Principal Component Analysis

Dataset testDF_imputed$completeObs

This dataset contains 217 individuals and 32 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 **50.52%** of the total dataset inertia ; that means that 50.52% 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 **11.39%**, 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 4619 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.

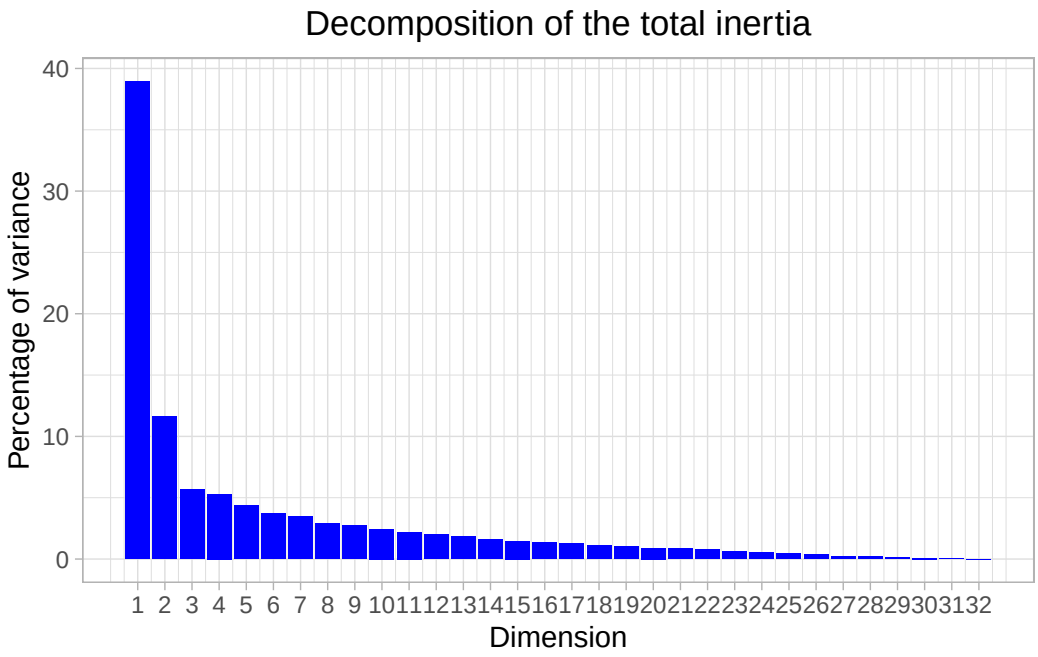

Figure 2 - Decomposition of the total inertia

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 (61.46% against 21.4%). 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

```
## Warning: ggrepel: 18 unlabeled data points (too many overlaps). Consider increasing max.overlaps
```

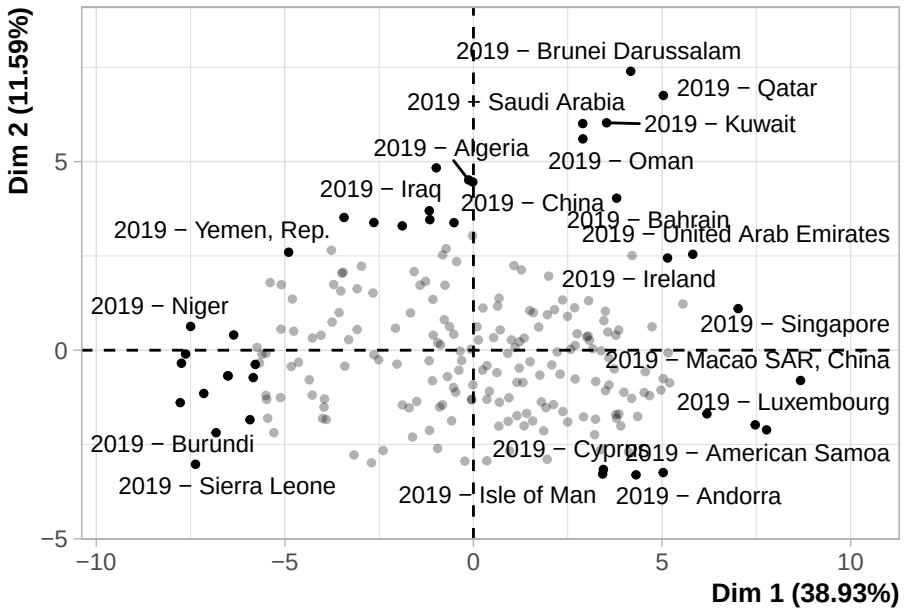

Figure 3.1 - Individuals factor map (PCA) *The labeled individuals are those with the higher contribution to the plane construction.*

