

La poule qui chante

Étude de marché pour le lancement à l'international



SOMMAIRE

Contexte et objectifs

Choix et nettoyage des données

Analyse de la donnée

- Normalisation
- Classification hiérarchique
- Algorithme K-mean
- Analyse en composantes principales (ACP)

Conclusion



Contexte et objectifs



Le contexte

L'entreprise " La poule qui chante " souhaite se développer à l'international.



Les objectifs

Trouver les pays les plus intéressants pour l'export des produits, via une étude et une analyse de marché global.



Choix des données

Indice de consommation

Fichier avec les disponibilités de protéines et alimentaire en 2022
(Sur la viande de poulet)

Indice économique

Fichier avec le PIB par habitant par en 2022

Indice commercial

Fichier avec l'importation de viande de poulet en 2022
(en tonnes)

Indice population

Fichier avec la population de 2000 à 2022

Indice politique

Fichier avec la stabilité politique de 2000 à 2022

Travail de la données

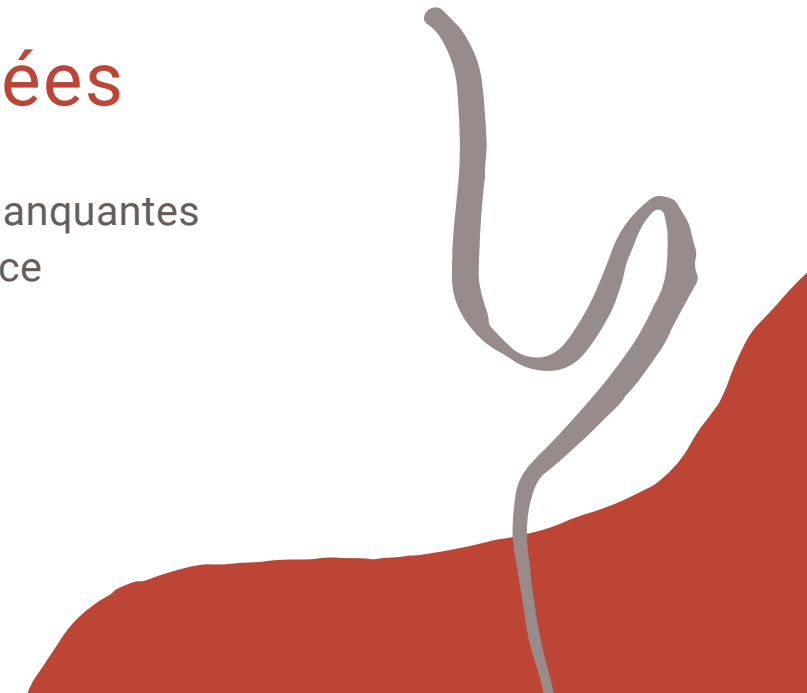
Nettoyage de la données

- Suppression des colonnes non utiles
- Vérification des doublons
- Unification de la nomenclature et de l'ordre des différentes colonnes

Consolidation

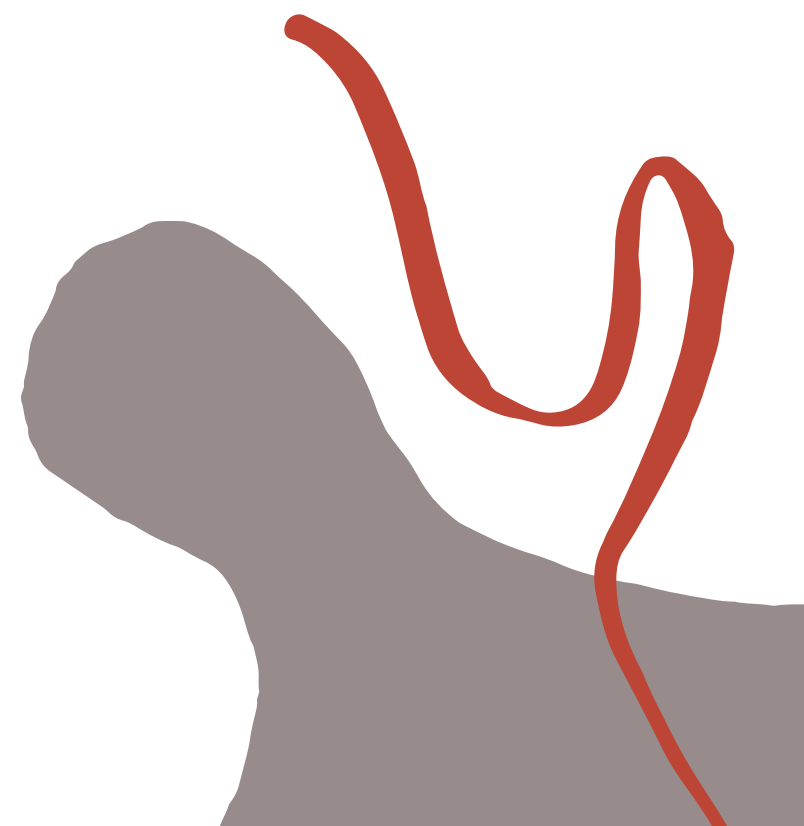
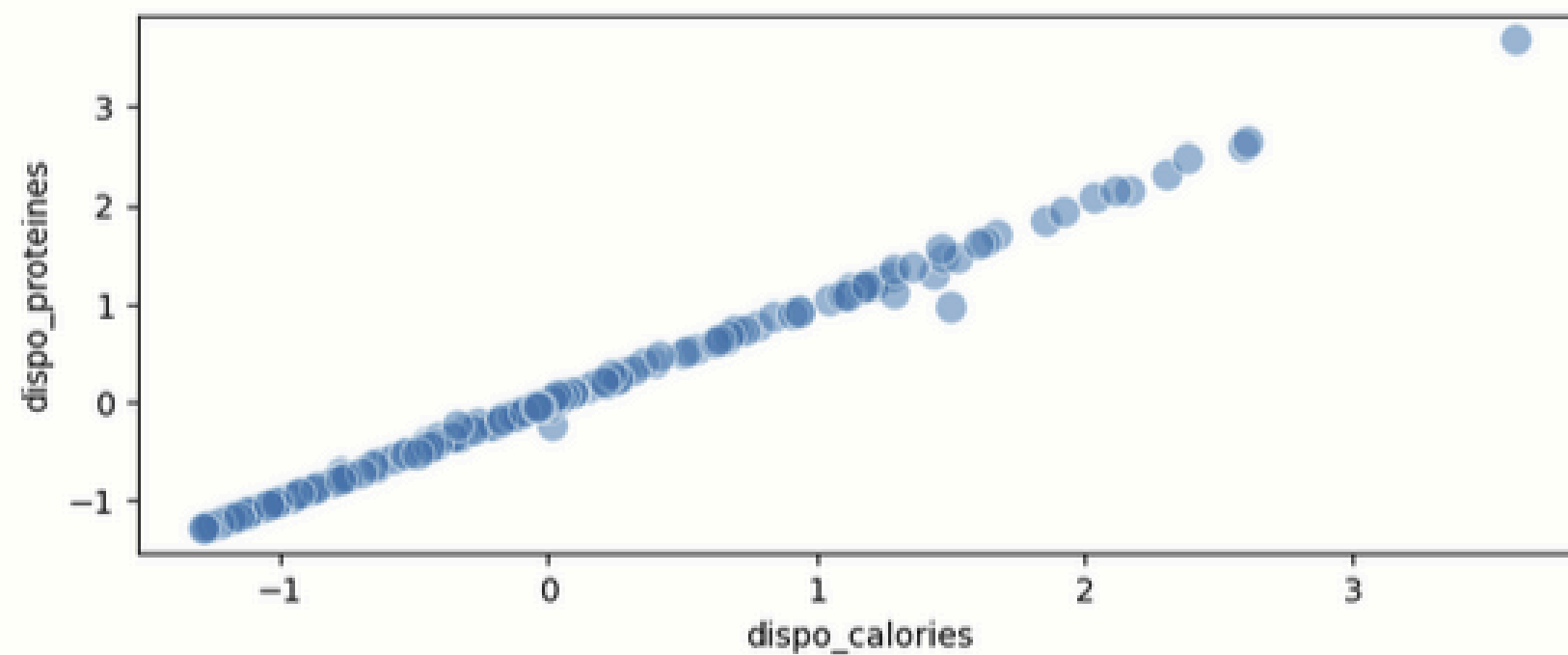
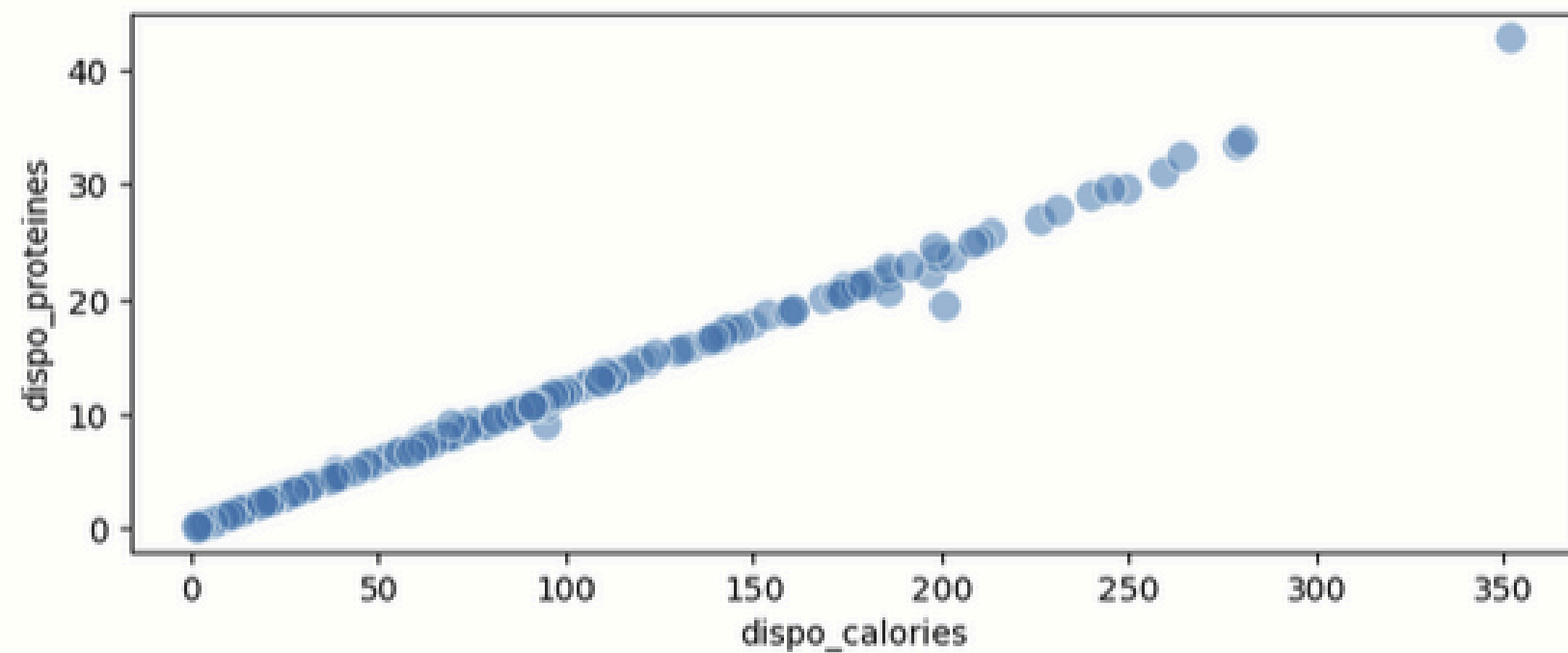
- Concaténation des différentes tables
- Pivot de la nouvelle table
- Renommage des colonnes

