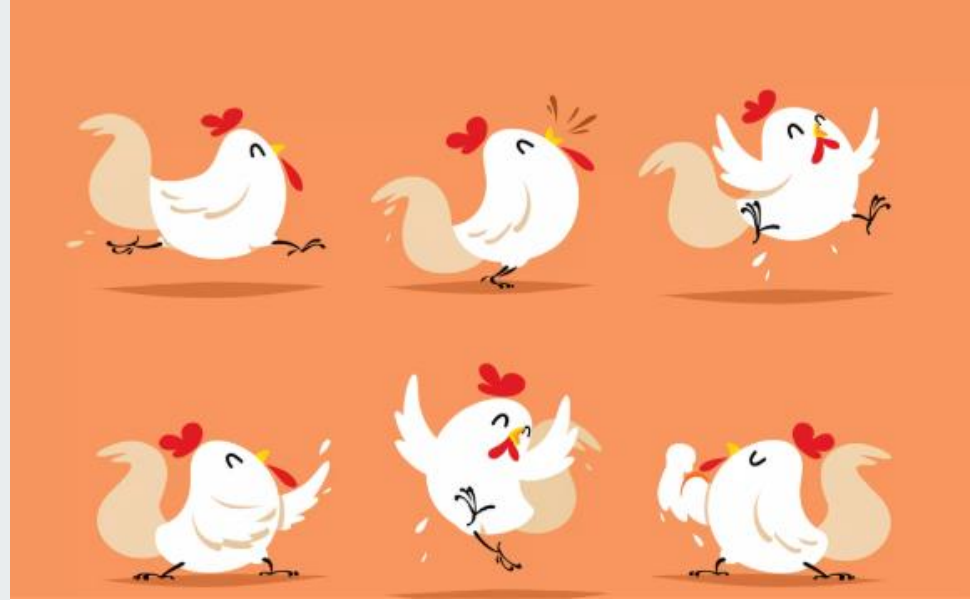




Étude de marché

Quels pays cibles pour
l'exportation de poulets ?





Plan

Contexte

Construction de
l'échantillon

Classification

Réduction des
variables

Sur la qualité des
données

Choix des pays

Conclusion





Contexte





Votre demande :

- Recherche de développement à l'international
- Secteur : agroalimentaire (poulet)
- Champ des possibles : tous les pays sont envisageables !
- Votre stratégie : exportation plutôt que production sur place



Construction de l'échantillon





Construction de l'échantillon

Source des données : [FAO](#)

L'échantillon contient l'ensemble des pays disponibles, chacun caractérisé par ces variables :

1. Taux d'évolution de la population 2013-2017 ;
2. Proportion de protéines d'origine animale par rapport à la quantité totale de protéines dans la disponibilité alimentaire du pays ;
3. Disponibilité alimentaire en protéines par habitant ;
4. Disponibilité alimentaire en calories par habitant.
5. PIB par habitant

Population : 236 lignes (pays et régions)

Code zone	Zone	pop_2013	pop_2017	evol_pop
1	Arménie	2897587	2944790	0.016290
2	Afghanistan	32269589	36296113	0.124778
3	Albanie	2903790	2884169	-0.006757
4	Algérie	38140133	41389189	0.085187

172 lignes

Code zone	Zone	ratio_evol_pop	dispo_PROT_(g/pers/j)	prop_prot_animaux	dispo_ALI_(kcal/pers/j)	PIB_par_hbt_(US\$)
96	136 Mauritanie	0.121828	79.82	0.631295	2842.0	1597.806534

Données ali : 173 lignes

country_code	country	dispo_PROT_(g/pers/j)	dispo_PROT_ANIMALES_(g/pers/j)	prop_prot_animaux
1	Arménie	97.33	52.77	0.542176
2	Afghanistan	54.09	43.53	0.804770
3	Albanie	119.50	53.18	0.445021
4	Algérie	92.85	67.15	0.723209
7	Angola	54.09	37.62	0.695507



Classification



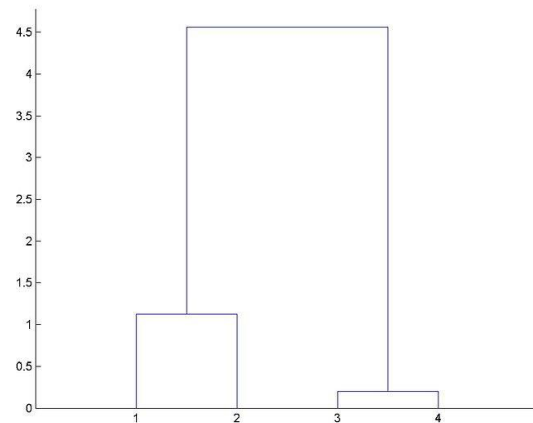
Classification hiérarchique

Objectif :

identifier des groupes d'observations ayant des caractéristiques similaires

Principe :

constituer des groupes (clusters) de manière à ce que les individus dans un même groupe se ressemblent, et les individus dans des groupes différents soient dissemblables.



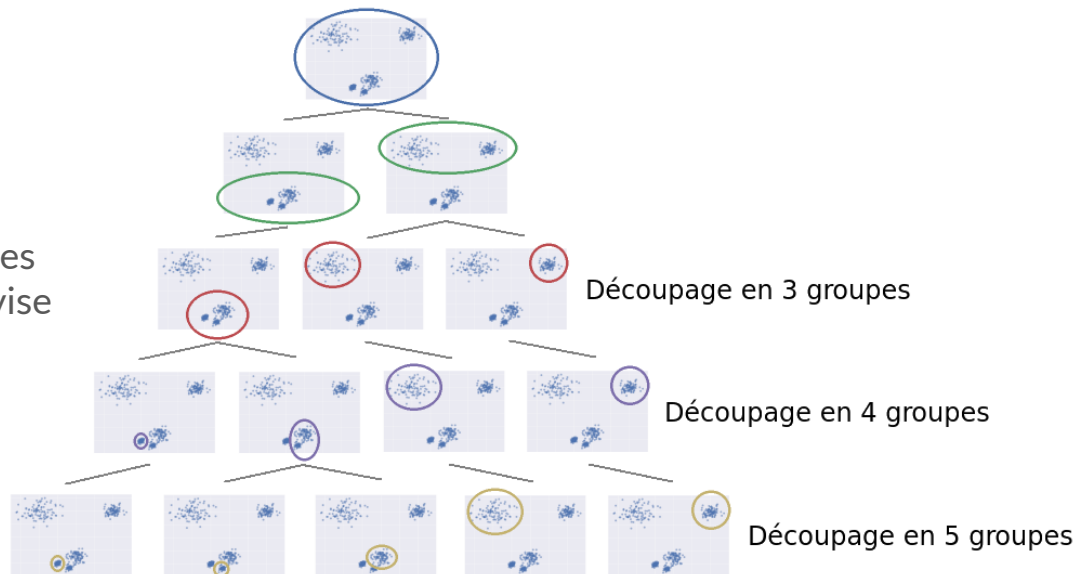
Source [Wikipedia](#)

