Principal Component Analysis

Dataset testDF_imputed$completeObs

This dataset contains 217 individuals and 66 variables.

---

1. Study of the outliers

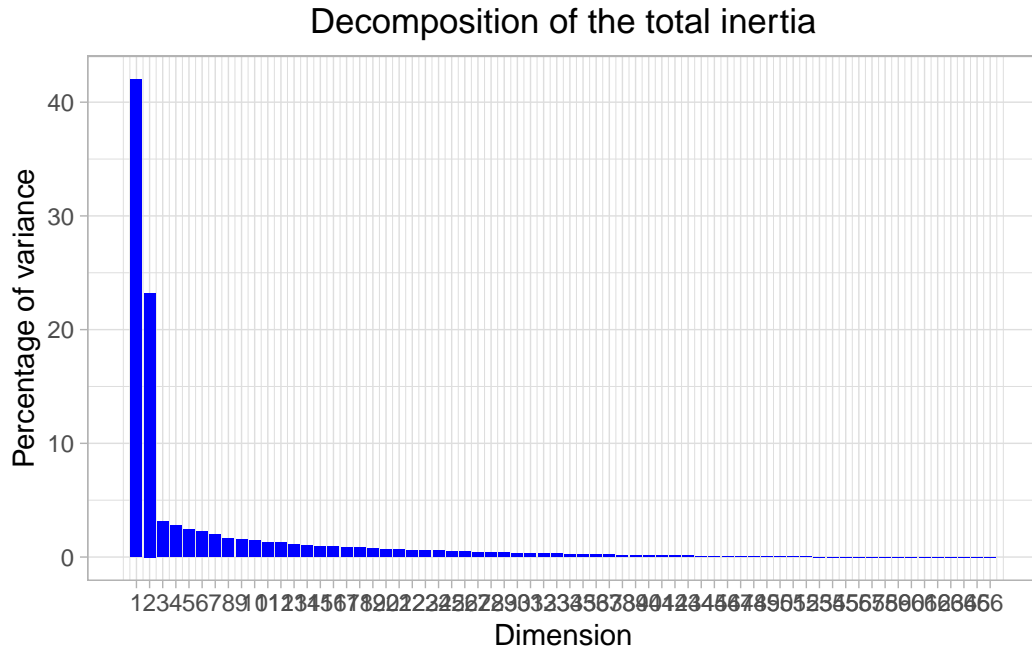The analysis of the graphs does not detect any outlier.

---

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 **65.21%** of the total dataset inertia ; that means that 65.21% 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 **7.04%**, 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 3612 data tables of equivalent size on the basis of a normal distribution).
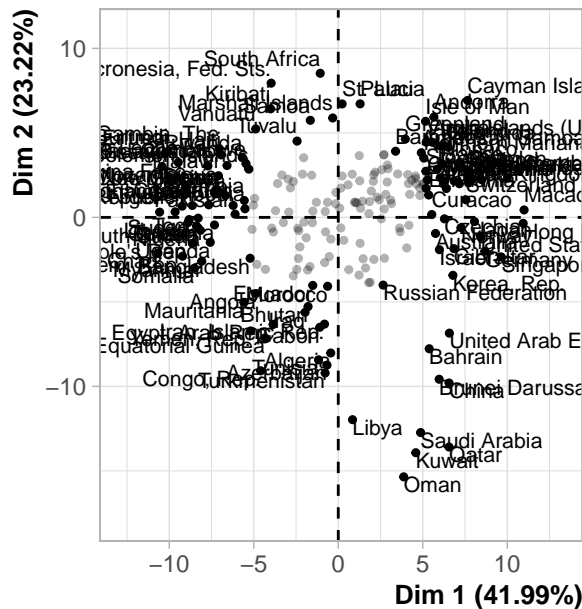
From these observations, it should be better to also interpret the dimensions greater or equal to the third one.
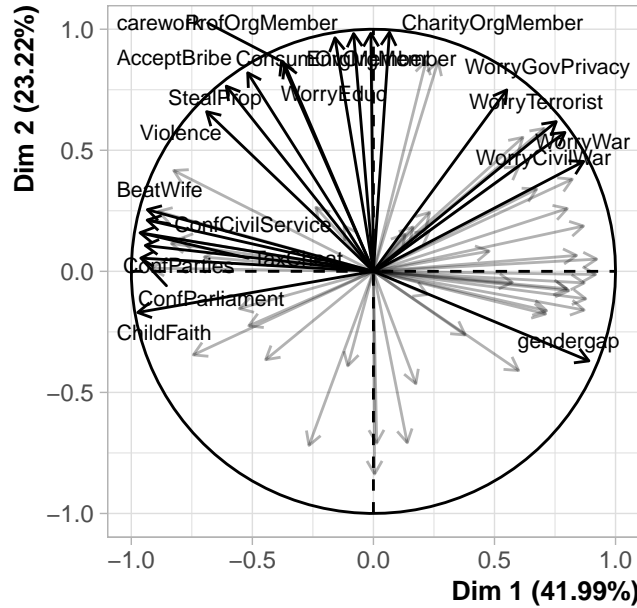
Decomposition of the total inertia

