



La poule qui chante

Etude de marché

Vers quel(s) marché(s) orienter
notre production ?



SOMMAIRE

1. CONTEXTE ET TRAITEMENT MÉTHODOLOGIQUE
2. METHODES DE CLUSTERING ET ANALYSE DES GROUPES
3. ANALYSE DES COMPOSANTES PRINCIPALES
4. BILAN ET RECOMMANDATIONS



1. CONTEXTE ET TRAITEMENT MÉTHODOLOGIQUE

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

En tant que Data Analyst chez *La poule qui chante*, l'équipe Marketing nous demande de :

- **Produire une analyse des groupements de pays ,**
- **Définir les groupements de pays à cibler pour la stratégie d'internationalisation.**

Les données utilisées sont issues de la Food and Agriculture Organization (FOA), à savoir :

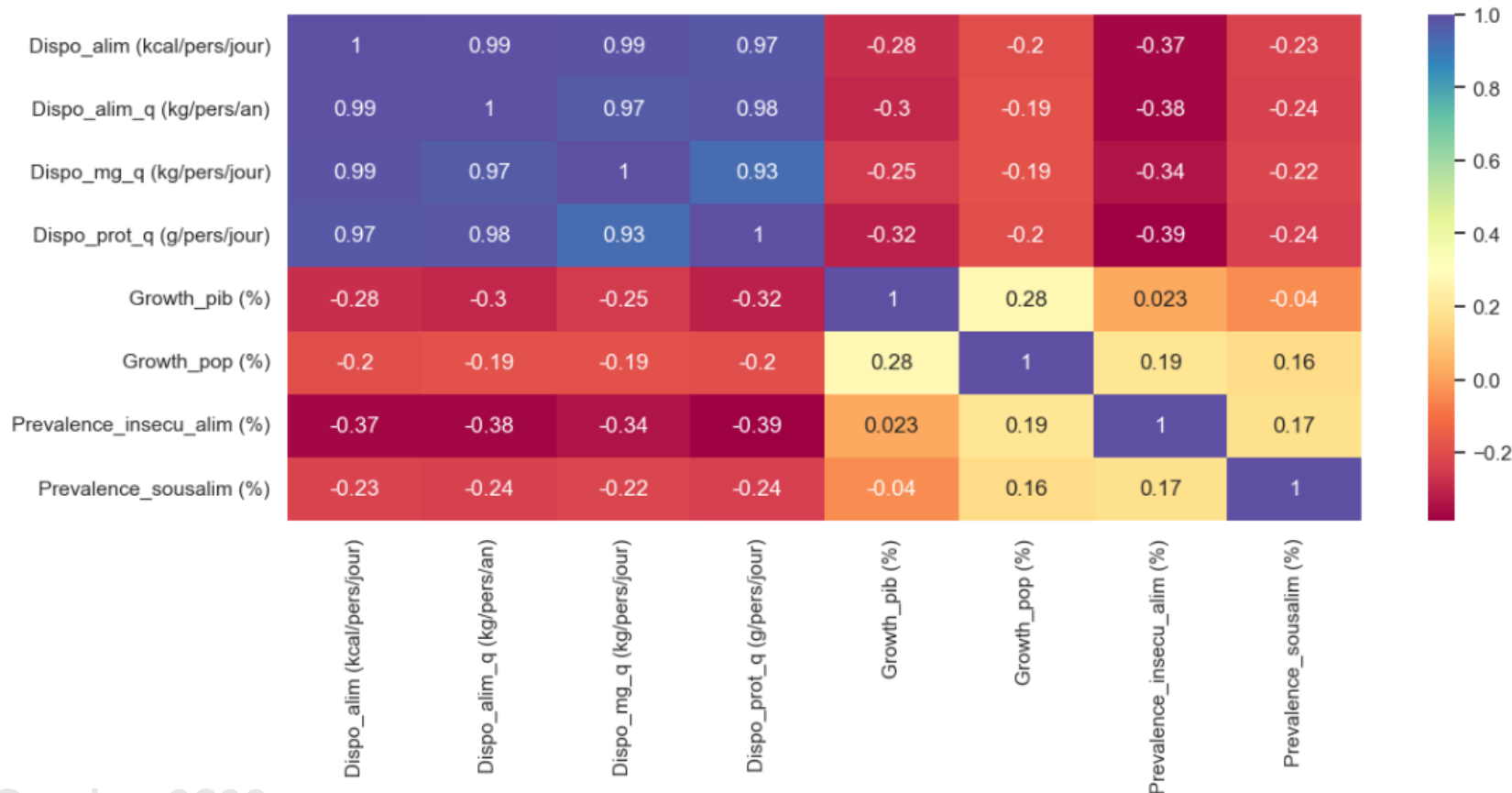
- **Les données de disponibilités alimentaires (2018),**
- **Les données de PIB et de croissance démographiques (2000-2018),**
- **Les données d'insécurité alimentaire (2018).**

A partir de ces données, certains indicateurs ont été créés comme le taux de croissance démographique et économique (2000-2018).



1. CONTEXTE ET TRAITEMENT MÉTHODOLOGIQUE

Après nettoyage des données, nous obtenons notre fichier d'analyse contenant près de **8 variables**

