Principal Component Analysis

Dataset testDF_imputed$completeObs

This dataset contains 209 individuals and 53 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 **8.2%**, 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 2114 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.
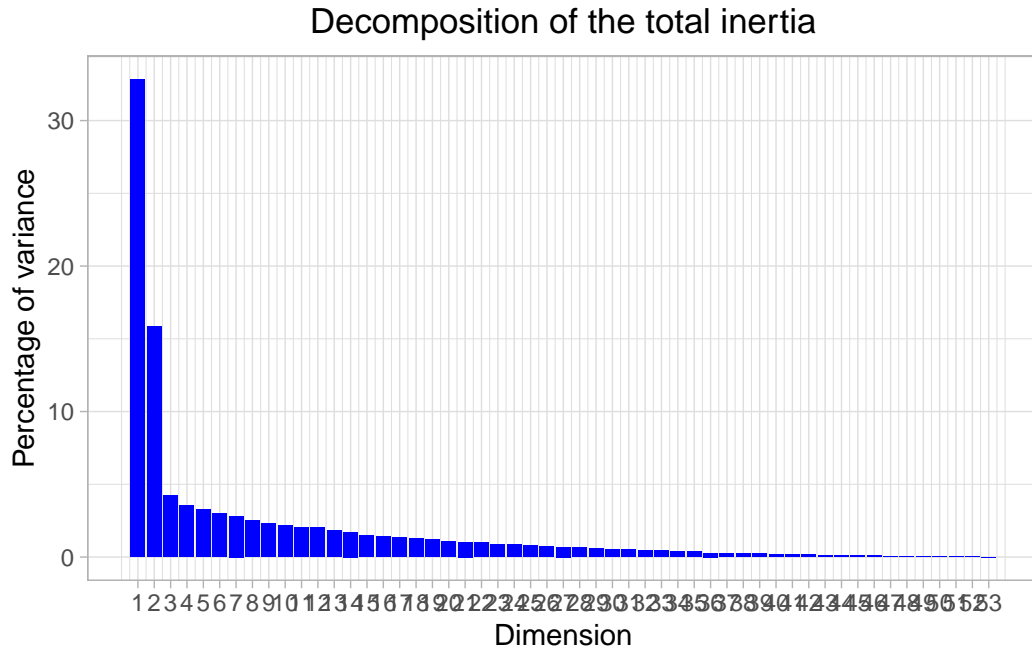
Decomposition of the total inertia

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 3 axis. These axis present an amount of inertia greater than those obtained by the 0.95-quantile of random distributions (52.89% against 11.95%). This observation suggests that only these axis are carrying a real information. As a consequence, the description will stand to these axis.