Figure 2 - Decomposition of the total inertia

An estimation of the right number of axis to interpret suggests to restrict the analysis to the description of the first 2 axis. These axis present an amount of inertia greater than those obtained by the 0.95-quantile of random distributions (65.21% against 7.04%). This observation suggests that only these axis are carrying a real information. As a consequence, the description will stand to these 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*, *Northern Mariana Islands*, *Isle of Man*, *San Marino*, *Monaco*, *Aruba*, *Virgin Islands (U.S.)*, *American Samoa*, *Netherlands* and *Denmark* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *Eritrea*, *Korea, Dem. People's Rep.*, *Congo, Dem. Rep.*, *Guinea*, *Niger*, *Afghanistan*, *Burkina Faso*, *Central African Republic*, *Guinea-Bissau* and *Myanmar* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *Gibraltar*, *Northern Mariana Islands*, *Isle of Man*, *San Marino*, *Monaco*, *Aruba*, *Virgin Islands (U.S.)*, *American Samoa*, *Netherlands* and *Denmark* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for variables like *WorryCivilWar*, *WorryWar*, *WorryTerrorist*, *BelieveGod*, *ecirank*, *goveffectiveness*, *WorryGovPrivacy*, *corruption*, *internet* and *unvotes* (variables are sorted from the strongest).
- low values for variables like *gini*, *ChildFaith*, *ConfParliament*, *ConfParties*, *homicide*, *selfemployed*, *TaxCheat*, *ConfCivilService*, *BeatWife* and *ConfGov* (variables are sorted from the weakest).

The group in which the individuals *Eritrea*, *Korea, Dem. People's Rep.*, *Congo, Dem. Rep.*, *Guinea*, *Niger*, *Afghanistan*, *Burkina Faso*, *Central African Republic*, *Guinea-Bissau* and *Myanmar* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for variables like *homicide*, *BeatWife*, *ConfCivilService*, *ConfGov*, *ConfCourts*, *ConfParties*, *ConfParliament*, *TaxCheat*, *selfemployed* and *agriculture* (variables are sorted from the strongest).
- low values for variables like *cleanfuel*, *gendergap*, *wageworker*, *electricity*, *divorce*, *internet*, *fossil*, *gnipercap*, *goveffectiveness* and *research* (variables are sorted from the weakest).

3

Note that the variables *ChildFaith*, *TaxCheat* and *ConfParliament* are highly correlated with this dimension (respective correlation of 0.95, 0.93, 0.92). These variables could therefore summarize themselve the dimension 1.