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

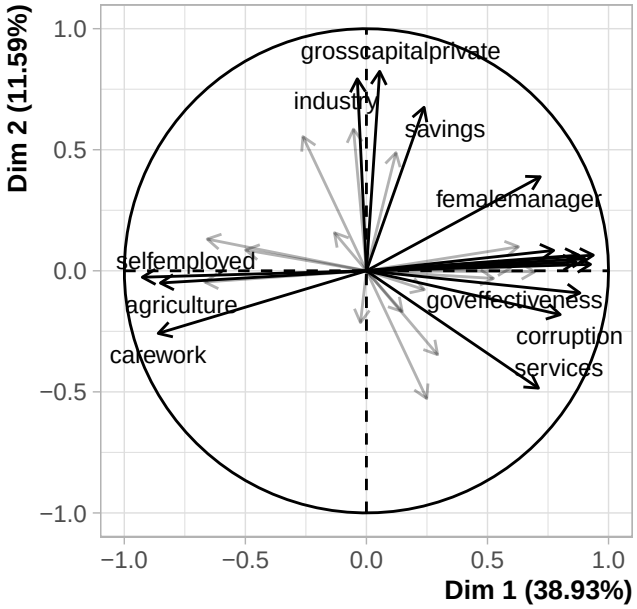

Figure 3.2 - Variables factor map (PCA) *The labeled variables are those the best shown on the plane.*

The **dimension 1** opposes individuals such as *2019 - American Samoa*, *2019 - Isle of Man*, *2019 - Andorra*, *2019 - Bermuda*, *2019 - United Arab Emirates*, *2019 - Hong Kong SAR, China*, *2019 - Singapore*, *2019 - Luxembourg*, *2019 - Macao SAR, China* and *2019 - Ireland* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *2019 - Liberia*, *2019 - Central African Republic*, *2019 - South Sudan*, *2019 - Somalia*, *2019 - Guinea*, *2019 - Congo, Dem. Rep.*, *2019 - Yemen, Rep.*, *2019 - Chad*, *2019 - Guinea-Bissau* and *2019 - Eritrea* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *2019 - American Samoa*, *2019 - Isle of Man*, *2019 - Andorra*, *2019 - Bermuda*, *2019 - United Arab Emirates*, *2019 - Hong Kong SAR, China*, *2019 - Singapore*, *2019 - Luxembourg*, *2019 - Macao SAR, China* and *2019 - Ireland* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for variables like *goveffectiveness*, *corruption*, *internet*, *wageworker*, *cleanfuel*, *services*, *gnipercap*, *gdppercap*, *urbanlevel* and *electricity* (variables are sorted from the strongest).
- low values for the variables *selfemployed*, *agriculture*, *urbanrate*, *carework*, *outofschool*, *outofpocket*, *resourcerents*, *industry*, *grosscapitalprivate* and *grosscapital* (variables are sorted from the weakest).

The group in which the individuals *2019 - Liberia*, *2019 - Central African Republic*, *2019 - South Sudan*, *2019 - Somalia*, *2019 - Guinea*, *2019 - Congo, Dem. Rep.*, *2019 - Yemen, Rep.*, *2019 - Chad*, *2019 - Guinea-Bissau* and *2019 - Eritrea* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *selfemployed*, *carework*, *agriculture*, *urbanrate*, *outofschool* and *outofpocket* (variables are sorted from the strongest).
- low values for variables like *cleanfuel*, *wageworker*, *internet*, *femalemanager*, *goveffectiveness*, *gnipercap*, *urbanlevel*, *electricity*, *gdppercap* and *corruption* (variables are sorted from the weakest).