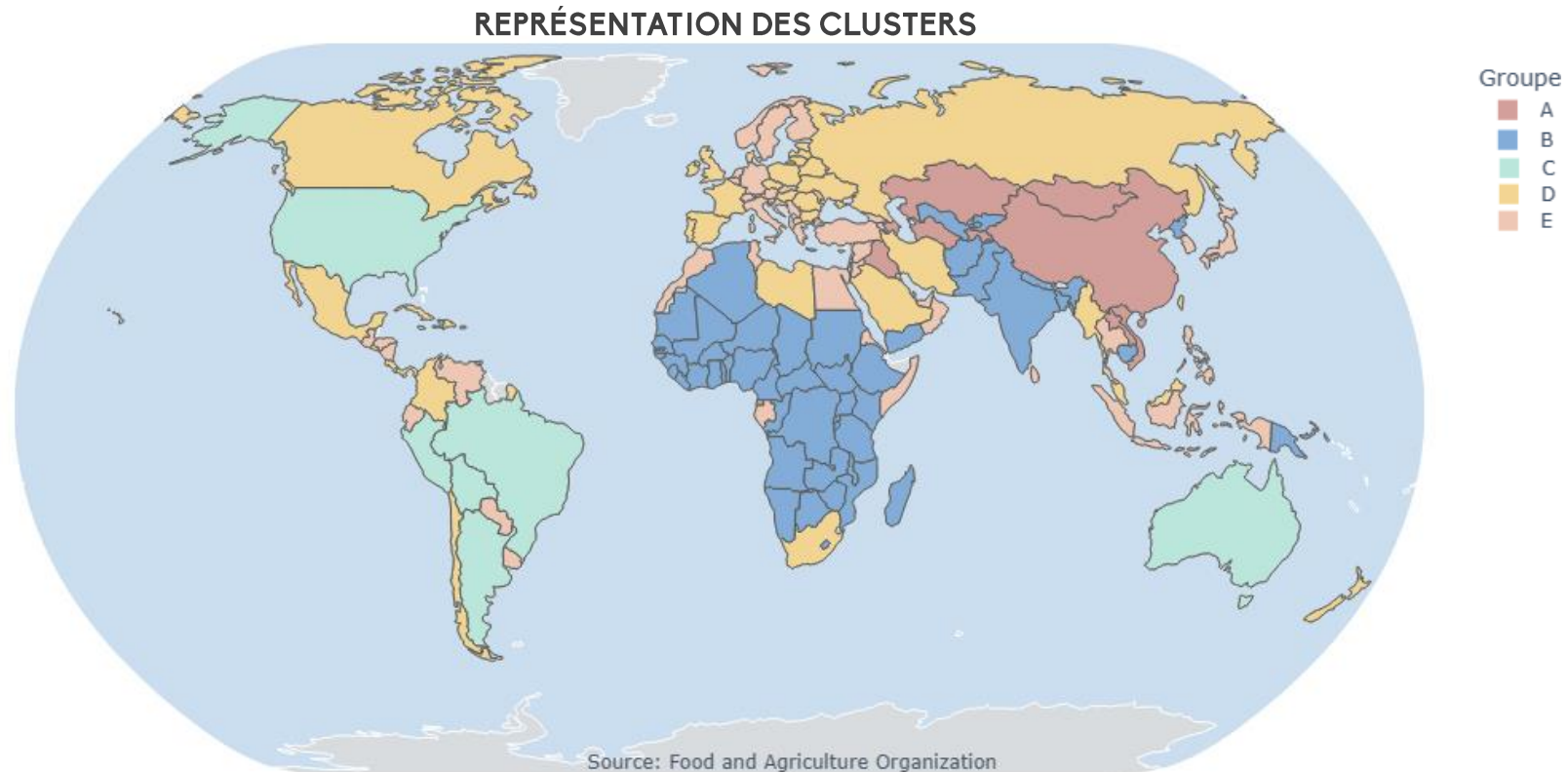
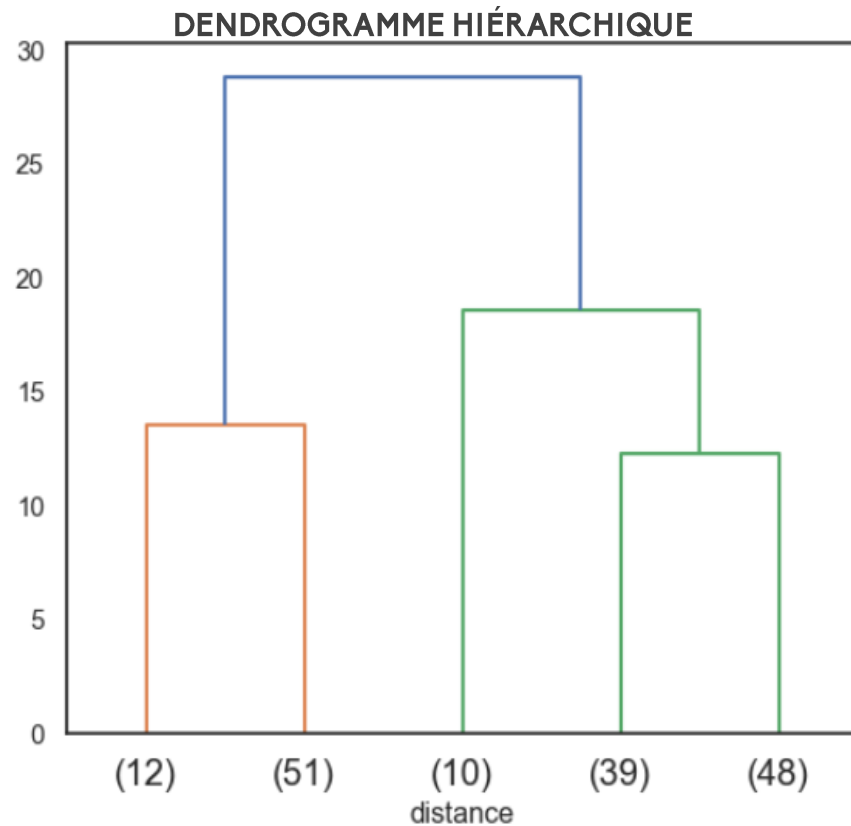


- Les **disponibilités** sont très corrélées entre elles.
- La **croissance démographique et économiques** sont négativement corrélées aux autres variables.
- La **sous-alimentation et l'insécurité alimentaire** sont négativement corrélées aux disponibilités alimentaires.



2. METHODES DE CLUSTERING ET ANALYSE DES GROUPEs

Méthode de Classification Ascendante Hiérarchique (CAH)



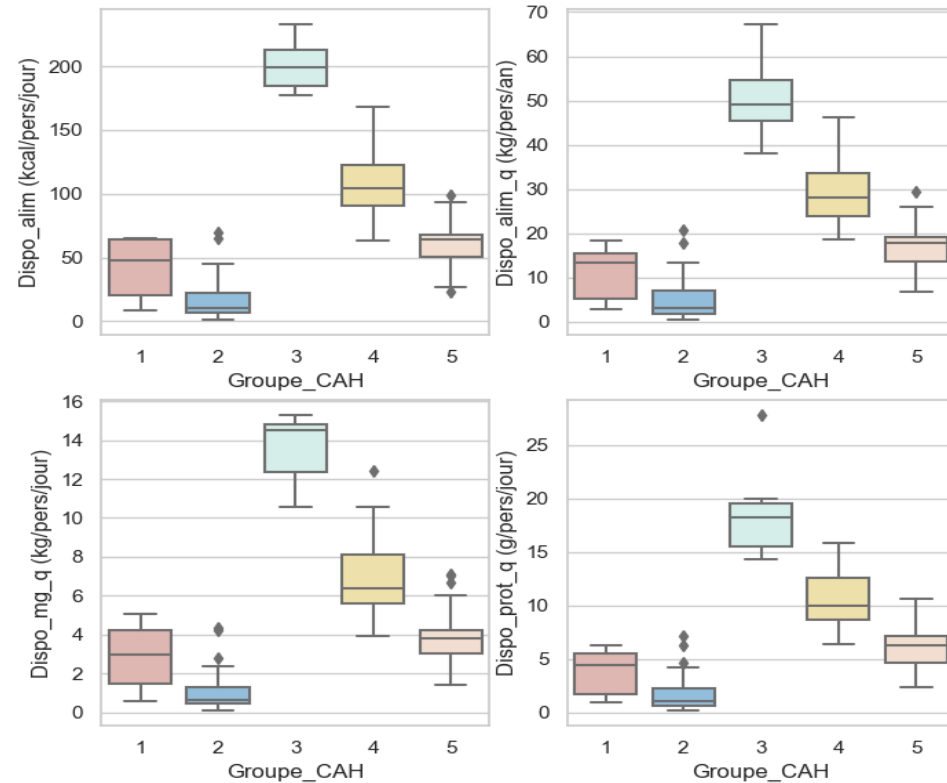


2. METHODES DE CLUSTERING ET ANALYSE DES GROUPES

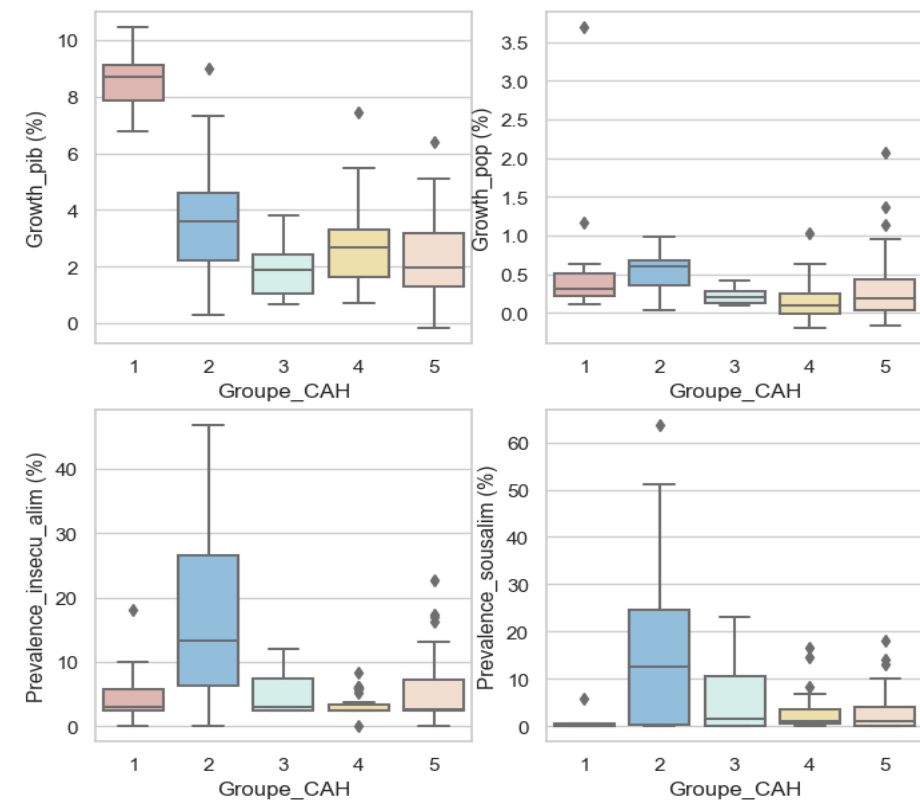
Méthode de Classification Ascendante Hiérarchique (CAH)

