Principal Component Analysis

Dataset testDF_imputed$completeObs

This dataset contains 58 individuals and 50 variables.

---

1. Study of the outliers

The analysis of the graphs does not detect any outlier.

---

2. Inertia distribution

The inertia of the first dimensions shows if there are strong relationships between variables and suggests the number of dimensions that should be studied.

The first two dimensions of analyse express **51.61%** of the total dataset inertia ; that means that 51.61% of the individuals (or variables) cloud total variability is explained by the plane. This percentage is relatively high and thus the first plane well represents the data variability. This value is strongly greater than the reference value that equals **13.88%**, the variability explained by this plane is thus highly significant (the reference value is the 0.95-quantile of the inertia percentages distribution obtained by simulating 5856 data tables of equivalent size on the basis of a normal distribution).

From these observations, it should be better to also interpret the dimensions greater or equal to the third one.

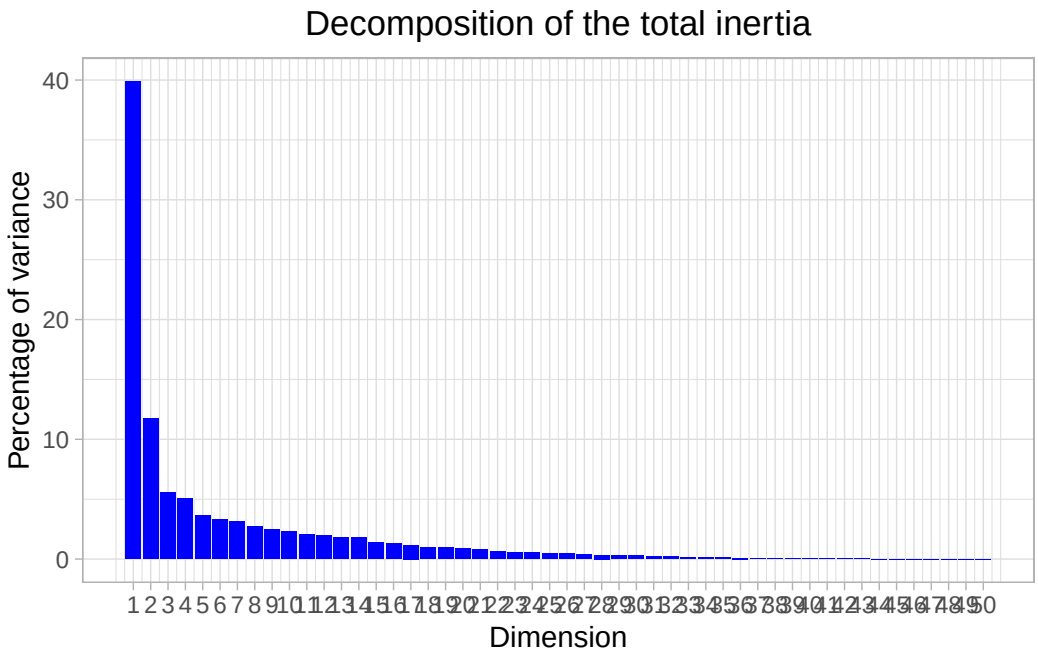

Figure 2 - Decomposition of the total inertia

We can observe that the first 2 axis present an amount of inertia greater than those obtained by the 0.95-quantile of random distributions (51.61% against 13.88%). Thus, a wise decision would be to restrict the description to these only axis. However, we choosed to describe the first 4 axis.

