

Étude de marché

Identification des pays propices à
une insertion dans le marché du
poulet

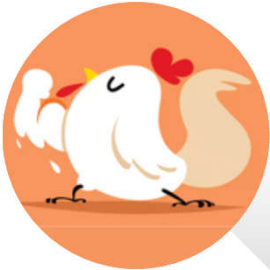
Ce travail est réalisé par :
Amina BOUBLENZA

Mentor :
Mohammed Ali Nacer



Contexte

- Entreprise d'agroalimentaire, spécialisée dans le poulet, souhaite se développer à l'international



Stratégie

- Exporter les produits plutôt que de produire sur place



Objectif

- Cibler un groupe de pays
- Approfondir l'étude de marché

Plan de l'étude de marché



Partie I	Analyse des données avec:	4 variables
		5 clusters
Partie II	Analyse des données avec :	7 variables
		7 clusters
Partie III	Analyse ACP	Projection des données
Partie IV	Statistique inférentielle :	Test d'adéquation
		Comparaison de population

Partie I

Analyse des données avec 4 variables

Préparation des données



- Variables de la première partie :
 - Différence de population entre 2017 et 2019 ;
 - Proportion de protéines d'origine animale par rapport à la quantité totale de protéines ;
 - Disponibilité alimentaire en protéines par habitant ;
 - Disponibilité alimentaire en calories par habitant.
- Sources



Food and Agriculture
Organization of the
United Nations

Classification ascendante hiérarchique (CAH)



- La classification hiérarchique (*hierarchical clustering en anglais*) :
 - Une des approches les plus importantes pour l'exploration des données multivariées.
 - L'objectif est d'identifier des groupes (i.e., clusters) d'objets similaires dans un jeu de données.
- Il y a deux approches pour créer une partition hiérarchique :
 - L'approche ascendante, aussi appelée clustering agglomératif
 - L'approche descendante, aussi appelée clustering divisif
- Un dendrogramme est la représentation graphique d'une classification ascendante hiérarchique
 - Il se présente souvent comme un arbre binaire dont les feuilles sont les individus alignés sur l'axe des abscisses.

Stratégies d'agrégation



➤ La méthode de Ward

- Regrouper les classes de façon que l'augmentation de l'inertie interclasse soit maximale, ou bien l'augmentation de l'inertie intraclasse soit minimum (théorème de Huygens)

➤ Le lien simple

- La distance entre 2 clusters égale à la distance entre leurs 2 points les plus proches.

➤ Le lien complet

- La distance entre 2 clusters est égale à la distance entre leurs 2 points les plus éloignés

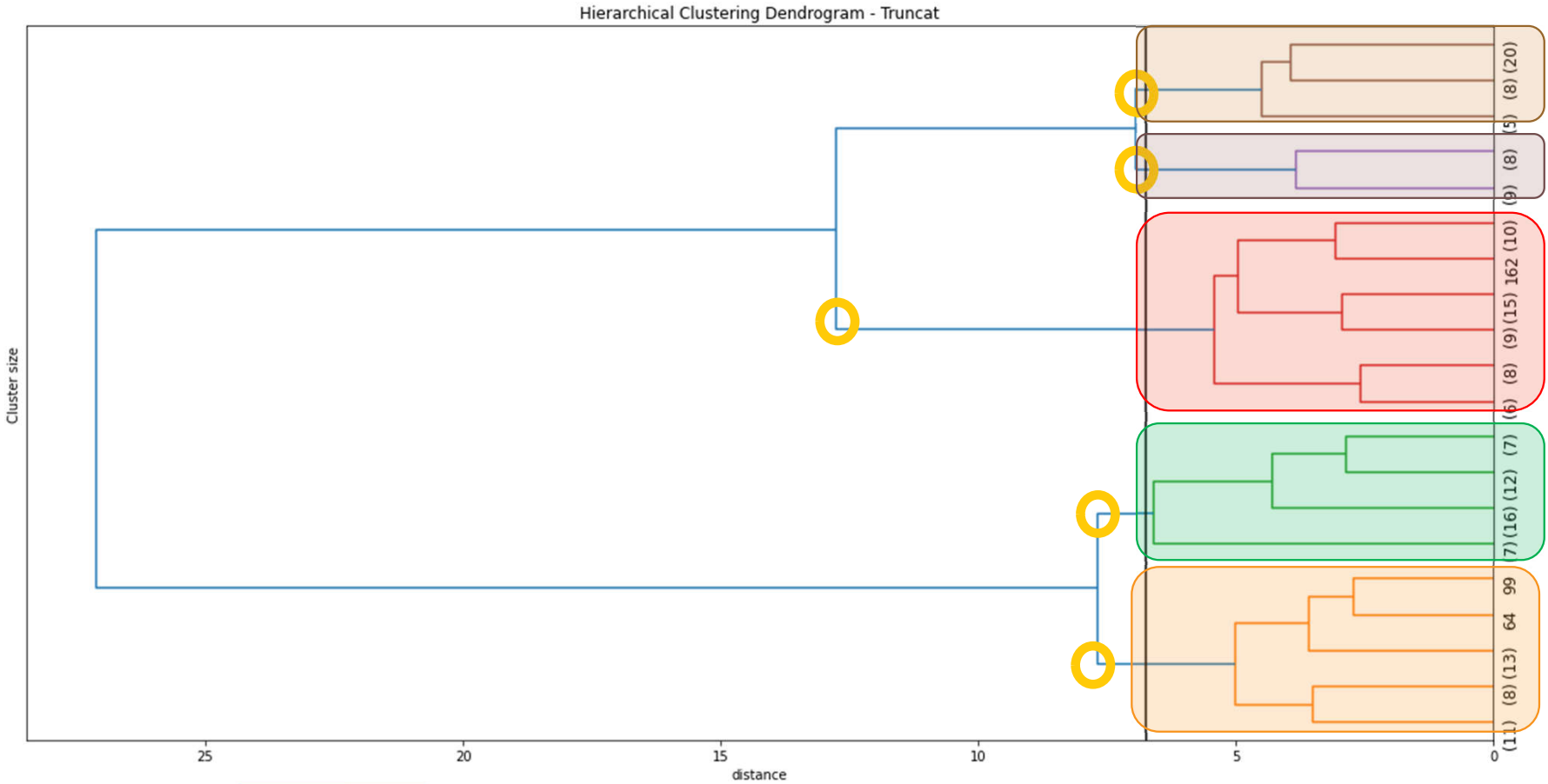
➤ Le lien moyen

- La distance entre 2 clusters est la moyenne de toutes les distances entre les points d'un cluster et les points de l'autre cluster.

➤ Le lien centroïdal :

- La distance entre 2 clusters est égale la distance entre les centroïdes de ceux-ci.