Classification hiérarchique

2 approches

L'approche **descendante** :

On part d'un grand cluster contenant tous les points (ici, un point = un pays), puis on le divise successivement jusqu'à obtenir autant de clusters que de points.



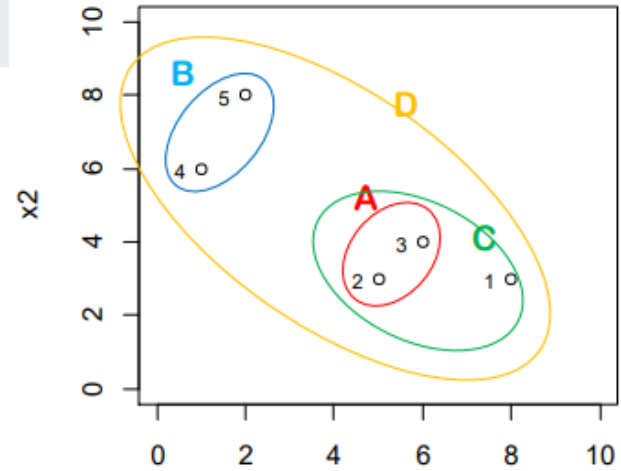
Source [Openclassrooms](https://openclassrooms.com)

Classification hiérarchique

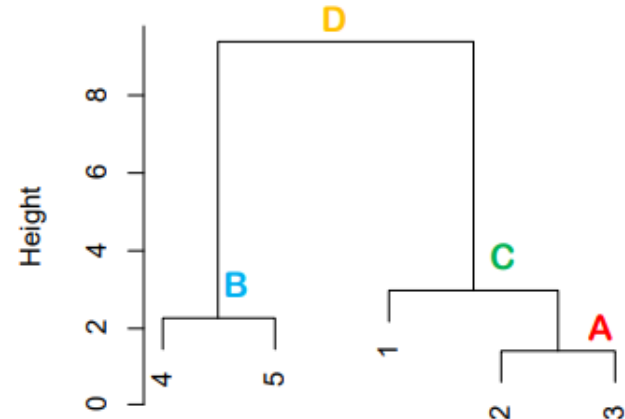
2 approches

L'approche **ascendante** :

On considère tout d'abord que chaque point est un cluster. Il y a donc autant de clusters que de points. Ensuite, on cherche les deux clusters les plus proches, et on les agglomère en un seul cluster. On répète cette étape jusqu'à ce que tous les points soient regroupés en un seul grand cluster.



