Principal Component Analysis

Dataset testDF_imputed$completeObs

This dataset contains 217 individuals and 43 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 **48.64%** of the total dataset inertia ; that means that 48.64% of the individuals (or variables) cloud total variability is explained by the plane. This is an intermediate percentage and the first plane represents a part of the data variability. This value is strongly greater than the reference value that equals **9.3%**, 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 2045 data tables of equivalent size on the basis of a normal distribution).

From these observations, it may be interesting to consider the next dimensions which also express a high percentage of the total inertia.

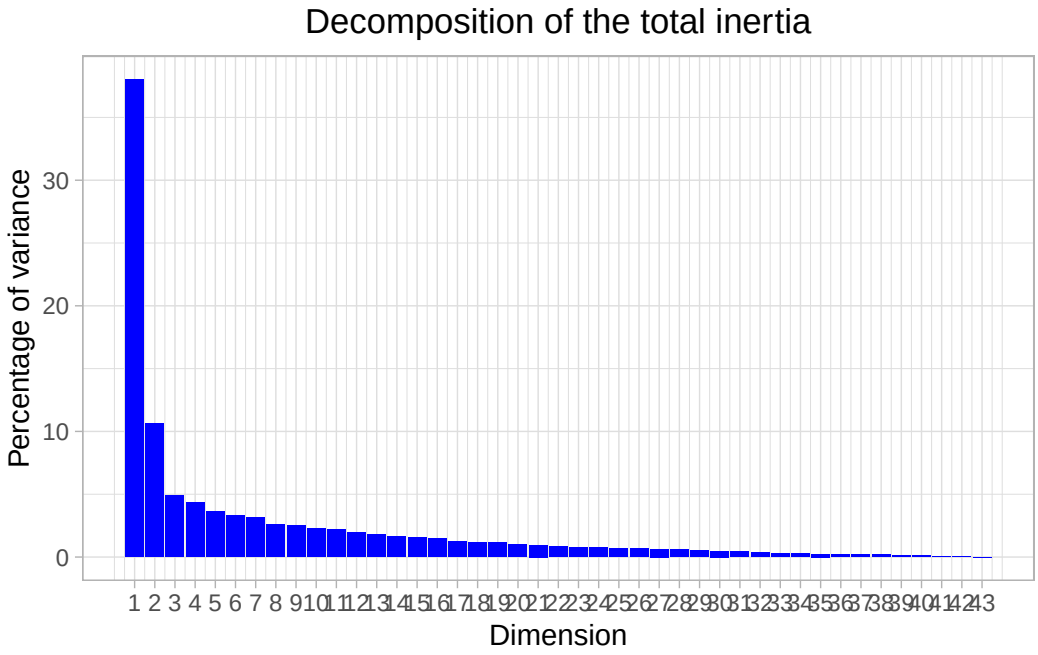

Figure 2 - Decomposition of the total inertia

An estimation of the right number of axis to interpret suggests to restrict the analysis to the description of the first 4 axis. These axis present an amount of inertia greater than those obtained by the 0.95-quantile of random distributions (57.89% against 17.56%). 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

