

# Etude de marché



La poule qui chante





## Présentation du projet

La poule qui chante, une entreprise française d'agroalimentaire, souhaite se développer à l'international.

### Mission :

Réaliser une première analyse des groupements de pays à cibler pour l'exportation de poulets.

### Données :

Open data sur le site de la FAO, critères analyse PESTEL.

### Demandes :

Clustering par - classification ascendante hiérarchique  
- méthode des Kmeans

Analyse des variables et des corrélations ACP.



## Plan de la présentation



Préparation des données.



Clusterisation des pays.



Analyse des groupes et des variables.



Choix des pays cibles.



## Préparation des données

**df population** (source FAO, version en anglais, pour cartographie ultérieure)

Code pays

Pays

Année

Nombre d'habitant

+ une colonne accroissement de la population entre 2012 et 2017

⌵ Année 2017

pop\_2017

| Code pays |   | Pays           | Nombre d'habitant_2017 | Accroiss_pop |
|-----------|---|----------------|------------------------|--------------|
| 0         | 2 | Afghanistan    | 36296113               | 16.48        |
| 1         | 3 | Albania        | 2884169                | -1.03        |
| 2         | 4 | Algeria        | 41389189               | 10.71        |
| 3         | 5 | American Samoa | 55620                  | -0.08        |
| 4         | 6 | Andorra        | 77001                  | -6.58        |

**df PIB** (source Banque Mondiale)

Pays

Année

⌵ Année 2017

Harmonisation des intitulés des pays

PIB\_2017

|   | Pays        | PIB          |
|---|-------------|--------------|
| 0 | Aruba       | 3.092179e+09 |
| 1 | NaN         | 1.020647e+12 |
| 2 | Afghanistan | 1.875347e+10 |
| 3 | NaN         | 6.837487e+11 |
| 4 | Angola      | 6.897276e+10 |



## Préparation des données

### df DisponibiliteAlimentaire 2017

Code pays

Pays

Élément

Produit

Quantité

#### ▼ Produit

##### 16 sources protéiques

Viande de Bovins  
Viande d'Ovins/Caprins  
Viande de Suides  
Viande de Volailles  
Viande, Autre  
Abats Comestibles  
Œufs  
Poissons Eau Douce  
Perciform  
Poissons Pelagiques  
Poissons Marins, Autres  
Crustacés  
Cephalopodes  
Mollusques, Autres  
Animaux Aquatiques Autre  
Viande de Anim Aquatiq

#### ▼ Element

##### 4 éléments

Production  
Importations - Quantité  
Disponibilité alimentaire en quantité  
(kg/personne/an)  
Exportations - Quantité

#### Pivot :

**63 nouvelles colonnes**  
chaque élément pour chacun des produits

### dispo\_final

|     | Code pays | Pays        | Prod Viande de Bovins | Import Viande de Bovins | Dispo_alim Viande de Bovins | Prod Viande d'Ovins/Caprins | Import Viande d'Ovins/Caprins |
|-----|-----------|-------------|-----------------------|-------------------------|-----------------------------|-----------------------------|-------------------------------|
| 0   | 1         | Arménie     | 71.0                  | 3.0                     | 17.82                       | 11.0                        | 0.0                           |
| 1   | 2         | Afghanistan | 94.0                  | 6.0                     | 2.61                        | 160.0                       | 0.0                           |
| 2   | 3         | Albanie     | 39.0                  | 3.0                     | 14.67                       | 26.0                        | 0.0                           |
| 3   | 4         | Algérie     | 151.0                 | 48.0                    | 4.63                        | 289.0                       | 2.0                           |
| 4   | 7         | Angola      | 100.0                 | 34.0                    | 4.33                        | 24.0                        | 0.0                           |
| ... | ...       | ...         | ...                   | ...                     | ...                         | ...                         | ...                           |

Unités :

Importations, Production = millier de tonnes

Dispo\_alimentaire = kg/pers/an



## Préparation des données

### Fusion des tables

Fusion `dispo_final` et `pop_2017` = `df_interm`

Fusion `df_interm` et `PIB_2017` = `df_final`

`df_final`

|     | Code<br>pays | Pays        | Prod<br>Viande<br>de<br>Bovins | Import<br>Viande<br>de<br>Bovins | Dispo_alim<br>Viande de<br>Bovins | Prod Viande<br>d'Ovins/Caprins | Import Viande<br>d'Ovins/Caprins | Dispo_alim<br>Viande<br>d'Ovins/Caprins | Import<br>Viande<br>de<br>Suides | Dispo_alim<br>Viande de<br>Suides | ... | Import<br>Animaux<br>Aquatiques<br>Autre | Export<br>Animaux<br>Aquatiques<br>Autre | Dispo<br>Ani<br>Aquat |
|-----|--------------|-------------|--------------------------------|----------------------------------|-----------------------------------|--------------------------------|----------------------------------|---|----------------------------------|-----------------------------------|-----|--|--|-----------------------|
| 0   | 1            | Armenia     | 71.0                           | 3.0                              | 17.82                             | 11.0                           | 0.0                              | 2.95                                    | 10.0                             | 8.79                              | ... | 0.0                                      | 0.0                                      |                       |
| 1   | 2            | Afghanistan | 94.0                           | 6.0                              | 2.61                              | 160.0                          | 0.0                              | 4.15                                    | 0.0                              | 0.00                              | ... | 0.0                                      | 0.0                                      |                       |
| 2   | 3            | Albania     | 39.0                           | 3.0                              | 14.67                             | 26.0                           | 0.0                              | 9.11                                    | 9.0                              | 7.20                              | ... | 0.0                                      | 0.0                                      |                       |
| 3   | 4            | Algeria     | 151.0                          | 48.0                             | 4.63                              | 289.0                          | 2.0                              | 6.69                                    | 0.0                              | 0.00                              | ... | 0.0                                      | 0.0                                      |                       |
| 4   | 7            | Angola      | 100.0                          | 34.0                             | 4.33                              | 24.0                           | 0.0                              | 0.77                                    | 108.0                            | 7.50                              | ... | 0.0                                      | 0.0                                      |                       |
| ... | ...          | ...         | ...                            | ...                              | ...                               | ...                            | ...                              | ...                                     | ...                              | ...                               | ... | ...                                      | ...                                      | ...                   |
| 169 | 255          | Belgium     | 282.0                          | 99.0                             | 13.76                             | 3.0                            | 23.0                             | 1.04                                    | 209.0                            | 36.54                             | ... | 1.0                                      | 0.0                                      |                       |
| 170 | 256          | Luxembourg  | 10.0                           | 8.0                              | 24.64                             | 0.0                            | 1.0                              | 1.34                                    | 12.0                             | 33.98                             | ... | 0.0                                      | 0.0                                      |                       |
| 171 | 272          | Serbia      | 62.0                           | 2.0                              | 6.58                              | 32.0                           | 0.0                              | 3.18                                    | 36.0                             | 33.38                             | ... | 0.0                                      | 0.0                                      |                       |
| 172 | 273          | Montenegro  | 5.0                            | 5.0                              | 12.96                             | 1.0                            | 0.0                              | 2.15                                    | 27.0                             | 41.89                             | ... | 0.0                                      | 0.0                                      |                       |
| 173 | 276          | Sudan       | 332.0                          | 1.0                              | 7.72                              | 364.0                          | 0.0                              | 8.47                                    | 0.0                              | 0.00                              | ... | 0.0                                      | 0.0                                      |                       |

174 rows × 69 columns



# Clusterisation

## Traitement préalable : Centrage et réduction des données

Centrage : on soustrait à chaque valeur la moyenne des valeurs

Réduction : on divise chaque valeur par son écart type



les données sont indépendantes de l'unité ou de l'échelle choisie

**Principe**: Regroupement des individus par - variance / inertie interclasse maximale  
- variance / inertie intraclasse minimale



clusters les plus éloignés possibles  
clusters les plus resserrés possibles

### **Méthode 1 : Classification Ascendante Hierarchique**

Au sein de l'échantillon, appariement des individus en fonction de leur distance  
(méthode des liens)