DISTRIBUTION DES VALEURS PAR REGROUPEMENT CAH

Variables de disponibilité alimentaire



Variables de croissance et d'insécurité

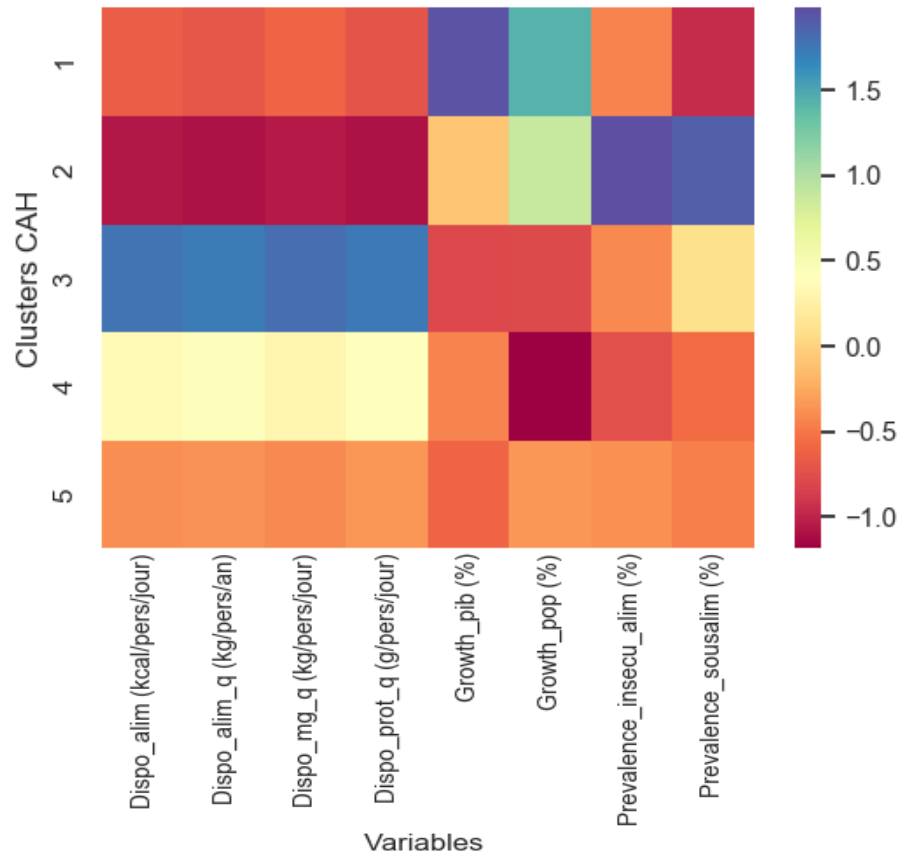




2. METHODES DE CLUSTERING ET ANALYSE DES GROUPES

Méthode de Classification Ascendante Hiérarchique (CAH)

HEATMAP DES VALEURS MOYENNES POUR LA CAH



CARACTÉRISTIQUES DES 5 GROUPES:

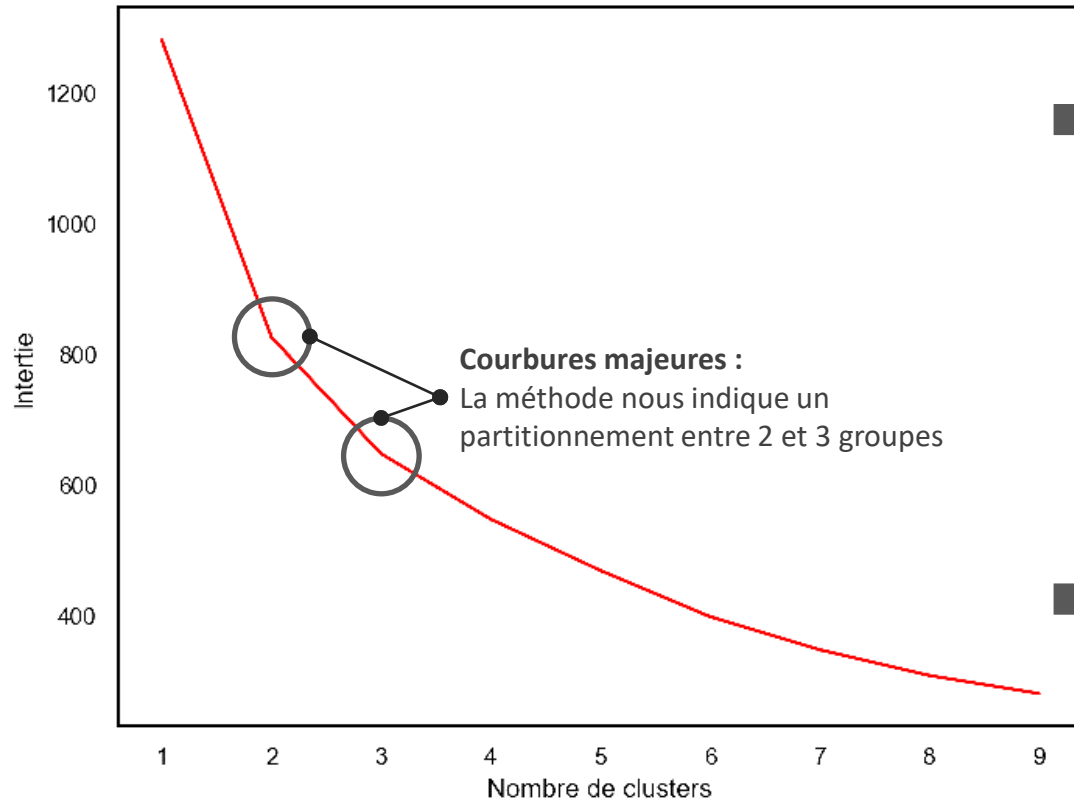
- **Groupe n°1** : Faible disponibilité alimentaire, croissance élevée, faible prévalence de la sous alimentation ;
- **Groupe n°2** : Très faible disponibilité alimentaire, croissance économique modérée, prévalence de la sous alimentation élevée ;
- **Groupe n°3** : Très forte disponibilité alimentaire, croissance économique modérée, faible prévalence de l'insécurité alimentaire ;
- **Groupe n°4** : Forte disponibilité alimentaire, croissance économique modérée, faible croissance démographique, faible prévalence de l'insécurité alimentaire ou de la sous-alimentation.
- **Groupe n°5** : faible disponibilité alimentaire, croissance économique modérée, faible croissance démographique, faible prévalence de l'insécurité alimentaire ou de la sous-alimentation.