The **dimension 2** opposes individuals such as *2019 - Brunei Darussalam*, *2019 - Qatar*, *2019 - Bahrain*, *2019 - Kuwait*, *2019 - Saudi Arabia*, *2019 - Turkmenistan*, *2019 - Algeria*, *2019 - Oman*, *2019 - Iran, Islamic Rep.* and *2019 - Iraq* (to the top of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *2019 - American Samoa*, *2019 - Liberia*, *2019 - Isle of Man*, *2019 - Central African Republic*, *2019 - Andorra*, *2019 - South Sudan*, *2019 - Somalia*, *2019 - Guinea*, *2019 - Congo, Dem. Rep.* and *2019 - Yemen, Rep.* (to the bottom of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *2019 - Brunei Darussalam*, *2019 - Qatar*, *2019 - Bahrain*, *2019 - Kuwait*, *2019 - Saudi Arabia*, *2019 - Turkmenistan*, *2019 - Algeria*, *2019 - Oman*, *2019 - Iran, Islamic Rep.* and *2019 - Iraq* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for variables like *resourcerents*, *industry*, *grosscapitalprivate*, *militaryexpenditure*, *savings*, *grosscapital*, *femalemanager*, *cleanfuel*, *wageworker* and *electricity* (variables are sorted from the strongest).
- low values for the variables *femaleminister*, *carework*, *femaleparliament*, *selfemployed* and *services* (variables are sorted from the weakest).

The group in which the individuals *2019 - Liberia*, *2019 - Central African Republic*, *2019 - South Sudan*, *2019 - Somalia*, *2019 - Guinea*, *2019 - Congo, Dem. Rep.*, *2019 - Yemen, Rep.*, *2019 - Chad*, *2019 - Guinea-Bissau* and *2019 - Eritrea* stand (characterized by a negative coordinate on the axis) is sharing :

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

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

The group in which the individuals *2019 - American Samoa*, *2019 - Isle of Man*, *2019 - Andorra*, *2019 - Bermuda*, *2019 - United Arab Emirates*, *2019 - Hong Kong SAR, China*, *2019 - Singapore*, *2019 - Luxembourg*, *2019 - Macao SAR, China* and *2019 - Ireland* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for variables like *goveffectiveness*, *corruption*, *internet*, *wageworker*, *cleanfuel*, *services*, *gnipercap*, *gdppercap*, *urbanlevel* and *electricity* (variables are sorted from the strongest).
- low values for the variables *selfemployed*, *agriculture*, *urbanrate*, *carework*, *outofschool*, *outofpocket*, *resourcerents*, *industry*, *grosscapitalprivate* and *grosscapital* (variables are sorted from the weakest).

---

4. Description of the plane 3:4

```
## Warning: ggrepel: 25 unlabeled data points (too many overlaps). Consider increasing max.overlaps
```

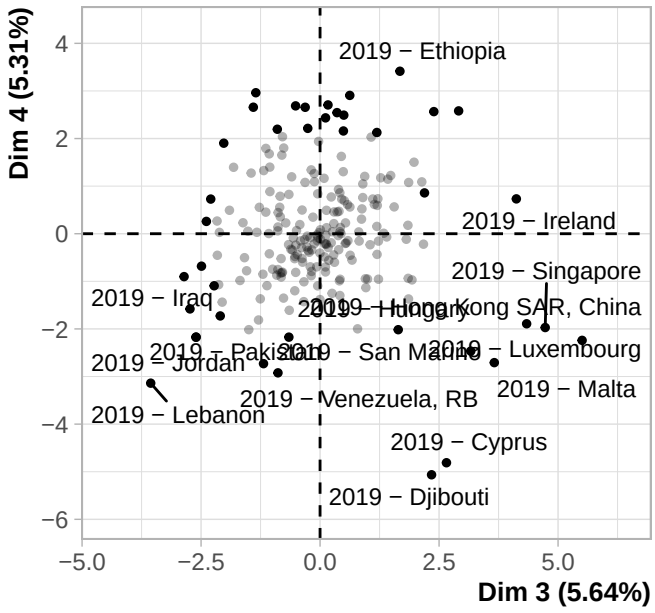

Figure 4.1 - Individuals factor map (PCA) *The labeled individuals are those with the higher contribution to the plane construction.*