---

3. Description of the plane 1:2

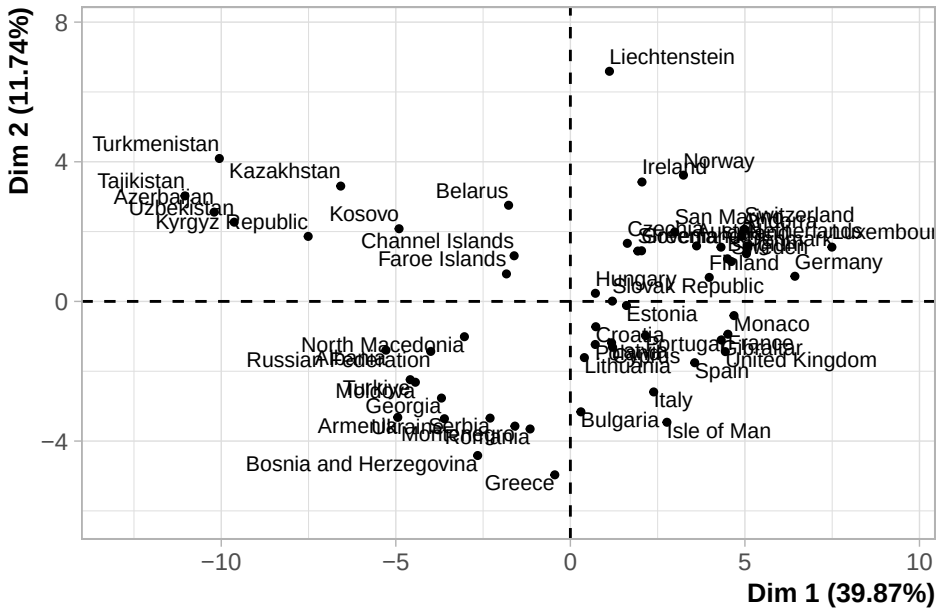

Figure 3.1 - Individuals factor map (PCA) *The labeled individuals are those with the higher contribution to the plane construction.*

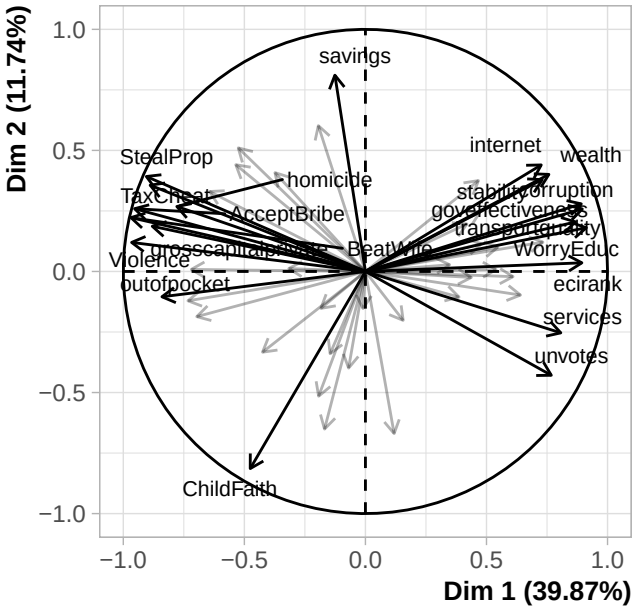

Figure 3.2 - Variables factor map (PCA) *The labeled variables are those the best shown on the plane.*

The **dimension 1** opposes individuals such as *Gibraltar*, *Netherlands*, *Denmark*, *Austria*, *Belgium*, *Switzerland*, *Andorra*, *Monaco*, *Isle of Man* and *Finland* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *Turkmenistan*, *Kyrgyz Republic*, *Uzbekistan*, *Tajikistan*, *Azerbaijan*, *Kosovo*, *Kazakhstan*, *Armenia*, *Channel Islands* and *Georgia* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *Gibraltar*, *Netherlands*, *Denmark*, *Austria*, *Belgium*, *Switzerland*, *Andorra*, *Monaco*, *Isle of Man* and *Finland* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for variables like *WorryEduc*, *transportquality*, *goveffectiveness*, *corruption*, *ecirank*, *stability*, *femaleminister*, *internet*, *femaleparliament* and *services* (variables are sorted from the strongest).
- low values for variables like *Violence*, *BeatWife*, *grosscapitalprivate*, *AcceptBribe*, *outofpocket*, *TaxCheat*, *StealProp*, *prison*, *homicide* and *agriculture* (variables are sorted from the weakest).