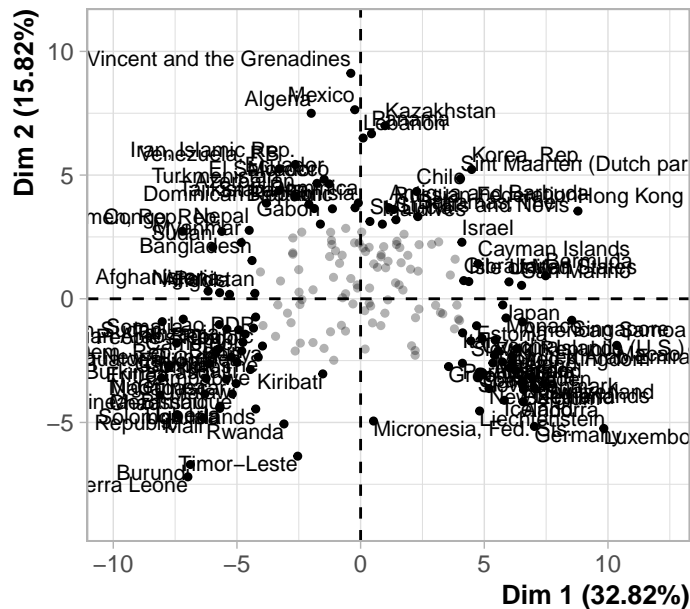
3. Description of the plane 1:2

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

Warning: ggrepel: 6 unlabeled data points (too many overlaps). Consider increasing max.overlaps
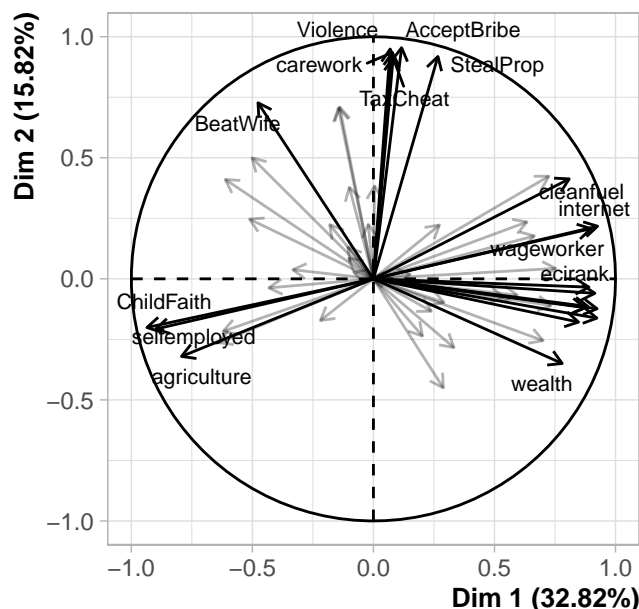

Figure 3.2 - Variables factor map (PCA) *The labeled variables are those the best shown on the plane.*

The **dimension 1** opposes individuals such as *Gibraltar*, *Netherlands*, *Denmark*, *American Samoa*, *Andorra*, *Austria*, *Belgium*, *Monaco*, *Iceland* and *Germany* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *Central African Republic*, *Guinea-Bissau*, *Korea, Dem. People's Rep.*, *Burundi*, *Congo, Dem. Rep.*, *Yemen, Rep.*, *Guinea*, *Eritrea*, *Morocco* and *Sierra Leone* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *Gibraltar*, *Netherlands*, *Denmark*, *American Samoa*, *Andorra*, *Austria*, *Belgium*, *Monaco*, *Iceland* and *Germany* stand (characterized by a positive coordinate on the axis) is sharing :

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

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

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

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

The group in which the individuals *Yemen, Rep.*, *Morocco*, *Panama*, *St. Vincent and the Grenadines*, *Algeria*, *Tajikistan*, *Iran, Islamic Rep.*, *Tunisia*, *Sudan* and *Sri Lanka* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for variables like *grosscapitalprivate*, *BeatWife*, *TaxCheat*, *prison*, *carework*, *AcceptBribe*, *Violence*, *homicide*, *gini* and *grosscapital* (variables are sorted from the strongest).
- low values for variables like *wealthincome*, *unvotes*, *wealth*, *femaleminister*, *corruption*, *WorryEduc*, *ecirank*, *gnipercap*, *transportquality* and *gdppercap* (variables are sorted from the weakest).

The **dimension 2** opposes individuals such as *Yemen, Rep.*, *Morocco*, *Panama*, *St. Vincent and the Grenadines*, *Algeria*, *Tajikistan*, *Iran, Islamic Rep.*, *Tunisia*, *Sudan* and *Sri Lanka* (to the top of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *Central African Republic*, *Guinea-Bissau*, *Korea, Dem. People's Rep.*, *Burundi*, *Congo, Dem. Rep.*, *Guinea*, *Eritrea*, *Sierra Leone*, *Niger* and *Somalia* (to the bottom of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *Yemen, Rep.*, *Morocco*, *Panama*, *St. Vincent and the Grenadines*, *Algeria*, *Tajikistan*, *Iran, Islamic Rep.*, *Tunisia*, *Sudan* and *Sri Lanka* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for variables like *grosscapitalprivate*, *BeatWife*, *TaxCheat*, *prison*, *carework*, *AcceptBribe*, *Violence*, *homicide*, *gini* and *grosscapital* (variables are sorted from the strongest).
- low values for variables like *wealthincome*, *unvotes*, *wealth*, *femaleminister*, *corruption*, *WorryEduc*, *ecirank*, *gnipercap*, *transportquality* and *gdppercap* (variables are sorted from the weakest).

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

- high values for variables like *selfemployed*, *ChildFaith*, *agriculture*, *urbanrate*, *outofschool*, *homicide*, *resourcerents*, *concentration*, *outofpocket* and *ideology* (variables are sorted from the strongest).
- low values for variables like *cleanfuel*, *wageworker*, *electricity*, *internet*, *goveffectiveness*, *WorryEduc*, *ecirank*, *gnipercap*, *transportquality* and *stability* (variables are sorted from the weakest).

