

# La poule qui chante

Etude de marché



# CONTEXTE

**La poule qui chante, une entreprise française d'agroalimentaire exerçant dans la vente de viande de poulet, souhaite se développer à l'international.**

**Actuellement, aucun pays ni continent n'est choisi (tout est envisageable).**

# MISSION

**L'objectif est de proposer une première analyse des groupements de pays que l'on peut cibler pour l'exportation de poulet, à partir des données fournies, mais également en utilisant les critères de l'analyse PESTEL, en utilisant les données en open data sur le site de la F.A.O.**

# METHODOLOGIE

J'ai choisi d'utiliser des données légèrement différentes de celles fournies au départ (données alimentaires 2017). Celles-ci concernant la viande de volaille au sens large, il me paraissait utile de se focaliser sur la viande de poulet, à proprement dit, à moins de se retrouver avec des produits n'ayant pas forcément les mêmes caractéristiques (Dinde, canard ,oie, pintade, caille, etc.).

Toutes les données que j'ai pu collecter ne concerne uniquement que de la viande de poulet, sauf pour ce qui est du cout de la viande par personne et par an, qui traite de la viande en général, ainsi que le taux d'inflation entre 2015 et 2017 qui, lui, traite de la nourriture, dans sa globalité.

J'ai ensuite regroupé, dans un premier temps, les données en 3 catégories (population, produit poulet et pays), pour, au final, ne faire qu'un dataframe.

Un grand nombre d'indicateurs a été intégré au départ de l'analyse, dans un souci de "ratisser large", mais un écrémage s'opère dès la première visualisation.

Le choix des pays à prendre en compte se fait au fur et à mesure de "cycles" de sélection successifs, pour arriver à se focaliser sur ceux qui arrivent à satisfaire au mieux aux différents indicateurs (ayant tous plus ou moins d'importance dans la décision).

# DONNEES UTILISEES

- **Population des pays.**
- **Disponibilité alimentaire des pays (viande de poulet).**
- **Consommation de viande de poulet.**
- **Taux d'inflation alimentaire entre 2015 et 2017.**
- **PIB par habitant.**
- **Coordonnées géographique des pays (distance avec la France).**
- **Cout de production de la viande de poulet (Dollar US).**
- **Cout de passage en douane.**
- **Délai d'attente aux douanes.**
- **Abordabilité de la nourriture carnée de qualité pour la population.**
- **Echanges commerciaux de poulet en quantités et valeurs(global).**
- **Echanges commerciaux de poulet avec la France.**
- **Indices de stabilité politique.**
- **Disponibilité alimentaire par habitant**

# **LES 32 INDICATEURS DU DF FINAL**

- **pays**
- **population**
- **%evolution\_2007/2017**
- **cout\_viande\_pers/an(USD)**
- **PIB\_habitant(USD)**
- **dispo\_alimentaire(kg/pers/an)**
- **Tx\_inflation**
- **production**
- **cout\_production(USD/t)**
- **import\_tonnes**
- **import\_USD**
- **prix\_import\_USD/tonne**
- **importFrance\_tonnes**
- **importFrance\_USD**
- **prix\_importFrance\_USD/tonne**
- **export\_tonnes**
- **export\_USD**
- **prix\_export\_USD/tonne**
- **exportFrance\_tonnes**
- **exportFrance\_USD**
- **prix\_exportFrance\_USD/tonne**
- **excedent\_commercial\_fr\_tonnes**
- **excedent\_commercial\_fr\_USD**
- **variation\_stock**
- **nourriture**
- **disponibilite\_interieure**
- **emission\_co2(tonnes)**
- **tonne\_co2/tonne\_poulet**
- **cout\_douanes(USD)**
- **delai\_douanes(heures)**
- **stabilite**
- **Distance\_france(km)**

**La chine, comprenant le cumul de ses régions, a été supprimée. Seules ces dernières sont conservées ( continentale, Taiwan, Hong-Kong et Makao).La France est également supprimée.**

## **indicateurs les plus décisifs**

- Population (+).
- Evolution démographique sur 10 ans (+).
- Taux d'inflation sur la nourriture (-).
- PIB par habitant (++) .
- Echanges commerciaux - Import/export(++) .
- Consommation de viande de poulet - nourriture (+).
- Distance avec la France (++) .
- Stabilité politique (+).
- Cout viande pers/an (--) .

## Tri parmi les clusters :

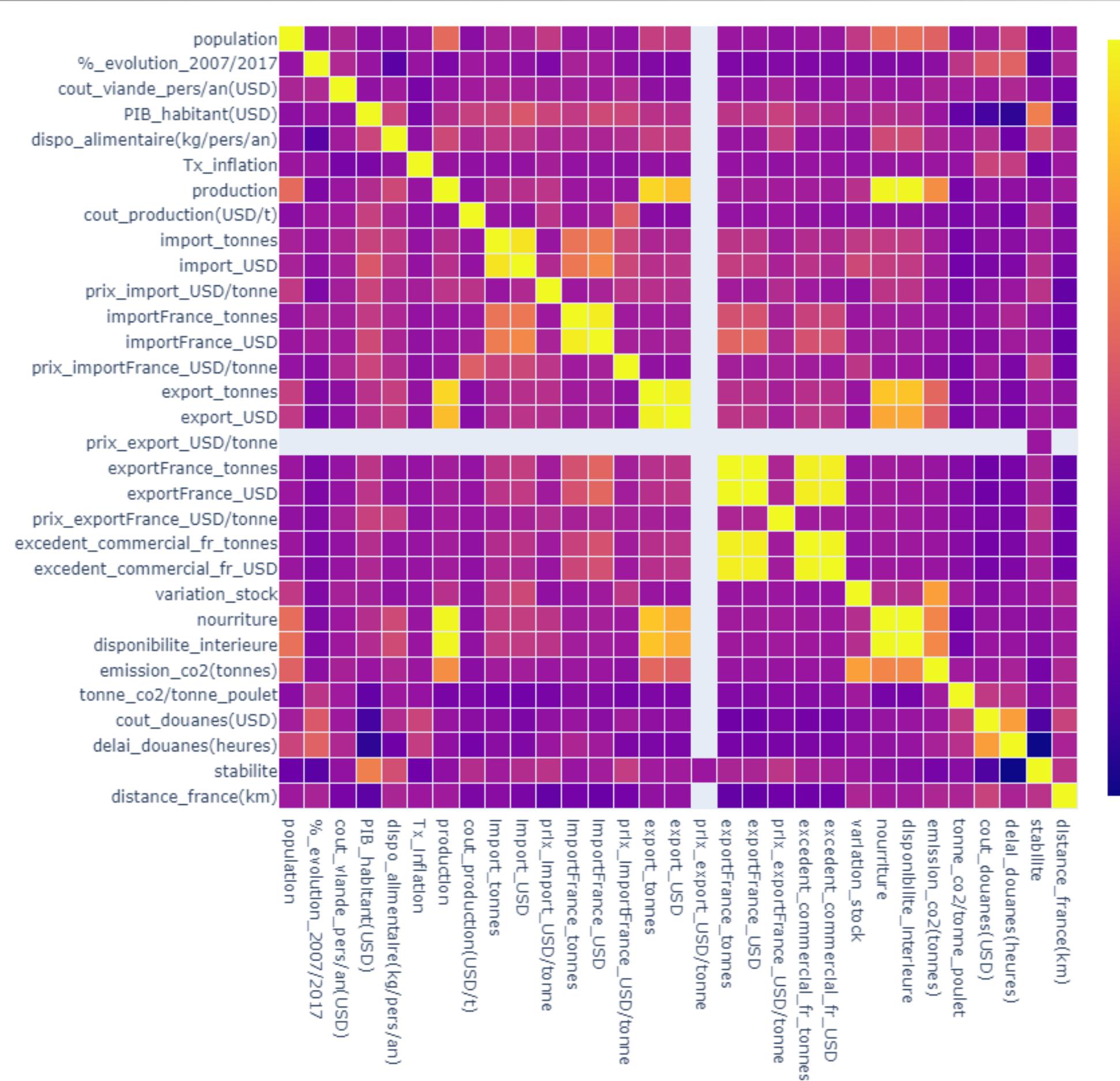
Clusters prometteurs à conserver :