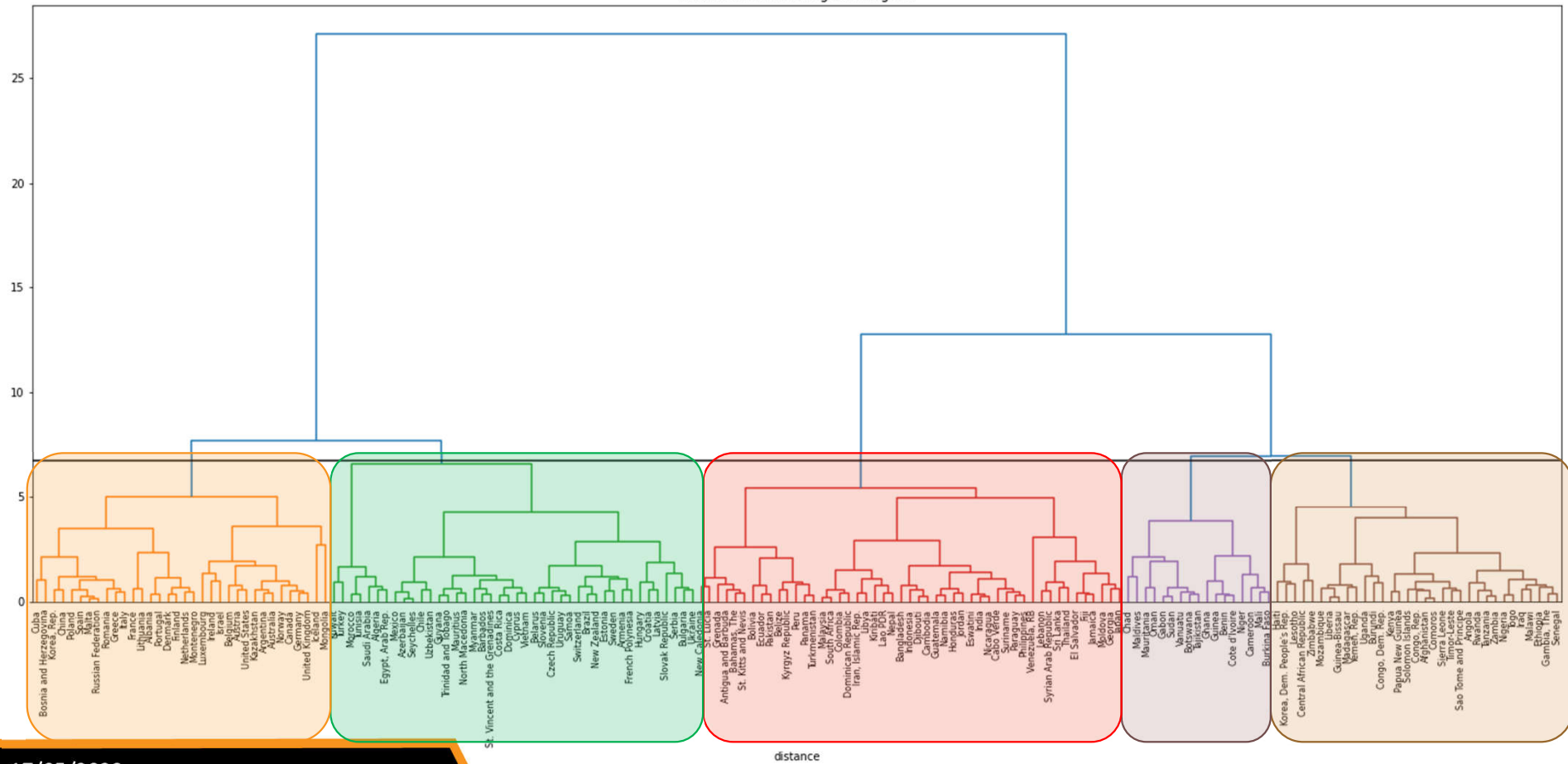
Dendrogramme tronqué



Dendrogramme



Hierarchical Clustering Dendrogram



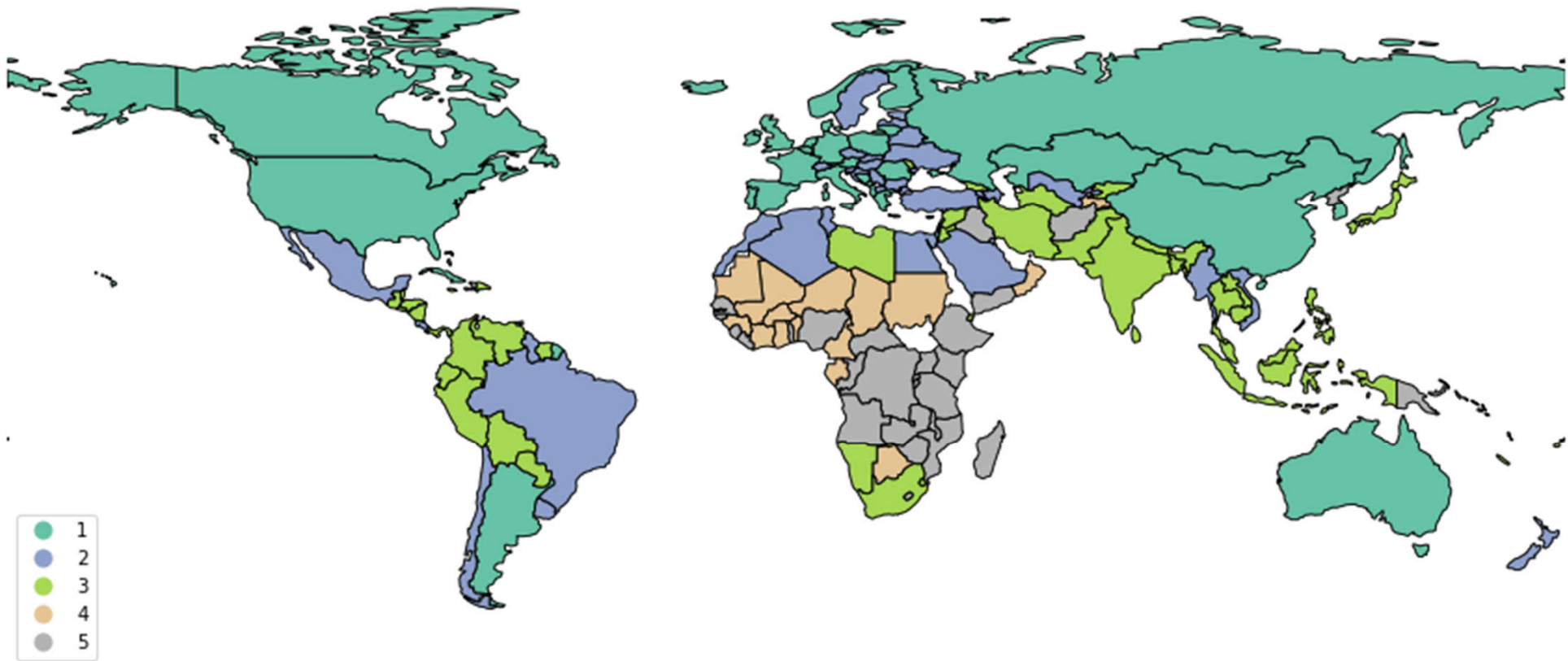
17/05/2022

Cluster selon la région



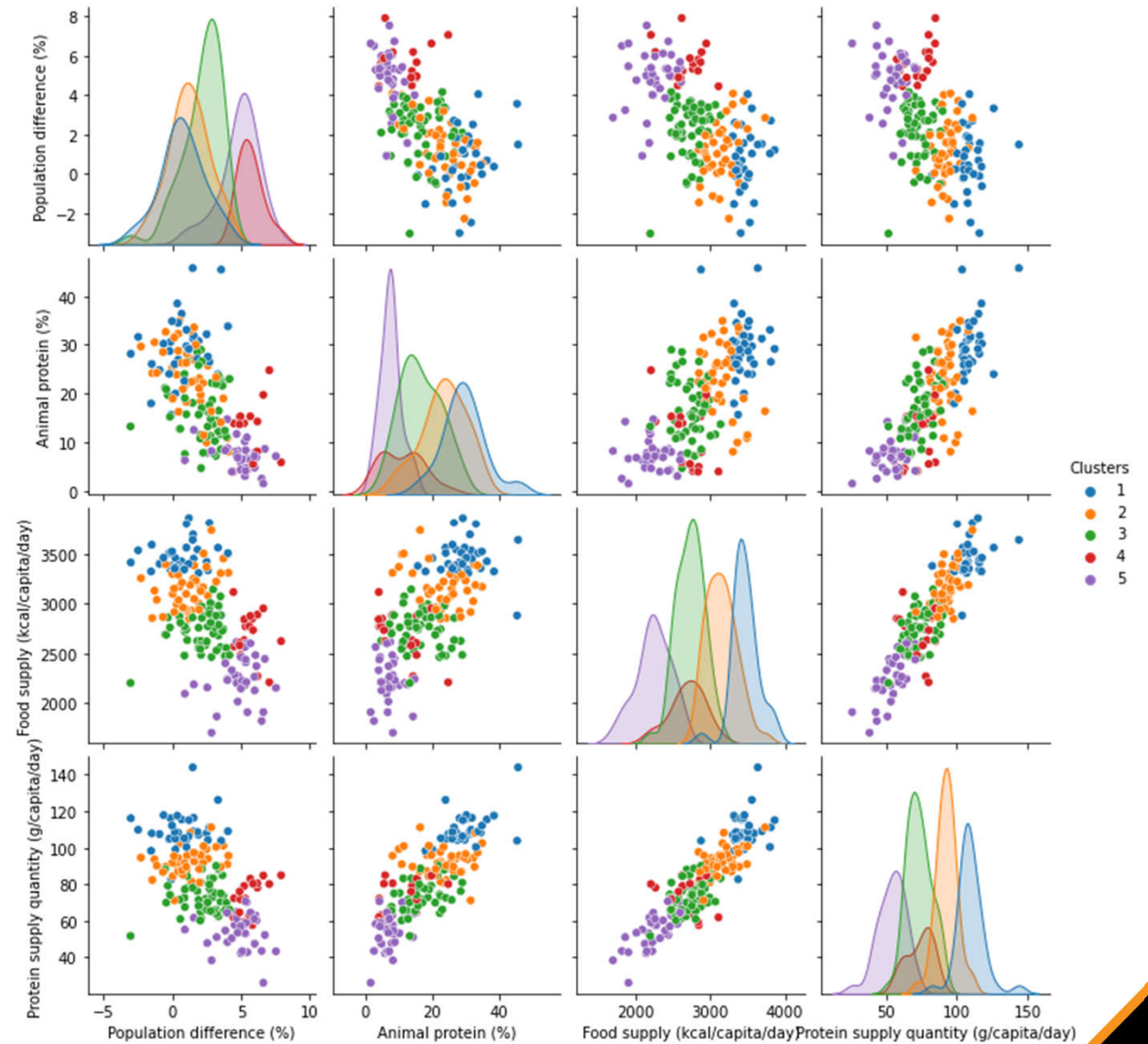
Région / Cluster	1	2	3	4	5
East Asia & Pacific	4	5	10	1	4
Europe & Central Asia	24	19	4	1	0
Latin America & Caribbean	2	10	20	0	1
Middle East & North Africa	2	6	6	1	2
North America	2	0	0	0	0
South Asia	0	0	5	1	1
Sub-Saharan Africa	0	2	4	13	25

Carte géographique selon le cluster



Relation bivarié

1. Différence de population (%)
2. protéines animale (%)
3. Disponibilité alimentaire en calories (kcal/personne/j)
4. Disponibilité alimentaire en protéines (g/personne/j)



Partie II

7 variables

Préparation des données

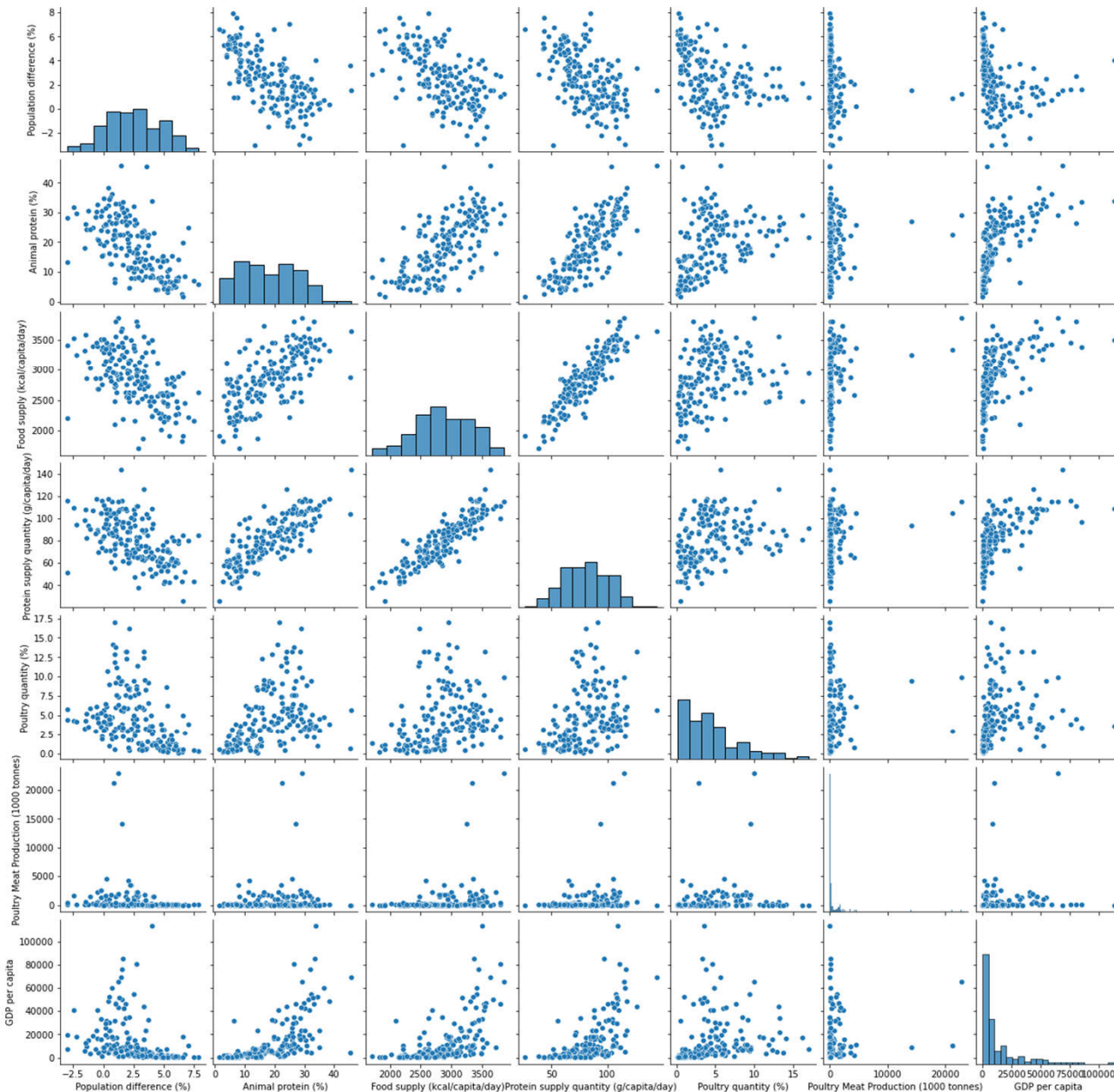


➤ Les variables utilisées :