Note that the variables *AcceptBribe* and *Violence* are highly correlated with this dimension (respective correlation of 0.01, 0). These variables could therefore summarize themselve the dimension 2.
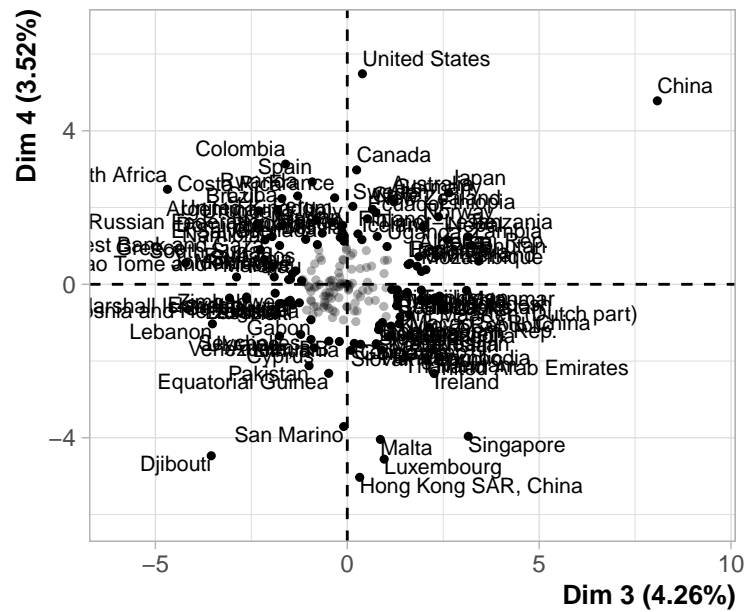
---