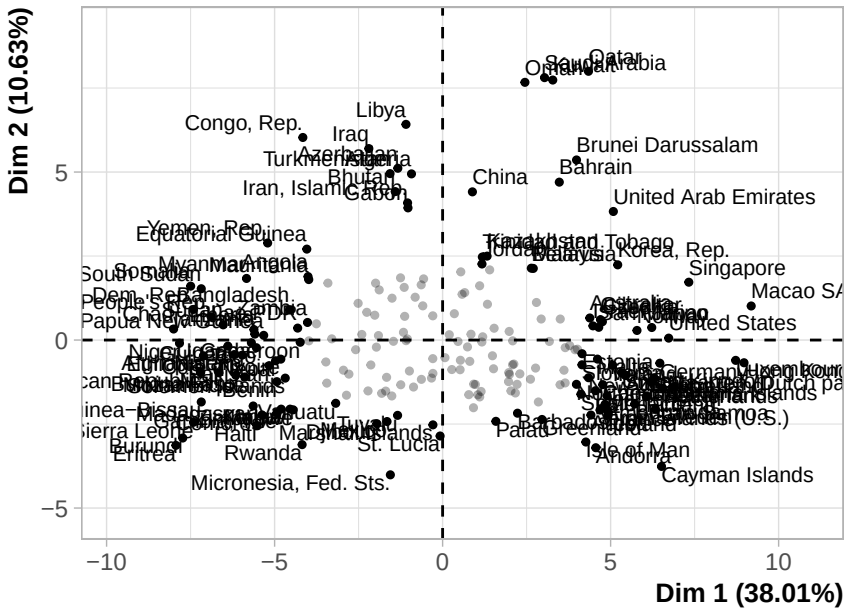

Figure 3.1 - Individuals factor map (PCA) *The labeled individuals are those with the higher contribution to the plane construction.*

```
## Warning: ggrepel: 7 unlabeled data points (too many overlaps). Consider increasing max.overlaps
```

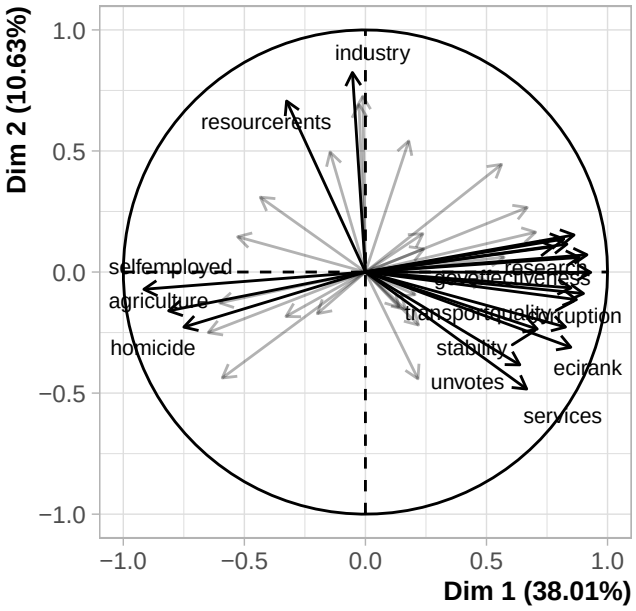

Figure 3.2 - Variables factor map (PCA) *The labeled variables are those the best shown on the plane.*

The **dimension 1** opposes individuals such as *Northern Mariana Islands*, *Gibraltar*, *American Samoa*, *Isle of Man*, *Monaco*, *Aruba*, *Virgin Islands (U.S.)*, *Sint Maarten (Dutch part)*, *Belgium* and *Netherlands* (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.*, *Somalia*, *Guinea-Bissau*, *Congo, Dem. Rep.*, *Central African Republic*, *Guinea*, *Chad*, *Comoros* and *Madagascar* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

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

- high values for variables like *ecirank*, *goveffectiveness*, *internet*, *transportquality*, *corruption*, *unvotes*, *research*, *wageworker*, *cleanfuel* and *services* (variables are sorted from the strongest).
- low values for variables like *selfemployed*, *homicide*, *urbanrate*, *agriculture*, *gini*, *outofschool*, *outofpocket*, *resourcerents*, *carework* and *concentration* (variables are sorted from the weakest).

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

- high values for the variables *homicide*, *selfemployed*, *agriculture*, *gini*, *carework*, *outofschool*, *urbanrate*, *outofpocket*, *concentration* and *ideology* (variables are sorted from the strongest).

- low values for variables like *cleanfuel*, *wageworker*, *internet*, *electricity*, *goveffectiveness*, *divorce*, *transportquality*, *gnipercap*, *research* and *urbanlevel* (variables are sorted from the weakest).

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

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

- high values for variables like *industry*, *resourcerents*, *militaryexpenditure*, *grosscapitalprivate*, *savings*, *fossil*, *grosscapital*, *cleanfuel*, *urbanrate* and *mobilecellular* (variables are sorted from the strongest).
- low values for the variables *carework*, *unvotes*, *femaleminister*, *services*, *femaleparliament*, *homicide*, *ecirank*, *agriculture*, *gini* and *selfemployed* (variables are sorted from the weakest).

