Statistical Learning project: Why are some countries free?

LUCA DONGHI
luca.donghi1@studenti.unimi.it, 982862

22/07/2022

Abstract

This project requires answering a research question through the use of supervised and unsupervised learning techniques. My research question is "what are the exogenous factors creating the conditions for individuals to enjoy more political, civil, and economic freedoms in some countries rather than in others?". The unsupervised learning techniques I will use are Principal Component Analysis and Hierarchical Clustering respectively to create an overall freedom index and to understand which exogenous factors are shared by countries with similar levels of freedom. The supervised learning techniques are instead Decision Tree and Random Forest: the first one will be mainly used for interpretation purposes, while the second to improve the decision tree predictive performances.

Contents

1. Introduction

Individual freedom in the world is extremely heterogeneous. Traveling from one continent to another, but also simply moving from one country to another, it is evident how the restrictions on freedom of choice change radically.
Freedom is such an abstract and controversial concept that it is difficult to think that it could be a research field for statisticians rather than just philosophers, politicians, and other humanistic thinkers of various kinds. However, this is not true. It is possible to find a specific and widely shared definition of freedom that can be measured and, whenever something can be measured, then the statistical tools can be powerful means of learning.
In this research, I will use the classical concept of freedom used in political science in the three main fields: political freedom, civil liberties, and economic freedom. Freedom is therefore treated as real-world rights and freedoms enjoyed by individuals. It is not enough for homosexual couples in a state to be legally recognized or protected to establish the right to free sex life. Exactly as it is not enough for contracts to be legally recognized and protected to establish the right to free trade. Indeed, a lot of non-state actors such as religious organizations, the neighborhood you live in, and traditions can substantially influence your freedom.
The aim of this research is therefore to try to explain why some countries are freer than others using different statistical learning techniques.
To explain freedom only exogenous variables will be considered. It is almost a simple task to try to predict a country's degree of freedom using variables such as GDP per capita, population density, fertility rate, or economic inequalities.
In doing so, however, we would learn very little about the causes of freedom. Indeed, it is not so much wealth that causes economic freedom or political freedom, but rather the opposite is true.

The research, and therefore this report, is structured as follows:
In Section 2 I will describe the datasets selected for the freedom variables and for the explanatory variables discussing why the selected explanatory variables can be assumed as exogenous.
In Section 3 I will implement two unsupervised learning techniques for two different purposes. I will first implement Principal Component Analysis on the freedom data (Subsection 3.1) in order to create an overall freedom index that will subsume as much information as possible from the 37 different freedom variables. This new variable will be the quantitative representation of the concept I want to study and

for this reason, it will be used as the dependent variable in the whole supervised learning section. Principal Components interpretation and visualization will also be useful for the graphical representation of the hierarchical Clustering outcome. Hierarchical Clustering (Subsection 3.2) will be performed on the freedom data to learn something useful about which countries have similar levels of freedom: what else do they have in common? Which exogenous variables could we use?
In Section 4 I will try to explain the differences in the levels of freedom using the exogenous variables. To interpret the underlying relationship, but also of predicting the value of the dependent variable, two techniques have been chosen. Decision Trees (Subsection 4.1) will be used mainly for their interpretability and ability to fit the highly non-linear underlying relationship. Random Forest (Subsection 4.2) will instead be used mainly to overcome the poor predictive accuracy of decision tree models.
Section 5 will conclude the report by summarizing the most important findings of this research.

2. Data

This section describes the datasets on which the statistical techniques will be implemented. I will focus in particular on the sources' choice for the freedom variables and on the exogeneity assumption, I made over the explanatory variables.

2.1. *Freedom datasets*

Many different sources try to quantify democratic, civil and economic freedoms. Different definitions of democracy and freedom gave birth to different scores. However, some institutions are more authoritative than others and are therefore more used in political science. Among these, two institutions are the most authoritative, used and at the same time fit perfectly the definition of freedom I discussed in the introduction 1.
FredomHouse (Edition 2019) is the source used for the political freedom and Civil Liberties variables. their dataset is composed of 10 political freedom variables and 15 civil liberties variables recorded on 209 observations (countries). Each variable takes value from 0 (no freedom) to 4 (very large freedom).

country	Gov_ Head_ Election	Leg_ Repres_ Election	Electoral _laws	Parties_ Organization	Opposition _Opportunity
Abkhazia	3	2	1	2	3
Afghanistan	1	1	1	2	2
Albania	3	3	2	3	4
Algeria	1	1	1	1	1
Andorra	4	4	4	4	4
Angola	0	2	1	2	1
Antigua and Barbuda	4	4	4	3	4
Argentina	4	4	3	4	4
Armenia	2	2	2	3	3
Australia	4	4	4	4	4

Table 1: FreedomHouse political freedom variables.
Snapshot of the first 10 rows and 5 columns of the FreedomHouse political freedom dataset.

country	Media_ Freedom	Religious_ Expression	Accademic _Freedom	Political_ Expression	Assembly _Freedom
Abkhazia	2	2	1	3	3
Afghanistan	2	1	1	2	2
Albania	2	4	3	4	4
Algeria	1	1	2	3	2
Andorra	3	3	4	4	4
Angola	1	2	2	2	2
Antigua and Barbuda	3	4	4	4	3
Argentina	3	4	4	4	4
Armenia	2	2	2	3	3
Australia	4	4	4	4	4

Table 2: FreedomHouse civil liberties variables.
Snapshot of the first 10 rows and 5 columns of the FreedomHouse civil liberties dataset.

The Heritage Foundation has been chosen for the economic freedom variables. Their "economic freedom index" (Edition 2019) dataset is composed of 12

variables ranging from 0 (no freedom) to 100 (very large freedom). These variables have been recorded on 186 observations (countries).

country	Property _Rights	Judical_ Effectiveness	Government _Integrity	Business_ Freedom	Labor_ Freedom
Afghanistan	19.6	29.6	25.2	49.2	60.4
Albania	54.8	30.6	40.4	69.3	52.7
Algeria	31.6	36.2	28.9	61.6	49.9
Angola	35.9	26.6	20.5	55.7	58.8
Argentina	47.8	44.5	33.5	56.4	46.9
Armenia	57.2	46.3	38.6	78.3	71.4
Australia	79.1	86.5	79.9	88.3	84.1

Table 3: Economic Freedom Index dataset.
Snapshot of the first 7 rows and 5 columns of the Economic Freedom Index dataset.

The final freedom dataset has been created by inner joining the two datasets to take into consideration only the countries present in both of them. To create the overall freedom index I definitely need for each country both the groups of freedom variables.
Before joining the two datasets, meticulous work was carried out to standardize the names of the countries so that many observations were not lost just because they were called in different ways.
The final Freedom dataset is therefore composed of 37 variables for 185 observations.

2.2. *Datasets for explanatory variables*

The explanatory variables selected are such that, with a certain level of validity, they could be assumed as exogenous. For this purpose, the scientific literature and the results of the clustering techniques have been used.
From the scientific literature, I got some factors known for influencing the rise and stability of democracy.
From the results of the clustering techniques, it was instead possible, through a good knowledge of the countries' history, to understand which exogenous factors are shared by countries with similar levels of freedom. Some of these are consol-

idated social factors that can influence freedom in the specific sense I'm dealing with.
The specific ways in which these variables could affect the dependent variable will be discussed in much more detail in Section 4.

- **Geographical data:**

Name	**Unit of measure**	**Source**	**Year**
temp	Celsius	Wikipedia	Average over 1961-1990
precip	mm	NationMaster	2008
coast	km	ChartsBin	2010
resources	Revenue % of GDP	WorldBank	2020
region_Americas	Binary	FreedomHouse	2019
region_Asia.Pacific	Binary	FreedomHouse	2019
region_Europe	Binary	FreedomHouse	2019
region_Middle.East.and.North.Africa	Binary	FreedomHouse	2019
region_Sub.Saharan.Africa	Binary	FreedomHouse	2019

Table 4: Geographical variables summary.
Names, unit of measures, sources, and year of data publication of the geographical variables.

The average temperature ($temp$) of a country and the amount of rainfall ($precip$) have been taken into account to represent the climatic conditions of a country. Indeed, it is widely accepted that the climatic conditions of a country, especially the extreme ones, can influence its economic development and the birth of democracy through different channels. Certainly, it would have been useful to have other additional data such as the variance of the annual temperature and the daily temperature range. Unfortunately, contrary to what is expected, they are not easy to be collected. For the purpose of this research, they can be assumed as exogenous as the human effect on climate is little (even if very relevant in other settings). More specifically the effect of freedom on climate can be assumed as irrelevant in explaining the variance of temperature and precipitations over the world.

Coastline length ($coast$) is another geographic variable that can affect freedoms across different channels. Indeed, countries having more access to the sea have had more opportunities for international trade and communication and the possibility to fish (which is an economic activity that is reasonably easy to be hidden from the state). We will see in much more detail how these factors could affect economic and political freedom in Section 4. The rivers length would have been another useful variable for similar reasons, but it was only available for very few countries.
The natural resources revenue ($resources$) is one of the most important variables that affect freedom. There are hundreds of scientific papers on the "resource curse" which tell us that the larger the revenue coming from resources such as oil, minerals, and gas, the smaller the probability for a country to become a stable democracy. This variable is not completely exogenous as it is expressed as a percentage of GDP but for the purpose of this research, it can be assumed as such. Indeed, this variable strongly depends on the available natural resources which are exogenous and it is expressed as a percentage of GDP to make it comparable across countries.
The $region\_X$ variables have been taken into consideration mostly as control variables. If the macro-region to which countries belong would be relevant in the output of the supervised learning techniques it would mean that I'm not taking into account some other relevant variables to explain the variance in freedom.