- La taille de la population (population) peut avoir un impact significatif sur la demande de produits ou services, car une population plus importante signifie généralement une base plus large de consommateurs potentiels.
- Croissance de la population (%\_evolution\_2007/2017): Un cluster avec une population en augmentation peut indiquer une demande croissante pour les produits offerts.
- Un budget annuel pour une consommation de produits carnés de qualité par habitant (cout\_viande\_pers/an(USD)) plus bas peut encourager une plus grande consommation de produits carnés et stimuler la demande.
- Pouvoir d'achat élevé (PIB\_habitant(USD)): Un PIB par habitant élevé suggère une capacité à acheter des produits à valeur ajoutée, ce qui peut être bénéfique pour les exportations.
- Un taux d'inflation (Tx\_inflation) plus faible peut stimuler la demande en rendant les produits alimentaires plus abordables.
- Faible production locale (production): Si un cluster a une capacité de production limitée pour les produits que l'on souhaite importer, il peut être avantageux de le conserver afin de répondre à la demande du marché.
- Forts coûts de production (cout\_production(USD/t)) : Si un cluster présente des coûts de production relativement élevés, il peut offrir un avantage concurrentiel en termes de prix et il peut être plus rentable de les importer plutôt que de les produire localement.
- Demande élevée (import\_tonnes): Si un cluster présente une demande significative pour les produits que vous souhaitez importer, il peut être judicieux de le conserver pour saisir les opportunités commerciales.

Clusters moins utiles à écarter :

- Faible croissance de la population : Un cluster avec une population stagnante ou en déclin peut indiquer une demande limitée à long terme.
- Pouvoir d'achat faible : Si le PIB par habitant est faible, cela peut limiter la capacité des consommateurs à acheter des produits à valeur ajoutée.

# 1ERE MATRICE DES CORRELATIONS



## **Nombre d'indicateurs trop élevés**

**On se rend compte qu'un grand nombre d'indicateurs est utilisé, mais qu'une grande partie d'entre eux semble être inutilement présente.**

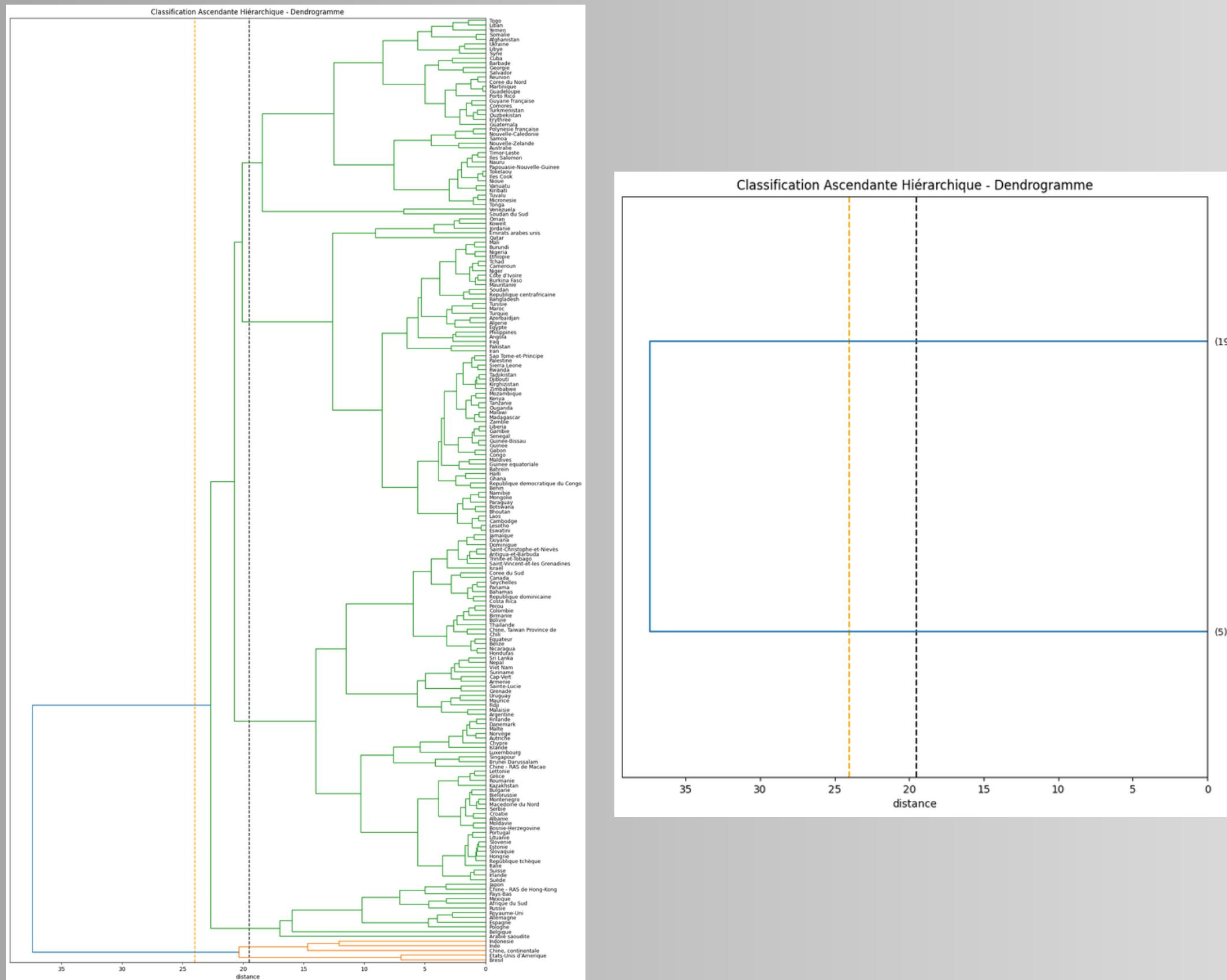
**Je choisi donc de n'en conserver que 17 sur les 32 initiaux :**