Traitement de la données

- Suppression des lignes aux valeurs manquantes
 - Suppression de la données de la France
 - on obtient 168 pays
- 

Analyse de données

Normalisation



Analyse de données

Classification hiérarchique

Définition

Théorique

La CAH :

- méthodes d'analyse exploratoire
- variables en clusters
- forme de dendrogramme

Distance

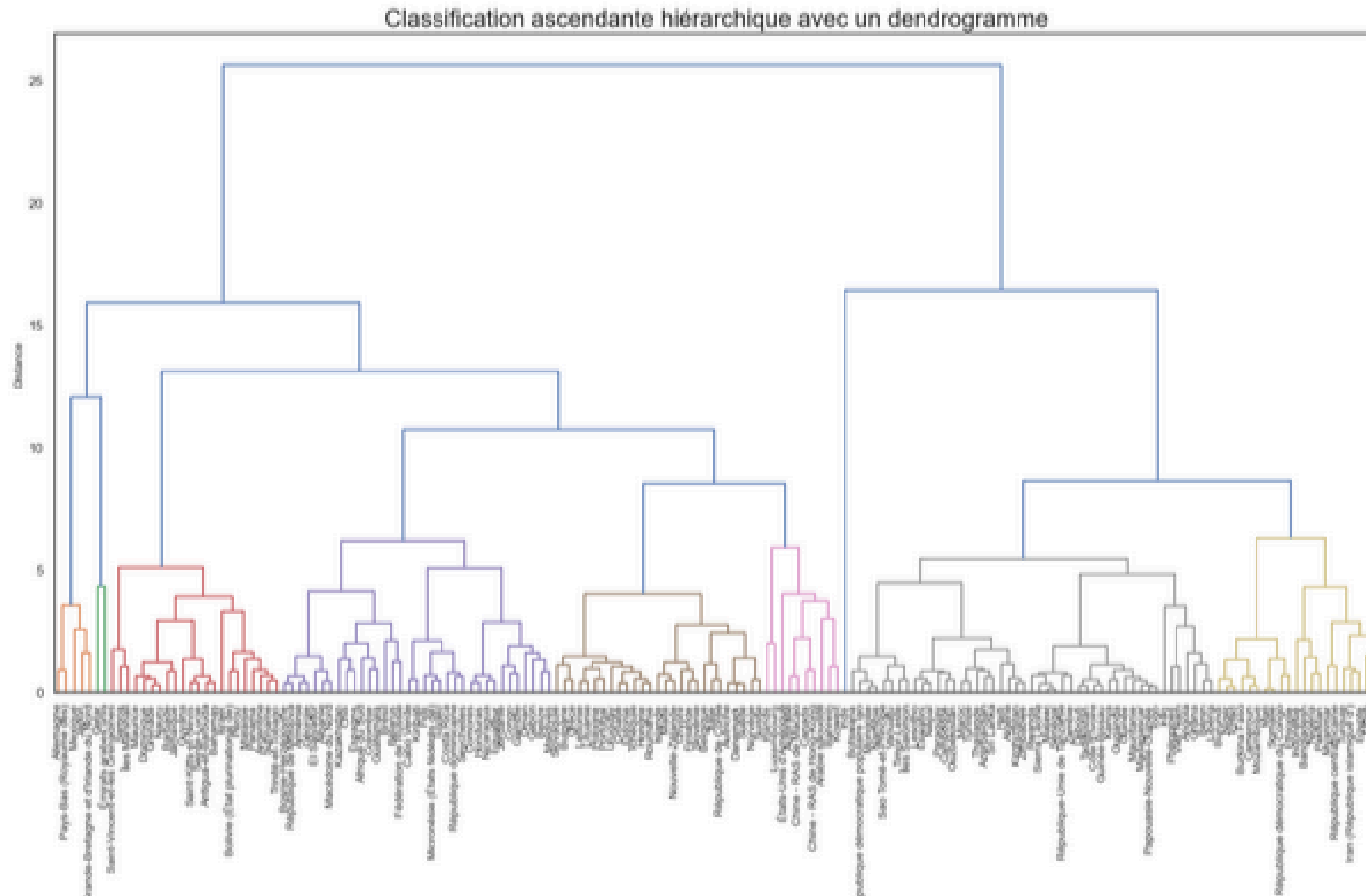
Le calcul des distances :

- distance euclidienne (géométrique)
- distance de Manhattan (somme des différences absolues)
- distances adaptées selon le contexte (cosinus, corrélation)
- fusion avec méthode de Ward