- **Colonization data:**

Name	**Unit of Measure**	**Source**	**Year**
belgium	Binary	Harvard Dataverse	2019
britain	Binary	Harvard Dataverse	2019
france	Binary	Harvard Dataverse	2019
germany	Binary	Harvard Dataverse	2019
italy	Binary	Harvard Dataverse	2019
netherlands	Binary	Harvard Dataverse	2019
spain	Binary	Harvard Dataverse	2019
portugal	Binary	Harvard Dataverse	2019

Table 5: Colonization variables summary.
Names, unit of measures, sources, and year of data publication of the colonization variables.

Each of these variables is a dummy variable telling if the observations were a colony of the country in the variable name or not. They have been taken into consideration in this research because, even if the type of effect is debated, there is agreement that colonization has a strong impact on freedoms. It is also clear that the effect strongly depends on the colonizer country: for example, Great Britain and Belgium have had two completely different approaches which lead to very different results.
These variables can be assumed as exogenous mainly due to the temporal distance that separates freedoms in 2019 from the colonial era.

- **Religious and ethnic data:**

As I want to explain substantial freedom religious and ethnic data are taken into account. Indeed, as already said in Section 1, the state, even if it is a very important actor in determining the amount of freedom individuals can enjoy, is not the only one. There are a lot of non-state actors, such as the kind of neighborhood you live in, the kinds of restrictions parents are used to imposing, and many other traditions and institutions.
Among the most important ones are the religious institutions and the relationship people have with diversity. There is a lot of scientific literature on the different effects that different religions have on democracy, trade, and self-determination. Some religions create a less suitable underlying for the rise and stability of democracy, civil liberties, and free trade than others. It is also very important how people of different ethnic or religious groups accept the other groups. Ethnic fractionalization and religious diversity in a country are often more important factors than the type of religion or ethnic group.

Name	**Unit of measure**	**Source**	**Year**
fract	Index range: 0-1	Harvard Dataverse	2019
RDI	Index range: 0-10	Pew Research Center	2010
christian	% of the population	Pew Research Center	2010
muslim	% of the population	Pew Research Center	2010
unaffiliated	% of the population	Pew Research Center	2010
jewish	% of the population	Pew Research Center	2010
hindu	% of the population	Pew Research Center	2010
buddhist	% of the population	Pew Research Center	2010
folk	% of the population	Pew Research Center	2010
other_rel	% of the population	Pew Research Center	2010

Table 6: Religious and ethnic variables summary.
Names, unit of measures, sources, and year of data publication of the Religious and ethnic variables.

I collected a variable for each macro-religious group representing the percentage of the population affiliated with each of them. Moreover, two important variables $fract$ and RDI, respectively expressing the ethnic fractionalization index and the religious diversity index, have been collected. All these variables are not completely exogenous. Indeed countries with large mobility rights could have a higher number of minorities. At the same time the richest countries, which are also the freest, are experiencing great migration flows. Even if in other setting these factors would be very relevant, in this framework they are marginal in explain-

ing the world variance in the religious and ethnic population. There is another source of endogeneity for some of these variables. The data have been taken based on affiliation and in some countries is very difficult if you are atheist, Buddhist, or Hindu to express it. This problem will be tackled directly in the supervised learning section 4 as it will be visible when the problem occurs.

- **Communism and socialism data:**

Name	**Unit of measure**	**Source**	**Year**
current_comm	binary	Wikipedia	2022
previous_comm	binary	Wikipedia	2022
current_constitut_ref_socialism	binary	Wikipedia	2022
previous_constitut_ref_socialism	binary	Wikipedia	2022
governing_social_commun	binary	Wikipedia	2022
communism	binary	Wikipedia	2022

Table 7: Communism and socialism variables summary.
Names, unit of measures, sources, and year of data publication of the communism and socialism variables.

The communism and socialism data are fundamental for this research for many reasons. Communism and socialism for their own nature aim at reducing individual freedom to achieve what they claim to be "collective freedom". Independently from the discussion on whether collective freedoms are actual freedoms or not and also regardless of whether they have positive effects or not, we are sure that they reduce the kind of freedom we are discussing in this research. As it is evident that communism and socialism influence freedom rather than the opposite, they can be assumed as exogenous.
The $current\_comm$ variable represents in a binary way if a country considers itself as a communist country in 2022: they have the communist party as the only party.
The $previous\_comm$ variable represents in a binary way if a country has been a communist country in the past.
The $current\_constitut\_ref\_socialism$ variable represents in a binary way if a country still has in its constitution references to socialism or communism.
The $previous\_constitut\_ref\_socialism$ variable represents in a binary way if a country has had in its constitution references to socialism or communism.

The $governing\_social\_commun$ variable represents in a binary way if a country is currently governed by communist or socialist parties without having constitutional references to communism or socialism.
In the end, the $communism$ variable simply represents in a binary way if a country has 1 in at least one of the previous variables. It tells if a country has had anything to do with communism or socialism.

- **Legal system data:**

Name	**Unit of measure**	**Source**	**Year**
legal_Civil.and.Common.Law	binary	Wikipedia	2022
legal_Civil.and.Sharia.Law	binary	Wikipedia	2022
legal_Civil.Law	binary	Wikipedia	2022
legal_Common.and.Sharia.Law	binary	Wikipedia	2022
legal_Common.Law	binary	Wikipedia	2022
legal_Religious.Law	binary	Wikipedia	2022

Table 8: Legal system variables summary.
Names, unit of measures, sources, and year of data publication of the legal system variables.

The legal system variables are useful to represent the cultural influences among countries. Indeed the legal systems have been historically implemented in two ways: imposition by colonization or cultural influence. Even if I already took into consideration the colonization variables, they are very useful to account for the many cases in which a country, to improve its own legal system, incorporates elements from other countries' legal systems. Even if these are the less exogenous variables I will use they still can be assumed as such. It is true that usually, free countries introduce legal elements which are inspired by free countries and vice versa. However, the legal family changes very slowly and the actual one has been consolidated for a very long time. For this reason, we can say that it is the legal system influencing freedom rather than the opposite. The $legal\_X$ variables simply represent in a binary way if the countries belong to that legal system or not.

3. Unsupervised learning

In this section, I will discuss the implementation of two unsupervised learning techniques: Principal Component Analysis and Hierarchical clustering. While

these two techniques have some connection points in this report, the purposes for which they are used are completely different.

3.1. *Principal Component Analysis*

Principal Component Analysis (PCA) is commonly used for two purposes: data visualization and PCA regression. In this research, the main goal of this technique is to create a new index of overall freedom able to retain as much information as possible from the freedom datasets. There are different ways to create an overall freedom index. One of the most simple and common choices is to take the average or the median values for each observation on all the normalized values. Another less simple but still common choice is to aggregate the values based on some arbitrary weights. Most of the freedom institutions assign the weights depending on some knowledge and personal evaluations. These two strategies have two different problems. The first loses a lot of information contained in the original variables, while the second is largely arbitrary and subject to researchers' biases. If we agree on some basic concepts PCA can overcome both these problems. The variance of a variable can be thought of as the amount of information it contains. We can all agree that a variable that quantifies the height of people in a state contains a huge amount of information compared to a variable that represents the number of fingers people have. Few people have a number of fingers other than 20 due to malformations or accidents, while the height of the people varies considerably.
Principal Component Analysis aggregate the variables weighing them based on their variances. It is definitely less arbitrary than choosing the weights directly and the first principal component is built such that it retains as much variance as possible from the original variables. However not all the datasets are suitable for PCA and they must be checked in advance. In the next subsections, I will describe the process required to achieve this aim and I will interpret the results. PCA will also be used for data visualization to allow the representation of the clusters that I will create on the two principal components.

3.1.1. *Preprocessing and pre-analysis*

The preprocessing part for the application of the PCA algorithm requires missing value handling and data scaling on the Freedom Dataset. There are different ways

to handle missing values. The most common are: deleting the observations having them or imputing the mean or median of the variable in which they are. However, these two strategies are not suitable for our setting. The missing values only occur in 6 observations: Iraq, Libya, Liechtenstein, Somalia, Syria, and Yemen. Except for Liechtenstein, these countries are among the least free in the world. The reason why the missing values are present is probably due to this. For this reason, by eliminating these observations we could introduce a small bias in our analysis, and by imputing the mean or median value we would make a big mistake with respect to their real values. Furthermore, 6 observations out of a total of 185 are still a sufficient number to influence the result. For these reasons, I decided to handle the missing values through *KNN-Imputation*. The missing value will be imputed as the average value for that variable among the 5 nearest neighbors concerning all the other variables. I think that this is the best way to address this problem because the 37 variables are very highly correlated and it is very likely that the missing values would be very similar to the one of their neighbors. At this point, the data is ready to be used.

The pre-analysis part consists of the *correlogram* interpretation and the tests on the dataset for adequacy to PCA. The Correlation Plot in Figure 1 already tells us that the Freedom Dataset is very likely to be suitable for PCA: variables are highly correlated. However, the most important facts that we can see in this plot concern the relationships among the groups of variables. The first obvious relationship we can see in the plot is between civil liberties and political freedoms: they are very highly correlated. Indeed, it is almost impossible to have civil liberties without democracy. If minorities cannot take part in the electoral process it is very unlikely they will be able to express themselves freely. Even if it is less evident compared to the previous relationship, also the economic freedom variables are highly correlated with the political and civil ones. The scientific literature tells us that a certain amount of economic freedom is necessary for political freedom to rise.