2. METHODES DE CLUSTERING ET ANALYSE DES GROUPES

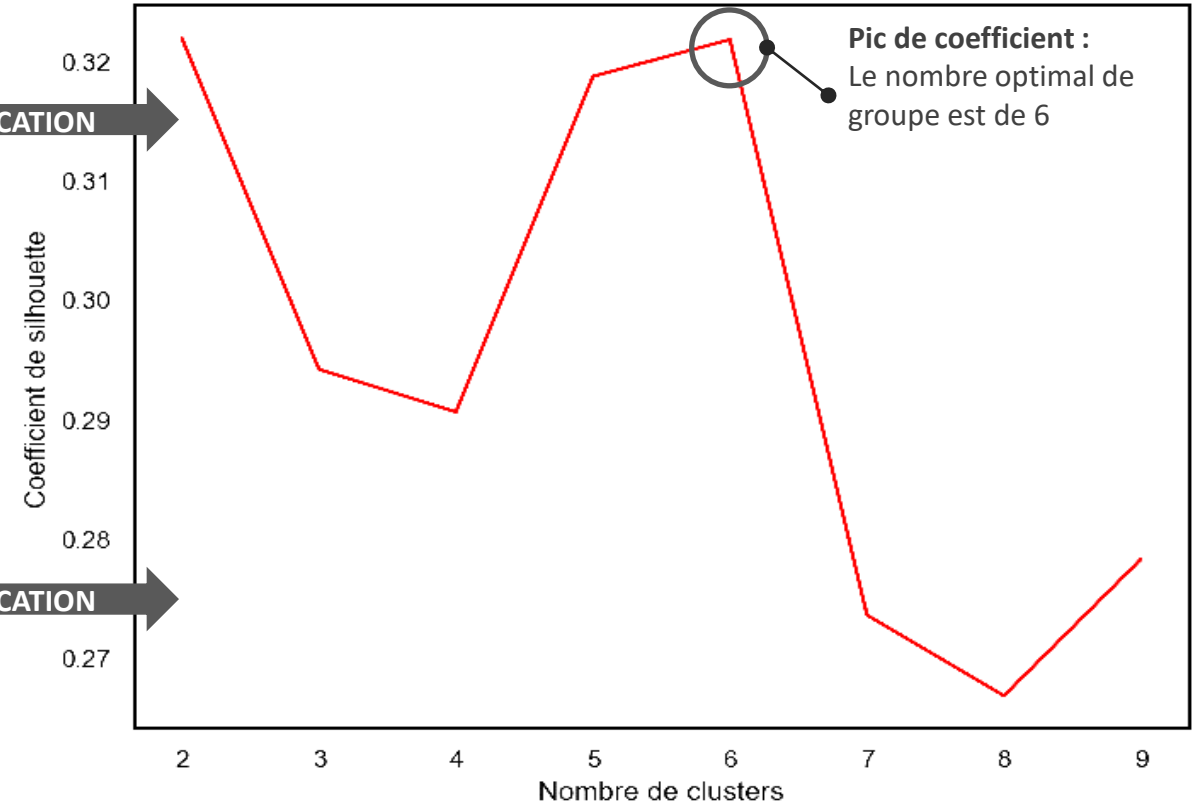
Méthode des centroïdes KMeans

VISUALISATION DE L' INERTIE PAR LA MÉTHODE DU COUDE



VERIFICATION

VISUALISATION DES COEFFICIENTS DE SILHOUETTES



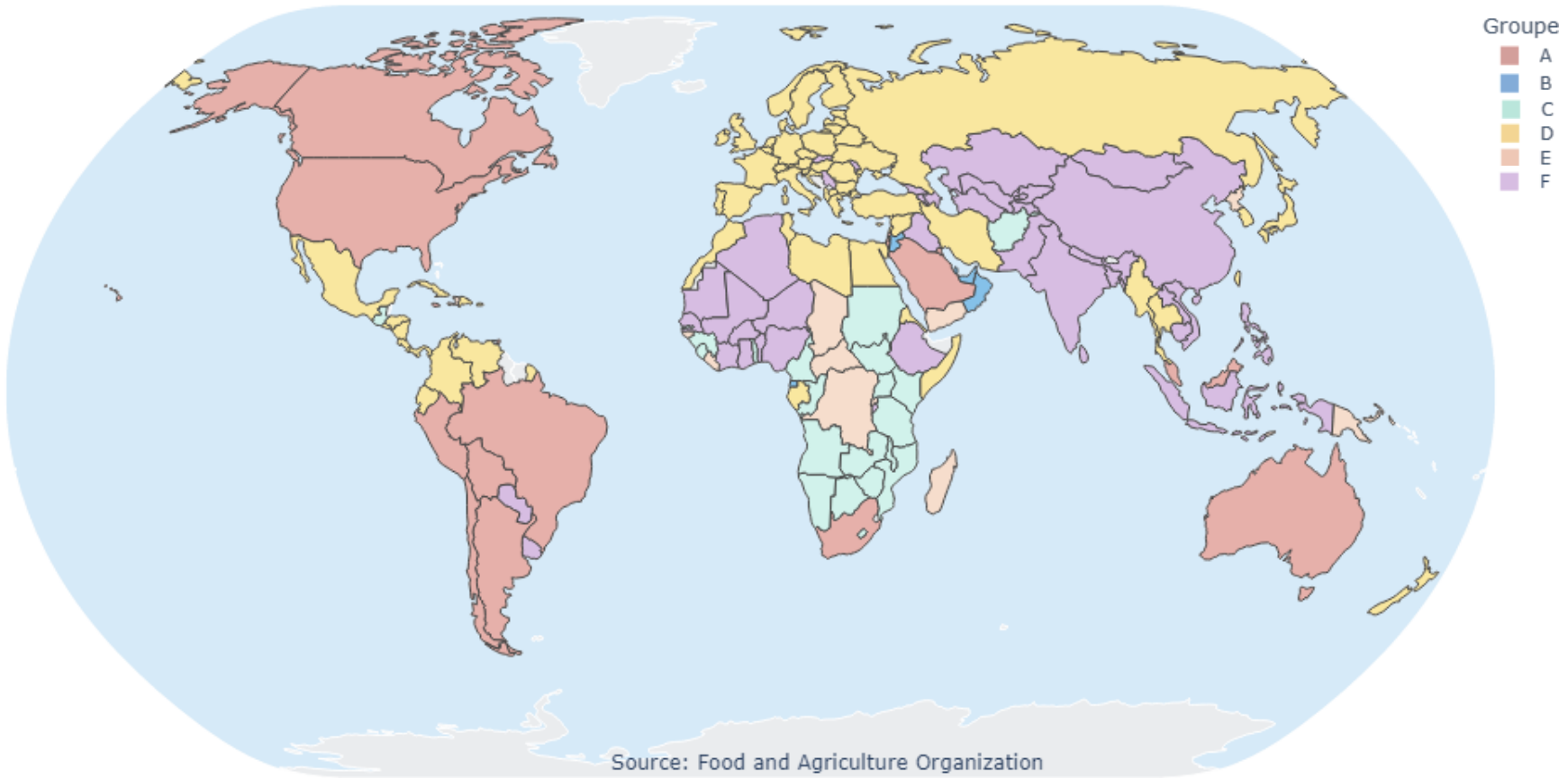
VERIFICATION



2. METHODES DE CLUSTERING ET ANALYSE DES GROUPES

Méthode des centroïdes KMeans

REPRÉSENTATION DES CLUSTERS



Source: Food and Agriculture Organization