- **population**
  - **%\_evolution\_2007\_2017**
  - **cout\_vianne\_pers/an(USD)**
  - **PIB\_habitant(USD)**
  - **Tx\_inflation**
  - **production**
  - **cout\_production(USD/t)**
  - **import\_tonnes**
  - **export\_tonnes**
  - **importFrance\_tonnes**
  - **exportFrance\_tonnes**
  - **emission\_co2(tonnes)**
  - **stabilite**
  - **distance\_france(km)**
  - **nourriture**
  - **disponibilite\_interieure**
  - **dispo\_alimentaire(kg/pers/an)**

# DENDROGRAMME

Après quelques essais de dendrogrammes, sur différentes quantités de lignes, j'ai pu m'apercevoir qu'un groupe de 5 feuilles se détachait des autres. Je vais donc approfondir les analyses concernant ces 5 feuilles.

Suite à la séparation des données en 2 clusters, on peut voir que les 5 feuilles se détachant des autres sont, à première vue, des pays à forte population.



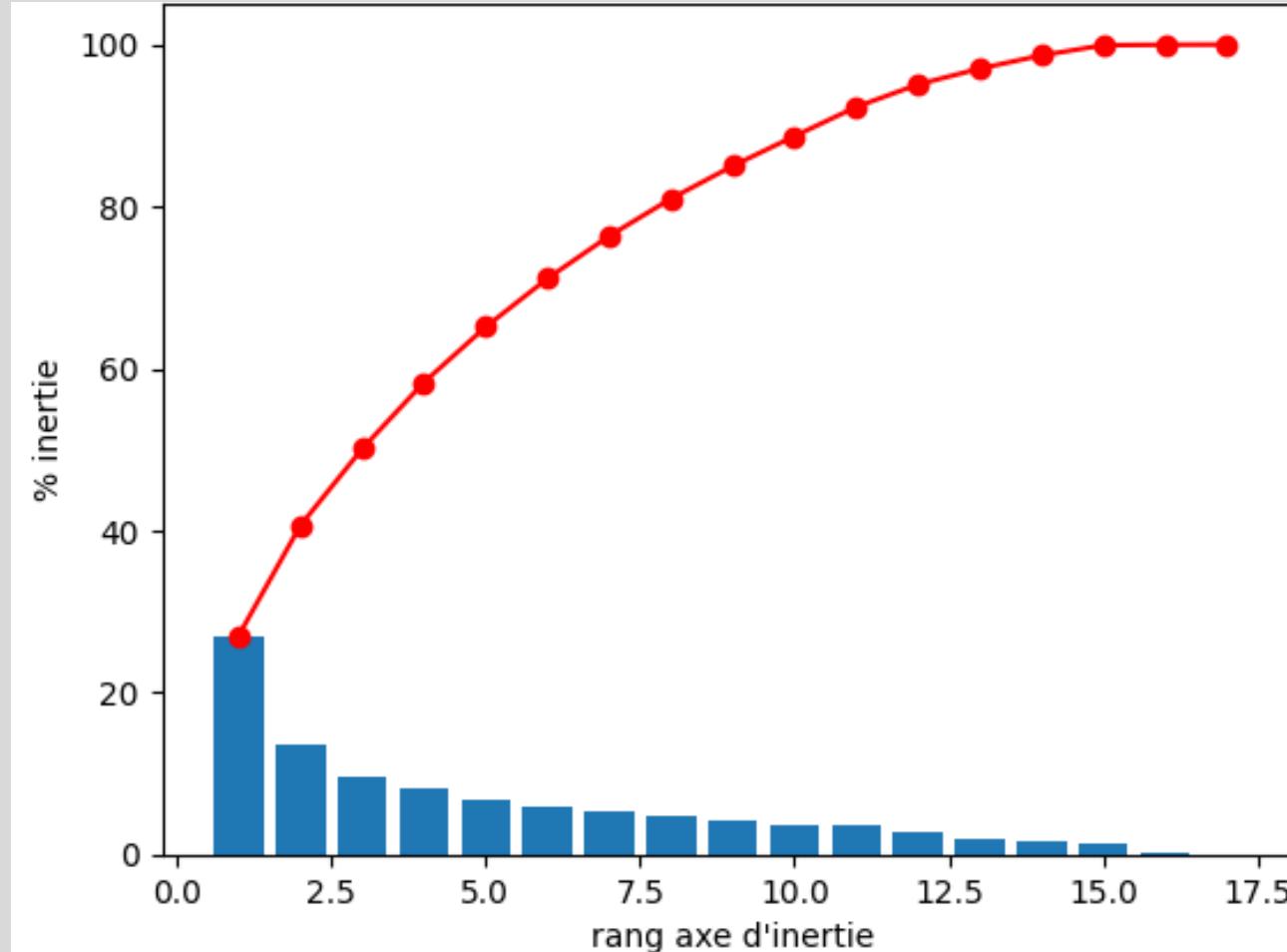
cluster	pays
25	Bresil
37	Chine, continentale
58	Etats-Unis d'Amérique
81	Inde
82	Indonesia

J'ai effectué plusieurs tests, en appelant les différents indicateurs, et il s'avère que ces 5 pays se retrouvent parmi les top de valeurs dans 6 d'entre eux, sur les 17 (soit plus de 35%).

- population
- production
- export
- émission c02
- nourriture
- dispo

Je décide donc de supprimer ces 5 pays des données, afin qu'ils ne faussent pas celles-ci.

