X				
Trust in Elections	X					
Relative Income						X
Age						X

Causal Effects on Satisfaction with Life

Variable	Lower Bound	Higher Bound	Number of NA	Test Deletion
Economic Hardship	-0.167	-0.157	1	NA
Religiosity	0.014	0.022	0	0.007
Interpersonal Trust	0.086	0.103	0	0.092
Trust in Political Parties	0.014	0.021	2	0.018
Trust in Political Leader	0.015	0.020	2	NA
Trust in Parliament	0.008	0.020	3	NA
Trust in Elections	0.019	0.030	0	0.029

Causal Effects on Subjective Economic Hardship

Variable	Lower Bound	Higher Bound	Number of NA	Test Deletion
Religiosity	0.028	0.040	1	0.044
Interpersonal Trust	-0.118	-0.090	0	-0.104
Trust in Political Parties	-0.001	0.001	2	-0.001
Trust in Political Leader	-0.001	0.002	2	NA
Trust in Parliament	0.001	0.003	3	NA
Trust in Elections	-0.048	-0.0296	0	-0.036

7.4 Tables and Figures

List of Figures

List of Tables

8 List of references

Andrews, R. M., Foraita, R., Didelez, V., and Witte, J. (2021). A practical guide to causal discovery with cohort data. arXiv preprint arXiv:2108.13395.

Bentham, J. [1780] 1970. An introduction to the principles of morals and legislation. In Collected works of Jeremy Bentham, edited by J. H. Burns and H. L. A. Hart. London: Athlone

Bértola, L., and Ocampo, J. A. (2012). The economic development of Latin America since independence. OUP Oxford.

Collard, D. (2006). Research on well-being: Some advice from Jeremy Bentham. Philosophy of the Social Sciences, 36(3), 330-354.

Colombo, D., and Maathuis, M. H. (2014). Order-independent constraint-based causal structure learning. J. Mach. Learn. Res., 15(1), 3741-3782.

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In Nations and households in economic growth (pp. 89-125). Elsevier.

Frey, B. S., and Stutzer, A. (2002). What can economists learn from happiness research? Journal of Economic literature, 40(2), 402-435.

Frey, B. S., and Stutzer, A. (2005). Happiness research: State and prospects. Review of social economy, 63(2), 207-228.

Frey, B. S., Stutzer, A., Benz, M., Meier, S., Luechinger, S., and Benesch, C. (2008). Happiness : a revolution in economics. MIT Press.

Glymour, C., Zhang, K., and Spirtes, P. (2019). Review of causal discovery methods based on graphical models. Frontiers in genetics, 10, 524.

Kalisch, M., Hauser, A., Maathuis, M., and Mächler, M. (2020). An Overview of the pcalg Package for R.

Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., and Bühlmann, P. (2012). Causal inference using graphical models with the R package pcalg. Journal of statistical software, 47, 1-26.

Li, C. (2013). Little's test of missing completely at random. The Stata Journal, 13(4), 795-809.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. Journal of the American statistical Association, 83(404), 1198-1202.

Maathuis, M. H., Kalisch, M., and Bühlmann, P. (2009). Estimating high-dimensional intervention effects from observational data. The Annals of Statistics, 37(6A), 3133-3164.

Pearl, J. (2009). Causality : Models, Reasoning and Inference (2nd ed.). Cambridge University Press.

Powdthavee, N. (2007). Causal analysis in happiness research. Southeast Asian Journal of Economics, 215-223.

Powdthavee, N. (2010). How much does money really matter? Estimating the causal effects of income on happiness. Empirical economics, 39(1), 77-92.

Ramsey, J., Zhang, J., and Spirtes, P. L. (2012). Adjacency-faithfulness and conservative

causal inference. arXiv preprint arXiv:1206.6843.

Reeskens, T., and Vandecasteele, L. (2017). Economic hardship and well-being: Examining the relative role of individual resources and welfare state effort in resilience against economic hardship. Journal of Happiness Studies, 18(1), 41-62.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. Advances in methods and practices in psychological science, 1(1), 27-42.

Rojas, M. (2019a). The economics of happiness : how the Easterlin Paradox transformed our understanding of well-being and progress. Springer.

Rojas, M. (2019b). Pioneer in Happiness Research in Latin America. Applied Research in Quality of Life, 14(5), 1435-1437.

Rose, D. (2017). A Modern History of Happiness as Economic Policy. University of Guelph, Canada, Researchgate. net www.researchgate. net/publication/316091499 A Modern History of Happiness as Economic Policy.

Spirtes, P., Glymour, C. N., Scheines, R., and Heckerman, D. (2000). Causation, prediction, and search. MIT press.

Spirtes, P., and Zhang, K. (2016). Causal discovery and inference: concepts and recent methodological advances. Applied informatics.

Van Buuren, S. (2018). Flexible imputation of missing data. CRC press.

Van Buuren, S., and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. Journal of statistical software, 45, 1-67.

Von Hippel, P. T. (2020). How many imputations do you need? A two-stage calculation using a quadratic rule. Sociological Methods and Research, 49(3), 699-718.

Witte, J., Foraita, R., and Didelez, V. (2021). Multiple imputation and test-wise deletion for causal discovery with incomplete cohort data. arXiv preprint arXiv:2108.13331.

Wulff, J. N., and Jeppesen, L. E. (2017). Multiple imputation by chained equations in praxis: guidelines and review. Electronic Journal of Business Research Methods, 15(1), 41-56.