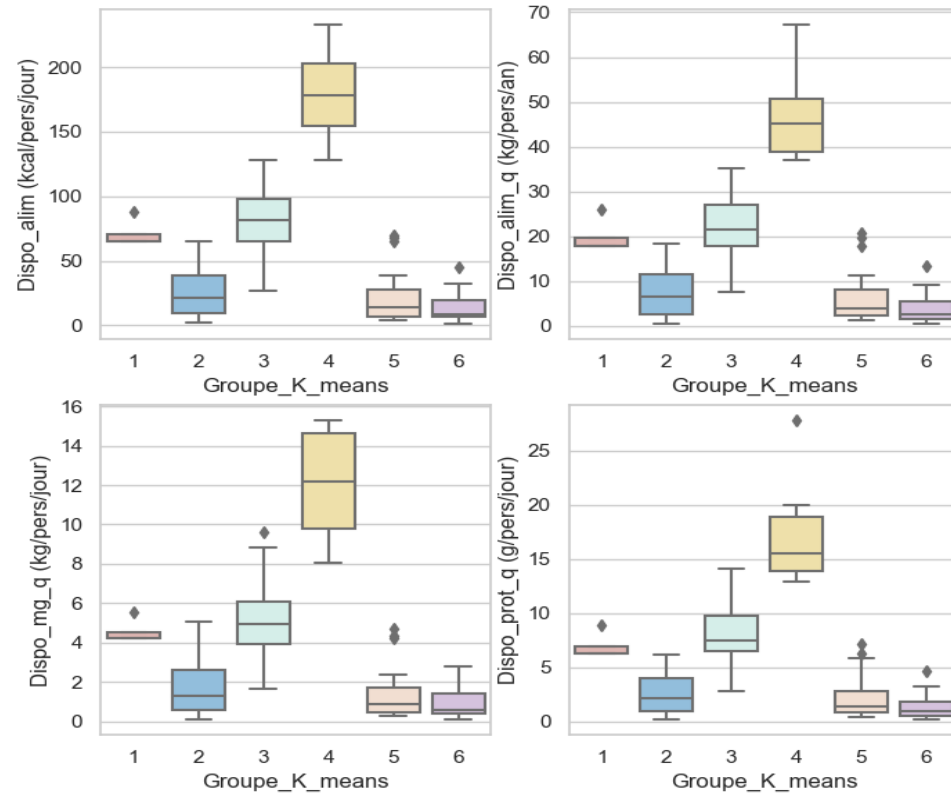


2. METHODES DE CLUSTERING ET ANALYSE DES GROUPES

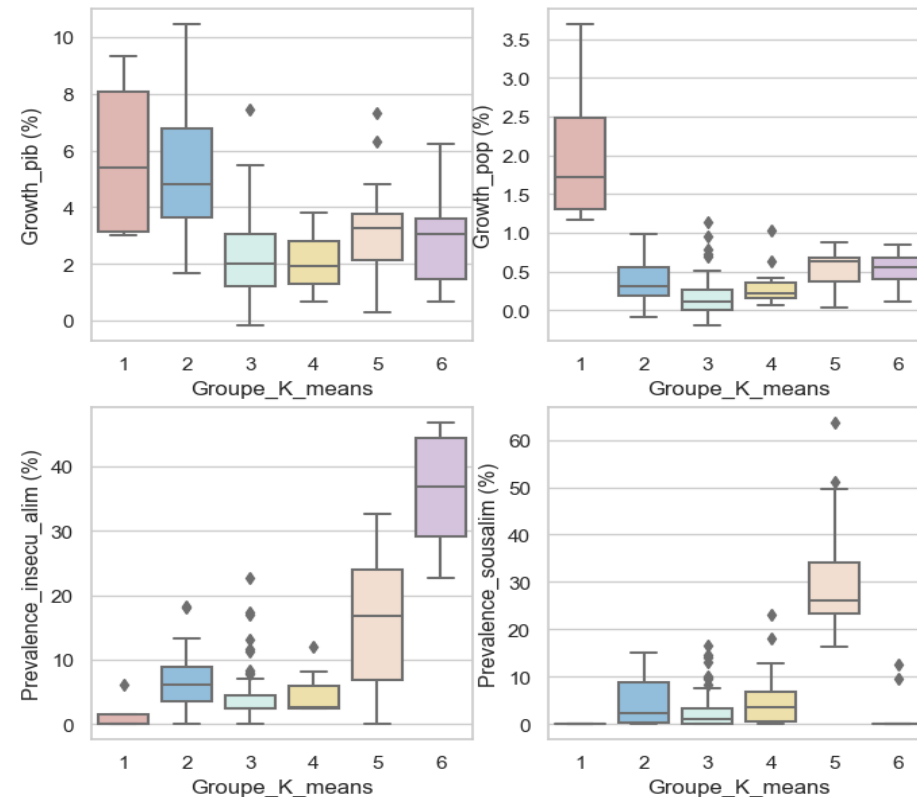
Méthode des centroïdes KMeans

DISTRIBUTION DES VALEURS PAR REGROUPEMENT KMEANS

Variables de disponibilité alimentaire



Variables de croissance et d'insécurité

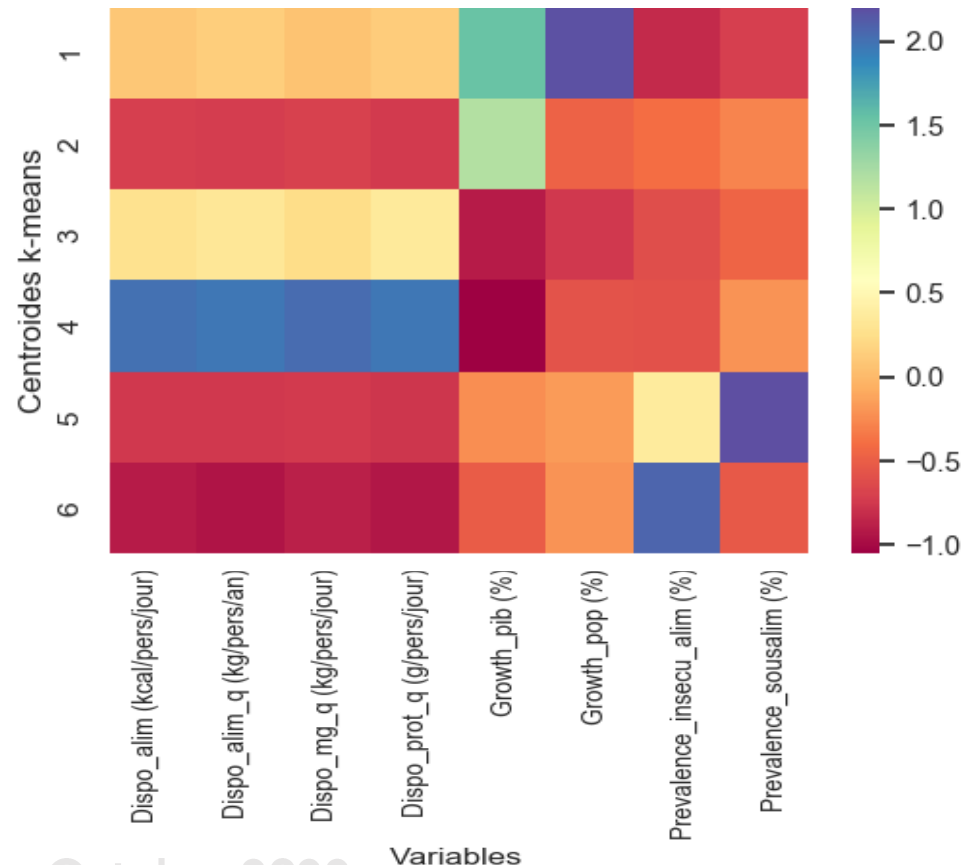




2. METHODES DE CLUSTERING ET ANALYSE DES GROUPES

Méthode des centroïdes K-means

HEATMAP DES CENTROÏDES K-MEANS



CARACTÉRISTIQUES DES 6 GROUPES:

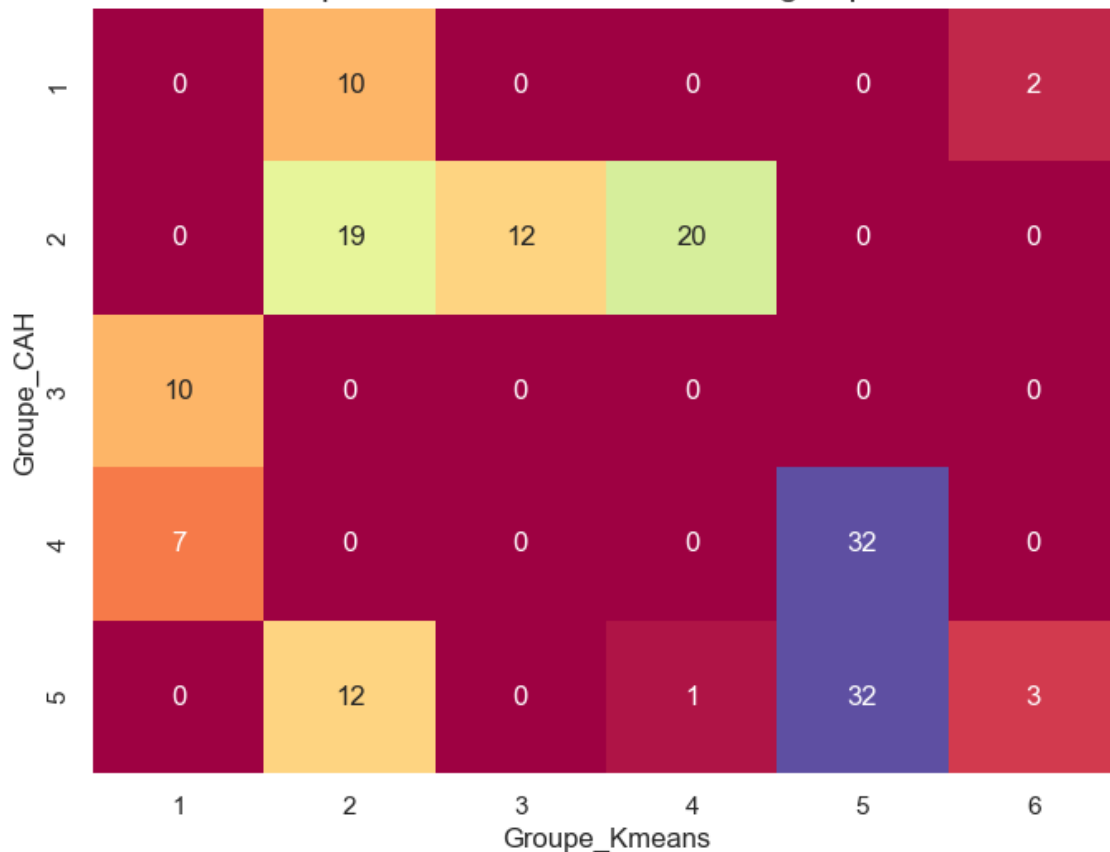
- **Groupe n°1** : Disponibilité alimentaire modérée, croissance économique élevée, croissance démographique élevée, faible prévalence de l'insécurité alimentaire ou de la sous-alimentation.
- **Groupe n°2** : Disponibilité alimentaire très forte ,croissance économique modérée et démographique faible, faible prévalence de l'insécurité et de la sous-alimentation faible.
- **Groupe n°3** : Disponibilité alimentaire modérée, croissance économique faible et démographique faible, insécurité alimentaire faible.
- **Groupe n°4** : Disponibilité alimentaire très forte ,croissance économique modérée et démographique faible, faible prévalence de l'insécurité et de la sous-alimentation faible.
- **Groupe n°5** : Disponibilité alimentaire faible, croissance économique modérée et croissance démographique moyenne, forte prévalence de l'insécurité alimentaire et de la sous-alimentation.
- **Groupe n°6** : Très faible disponibilité alimentaire, croissance économique modérée, croissance démographique faible , forte prévalence de l'insécurité alimentaire.



2. METHODES DE CLUSTERING ET ANALYSE DES GROUPEs

Similitudes et différences entre es méthodes

Compraraision des méthodes de regroupement



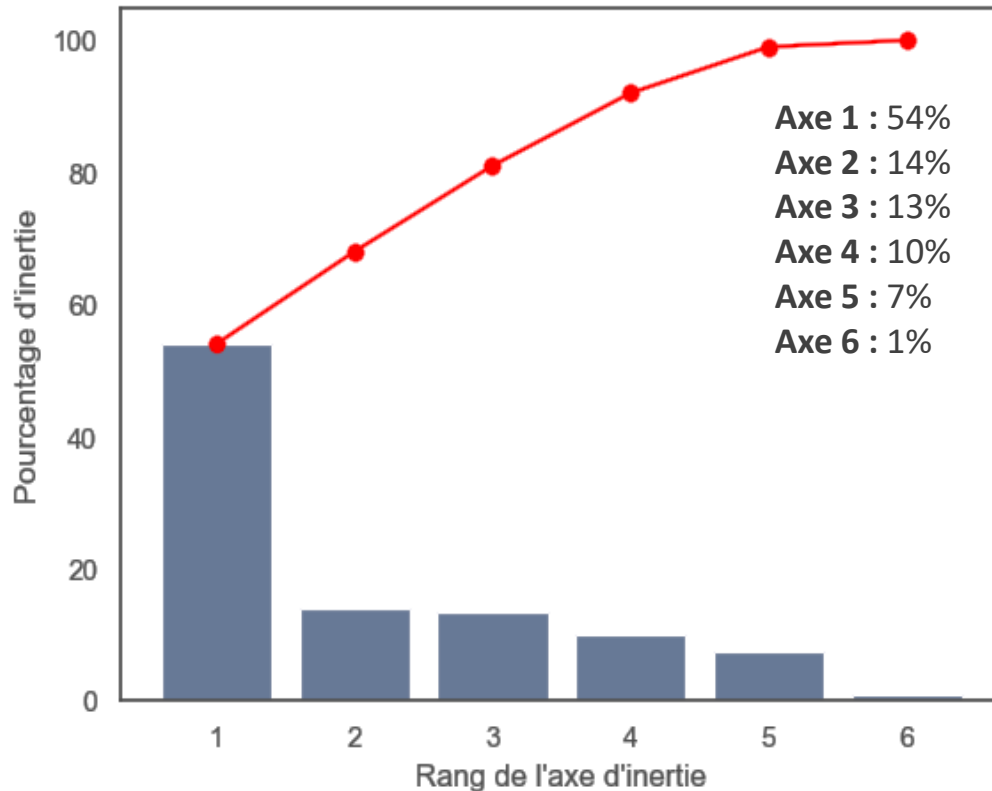
- Le partitionnement des données est relativement différent entre les deux méthodes.
- Pour la majorité des clusters formés, ils sont constitués d'une base commune de pays.
- On note des distinctions plus flagrantes pour le groupe 2 et de la CAH et du Kmeans ou encore le groupe 5 du Kmeans.
- Après avoir vu les deux méthodes, nous faisons le choix de sélectionner la méthode du K-means pour développer l'Analyse en Composante Principale (ACP).



3. ANALYSE DES COMPOSANTES PRINCIPALES (ACP)

Création des composantes

ÉBOULIE DES VALEURS PROPRES



MATRICE DES CORÉLATION ENTRE LES VARIABLES ET LES COMPOSANTES

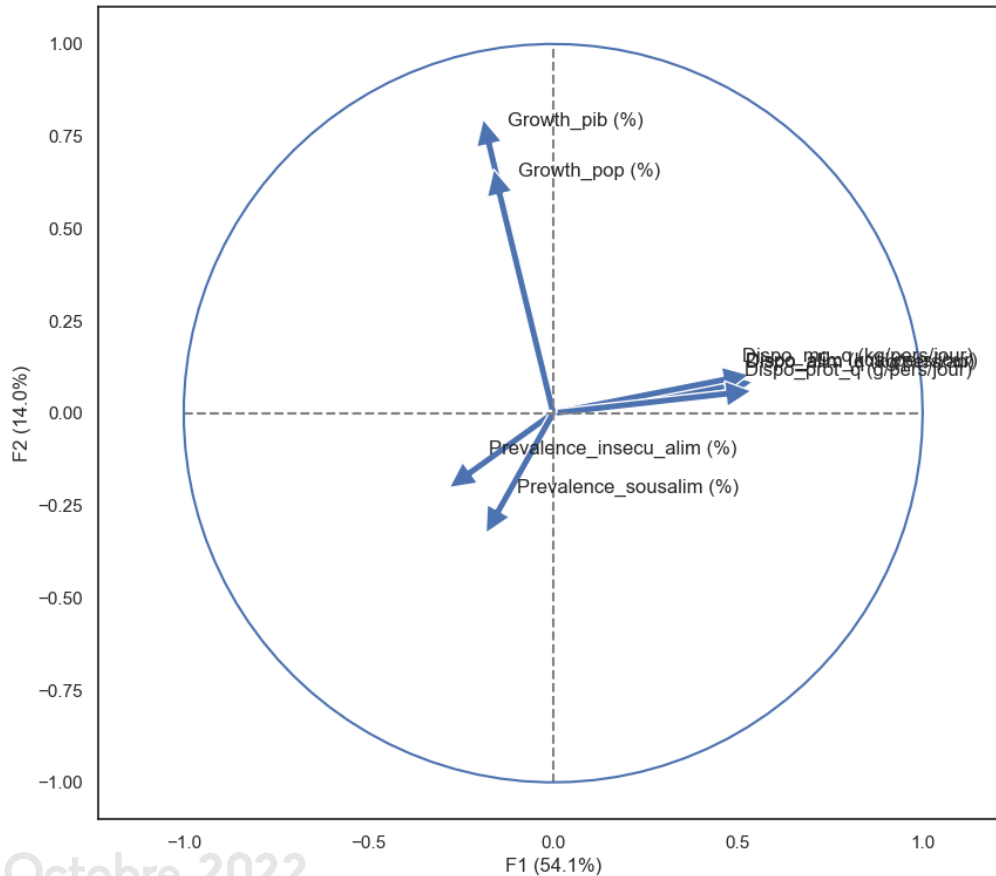
	F1	F2	F3	F4	F5	F6
Dispo_alim (kcal/pers/jour)	0.47	0.08	0.13	-0.04	-0.10	-0.13
Dispo_alim_q (kg/pers/an)	0.47	0.07	0.13	-0.04	-0.07	0.16
Dispo_mg_q (kg/pers/jour)	0.46	0.09	0.14	-0.04	-0.14	-0.68
Dispo_prot_q (g/pers/jour)	0.47	0.05	0.11	-0.03	-0.03	0.69
Growth_pib (%)	-0.17	0.73	-0.17	0.15	-0.62	0.07
Growth_pop (%)	-0.14	0.59	0.55	-0.09	0.56	-0.02
Prevalence_insecu_alim (%)	-0.22	-0.16	0.41	-0.76	-0.41	0.05
Prevalence_sousalim (%)	-0.15	-0.26	0.66	0.62	-0.30	0.02