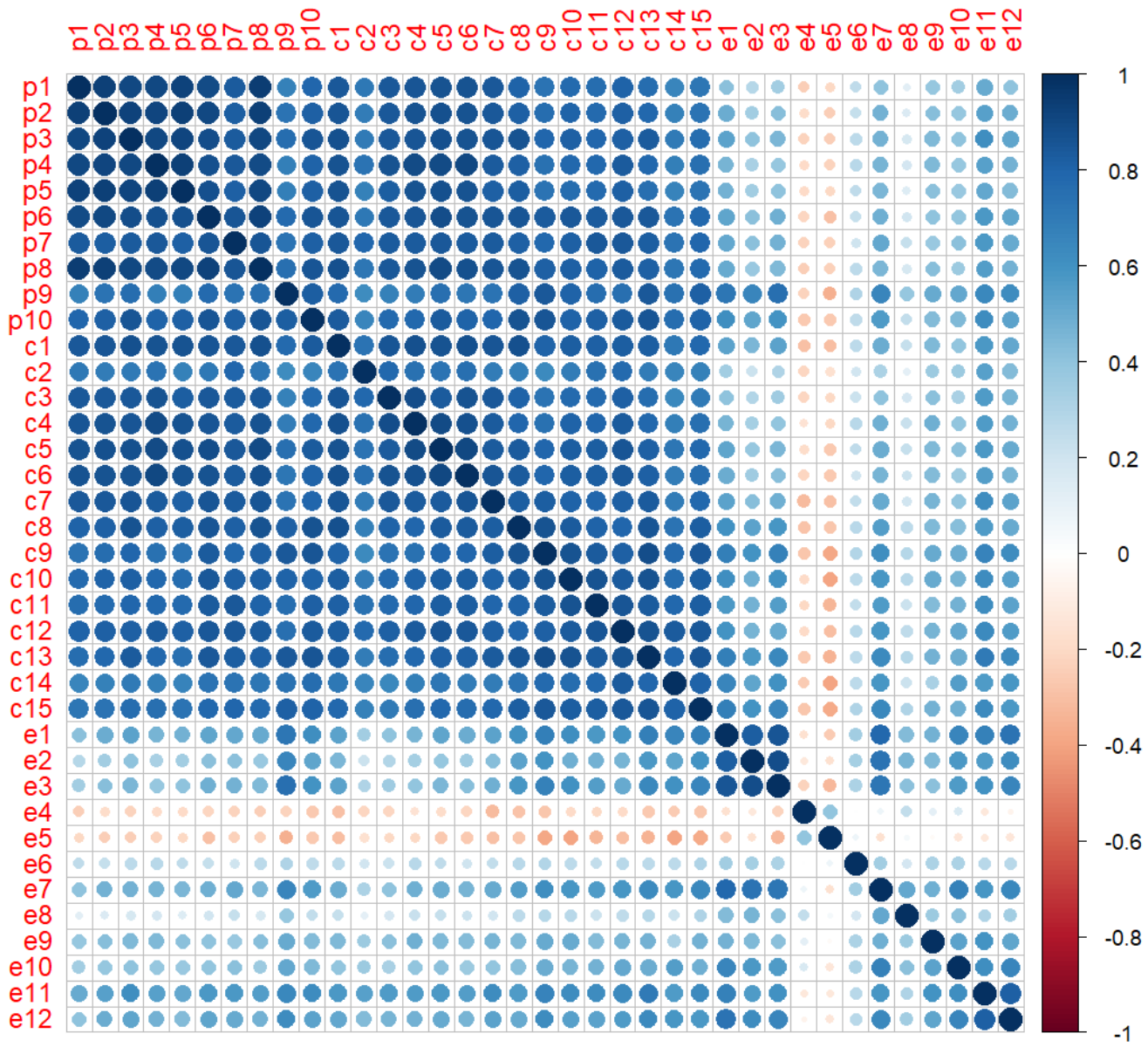

Figure 1. Correlogram Plot from freedom dataset.

Indeed, for democracy to rise and remain stable a large part of the population must be able to enrich itself independently from the sovereign. The sovereign always needs resources to maintain power by distributing wealth to the cohort of people close to him. For this reason, a large group of people must be able to threaten the sovereign not to give him the resources in exchange for political concessions. To make this happen the population must have a certain level of freedom

to undertake their own business independently of the sovereign and to be able to trade. Furthermore, the activities they carry out must be easy to be hidden from the sovereign: fishing and weaving are easier to be hidden than landed properties. These dynamics are studied in political science in a game theory setting. Emblematic is the example of the early rise of democracy in Great Britain compared to the French experience. There are only two exceptions: e4 and e5. Taxation and government spending are indeed negatively correlated with political and civil liberties. Democracy is a great way to settle social conflicts without the use of violence, but at the same time, it is subject to requests from all possible social groups to receive all sorts of subsidies. This led almost every democracy to increase the government size.

To test the adequacy of the Freedom Dataset for PCA implementation I will use the *Kaiser–Meyer–Olkin criterion* and the *Bartlett's test of Sphericity*. The KMO criterion is a measure of the proportion of variance among variables that might be common variance. It ranges from 0 to 1 and the highest the score the more suitable the dataset. The freedom dataset received a score of 0.97 which as stated by the author of this criterion is a marvelous score of suitability. Bartlett's test of Sphericity is used to test the null hypothesis that the correlation matrix is an identity matrix. An identity matrix means that all the variables are independent and that the dataset is not suitable for PCA. The null hypothesis is rejected at the 99.9% confidence interval with a p-value of $2.22e^{-}16$. The Freedom Dataset is perfect for PCA implementation.

3.1.2. *Interpretation*

It is usually very important to choose the number of components to retain using the *screeplot*. In this case, I already know I want to keep only the first principal component to create the new overall freedom variable. My only concern is the amount of variance of the original variables that this new variable retains: the higher it is, the better this can represent the concept I want. However, for the purpose of graphical representation, I kept the first three principal components to be interpreted.

Variable Name	**Comp1**	**Comp2**	**Comp3**
Gov_Head_Election	**-0.879**	-0.341	0.109
Leg_Repres_Election	**-0.901**	-0.255	0.11
Electoral_laws	**-0.918**	-0.202	0.067
Parties_Organization	**-0.901**	-0.272	0.14
Opposition_Opportunity	**-0.887**	-0.277	0.123
Free_From_Domination	**-0.923**	-0.198	0.013
Minorities_Electoral_Opport	**-0.899**	-0.166	0.024
Policy_Determination	**-0.919**	-0.253	0.06
Corruption_Safeguards	**-0.875**	0.21	-0.129
Government_Transparency	**-0.908**	-0.036	-0.061
Media_Freedom	**-0.917**	-0.16	-0.019
Religious_Expression	**-0.77**	-0.248	0.087
Accademic_Freedom	**-0.868**	-0.29	0.119
Political_Expression	**-0.892**	-0.227	0.17
Assembly_Freedom	**-0.916**	-0.194	0.082
NGO_Freedom	**-0.905**	-0.239	0.066
Trade_Unions_Freedom	**-0.901**	-0.152	-0.023
Judiciary_Independence	**-0.914**	-0.062	-0.06
Due_Process	**-0.911**	0.096	-0.148
Protection_Illegittimate_Force	**-0.909**	-0.006	-0.065
Treatment_Equality	**-0.899**	-0.052	-0.053
Mobility_Freedom	**-0.919**	-0.06	0.017
Interference_In_Private_business	**-0.928**	0.065	-0.104
Personal_Social_Freedom	**-0.83**	0.094	-0.194
Equality_Of_Opportunities	**-0.894**	0.099	-0.143
Property_Rights	**-0.697**	**0.59**	-0.15
Judical_Effectiveness	**-0.576**	**0.668**	-0.161
Government_Integrity	**-0.646**	**0.603**	-0.294
Tax_Burden	0.243	0.229	**0.736**
Gov_Spending	0.319	0.023	**0.686**
Fiscal_Health	-0.324	0.263	0.233
Business_Freedom	**-0.654**	**0.569**	0.079
Labor_Freedom	-0.309	**0.521**	0.312
Monetary_Freedom	**-0.541**	**0.334**	0.358
Trade_Freedom	**-0.551**	**0.549**	0.201
Investment_Freedom	**-0.717**	0.348	0.088
Financial_Freedom	**-0.655**	**0.483**	0.05
% of VAR explained	**0.638601**	**0.096624**	**0.046903**

Table 9: Principal Components Loadings.
Loadings for principal components interpretation. Values over |0.4| are in bold.

Looking at Table 9 the first thing to notice is that the first principal component alone retains 64% of the variance of the original variables. This is a huge amount and I couldn't be happier about it. Furthermore, we can see that the first principal component highly correlates with almost every original freedom variable. Simply changing the sign of the scores of this new variable would be perfect for representing the concept of **overall freedom** of a country. The second principal component, instead, highly correlates with most of the economic freedom variables and it can be interpreted as a **strengthening of economic freedom**. The third variable mainly accounts for the two variables not considered by the first two principal components: Tax_Burden and $Gov\_spending$. For this reason, the third principal component can be simply interpreted as the **government size**.

3.1.3. *New variable*

In Table 10 is represented the amount of variance that the first principal component alone is retaining from each of the original variables. As we can see the new overall freedom variable retains a lot of variance from the political and civil variables. It also retains a good amount of variance from the economic variables with some exceptions. For the reason we discussed in Subsection 3.1.1 while analyzing the correlogram plot, democratic countries are subject to every kind of request from all the social categories bringing them to have huge government sizes, extensive labor regulations, frequent expansionary policies, and high public debts. For this reason, the variables whose values are not in bold in the table failed to be largely represented by a single variable expressing the overall freedom in a country.
Taking everything into consideration, the new variable is an excellent overall freedom index: it is almost completely not arbitrary and represents 64% of the information that was contained in 37 freedom variables. The new variable was then scaled to range from 0 to 100.