4. Description of the dimension 3

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
```
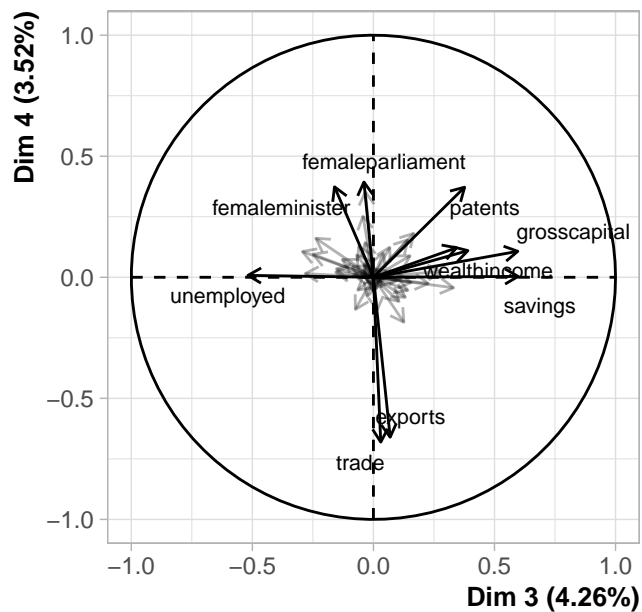

Figure 4.2 - Variables factor map (PCA) *The labeled variables are those the best shown on the plane.*

The **dimension 3** opposes individuals such as *China*, *Vietnam*, *Malta*, *Cambodia*, *Bhutan*, *Singapore*, *San Marino*, *Tanzania*, *Japan* and *Zambia* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *Djibouti*, *South Africa*, *Greece*, *Costa Rica*, *United States*, *Uruguay*, *Argentina*, *Colombia*, *Eswatini* and *Namibia* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *Vietnam*, *Malta*, *Cambodia*, *Singapore*, *San Marino*, *Ireland*, *Hong Kong SAR*, *China*, *Kyrgyz Republic*, *Thailand* and *Hungary* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for variables like *exports*, *trade*, *gdppercap*, *mobilecellular*, *gnipercap*, *wealth*, *savings*, *WorryEduc*, *stability* and *ecirank* (variables are sorted from the strongest).
- low values for the variables *gini, unemployed, femaleminister, grosscapitalprivate, homicide, ChildFaith* and *femaleparliament* (variables are sorted from the weakest).

The group in which the individuals *China*, *Bhutan*, *Tanzania*, *Japan*, *Zambia*, *Myanmar*, *Uzbekistan*, *Ethiopia*, *Nepal* and *India* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for variables like *grosscapital, savings, grosscapitalprivate, agriculture, selfemployed, onepcwealth, wealthincome, urbanrate, industry* and *outofpocket* (variables are sorted from the strongest).
- low values for the variables *services, trade, exports, wageworker, unemployed, urbanlevel, internet, govexp, cleanfuel* and *WorryEduc* (variables are sorted from the weakest).

The group in which the individuals *Djibouti*, *Greece*, *Eswatini*, *Namibia*, *West Bank and Gaza*, *Marshall Islands*, *Serbia*, *Ukraine*, *Belize* and *Jordan* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *unemployed, govexp, militaryexpenditure, outofschool* and *services* (variables are sorted from the strongest).
- low values for the variables *savings, wealthincome, grosscapital, industry, wealth, urbanrate* and *goveffectiveness* (variables are sorted from the weakest).

The group in which the individuals *South Africa*, *Costa Rica*, *United States*, *Uruguay*, *Argentina*, *Colombia*, *Cuba*, *Spain*, *Brazil* and *Canada* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *femaleminister, femaleparliament, netmigration, gini, prison, onepcincome* and *urbanlevel* (variables are sorted from the strongest).
- low values for the variables *savings, trade, exports, grosscapital* and *outofpocket* (variables are sorted from the weakest).
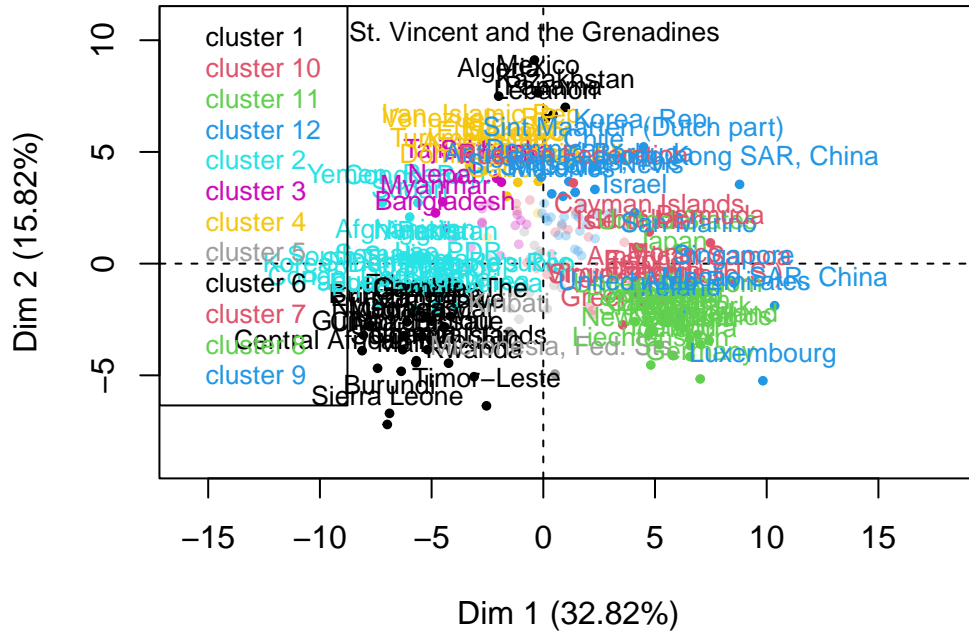
5. Classification

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, Burkina Faso, Central African Republic, Comoros, Eritrea, Ethiopia, Guinea, Gambia, The, Guinea-Bissau* and *Haiti.* This group is characterized by :

- high values for the variables *agriculture, selfemployed, ChildFaith, outofschool, urbanrate, resourcerents, ideology* and *concentration* (variables are sorted from the strongest).
- low values for variables like *electricity, cleanfuel, internet, wageworker, StealProp, AcceptBribe, carework, Violence, TaxCheat* and *urbanlevel* (variables are sorted from the weakest).

The **cluster 2** is made of individuals such as *Afghanistan, Angola, Benin, Cameroon, Congo, Dem. Rep., Congo, Rep., Equatorial Guinea, Kenya, Lao PDR* and *Nigeria.* This group is characterized by :

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

The **cluster 3** is made of individuals such as *Bangladesh, Bhutan, Sri Lanka, Myanmar, Nepal* and *Tajikistan.* This group is characterized by :

- high values for variables like *grosscapital, grosscapitalprivate, outofpocket, savings, BeatWife, industry, ChildFaith, selfemployed, gini* and *urbanrate* (variables are sorted from the strongest).
- low values for variables like *services, WorryEduc, urbanlevel, unvotes, gnipercap, gdppercap, govexp, wageworker, femaleminister* and *corruption* (variables are sorted from the weakest).

The **cluster 4** is made of individuals such as *Azerbaijan, Dominican Republic, Ecuador, Gabon, Iran, Islamic Rep., Morocco, El Salvador, Turkmenistan, Tunisia* and *Venezuela, RB.* This group is characterized by :