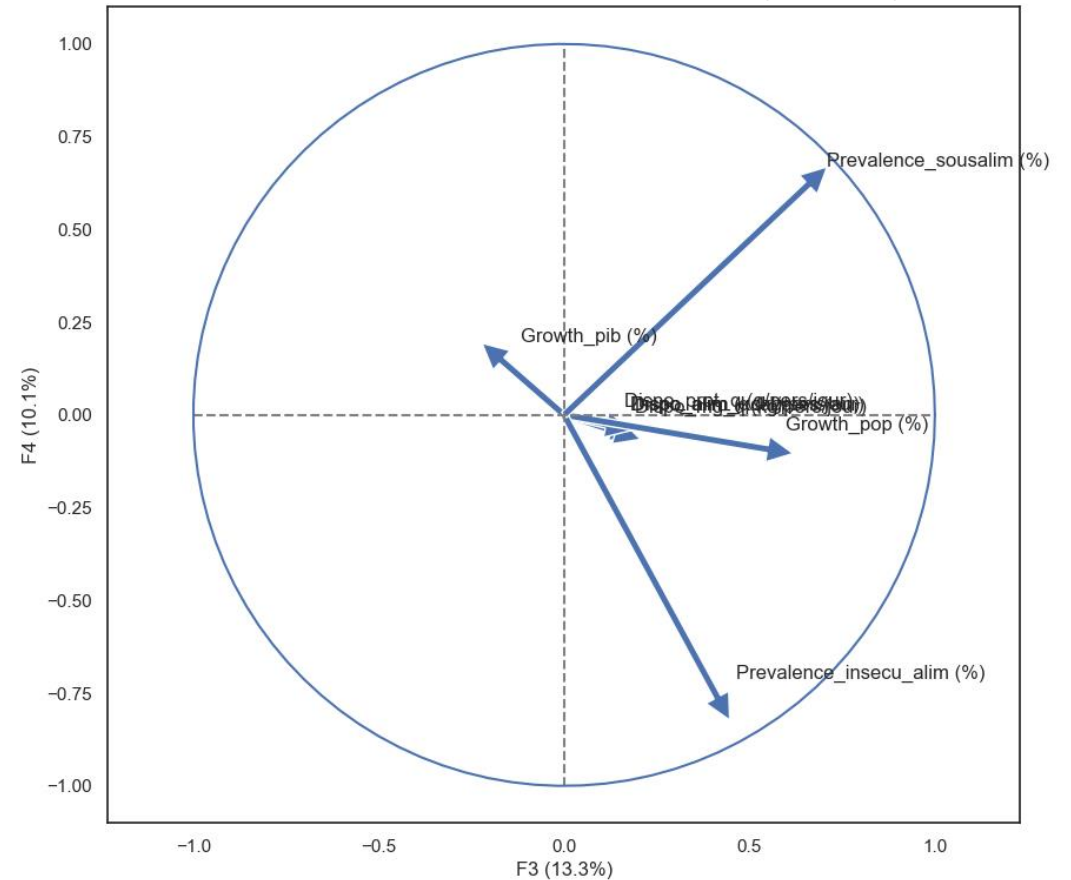
3. ANALYSE DES COMPOSANTES PRINCIPALES (ACP)

Corrélation entre les variables et les composantes

CERCLE DES CORRÉLATIONS (F1 ET F2)



CERCLE DES CORRÉLATIONS (F3 ET F4)