Cluster Dendrogram

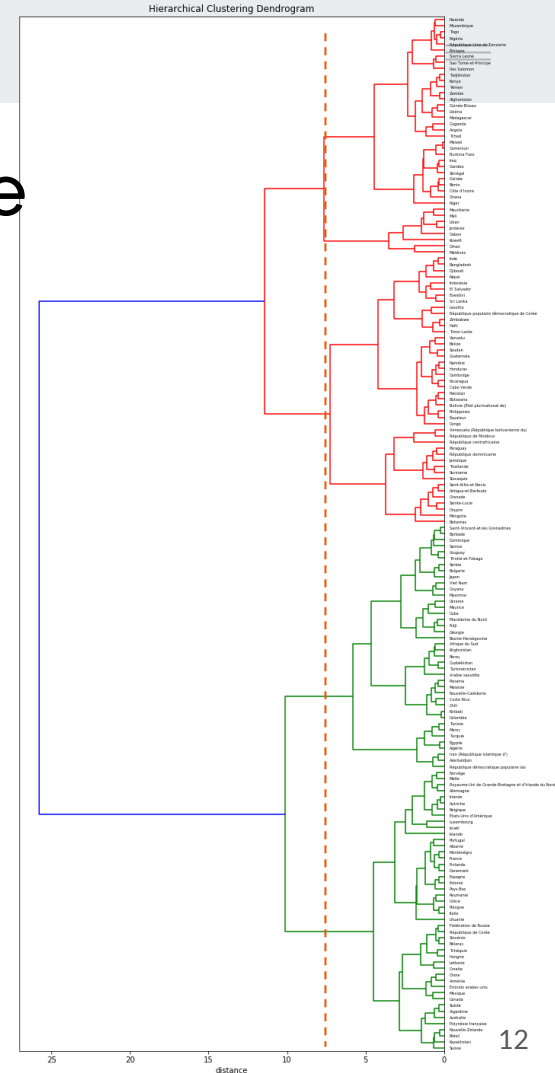


Classification ascendante hiérarchique

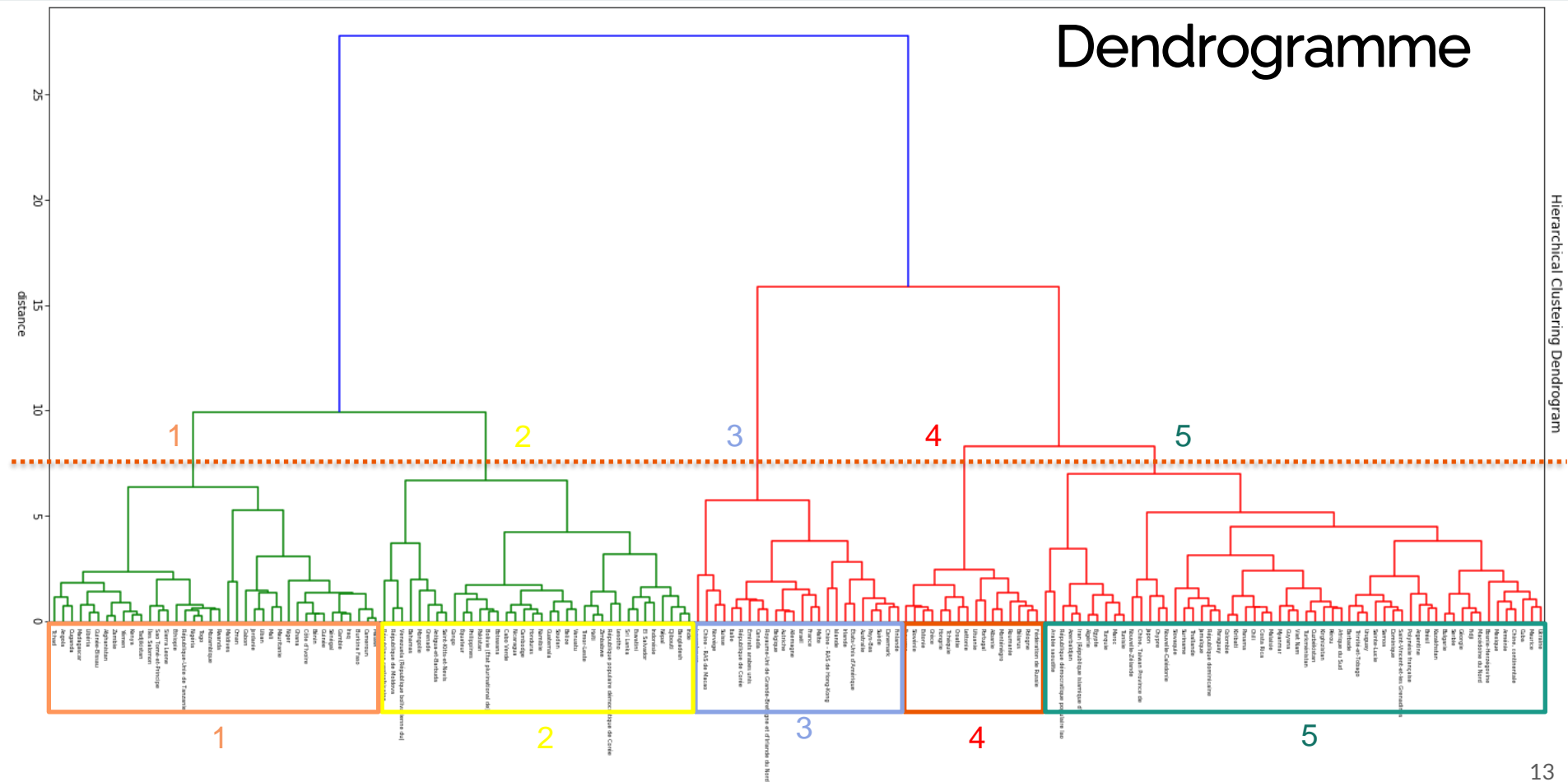
Dendrogramme

Découpage en 5 clusters

Zone	ratio_evol_pop	dispo_PROT_(g/pers/j)	prop_prot_animales	dispo_ALI_(kcal/pers/j)	PIB_par_hbt_(US\$)	cluster
Arménie	0.016290	97.33	0.542176	3072.0	3933.682101	5
Afghanistan	0.124778	54.09	0.804770	1997.0	605.557362	1
Albanie	-0.006757	119.50	0.445021	3400.0	4445.132198	4
Algérie	0.085187	92.85	0.723209	3345.0	4051.244377	5
Angola	0.146103	54.09	0.695507	2266.0	4100.291004	1



13



Group 1:

Afghanistan, Angola, Burkina Faso, Bénin, Cameroun, Côte d'Ivoire, Gabon, Gambie, Ghana, Guinée, Guinée-Bissau, Iraq, Jordanie, Kenya, Liban, Libéria, Madagascar, Malawi, Maldives, Mali, Mauritanie, Mozambique, Niger, Nigéria, Oman, Ouganda, Rwanda, République-Unie de Tanzanie, Sao Tomé-et-Principe, Sierra Leone, Sénégal, Tadjikistan, Tchad, Togo, Yémen, Zambie, Éthiopie, Îles Salomon

Group 2:

Antigua-et-Barbuda, Bahamas, Bangladesh, Belize, Bolivie (État plurinational de), Botswana, Cabo Verde, Cambodge, Congo, Djibouti, El Salvador, Eswatini, Grenade, Guatemala, Haïti, Honduras, Inde, Indonésie, Lesotho, Mongolie, Namibie, Nicaragua, Népal, Pakistan, Philippines, République centrafricaine, République de Moldova, République populaire démocratique de Corée, Saint-Kitts-et-Nevis, Soudan, Sri Lanka, Timor-Leste, Vanuatu, Venezuela (République bolivarienne du), Zimbabwe, Équateur

Group 3:

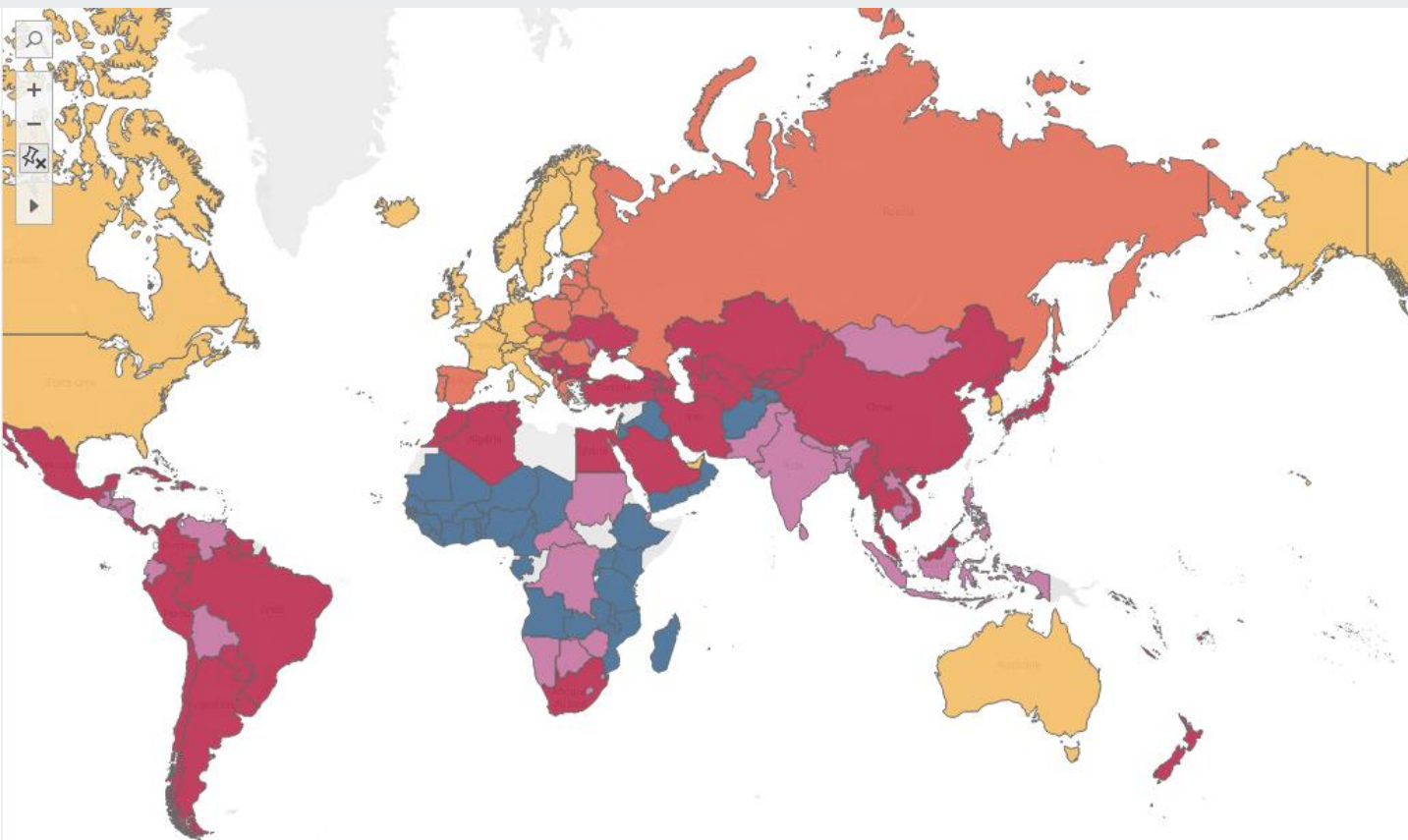
Allemagne, Australie, Autriche, Belgique, Canada, Chine - RAS de Hong-Kong, Chine - RAS de Macao, Danemark, Finlande, France, Irlande, Islande, Israël, Italie, Luxembourg, Malte, Norvège, Pays-Bas, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, République de Corée, Suisse, Suède, Émirats arabes unis, États-Unis d'Amérique

Group 4:

Albanie, Bélarus, Croatie, Espagne, Estonie, Fédération de Russie, Grèce, Hongrie, Lettonie, Lituanie, Monténégro, Pologne, Portugal, Roumanie, Slovaquie, Tchéquie

Group 5:

Afrique du Sud, Algérie, Arabie saoudite, Argentine, Arménie, Azerbaïdjan, Barbade, Bosnie-Herzégovine, Brésil, Bulgarie, Chili, Chine, Taiwan Province de, Chine, continentale, Chypre, Colombie, Costa Rica, Cuba, Dominique, Fidji, Guyana, Géorgie, Iran (République islamique d'), Jamaïque, Japon, Kazakhstan, Kirghizistan, Kiribati, Koweït, Macédoine du Nord, Malaisie, Maroc, Maurice, Mexique, Myanmar, Nouvelle-Calédonie, Nouvelle-Zélande, Ouzbékistan, Panama, Paraguay, Polynésie française, Pérou, République dominicaine, République démocratique populaire lao, Saint-Vincent-et-les Grenadines, Sainte-Lucie, Samoa, Serbie, Slovaquie, Suriname, Thaïlande, Trinité-et-Tobago, Tunisie, Turkménistan, Turquie, Ukraine, Uruguay, Viet Nam, Égypte



N° de cluster

1

2

3

4

5

Lien vers
[l'outil en ligne](#)

Ce que l'on constate (grandes lignes) :

Groupe 1 : pays relativement pauvres et/ou dans des situations géopolitiques complexes (Afghanistan, Iraq). Forte représentation de l'Afrique.

Groupe 2 : pays relativement pauvres et/ou en crise économique (Vénézuela, Zimbabwe).

Groupe 3 : pays riches. ex : France, États-Unis

Groupe 4 : pays un peu moins riches que ceux du groupe 3, forte représentation de l'Europe de l'Est et du Sud.

Groupe 5 : groupe visiblement le plus mélangés à la fois au niveau des richesses (Koweït, Suriname) et géographique.

A priori, on peut déjà se dire que le groupe 1 et 2 seront peu intéressants pour notre projet. Idem pour les pays du groupe 3 (pays riches déjà pourvus, marchés saturés).

Resteront probablement les **groupes 4 et 5**.



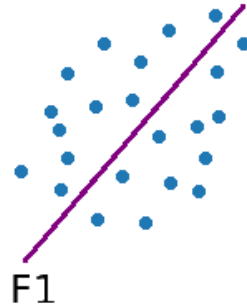
Réduction des variables



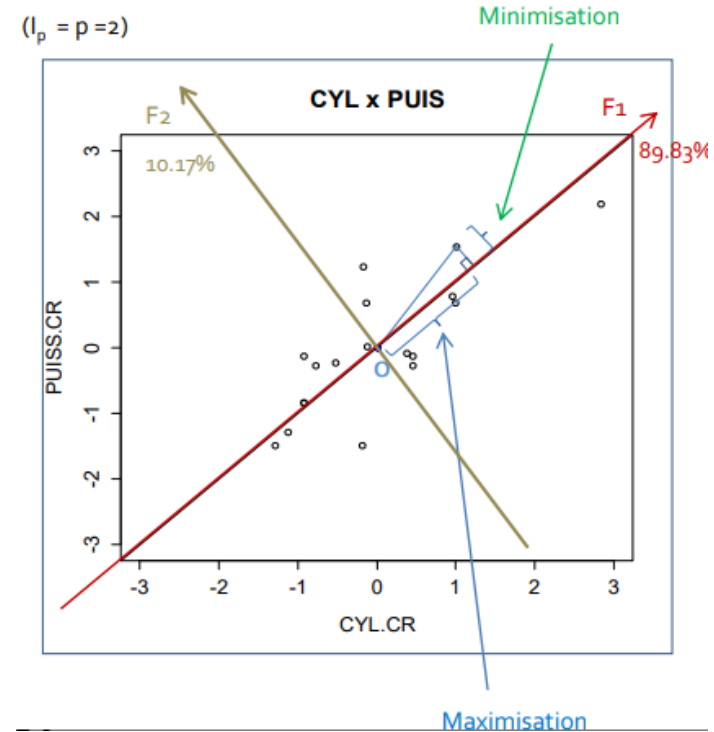
Analyse en composantes principales (ACP)