Analyse de données

Classification hiérarchique

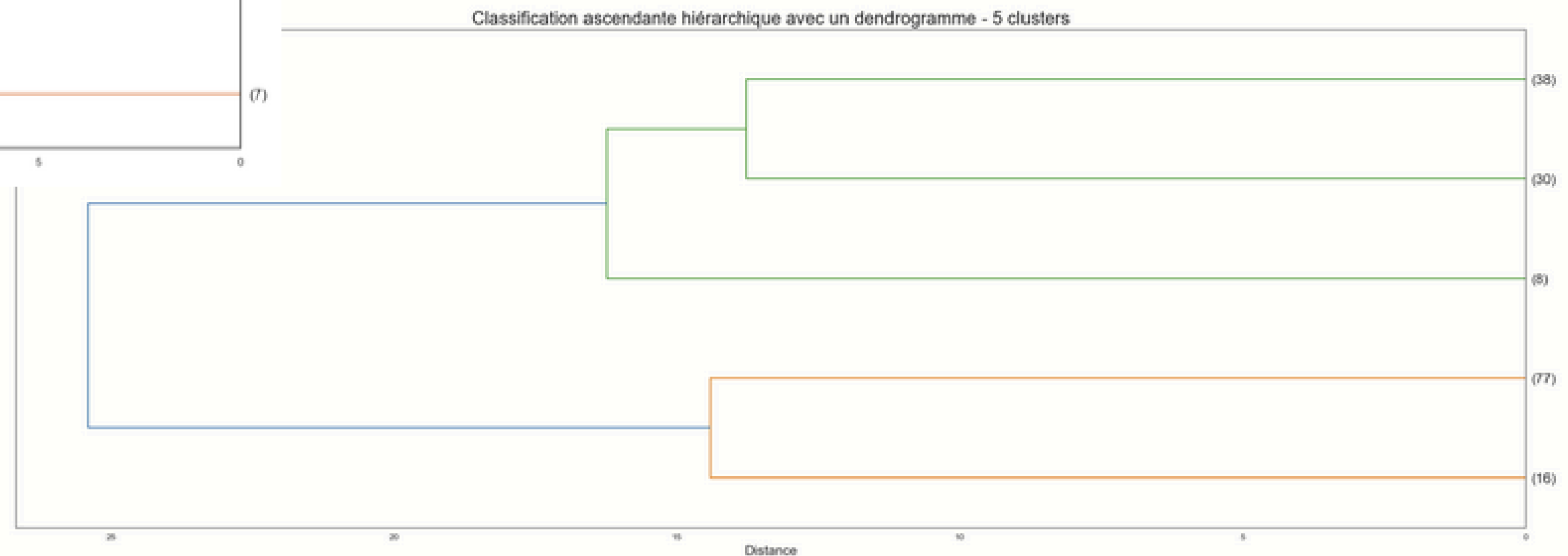
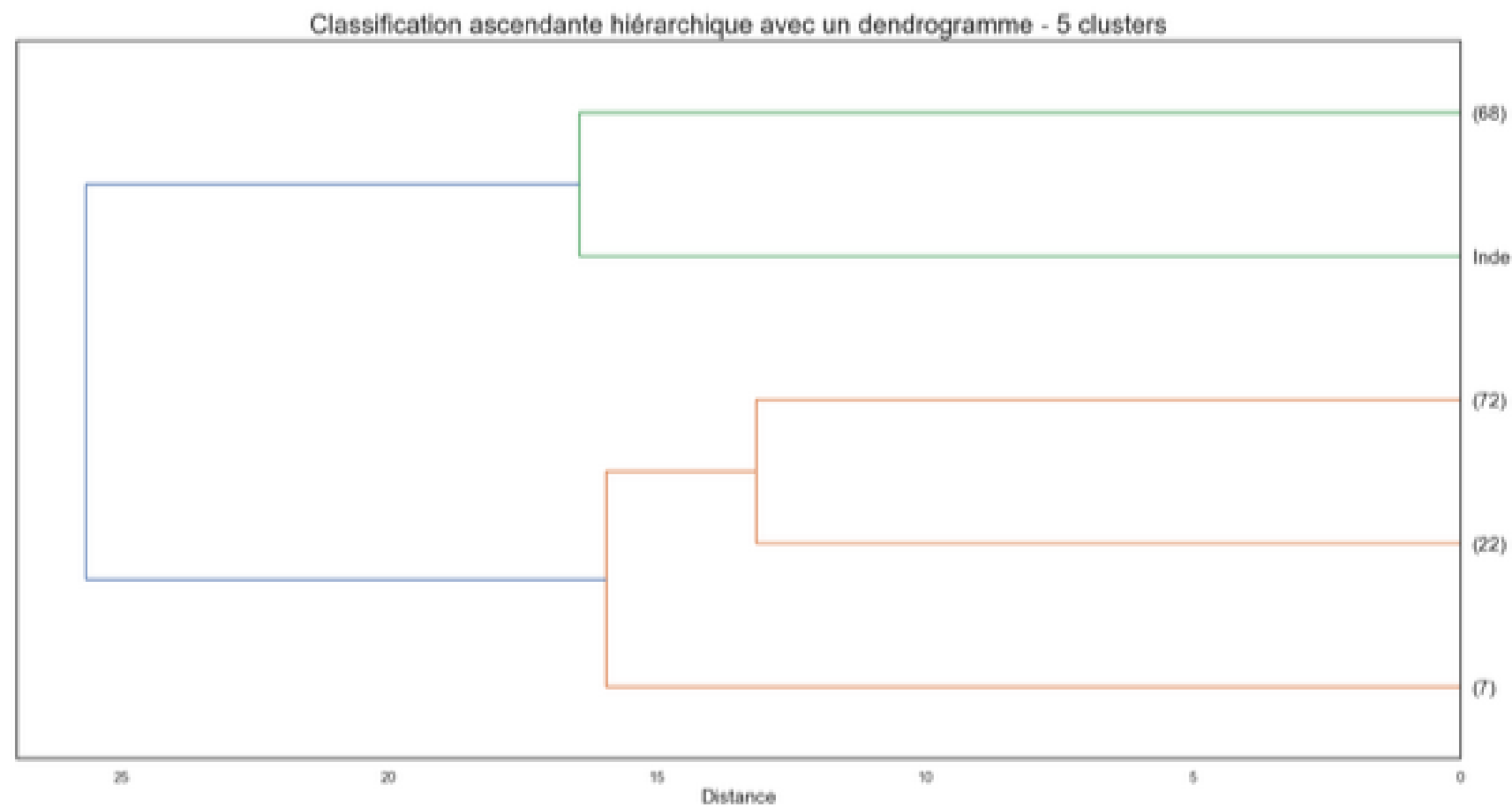
Mes données