- Les 4 variables de la première partie
- Production de volaille de chaque pays
- Rapport de consommation du volaille par rapport à son alimentation
- PIB par habitant en 2019

➤ Sources





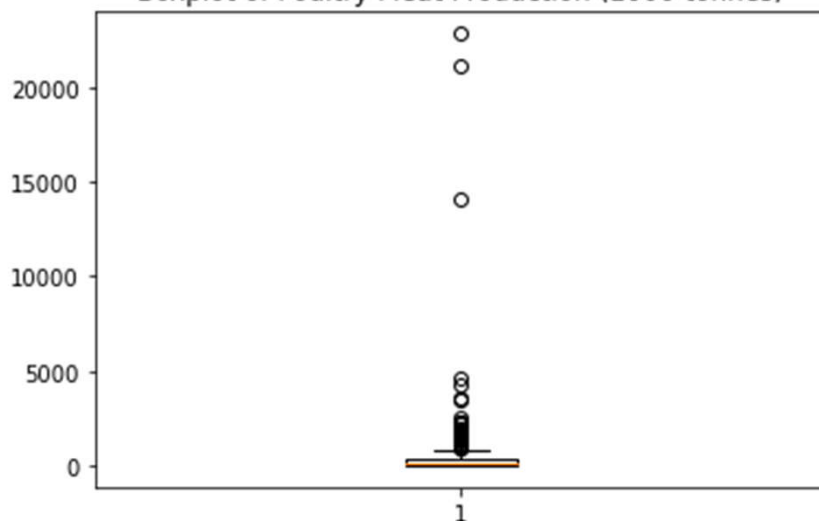
Relation bivarié

1. Différence de population (%)
2. protéines animale (%)
3. Disponibilité alimentaire en calories (kcal/personne/j)
4. Disponibilité alimentaire en protéines (g/personne/j)
5. Quantité de volaille (%)
6. Production de volaille (1000 tonnes)
7. PIB par personne

Boîte à moustache de production de volailles



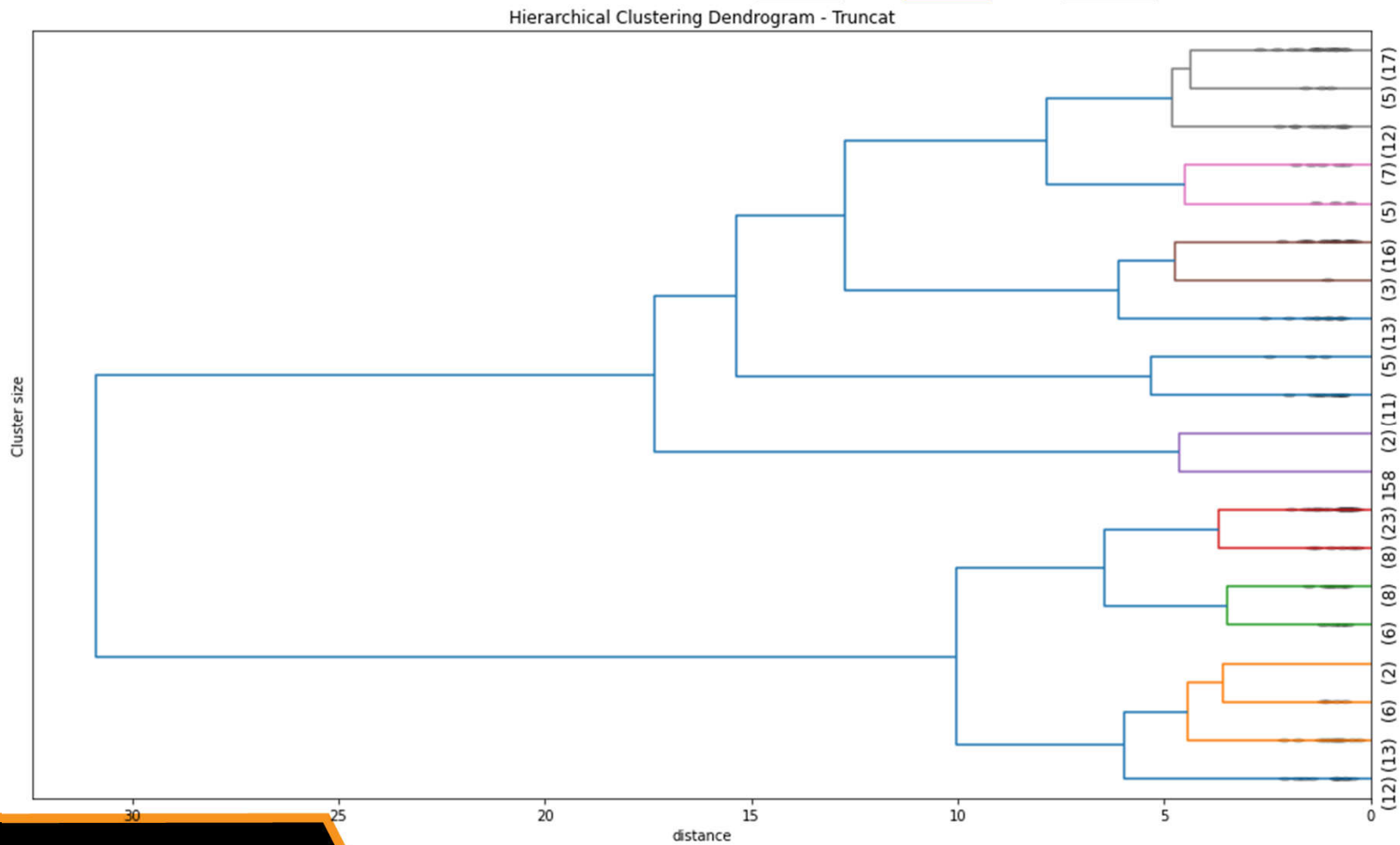
Boxplot of Poultry Meat Production (1000 tonnes)



Moyenne \approx 750
Écart type \approx 2644
Médiane \approx 75
min = 0
max = 22858

Pays	Production de volailles (1000 Tonnes)
United States	22858
China	21195
Brazil	14137
Russian Federation	4606
India	4232
Indonesia	3539
Mexico	3515