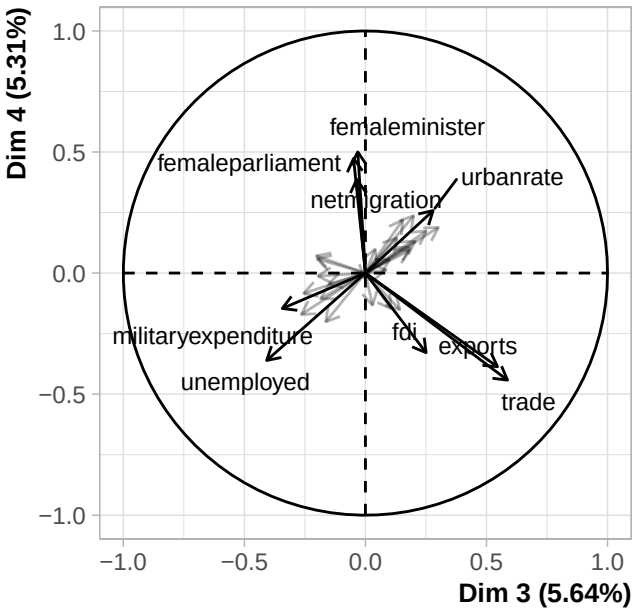

Figure 4.2 - Variables factor map (PCA) *The labeled variables are those the best shown on the plane.*

The **dimension 3** opposes individuals such as *2019 - Malta*, *2019 - Djibouti*, *2019 - Luxembourg*, *2019 - Singapore*, *2019 - San Marino*, *2019 - Ireland*, *2019 - Hungary*, *2019 - Hong Kong SAR, China* and *2019 - Cyprus* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *2019 - Brazil*, *2019 - Lebanon*, *2019 - Greece*, *2019 - Armenia*, *2019 - Jordan*, *2019 - Argentina*, *2019 - St. Vincent and the Grenadines*, *2019 - Libya*, *2019 - Iraq* and *2019 - Venezuela, RB* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *2019 - Malta*, *2019 - Djibouti*, *2019 - Luxembourg*, *2019 - Singapore*, *2019 - San Marino*, *2019 - Ireland*, *2019 - Hungary*, *2019 - Hong Kong SAR, China* and *2019 - Cyprus* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for the variables *exports*, *trade*, *fdi*, *mobilecellular*, *gdppercap*, *savings*, *gnipercap* and *femalemanager* (variables are sorted from the strongest).
- low values for the variables *carework*, *femaleminister*, *femaleparliament* and *unemployed* (variables are sorted from the weakest).

The group in which the individuals *2019 - Brazil*, *2019 - Lebanon*, *2019 - Greece*, *2019 - Armenia*, *2019 - Jordan*, *2019 - Argentina*, *2019 - St. Vincent and the Grenadines*, *2019 - Libya*, *2019 - Iraq* and *2019 - Venezuela, RB* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *unemployed*, *outofpocket*, *militaryexpenditure* and *electricity* (variables are sorted from the strongest).
- low values for variables like *goveffectiveness*, *corruption*, *femaleminister*, *femaleparliament*, *urbanrate*, *gdppercap*, *gnipercap*, *netmigration*, *grosscapital* and *exports* (variables are sorted from the weakest).

The **dimension 4** opposes individuals such as *2019 - Colombia*, *2019 - Ethiopia*, *2019 - Spain*, *2019 - Australia*, *2019 - Mozambique*, *2019 - Tanzania*, *2019 - Canada*, *2019 - New Zealand*, *2019 - Austria* and *2019 - Finland* (to the top of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *2019 - Brazil*, *2019 - Lebanon*, *2019 - Malta*, *2019 - Greece*, *2019 - Armenia*, *2019 - Jordan*, *2019 - Argentina*, *2019 - Djibouti*, *2019 - Luxembourg* and *2019 - Singapore* (to the bottom of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *2019 - Colombia*, *2019 - Ethiopia*, *2019 - Spain*, *2019 - Australia*, *2019 - Mozambique*, *2019 - Tanzania*, *2019 - Canada*, *2019 - New Zealand*, *2019 - Austria* and *2019 - Finland* stand (characterized by a positive coordinate on the axis) is sharing :

- high values for the variables *femaleminister*, *femaleparliament*, *corruption*, *netmigration*, *goveffectiveness*, *urbanrate* and *selfemployed* (variables are sorted from the strongest).
- low values for the variables *unemployed*, *trade*, *exports*, *outofpocket*, *electricity*, *X*, *militaryexpenditure* and *wageworker* (variables are sorted from the weakest).