Formation d'une arborescence

### **Méthode 2 : Algorithme des Kmeans ( ou des centres mobiles)**

Calcul d'un nombre optimal de cluster

Positionnement aléatoire des centroïdes au sein de l'échantillon

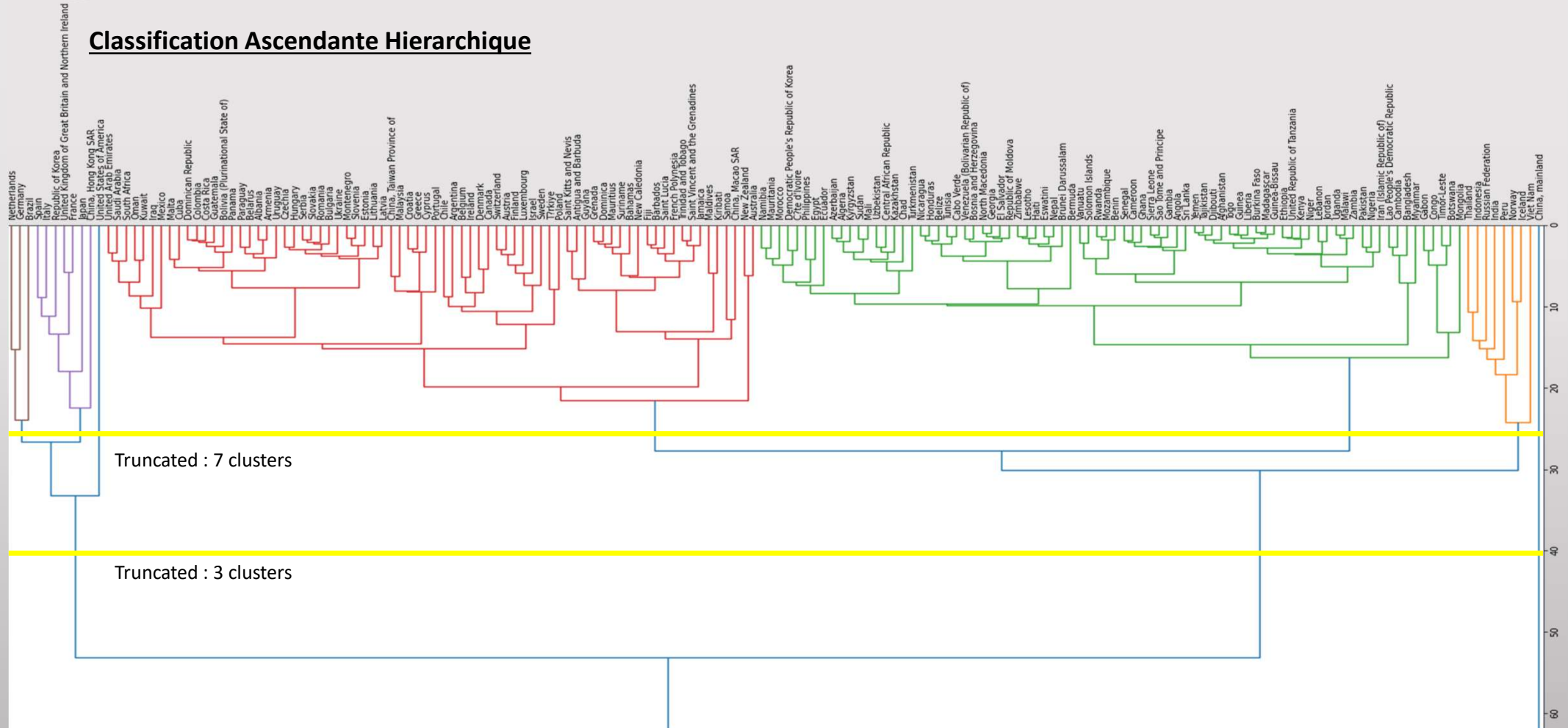
Regroupement des individus autour du centroïde le plus proche



# Clusterisation

**1<sup>ère</sup> approche** : étude sur les habitudes alimentaires en ressource protéique

## Classification Ascendante Hierarchique



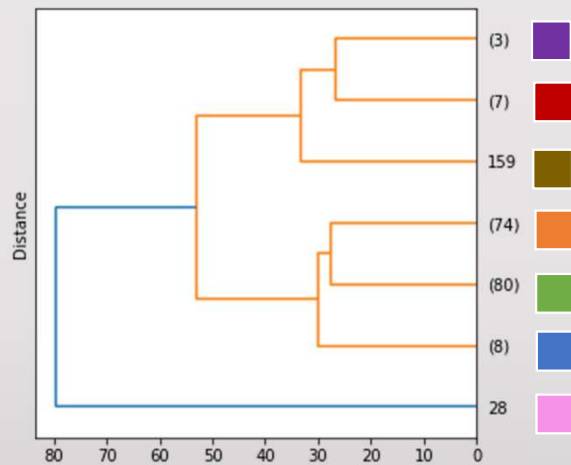




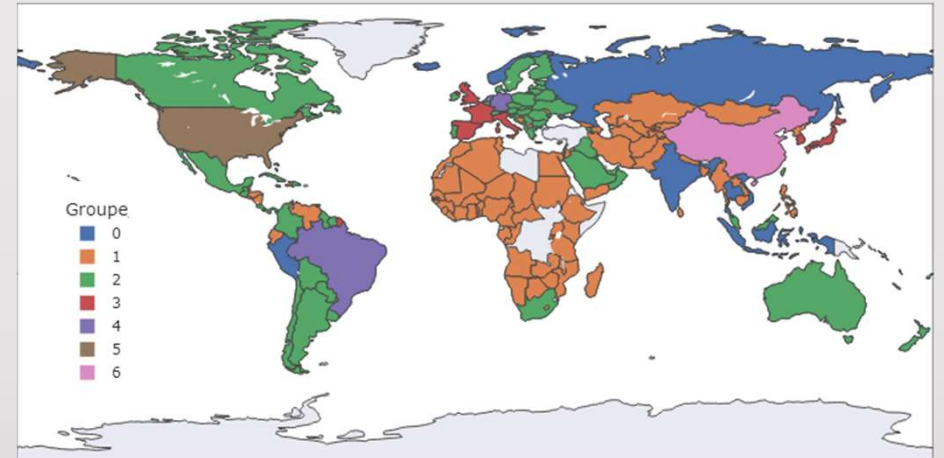
# Clusterisation

1<sup>ère</sup> approche : étude sur les habitudes alimentaires en ressource protéique

## Hierarchical Clustering Dendrogram

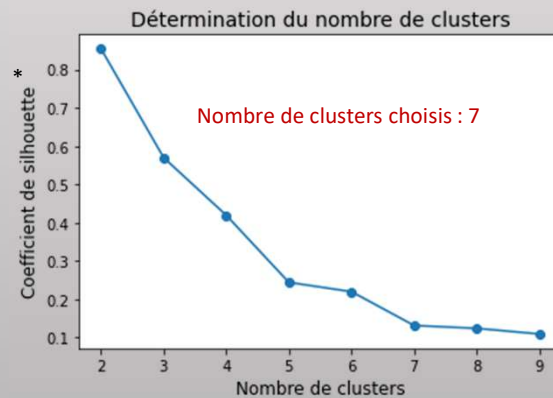


CAH

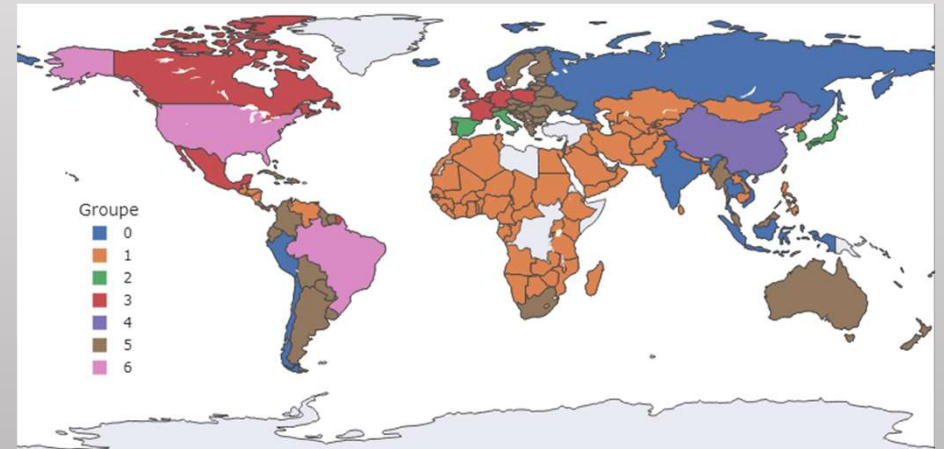


## Méthode des Kmeans

Methode du coude



Kmeans



\* Le coefficient de silhouette mesure le degré de séparation des clusters



# Analyse en composante principale APC

**Traitement préalable** : Centrage et réduction des données

**Principe** : - Créer de nouvelles variables en synthétisant les variables existantes  
- Projeter les données (individus) sur les plans définis par les nouvelles variables



Meilleure visualisation des données  
Corrélation entre les variables

**1 ère étape** : **Choix du nombre composantes principales** : Eboulis des valeurs propres  
on cherche à obtenir un maximum de variance avec un minimum de composantes

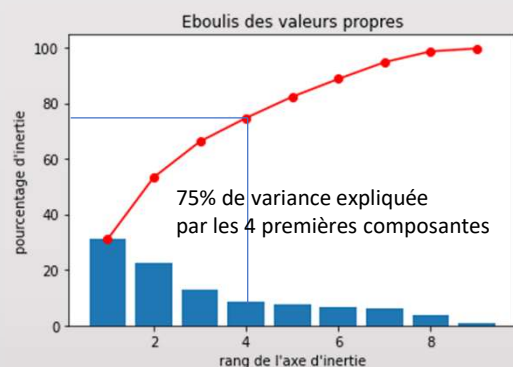
**2ème étape** : **Projection des variables sur des plans d'inertie maximale** (variance maximale) : Cercles des corrélations  
on observe les corrélations entre les variables et les composantes pour leur attribuer une signification (variables synthétiques)

**3ème étape** : **Projection des individus sur les différents plans factoriels**  
on observe les corrélations entre les individus et les composantes pour déterminer leurs caractéristiques  
on cherche à déterminer le meilleur plan de représentation (plan d'inertie maximale )



