

Produisez une étude de marché avec R ou Python

Projet 9 - OpenClassroom

Jules-Yann MILLERET - Janvier 2023

Présentation

Produisez une étude de marché avec R ou Python

1. Objectif du projet
2. Création du jeu de données
3. Analyse en composantes principales
4. Le clustering
5. Caractérisation des clusters
6. Résultats - Cartes
7. Conclusion

Objectif du projet

Produisez une étude de marché avec R ou Python

- Entreprise de vente de poulet qui souhaite se développer à l'international.
- Nous devons proposer un groupe de pays qui pourrait intéresser l'entreprise pour s'implanter.
- Enfin nous devons réaliser une analyse plus poussée pour proposer une liste de quelques pays.
- Algorithmes de clustering (CAH & K-means) et réaliser une ACP.

Création du jeu de données

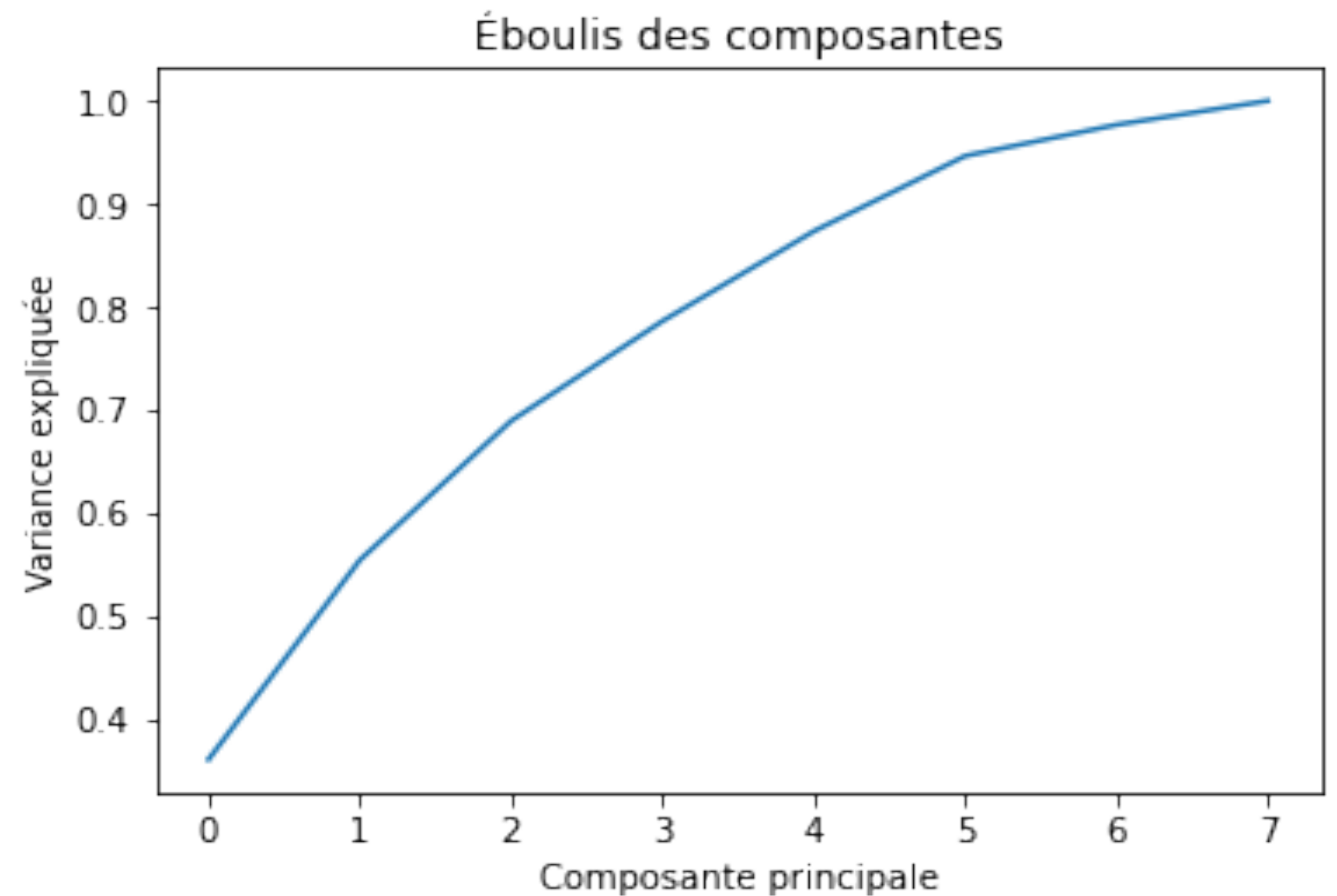
- Méthode PESTEL : Distance, Stabilité politique, Population, PIB
- Création d'un scrapper pour les distances
- Variables : Importations*, Nourriture*, Production*, Population, Variation de la population (2012 - 2017), Distance, PIB*, Stabilité
- « Correction » de la table créé (valeurs manquantes, normalisation)

* par habitant

ACP

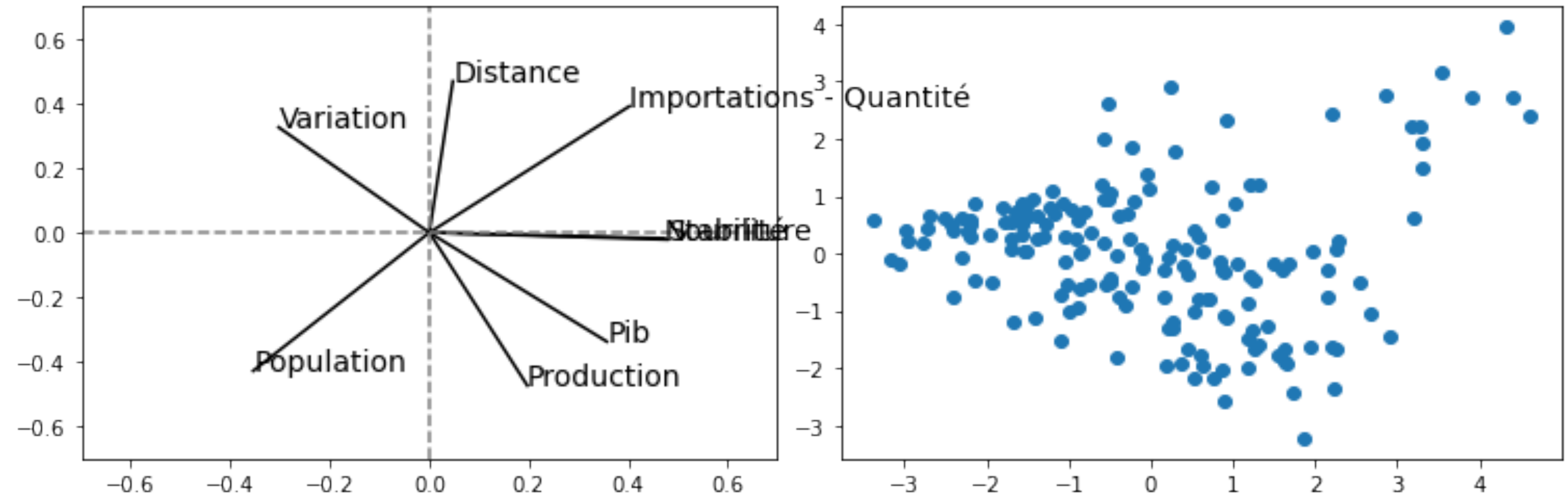
Analyse en composantes principales

- Résumé l'information en réduisant le nombre de variables
- Éboulis des composantes pour le nombre de composantes optimales
- Permet de réduire les temps de calcul
- Cercle des corrélations



ACP

Produisez une étude de marché avec R ou Python



Le clustering

Le clustering est le fait de regrouper les individus d'un jeu de données dans des groupes homogènes, c'est à dire qui partage des caractéristiques communes.

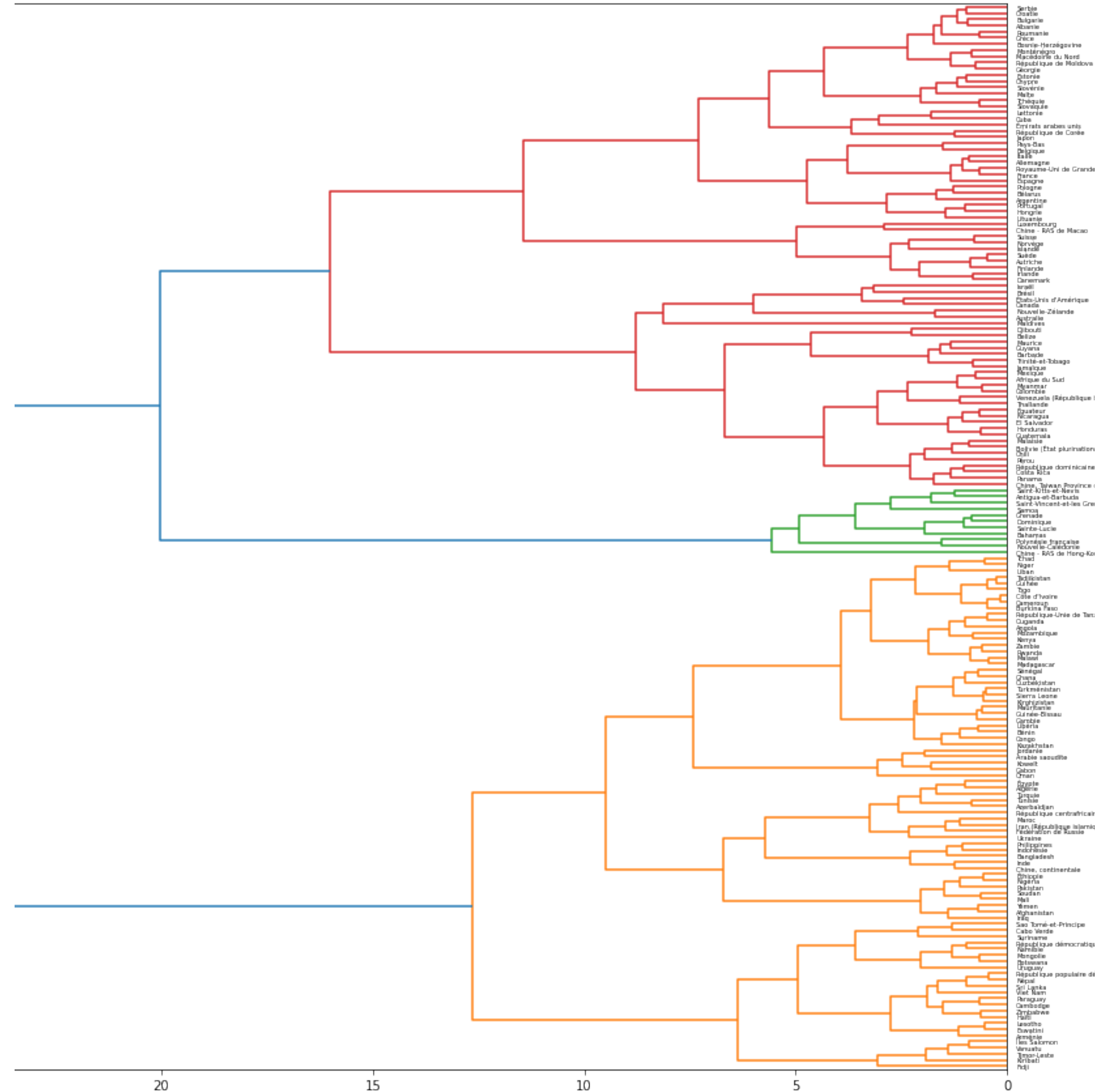
Deux algorithmes :