The group in which the individuals *Turkmenistan*, *Kyrgyz Republic*, *Uzbekistan*, *Tajikistan*, *Azerbaijan* and *Kazakhstan* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for variables like *StealProp*, *TaxCheat*, *AcceptBribe*, *BeatWife*, *Violence*, *homicide*, *resourcerents*, *prison*, *grosscapitalprivate* and *industry* (variables are sorted from the strongest).
- low values for variables like *unvotes*, *ecirank*, *corruption*, *WorryEduc*, *goveffectiveness*, *femaleminister*, *services*, *transportquality*, *wageworker* and *cleanfuel* (variables are sorted from the weakest).

The group in which the individuals *Kosovo*, *Armenia*, *Channel Islands*, *Georgia* and *Ukraine* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *ChildFaith*, *unemployed*, *outofschool* and *outofpocket* (variables are sorted from the strongest).
- low values for the variables *stability*, *femaleparliament*, *transportquality*, *wealth*, *WorryEduc* and *goveffectiveness* (variables are sorted from the weakest).

Note that the variables *AcceptBribe*, *Violence* and *BeatWife* are highly correlated with this dimension (respective correlation of 0.91, 0.93, 0.94). These variables could therefore summarize themselve the dimension 1.

The **dimension 2** opposes individuals such as *Turkmenistan*, *Kyrgyz Republic*, *Uzbekistan*, *Tajikistan*, *Azerbaijan* and *Kazakhstan* (to the top of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *Kosovo*, *Armenia*, *Channel Islands*, *Georgia* and *Ukraine* (to the bottom of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *Turkmenistan*, *Kyrgyz Republic*, *Uzbekistan*, *Tajikistan*, *Azerbaijan* and *Kazakhstan* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for variables like *StealProp*, *TaxCheat*, *AcceptBribe*, *BeatWife*, *Violence*, *homicide*, *resourcerents*, *prison*, *grosscapitalprivate* and *industry* (variables are sorted from the strongest).
- low values for variables like *unvotes*, *ecirank*, *corruption*, *WorryEduc*, *goveffectiveness*, *femaleminister*, *services*, *transportquality*, *wageworker* and *cleanfuel* (variables are sorted from the weakest).

The group in which the individuals *Kosovo*, *Armenia*, *Channel Islands*, *Georgia* and *Ukraine* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *ChildFaith*, *unemployed*, *outofschool* and *outofpocket* (variables are sorted from the strongest).
- low values for the variables *stability*, *femaleparliament*, *transportquality*, *wealth*, *WorryEduc* and *goveffectiveness* (variables are sorted from the weakest).