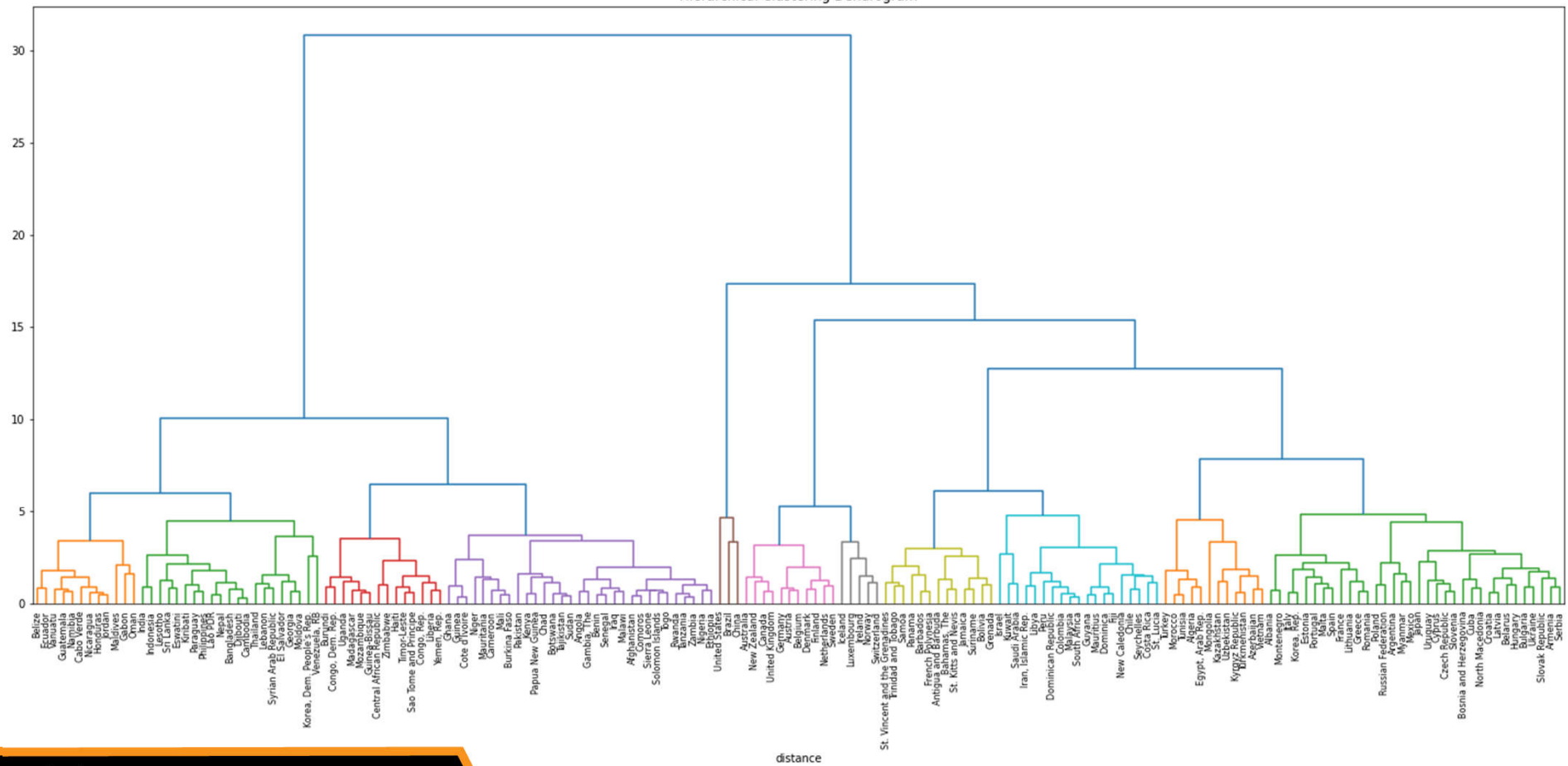
Dendrogramme tronqué



Dendrogramme

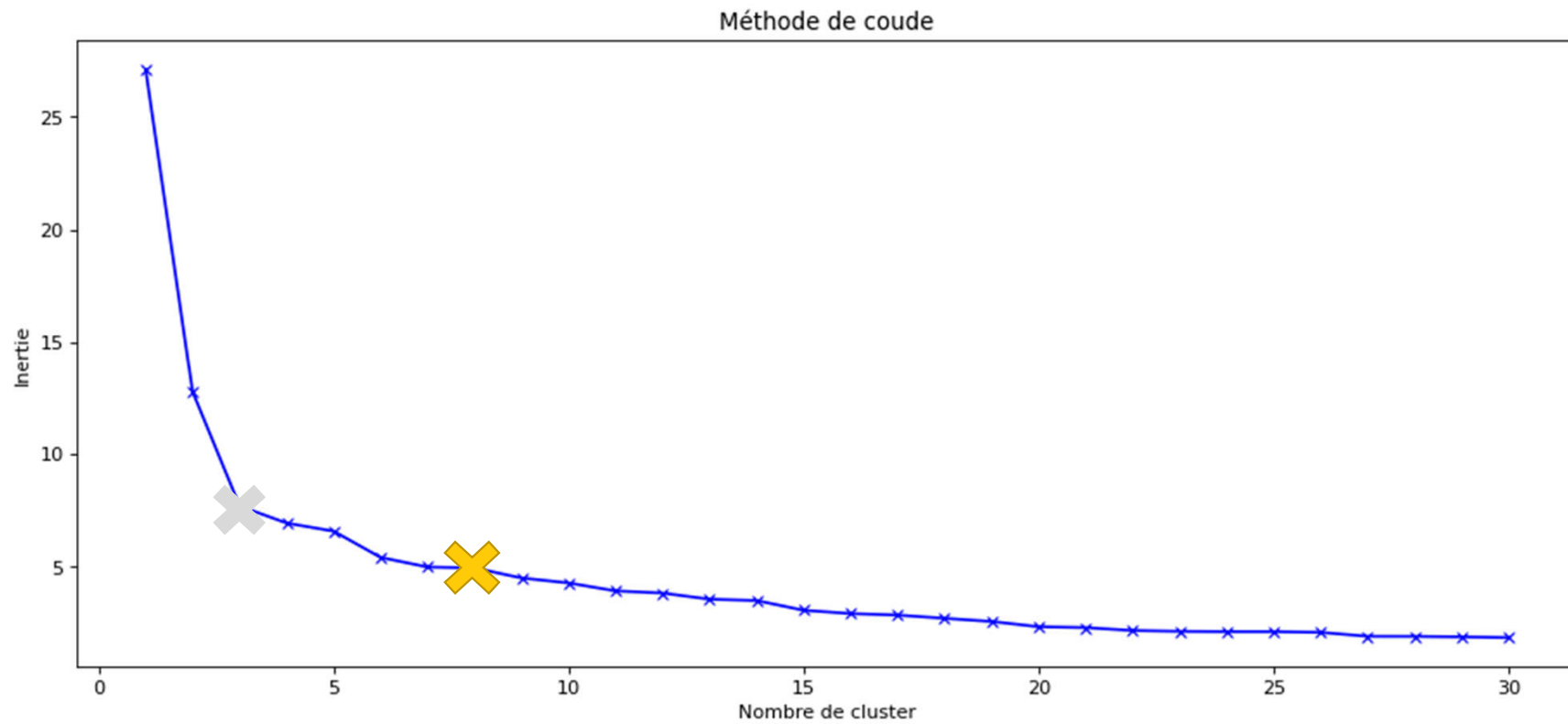


Hierarchical Clustering Dendrogram

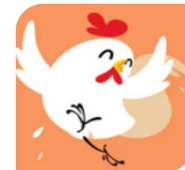


17/05/2022

Méthode de coude

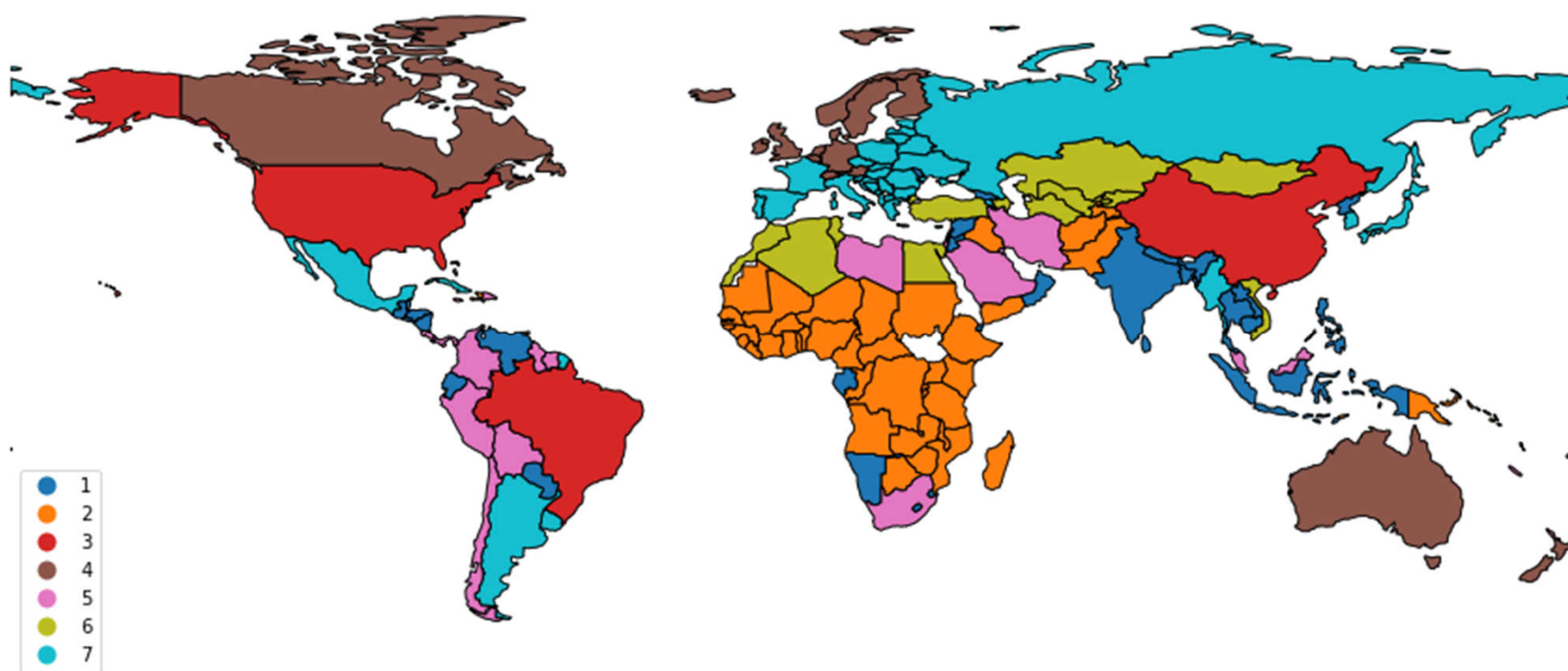


Cluster selon la région

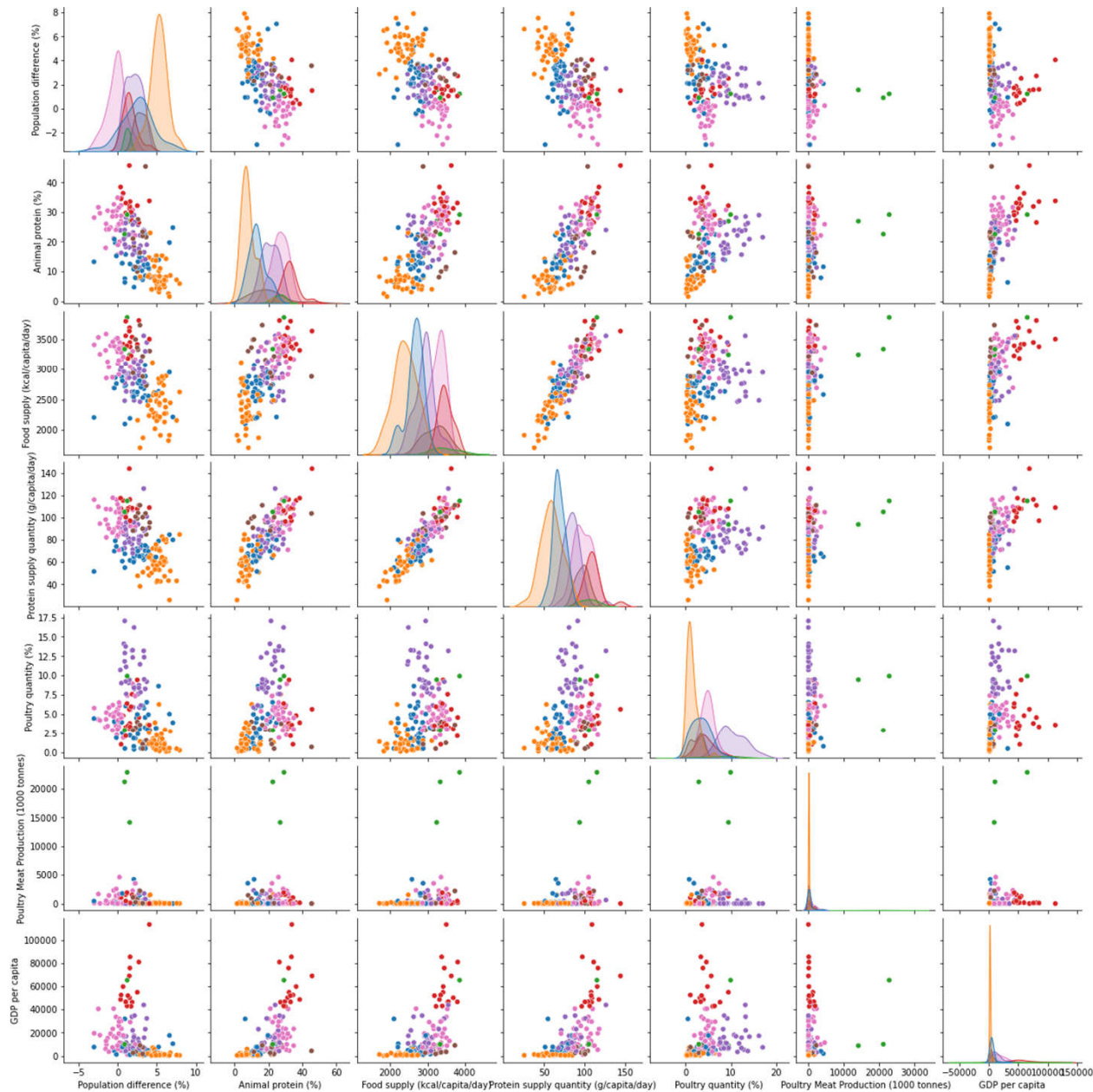


Région / Cluster	1	2	3	4	5	6	7
East Asia & Pacific	8	3	1	2	5	2	3
Europe & Central Asia	2	1	0	13	0	6	26
Latin America & Caribbean	8	1	1	0	19	0	4
Middle East & North Africa	5	2	0	0	5	4	1
North America	0	0	1	1	0	0	0
South Asia	5	2	0	0	0	0	0
Sub-Saharan Africa	5	36	0	0	3	0	0