Variable Name	Communality
Gov_Head_Election	**0.77**
Leg_Repres_Election	**0.81**
Electoral_laws	**0.84**
Parties_Organization	**0.81**
Opposition_Opportunity	**0.79**
Free_From_Domination	**0.85**
Minorities_Electoral_Opport	**0.81**
Policy_Determination	**0.84**
Corruption_Safeguards	**0.77**
Government_Transparency	**0.82**
Media_Freedom	**0.84**
Religious_Expression	**0.59**
Accademic_Freedom	**0.75**
Political_Expression	**0.80**
Assembly_Freedom	**0.84**
NGO_Freedom	**0.82**
Trade_Unions_Freedom	**0.81**
Judiciary_Independence	**0.84**
Due_Process	**0.83**
Protection_Illegittimate_Force	**0.83**
Treatment_Equality	**0.81**
Mobility_Freedom	**0.84**
Interference_In_Private_business	**0.86**
Personal_Social_Freedom	**0.69**
Equality_Of_Opportunities	**0.80**
Property_Rights	**0.49**
Judical_Effectiveness	**0.33**
Government_Integrity	**0.42**
Tax_Burden	0.06
Gov_Spending	0.10
Fiscal_Health	0.10
Business_Freedom	**0.43**
Labor_Freedom	0.10
Monetary_Freedom	0.29
Trade_Freedom	**0.30**
Investment_Freedom	**0.51**
Financial_Freedom	**0.43**

Table 10: First Principal Component Communalities.
Communalities of the first principal component with the original variables. Values higher than 0.3 in bold

3.1.4. *Representation*

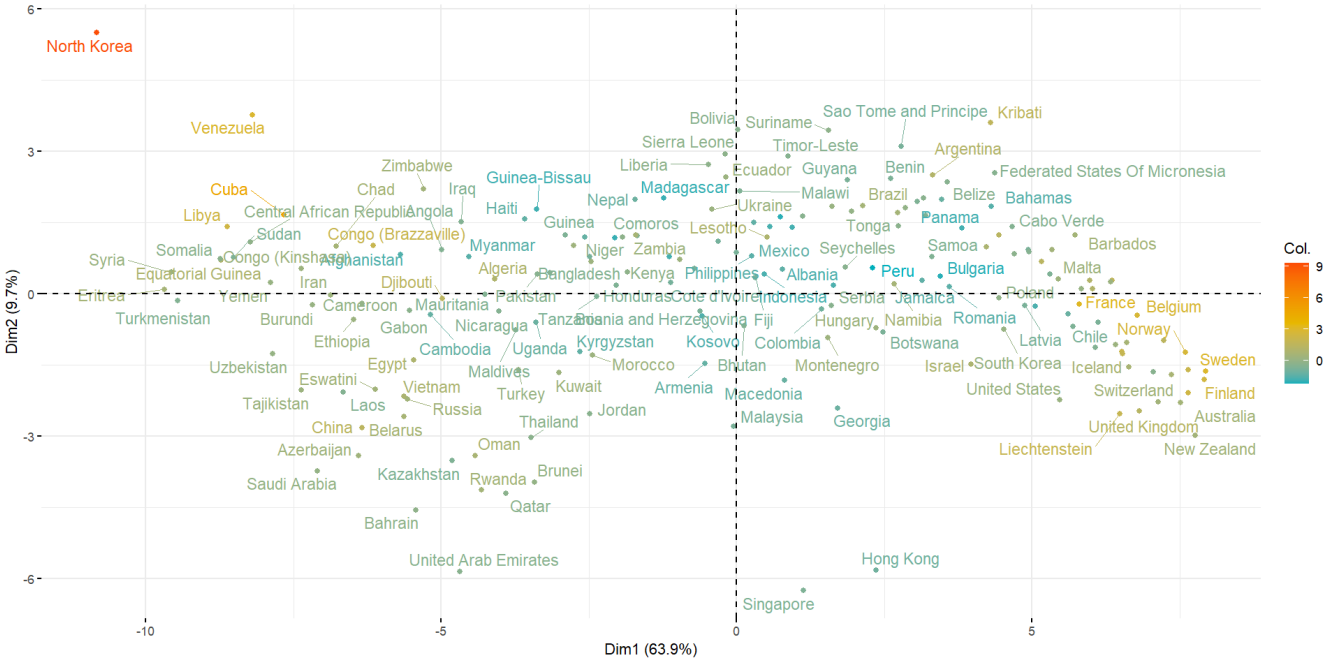

Figure 2. 2-Dimensional Graphical representation of the 185 countries.
Countries are represented in the first 2 principal component space. The color represents the value of the third component.

Figure 2 represents all 185 countries in the space of the first two principal components. This representation will be useful to visualize and interpret correctly the clusters that will be created. Indeed, by exploiting the *loadings* representation in Figure 3 we can understand the freedom of the countries. The higher the value of the x-axis the higher the overall freedom. At the same time, the lower the value of the y-axis, the higher the additional economic freedom enjoyed by the individuals with respect to the amount already taken into account by the first component. Furthermore, we have to remember that the x-axis is far more important than the y-axis as it represents much more variance.

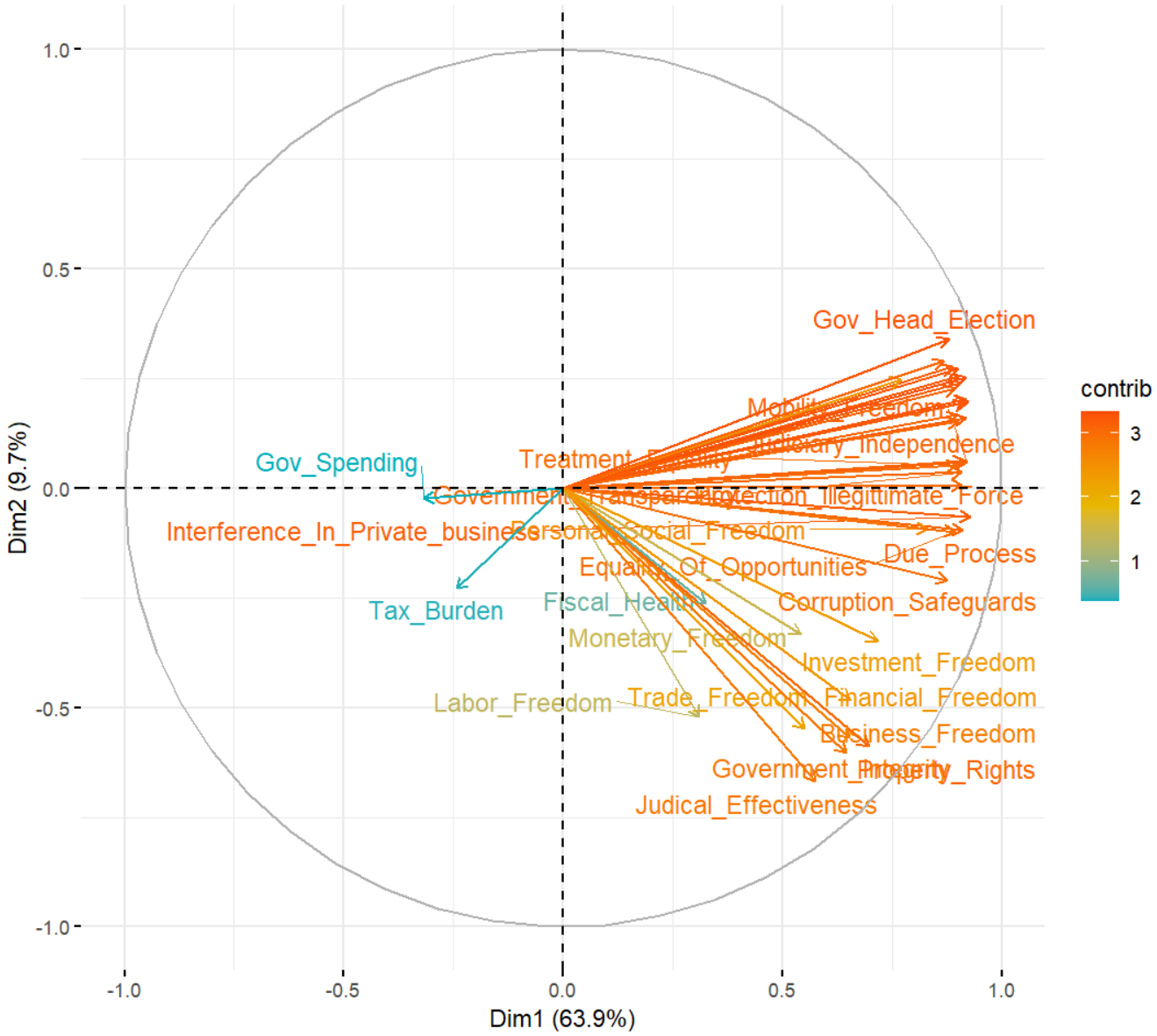

Figure 3. 2-Dimensional Graphical representation of the loadings.

3.2. *Hierarchical Clustering*

The clustering section is useful for my analysis mainly for one reason. With a little knowledge of the history and characteristics of the countries of the world, it is possible to understand which other exogenous factors are shared by countries with similar levels of freedom. In fact, this technique was performed before the

construction of the dataset for supervised learning, and some variables such as colonization or communism were added based on the output. *Hierarchical clustering* was chosen over the classic kmeans because it allows us to build clusters without having to decide the number in advance. I will set the number of clusters based on their interpretability. Furthermore, hierarchical clustering allows us to use many different distance measures. To perform a meaningful interpretation of the clusters I used both the loadings of the PCA from Subsection 3.1 and the clustering *heatmaps* showing how the variables behave in the different clusters (heatmaps are not shown in this report for brevity). For each of the hierarchical clustering implementations, I used the complete linkage and I only changed the distance measure.