Analyse de données

Classification hiérarchique

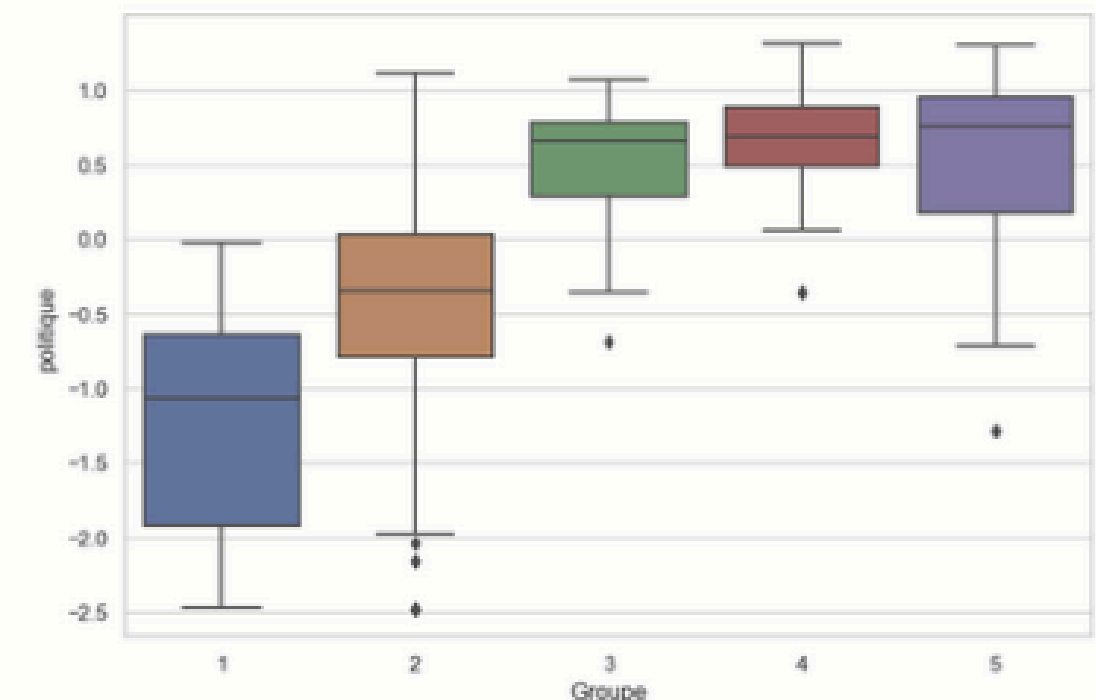
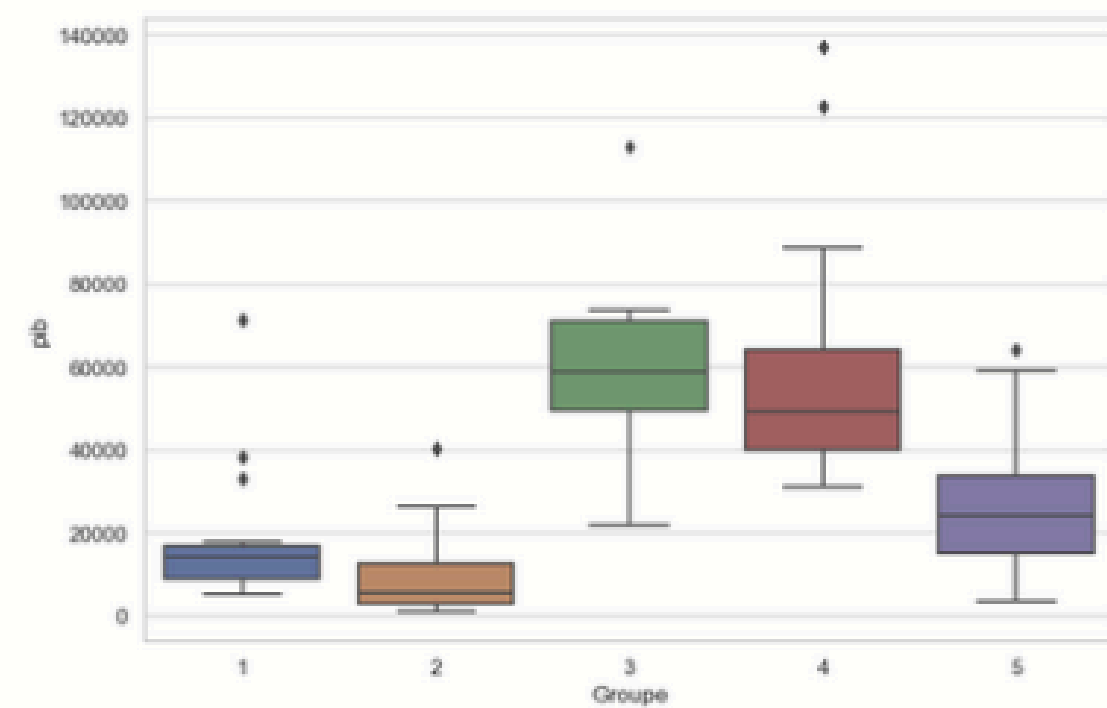
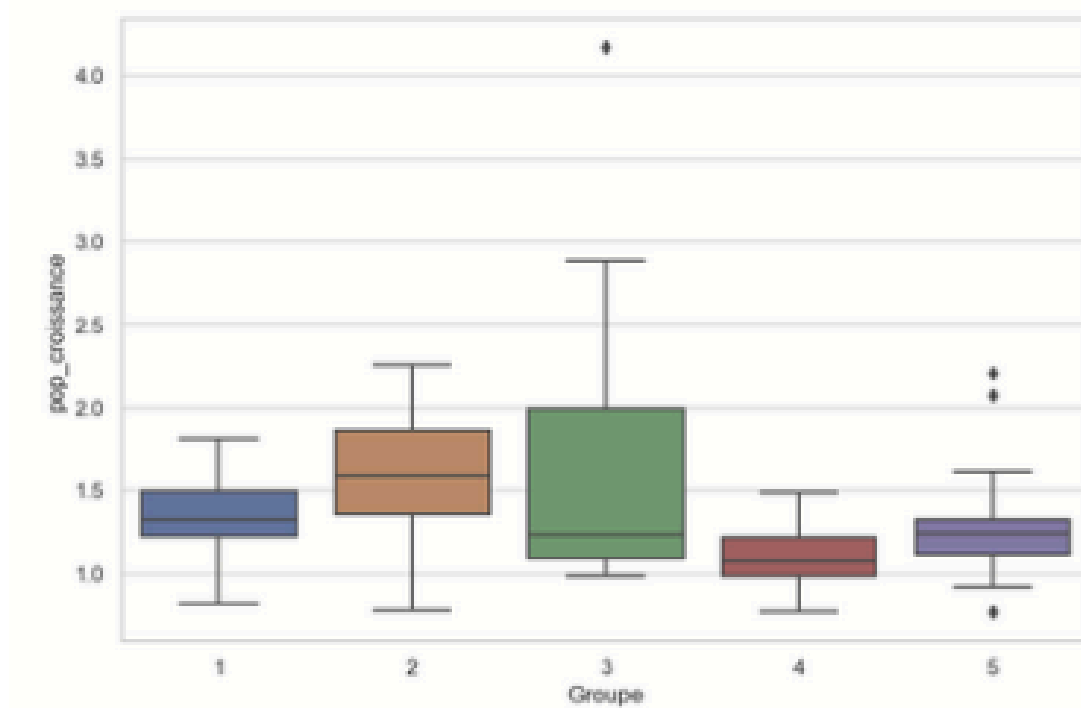
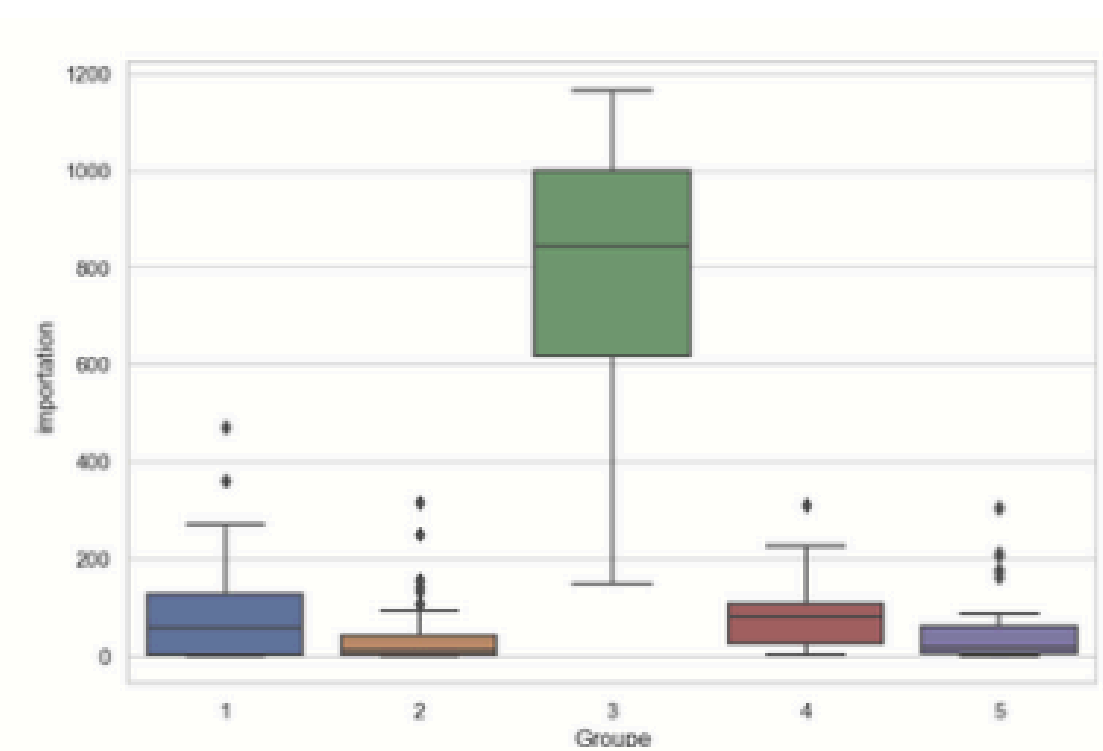
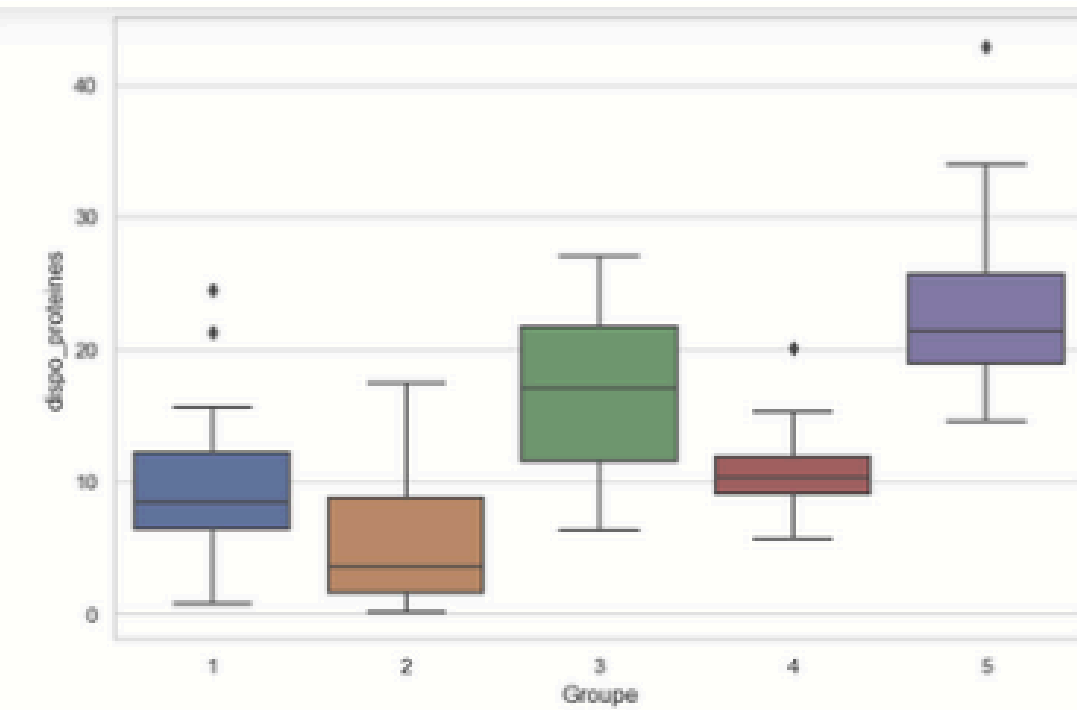
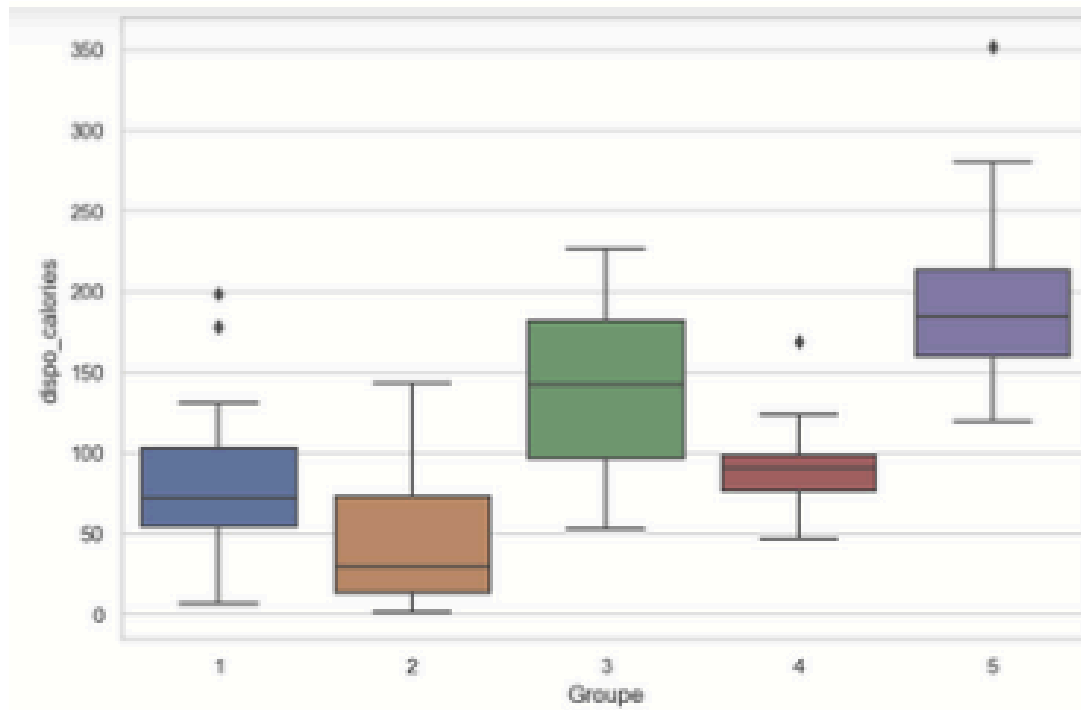
Regroupement de mes données