---

4. Description of the plane 3:4

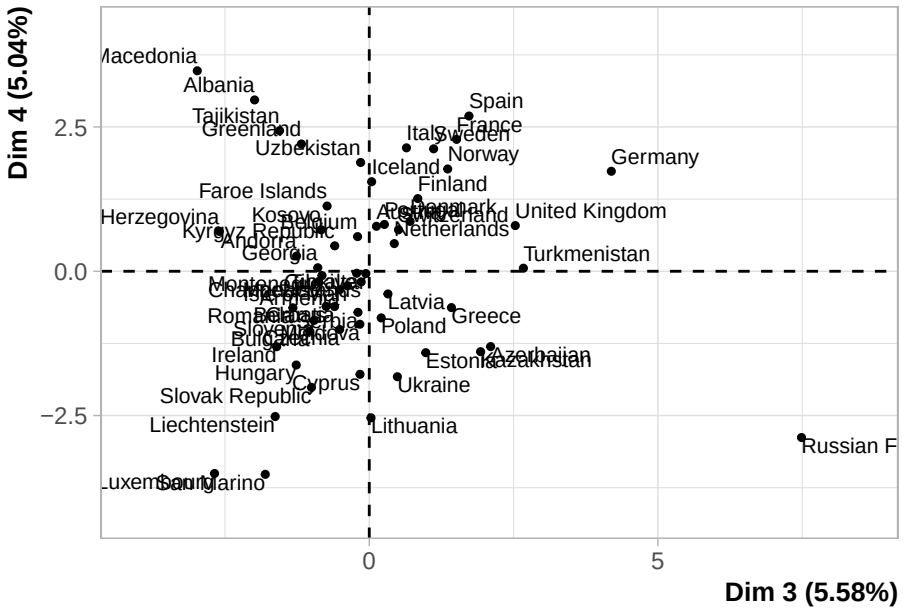

Figure 4.1 - Individuals factor map (PCA) *The labeled individuals are those with the higher contribution to the plane construction.*

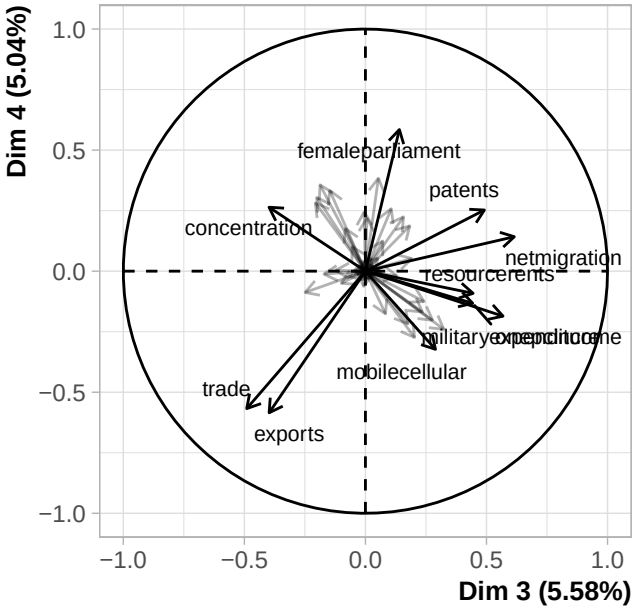

Figure 4.2 - Variables factor map (PCA) *The labeled variables are those the best shown on the plane.*

The **dimension 3** opposes individuals such as *Russian Federation*, *Spain*, *Germany*, *France*, *United Kingdom*, *Sweden*, *Norway* and *Italy* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *San Marino*, *Slovak Republic*, *Lithuania*, *North Macedonia*, *Hungary*, *Albania*, *Luxembourg*, *Cyprus*, *Ireland* and *Estonia* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *Spain*, *Germany*, *France*, *United Kingdom*, *Sweden*, *Norway* and *Italy* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for the variables *transportquality*, *femaleparliament*, *patents*, *femaleminister*, *WorryEduc*, *wealthincome*, *netmigration*, *corruption* and *goveffectiveness* (variables are sorted from the strongest).
- low values for the variables *StealProp*, *TaxCheat*, *AcceptBribe*, *BeatWife*, *grosscapitalprivate*, *Violence*, *trade*, *exports*, *outofpocket* and *homicide* (variables are sorted from the weakest).

The group in which the individual *Russian Federation* stands (characterized by a positive coordinate on the axis) is sharing :

- high values for the variables *onepcincome*, *resourcerents*, *militaryexpenditure*, *prison*, *homicide*, *femaleincome*, *mobilecellular* and *netmigration* (variables are sorted from the strongest).
- low values for the variable *unvotes*.

The group in which the individuals *North Macedonia*, *Albania*, *Greenland*, *Faroe Islands* and *Bosnia and Herzegovina* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for variables like *agriculture*, *concentration*, *selfemployed*, *Violence*, *BeatWife*, *AcceptBribe*, *TaxCheat*, *grosscapitalprivate*, *unemployed* and *StealProp* (variables are sorted from the strongest).
- low values for variables like *cleanfuel*, *wageworker*, *goveffectiveness*, *WorryEduc*, *mobilecellular*, *transportquality*, *internet*, *services*, *urbanlevel* and *ecirank* (variables are sorted from the weakest).