7

- high values for variables like *Violence*, *AcceptBribe*, *homicide*, *carework*, *TaxCheat*, *industry*, *BeatWife*, *StealProp*, *resourcerents* and *prison* (variables are sorted from the strongest).
- low values for the variables *unvotes*, *corruption*, *ecirank*, *stability*, *transportquality*, *wealthincome*, *net-migration*, *goveffectiveness*, *WorryEduc* and *wealth* (variables are sorted from the weakest).

The **cluster 5** is made of individuals such as *Micronesia, Fed. Sts.* and *Kiribati*. This group is characterized by :

- high values for the variables *govexp*, *unemployed*, *ideology* and *outofschool* (variables are sorted from the strongest).
- low values for the variables *outofpocket*, *gdppercap*, *urbanlevel*, *cleanfuel*, *mobilecellular*, *gnipercap*, *grosscapital* and *industry* (variables are sorted from the weakest).

The **cluster 6** is made of individuals such as *Algeria*, *Kazakhstan*, *Lebanon*, *Mexico*, *Panama* and *St. Vincent and the Grenadines*. This group is characterized by :

- high values for variables like *carework*, *AcceptBribe*, *TaxCheat*, *StealProp*, *Violence*, *BeatWife*, *prison*, *grosscapitalprivate*, *homicide* and *militaryexpenditure* (variables are sorted from the strongest).

The **cluster 7** is made of individuals such as *Russian Federation*. This group is characterized by :

- high values for the variables *femaleincome*, *onepcincome*, *unemployed*, *urbanlevel*, *cleanfuel*, *military-expenditure*, *TaxCheat*, *femaleminister* and *electricity* (variables are sorted from the strongest).
- low values for the variables *grosscapital*, *savings*, *trade*, *concentration* and *exports* (variables are sorted from the weakest).

The **cluster 8** is made of individuals sharing :

- high values for the variables *patents*, *grosscapitalprivate*, *grosscapital*, *savings* and *wealthincome* (variables are sorted from the strongest).
- low values for the variable *ChildFaith*.

The **cluster 9** is made of individuals such as *Antigua and Barbuda*, *Bahamas, The*, *Chile*, *Dominica*, *Israel*, *St. Kitts and Nevis*, *Korea, Rep.*, *St. Lucia*, *Maldives* and *Sint Maarten (Dutch part)*. This group is characterized by :

- high values for variables like *StealProp*, *Violence*, *AcceptBribe*, *carework*, *prison*, *TaxCheat*, *cleanfuel*, *wageworker*, *internet* and *electricity* (variables are sorted from the strongest).
- low values for the variables *agriculture*, *ChildFaith*, *wealthincome*, *selfemployed*, *outofschool* and *resourcerents* (variables are sorted from the weakest).

The **cluster 10** is made of individuals such as *American Samoa*, *Bermuda*, *Cayman Islands*, *Czechia*, *Estonia*, *Gibraltar*, *Greenland*, *Isle of Man*, *Italy* and *Monaco*. This group is characterized by :

- high values for variables like *unvotes*, *ecirank*, *wageworker*, *WorryEduc*, *internet*, *stability*, *gnipercap*, *gdppercap*, *services* and *cleanfuel* (variables are sorted from the strongest).
- low values for variables like *urbanrate*, *selfemployed*, *homicide*, *ChildFaith*, *agriculture*, *gini*, *BeatWife*, *resourcerents*, *outofschool* and *concentration* (variables are sorted from the weakest).

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

- high values for variables like *transportquality*, *wealth*, *corruption*, *goveffectiveness*, *gnipercap*, *WorryEduc*, *unvotes*, *gdppercap*, *ecirank* and *wealthincome* (variables are sorted from the strongest).
- low values for variables like *BeatWife*, *ChildFaith*, *homicide*, *gini*, *selfemployed*, *Violence*, *carework*, *outofpocket*, *prison* and *agriculture* (variables are sorted from the weakest).

The **cluster 12** is made of individuals such as *United Arab Emirates*, *Hong Kong SAR, China*, *Ireland*, *Luxembourg*, *Macao SAR, China*, *Malta*, *Singapore* and *San Marino*. This group is characterized by :

- high values for variables like *exports*, *trade*, *gdppercap*, *gnipercap*, *wealth*, *mobilecellular*, *WorryEduc*, *goveffectiveness*, *transportquality* and *ecirank* (variables are sorted from the strongest).
- low values for the variables *ChildFaith*, *selfemployed*, *homicide*, *gini*, *agriculture*, *BeatWife*, *outofpocket*, *prison*, *concentration* and *outofschool* (variables are sorted from the weakest).

Annexes