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

- high values for the variables *homicide*, *selfemployed*, *agriculture*, *gini*, *carework*, *outofschool*, *urbanrate*, *outofpocket*, *concentration* and *ideology* (variables are sorted from the strongest).
- low values for variables like *cleanfuel*, *wageworker*, *internet*, *electricity*, *goveffectiveness*, *divorce*, *transportquality*, *gnipercap*, *research* and *urbanlevel* (variables are sorted from the weakest).

The group in which the individuals *Northern Mariana Islands*, *Gibraltar*, *American Samoa*, *Isle of Man*, *Monaco*, *Aruba*, *Virgin Islands (U.S.)*, *Sint Maarten (Dutch part)*, *Belgium* and *Netherlands* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for variables like *ecirank*, *goveffectiveness*, *internet*, *transportquality*, *corruption*, *unvotes*, *research*, *wageworker*, *cleanfuel* and *services* (variables are sorted from the strongest).
- low values for variables like *selfemployed*, *homicide*, *urbanrate*, *agriculture*, *gini*, *outofschool*, *outofpocket*, *resourcerents*, *carework* and *concentration* (variables are sorted from the weakest).

---

4. Description of the plane 3:4

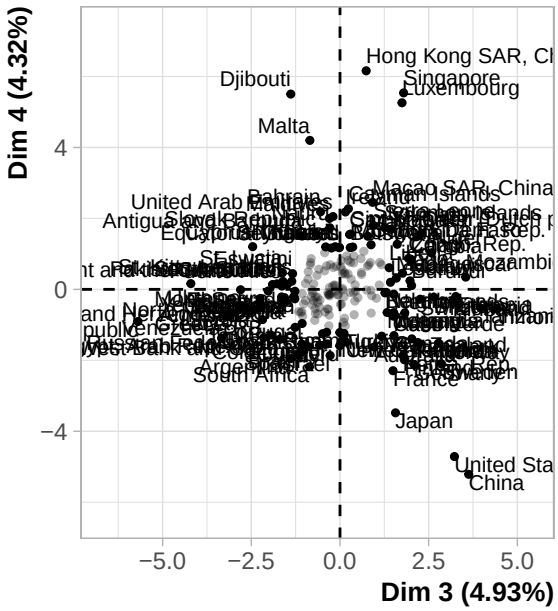

Figure 4.1 - Individuals factor map (PCA) *The labeled individuals are those with the higher contribution to the plane construction.*

```
## Warning: ggrepel: 1 unlabeled data points (too many overlaps). Consider increasing max.overlaps
```

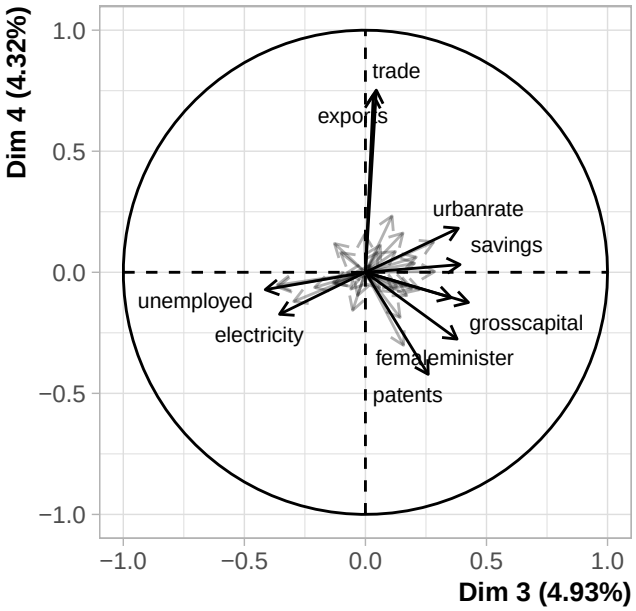

Figure 4.2 - Variables factor map (PCA) *The labeled variables are those the best shown on the plane.*