The group in which the individuals *San Marino*, *Slovak Republic*, *Lithuania*, *Hungary*, *Luxembourg*, *Cyprus*, *Ireland* and *Estonia* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *trade* and *exports* (variables are sorted from the strongest).
- low values for the variable *femaleparliament*.

The **dimension 4** opposes individuals such as *North Macedonia*, *Spain*, *Germany*, *France*, *Albania*, *United Kingdom*, *Sweden*, *Norway*, *Greenland* and *Faroe Islands* (to the top of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *San Marino*, *Slovak Republic*, *Lithuania*, *Hungary*, *Luxembourg*, *Cyprus*, *Ireland* and *Estonia* (to the bottom of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *North Macedonia*, *Albania*, *Greenland*, *Faroe Islands* and *Bosnia and Herzegovina* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for variables like *agriculture*, *concentration*, *selfemployed*, *Violence*, *BeatWife*, *AcceptBribe*, *TaxCheat*, *grosscapitalprivate*, *unemployed* and *StealProp* (variables are sorted from the strongest).
- low values for variables like *cleanfuel*, *wageworker*, *goveffectiveness*, *WorryEduc*, *mobilecellular*, *transportquality*, *internet*, *services*, *urbanlevel* and *ecirank* (variables are sorted from the weakest).

The group in which the individuals *Spain*, *Germany*, *France*, *United Kingdom*, *Sweden*, *Norway* and *Italy* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for the variables *transportquality*, *femaleparliament*, *patents*, *femaleminister*, *WorryEduc*, *wealthincome*, *netmigration*, *corruption* and *goveffectiveness* (variables are sorted from the strongest).
- low values for the variables *StealProp*, *TaxCheat*, *AcceptBribe*, *BeatWife*, *grosscapitalprivate*, *Violence*, *trade*, *exports*, *outofpocket* and *homicide* (variables are sorted from the weakest).

The group in which the individuals *San Marino*, *Slovak Republic*, *Lithuania*, *Hungary*, *Luxembourg*, *Cyprus*, *Ireland* and *Estonia* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *trade* and *exports* (variables are sorted from the strongest).
- low values for the variable *femaleparliament*.

---

5. Classification

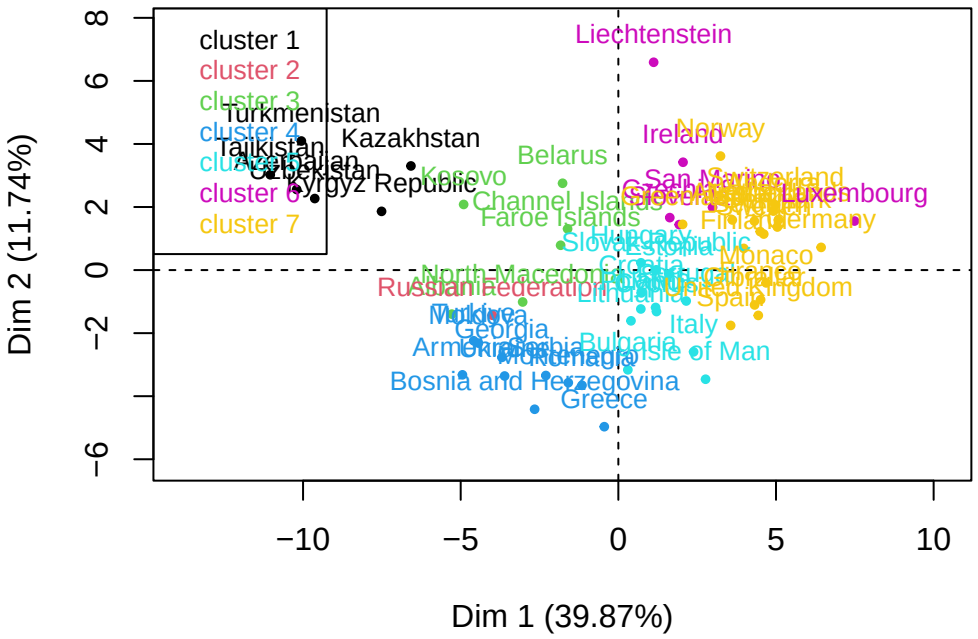

Figure 5 - Ascending Hierarchical Classification of the individuals. *The classification made on individuals reveals 7 clusters.*