The **dimension 2** opposes individuals such as *Gibraltar*, *Northern Mariana Islands*, *Eritrea*, *Isle of Man*, *San Marino*, *Monaco*, *Aruba*, *Virgin Islands (U.S.)*, *American Samoa* and *Korea, Dem. People's Rep.* (to the top of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *Saudi Arabia*, *Qatar*, *Oman*, *Brunei Darussalam*, *Kuwait*, *Bahrain*, *Turkmenistan*, *Yemen, Rep.*, *Azerbaijan* and *Algeria* (to the bottom of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *Eritrea*, *Korea, Dem. People's Rep.*, *Congo, Dem. Rep.*, *Guinea*, *Niger*, *Afghanistan*, *Burkina Faso*, *Central African Republic*, *Guinea-Bissau* and *Myanmar* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for variables like *homicide*, *BeatWife*, *ConfCivilService*, *ConfGov*, *ConfCourts*, *ConfParties*, *ConfParliament*, *TaxCheat*, *selfemployed* and *agriculture* (variables are sorted from the strongest).
- low values for variables like *cleanfuel*, *gendergap*, *wageworker*, *electricity*, *divorce*, *internet*, *fossil*, *gnipercap*, *goveffectiveness* and *research* (variables are sorted from the weakest).

The group in which the individuals *Gibraltar*, *Northern Mariana Islands*, *Isle of Man*, *San Marino*, *Monaco*, *Aruba*, *Virgin Islands (U.S.)*, *American Samoa*, *Netherlands* and *Denmark* stand (characterized by a positive coordinate on the axis) is sharing :
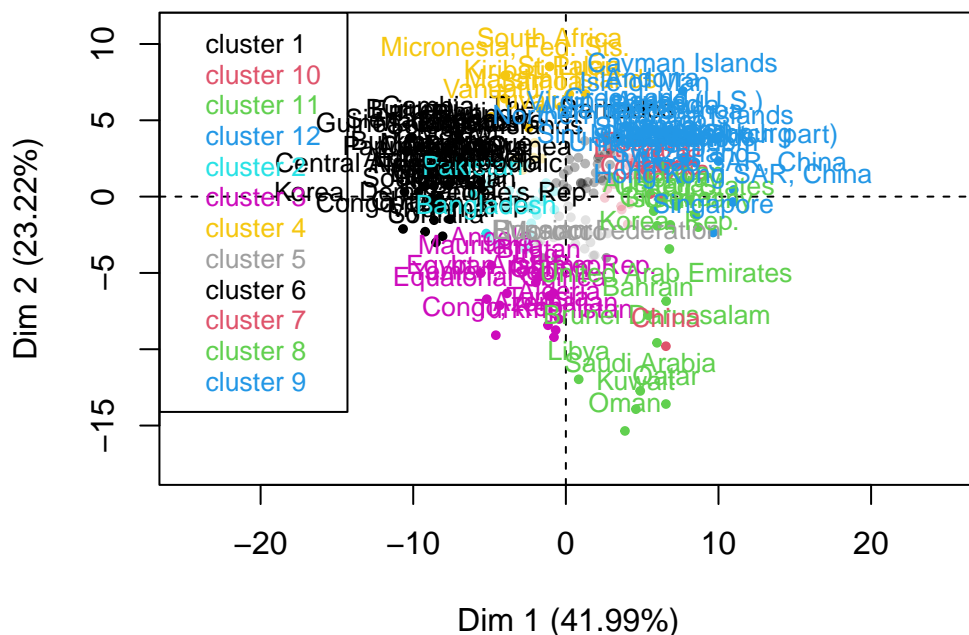
- high values for variables like *WorryCivilWar*, *WorryWar*, *WorryTerrorist*, *BelieveGod*, *ecirank*, *goveffectiveness*, *WorryGovPrivacy*, *corruption*, *internet* and *unvotes* (variables are sorted from the strongest).
- low values for variables like *gini*, *ChildFaith*, *ConfParliament*, *ConfParties*, *homicide*, *selfemployed*, *TaxCheat*, *ConfCivilService*, *BeatWife* and *ConfGov* (variables are sorted from the weakest).

The group in which the individuals *Saudi Arabia*, *Qatar*, *Oman*, *Brunei Darussalam*, *Kuwait*, *Bahrain*, *Turkmenistan*, *Yemen, Rep.*, *Azerbaijan* and *Algeria* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for variables like *industry*, *resourcerents*, *militaryexpenditure*, *savings*, *grosscapitalprivate*, *grosscapital*, *fossil*, *gini*, *urbanrate* and *gendergap* (variables are sorted from the strongest).
- low values for variables like *CharityOrgMember*, *EnvOrgMember*, *ProfOrgMember*, *ConsumOrgMember*, *femalemanager*, *SelfHelpOrgMember*, *carework*, *WorryEduc*, *AcceptBribe* and *WorryGovPrivacy* (variables are sorted from the weakest).