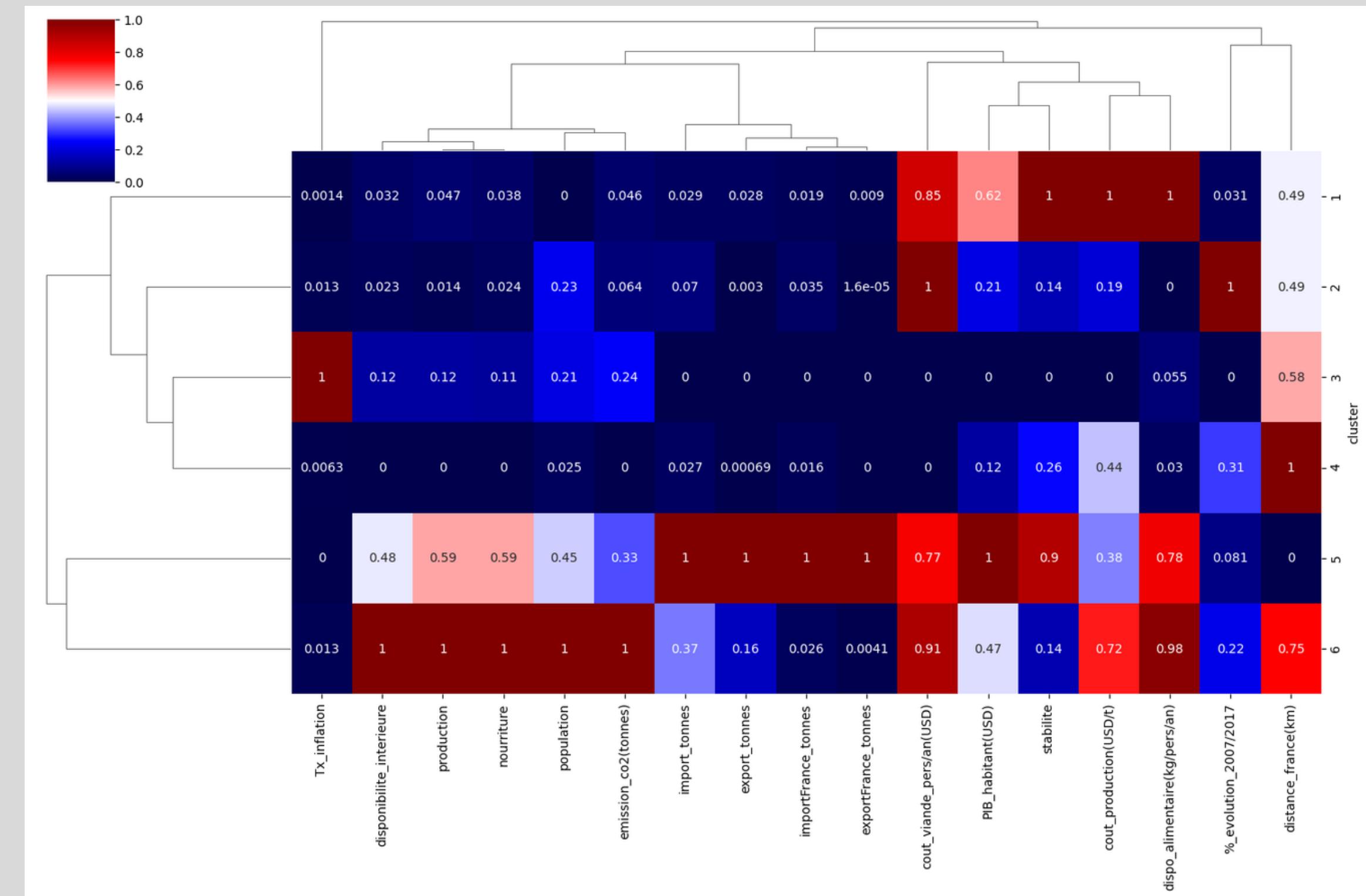
# EBOULIS DES VALEURS PROPRES



```
array([0.27010421, 0.40614182, 0.50075697, 0.5828792 , 0.6512982 ,
0.71107526, 0.76353112, 0.80964166, 0.85011108, 0.88686298,
0.92293052, 0.95074013, 0.97027144, 0.98662704, 0.99914788,
0.99978221, 1.        ])
```

au regard de l'ébouli et du cumul des variances expliquées, je choisis de retenir 6 composantes principales, celles-ci représentant un peu plus de 71% de l'inertie.

# CLUSTERMAP



On peut observer les clusters et les variables y étant représentées. Les clusters 1, 5 et 6 semblent être les plus "avantageux" à la lecture de ce clustermap.

# AUTOMATISATION DE LA LECTURE DES RESULTATS DE CLUSTERING

```
cluster 1:  
valeurs Forte : cout_viande_pers/an(USD), stabilité, cout_production(USD/t), dispo_alimentaire(kg/pers/an)  
valeurs Moyenne : PIB_habitant(USD), distance_france(km)  
valeurs Faible : Tx_inflation, disponibilité_interieure, production, nourriture, population, emission_co2(tonnes), import_tonnes, export_tonnes, importFrance_tonnes, exportFrance_tonnes, %_evolution_2007/2017  
  
cluster 2:  
valeurs Forte : cout_viande_pers/an(USD), %_evolution_2007/2017  
valeurs Moyenne : distance_france(km)  
valeurs Faible : Tx_inflation, disponibilité_interieure, production, nourriture, population, emission_co2(tonnes), import_tonnes, export_tonnes, importFrance_tonnes, exportFrance_tonnes, PIB_habitant(USD), stabilité, cout_production(USD/t), dispo_alimentaire(kg/pers/an)  
  
cluster 3:  
valeurs Forte : Tx_inflation  
valeurs Moyenne : distance_france(km)  
valeurs Faible : disponibilité_interieure, production, nourriture, population, emission_co2(tonnes), import_tonnes, export_tonnes, importFrance_tonnes, exportFrance_tonnes, cout_viande_pers/an(USD), PIB_habitant(USD), stabilité, cout_production(USD/t), dispo_alimentaire(kg/pers/an), %_evolution_2007/2017  
  
cluster 4:  
valeurs Forte : distance_france(km)  
valeurs Moyenne : cout_production(USD/t), %_evolution_2007/2017  
valeurs Faible : Tx_inflation, disponibilité_interieure, production, nourriture, population, emission_co2(tonnes), import_tonnes, export_tonnes, importFrance_tonnes, exportFrance_tonnes, cout_viande_pers/an(USD), PIB_habitant(USD), stabilité, dispo_alimentaire(kg/pers/an)  
  
cluster 5:  
valeurs Forte : import_tonnes, export_tonnes, importFrance_tonnes, exportFrance_tonnes, cout_viande_pers/an(USD), PIB_habitant(USD), stabilité, dispo_alimentaire(kg/pers/an)  
valeurs Moyenne : disponibilité_interieure, production, nourriture, population, emission_co2(tonnes), cout_production(USD/t)  
valeurs Faible : Tx_inflation, %_evolution_2007/2017, distance_france(km)  
  
cluster 6:  
valeurs Forte : disponibilité_interieure, production, nourriture, population, emission_co2(tonnes), cout_viande_pers/an(USD), cout_production(USD/t), dispo_alimentaire(kg/pers/an), distance_france(km)  
valeurs Moyenne : import_tonnes, PIB_habitant(USD)  
valeurs Faible : Tx_inflation, export_tonnes, importFrance_tonnes, exportFrance_tonnes, stabilité, %_evolution_2007/2017
```

Colonnes	Valeurs Fortes (Clusters)	Valeurs Moyennes (Clusters)	Valeurs Faibles (Clusters)
population	6	5	1, 2, 3, 4
%_evolution_2007/2017	2	4	1, 3, 5, 6
cout_viande_pers/an(USD)	1, 2, 5, 6		3, 4
PIB_habitant(USD)	5	1, 6	2, 3, 4
Tx_inflation	3		1, 2, 4, 5, 6
production	6	5	1, 2, 3, 4
cout_production(USD/t)	1, 6	4, 5	2, 3
import_tonnes	5	6	1, 2, 3, 4
export_tonnes	5		1, 2, 3, 4, 6
importFrance_tonnes	5		1, 2, 3, 4, 6
exportFrance_tonnes	5		1, 2, 3, 4, 6
emission_co2(tonnes)	6	5	1, 2, 3, 4
stabilité	1, 5		2, 3, 4, 6
distance_france(km)	4, 6	1, 2, 3	5
nourriture	6	5	1, 2, 3, 4
disponibilité_interieure	6	5	1, 2, 3, 4
dispo_alimentaire(kg/pers/an)	1, 5, 6		2, 3, 4

Après le classement des indicateurs dans chaque cluster (fig gauche), ainsi que la représentativité des clusters dans chaque indicateur (fig droite), je choisis d'écarter les clusters 2, 3 et 4.