The **dimension 3** opposes individuals such as *Tanzania*, *China*, *United States*, *Mozambique*, *Sweden*, *Norway*, *France*, *Cabo Verde*, *Timor-Leste* and *Japan* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *Egypt, Arab Rep.*, *Tunisia*, *Serbia*, *Greece*, *Armenia*, *Argentina*, *Syrian Arab Republic*, *Bosnia and Herzegovina*, *Ukraine* and *West Bank and Gaza* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *Tanzania*, *China*, *United States*, *Mozambique*, *Sweden*, *Norway*, *France*, *Cabo Verde*, *Timor-Leste* and *Japan* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for variables like *grosscapital*, *femaleminister*, *savings*, *patents*, *corruption*, *urbanrate*, *grosscapitalprivate*, *selfemployed*, *govexp* and *research* (variables are sorted from the strongest).
- low values for the variables *electricity*, *fossil*, *unemployed*, *outofpocket*, *cleanfuel*, *mobilecellular*, *wageworker*, *exports* and *trade* (variables are sorted from the weakest).

The group in which the individuals *Egypt, Arab Rep.*, *Tunisia*, *Serbia*, *Greece*, *Armenia*, *Argentina*, *Syrian Arab Republic*, *Bosnia and Herzegovina*, *Ukraine* and *West Bank and Gaza* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *outofpocket*, *unemployed*, *electricity*, *carework*, *fossil*, *gini* and *cleanfuel* (variables are sorted from the strongest).
- low values for variables like *corruption*, *savings*, *urbanrate*, *grosscapital*, *stability*, *trade*, *transportquality*, *exports*, *grosscapitalprivate* and *goveffectiveness* (variables are sorted from the weakest).

The **dimension 4** opposes individuals such as *Djibouti*, *Malta*, *Singapore*, *Hong Kong SAR, China*, *Luxembourg*, *Maldives*, *Antigua and Barbuda*, *Vanuatu*, *Nauru* and *Slovak Republic* (to the top of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *Egypt, Arab Rep.*, *Tunisia*, *Serbia*, *Greece*, *Armenia*, *Argentina*, *Syrian Arab Republic*, *Bosnia and Herzegovina*, *Ukraine* and *West Bank and Gaza* (to the bottom of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *Djibouti*, *Malta*, *Singapore*, *Hong Kong SAR, China*, *Luxembourg*, *Maldives*, *Antigua and Barbuda*, *Vanuatu*, *Nauru* and *Slovak Republic* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for the variables *exports*, *trade*, *services*, *gdppercap*, *gnipercap*, *ecirank*, *stability* and *mobilecellular* (variables are sorted from the strongest).
- low values for the variables *femaleparliament*, *carework*, *gini*, *outofpocket* and *agriculture* (variables are sorted from the weakest).

The group in which the individuals *Egypt, Arab Rep.*, *Tunisia*, *Serbia*, *Greece*, *Armenia*, *Argentina*, *Syrian Arab Republic*, *Bosnia and Herzegovina*, *Ukraine* and *West Bank and Gaza* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *outofpocket*, *unemployed*, *electricity*, *carework*, *fossil*, *gini* and *cleanfuel* (variables are sorted from the strongest).
- low values for variables like *corruption*, *savings*, *urbanrate*, *grosscapital*, *stability*, *trade*, *transportquality*, *exports*, *grosscapitalprivate* and *goveffectiveness* (variables are sorted from the weakest).

The group in which the individuals *Tanzania*, *China*, *United States*, *Mozambique*, *Sweden*, *Norway*, *France*, *Cabo Verde*, *Timor-Leste* and *Japan* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for variables like *grosscapital*, *femaleminister*, *savings*, *patents*, *corruption*, *urbanrate*, *grosscapitalprivate*, *selfemployed*, *govexp* and *research* (variables are sorted from the strongest).
- low values for the variables *electricity*, *fossil*, *unemployed*, *outofpocket*, *cleanfuel*, *mobilecellular*, *wageworker*, *exports* and *trade* (variables are sorted from the weakest).