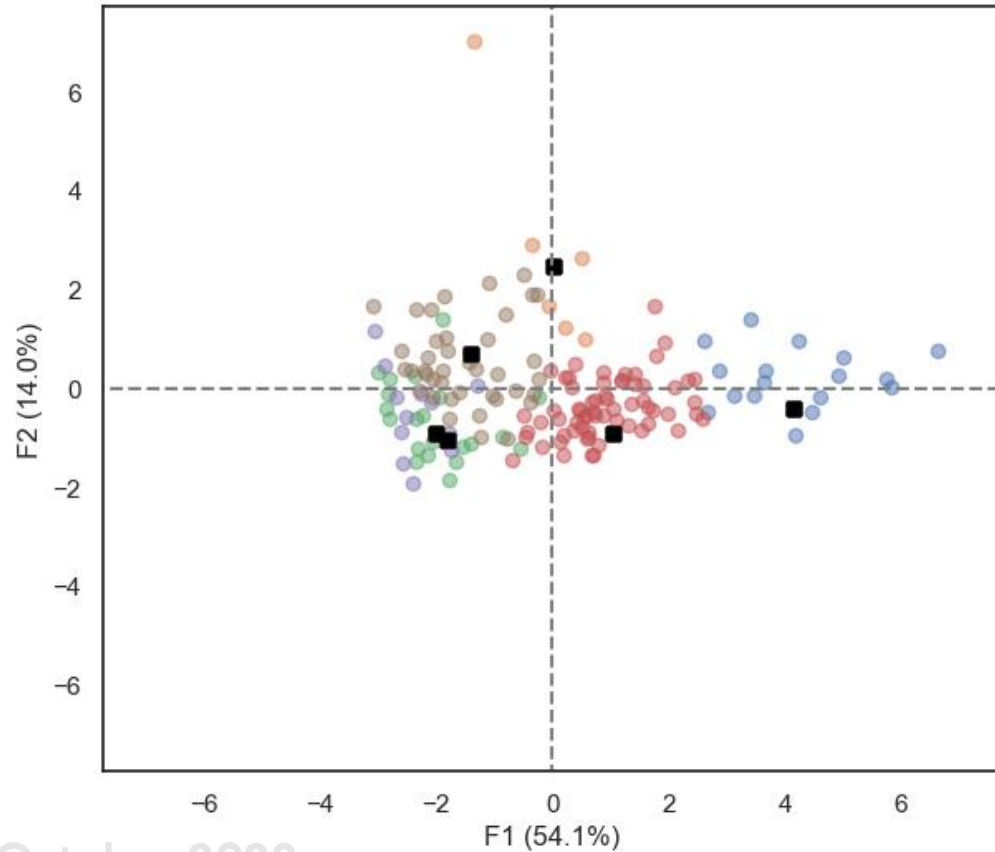




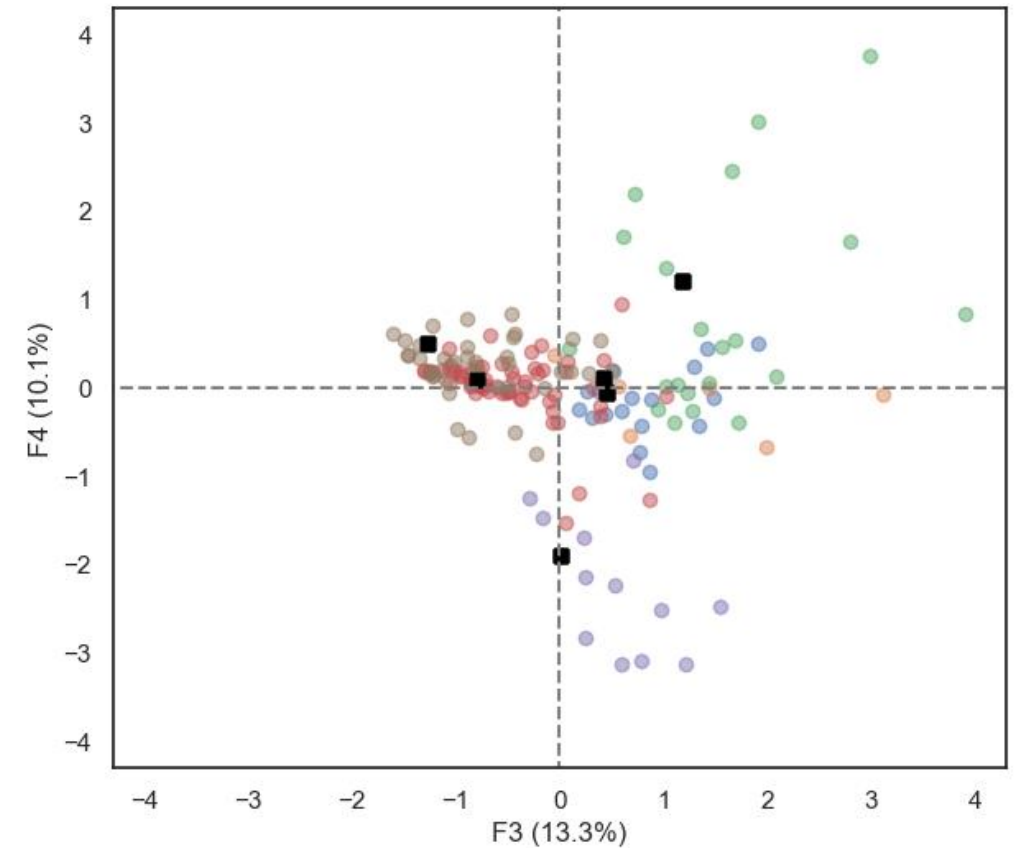
3. ANALYSE DES COMPOSANTES PRINCIPALES (ACP)

Représentation des individus sur les plans factoriels

PROJECTION DES INDIVIDUS PAR GROUPE (F1 ET F2)



PROJECTION DES INDIVIDUS PAR GROUPE (F3 ET F4)





3. ANALYSE DES COMPOSANTES PRINCIPALES (ACP)

Interprétation des composantes

Observations :

- Les variables les plus corrélées à **F1** sont les données de **disponibilité alimentaire (+)** et la **croissance démographique (-)**.
- Les variables les plus corrélées à **F2** sont la prévalence de la **sous-alimentation (+)**, l'**insécurité alimentaire (-)** et la **croissance économique (-)**.
- Les variables les plus corrélées à **F3** sont les données de **disponibilité alimentaire (-)** et la **croissance démographique (-)**.
- Les variables les plus corrélées à **F4** sont les données de **sous-alimentation (-)** et la **croissance économique (+)**.

Hypothèses d'analyse :

- **F1** correspond à la notion de **ressources alimentaires accessibles sur le territoire**.
- **F2** correspond à la notion de **malnutrition**.
- **F3** correspond à la notion de **développement démographique**.
- **F4** correspond à la notion de **développement économique**.

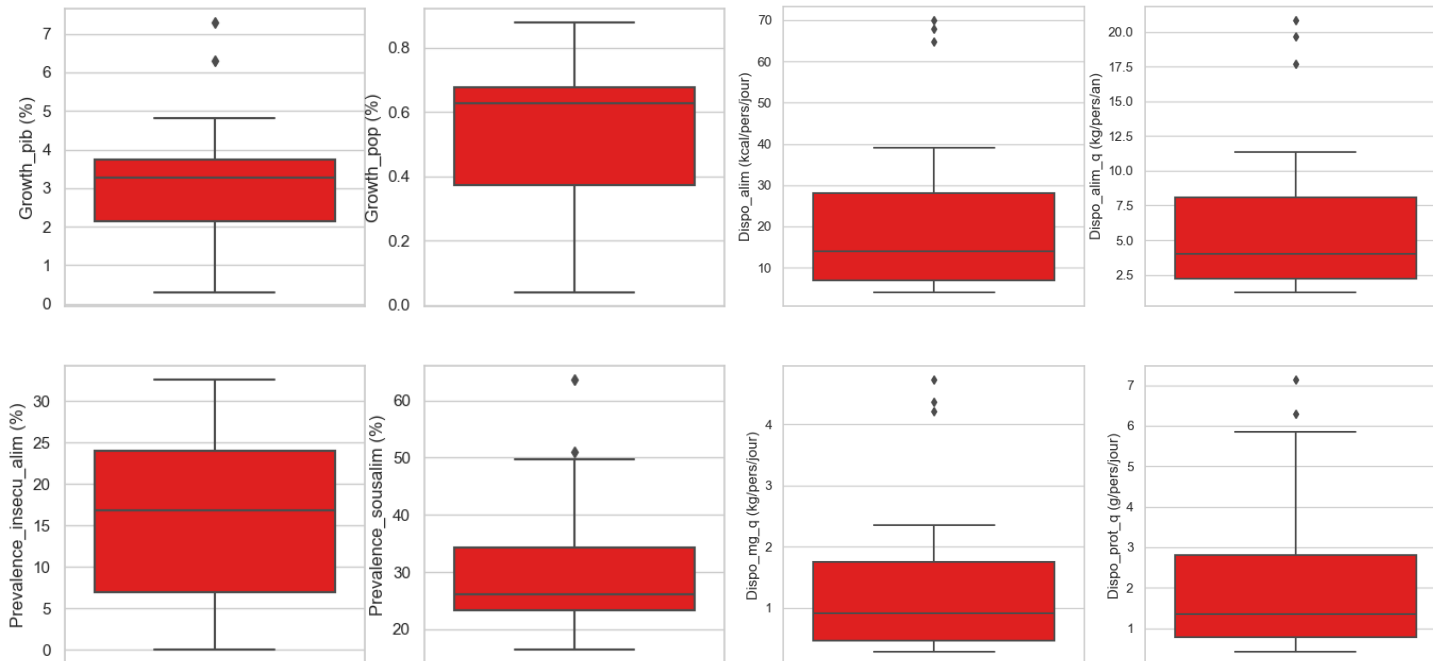


4. BILAN ET RECOMMANDATIONS

Synthèse des résultats

Notre choix de sélection de cluster s'oriente à priori davantage vers **le groupe 4**. Les statistiques de disponibilité alimentaires sont avantageuses et complétées par une croissance économique moyenne et peu d'insécurité alimentaire. La composante qui explique le mieux la répartition de notre échantillon est la composante liée aux **ressources alimentaires disponibles sur le territoire**. Notre choix de cluster s'est prioritairement porté sur cet élément.

DISTRIBUTION DES VARIABLES DU GROUPE SÉLECTIONNÉ



VISUALISATION DU GROUPE SÉLECTIONNÉ



Source: Food and Agriculture Organization



4. BILAN ET RECOMMANDATIONS

Recommandations

- L'étude permet une première analyse et segmentation des pays pour la stratégie d'internationalisation.
- Elle doit être complétée par des informations contextuelles pour chaque pays, afin d'affiner l'analyse. Les données utilisées nous donnent une vision sur l'année 2018, mais aucune information au-delà de cette date.
- Il aurait été intéressant d'aller plus loin en rajoutant divers éléments :
 - **Des éléments de mesure du pouvoir d'achat à l'international (ex: indice Big Mac).**
On remarque que dans nos clusters, certains regroupements semblent peu cohérent si l'on élargissait les données prises en comptes.
 - **Des éléments sur la sécurité intérieure des pays pour rendre compte des difficultés d'implantation:**
Ces éléments pourraient être d'une grande aide en complément des éléments de sous-alimentation et d'insécurité alimentaire.