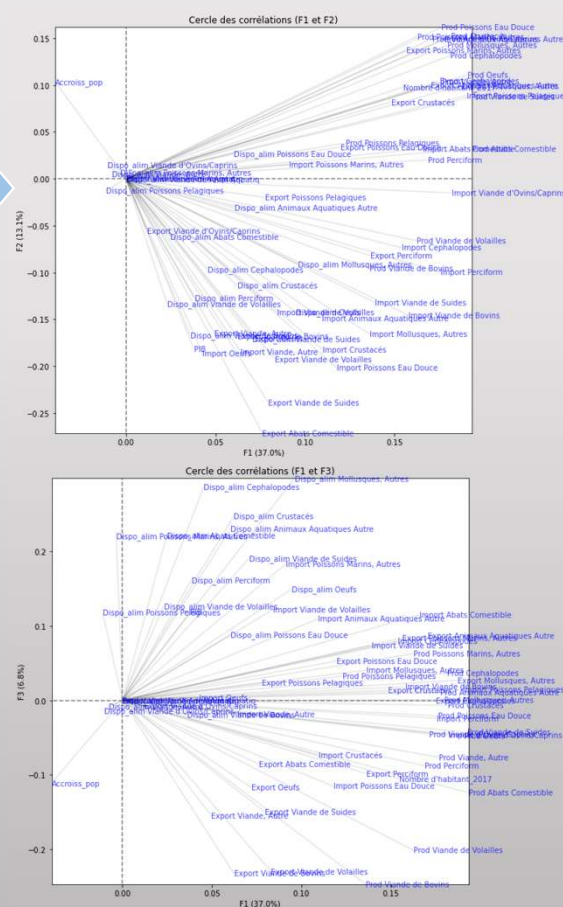
# Analyse en composante principale APC

## Choix du nombre composantes : 4

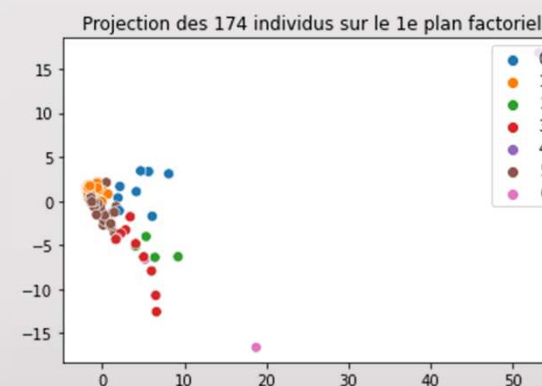


APC

## Cercles des corrélations



## Projection des individus sur le 1er plan factoriel



APC difficile à exploiter  
Clusters peu différenciés

Limitation de l'étude au produit volaille



# Préparation des données

2ème approche : étude sur le produit volaille

## df\_final

Pays  
Nombre d'habitant\_2017  
Accroiss\_pop  
PIB  
Prod Viande de Volailles  
Import Viande de Volailles  
Dispo\_alim Viande de Volailles  
Export Viande de Volailles

### + Calcul de nouveaux indicateurs

#### **Consommation apparente**

= Production + importation - exportation

#### **Dépendance aux importations**

= importation/ consommation apparente

#### **Taux de couverture de la consommation**

= (production-exportation)/consommation

#### **Taux d'auto-apvisionnement**

= production/consommation

## df\_volaille

|     | Pays        | Nombre d'habitant_2017 | Accroiss_pop | PIB       | Prod Viande de Volailles | Import Viande de Volailles | Dispo_alim Viande de Volailles | Export Viande de Volailles | Conso | Dependance | Taux de couverture | Auto_approv |
|-----|-------------|------------------------|--------------|-----------|--------------------------|----------------------------|--------------------------------|----------------------------|-------|------------|--------------------|-------------|
| 0   | Armenia     | 2944791                | 2.10         | 3914.53   | 11.0                     | 35.0                       | 16.06                          | 0.0                        | 46.0  | 0.760870   | 0.239130           | 0.239130    |
| 1   | Afghanistan | 36296113               | 16.48        | 516.68    | 28.0                     | 29.0                       | 1.53                           | 0.0                        | 57.0  | 0.508772   | 0.491228           | 0.491228    |
| 2   | Albania     | 2884169                | -1.03        | 4514.19   | 13.0                     | 38.0                       | 16.36                          | 0.0                        | 51.0  | 0.745098   | 0.254902           | 0.254902    |
| 3   | Algeria     | 41389189               | 10.71        | 4109.70   | 275.0                    | 2.0                        | 6.38                           | 0.0                        | 277.0 | 0.007220   | 0.992780           | 0.992780    |
| 4   | Angola      | 29816766               | 18.75        | 2313.22   | 42.0                     | 277.0                      | 10.56                          | 0.0                        | 319.0 | 0.868339   | 0.131661           | 0.131661    |
| ... | ...         | ...                    | ...          | ...       | ...                      | ...                        | ...                            | ...                        | ...   | ...        | ...                | ...         |
| 169 | Belgium     | 11419748               | 3.02         | 44025.90  | 463.0                    | 338.0                      | 12.65                          | 656.0                      | 145.0 | 2.331034   | -1.331034          | 3.193103    |
| 170 | Luxembourg  | 591910                 | 11.50        | 111017.18 | 0.0                      | 11.0                       | 18.33                          | 1.0                        | 10.0  | 1.100000   | -0.100000          | 0.000000    |
| 171 | Serbia      | 8829628                | -1.24        | 5003.50   | 85.0                     | 12.0                       | 10.16                          | 7.0                        | 90.0  | 0.133333   | 0.866667           | 0.944444    |
| 172 | Montenegro  | 627563                 | 0.26         | 7738.83   | 4.0                      | 8.0                        | 15.98                          | 0.0                        | 12.0  | 0.666667   | 0.333333           | 0.333333    |

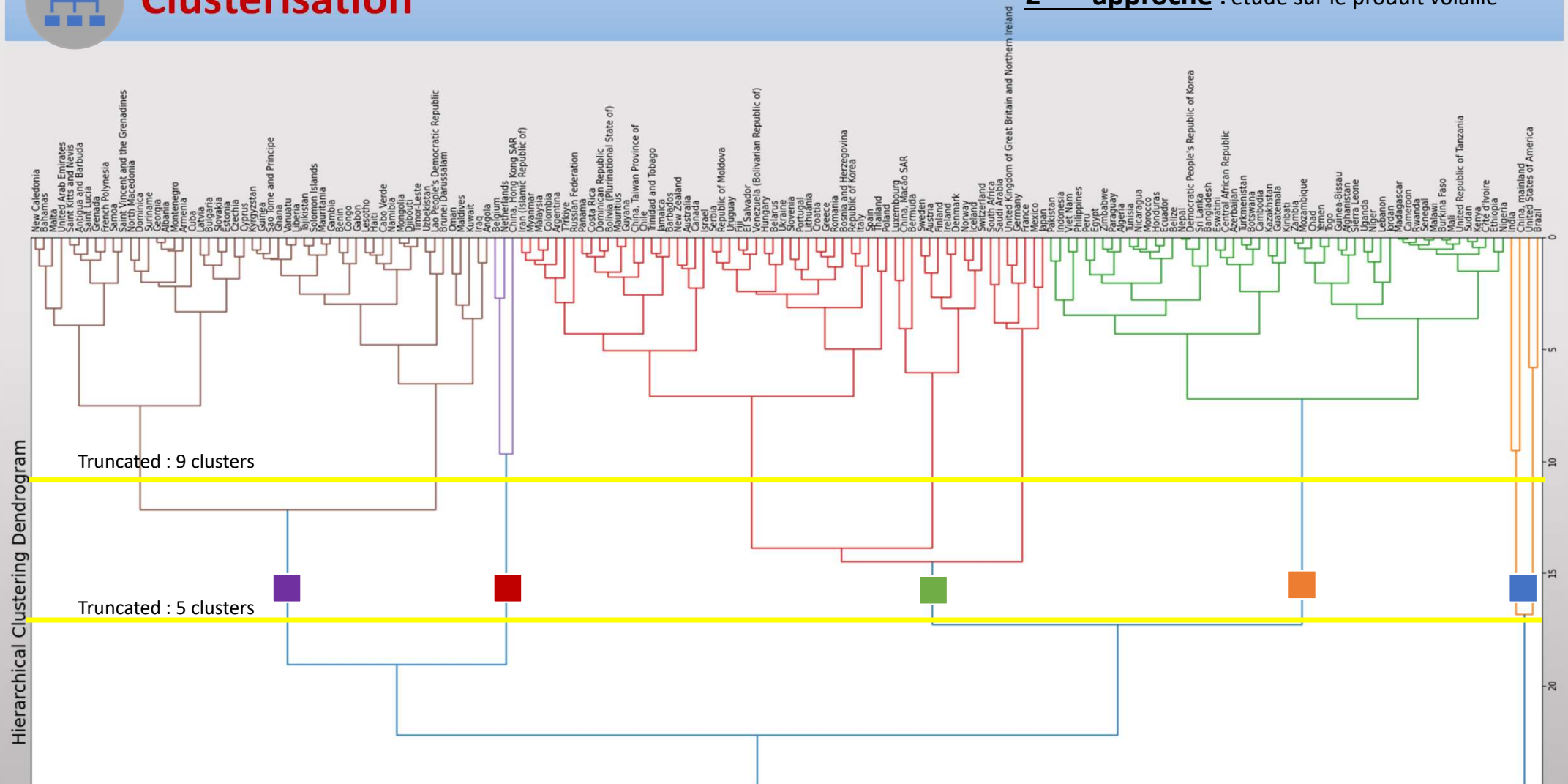
Unités :