On peut voir l'ACP comme une compression avec perte (contrôlée) de l'information.
Permet de Synthétiser.

Par exemple ici, nous chercherons une projection sur un axe (à 1 dimension) plutôt que sur un plan.
Les axes factoriels sont des axes virtuels issus d'une synthèse entre les variables de l'analyse.



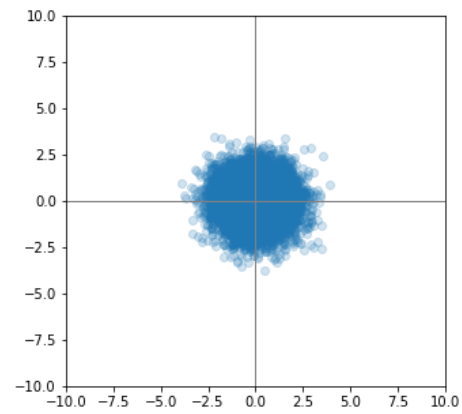
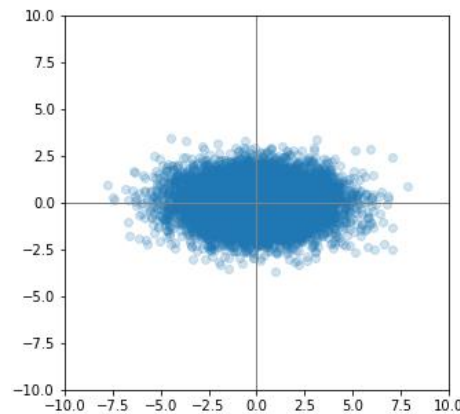
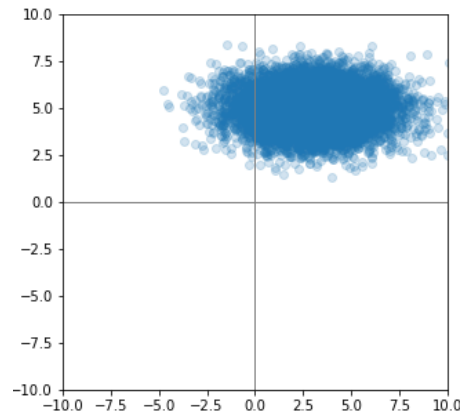
Source [Openclassrooms](#)



Source [Univ.Lyon2](#)

Analyse en composantes principales (ACP)

Il s'agira aussi de **centrer** et **réduire** les données (ACP normée).



Source [Openclassrooms](#)

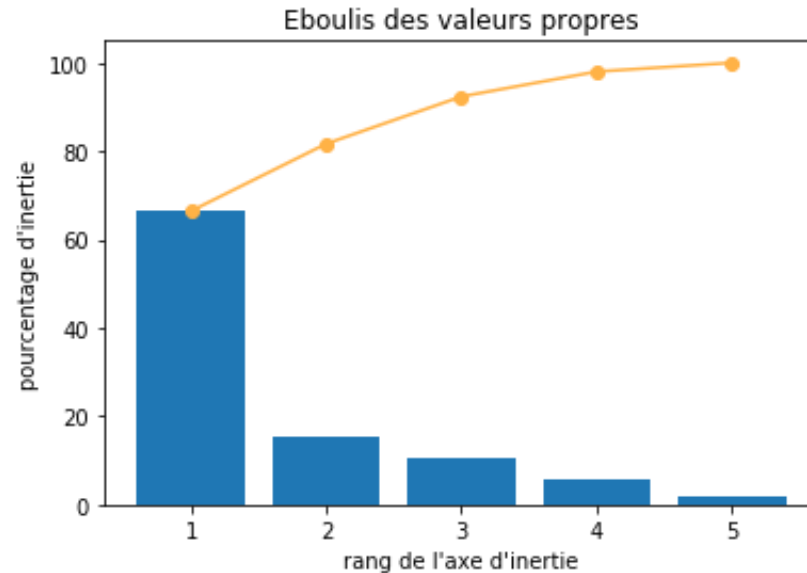
ACP – Éboulis des valeurs propres



Combien d'axes ?

Pourcentage de variance expliquée =

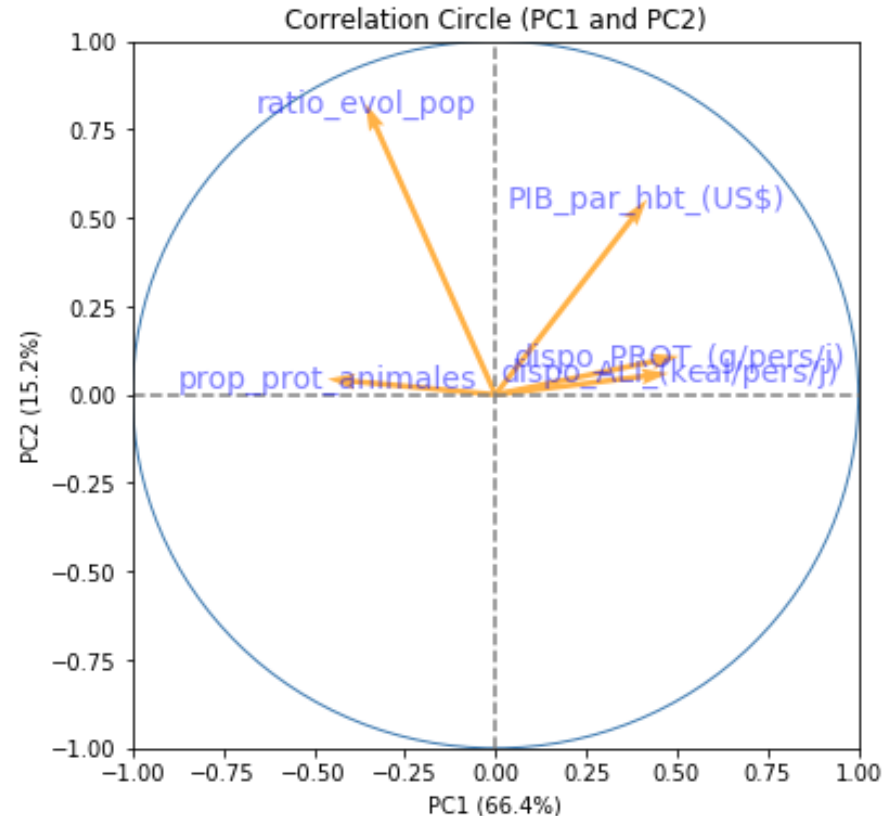
La part de l'information que l'on retrouve sur chaque axe