The group in which the individuals *2019 - Brazil*, *2019 - Lebanon*, *2019 - Greece*, *2019 - Armenia*, *2019 - Jordan*, *2019 - Argentina*, *2019 - St. Vincent and the Grenadines*, *2019 - Libya*, *2019 - Iraq* and *2019 - Venezuela, RB* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *unemployed*, *outofpocket*, *militaryexpenditure* and *electricity* (variables are sorted from the strongest).
- low values for variables like *goveffectiveness*, *corruption*, *femaleminister*, *femaleparliament*, *urbanrate*, *gdppercap*, *gnipercap*, *netmigration*, *grosscapital* and *exports* (variables are sorted from the weakest).

The group in which the individuals *2019 - Malta*, *2019 - Djibouti*, *2019 - Luxembourg*, *2019 - Singapore*, *2019 - San Marino*, *2019 - Ireland*, *2019 - Hungary*, *2019 - Hong Kong SAR, China* and *2019 - Cyprus* stand (characterized by a negative coordinate on the axis) is sharing :

- high values for the variables *exports*, *trade*, *fdi*, *mobilecellular*, *gdppercap*, *savings*, *gnipercap* and *femalemanager* (variables are sorted from the strongest).
- low values for the variables *carework*, *femaleminister*, *femaleparliament* and *unemployed* (variables are sorted from the weakest).

---

5. Classification

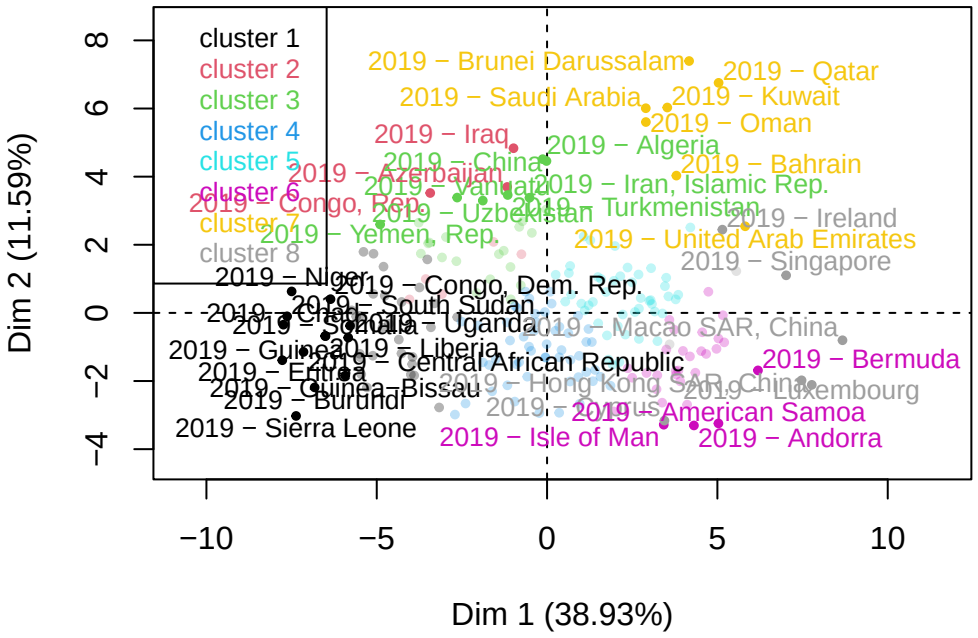

Figure 5 - Ascending Hierarchical Classification of the individuals. *The classification made on individuals reveals 8 clusters.*

The **cluster 1** is made of individuals such as *2019 - Burundi*, *2019 - Central African Republic*, *2019 - Congo, Dem. Rep.*, *2019 - Eritrea*, *2019 - Guinea*, *2019 - Guinea-Bissau*, *2019 - Liberia*, *2019 - Niger*, *2019 - Sierra Leone* and *2019 - Somalia*. This group is characterized by :

- high values for the variables *selfemployed*, *agriculture*, *carework*, *urbanrate*, *outofschool* and *outofpocket* (variables are sorted from the strongest).
- low values for variables like *electricity*, *internet*, *cleanfuel*, *wageworker*, *goveffectiveness*, *femalemanager*, *gnipercap*, *urbanlevel*, *services* and *gdppercap* (variables are sorted from the weakest).