Ces clusters ont des résultats ne permettant pas de les conserver.

Par exemple, le cluster 2 a un cout élevé par habitant, pour ce qui est de la nourriture carnée, tout en ayant un faible PIB par habitant, un indice de stabilité politique bas et de faibles consommations de viande de poulet

Il reste 96 pays en course, sur les 195, au départ.

# DIAGRAMMES DE KIVIAT

La lecture des diagramme de Kiviat confirme, de manière parlante, les informations m'ayant amené à écarter les clusters 2, 3 et 4 .



# AUTOMATISATION DE CLASSEMENT DU TOP 3 DES CLUSTERS

## 5.2.3.4 - top 3 clusters pour l'ensemble des indicateurs les plus pertinents (besoins hauts et besoins bas)

```
meilleures_colonnes = ['population', '%_evolution_2007/2017', 'PIB_habitant(USD)', 'stabilite', 'cout_production(USD/t)',  
    'import_tonnes', 'nourriture']  
moins_bonnes_colonnes = ['cout_viande_pers/an(USD)', 'Tx_inflation', 'production', 'distance_france(km)',  
    'disponibilite_interieure', 'dispo_alimentaire(kg/pers/an)']  
  
df_annot['moyenne_meilleures'] = df_annot[meilleures_colonnes].mean(axis=1)  
  
df_annot['moyenne_moins_bonnes'] = df_annot[moins_bonnes_colonnes].mean(axis=1)  
  
indices_meilleures = df_annot['moyenne_meilleures'].nlargest(3).index.tolist()  
indices_moins_bonnes = df_annot['moyenne_meilleures'].nsmallest(3).index.tolist()  
  
indices_meilleures_moins_bonnes = df_annot['moyenne_moins_bonnes'].nlargest(3).index.tolist()  
indices_moins_bonnes_moins_bonnes = df_annot['moyenne_moins_bonnes'].nsmallest(3).index.tolist()  
  
meilleurs_clusters = [str(cluster) for cluster in indices_meilleures]  
moins_bons_clusters = [str(cluster) for cluster in indices_moins_bonnes]  
meilleurs_clusters_moins_bons = [str(cluster) for cluster in indices_meilleures_moins_bonnes]  
moins_bons_clusters_moins_bons = [str(cluster) for cluster in indices_moins_bonnes_moins_bonnes]  
  
phrase_finale = f"Les 3 clusters ayant les meilleures moyennes sur les colonnes \"{''.join(meilleures_colonnes)}\" sont les cl  
    f"Les 3 clusters ayant les meilleures moyennes sur les colonnes \"{''.join(moins_bonnes_colonnes)}\" sont les cl  
  
print(phrase_finale)
```