---

5. Classification

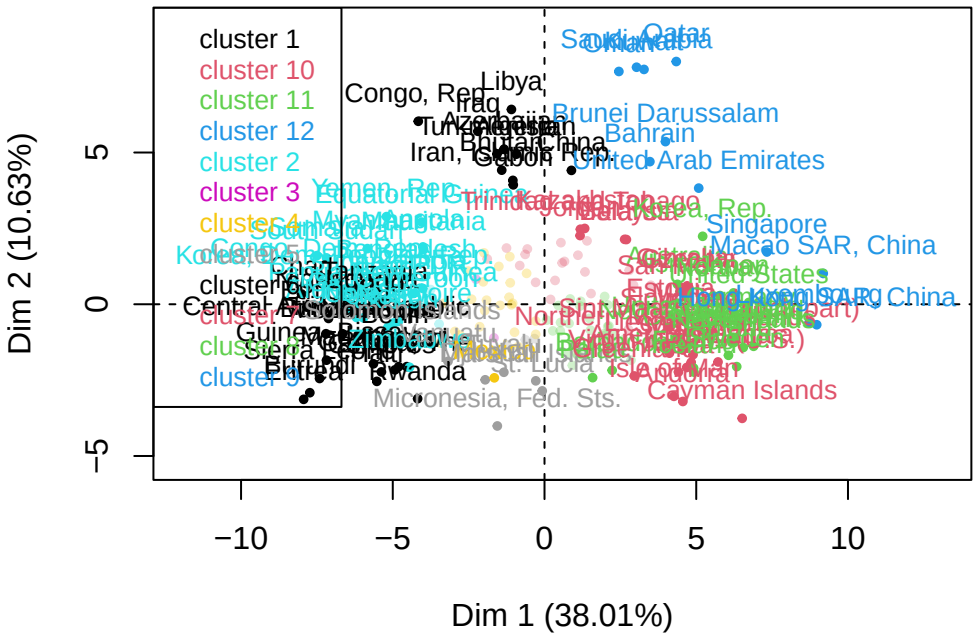

Figure 5 - Ascending Hierarchical Classification of the individuals. *The classification made on individuals reveals 12 clusters.*

The **cluster 1** is made of individuals such as *Burundi*, *Benin*, *Burkina Faso*, *Central African Republic*, *Comoros*, *Eritrea*, *Ethiopia*, *Gambia, The*, *Guinea-Bissau* and *Haiti*. This group is characterized by :

- high values for the variables *agriculture*, *selfemployed*, *outofschool*, *homicide*, *gini*, *carework*, *urbanrate*, *femaleminister*, *concentration* and *ideology* (variables are sorted from the strongest).
- low values for variables like *electricity*, *cleanfuel*, *wageworker*, *fossil*, *internet*, *divorce*, *research*, *mobilecellular*, *urbanlevel* and *goveffectiveness* (variables are sorted from the weakest).

The **cluster 2** is made of individuals such as *Afghanistan*, *Angola*, *Bangladesh*, *Cote d'Ivoire*, *Cameroon*, *Congo, Dem. Rep.*, *Guinea*, *Equatorial Guinea*, *Lao PDR* and *Liberia*. This group is characterized by :

- high values for the variables *selfemployed*, *outofpocket*, *homicide*, *outofschool*, *agriculture*, *urbanrate*, *gini*, *resourcerents*, *carework* and *industry* (variables are sorted from the strongest).
- low values for variables like *wageworker*, *cleanfuel*, *ecirank*, *internet*, *electricity*, *stability*, *goveffectiveness*, *research*, *corruption* and *transportquality* (variables are sorted from the weakest).

The **cluster 3** is made of individuals sharing :

- high values for the variables *govexp*, *savings* and *grosscapital* (variables are sorted from the strongest).
- low values for the variables *cleanfuel* and *outofpocket* (variables are sorted from the weakest).

The **cluster 4** is made of individuals such as *Mexico*. This group is characterized by :

- high values for the variables *gini*, *outofpocket*, *carework*, *homicide*, *electricity* and *ideology* (variables are sorted from the strongest).
- low values for variables like *corruption*, *unvotes*, *exports*, *stability*, *gnipercap*, *trade*, *gdppercap*, *transportquality*, *goveffectiveness* and *netmigration* (variables are sorted from the weakest).

The **cluster 5** is made of individuals such as *Djibouti*, *Micronesia, Fed. Sts.*, *St. Lucia*, *Marshall Islands*, *Solomon Islands*, *Tuvalu* and *Vanuatu*. This group is characterized by :

- high values for the variables *concentration*, *homicide*, *govexp* and *carework* (variables are sorted from the strongest).
- low values for variables like *femaleparliament*, *mobilecellular*, *outofpocket*, *urbanlevel*, *industry*, *cleanfuel*, *gdppercap*, *internet*, *grosscapitalprivate* and *gnipercap* (variables are sorted from the weakest).