Importations, Production, Consommation = millier de tonnes

Dispo\_alimentaire = kg/pers/an



## 2ème approche : étude sur le produit volaille

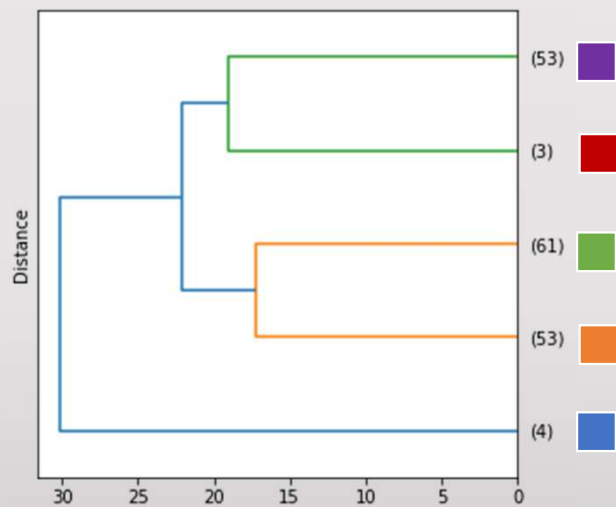




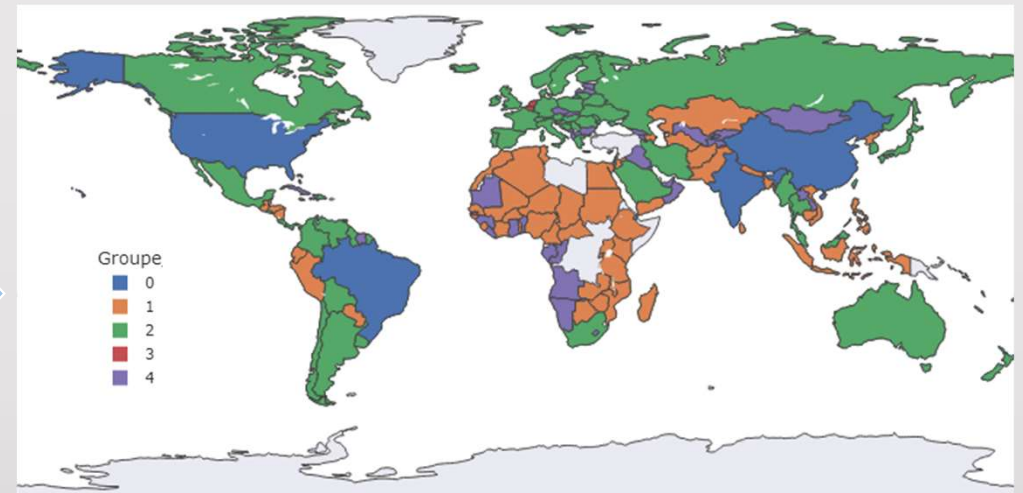
# Clusterisation

2<sup>ème</sup> approche : étude sur le produit volaille

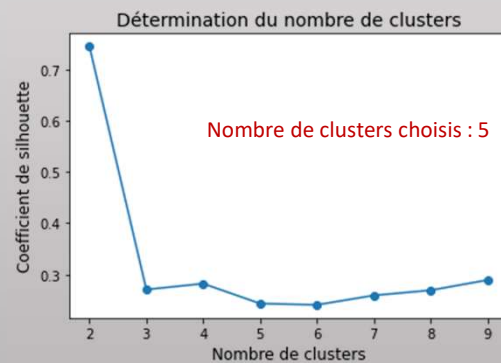
## Hierarchical Clustering Dendrogram



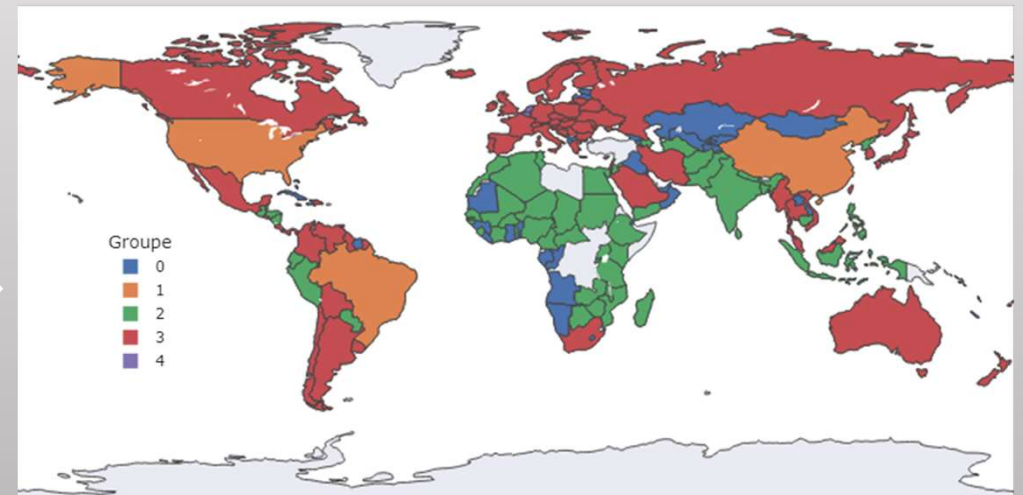
CAH



## Méthode des Kmeans



Kmeans



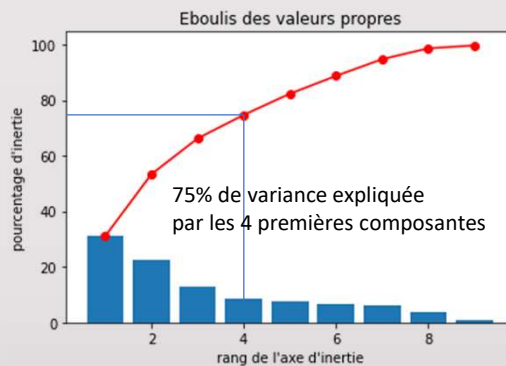




# Analyse en composante principale APC

2<sup>ème</sup> approche : étude sur le produit volaille

## Choix du nombre composantes : 4

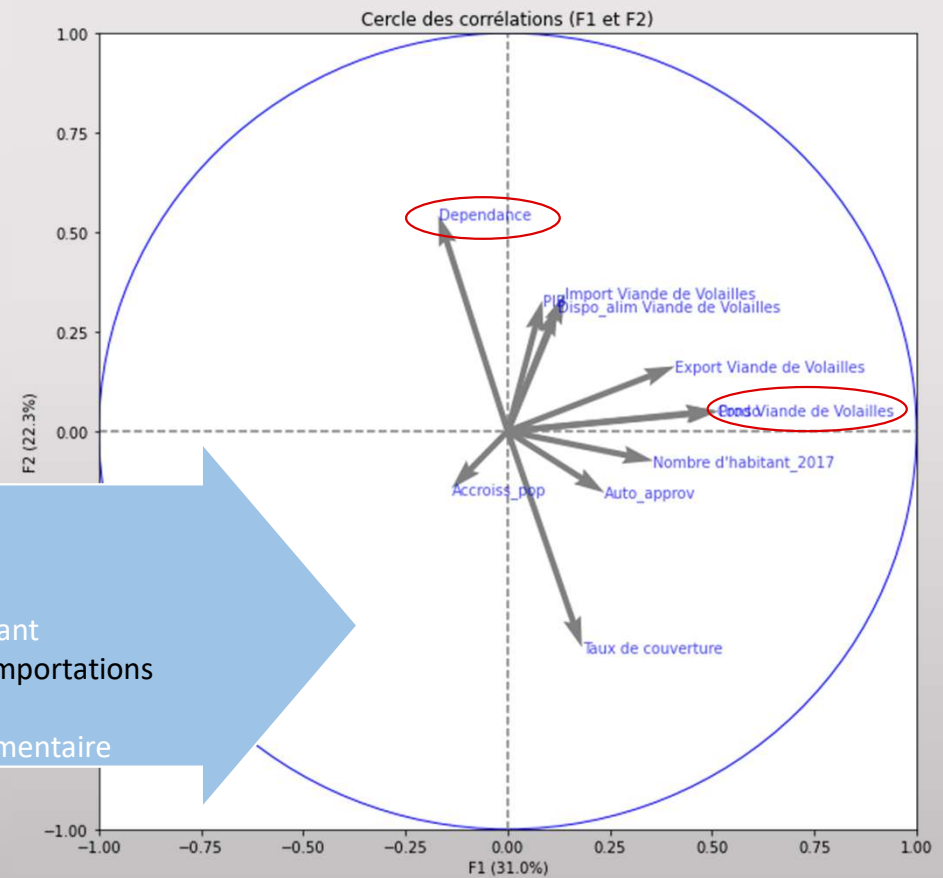


APC

## Cercles des corrélations

F1 : représente la **production**  
elle est liée à la consommation  
l'export  
le nombre d'habitant

F2 : représente la **dépendance aux importations**  
elle est liée à l'importation  
la disponibilité alimentaire