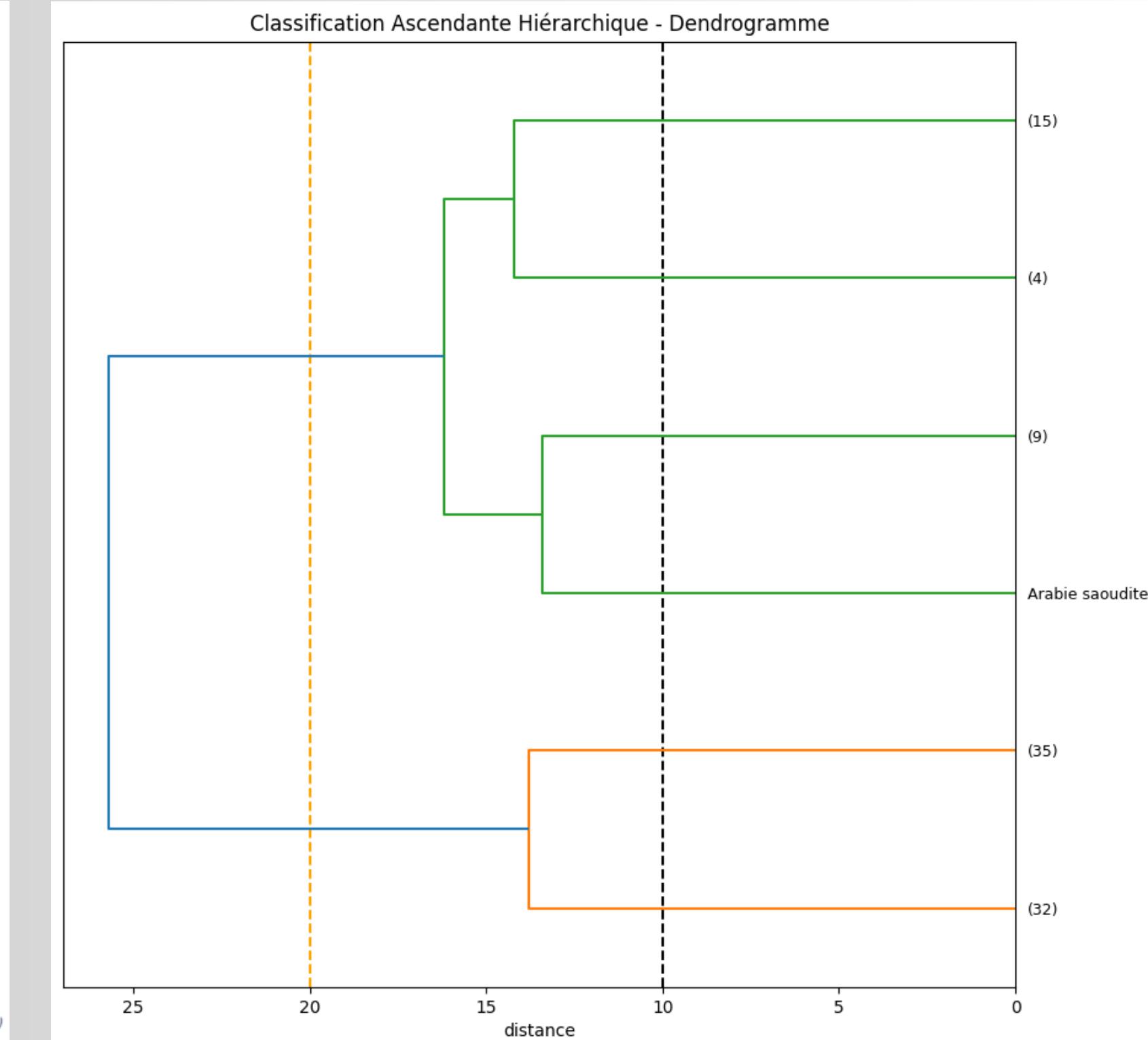
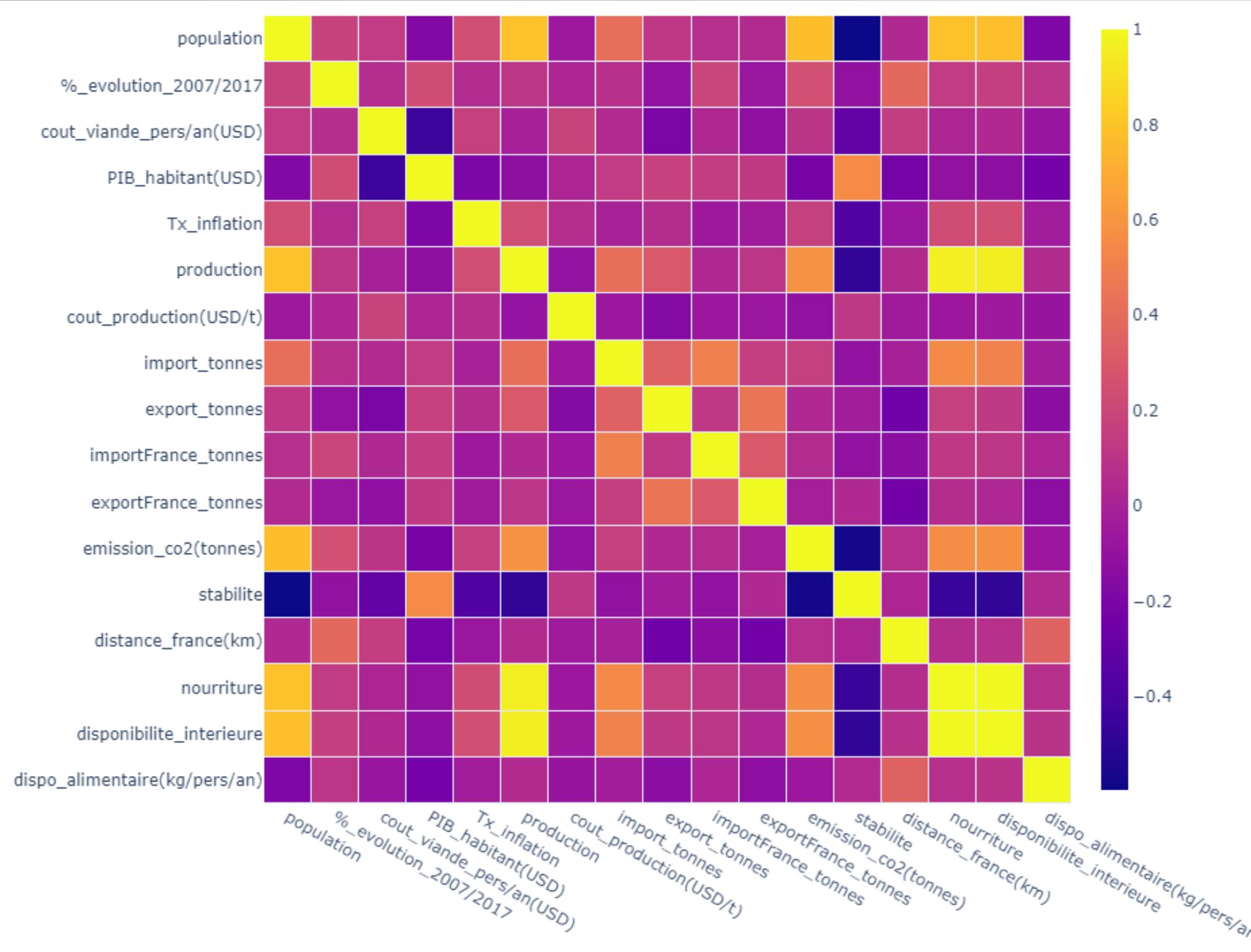
```
Les 3 clusters ayant les meilleures moyennes sur les colonnes "population, %_evolution_2007/2017, PIB_habitant(USD), stabilite,  
cout_production(USD/t), import_tonnes, nourriture" sont les clusters "5, 6, 1".  
Les 3 clusters ayant les meilleures moyennes sur les colonnes "cout_viande_pers/an(USD), Tx_inflation, production, distance_fra  
nce(km), disponibilite_interieure, dispo_alimentaire(kg/pers/an)" sont les clusters "6, 5, 1".
```

En faisant en sorte de sélectionner les meilleurs valeurs pour les indicateurs les plus pertinents.

Selon que celles-ci soient, au besoin, hautes comme pour la population, le PIB par habitant ou la stabilité politique ,ou bien basses comme pour le cout de la viande par habitant, le taux d'inflation, la distance avec la France,etc, les résultats ammènent toujours les 3 mêmes clusters, à savoir 1, 5 et 6.

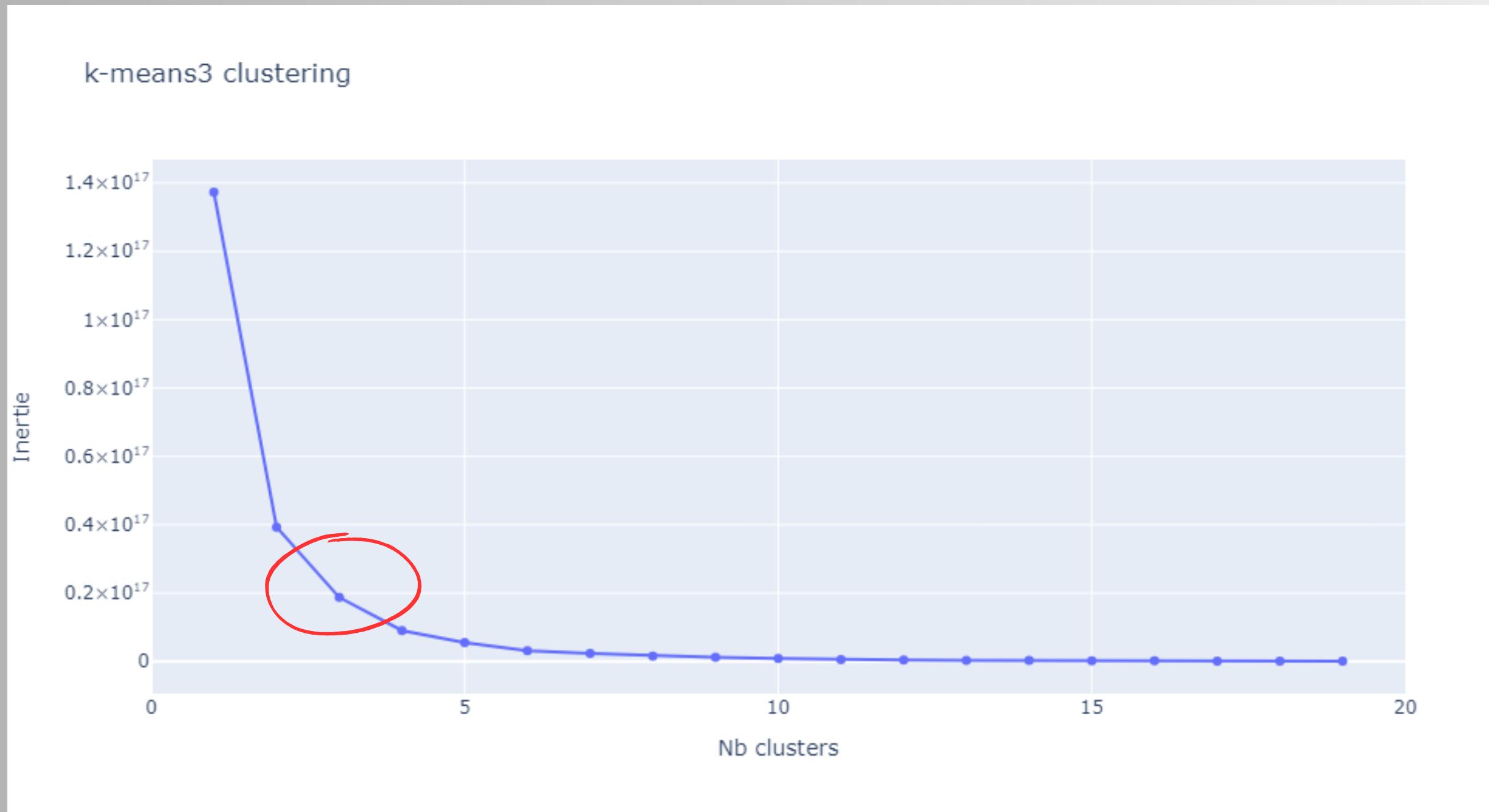
Ceci me conforte, une fois de plus, dans mon choix d'écartier les clusters 2, 3 et 4.