Carte graphique selon cluster

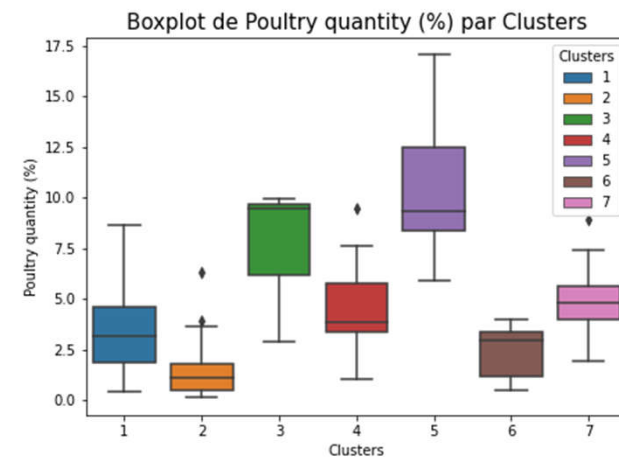
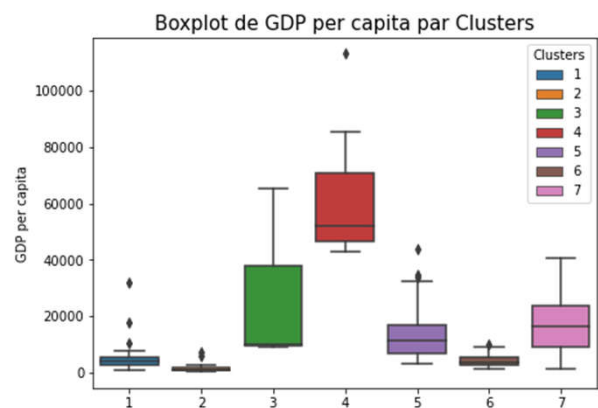
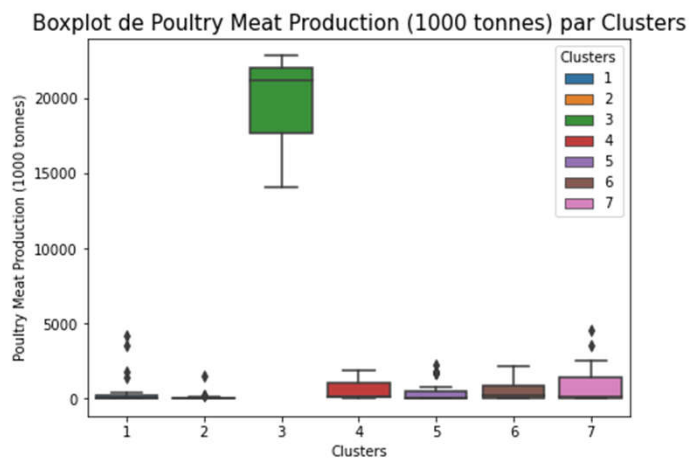
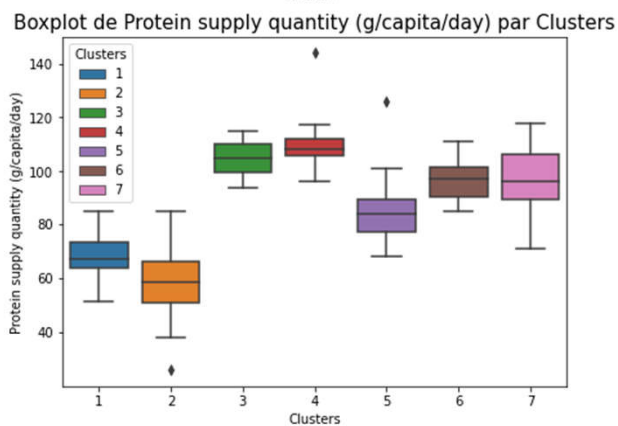
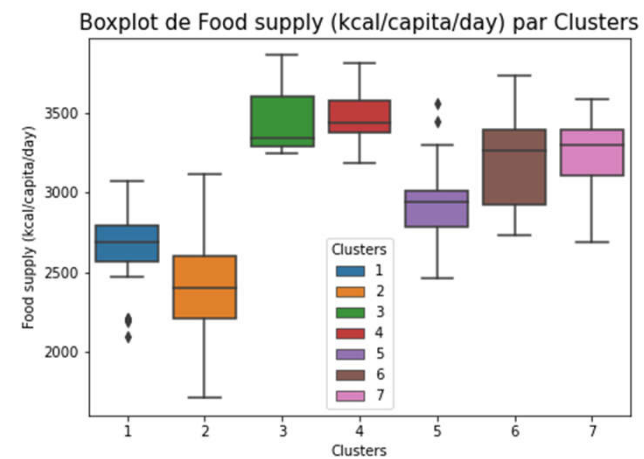
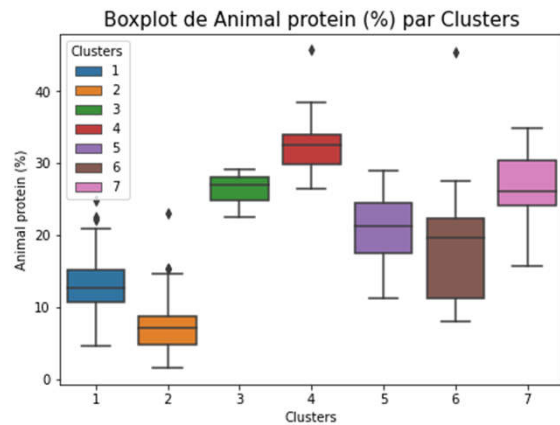
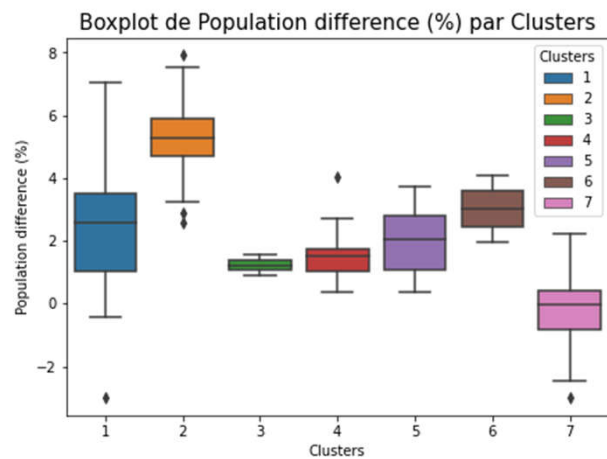


17/05/2022



Relation bivarié selon cluster

1. Différence de population (%)
2. protéines animale (%)
3. Disponibilité alimentaire en calories (kcal/personne/j)
4. Disponibilité alimentaire en protéines (g/personne/j)
5. Quantité de volaille (%)
6. Production de volaille (1000 tonnes)
7. PIB par personne



Centroïdes des clusters pour chaque dimension



Clusters	Population difference (%)	Animal protein (%)	Food supply (kcal/capita/day)	Protein supply quantity (g/capita/day)	Poultry quantity (%)	Poultry Meat Production (1000 tonnes)	GDP per capita
1	-0.007856	-0.523890	-0.493010	-0.617421	-0.329670	-0.126525	-0.438445
2	1.203432	-1.075569	-1.065051	-1.061099	-0.860509	-0.254038	-0.647075
3	-0.571254	0.812374	1.326364	1.144317	0.804594	7.071132	0.767085
4	-0.415296	1.502914	1.321484	1.401324	0.000108	-0.073264	2.486086
5	-0.217825	0.273014	0.098850	0.173300	1.606581	-0.127718	0.023755
6	0.217464	0.130685	0.730918	0.775726	-0.591181	-0.078123	-0.480558
7	-1.211055	0.850016	0.798223	0.806349	0.084414	0.017362	0.191617

Choix du meilleur groupe de pays



Clusters	Population difference (%)	Animal protein (%)	Food supply (kcal/capita/day)	Protein supply quantity (g/capita/day)	Poultry quantity (%)	Poultry Meat Production (1000 tonnes)	GDP per capita
1	-0.007856	-0.523890	-0.493010	-0.617421	-0.329670	-0.126525	-0.438445
2	1.203432	-1.075569	-1.065051	-1.061099	-0.860509	-0.254038	-0.647075
3	-0.571254	0.812374	1.326364	1.144317	0.804594	7.071132	0.767085
4	-0.415296	1.502914	1.321484	1.401324	0.000108	-0.073264	2.486086
5	-0.217825	0.273014	0.098850	0.173300	1.606581	-0.127718	0.023755
6	0.217464	0.130685	0.730918	0.775726	-0.591181	-0.078123	-0.480558
7	-1.211055	0.850016	0.798223	0.806349	0.084414	0.017362	0.191617

Choix du meilleur groupe de pays



Clusters	Population difference (%)	Animal protein (%)	Food supply (kcal/capita/day)	Protein supply quantity (g/capita/day)	Poultry quantity (%)	Poultry Meat Production (1000 tonnes)	GDP per capita
1	-0.007856	-0.523890	-0.493010	-0.617421	-0.329670	-0.126525	-0.438445
2	1.203432	-1.075569	-1.065051	-1.061099	-0.860509	-0.254038	-0.647075
3	-0.571254	0.812374	1.326364	1.144317	0.804594	7.071132	0.767085
4	-0.415296	1.502914	1.321484	1.401324	0.000108	-0.073264	2.486086
5	-0.217825	0.273014	0.098850	0.173300	1.606581	-0.127718	0.023755
6	0.217464	0.130685	0.730918	0.775726	-0.591181	-0.078123	-0.480558
7	-1.211055	0.850016	0.798223	0.806349	0.084414	0.017362	0.191617