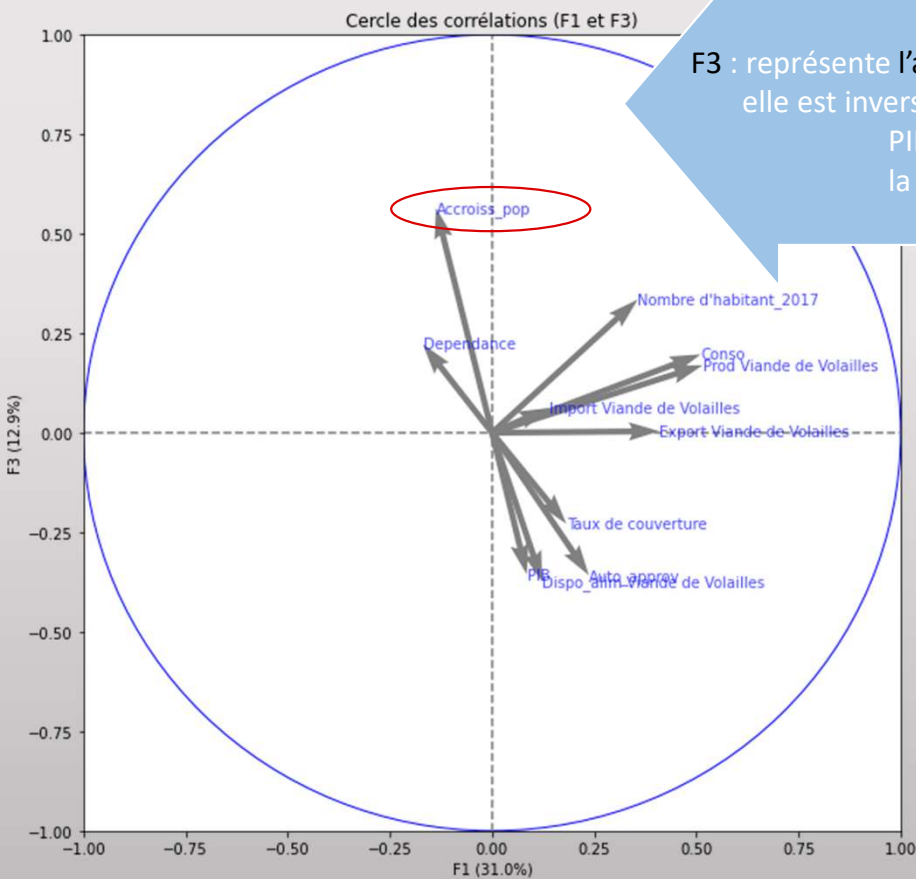




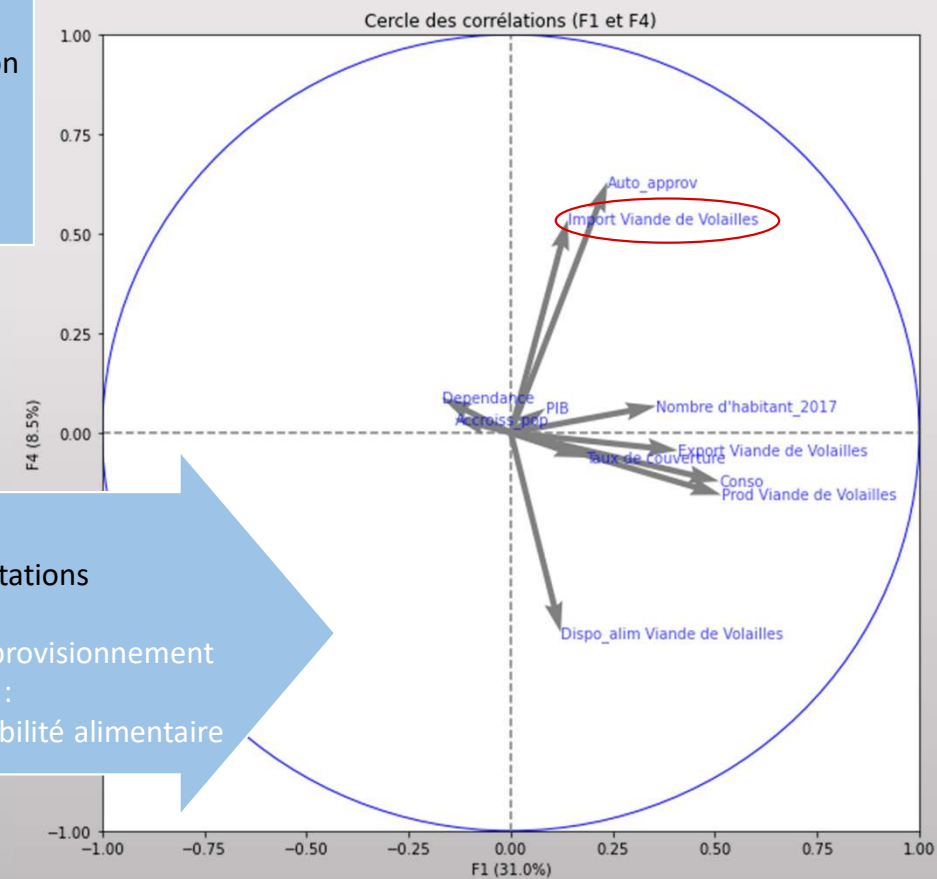
# Analyse en composante principale APC

2<sup>ème</sup> approche : étude sur le produit volaille

## Cercles des corrélations



F3 : représente l'accroissement de population  
elle est inversement liée au :  
PIB  
la disponibilité alimentaire



F4 : représente les importations  
elle est liée à :  
l'auto-approvisionnement  
et inversement liée à :  
la disponibilité alimentaire

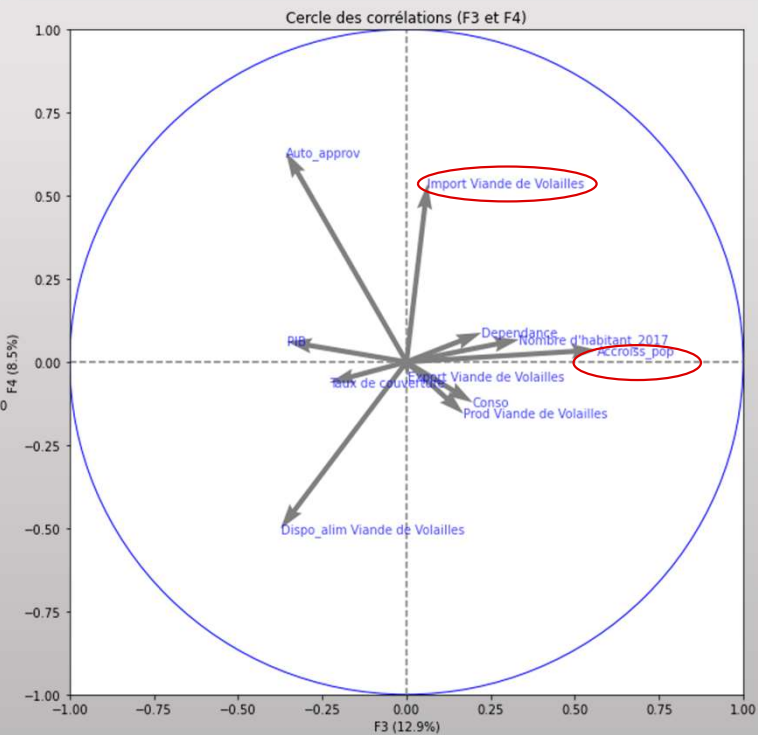
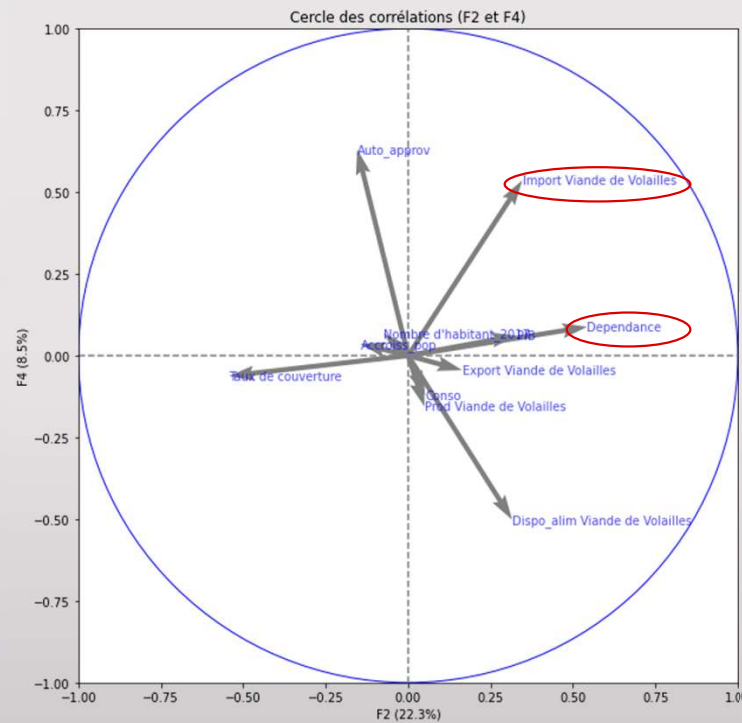
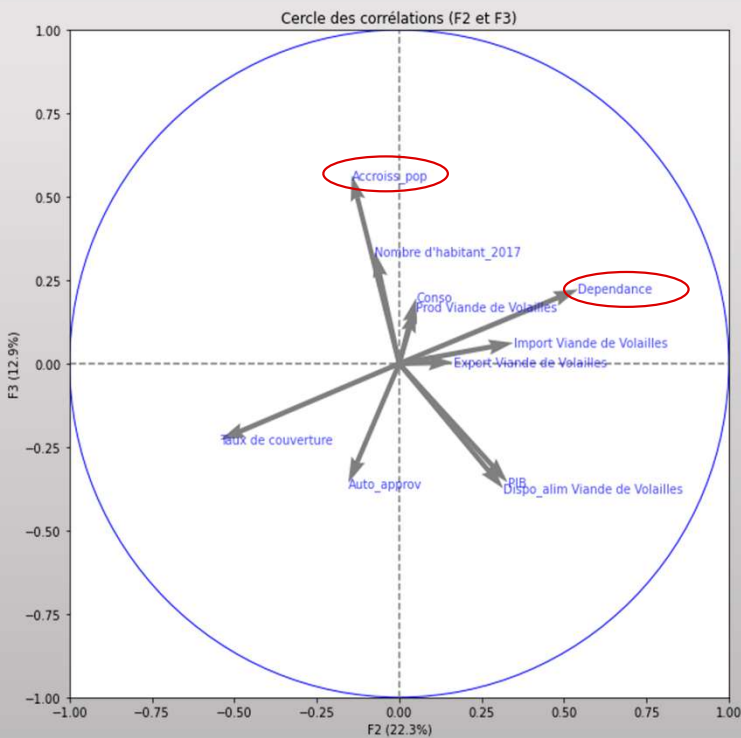




# Analyse en composante principale APC

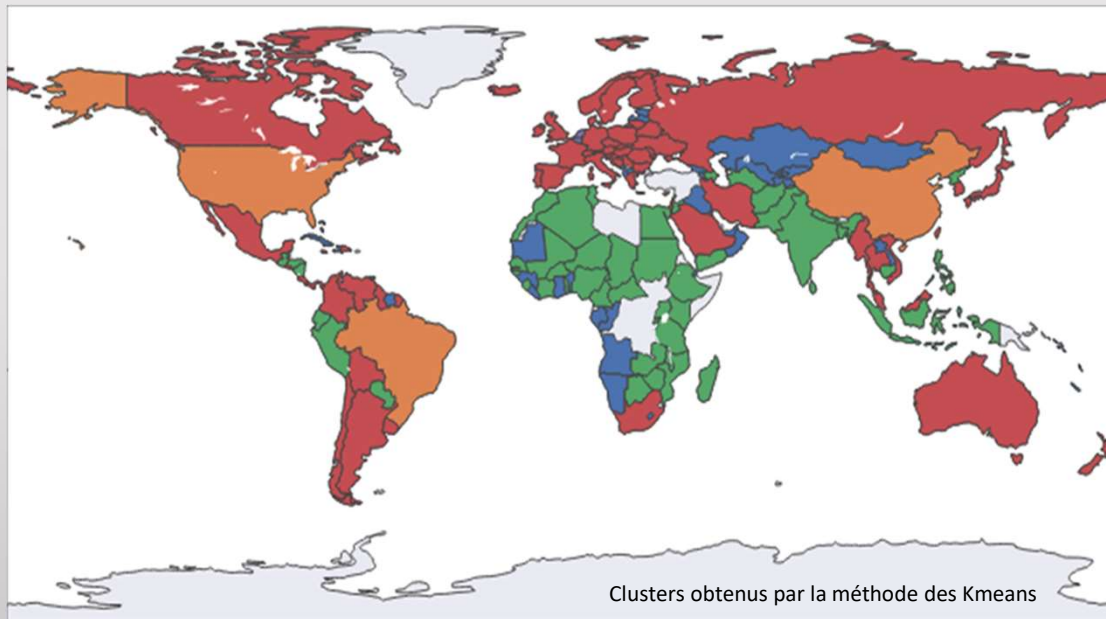
2<sup>ème</sup> approche : étude sur le produit volaille