3.2.1. *Preprocessing*

As hierarchical clustering preprocessing requirements are the same for the PCA implementation I will use the scaled Freedom dataset with the same imputation technique for missing values (KNN-Imputation). The only additional step is to create the distance matrix using each of the three different distance measures. On this distance matrices will be applied the clustering algorithm.

3.2.2. *Euclidean distances*

The first distance measure I used is the *Euclidean distance*. It is by far the most used distance measure and it is the one closest to the concept of distance that each of us has in mind. As I was able to make a reasonable interpretation of the clusters up to the number 6, this was the selected number.

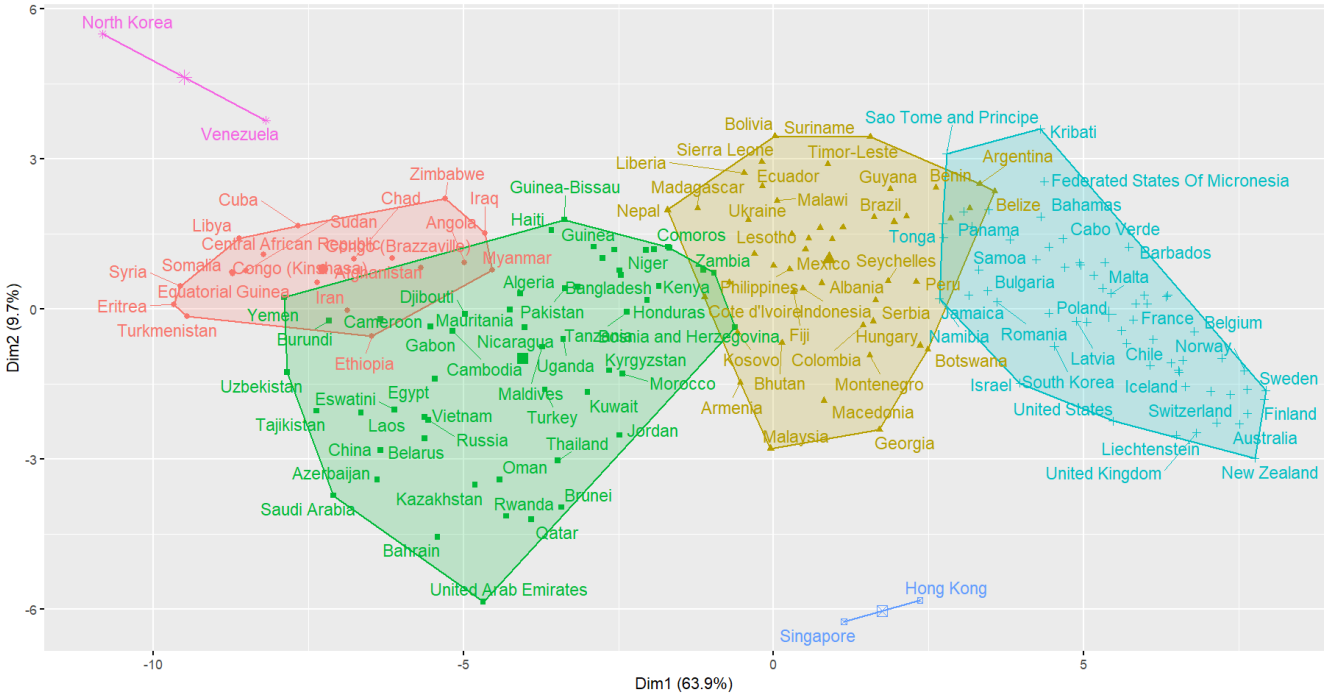

Figure 4. Hierarchical Clusters with Euclidean Distances representation
Countries are represented in the first 2 principal component space grouped based on the clusters output by hierarchical clustering with euclidean distances.

In Figure 4 we can see that the countries belonging to the same cluster have a great deal in common. The 4 most populated clusters almost perfectly coincide with the classification of *political compass*.

The **blue cluster** is populated by the so-called libertarian-right meaning countries which at the same time favor political, civil, and economic freedoms. This cluster is composed almost exclusively of Western countries that share many historical phenomena (wars, plagues, industrial revolutions, etc.) and have a long history of democracy and economic interdependence. There are only a few peculiar exceptions as Japan, South Korea, Uruguay, and Chile which have very particular histories that well explain their high level of freedom. Although I would very much like to make a historical digression on these countries, I don't think this is the right framework.

The **yellow cluster** is populated by the so-called libertarian-left meaning countries which have a good level of political and civil liberties with respect to the rest of the world. At the same time, they are characterized by a poor level of economic freedom. It is composed largely of countries of Eastern Europe and South America.

Eastern European countries are those countries that have been part of the Soviet Union or Yugoslavia. After their fall they developed political and civil liberties but remained characterized by a cultural heritage that was not very favorable to the free market. The South American countries in this cluster similarly have a historical background linked to socialist or communist governments.
The **green cluster** is populated by the so-called authoritarian-right meaning countries which are not politically and civilly free, but they have a certain level of economic freedom. It is composed largely of countries that have an enormous amount of natural resources. Indeed, in these countries, it is very difficult for democracy to rise because the sovereigns have plenty of monetary resources to keep their power as they control natural resources. People have no credible threat to the sovereign as he already has at his disposal what is necessary to remain in his position. However, a little but relevant amount of economic freedom is guaranteed as they need international trade to sell natural resources.
The **red cluster** is populated by the so-called authoritarian-left meaning countries which are neither politically nor civilly nor economically free. The countries composing this cluster have a large combination of factors. We have communist countries or countries that experienced it. A large part of them have a large amount of natural resources and finally, they are very homogeneous from a religious point of view: the Muslim population is often over 90%.
In the end, we have two very little and special clusters. One is composed by **North Korea and Venezuela** which are the least free countries in the world and are currently under ferocious communist dictatorships. The other is composed by **Singapore and Hong Kong** which are extremely peculiar. They are two of the freest country from an economical point of view and have a good amount of political e civil liberties. They both have a large part of the population composed of Chinese people, and they have been British colonies for a long period (Hong Kong for more than 150 years).

3.2.3. *Manhattan distances*

The *Manhattan distance* is a distance measure often used instead of Euclidean distance when dealing with geographic coordinates. This is obviously not our case and the only reason it is present in this research is to make a comparison between some different distance measures we can use in hierarchical clustering. Even in this case, the number of clusters I could manage to give a reasonable interpretation

is 6.

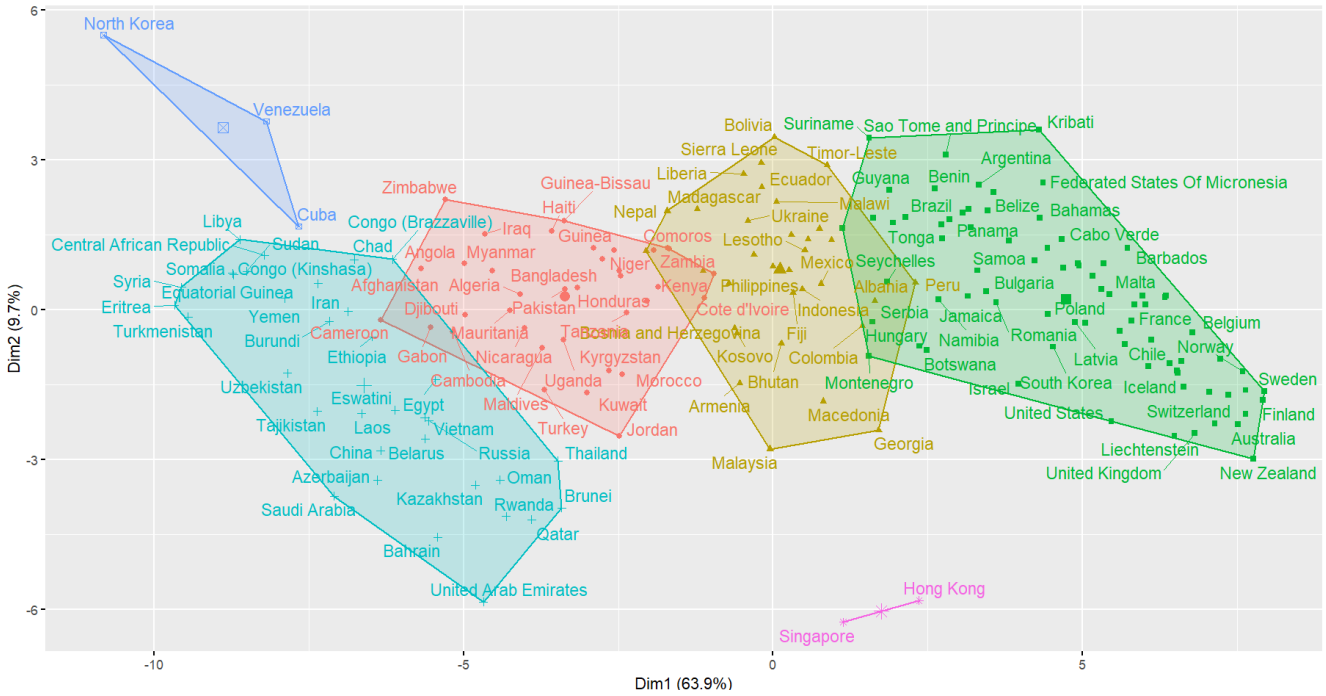

Figure 5. Hierarchical Clusters with Manhattan Distances representation
Countries are represented in the first 2 principal component space grouped based on the clusters output by hierarchical clustering with Manhattan distances.

The cluster interpretation is the same, but the approximation of the clusters to the political compass classifications is less precise.