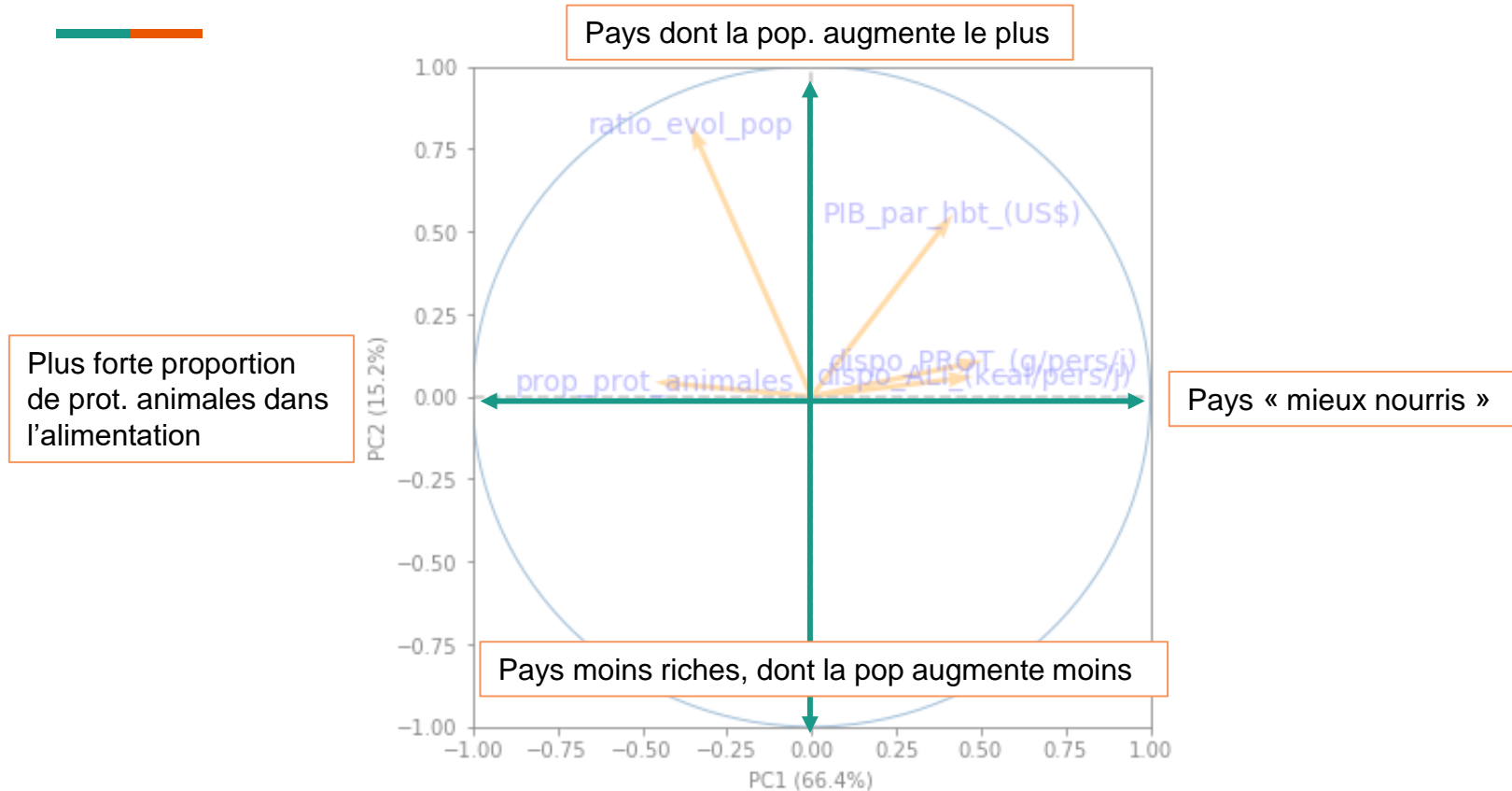
Pourcentage de variance expliquée : [66.43 15.23 10.7 5.65 2.]

Pourcentage de variance expliquée (somme cumulative) : [66.43 81.66 92.36 98. 100.]