Analyse de données

Classification hiérarchique

Comparaison par boîtes à moustaches





Analyse de données

Algorithme K-mean

Définition

Théorique

La K-mean :

- un algorithme de clustering
- variables en clusters autour de centre de gravité
- pré définition du nombre de cluster

Distance

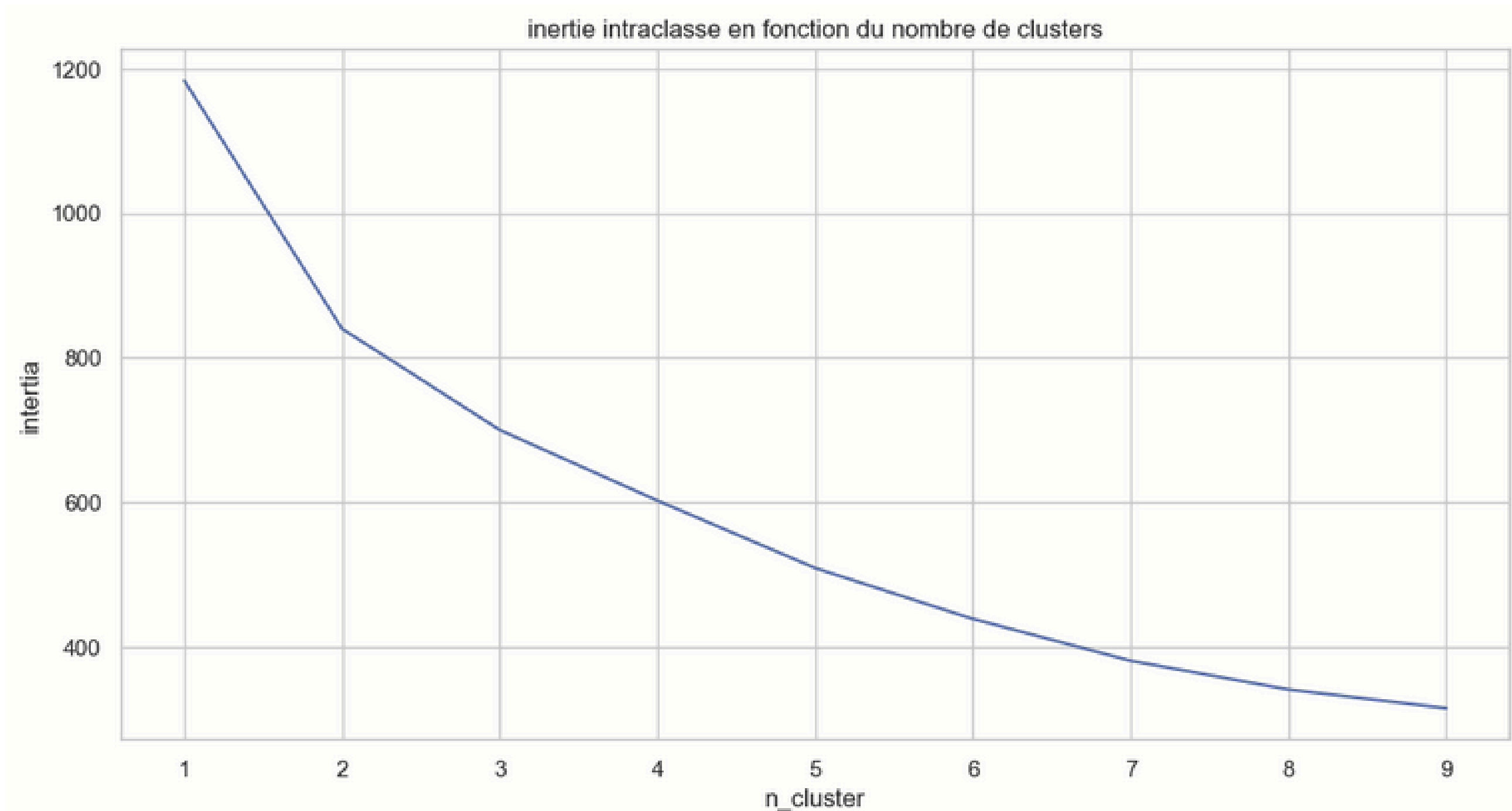
Le calcul des distances :

- distance euclidienne entre variables et centre de gravité
- minimisation des sommes des distance intra-cluster
- centre de gravité recalculés à chaque itération jusqu'à stabilisation

Analyse de données

Algorithme K-mean

Identification du nombre de clusters



Analyse de données

Algorithme K-mean

Attribution des clusters

	cluster	pays
0	d	Afrique du Sud
1	d	Albanie
2	a	Algérie
3	e	Allemagne
4	a	Angola
...
164	d	Équateur
165	c	États-Unis d'Amérique
166	a	Éthiopie
167	b	Îles Marshall
168	a	Îles Salomon