3.2.4. *Correlation distances*

Correlation-based distance considers two objects to be similar if their features are highly correlated, even though the observed values may be far apart in terms of Euclidean distance. For example, Sweden and North Korea are in the same cluster because they both have almost the same freedom level for all the freedom variables: while Sweden has a very high value for all the variables, North Korea has almost zero in every variable. It is used to identify clusters of observations with the same overall profiles regardless of their magnitudes. This is particularly the case in gene expression data analysis, where we might want to consider genes as similar when

they are up and down together. It is certainly not our case and once again I have used this distance measure for comparison. In this case, I choose 3 as the number of clusters for which a reasonable interpretation could be done.

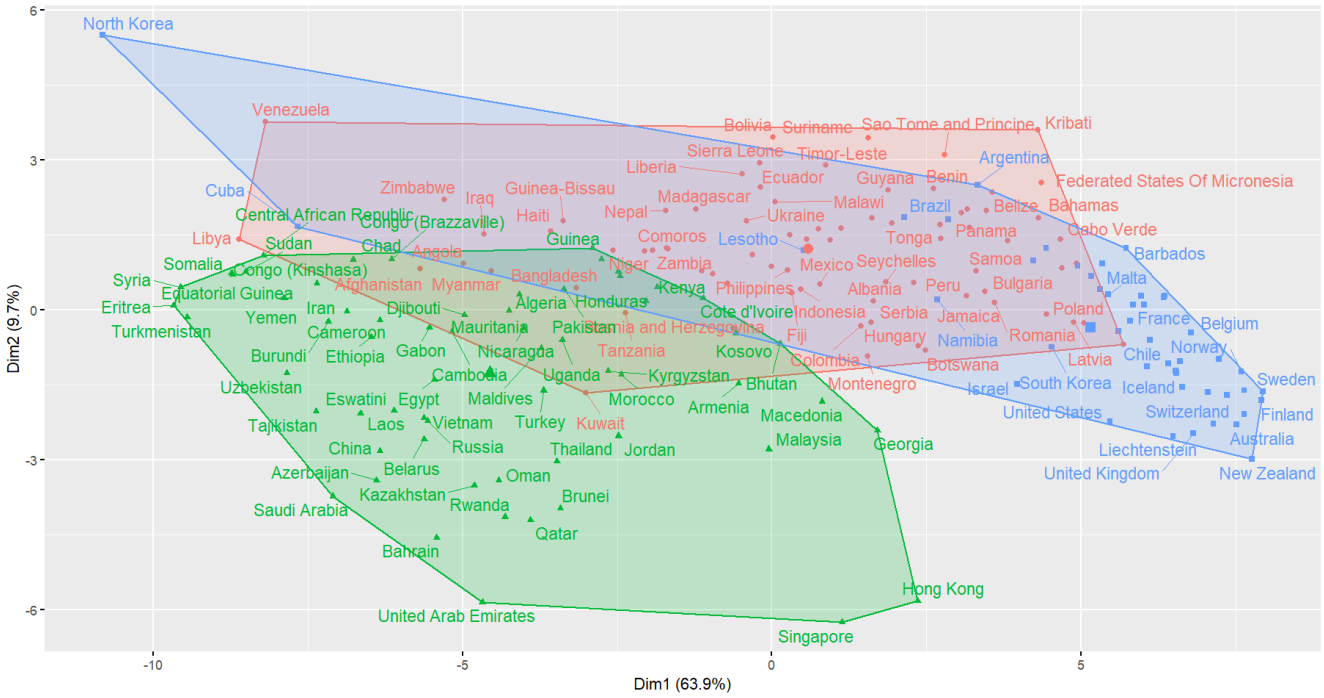

Figure 6. Hierarchical Clusters with Correlation Distances representation
Countries are represented in the first 2 principal component space grouped based on the clusters output by hierarchical clustering with correlation distances.

The **Blue cluster** is populated by countries having similar values for all the freedom variables peaking down for *tax_burden* and *gov_spending*.
The **green cluster** is populated by countries having higher values for economic freedom variables than for the political a civil liberties ones.
The **red cluster** is the opposite of the green one: higher values for the political freedom variables than for the economic freedom ones.

3.3. *Findings from unsupervised learning*

This subsection, as Subsection "Findings from supervised learning" 4.3 is simply a list of the main findings/keypoints from the techniques used until now.

- Using the KNN-Imputation I handled the missing data so that the imputed values are very similar to the real values.

- I found out that the freedom dataset is perfectly adequate for PCA implementation and through the correlogram, I was able to analyze the relationships among the 37 variables.

- I found out that it is possible to create an overall freedom index through the use of PCA which retains 64% of the information that was contained in 37 variables of freedom. This new variable has been constructed in such a way that it is very little arbitrary and the loadings' interpretation tells us it perfectly fits with the concept I wanted.

- Using the first two principal components from PCA I represented the clusters resulting from the hierarchical clustering with 3 different distance measures: the one that is suitable for our setting is definitely the euclidean distance.

- From hierarchical clustering (euclidean distance) I understood some exogenous factors common to the countries belonging to the same cluster. Variables related to communism and colonization will be used in supervised learning because of this.

4. Supervised learning

In this section, I will discuss the implementation of two supervised learning techniques: Decision Tree and Random Forest. Even if decision tree is used mainly for interpretation and random forest mainly for predictions, the two aims will be addressed for both of them.

4.1. *Decision Tree*

Decision Tree is a supervised learning approach used both for classification and regression problems. As the dependent variable I'm trying to explain and predict is a continuous variable I will use the decision tree model for regression. The

decision tree model was chosen because it is more functional to our problem than all the other techniques. I want to be able to interpret the relationship between the exogenous explanatory variables and the overall freedom variable. I am therefore very interested in a highly interpretable model. The choice was between linear regression and decision tree. The decision tree model won over linear regression because the underlying relationship I want to explain is highly non-linear and this model is much better at handling non-linearity. Furthermore, the decision tree does not require testing the many fundamental assumptions for linear regression. In our case, the dependent variable does not have a completely normal distribution, and although some of the independent variables are correlated I still want to keep them in the model. Furthermore, to deal with multicollinearity, I could not use PCA because the resulting independent variables would be difficult to interpret. Moreover, the few high leverage points (North Korea and Venezuela mainly) are not data collection errors but are perfectly explainable with a highly non-linear function. The decision tree is, however, definitely inadequate for prediction purposes. Random Forest Will be used later to overcome this issue.

4.1.1. *Preprocessing*

The decision tree algorithm requires almost no preprocessing. There is no need for data scaling or missing data handling. However, to compare the performance of the different models they must be implemented on the same dataset. The Random Forest algorithm requires the missing values to be handled. I decided to handle the missing values using again KNN-Imputation for almost the same reasons I discussed in the previous section. Even in this case, the missing values are mostly concentrated in countries where there is less freedom. The missing values are few but distributed over many observations and their elimination would lead to a significant reduction in the number of observations. Furthermore, the independent variables are quite correlated with each other. For all these reasons KNN-Imputation is again the most suitable choice. Indeed, after checking the imputed values for some observations for which I could find on Wikipedia the real missing values, I can confirm the previous statement. The final dataset will be composed of 185 observations, the overall freedom variable (dependent variable), and 41 exogenous explanatory variables.

To evaluate the performance of the models the final dataset has been randomly split into a training set (141 observations) and a test set (44 observations). The same training set and test set have been used for all the models (decision tree and random forest) to be able to compare them.

4.1.2. *Interpretation*

The result of the implementation of the decision tree algorithm giving as input all the independent variables let a problem, that we had already discussed, arise. The variables that represent the percentage of people affiliated with minority religions (Buddhism, Hinduism, folk religions, and others) appear to have endogeneity. In fact, in some not free states, it is extremely difficult to freely declare affiliation to one of these religions, and the decision tree in some nodes used precisely these variables with splits that did not make sense otherwise ($Buddhists < 0.0005$) to recognize these countries as very little free. For this reason, these variables were not used in the models. The decision tree model trained on the training set without any pruning is represented in Figure 7.