## Cercles des corrélations

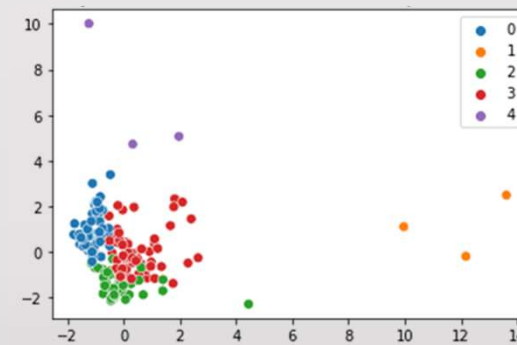




# Analyse en composante principale APC

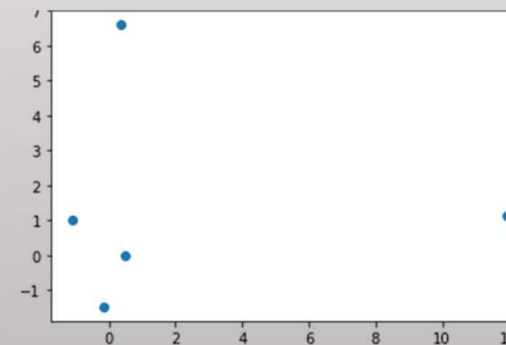


## Projection des individus sur le 1er plan factoriel



Les 5 clusters  
apparaissent  
clairement

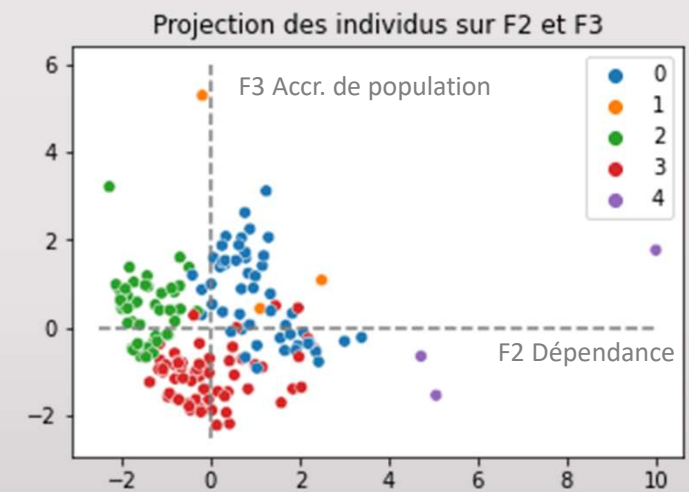
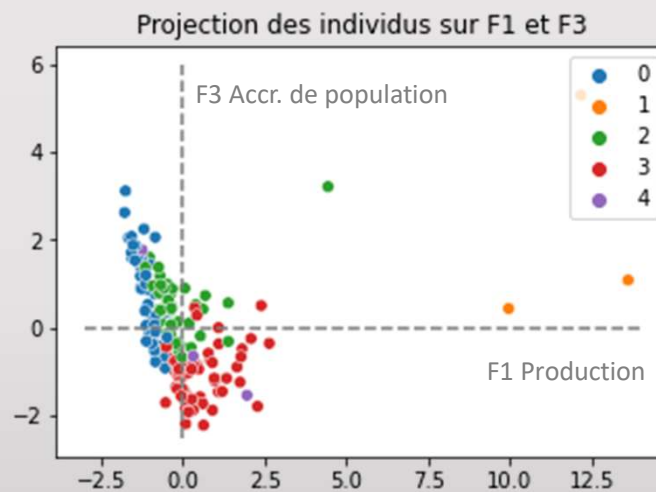
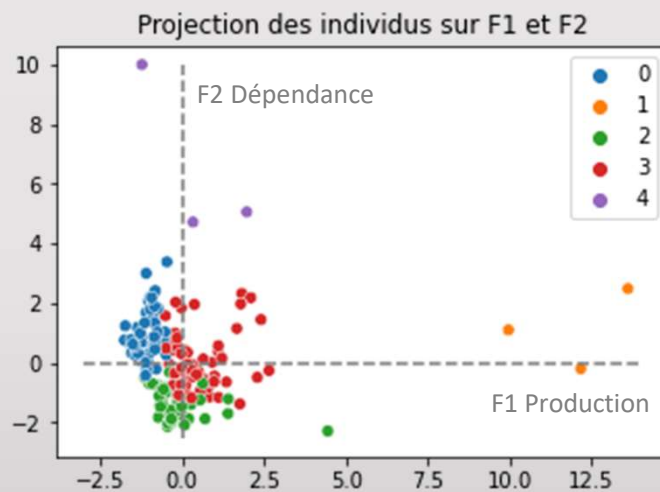
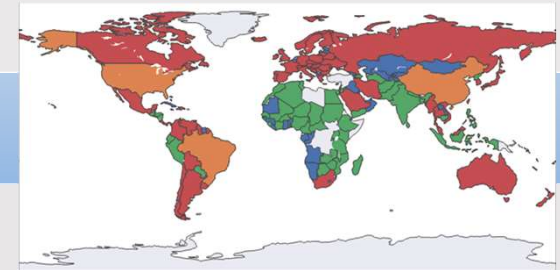
## Projection des centroïdes sur le 1er plan factoriel



Les centroïdes  
sont bien séparés



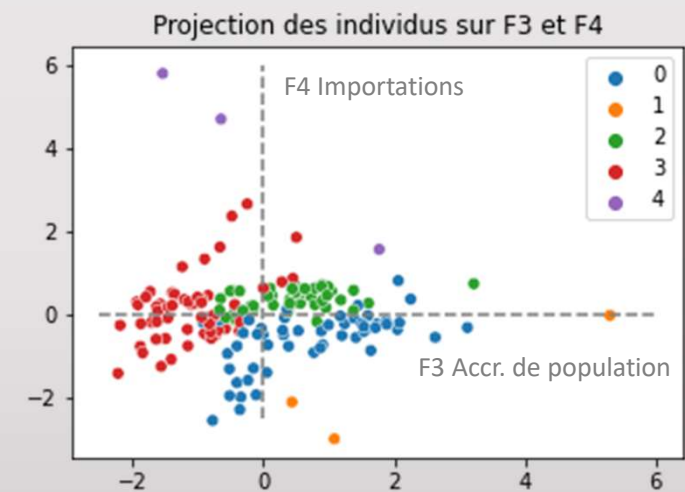
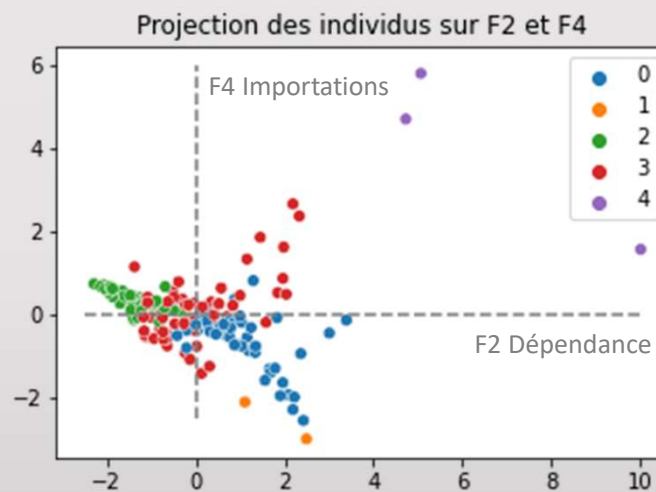
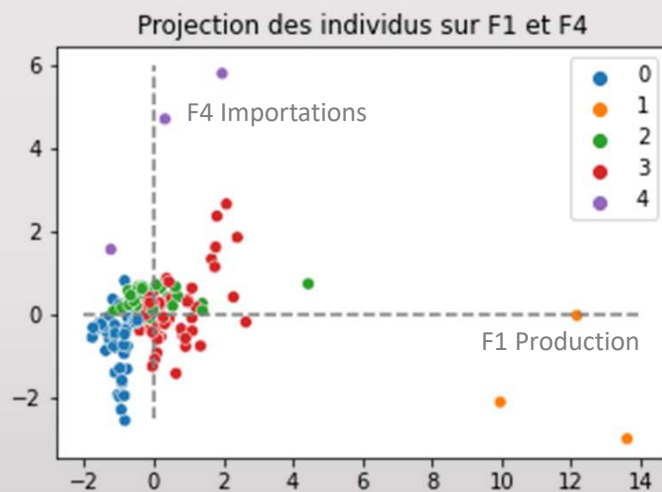
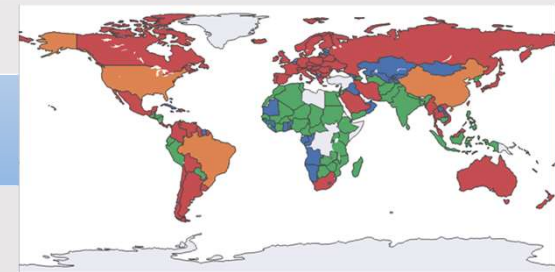
## Analyse des groupes