Choix du meilleur groupe de pays



Clusters	Population difference (%)	Animal protein (%)	Food supply (kcal/capita/day)	Protein supply quantity (g/capita/day)	Poultry quantity (%)	Poultry Meat Production (1000 tonnes)	GDP per capita
1	-0.007856	-0.523890	-0.493010	-0.617421	-0.329670	-0.126525	-0.438445
2	1.203432	-1.075569	-1.065051	-1.061099	-0.860509	-0.254038	-0.647075
3	-0.571254	0.812374	1.326364	1.144317	0.804594	7.071132	0.767085
4	-0.415296	1.502914	1.321484	1.401324	0.000108	-0.073264	2.486086
5	-0.217825	0.273014	0.098850	0.173300	1.606581	-0.127718	0.023755
6	0.217464	0.130685	0.730918	0.775726	-0.591181	-0.078123	-0.480558
7	-1.211055	0.850016	0.798223	0.806349	0.084414	0.017362	0.191617

Choix du meilleur groupe de pays



Clusters	Population difference (%)	Animal protein (%)	Food supply (kcal/capita/day)	Protein supply quantity (g/capita/day)	Poultry quantity (%)	Poultry Meat Production (1000 tonnes)	GDP per capita
1	-0.007856	-0.523890	-0.493010	-0.617421	-0.329670	-0.126525	-0.438445
2	1.203432	-1.075569	-1.065051	-1.061099	-0.860509	-0.254038	-0.647075
3	-0.571254	0.812374	1.326364	1.144317	0.804594	7.071132	0.767085
4	-0.415296	1.502914	1.321484	1.401324	0.000108	-0.073264	2.486086
5	-0.217825	0.273014	0.098850	0.173300	1.606581	-0.127718	0.023755
6	0.217464	0.130685	0.730918	0.775726	-0.591181	-0.078123	-0.480558
7	-1.211055	0.850016	0.798223	0.806349	0.084414	0.017362	0.191617

Choix du meilleur groupe de pays



Clusters	Population difference (%)	Animal protein (%)	Food supply (kcal/capita/day)	Protein supply quantity (g/capita/day)	Poultry quantity (%)	Poultry Meat Production (1000 tonnes)	GDP per capita
1	-0.007856	-0.523890	-0.493010	-0.617421	-0.329670	-0.126525	-0.438445
2	1.203432	-1.075569	-1.065051	-1.061099	-0.860509	-0.254038	-0.647075
3	-0.571254	0.812374	1.326364	1.144317	0.804594	7.071132	0.767085
4	-0.415296	1.502914	1.321484	1.401324	0.000108	-0.073264	2.486086
5	-0.217825	0.273014	0.098850	0.173300	1.606581	-0.127718	0.023755
6	0.217464	0.130685	0.730918	0.775726	-0.591181	-0.078123	-0.480558
7	-1.211055	0.850016	0.798223	0.806349	0.084414	0.017362	0.191617

Choix du meilleur groupe de pays



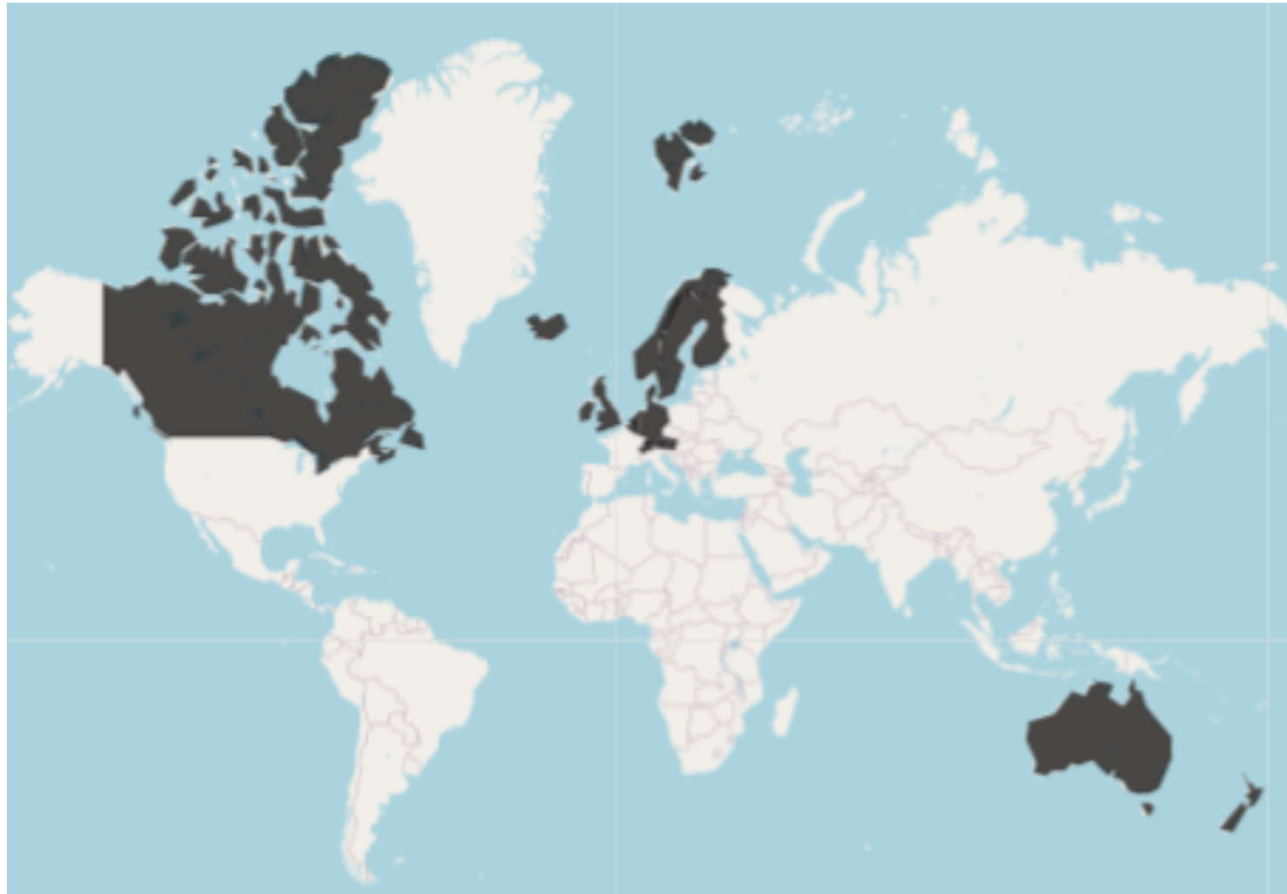
Clusters	Population difference (%)	Animal protein (%)	Food supply (kcal/capita/day)	Protein supply quantity (g/capita/day)	Poultry quantity (%)	Poultry Meat Production (1000 tonnes)	GDP per capita
1	-0.007856	-0.523890	-0.493010	-0.617421	-0.329670	-0.126525	-0.438445
2	1.203432	-1.075569	-1.065051	-1.061099	-0.860509	-0.254038	-0.647075
3	-0.571254	0.812374	1.326364	1.144317	0.804594	7.071132	0.767085
4	-0.415296	1.502914	1.321484	1.401324	0.000108	-0.073264	2.486086
5	-0.217825	0.273014	0.098850	0.173300	1.606581	-0.127718	0.023755
6	0.217464	0.130685	0.730918	0.775726	-0.591181	-0.078123	-0.480558
7	-1.211055	0.850016	0.798223	0.806349	0.084414	0.017362	0.191617

Choix du meilleur groupe de pays



Clusters	Population difference (%)	Animal protein (%)	Food supply (kcal/capita/day)	Protein supply quantity (g/capita/day)	Poultry quantity (%)	Poultry Meat Production (1000 tonnes)	GDP per capita
1	-0.007856	-0.523890	-0.493010	-0.617421	-0.329670	-0.126525	-0.438445
2	1.203432	-1.075569	-1.065051	-1.061099	-0.860509	-0.254038	-0.647075
3	-0.571254	0.812374	1.326364	1.144317	0.804594	7.071132	0.767085
4	-0.415296	1.502914	1.321484	1.401324	0.000108	-0.073264	2.486086
5	-0.217825	0.273014	0.098850	0.173300	1.606581	-0.127718	0.023755
6	0.217464	0.130685	0.730918	0.775726	-0.591181	-0.078123	-0.480558
7	-1.211055	0.850016	0.798223	0.806349	0.084414	0.017362	0.191617