- Classification ascendante hiérarchique
- K-means

Clustering

Classification ascendante hiérarchique

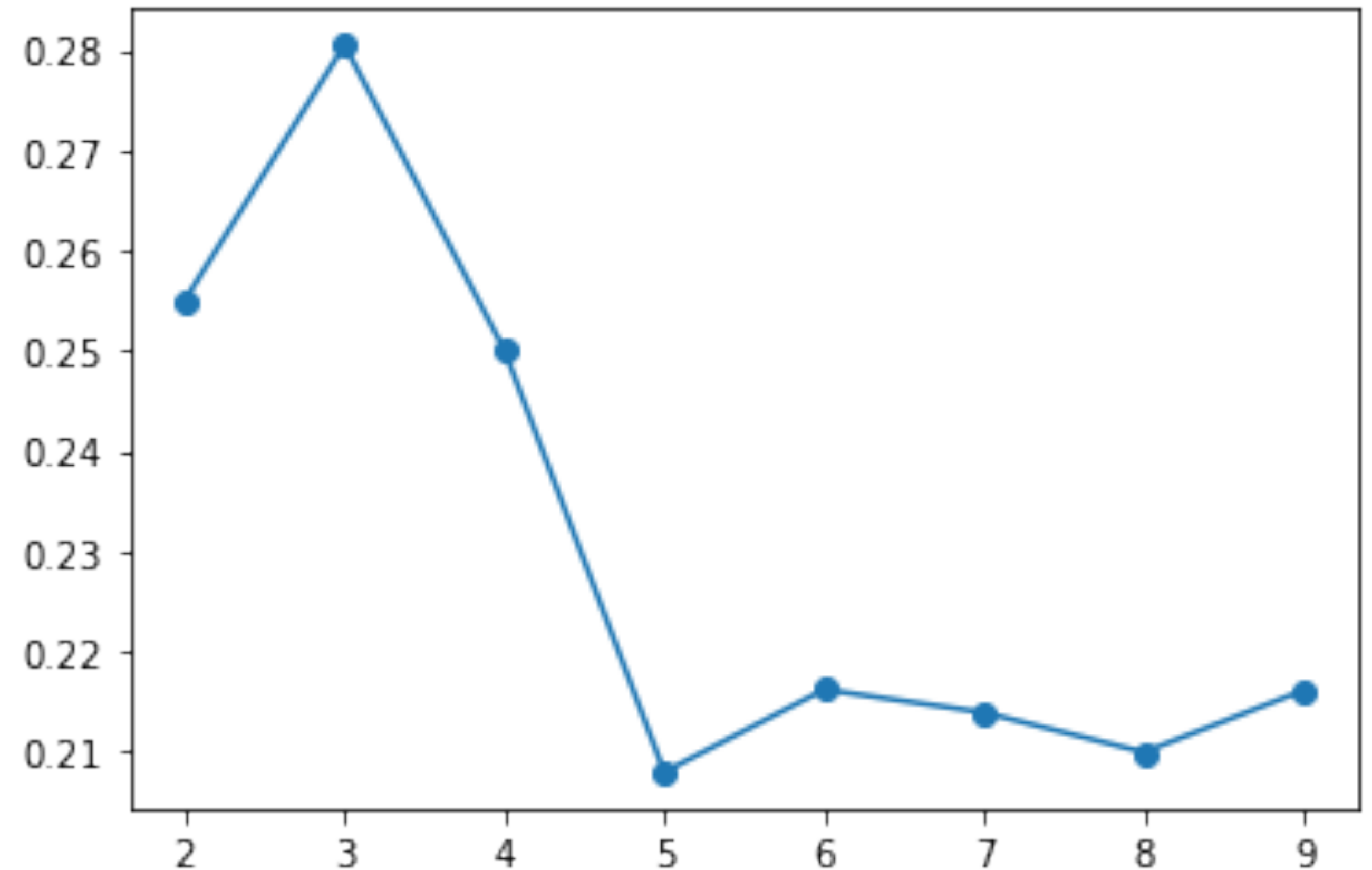
- Fonctionnement : chaque point est un cluster puis on les regroupe pour former les clusters
- Dendrogramme
- Permet de choisir le nombre de clusters
- Pas adapté à un grand volume de données