- Gr 0 : corrélé à la dépendance et inversement corrélé production
- Gr 1 : fortement corrélé à la production
- Gr 2 : corrélé à l'accroissement de population et inversement corrélé à la dépendance
- Gr 3 : inversement corrélé à l'accroissement de population
- Gr 4 : fortement corrélé à la dépendance alimentaire



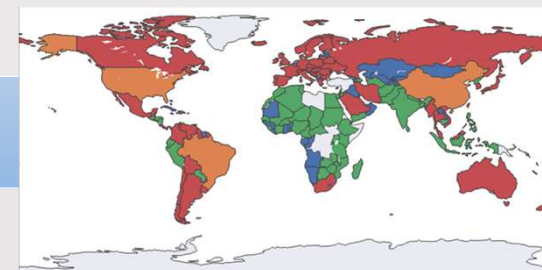
## Analyse des clusters



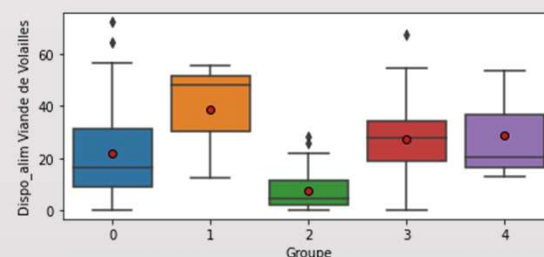
- Gr 0 : corrélé à la dépendance et l'accr. de population  
inversement corrélé à importation et production
- Gr 1 : fortement corrélé à la production
- Gr 2 : inversement corrélé à la dépendance alim.
- Gr 3 : corrélé aux importations,  
inversement corrélé à l'accr. de population
- Gr 4 : fortement corrélé aux importations



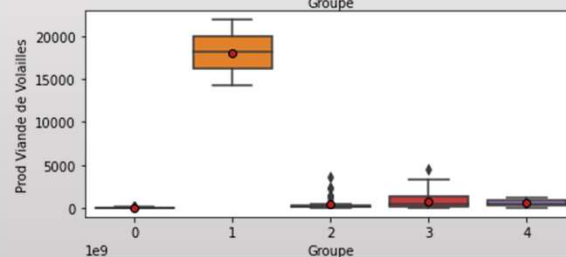
## Analyse des groupes



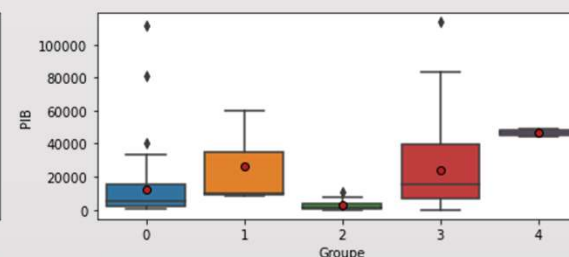
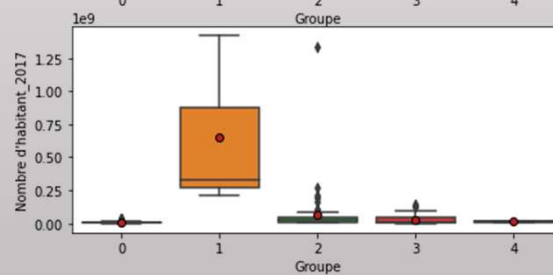
Gr2 : disponibilité alimentaire la plus faible



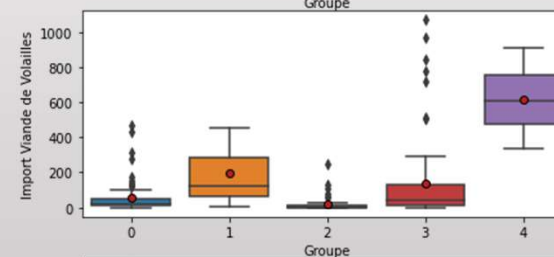
Gr1 : forte production



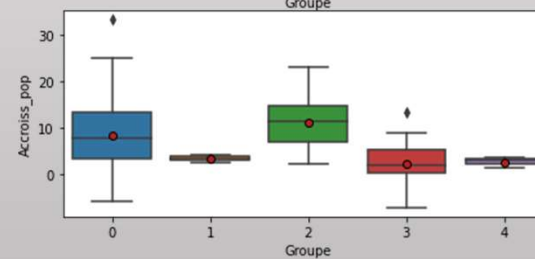
Gr1 : pays très peuplés



Gr4 : PIB élevé  
Gr 1 & 3 : PIB moyen  
Gr 0 & 2 : PIB faible



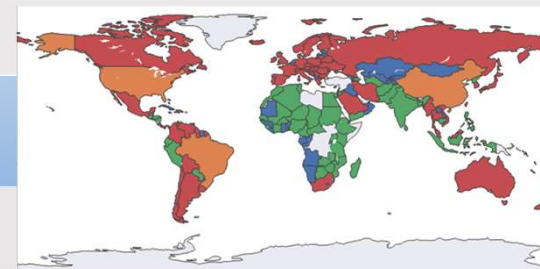
Gr4 : plus fortes importations



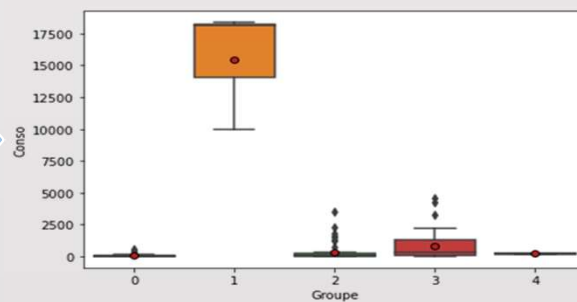
Gr0 & 2 : plus fort accroissement de population



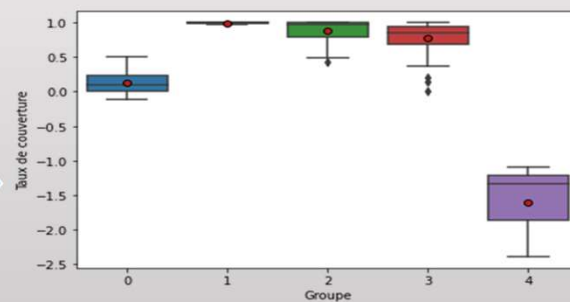
## Analyse des groupes



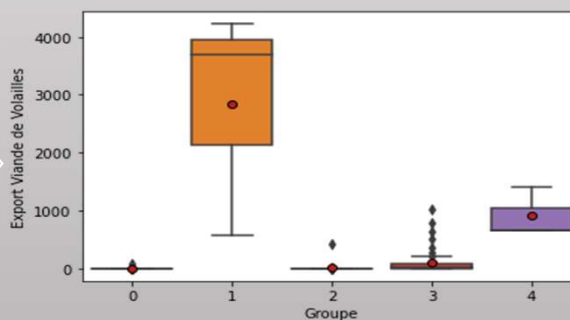
Gr1 : forte consommation



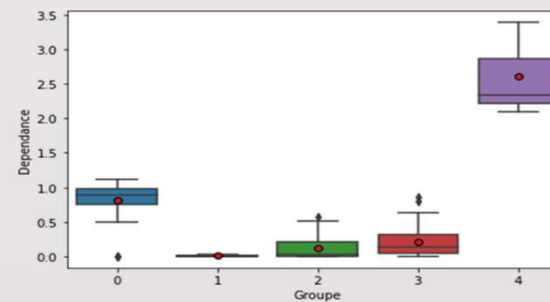
Gr4 : taux de couverture de la consommation par la production faible