# MATRICE DES CORRELATIONS ET DENDROGRAMME POUR LES 96 PAYS RESTANTS



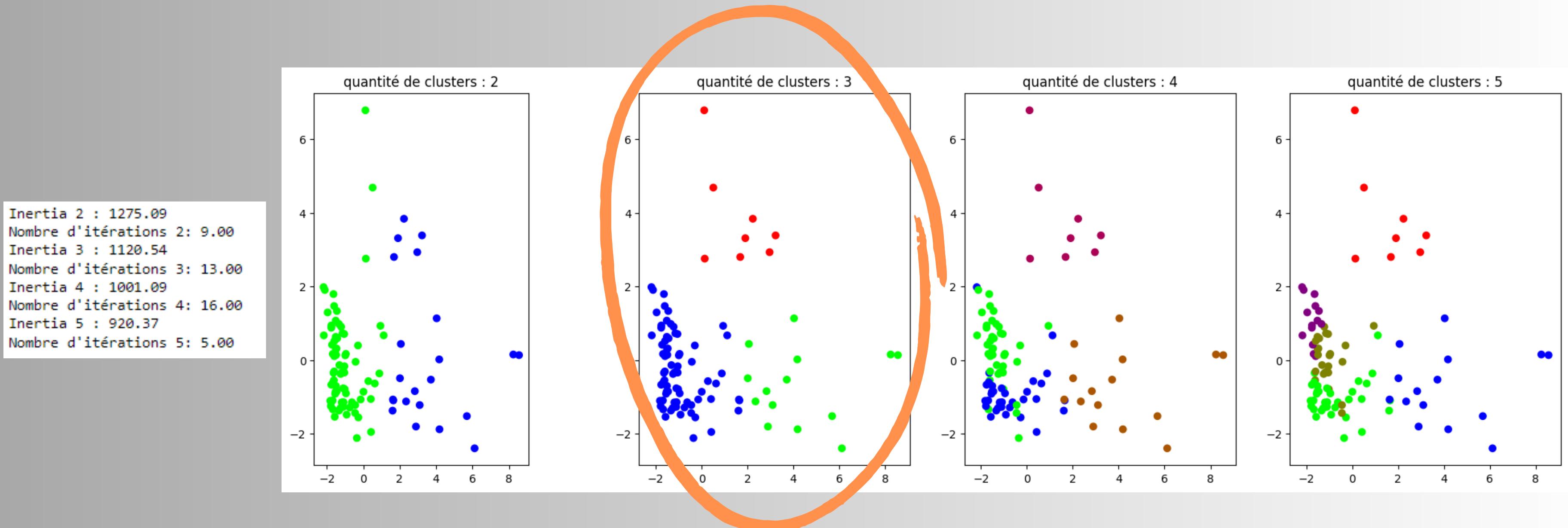
dendrogramme réparti sur 6 lignes

# **ALGORITHME DE KMEANS**



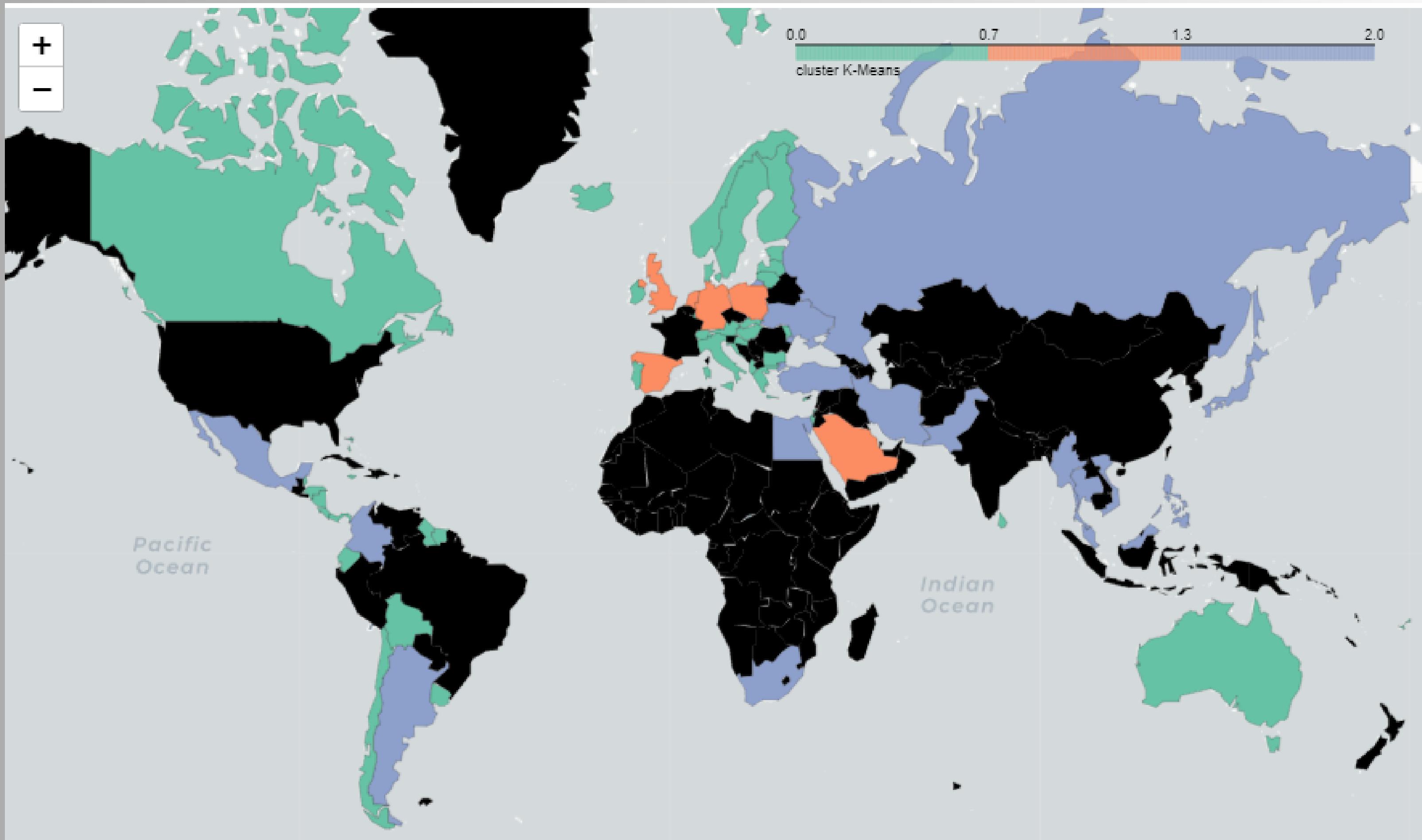
La méthode du coude (Elbow) me permet de visualiser le nombre optimal de clusters à utiliser.  
Ici, je choisis d'en sélectionner 3.