Clustering

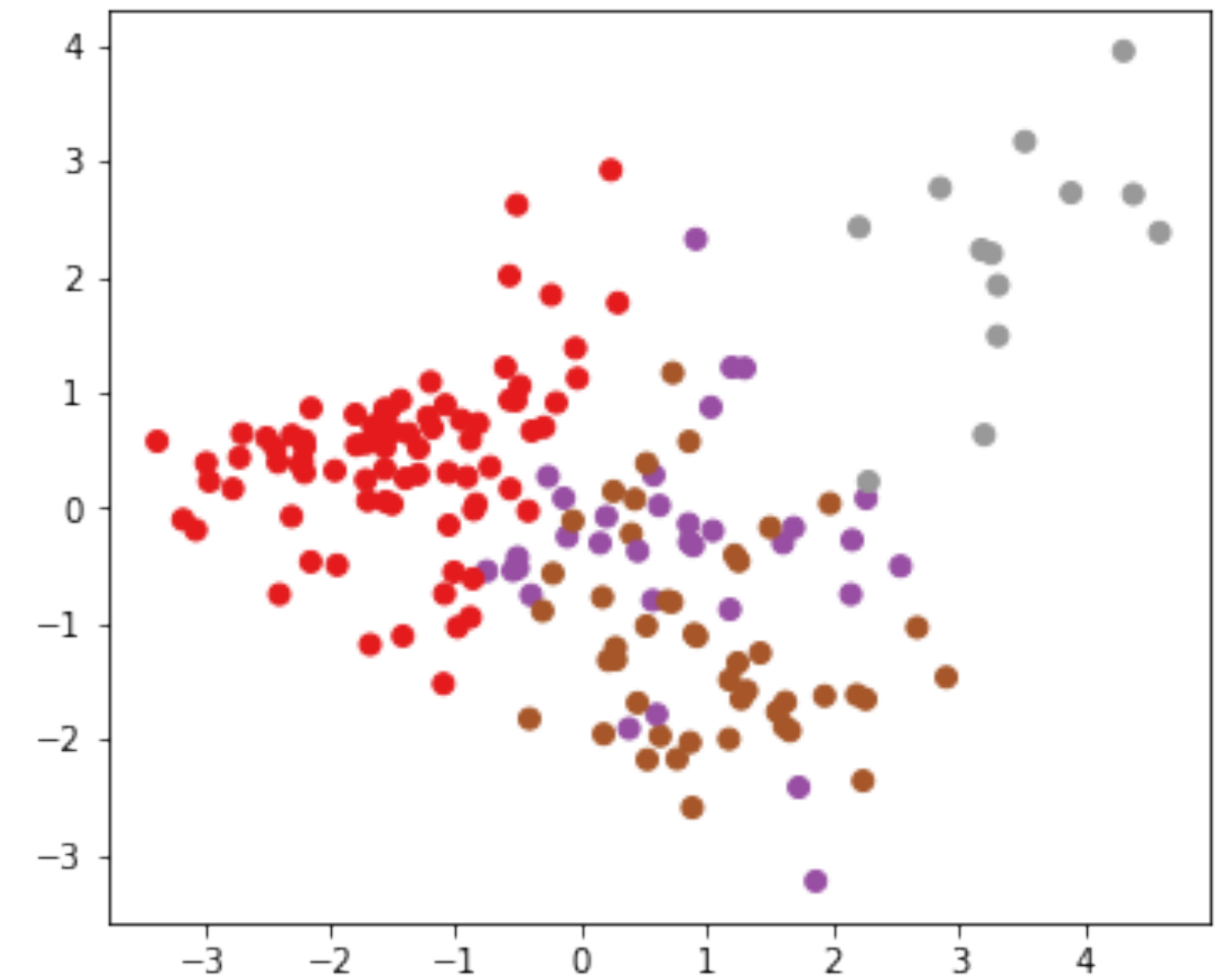
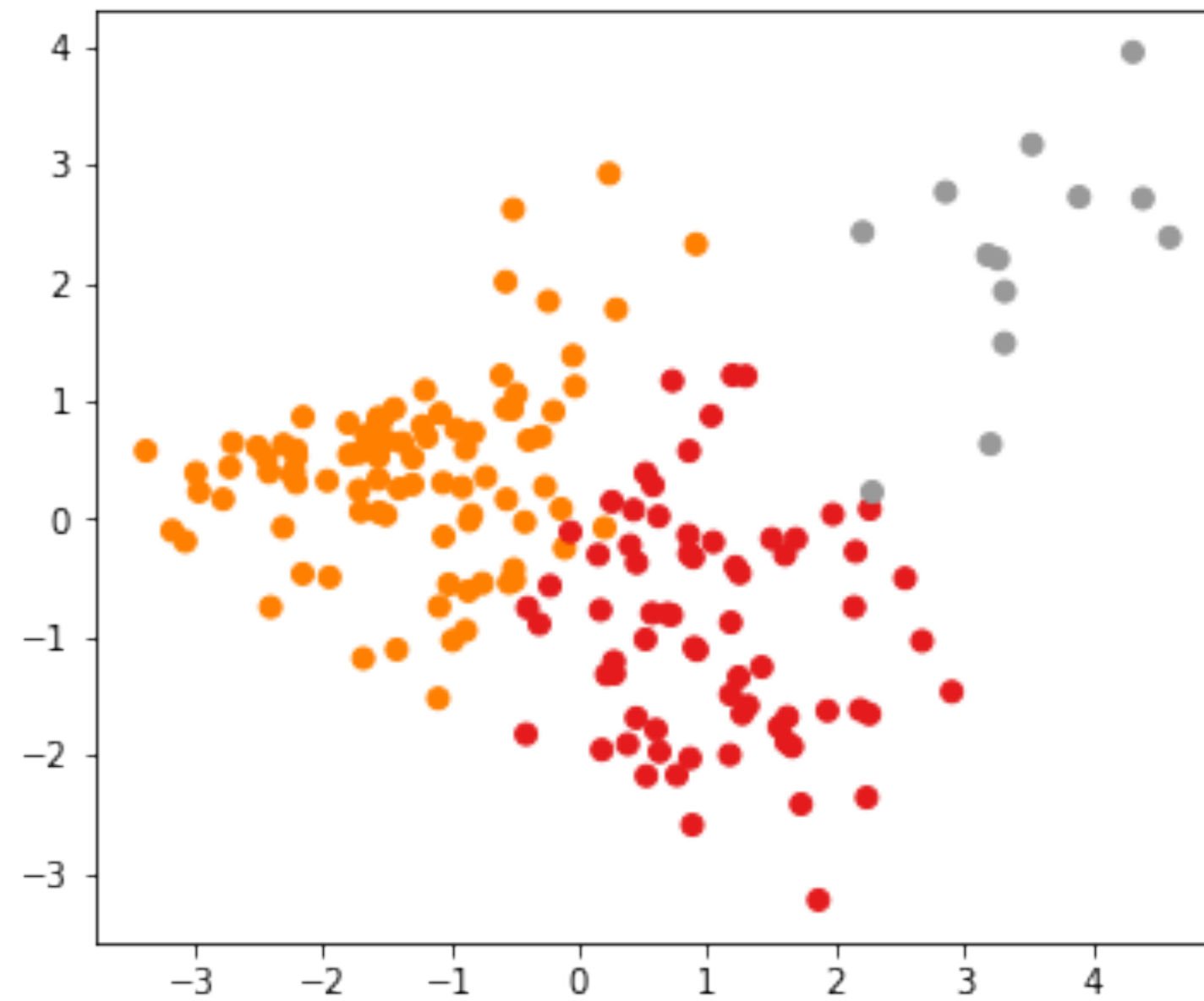
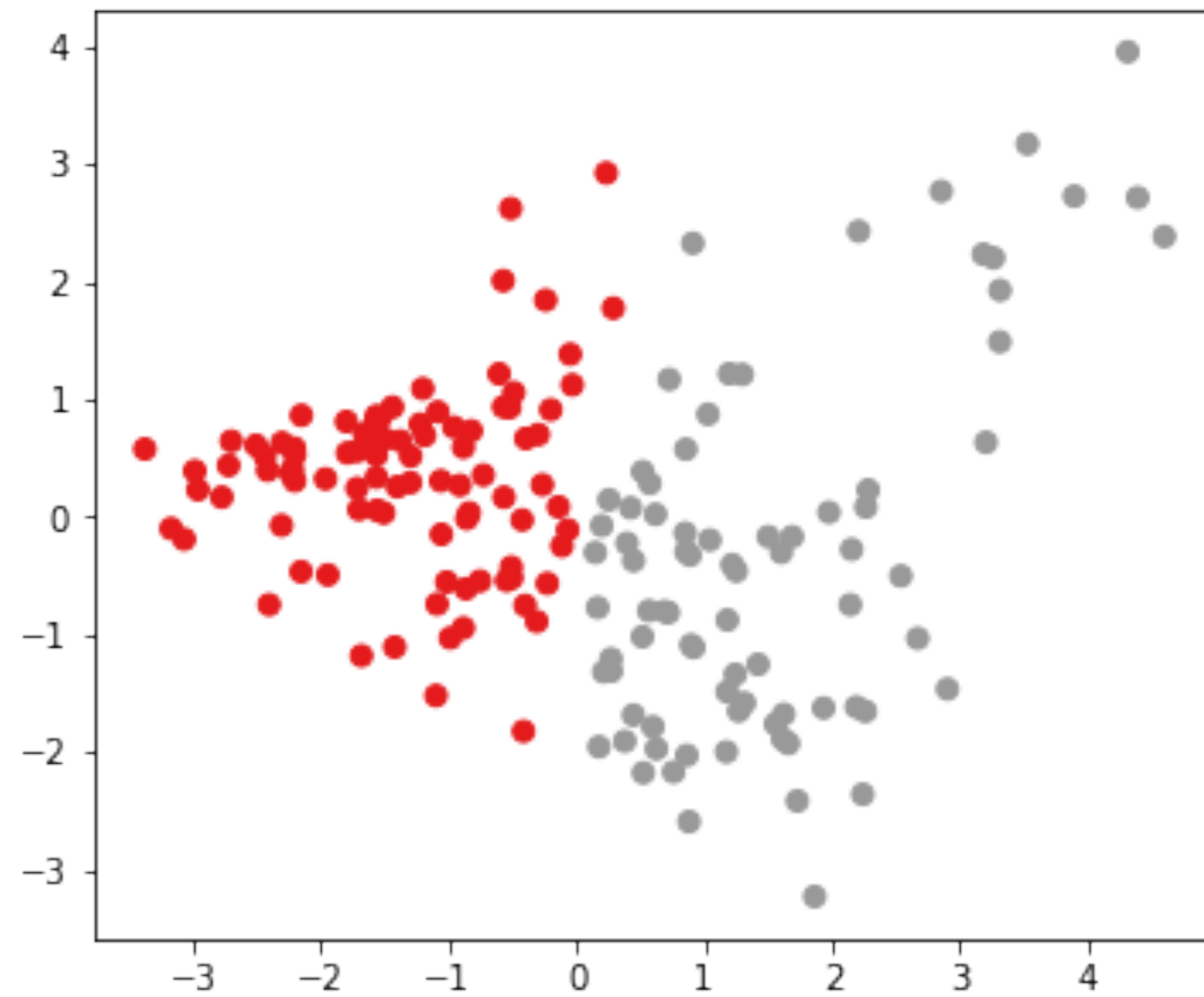
K-means

- Fonctionnement : minimiser la distance individu - centre de gravité
- Initialisation aléatoire résultats différents entre 2 exécutions
- Faible temps de calcul
- Choix du nombre de cluster
- Coefficient de silhouette