Note that the variables *EnvOrgMember*, *ConsumOrgMember*, *CharityOrgMember* and *ProfOrgMember* are highly correlated with this dimension (respective correlation of 0, 0.03, 0, 0.01). These variables could therefore summarize themselve the dimension 2.

---

4. Classification

Figure 4 - Ascending Hierarchical Classification of the individuals. *The classification made on individuals reveals 12 clusters.*

The **cluster 1** is made of individuals such as *Afghanistan*, *Burundi*, *Benin*, *Burkina Faso*, *Central African Republic*, *Cote d'Ivoire*, *Cameroon*, *Congo, Dem. Rep.*, *Comoros* and *Eritrea*. This group is characterized by :

- high values for variables like *agriculture*, *selfemployed*, *TaxCheat*, *BeatWife*, *ConfParliament*, *ConfCivilService*, *ConfParties*, *ChildFaith*, *ConfGov* and *ConfCourts* (variables are sorted from the strongest).
- low values for variables like *electricity*, *cleanfuel*, *wageworker*, *gendergap*, *fossil*, *divorce*, *internet*, *WorryCivilWar*, *research* and *mobilecellular* (variables are sorted from the weakest).

The **cluster 2** is made of individuals such as *Bangladesh* and *Pakistan*. This group is characterized by :

- high values for variables like *outofpocket*, *ChildFaith*, *TaxCheat*, *gini*, *ConfParliament*, *homicide*, *ConfParties*, *BeatWife*, *ConfCivilService* and *selfemployed* (variables are sorted from the strongest).
- low values for variables like *corruption*, *stability*, *netmigration*, *unvotes*, *gnipercap*, *internet*, *gdppercap*, *WorryCivilWar*, *WorryWar* and *WorryTerrorist* (variables are sorted from the weakest).

The **cluster 3** is made of individuals such as *Angola*, *Azerbaijan*, *Bhutan*, *Congo, Rep.*, *Algeria*, *Egypt, Arab Rep.*, *Gabon*, *Equatorial Guinea*, *Iran, Islamic Rep.* and *Iraq*. This group is characterized by :

- high values for variables like *industry*, *resourcerents*, *grosscapital*, *militaryexpenditure*, *ChildFaith*, *grosscapitalprivate*, *gini*, *outofpocket*, *savings* and *urbanrate* (variables are sorted from the strongest).
- low values for variables like *CharityOrgMember*, *EnvOrgMember*, *ProfOrgMember*, *SelfHelpOrgMember*, *ConsumOrgMember*, *WorryGovPrivacy*, *femalemanager*, *WorryTerrorist*, *BelieveGod* and *WorryWar* (variables are sorted from the weakest).

5

The **cluster 4** is made of individuals such as *Micronesia, Fed. Sts., Kiribati, St. Lucia, Marshall Islands, Palau, Tuvalu, Vanuatu, Samoa* and *South Africa*. This group is characterized by :

- high values for variables like *ConsumOrgMember, AcceptBribe, ProfOrgMember, StealProp, EnvOrgMember, Violence, concentration, CharityOrgMember, carework* and *WorryEduc* (variables are sorted from the strongest).
- low values for variables like *outofpocket, gendergap, grosscapitalprivate, industry, mobilecellular, femaleparliament, militaryexpenditure, gdppercap, fossil* and *internet* (variables are sorted from the weakest).

The **cluster 5** is made of individuals such as *Ecuador, Morocco* and *Russian Federation*. This group is characterized by :

- high values for the variables *fossil, mobilecellular, cleanfuel, electricity* and *industry* (variables are sorted from the strongest).
- low values for variables like *femalemanager, EnvOrgMember, ConsumOrgMember, CharityOrgMember, ProfOrgMember, AcceptBribe, SelfHelpOrgMember, StealProp, WorryEduc* and *Violence* (variables are sorted from the weakest).