The **cluster 2** is made of individuals such as *2019 - Azerbaijan*, *2019 - Congo, Rep.* and *2019 - Iraq*. This group is characterized by :

- high values for the variables *resourcerents*, *industry*, *unemployed*, *outofpocket*, *selfemployed* and *urbanrate* (variables are sorted from the strongest).
- low values for the variables *corruption*, *goveffectiveness*, *services* and *wageworker* (variables are sorted from the weakest).

The **cluster 3** is made of individuals such as *2019 - China*, *2019 - Algeria*, *2019 - Iran, Islamic Rep.*, *2019 - Turkmenistan*, *2019 - Uzbekistan*, *2019 - Vanuatu* and *2019 - Yemen, Rep.*. This group is characterized by :

- high values for the variables *grosscapitalprivate*, *savings*, *grosscapital*, *outofpocket*, *industry*, *agriculture*, *selfemployed*, *urbanrate* and *X* (variables are sorted from the strongest).
- low values for variables like *femaleminister*, *services*, *urbanlevel*, *gdppercap*, *gnipercap*, *femaleparliament*, *corruption*, *wageworker*, *goveffectiveness* and *netmigration* (variables are sorted from the weakest).

The **cluster 4** is made of individuals sharing :

- high values for the variables *unemployed*, *electricity* and *carework* (variables are sorted from the strongest).
- low values for the variables *savings*, *grosscapital*, *urbanrate*, *grosscapitalprivate*, *resourcerents*, *femalemanager*, *gdppercap*, *gnipercap* and *industry* (variables are sorted from the weakest).

The **cluster 5** is made of individuals sharing :

- high values for variables like *femalemanager*, *wageworker*, *internet*, *cleanfuel*, *electricity*, *goveffectiveness*, *mobilecellular*, *gnipercap*, *corruption* and *urbanlevel* (variables are sorted from the strongest).
- low values for the variables *selfemployed*, *carework*, *agriculture*, *urbanrate*, *outofschool*, *outofpocket* and *resourcerents* (variables are sorted from the weakest).

The **cluster 6** is made of individuals such as *2019 - Andorra*, *2019 - American Samoa*, *2019 - Bermuda* and *2019 - Isle of Man*. This group is characterized by :

- high values for variables like *corruption*, *goveffectiveness*, *gnipercap*, *femaleminister*, *gdppercap*, *services*, *femaleparliament*, *internet*, *wageworker* and *urbanlevel* (variables are sorted from the strongest).
- low values for variables like *selfemployed*, *carework*, *agriculture*, *outofpocket*, *outofschool*, *urbanrate*, *industry*, *grosscapitalprivate*, *X* and *resourcerents* (variables are sorted from the weakest).

The **cluster 7** is made of individuals such as *2019 - United Arab Emirates*, *2019 - Bahrain*, *2019 - Brunei Darussalam*, *2019 - Kuwait*, *2019 - Oman*, *2019 - Qatar* and *2019 - Saudi Arabia*. This group is characterized by :

- high values for variables like *militaryexpenditure*, *industry*, *resourcerents*, *grosscapitalprivate*, *gnipercap*, *femalemanager*, *gdppercap*, *savings*, *internet* and *wageworker* (variables are sorted from the strongest).
- low values for the variables *carework*, *selfemployed*, *femaleminister*, *outofpocket*, *agriculture* and *unemployed* (variables are sorted from the weakest).

The **cluster 8** is made of individuals such as *2019 - Cyprus*, *2019 - Hong Kong SAR, China*, *2019 - Ireland*, *2019 - Luxembourg*, *2019 - Macao SAR, China* and *2019 - Singapore*. This group is characterized by :

- high values for variables like *exports*, *trade*, *gdppercap*, *fdi*, *gnipercap*, *mobilecellular*, *services*, *goveffectiveness*, *corruption* and *urbanlevel* (variables are sorted from the strongest).
- low values for the variables *carework*, *selfemployed*, *agriculture*, *outofschool* and *industry* (variables are sorted from the weakest).

Annexes