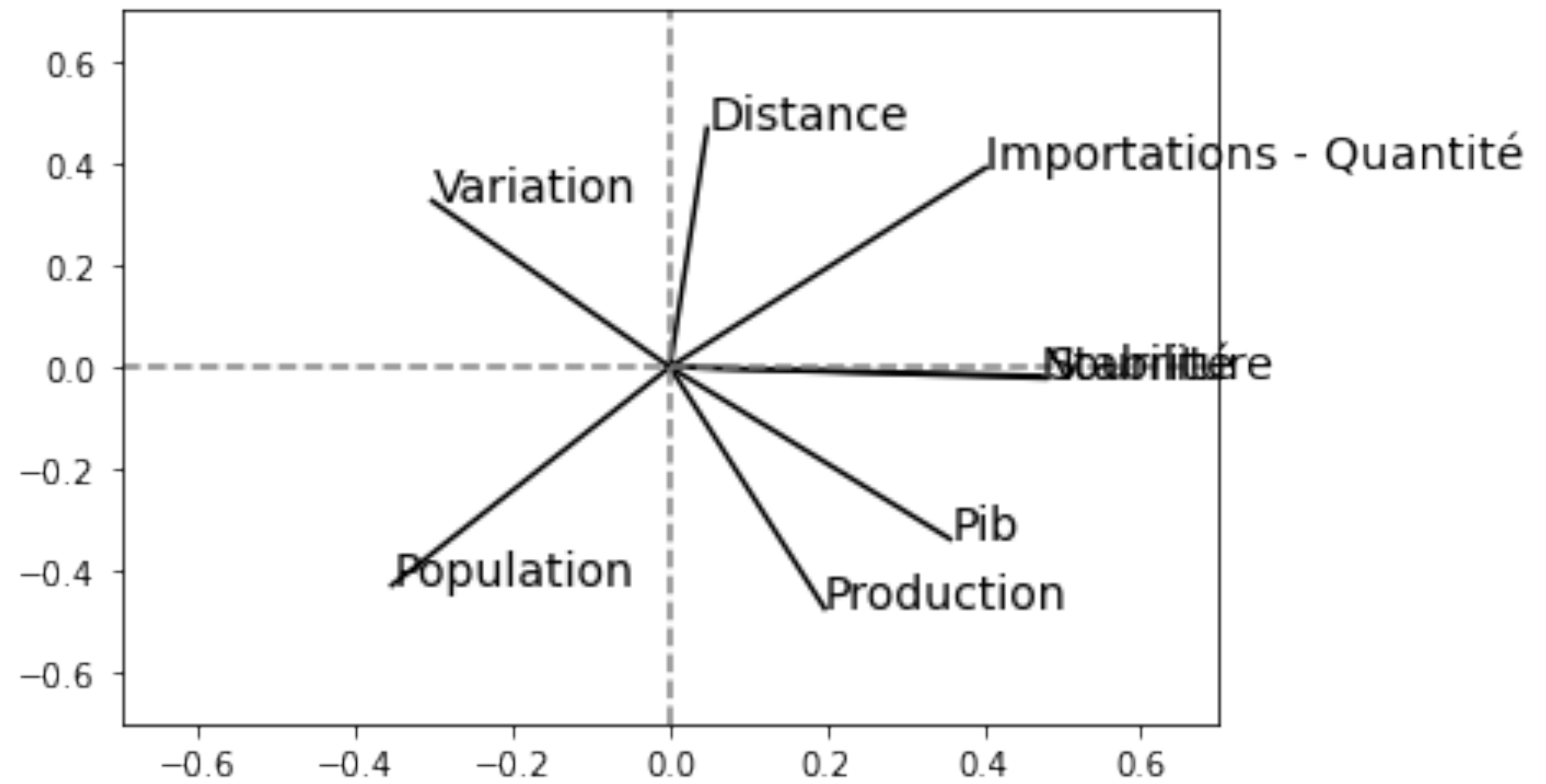
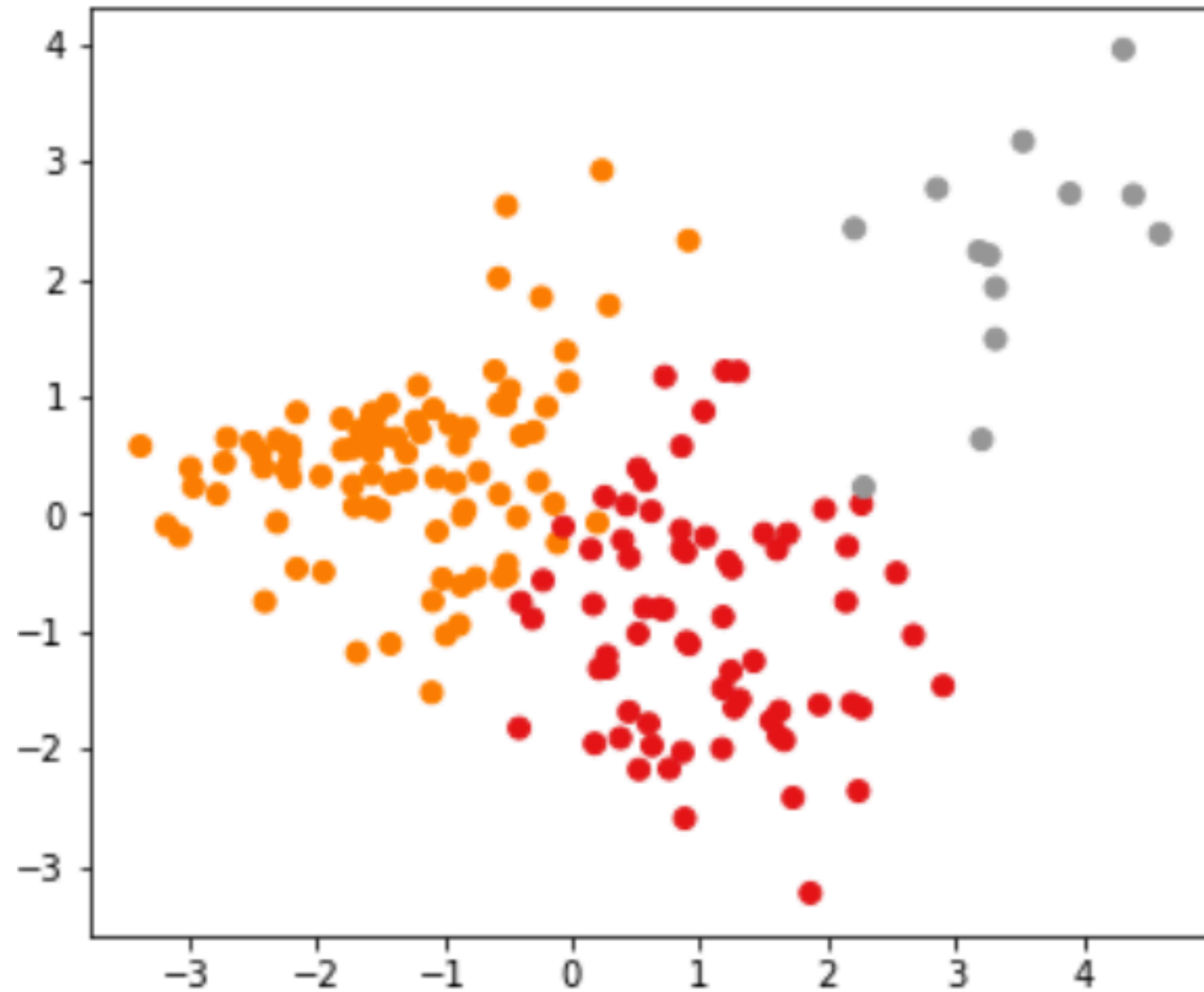
Clustering

K-means



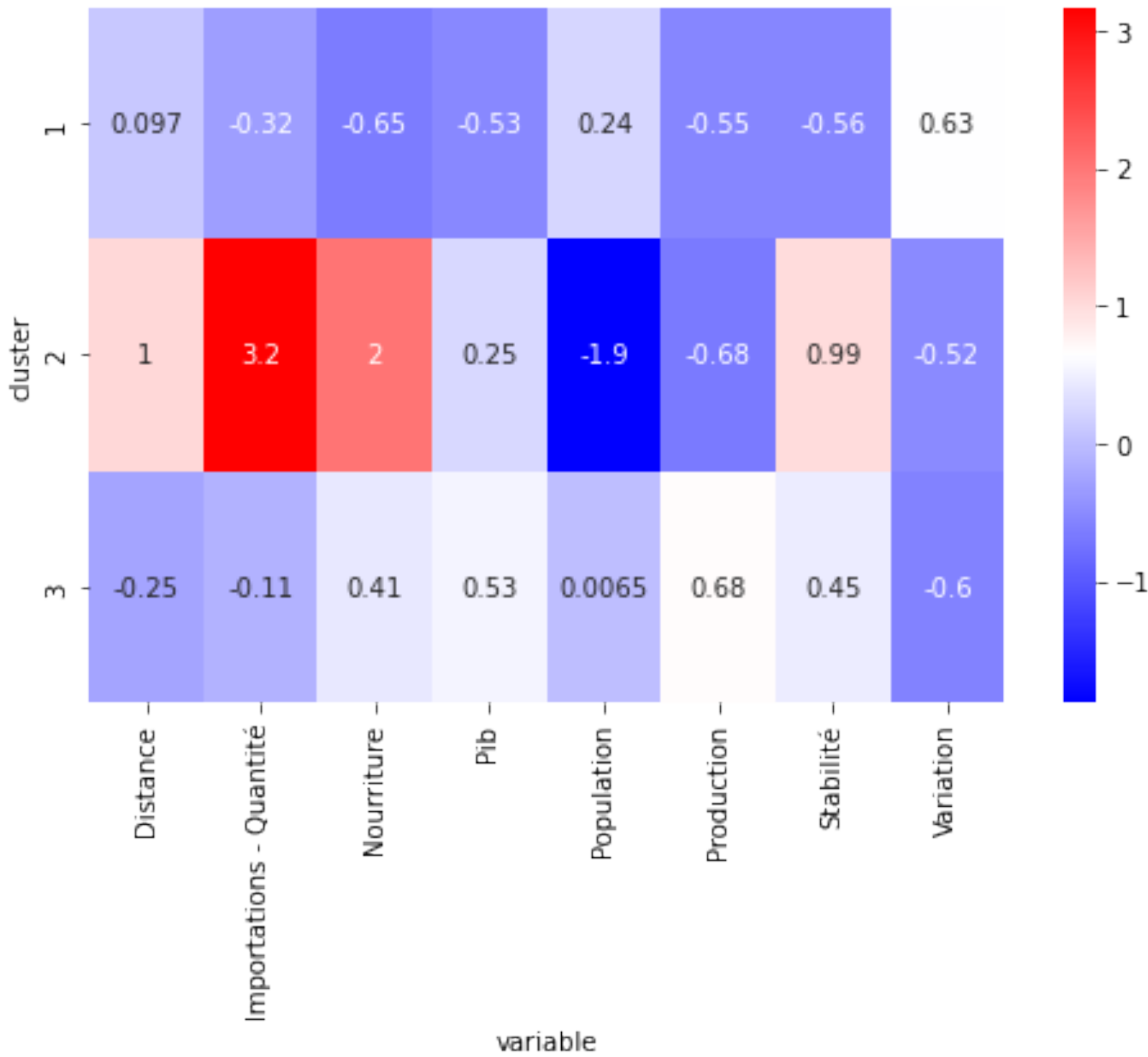
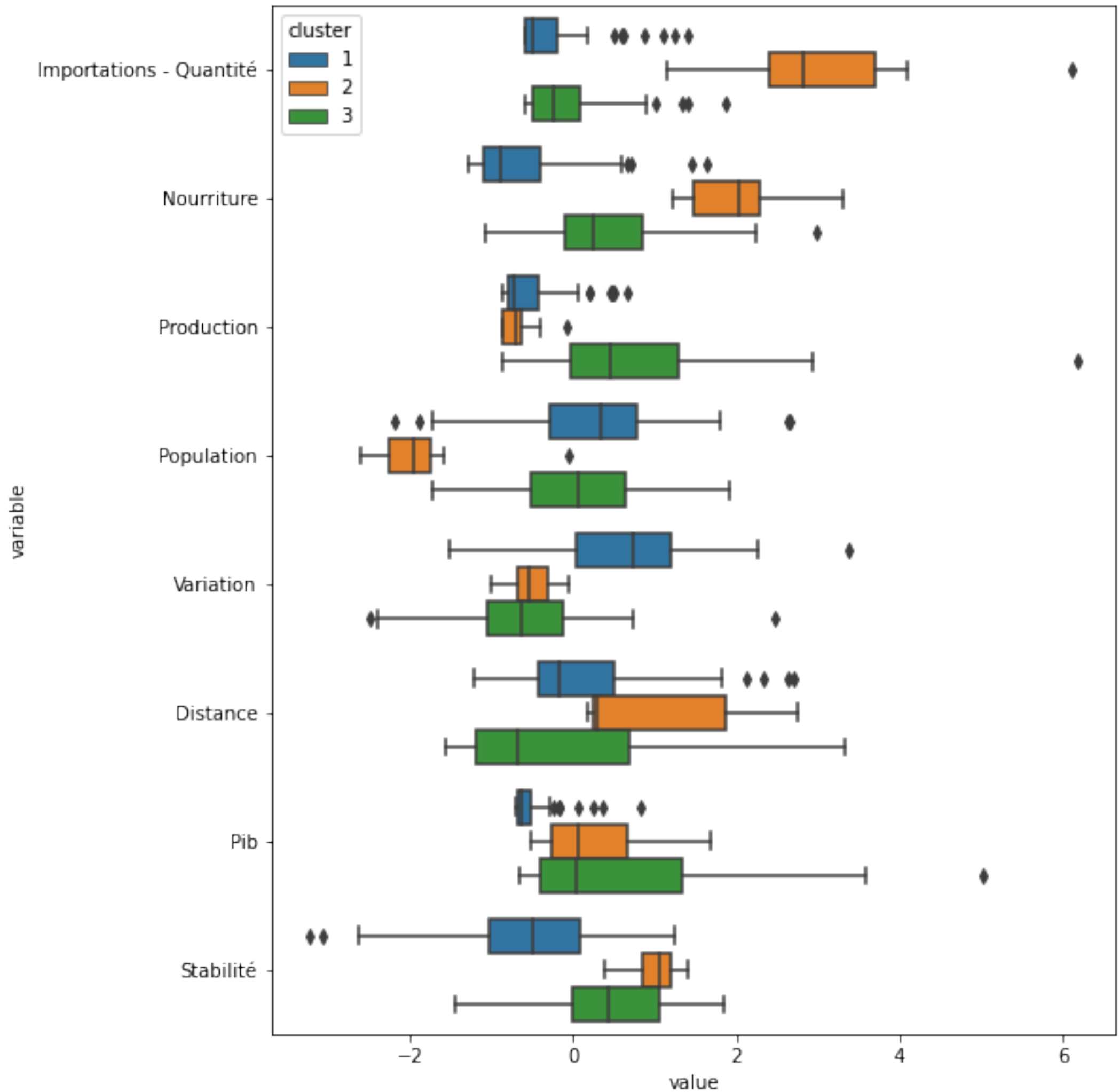
Clustering