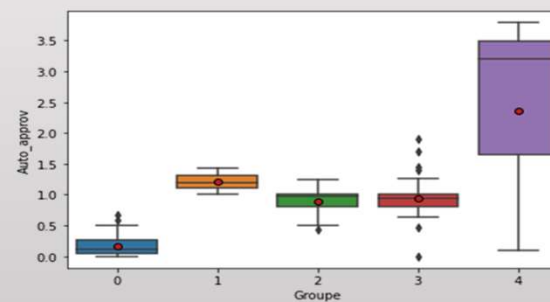
Gr1 : plus fortes exportations



Gr4 : forte importation, faible consommation

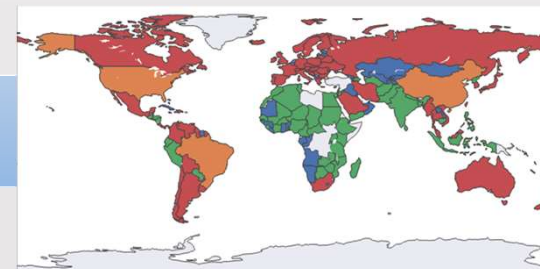


Gr0 : faible taux d'auto-approvisionnement

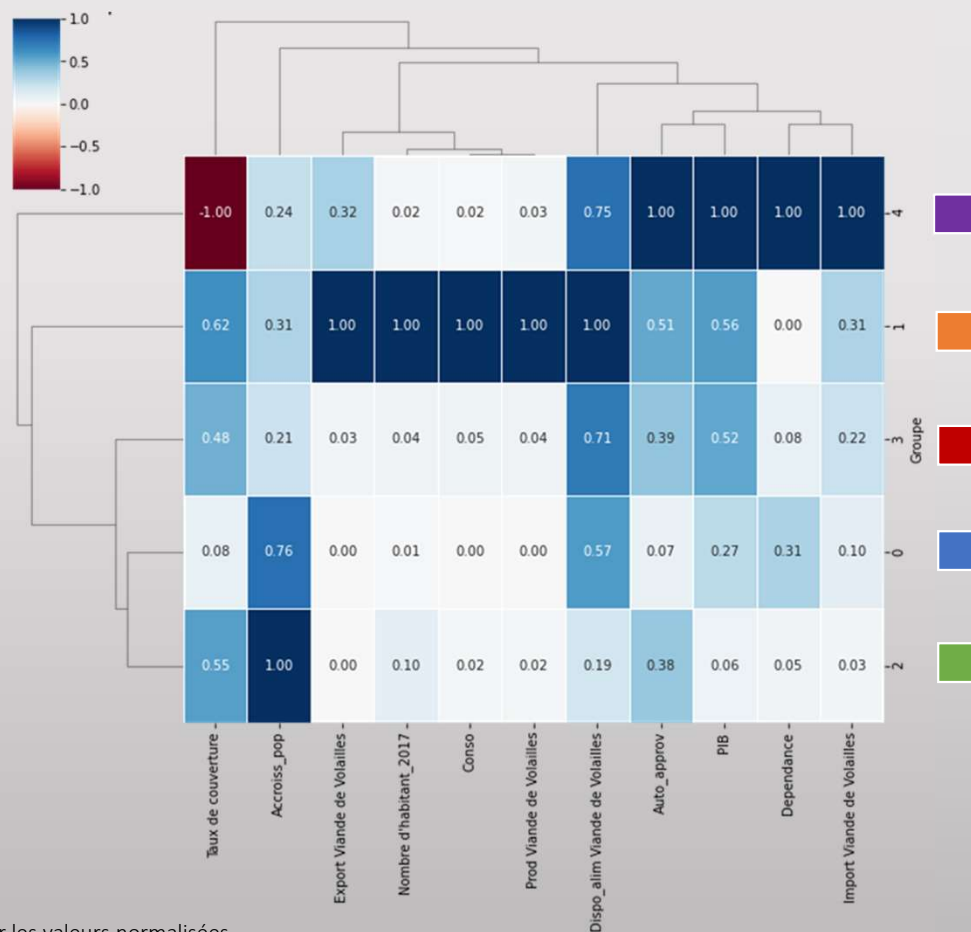




# Analyse des groupes



## Clustermap entre les groupes et les variables



Gr 4 : fortement corrélé au PIB et aux importations

Gr 1 : fortement corrélé à l'export, la consommation, la production et à la population

Gr 3 : corrélé à la disponibilité, au PIB

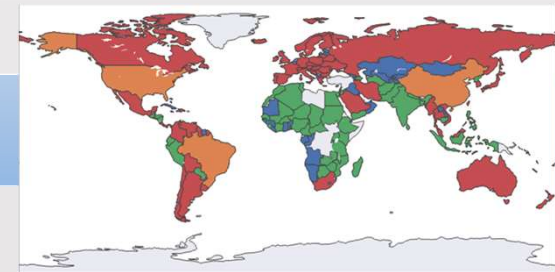
Gr 0 : fortement corrélé à l'accroissement de population et à la disponibilité alimentaire

Gr 2 : fortement corrélé à l'accroissement de population





# Analyse des variables



## Corrélations entre les variables



Variables fortement corrélées :



Consommation / production

Production / exportation

Consommation / exportation

Production / nombre d'habitant

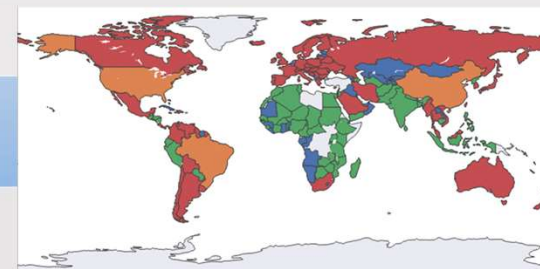


Dépendance / taux de couverture

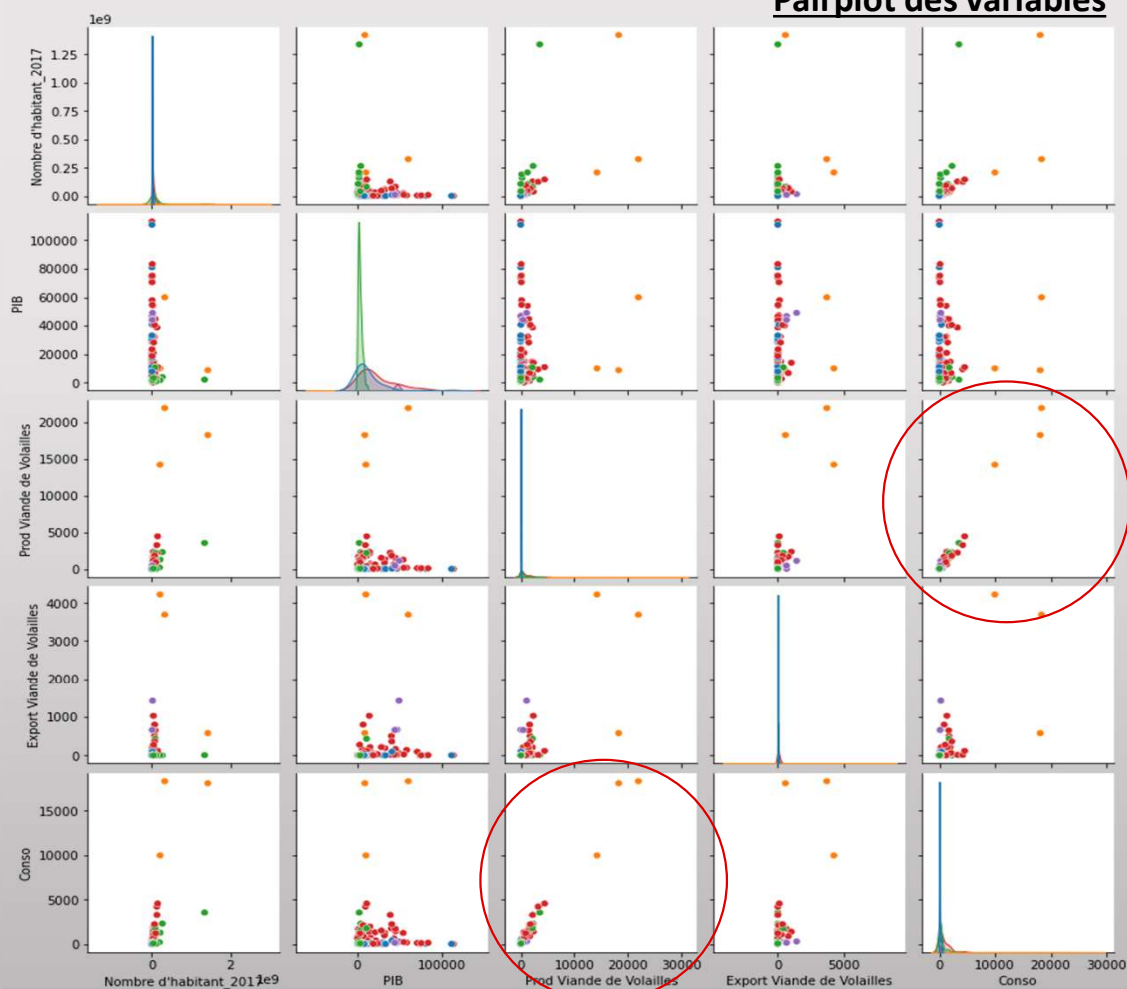




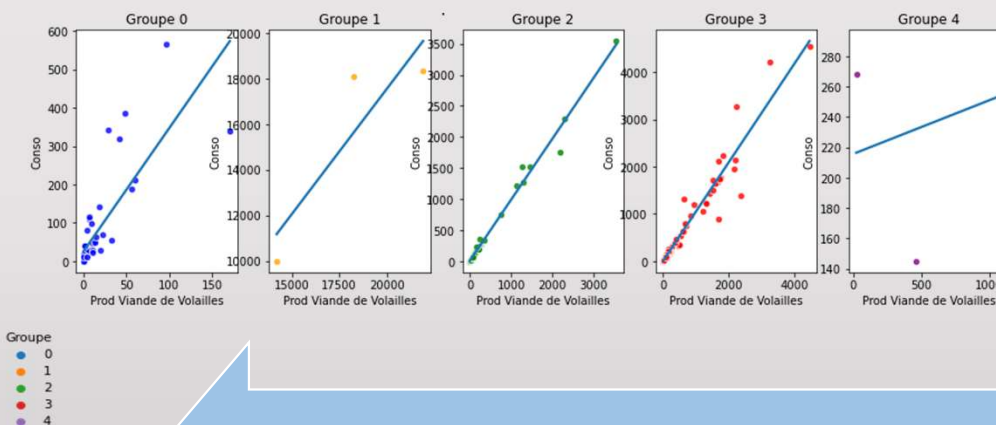
# Analyse des variables



**Pairplot des variables**



**Corrélations entre production et consommation**



Corrélation linéaire entre production et consommation

En particulier pour les Gr2 et Gr3

→ production pour consommation nationale



## Choix des pays cibles

### Bilan des études réalisées

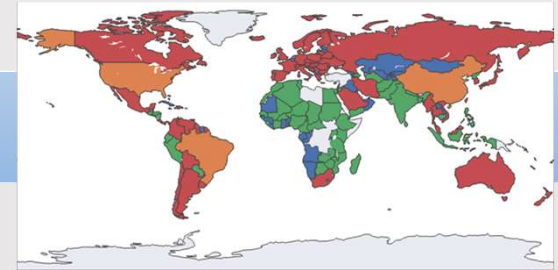
**Gr 0** : - PIB faible & plus fort accroissement de population  
- pays dépendants, faible production

**Gr 1** : - plus fortes exportations, forte production, forte consommation, PIB moyen & pays très peuplés

**Gr 2** : - plus fort accroissement de population, disponibilité alimentaire la plus faible, PIB faible

**Gr 3** : - PIB moyen, pays intermédiaires

**Gr 4** : - les plus fortes importations, exportations importantes, faible consommation, PIB élevé



**Gr 0 & Gr 2** : pays en demande d'importations mais faible PIB

**Gr 1** : pays producteurs et exportateurs

**Gr 3** : pays en demande d'importations et PIB moyen

**Gr 4** : pays transformateurs de volailles



## Choix des pays cibles

Pays cibles : **Groupe 3**

### Importation en viande de volaille