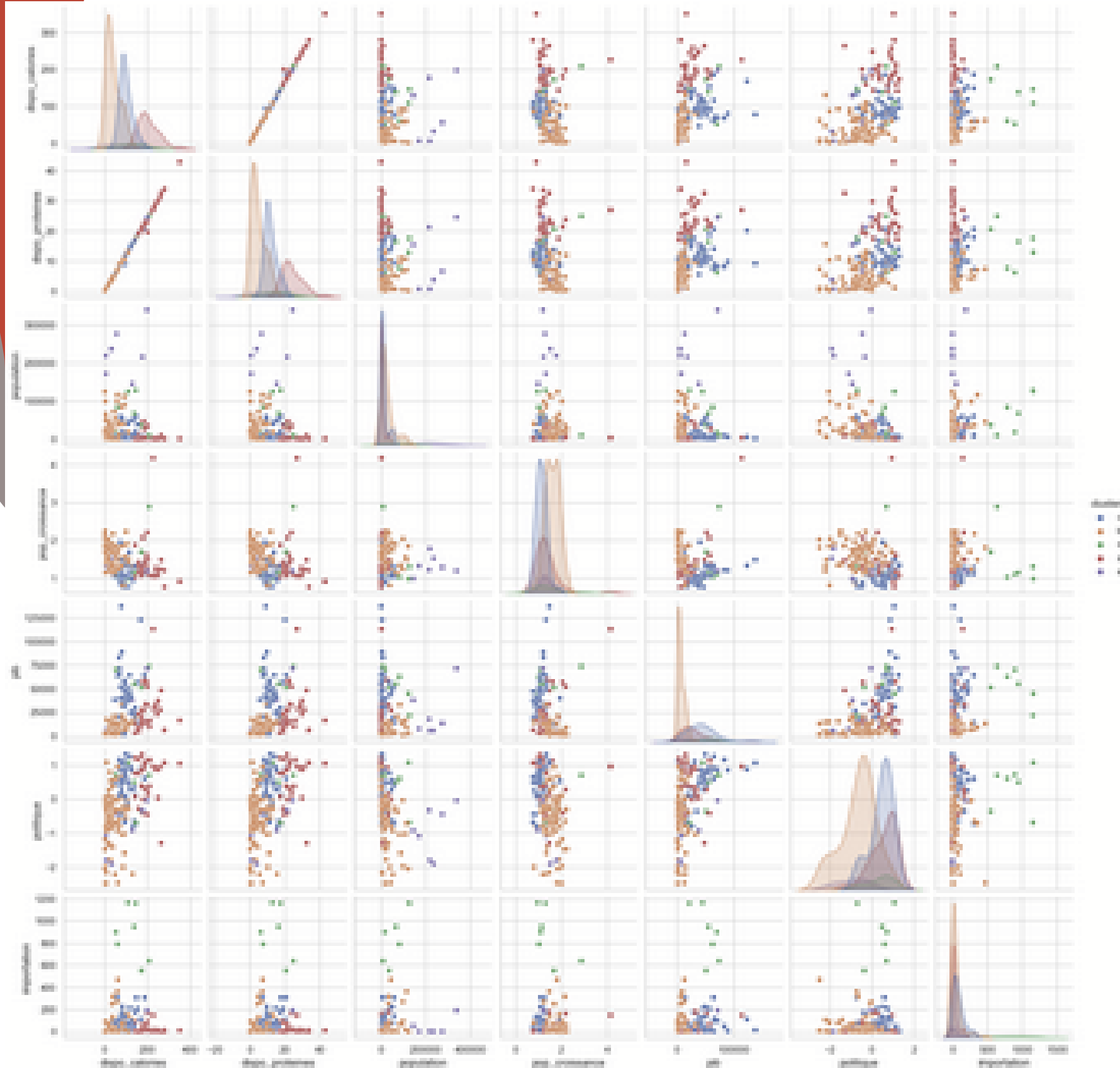
169 rows x 2 columns

	pays	dispo_calories	dispo_proteines	population	pop_croissance	pib	politique	importation	cluster
0	Afrique du Sud	149.24	17.86	59893.885	1.279421	14322.3	-0.72	304.0	d
1	Albanie	85.67	10.27	2842.321	0.893244	17261.0	0.11	47.0	d
2	Algérie	23.14	2.77	44803.225	1.458089	13080.9	-0.74	0.0	a
3	Allemagne	60.85	7.69	83369.843	1.022295	62605.5	0.61	789.0	e
4	Angola	39.62	4.75	35588.987	2.170846	7407.1	-0.65	316.0	a
...
164	Équateur	108.60	12.99	18001.000	1.425652	13530.4	-0.24	1.0	d
165	États-Unis d'Amérique	198.23	24.52	338289.857	1.197916	71308.9	-0.04	191.0	c
166	Éthiopie	1.49	0.18	123379.924	1.840616	2698.6	-2.04	0.0	a
167	Îles Marshall	280.32	33.92	41.569	0.766616	6625.5	1.06	3.0	b
168	Îles Salomon	37.84	4.43	724.273	1.684442	2480.3	0.38	7.0	a

Analyse de données

Algorithme K-mean

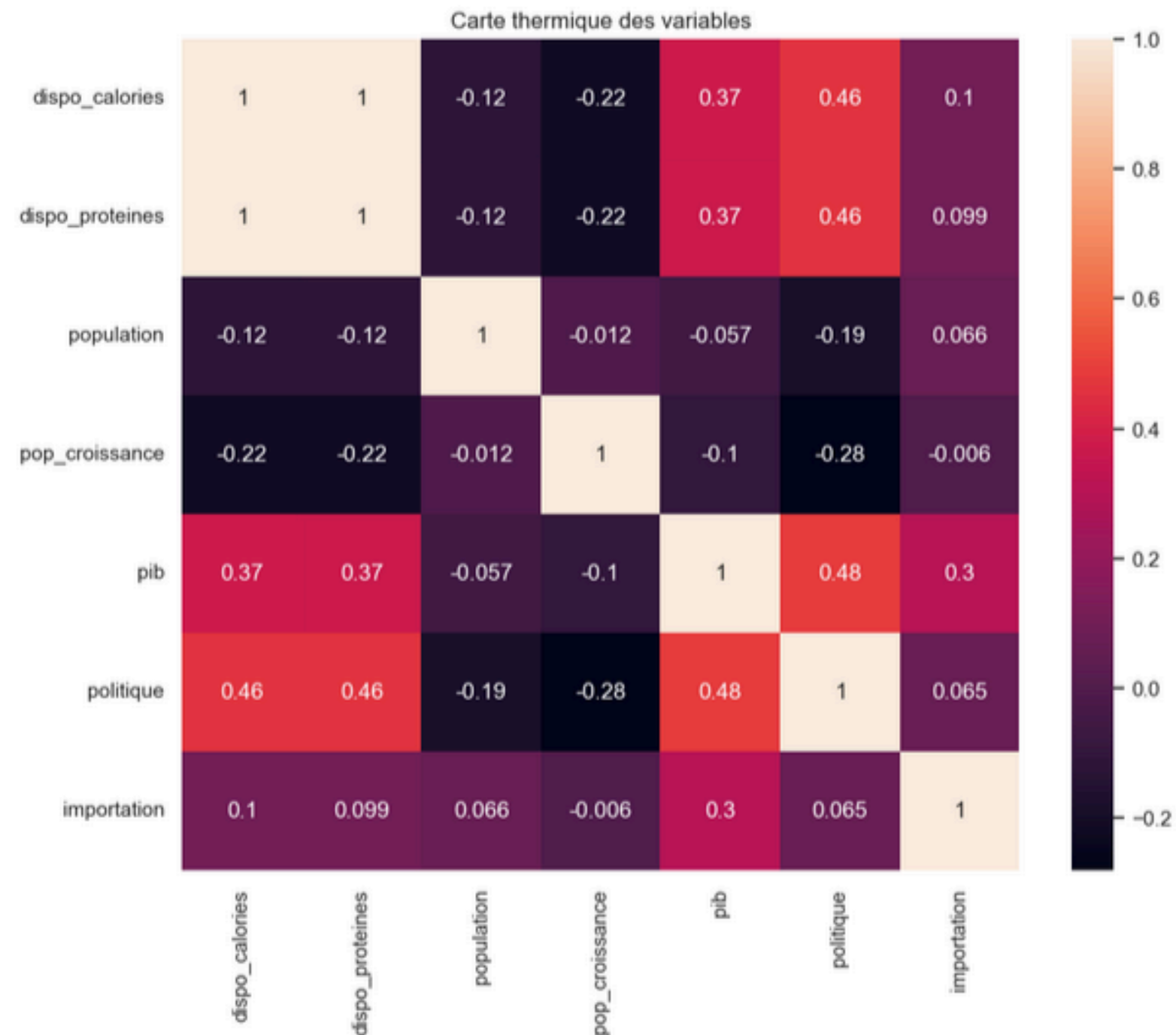
Comparaison par nuages de points



Analyse de données

Algorithme K-mean

Identification des corrélation par cartes thermiques



Analyse de données

ACP

Définition

Théorique

La ACP :

- méthode de réduction de dimensionnalité
- composantes principales
- capture de la variance

Distance

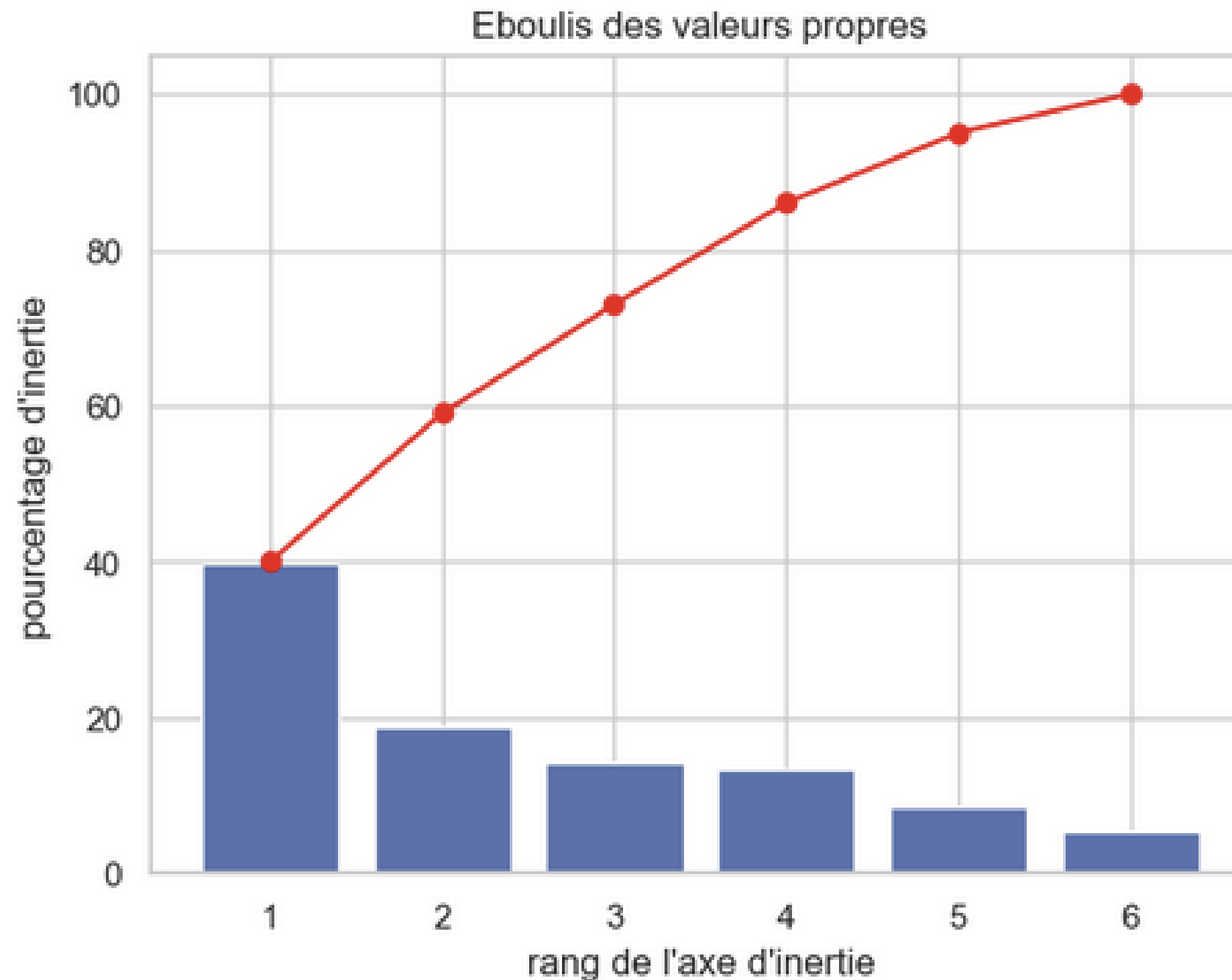
Le calcul des distances :