K-means



Caractérisation des clusters

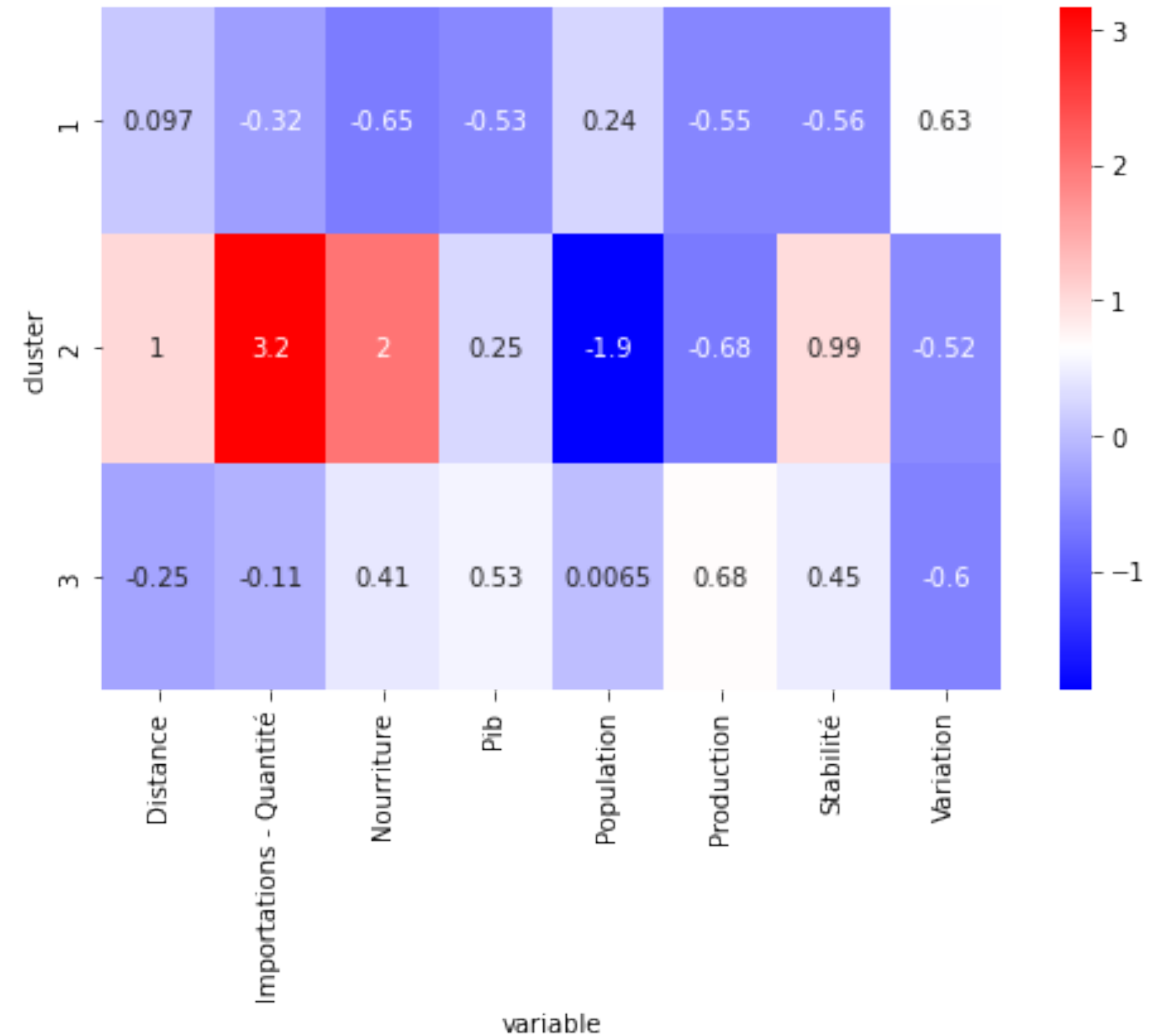
CAH



Caractérisation des clusters

CAH

- Cluster n°1 :
Peu de nourriture par habitant
PIB, stabilité faible
Forte variation de population
- Cluster n°2 :
Distance, stabilité élevé
Importation très élevé
Faible population
- Cluster n°3 :
Pays proche
PIB, production plutôt élevé