The **cluster 6** is made of individuals such as *Azerbaijan*, *Bhutan*, *China*, *Congo, Rep.*, *Algeria*, *Gabon*, *Iran, Islamic Rep.*, *Iraq*, *Libya* and *Turkmenistan*. This group is characterized by :

- high values for the variables *grosscapitalprivate*, *industry*, *resourcerents*, *grosscapital*, *militaryexpenditure*, *savings*, *patents*, *fossil* and *outofpocket* (variables are sorted from the strongest).
- low values for the variables *services*, *ecirank*, *unvotes*, *femaleminister*, *corruption*, *stability*, *carework* and *transportquality* (variables are sorted from the weakest).

The **cluster 7** is made of individuals such as *Belarus*, *Jordan*, *Kazakhstan*, *Malaysia* and *Trinidad and Tobago*. This group is characterized by :

- high values for the variables *fossil*, *mobilecellular*, *grosscapitalprivate*, *grosscapital*, *urbanrate* and *netmigration* (variables are sorted from the strongest).
- low values for the variables *unvotes* and *femaleminister* (variables are sorted from the weakest).

The **cluster 8** is made of individuals such as *Barbados* and *Palau*. This group is characterized by :

- high values for the variables *unemployed*, *electricity*, *wageworker*, *cleanfuel*, *services*, *internet*, *stability*, *ecirank* and *unvotes* (variables are sorted from the strongest).
- low values for the variables *urbanrate*, *resourcerents*, *selfemployed*, *savings*, *agriculture*, *grosscapital*, *industry*, *grosscapitalprivate* and *homicide* (variables are sorted from the weakest).

The **cluster 9** is made of individuals such as *United Arab Emirates*, *Bahrain*, *Brunei Darussalam*, *Kuwait*, *Oman*, *Qatar* and *Saudi Arabia*. This group is characterized by :

- high values for variables like *resourcerents*, *industry*, *militaryexpenditure*, *gdppercap*, *savings*, *gnipercap*, *mobilecellular*, *fossil*, *wageworker* and *grosscapitalprivate* (variables are sorted from the strongest).
- low values for variables like *carework*, *selfemployed*, *gini*, *services*, *homicide*, *unvotes*, *femaleminister*, *outofpocket*, *femaleparliament* and *ideology* (variables are sorted from the weakest).

The **cluster 10** is made of individuals such as *Aruba*, *Andorra*, *American Samoa*, *Bermuda*, *Cayman Islands*, *Czechia*, *Spain*, *Estonia*, *Gibraltar* and *Greenland*. This group is characterized by :

- high values for variables like *ecirank*, *unvotes*, *internet*, *goveffectiveness*, *services*, *wageworker*, *stability*, *trade*, *exports* and *corruption* (variables are sorted from the strongest).
- low values for variables like *selfemployed*, *homicide*, *agriculture*, *gini*, *urbanrate*, *resourcerents*, *outofschool*, *outofpocket*, *industry* and *carework* (variables are sorted from the weakest).

The **cluster 11** is made of individuals such as *Australia*, *Austria*, *Belgium*, *Canada*, *Switzerland*, *Germany*, *Denmark*, *Finland*, *France* and *United Kingdom*. This group is characterized by :

- high values for variables like *research*, *transportquality*, *corruption*, *goveffectiveness*, *unvotes*, *gnipercap*, *internet*, *femaleminister*, *ecirank* and *gdppercap* (variables are sorted from the strongest).
- low values for variables like *gini*, *homicide*, *selfemployed*, *carework*, *agriculture*, *outofpocket*, *outofschool*, *resourcerents*, *ideology* and *urbanrate* (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*, *transportquality*, *ecirank*, *goveffectiveness* and *divorce* (variables are sorted from the strongest).
- low values for the variables *carework*, *gini*, *homicide*, *agriculture*, *selfemployed* and *industry* (variables are sorted from the weakest).

Annexes