- distance indirectement
- entre variable et composantes principales

Analyse de données

ACP

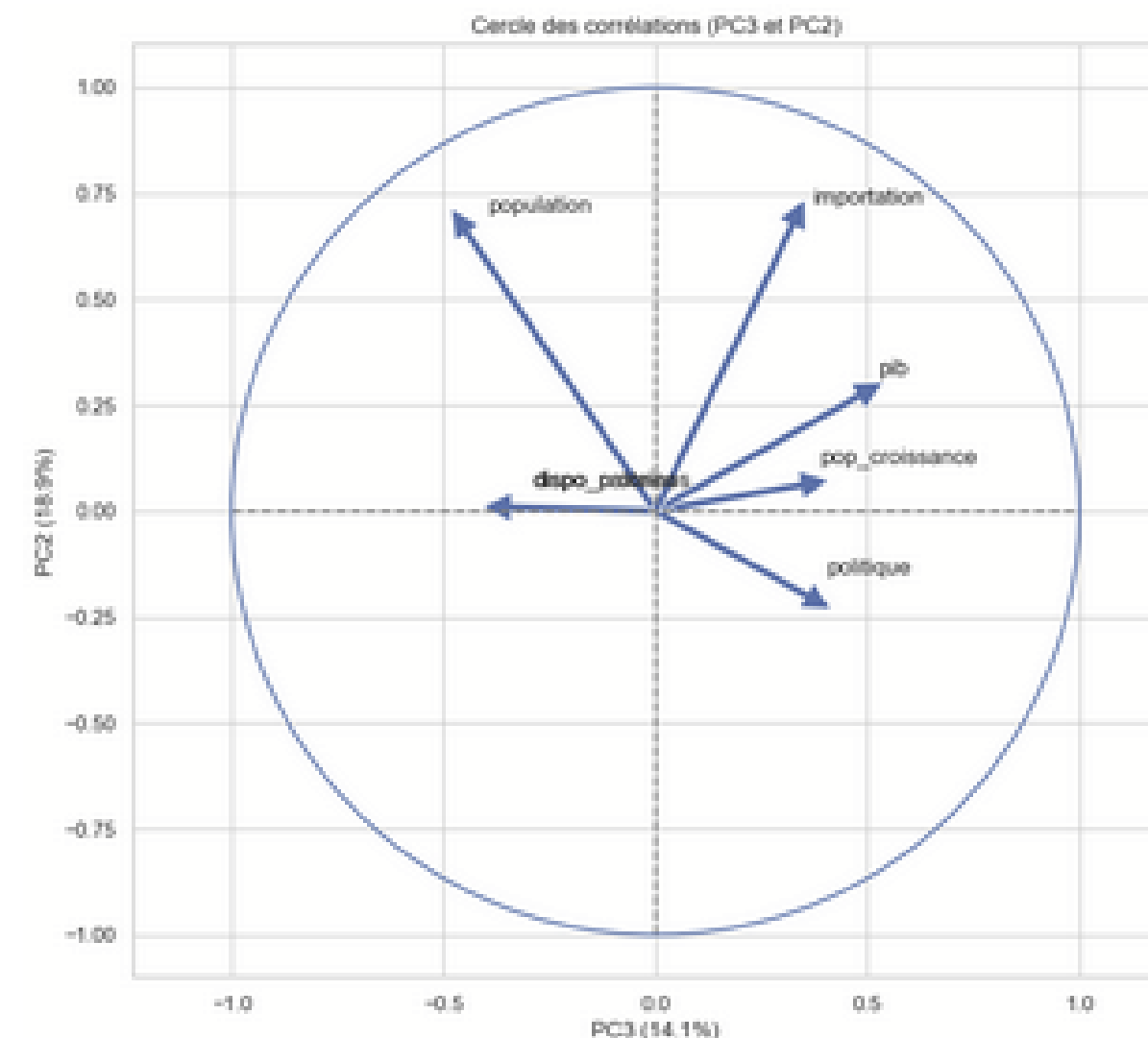
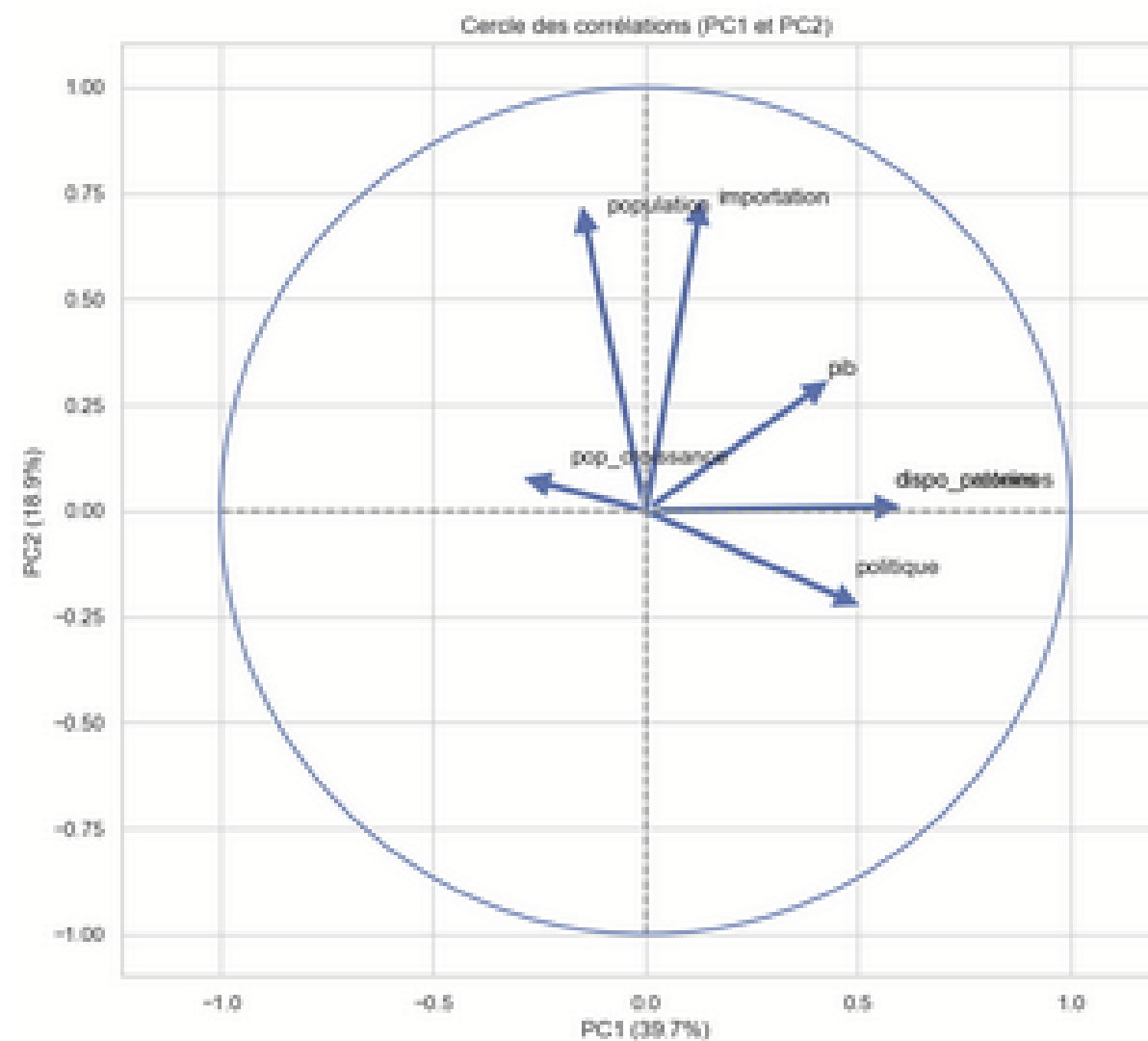
Mes données