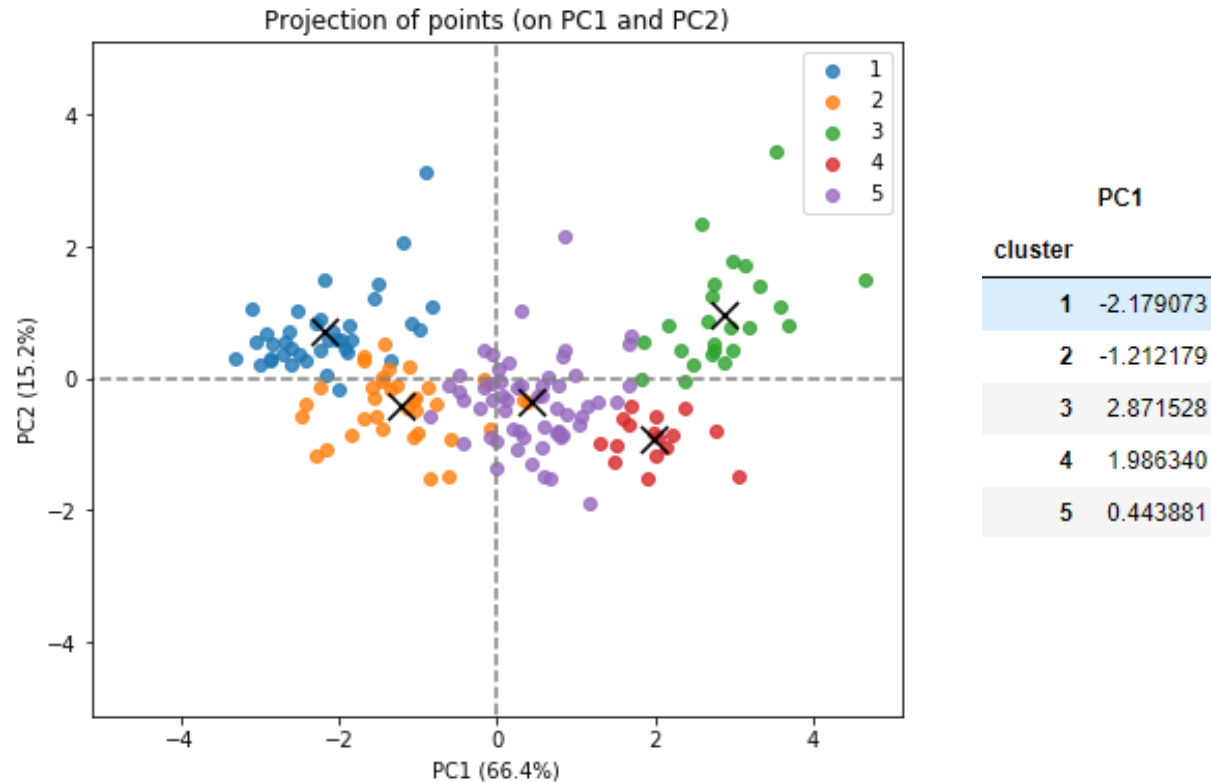
Résultats de notre ACP – Cercle des corrélations



Résultats de notre ACP – Cercle des corrélations



ACP – Projection des individus sur le 1^{er} plan factoriel

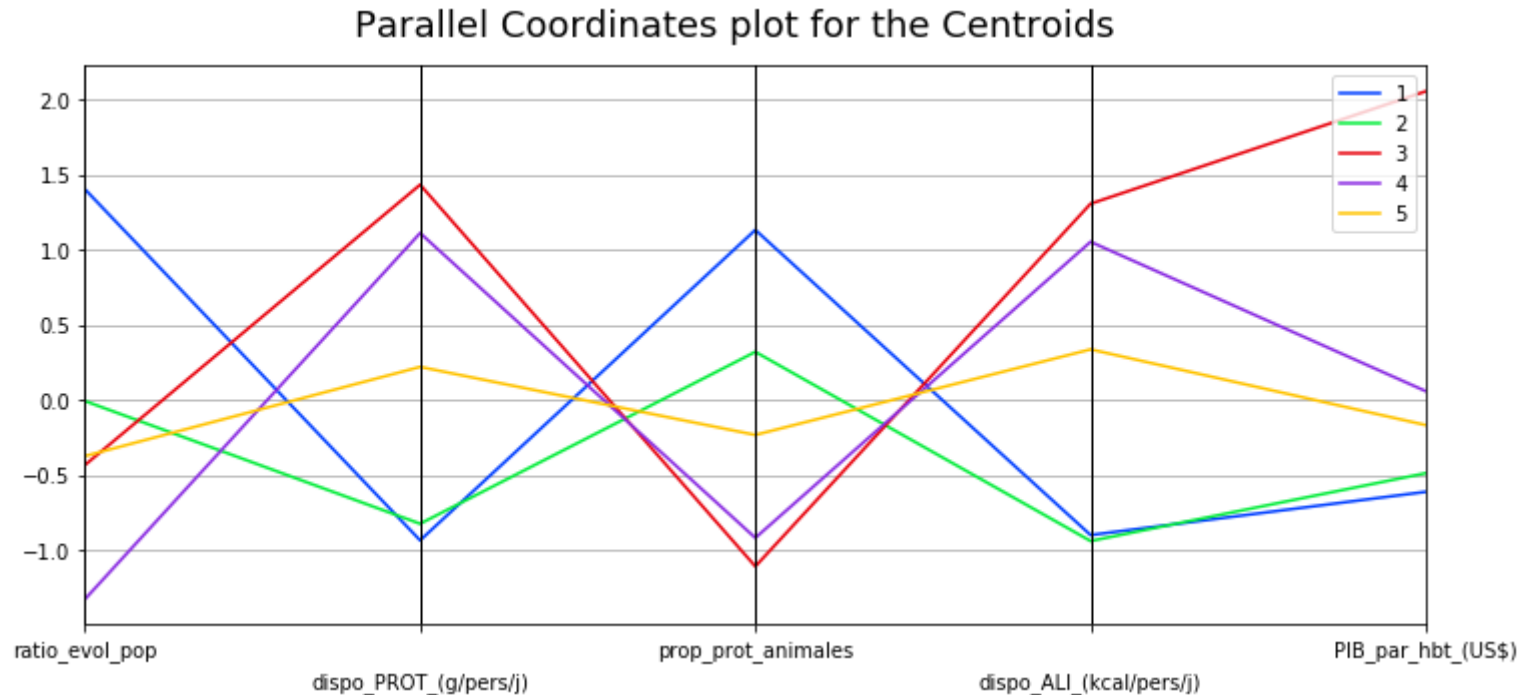


Caractérisation par les moyennes



	ratio_evol_pop	dispo_PROT_(g/pers/j)	prop_prot_animales	dispo_ALI_(kcal/pers/j)	PIB_par_hbt_(US\$)	PC1
cluster						
1	0.124917	62.425789	0.753536	2451.973684	2321.062882	-2.179073
2	0.054478	64.643889	0.624266	2433.972222	4732.389461	-1.212179
3	0.033161	110.044583	0.396816	3450.375000	54144.962421	2.871528
4	-0.011454	103.566875	0.427069	3335.625000	15341.566329	1.986340
5	0.036157	85.649655	0.536188	3011.500000	10955.881915	0.443881