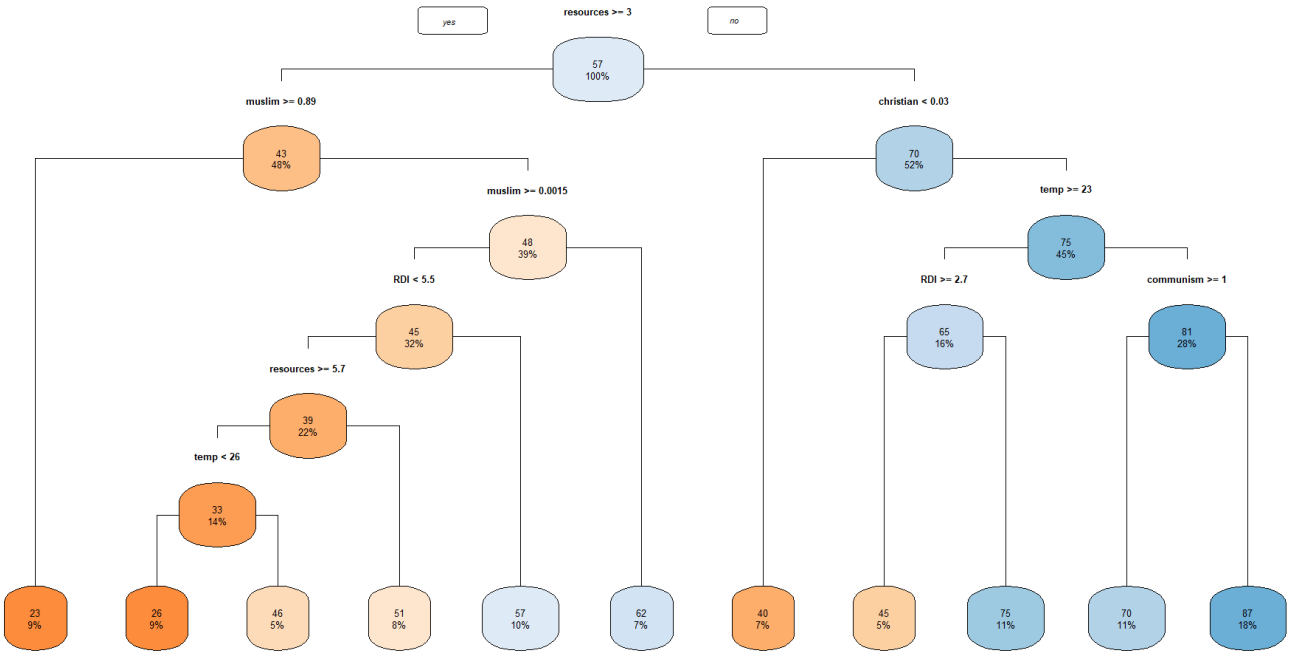

Figure 7. Decision Tree on the training set representation

The tree is perfectly consistent with the expectations I had about the effect of the explanatory variables.
The **first leaf** contains 9% of the observations having an average $overall\_freedom$ of 23 (the lowest of all the leaves). This leaf corresponds to the countries having more than 3% of the GDP coming from natural resources and more than 89% of the population affiliated with Islam. The first factor ensures that the sovereign has enough monetary resources to distribute to his cohort to maintain the power. In this way, democracy hardly manages to rise, and even more difficult to be stable. Furthermore, such a high percentage of people affiliated with Islam often leads religious dogmatism to become the law of the state, significantly reducing freedoms.
The **second and the third leaves** are separated evidently because of overfitting. It does not make any sense to think that having an average temperature lower than 26 degrees (Celsius), which implies no extreme conditions, would lower freedoms. For this reason, they would be interpreted together: they are composed of 14% of the countries and they have an average $overall\_freedom$ of 36. They all have more than 5.7% of the GDP coming from natural resources and a percentage of people affiliated with Islam between 0.15% and 89%. Moreover, they have a religious di-

versity index lower than 5.5 (it ranges between 0 and 10). The interpretation for this group of countries is very similar to the previous one. They have little more freedom only because the religious homogeneity is less than the previous one thus decreasing the probability of the religious dogmatism becoming state law.
The **fourth leaf** contains 8% of the observations having an average $overall\_freedom$ of 51. They differ from the countries in leaves two and three because they have a lower percentage of the GDP coming from natural resources (between 3 and 5.7): less stable power to the sovereign.
The **fifth leaf** contains 10% of the observations having an average $overall\_freedom$ of 57. They differ from leaf four only because they have a religious diversity index higher than 5.5: the probability of religious dogmatism becoming state law is even lower.
The **sixth leaf** contains 7% of the observations having an average $overall\_freedom$ of 62. The fundamental difference with the previous leaf is that this group of countries does not have people affiliated with Islam (lower than 0.15%) suggesting that Islam is a problematic religion by itself for individuals to enjoy freedom.
The **seventh leaf** contains 7% of the observations having an average $overall\_freedom$ of 40. This group of countries has less than 3% of the GDP coming from natural resources and less than 3% of the population affiliated with Christianity. This seems to suggest that having at least a little percentage of Christian people positively influences freedom. In my opinion, this tree built over this specific training set has this split because of the endogeneity issue. Indeed, what this split suggests does not seem reasonable to me. It is instead likely that this tree is picking those countries where being affiliated with Christianity is dangerous. These countries are less free. We still keep the Christian percentage variable in the models because in different trees it produces meaningful splits while the ones we removed never produced meaningful splits.
The **eighth leaf** contains 5% of the observations having an average $overall\_freedom$ of 45. This group of countries has more than 3% of Christians, an average temperature higher than 23 degrees (Celsius), and a religious diversity index higher than 2.7. An average temperature higher than 23 is likely to be an extreme climatic condition if you think that Italy (not a cold country) has an average temperature of 13 degrees. A very high temperature can negatively influence freedom in many different and complicated ways. Among the simplest we could think about is drought: it is well known that where the available water is scarce, social conflicts arise. Moreover, if we add a relevantly high religious diversity index, social conflicts could also be motivated by religious hate and minorities can be dominated.

The **ninth leaf** contains 11% of the observations having an average $overall\_freedom$ of 75. This group of countries differs from the previous one only because it is characterized by a lower religious diversity index. It seems to suggest that in very extreme climatic conditions a certain cultural homogeneity positively influences the overall freedom.
The **tenth leaf** contains 11% of the observations having an average $overall\_freedom$ of 70. This group of countries has more than 3% of Christians, an average temperature lower than 23 degrees (Celsius), and have had to do with socialism or communism (see how this variable is constructed in Section 2.2). These countries do not have extremely high average temperatures, but they have had to do with socialism or communism which by their very definition reduce the level of individual freedom.
To conclude, there is the freest group of countries: the one in **leaf eleven**. It contains 18% of the observations and has an average $overall\_freedom$ of 87. This group differently from the previous one never experienced communist or socialist regimes.

Variable Name	Importance
resources	28963
muslim	26097
christian	22107
fract	21505
temp	15566
region_Sub.Saharan.Africa	11140
RDI	11006
region_Europe	9362
precip	8327
unaffiliated	5686
legal_Civil.and.Sharia.Law	3235
communism	2681
legal_Religious.Law	2504
previous_comm	2502
region_Americas	2038
coast	962
britain	904
region_Middle.East.and.North.Africa	737
spain	514

Table 11: Decision tree variable importance.
Relative influence of each variable on the MSE

Table 11 shows the variable importance measured as the relative influence that each of the explanatory variables has in reducing the MSE (according to this tree model). The importance of the variables is fairly consistent with my initial guesses. What is worth noticing is that the $region\_X$ variables are still quite important, meaning that there is still something that countries in the same region have in common that influences the overall freedom. This is not a surprise as we are only using variables that have a certain degree of exogeneity. Indeed freedom has very complex relations with all the other factors. For example, even if we discarded the per capita GDP as explanatory variable because economic freedom causes economic development, it is also true that the relationship is not unidirectional.

4.1.3. *Pruning*

From the interpretation of the tree, it is evident that the split that separates leaf two from leaf three is probably due to **overfitting**. To tackle this problem, this section will address the *pruning* procedure. The tree model has been pruned using *cross-validation* over the training set to find the optimal value for the complexity parameter. The *complexity parameter* is the minimum improvement in the model needed at each node for growing the tree.

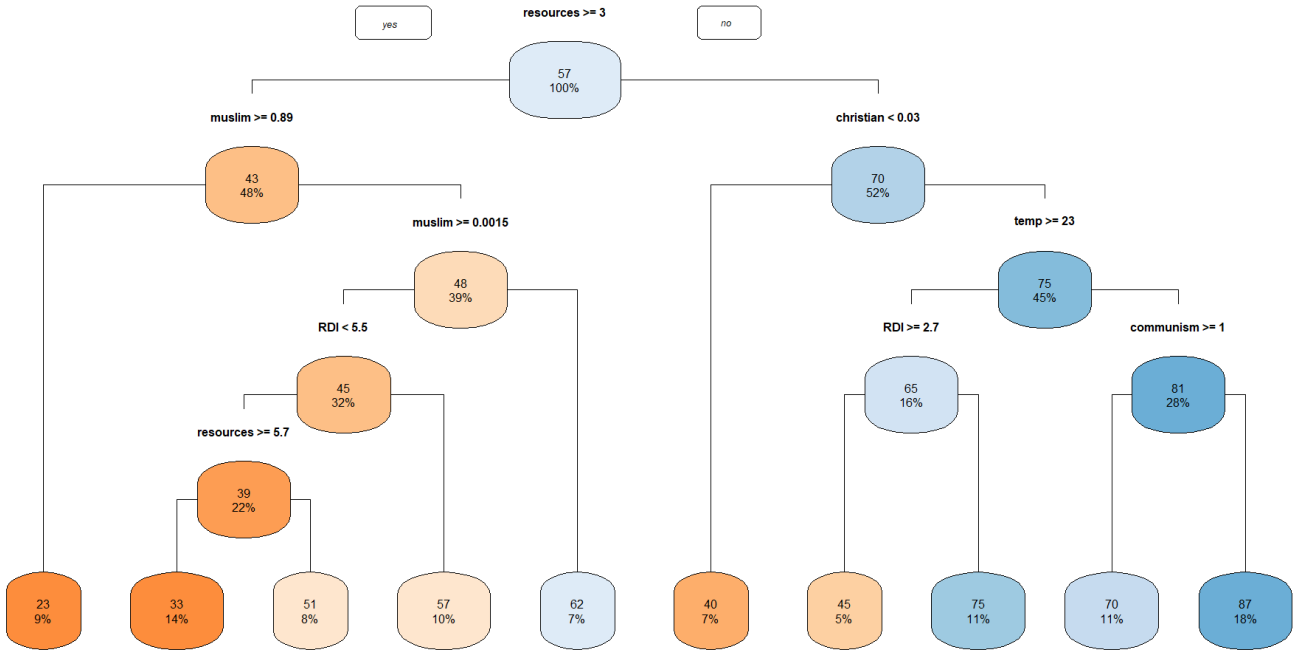

Figure 8. Pruned Decision Tree representation

In Figure 8 the pruned tree is displayed. As we can see the only split that has been eliminated is the only one without a meaningful interpretation.

4.1.4. *Forecasting*

In this Section, I will use the two previous models (pruned and not pruned) to make predictions over the test set and compare their performances. The tree model with-

out any pruning has a root mean squared error (RMSE) of **20.12**, while the pruned model has an RMSE of **19.52**. As expected these are not very good performances knowing that the value of the overall freedom variable ranges from 0 to 100. Decision trees are a very good tool for interpretation but they are not adequate for prediction purposes. However, we can still appreciate that the pruning of the tree reduced overfitting and slightly reduced the RMSE.