Carte graphique du groupe des pays retenus



Groupe des pays retenu

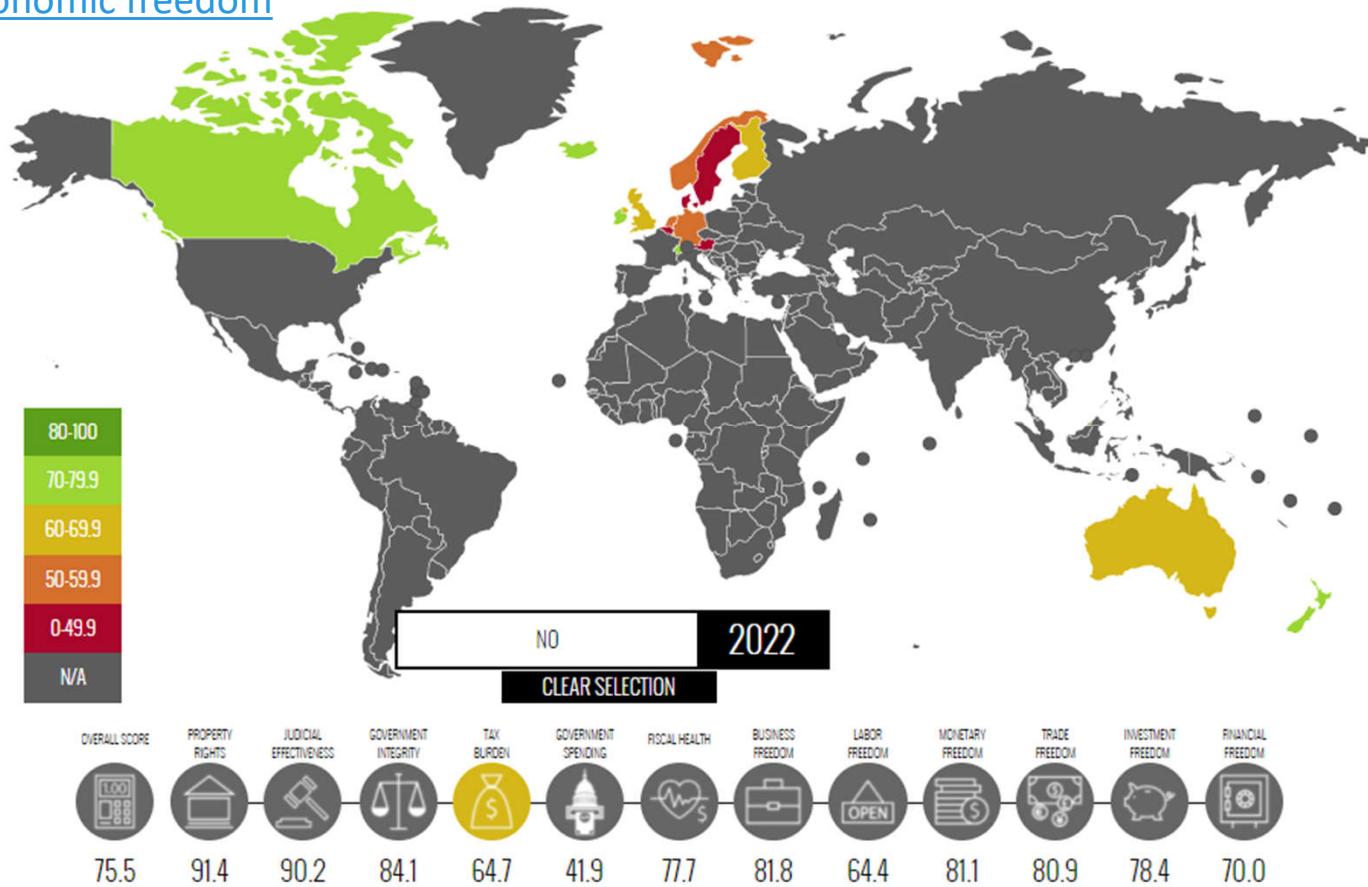


	Area	Country Code	Year	Region
1	Australia	AUS	2019	East Asia & Pacific
2	New Zealand	NZL	2019	East Asia & Pacific
3	Austria	AUT	2019	Europe & Central Asia
4	Denmark	DNK	2019	Europe & Central Asia
5	Finland	FIN	2019	Europe & Central Asia
6	Germany	DEU	2019	Europe & Central Asia
7	Iceland	ISL	2019	Europe & Central Asia
8	Ireland	IRL	2019	Europe & Central Asia
9	Netherlands	NLD	2019	Europe & Central Asia
10	Norway	NOR	2019	Europe & Central Asia
11	Sweden	SWE	2019	Europe & Central Asia
12	Switzerland	CHE	2019	Europe & Central Asia
13	United Kingdom	GBR	2019	Europe & Central Asia
14	Belgium	BEL	2019	Europe & Central Asia
15	Luxembourg	LUX	2019	Europe & Central Asia
16	Canada	CAN	2019	North America

Classification des pays selon le fardeau fiscal



Source : [index of economic freedom](https://www.economicfreedom.com/)

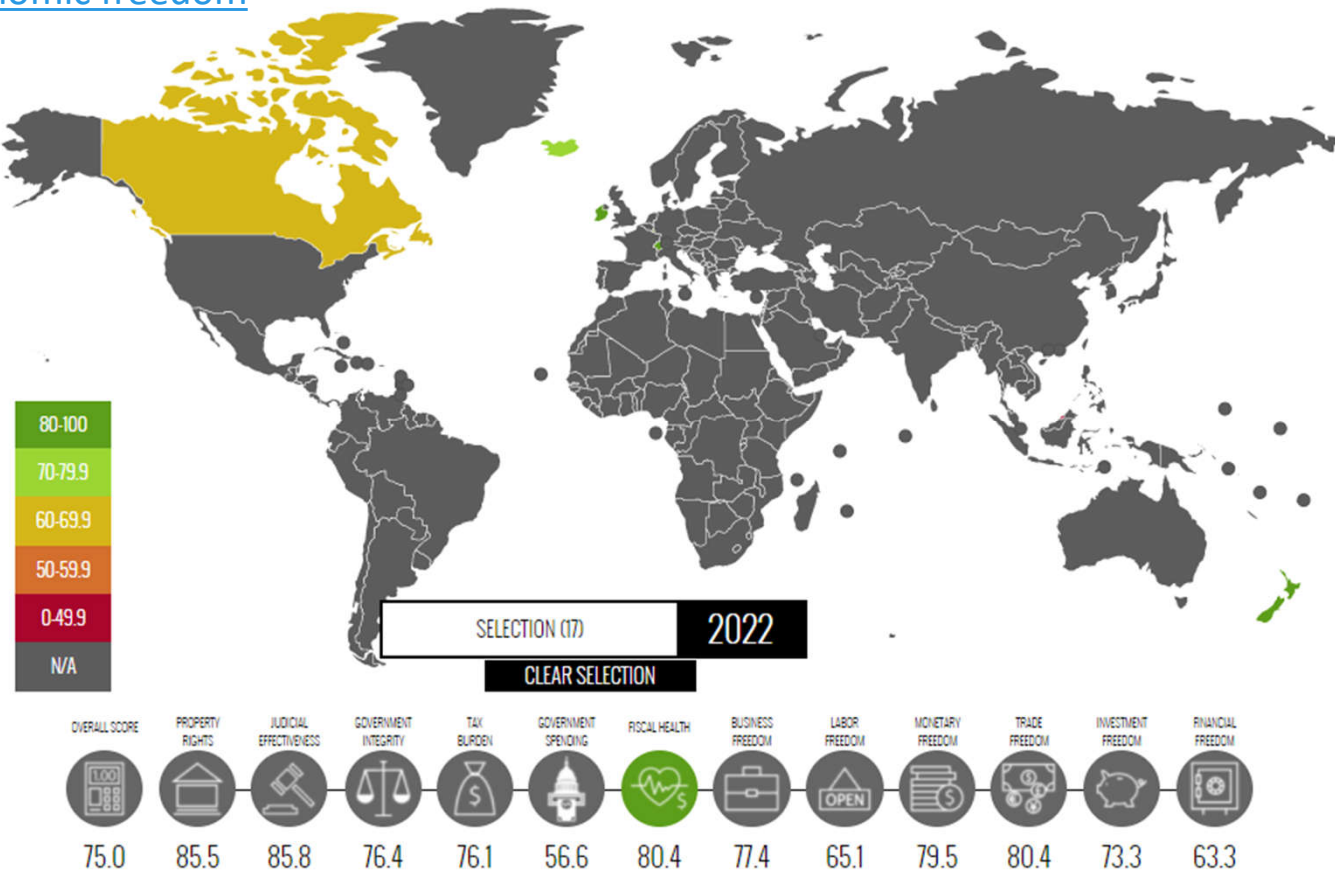


17/05/2022

Classification des pays selon la santé budgétaire



Source : [index of economic freedom](https://www.economicfreedom.com/)

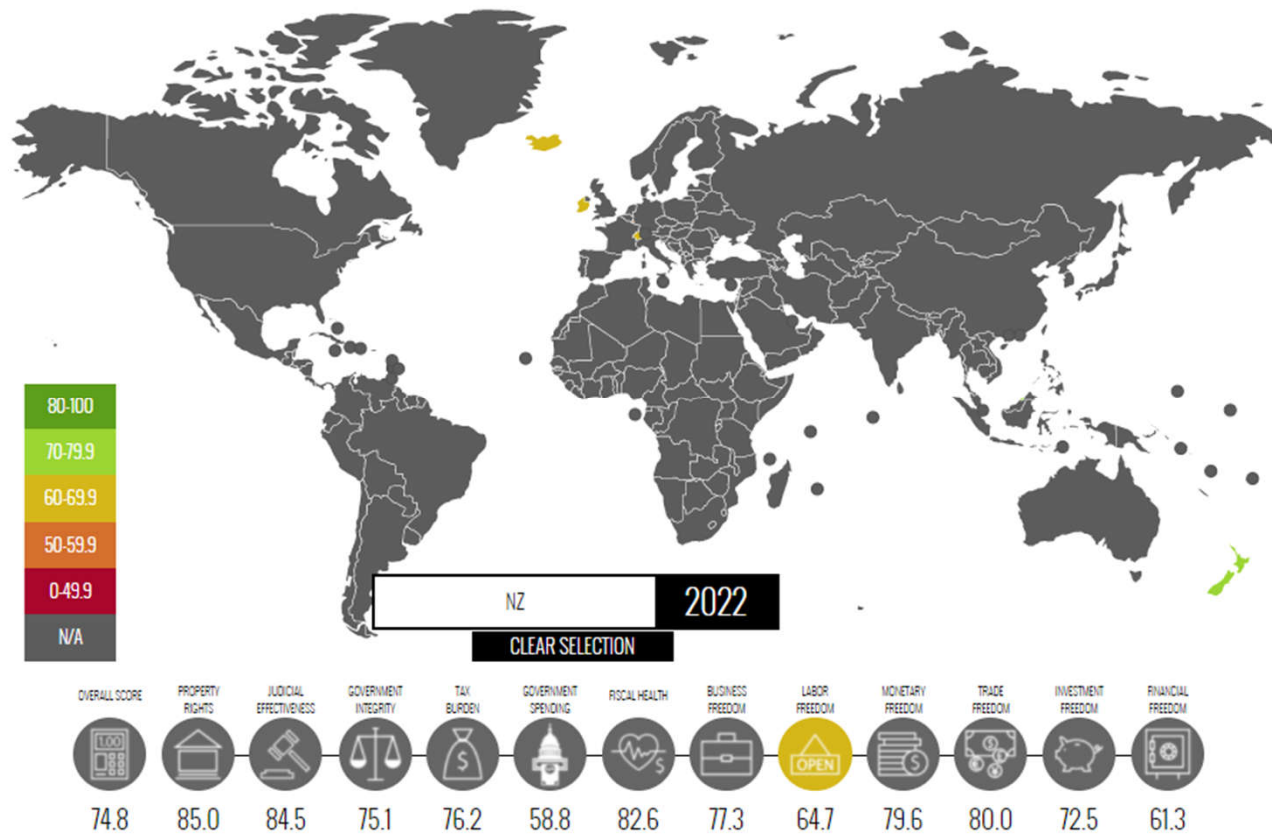


17/05/2022

Classification des pays selon la main-d'œuvre



Source : [index of economic freedom](https://www.economicfreedom.com/)



17/05/2022

Nouvelle-Zélande



- Population:
 - 5,1 millions
- PIB (PPA - parités de pouvoir d'achat):
 - 213,9 milliards de dollars
 - -3,0% de croissance
 - Croissance annuelle de 2,1 % sur 5 ans
 - 42 018 \$ par habitant
- Chômage:
 - 4,6 %
- Inflation (IPC):
 - 1,7 %
- Entrée d'IDE (Investissement direct à l'étranger) :
 - 4,2 milliards de dollars