Caractérisation par les moyennes



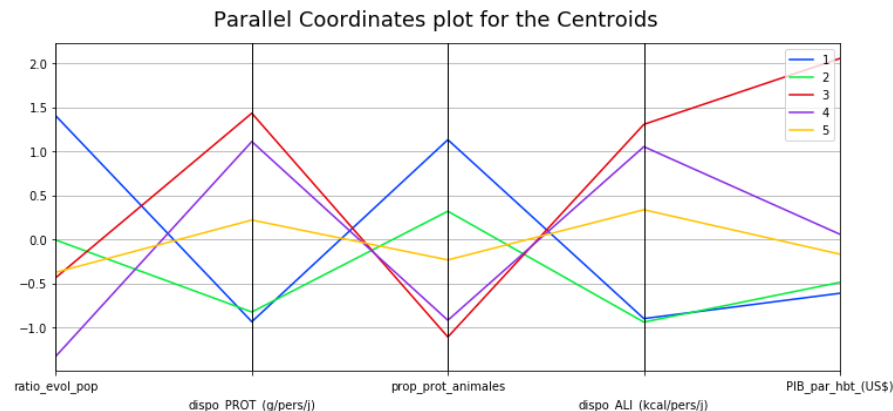
Caractérisation par les moyennes

Analyse

La France se situe dans le groupe 3.

- Les groupes 1 et 2 ont les plus faibles dispo ali et les plus faibles PIB.
- Le groupe 3 a la plus faible prop prot ani.
- Le groupe 4 a le plus faible ratio evol pop.
- Le groupe 5 a les valeurs les plus constantes.

=> Nous retenons le groupe 5. Dans celui-ci nous pourrions sélectionner les pays qui ont le meilleur ratio d'évolution de la population.





Sur la qualité des données



Vérifions que nos groupes diffèrent réellement

Test de comparaison de deux populations gaussiennes

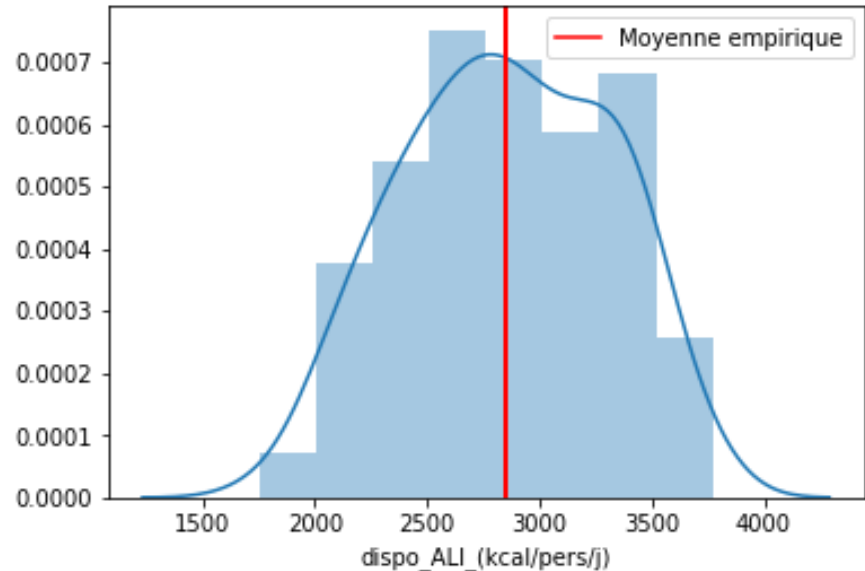
Recherche de variables gaussiennes

Le test **Kolmogorov-Smirnov** permet de tester l'adéquation à une loi normale.

H0 : l'échantillon est distribué selon une loi normale

Ici, $p\text{-value} > 0,05 \Rightarrow H_0$ non rejetée

Variable retenue : `dispo_ali`



Vérifions que nos groupes diffèrent réellement



Pour comparer deux échantillons gaussiens (nos variables suivent-elles la même loi normale ?) :

1. Un test d'égalité des variances (Bartlett).

Ici, $p\text{-value} > 0,05 \Rightarrow$ on ne rejette pas l'hypothèse d'égalité des variances au niveau de test 5%*

2. Test de Student d'égalité des moyennes.

$p\text{-value} < 0,05 \Rightarrow$ on rejette l'hypothèse d'égalité des moyennes : nos groupes diffèrent réellement.

**Si les variances ne sont pas considérées comme égales, les deux échantillons n'ont pas la même loi. Si les variances sont considérées comme égales, il est alors possible d'estimer cette variance sur les deux échantillons à la fois, et de tester l'égalité des moyennes en utilisant cette variance empirique globale.*

Notons qu'il est néanmoins possible d'effectuer un test de comparaison des moyennes sous hypothèse de variances différentes. Il ne s'agit pas d'une comparaison des lois, mais alors d'une comparaison simple des moyennes.



Choix des pays



Classement en fonction de la variable qui nous intéresse

...puis réduction au nombre de pays souhaités

```
(X_c
    # Je sélectionne le cluster qui m'intéresse
    .query('cluster == 5')

    # Sélection dans le cluster choisi en fonction d'une variable
    .sort_values(by='ratio_evol_pop', ascending=False)

    # Réduction du nb de pays
    .head(8)
)
```

Koweït
Arabie saoudite
Égypte
Algérie
Kirghizistan
Turkménistan
Panama
Turquie

Classement plus précis

Nous pouvons choisir les variables

```
# Si l'on veut réduire encore :  
(X_c  
    # Je sélectionne le cluster qui m'intéresse  
    .query('cluster == 5')  
  
    # Sélection dans le cluster choisi en fonction d'une variable  
    .sort_values(by='ratio_evol_pop', ascending=False)  
  
    # Réduction à 5 pays  
    .head(5)  
  
    # Sélection grâce à une autre variable  
    .sort_values(by='PIB_par_hbt_(US$)', ascending=False)  
)
```

Koweït

Arabie saoudite

Algérie

Égypte

Kirghizistan



Conclusion





Nous vous conseillons ces pays ...

...sachant que d'autres possibilités d'analyse sont possibles en fonction de vos besoins !



Merci !