Analyse de données

ACP

Analyse via cercle de corrélation



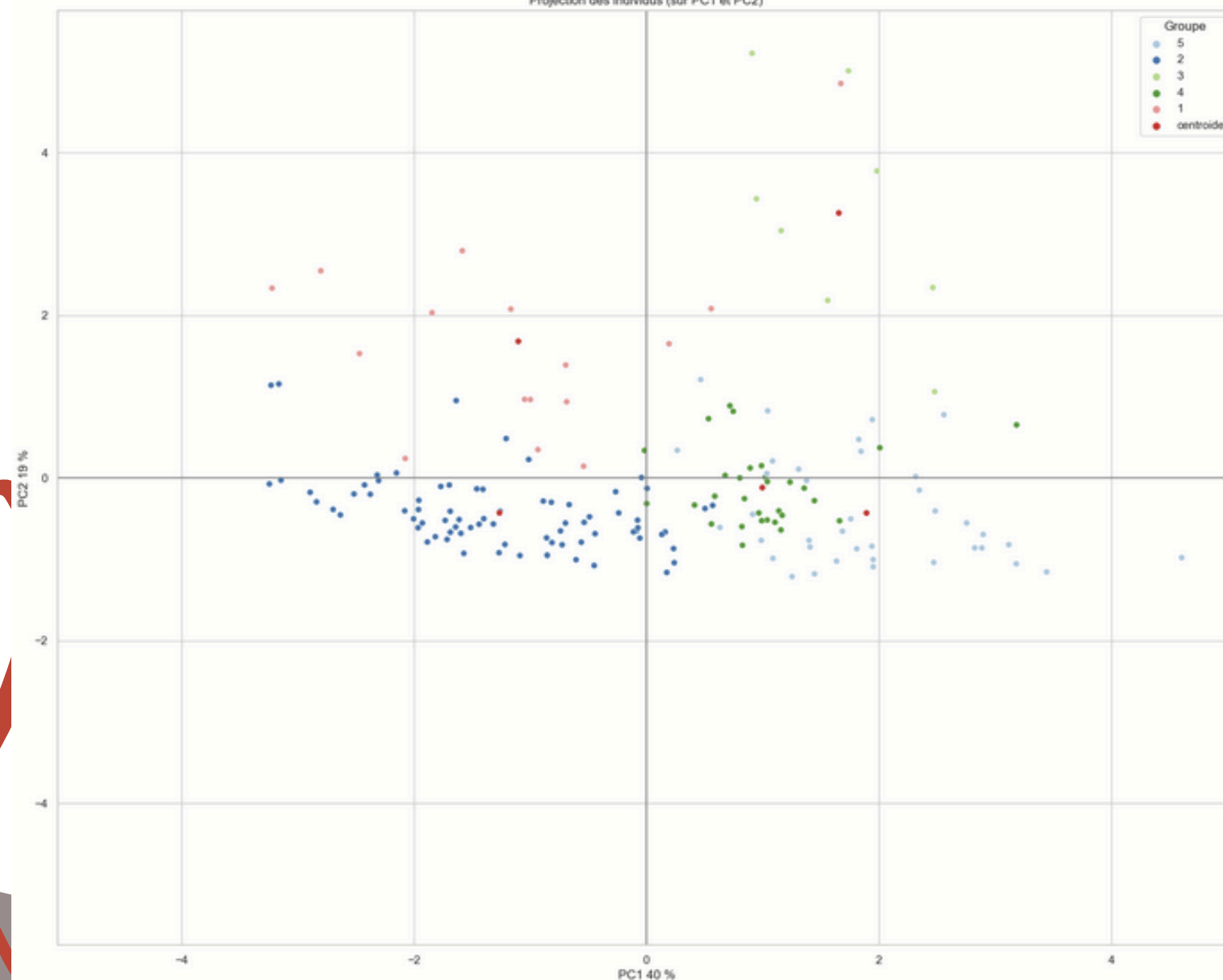
Analyse de données

ACP

Visualisation des données par méthodes

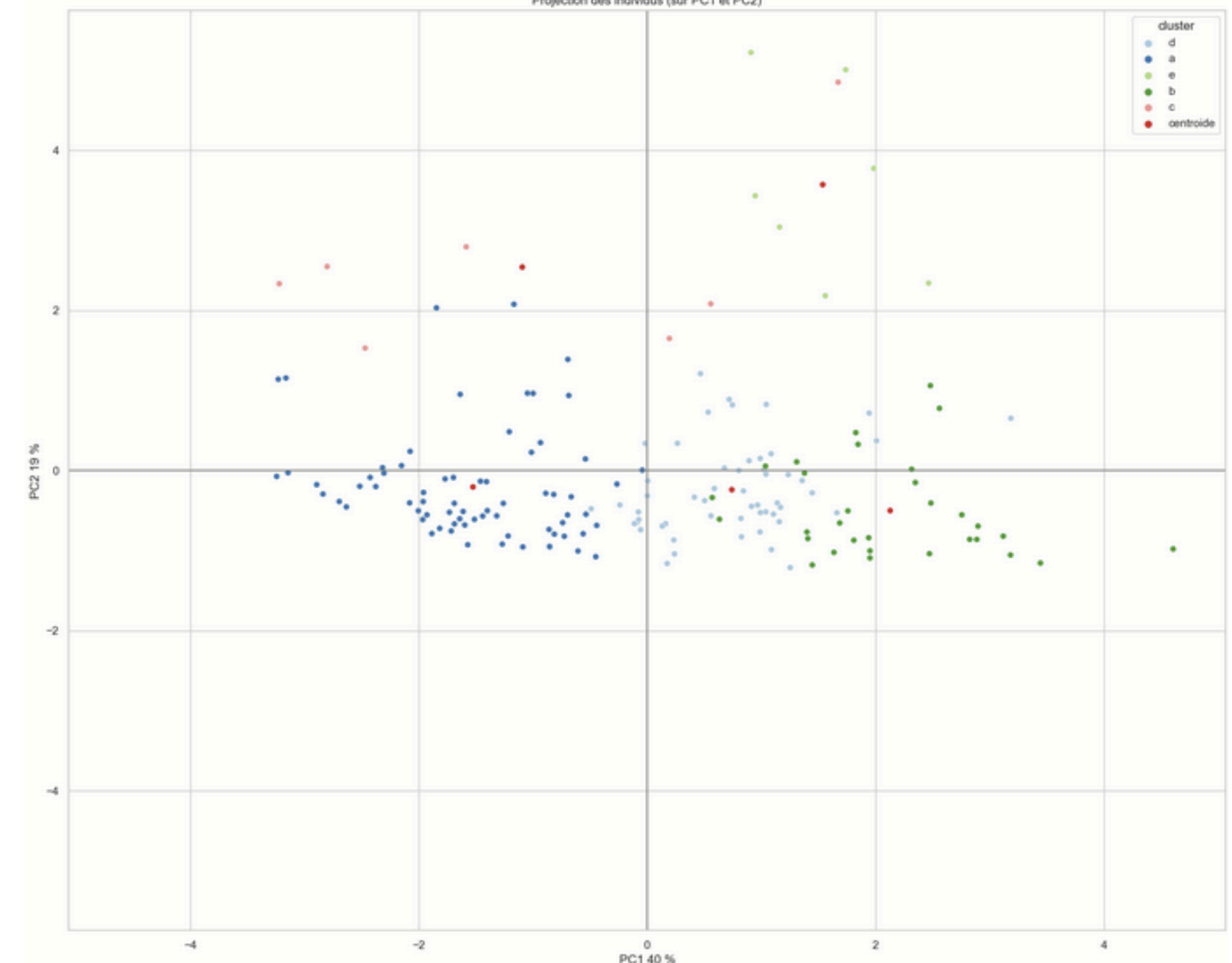
Avec CAH

Projection des individus (sur PC1 et PC2)



Avec K-mean

Projection des individus (sur PC1 et PC2)



Conclusion

Via la classification hiérarchique

Les groupes intéressants ont été le 3 en priorité puis le 4

Via l'algorithme K-mean

Les groupes intéressants ont été le d en priorité puis le c

Via l'ACP

Les groupes cités au-dessus sont les mêmes c'est deux groupes sont donc confirmés.

La liste des pays à développer en priorité

- Allemagne
- Arabie saoudite
- Japon
- Mexique
- Pays-Bas (Royaume des)
- Royaume-Uni de Grande-Bretagne et d'Irlande du Nord
- Émirats arabes unis

La liste des pays à développer dans un second temps

- | | | | |
|------------|--------------|-----------------------|------------|
| • Autriche | • Hongrie | • Norvège | |
| • Belgique | • Irlande | • Nouvelle-Zélande | |
| • Bulgarie | • Islande | • Pologne | |
| • Chypre | • Italie | • Portugal | • Suède |
| • Croatie | • Kazakhstan | • Roumanie | • Tchéquie |
| • Danemark | • Lettonie | • République de Corée | • Uruguay |
| • Estonie | • Lituanie | • Slovaquie | |
| • Finlande | • Luxembourg | • Slovénie | |
| • Grèce | • Malte | • Suisse | |



Merci pour
votre attention