The **cluster 6** is made of individuals such as *Barbados*. This group is characterized by :

- high values for variables like *unemployed, WorryGovPrivacy, electricity, WorryTerrorist, WorryWar, femalemanager, WorryCivilWar, BelieveGod, CharityOrgMember* and *wageworker* (variables are sorted from the strongest).
- low values for variables like *urbanrate, resourcerents, selfemployed, industry, grosscapital, ChildFaith, agriculture, savings, gini* and *ethnic* (variables are sorted from the weakest).

The **cluster 7** is made of individuals such as *Curacao, Czechia, Estonia* and *Lithuania*. This group is characterized by :

- high values for variables like *divorce, ecirank, WorryCivilWar, unvotes, WorryWar, wageworker, internet, WorryTerrorist, cleanfuel* and *BelieveGod* (variables are sorted from the strongest).
- low values for variables like *ChildFaith, homicide, ConfParliament, selfemployed, ConfParties, ConfCivilService, TaxCheat, BeatWife, gini* and *ConfGov* (variables are sorted from the weakest).

The **cluster 8** is made of individuals such as *United Arab Emirates, Bahrain, Brunei Darussalam, Kuwait, Libya, Oman, Qatar* and *Saudi Arabia*. This group is characterized by :

- high values for variables like *resourcerents, militaryexpenditure, industry, gendergap, grosscapitalprivate, gdppercap, gnipercap, fossil, savings* and *mobilecellular* (variables are sorted from the strongest).
- low values for variables like *carework, ConsumOrgMember, ProfOrgMember, EnvOrgMember, AcceptBribe, WorryEduc, CharityOrgMember, StealProp, Violence* and *femalemanager* (variables are sorted from the weakest).

The **cluster 9** is made of individuals such as *Aruba, Andorra, American Samoa, Austria, Belgium, Bermuda, Canada, Switzerland, Cayman Islands* and *Denmark*. This group is characterized by :

- high values for variables like *corruption, goveffectiveness, WorryWar, WorryCivilWar, WorryTerrorist, research, BelieveGod, ecirank, internet* and *WorryGovPrivacy* (variables are sorted from the strongest).
- low values for variables like *gini, ChildFaith, ConfParliament, ConfParties, TaxCheat, selfemployed, ConfCivilService, BeatWife, homicide* and *ConfGov* (variables are sorted from the weakest).

The **cluster 10** is made of individuals such as *China*. This group is characterized by :

- high values for the variables *patents*, *grosscapitalprivate*, *grosscapital* and *gendergap* (variables are sorted from the strongest).
- low values for variables like *ConfGov*, *ConfParties*, *ConfParliament*, *ConfCivilService*, *ConfCourts*, *WorryEduc*, *WorryGovPrivacy*, *carework*, *CharityOrgMember* and *ProfOrgMember* (variables are sorted from the weakest).

The **cluster 11** is made of individuals such as *Australia*, *Germany*, *Gibraltar*, *Israel*, *Japan*, *Korea, Rep.*, *Norway* and *United States*. This group is characterized by :

- high values for variables like *research*, *patents*, *unvotes*, *gendergap*, *goveffectiveness*, *corruption*, *gnipercap*, *internet*, *WorryCivilWar* and *netmigration* (variables are sorted from the strongest).
- low values for variables like *TaxCheat*, *BeatWife*, *ChildFaith*, *Violence*, *ConfCivilService*, *ConfCourts*, *ConfParliament*, *StealProp*, *ConfParties* and *ConfGov* (variables are sorted from the weakest).

The **cluster 12** is made of individuals such as *Hong Kong SAR, China*, *Luxembourg*, *Macao SAR, China* and *Singapore*. This group is characterized by :

- high values for variables like *exports*, *trade*, *gdppercap*, *gnipercap*, *mobilecellular*, *services*, *ecirank*, *gendergap*, *WorryCivilWar* and *goveffectiveness* (variables are sorted from the strongest).
- low values for variables like *ChildFaith*, *TaxCheat*, *ConfParliament*, *ConfParties*, *ConfCivilService*, *BeatWife*, *ConfGov*, *ConfCourts*, *Violence* and *gini* (variables are sorted from the weakest).

Annexes