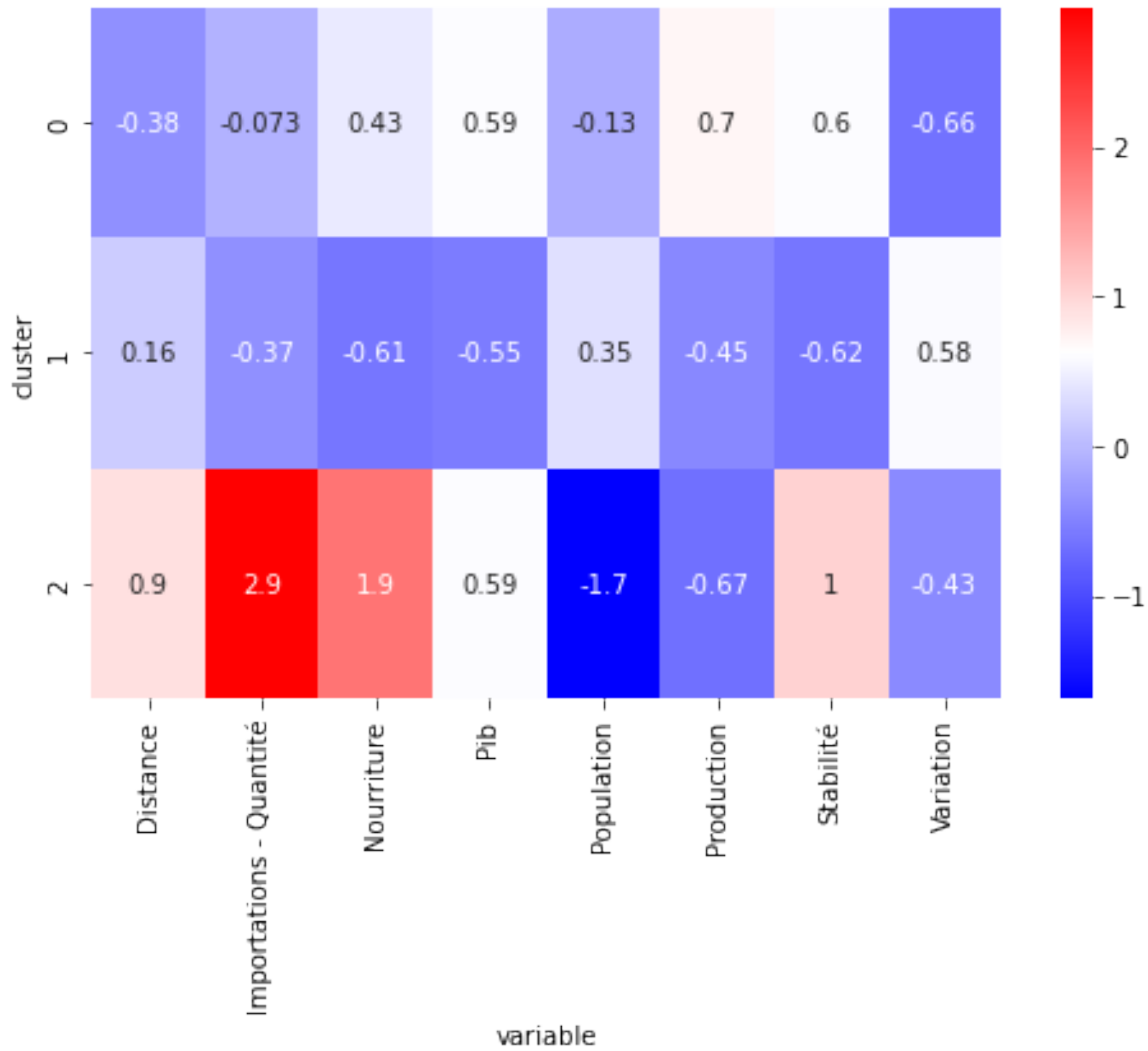
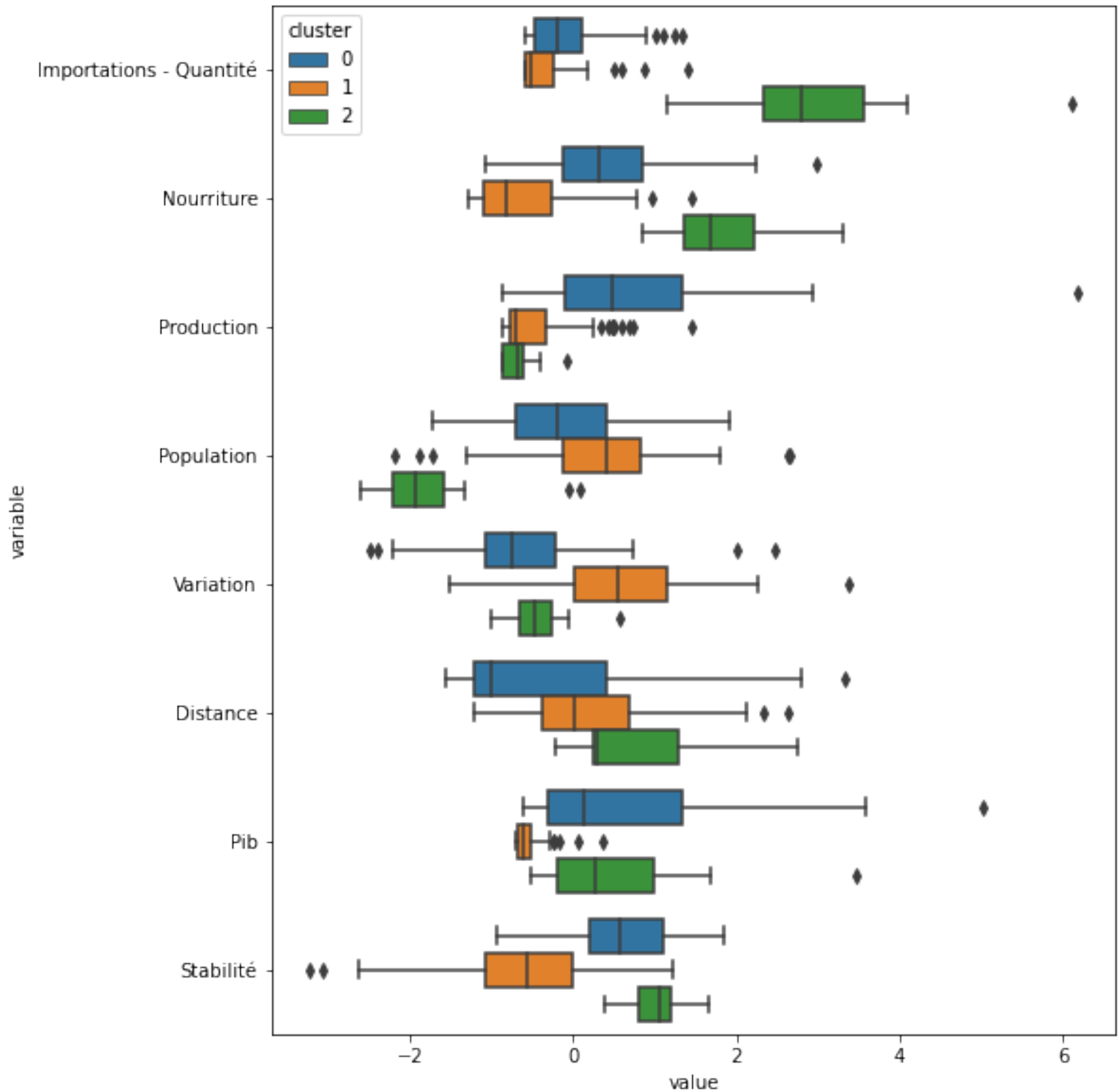
Caractérisation des clusters

CAH

	Importations - Quantité	Nourriture	Production	Population	Variation	Distance	Pib	Stabilité	cluster	Zone
6	3.367474	2.026835	-0.862621	-2.289803	-0.173881	1.283075	0.071774	0.910216	2	Antigua-et-Barbuda
13	2.801931	1.365201	-0.077974	-1.574601	-0.263469	0.294744	0.952113	1.179446	2	Bahamas
31	6.105954	2.097428	-0.698630	-0.051931	-0.481040	0.918370	1.694449	0.999959	2	Chine - RAS de Hong-Kong
44	2.430360	1.369682	-0.862621	-2.439011	-0.996197	0.196426	-0.347686	1.415022	2	Dominique
56	2.816339	1.566156	-0.412346	-2.212402	-0.563777	0.272095	-0.199206	1.201882	2	Grenade
108	1.162948	1.224985	-0.682488	-1.739790	-0.048776	2.743951	0.989686	0.382974	2	Nouvelle-Calédonie
120	2.341429	1.691594	-0.681804	-1.741744	-0.657185	2.478852	0.369358	0.382974	2	Polynésie française
133	3.556279	2.357864	-0.862621	-2.602539	-0.373911	0.178924	0.267190	0.798037	2	Saint-Kitts-et-Nevis
134	3.830781	3.316998	-0.862621	-2.217297	-0.892222	0.239150	-0.351512	1.067267	2	Saint-Vincent-et-les Grenadines
135	2.391844	2.207692	-0.586729	-1.959709	-0.691796	0.239150	-0.153520	1.179446	2	Sainte-Lucie
136	4.104465	2.919744	-0.862621	-1.920215	-0.543215	2.573310	-0.506387	1.392586	2	Samoa

Caractérisation des clusters

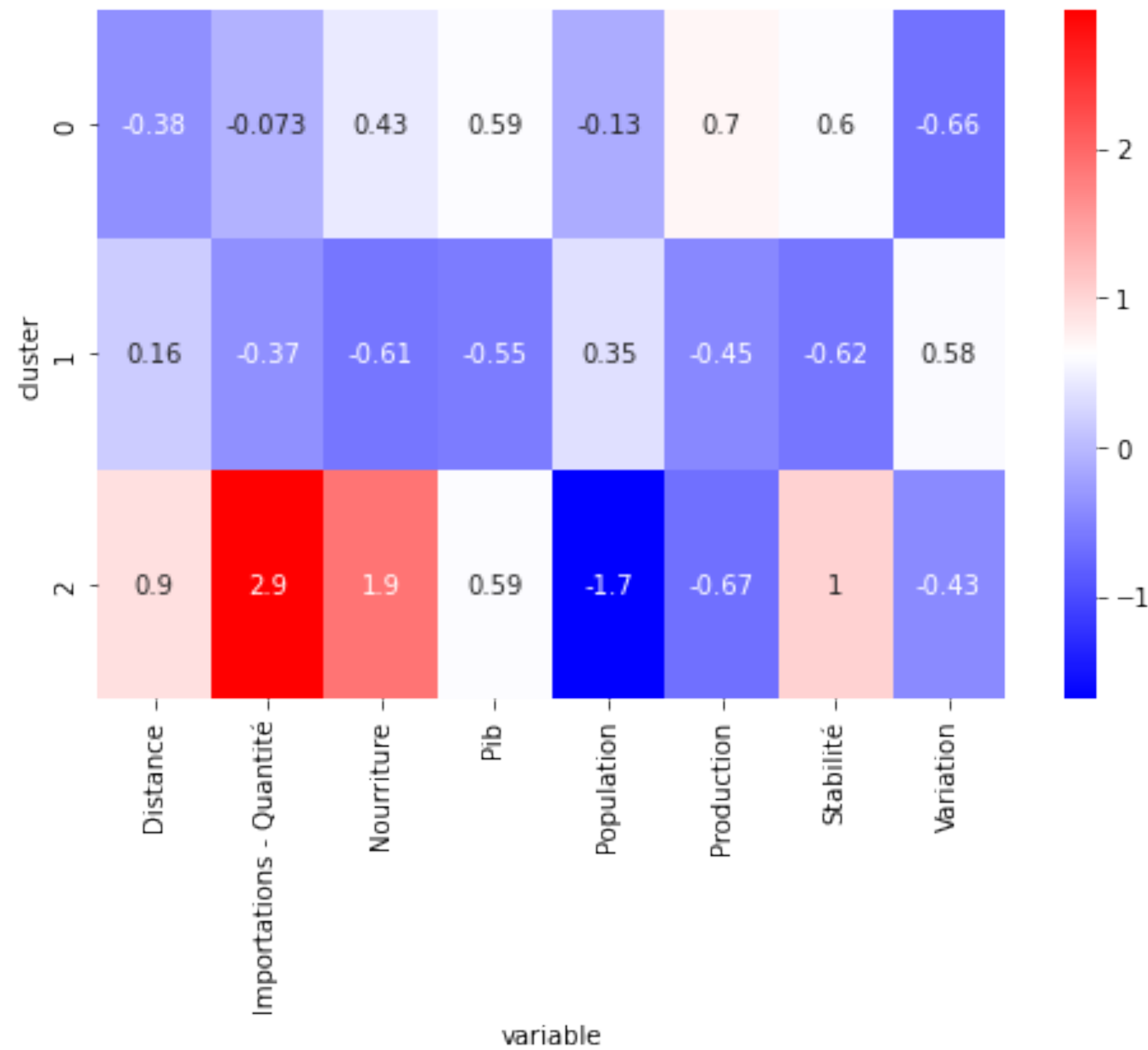
K-means



Caractérisation des clusters

K-means

Les résultats sont similaire aux résultats de l'algorithme CAH bien que les valeurs sont légèrement moins extrême. Il y à donc quelques différences entre les clusters.