4.2. *Random Forest*

Random Forest algorithm is a supervised learning technique created to overcome the low predictive performances of decision trees. It is based on the concept of ensemble learning, which is a process of combining multiple predictors to solve a complex problem. The Random forest algorithm does not rely on one decision tree: it builds many decision trees using different subsets of variables. It takes the prediction from each tree and based on the average of the predictions it predicts the final output.

4.2.1. *Preprocessing*

The preprocessing techniques I applied to the dataset to implement the random forest algorithm are the same I used for the decision tree implementation (Section 4.1.1). Even if the decision tree did not require missing value handling, it was still performed to implement the different models on the same dataset and be able to compare them. Furthermore, also the training set and the test set used to evaluate the performance of the model are the same partitions of the dataset.

4.2.2. *Tuning*

Random forest algorithm have some hyperparameters to be *tuned* to increase the model performances. This was achieved through the use of cross-validation on the training set. Two different R packages that implement the random forest algorithm have been applied: $randomForest$ and $ranger$. Both of them allow us to perform cross-validation directly in the training function. However, randomForest cross-validation only allows us to tune the "$mtry$" hyperparameter: the number of variables randomly sampled as candidates at each split. The ranger

package, instead, allow us to tune three hyperparameters: "$mtry$", "$splitrule$", and "$min.node.size$". The "splitrule" parameter is the splitting rule for growing a regression tree and has different possibilities: "$variance$", "$extratrees$", "$maxstat$", or "$beta$". The "min.node.size" parameter is instead the minimal node size allowed. Both the random forest models will be composed of **1000 trees**.
The first model created with the randomForest package has mtry = 2, while the model created with the ranger package has mtry = 18, splitrule = extratrees, and min.node.size = 5.

4.2.3. *Forecasting*

In this Section, I will use the two random forest models to make predictions over the test set and compare their performances also with the decision tree models. The random forest model built using the randomForest package and tuning only the mtry parameter has an RMSE of **14.83**. This is a great improvement compared to the decision tree models RMSE (around 20) but we can do even better with the ranger package. Indeed, the random forest model built tuning three hyperparameters has an RMSE of **13.69**. This is a small RMSE as the overall freedom variable ranges from 0 to 100 and it is incredible that we reach it only by using exogenous variables.

4.2.4. *Interpretation*

Even if the random forest algorithm was mainly used to improve the predictive accuracy of the decision tree models it still provides some useful information for interpretation. In Figure 9 and in Table 12 we can respectively see the relative influence of each variable on the MSE. These measures of importance, being provided by a more accurate algorithm, are more reliable than the ones provided by the decision tree models. Although the importance values are represented on a different scale for each model, what we are interested in is their relative order.

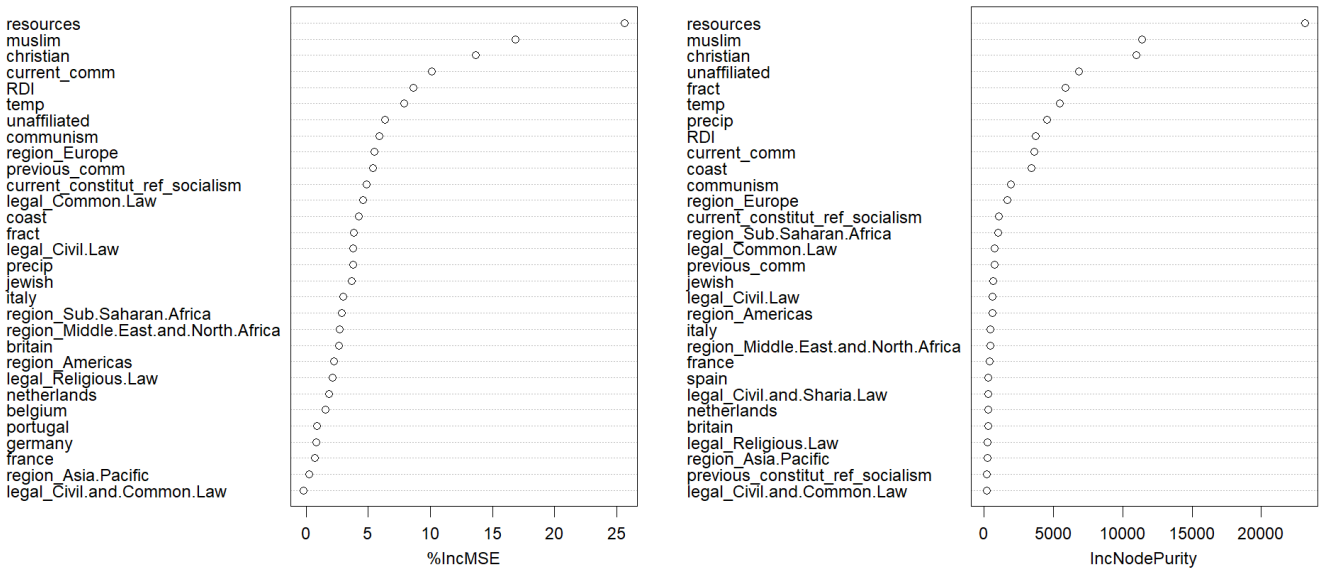

Figure 9. Random Forest Variable Importance representation (randomForest package) The %IncMSE represents the relative influence of each variable on the MSE, while IncNodePurity represents the relative influence of each variable on the Gini index.

In the first plot of Figure 9 we can see that the five most important variables are in order: the natural resources revenue, the percentage of the population affiliated with Islam, the percentage of the population affiliated with Christianity, being a communist country at this moment and the religious diversity index.

Variable Name	Overall	Variable Name	Overall
muslim	100.0	legal_Religious.Law	9.3
resources	68.2	legal_Civil.Law	8.8
region_Europe	58.1	legal_Civil.and.Sharia.Law	8.4
christian	51.4	italy	8.2
current_comm	47.0	region_Middle.East.and.North.Africa	7.9
region_Sub.Saharan.Africa	45.8	britain	7.2
communism	36.6	region_Asia.Pacific	7.0
unaffiliated	31.7	previous_constitut_ref_socialism	6.9
temp	25.2	spain	5.5
fract	23.2	jewish	4.6
previous_comm	19.7	governing_social_commun	4.3
precip	19.5	legal_Civil.and.Common.Law	4.2
RDI	19.0	netherlands	3.6
current_constitut_ref_socialism	17.4	germany	3.0
coast	15.2	portugal	2.7
legal_Common.Law	15.2	belgium	2.1
france	10.5	legal_Common.and.Sharia.Law	0.0
region_Americas	10.1		

Table 12: Random forest variable importance (Ranger package).
Relative influence of each variable on the MSE

In Table 12, showing the relative importance given by the best performing model, we can see instead that the five most important variables are in order: the percentage of the population affiliated with Islam, the natural resources revenue, being in the European region, the percentage of the population affiliated to Christianity and being a communist country at this moment. What is worth noticing is that the relative importance order is quite consistent among the different models and that each of the important variables has a very robust interpretation, which I discussed in previous sections, of their relationship with the dependent variable.

4.3. *Findings from supervised learning*

This subsection summarizes with a simple list of the main findings/keypoints of the supervised learning process.

- I used the decision tree model mainly to interpret the complex and highly non-linear relationship underlying between the exogenous explanatory variables and the overall freedom index I created.

- The exogenous explanatory variables have been found to be absolutely important in describing the conditions that can make freedoms develop and be stable in a country.

- The underlying relationship I found is very consistent with the political science literature and my additional guesses.

- Decision tree models, even if pruned, are inadequate for predictive purposes. Random forest model improved a lot the predictive accuracy.

- The most important exogenous variables for predicting the overall freedom according to the best performing model are: the percentage of the population affiliated to Islam and Christianity or unaffiliated, the natural resources revenue, having had anything to do with communism or socialism, the religious and ethnic heterogeneity, the climatic conditions and the belonging to the European or Sub Saharan regions.

5. Conclusions

Freedom is a complex concept, full of facets and many interacting causes. In this research, I tried to treat it as if it were a crop whose quality is evaluated based on the characteristics of the soil, the climatic conditions, and its biosphere. Although this is certainly a simplification, simplified models like this allow us to understand complex relationships piece by piece. I worked on a dataset built specifically for this research by combining data from different sources and applying the appropriate preprocessing techniques (scaling or knn-imputation). First I built an indicator that best suited the concept I wanted to explain using the PCA. I thus

obtained a variable that represented the overall freedom in a country. I continued the unsupervised learning process by applying the hierarchical clustering technique. From this, I discovered other exogenous variables that could be useful in the analysis. At this point, I moved on to the supervised learning part where I understood some fundamental parts of the underlying relationship by interpreting the output of the decision tree model. Finally, I tried to make predictions using the decision tree models which turned out to be inadequate. For this reason, I used the random forest algorithm suitably tuned to successfully increase the predictive performances.

From this whole process, I have learned that, from a certain perspective, treating freedom as if it were a cultivation is not wrong. Among other things, I have learned that some religions are a more or less favorable substrate than others for freedom to rise. That too many ethnic and religious differences, under certain conditions, can lead to harsh social conflicts undermining the freedoms of minorities. On the contrary, I also learned that, in other conditions, a certain level of ethnic and religious heterogeneity prevents the dogmatisms of one social group from becoming submission for others through the law. I learned that the presence of too many fixed and difficult-to-hide resources (oil, minerals, but also crops) makes the sovereign unlikely to be threatened and prevents the rise of democracy. Last but not least, I have learned that socialism and communism, regardless of what "higher" purpose they aim at, are direct enemies of individual freedoms.