# **CLUSTERING K-MEANS - SUBPLOTS**

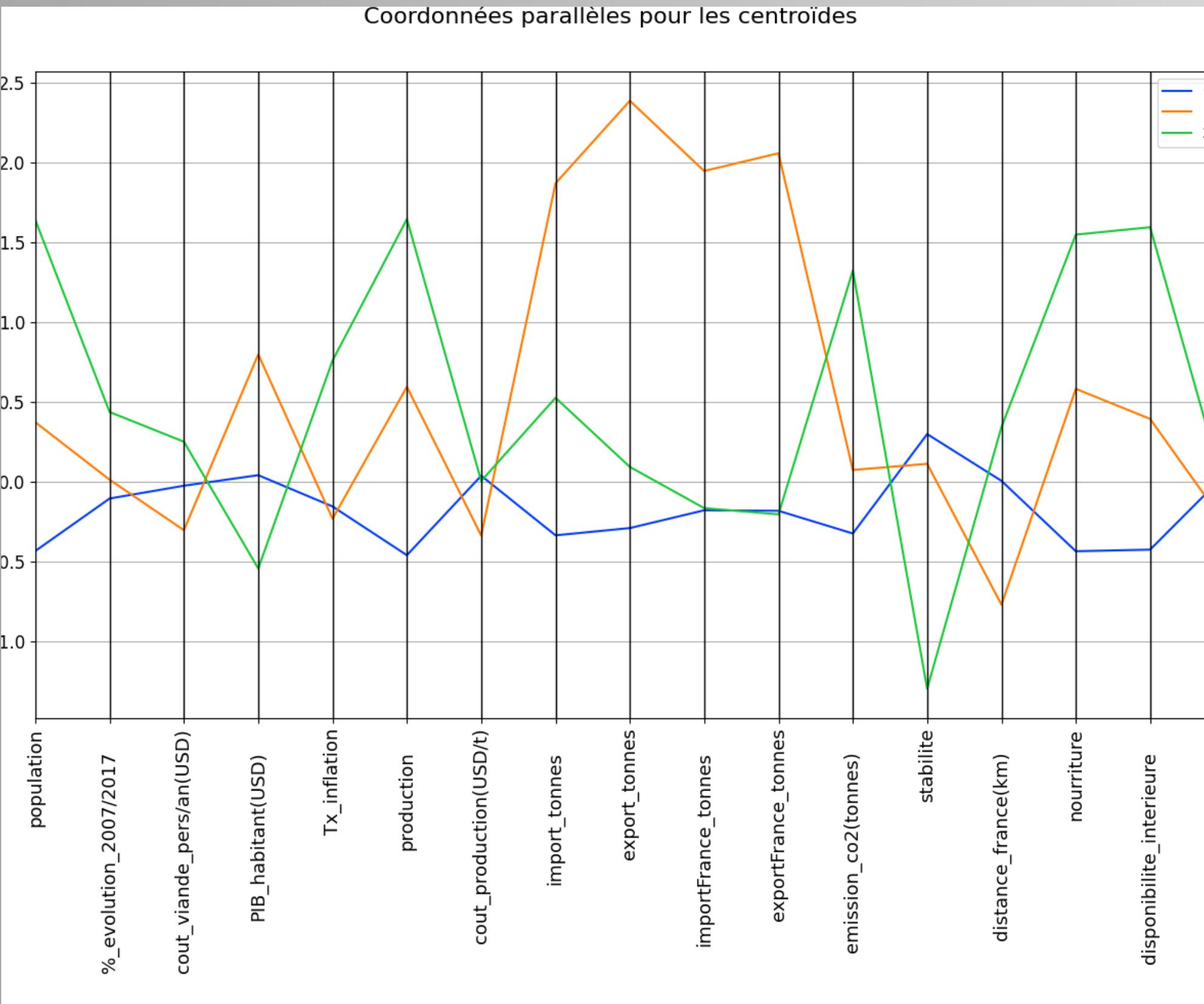


A l'aide de ces sous-graphiques représentant les possibilités diverses de segmentation des données en différents clusters, je remarque que le choix de 3 clusters sera le plus approprié : ils restent bien distincts les uns des autres et il n'y a pas de chevauchement entre eux.

# **CARTE DES 3 CLUSTERS RESTANTS**



# COORDONNEES PARALLELES POUR LES CENTROÏDES



Parmis les 3 clusters représentés, on voit que le cluster 1 représente le meilleur compromis, à la lecture des indicateurs, en comparaison des 2 autres clusters:

- La population n'est pas forcément la plus élevée, mais le coût de la viande par habitant et le taux d'inflation sur la nourriture, ainsi que la distance avec la France et la disponibilité alimentaire par habitant sont les plus bas, alors que leurs PIB par habitant, leur consommation de viande de poulet et leurs échanges commerciaux (y compris avec la France) sont les plus élevés.
- Ces données me confortent dans mon choix pour le cluster n°1, à savoir les pays suivants :

- **Allemagne.**
- **Arabie Saoudite.**
- **Belgique.**
- **Hong-Kong.**
- **Espagne.**
- **Pays-Bas.**
- **Pologne.**
- **Royaume-Uni**