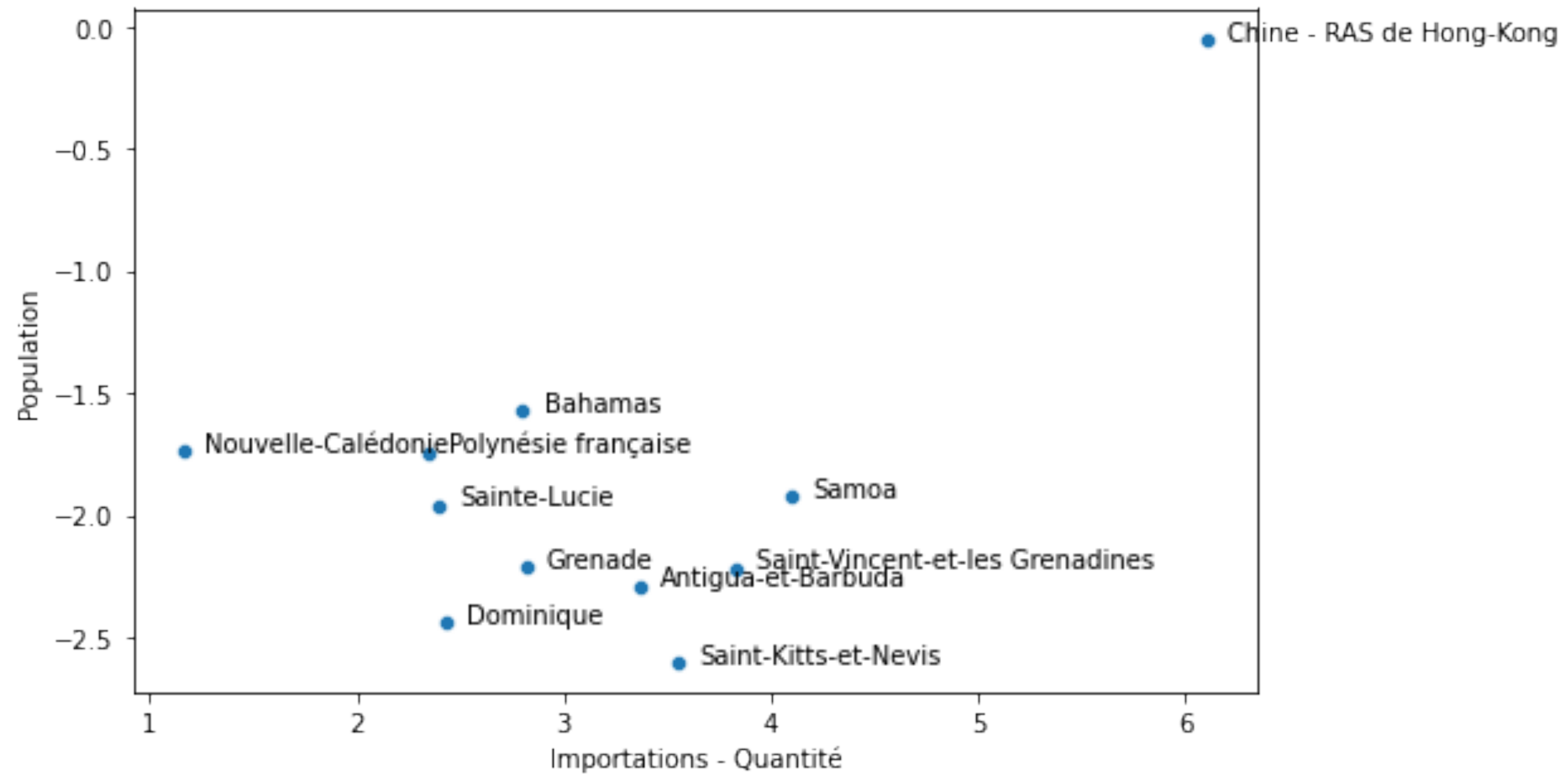
Caractérisation des clusters

CAH

	Importations - Quantité	Nourriture	Production	Population	Variation	Distance	Pib	Stabilité	cluster	Zone
6	3.367474	2.026835	-0.862621	-2.289803	-0.173881	1.283075	0.071774	0.910216	2	Antigua-et-Barbuda
13	2.801931	1.365201	-0.077974	-1.574601	-0.263469	0.294744	0.952113	1.179446	2	Bahamas
31	6.105954	2.097428	-0.698630	-0.051931	-0.481040	0.918370	1.694449	0.999959	2	Chine - RAS de Hong-Kong
32	1.403955	0.848929	-0.622057	-1.322294	0.572579	0.554953	3.473442	1.673034	2	Chine - RAS de Macao
44	2.430360	1.369682	-0.862621	-2.439011	-0.996197	0.196426	-0.347686	1.415022	2	Dominique
56	2.816339	1.566156	-0.412346	-2.212402	-0.563777	0.272095	-0.199206	1.201882	2	Grenade
108	1.162948	1.224985	-0.682488	-1.739790	-0.048776	2.743951	0.989686	0.382974	2	Nouvelle-Calédonie
120	2.341429	1.691594	-0.681804	-1.741744	-0.657185	2.478852	0.369358	0.382974	2	Polynésie française
133	3.556279	2.357864	-0.862621	-2.602539	-0.373911	0.178924	0.267190	0.798037	2	Saint-Kitts-et-Nevis
134	3.830781	3.316998	-0.862621	-2.217297	-0.892222	0.239150	-0.351512	1.067267	2	Saint-Vincent-et-les Grenadines
135	2.391844	2.207692	-0.586729	-1.959709	-0.691796	0.239150	-0.153520	1.179446	2	Sainte-Lucie
136	4.104465	2.919744	-0.862621	-1.920215	-0.543215	2.573310	-0.506387	1.392586	2	Samoa
167	1.872971	1.460817	-0.610034	0.082814	-0.464564	-0.209202	1.380487	0.764383	2	Émirats arabes unis

Résultats

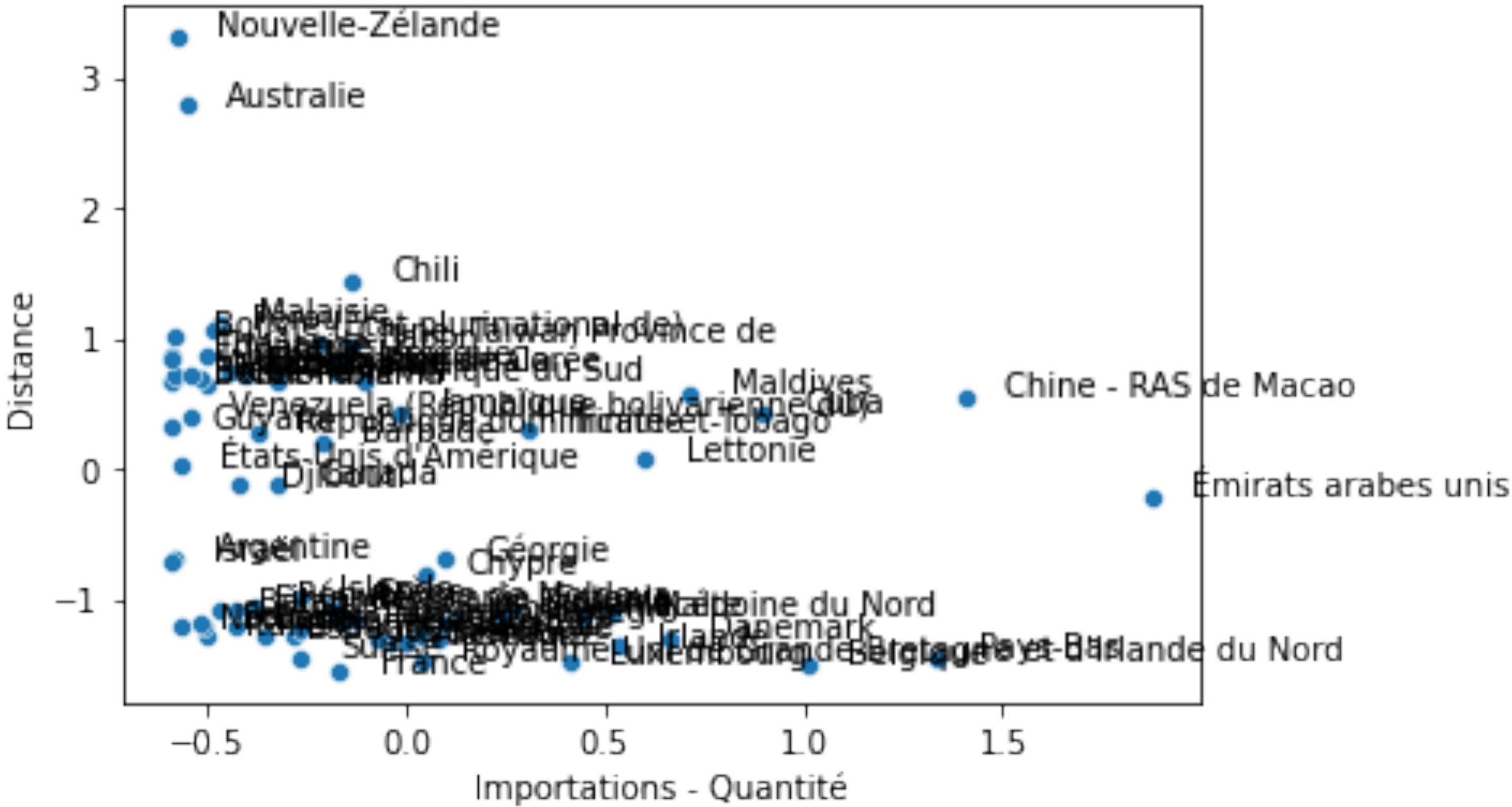
CAH



Résultats

CAH

- Cluster des pays proches
- Pays-Bas, Belgique intéressant dans les pays proches
- Émirats arabes unis semblent être intéressant malgré la distance