The **cluster 1** is made of individuals such as *Azerbaijan*, *Kazakhstan*, *Kyrgyz Republic*, *Tajikistan*, *Turkmenistan* and *Uzbekistan*. This group is characterized by :

- high values for variables like *TaxCheat*, *StealProp*, *resourcerents*, *AcceptBribe*, *BeatWife*, *Violence*, *homicide*, *grosscapitalprivate*, *outofpocket* and *selfemployed* (variables are sorted from the strongest).
- low values for variables like *unvotes*, *ecirank*, *services*, *wageworker*, *WorryEduc*, *goveffectiveness*, *corruption*, *femaleminister*, *transportquality* and *govexp* (variables are sorted from the weakest).

The **cluster 2** is made of individuals such as *Russian Federation*. This group is characterized by :

- high values for the variables *onepcincome*, *prison*, *netmigration*, *femaleincome*, *militaryexpenditure*, *mobilecellular*, *resourcerents* and *homicide* (variables are sorted from the strongest).

The **cluster 3** is made of individuals such as *Albania*, *Belarus*, *Channel Islands*, *Faroe Islands*, *North Macedonia* and *Kosovo*. This group is characterized by :

- high values for the variables *concentration*, *agriculture*, *grosscapital*, *grosscapitalprivate* and *onepcwealth* (variables are sorted from the strongest).
- low values for the variables *services* and *transportquality* (variables are sorted from the weakest).

The **cluster 4** is made of individuals such as *Armenia*, *Bosnia and Herzegovina*, *Georgia*, *Greece*, *Moldova*, *Montenegro*, *Romania*, *Serbia*, *Turkiye* and *Ukraine*. This group is characterized by :

- high values for the variables *ChildFaith*, *unemployed*, *outofschool*, *femaleincome*, *gini*, *selfemployed*, *militaryexpenditure* and *outofpocket* (variables are sorted from the strongest).
- low values for variables like *savings*, *stability*, *internet*, *cleanfuel*, *wealth*, *goveffectiveness*, *WorryEduc*, *transportquality*, *urbanrate* and *wealthincome* (variables are sorted from the weakest).

The **cluster 5** is made of individuals such as *Bulgaria*, *Cyprus*, *Estonia*, *Croatia*, *Hungary*, *Isle of Man*, *Italy*, *Lithuania*, *Latvia* and *Poland*. This group is characterized by :

- high values for the variable *unvotes*.
- low values for the variables *urbanrate*, *concentration* and *grosscapital* (variables are sorted from the weakest).

The **cluster 6** is made of individuals such as *Czechia*, *Ireland*, *Liechtenstein*, *Luxembourg*, *San Marino* and *Slovenia*. This group is characterized by :

- high values for the variables *exports*, *trade*, *stability*, *wealth*, *ecirank*, *goveffectiveness*, *corruption* and *industry* (variables are sorted from the strongest).
- low values for the variables *fdi*, *ChildFaith*, *militaryexpenditure*, *outofpocket* and *unemployed* (variables are sorted from the weakest).

The **cluster 7** is made of individuals such as *Andorra*, *Austria*, *Belgium*, *Switzerland*, *Germany*, *Denmark*, *Spain*, *Finland*, *France* and *United Kingdom*. This group is characterized by :

- high values for variables like *transportquality*, *femaleparliament*, *corruption*, *goveffectiveness*, *urbanlevel*, *WorryEduc*, *internet*, *wealth*, *femaleminister* and *govexp* (variables are sorted from the strongest).
- low values for variables like *Violence*, *BeatWife*, *AcceptBribe*, *prison*, *grosscapitalprivate*, *TaxCheat*, *outofpocket*, *StealProp*, *ChildFaith* and *homicide* (variables are sorted from the weakest).

Annexes