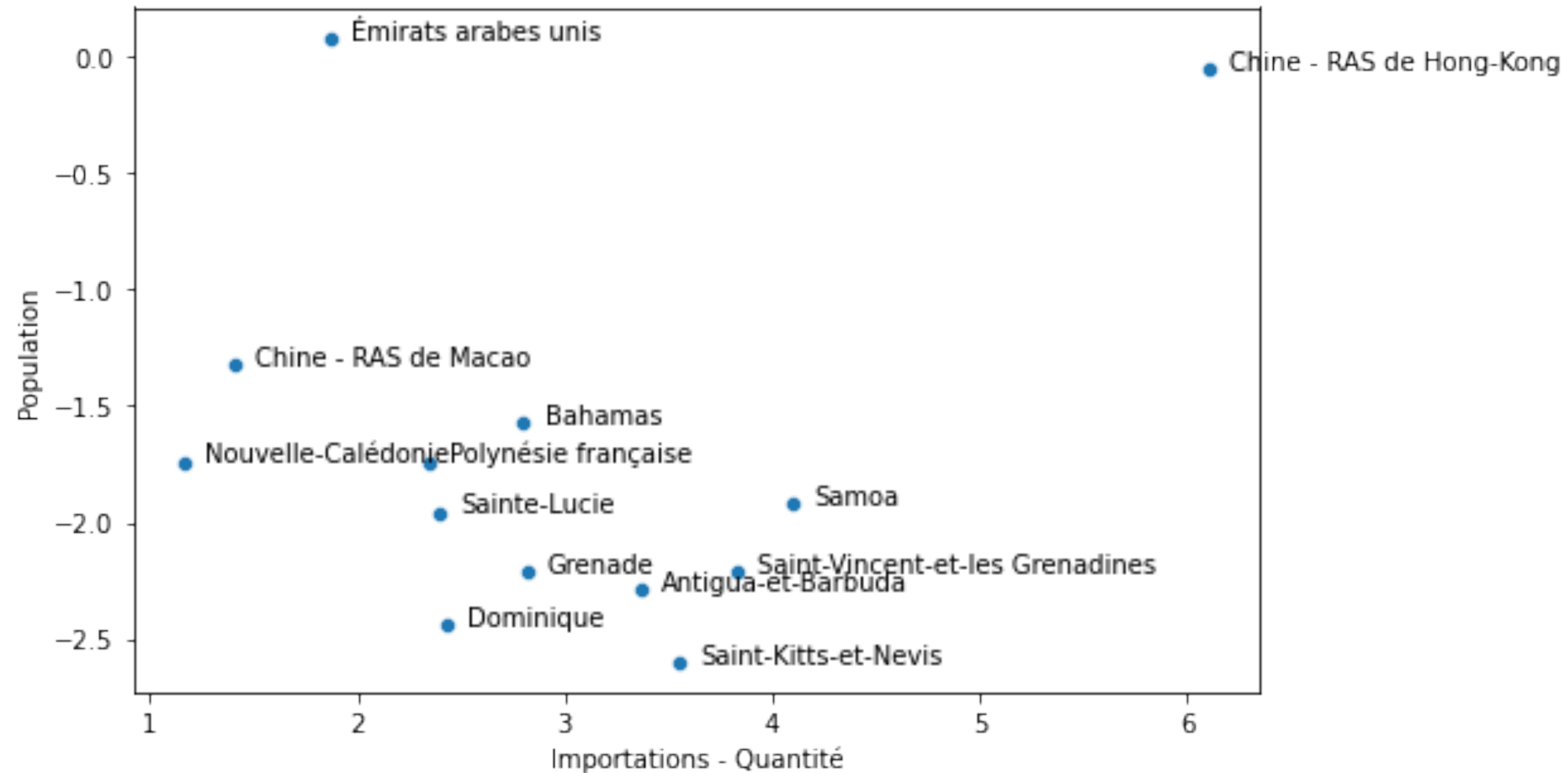


	Importations - Quantité	Nourriture	Production	Population	Variation	Distance	Pib	Stabilité	cluster	Zone
167	1.872971	1.460817	-0.610034	0.082814	-0.464564	-0.209202	1.380487	0.764383	3	Émirats arabes unis
32	1.403955	0.848929	-0.622057	-1.322294	0.572579	0.554953	3.473442	1.673034	3	Chine - RAS de Macao
117	1.338012	0.003159	2.363692	0.384346	-0.877657	-1.445130	1.812807	1.100920	3	Pays-Bas
16	1.007887	-0.483853	1.161480	0.178454	-0.593235	-1.487855	1.555652	0.551243	3	Belgique
40	0.895573	0.217426	-0.734941	0.174805	-0.989976	0.426521	-0.282699	0.798037	3	Cuba



Résultats

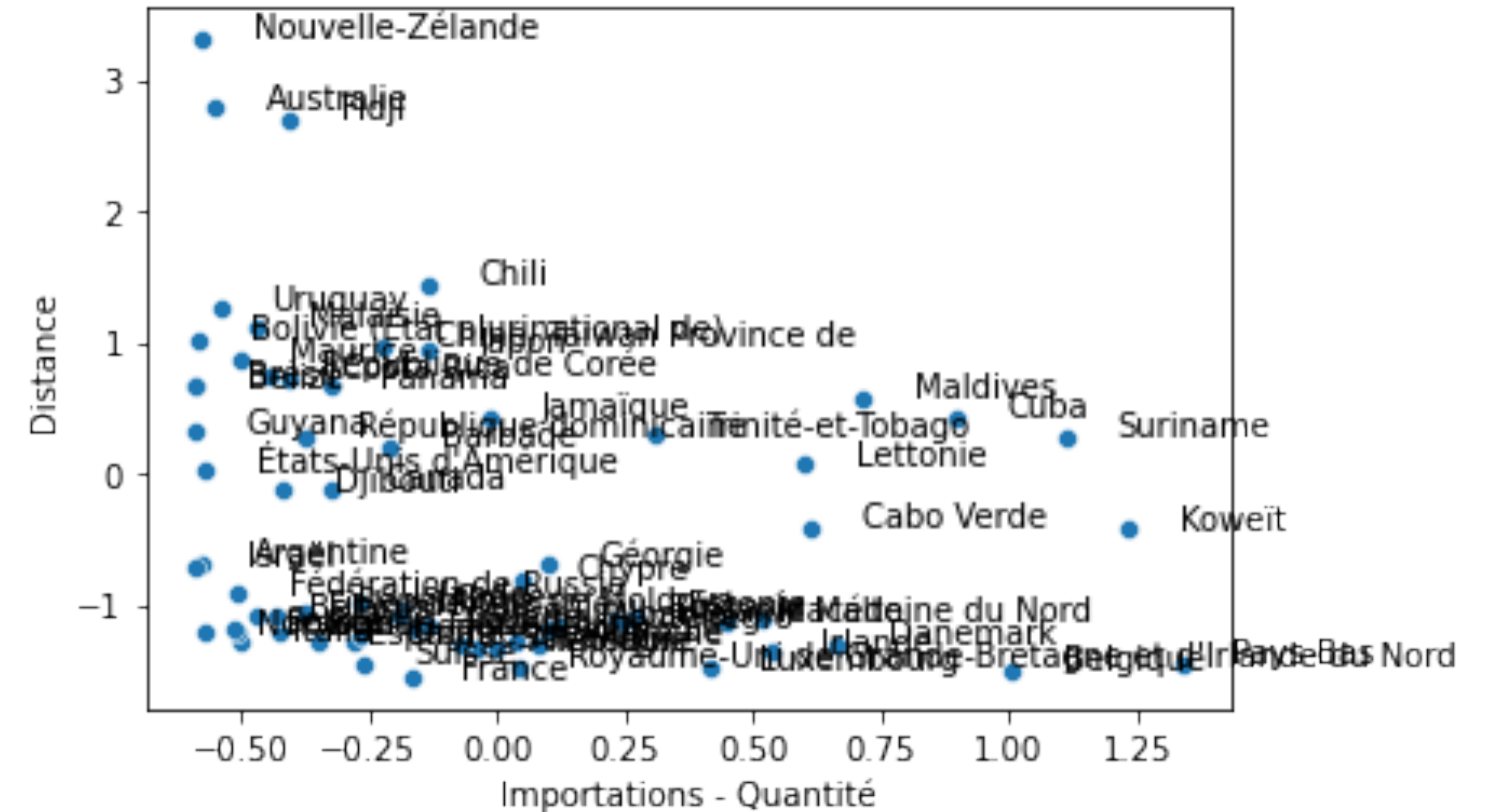
CAH



Résultats

K-means

- Cluster des pays proches
- Pays-Bas, Belgique intéressant dans les pays proches



	Importations - Quantité	Nourriture	Production	Population	Variation	Distance	Pib	Stabilité	cluster	Zone
117	1.338012	0.003159	2.363692	0.384346	-0.877657	-1.445130	1.812807	1.100920	0	Pays-Bas
81	1.233207	1.645270	-0.173354	-0.355523	2.018776	-0.422511	0.816302	0.012783	0	Koweït
145	1.113236	0.711436	0.012474	-1.367367	-0.197319	0.290883	-0.417078	0.214706	0	Suriname
16	1.007887	-0.483853	1.161480	0.178454	-0.593235	-1.487855	1.555652	0.551243	0	Belgique
40	0.895573	0.217426	-0.734941	0.174805	-0.989976	0.426521	-0.282699	0.798037	0	Cuba



Conclusion

Produisez une étude de marché avec R ou Python

- L'entreprise devrait exporter ses poulets vers Hong-Kong ou les Émirats Arabes Unis
- Les îles comme les Bahamas ou les Samoa pourraient être également intéressante
- En prenant en compte l'impact écologique dû à la distance les Pays-Bas ou la Belgique sont les plus intéressantes
- Finalement je conseillerais plutôt d'exporter vers les Émirats Arabes Unis car ils sont moins loin que Hong-Kong et plus exportateur que les Pays-Bas ou la Belgique