Classement des pays selon l'indice de liberté économique

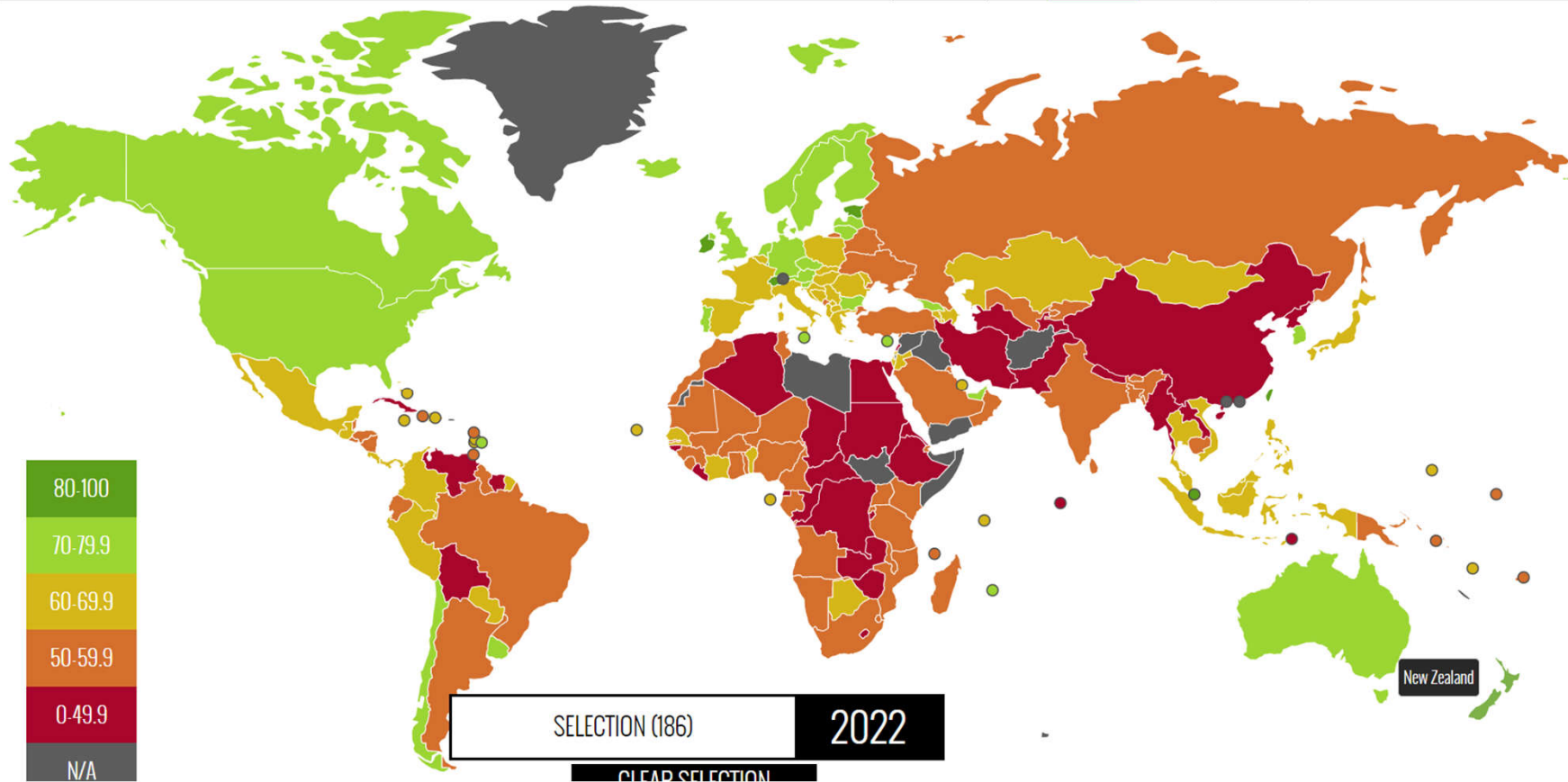


Suite ..

Area	Rang	Score
Switzerland	2	84,20
Ireland	3	82,00
New Zealand	4	80,60
Luxembourg	5	80,60
Netherlands	8	80,60
Finland	9	78,30
Denmark	10	78,00
Sweden	11	77,90

Area	Rang	Score
Australia	12	77,70
Iceland	13	77,00
Norway	14	76,90
Canada	15	76,60
Germany	16	76,10
Austria	22	73,80
United Kingdom	24	72,70
Belgium	37	69,60

Catégorisation mondiale selon l'indice de liberté économique



17/05/2022

Source : [index of economic freedom](https://www.economicfreedom.com/)

Partie III

ACP

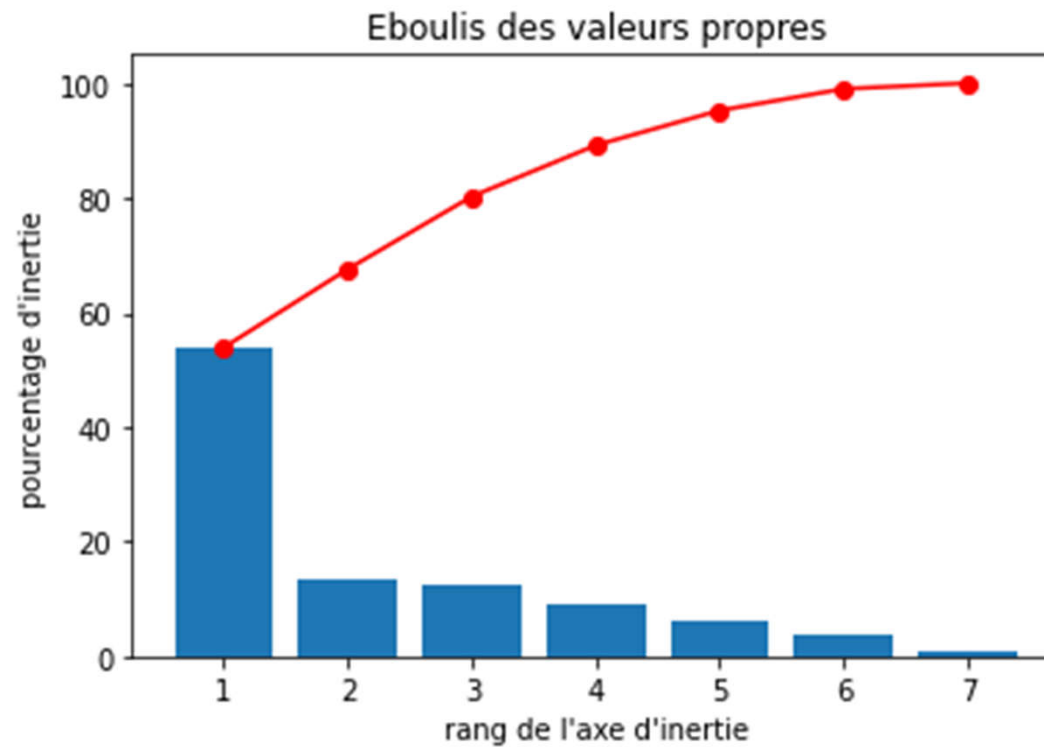
Analyse en Composantes Principales ACP

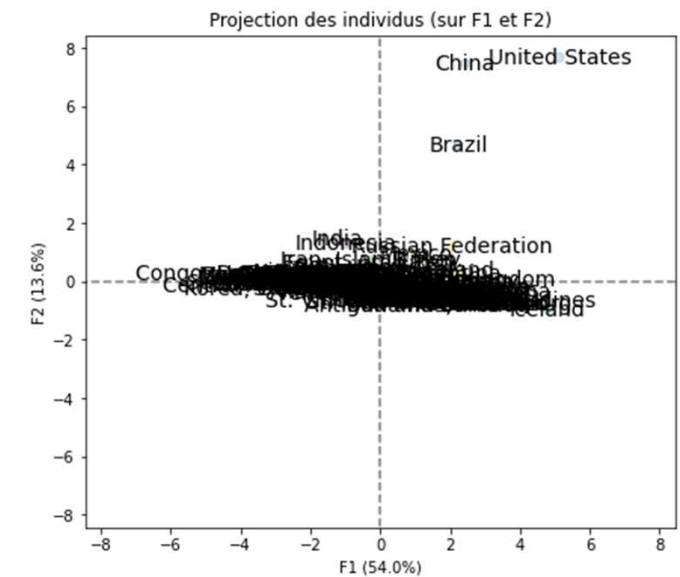
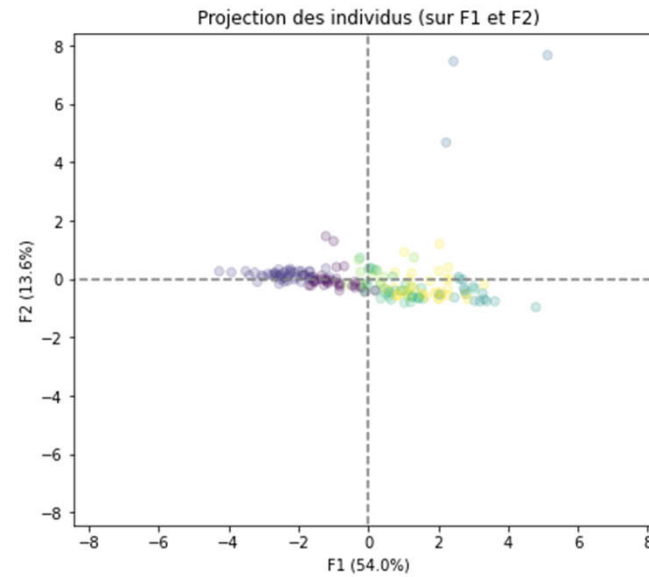


➤ Objectifs de l'ACP :

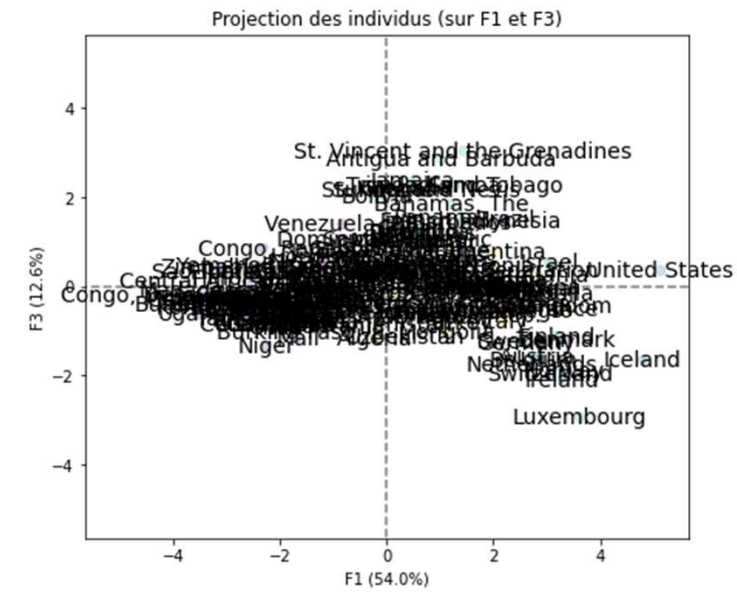
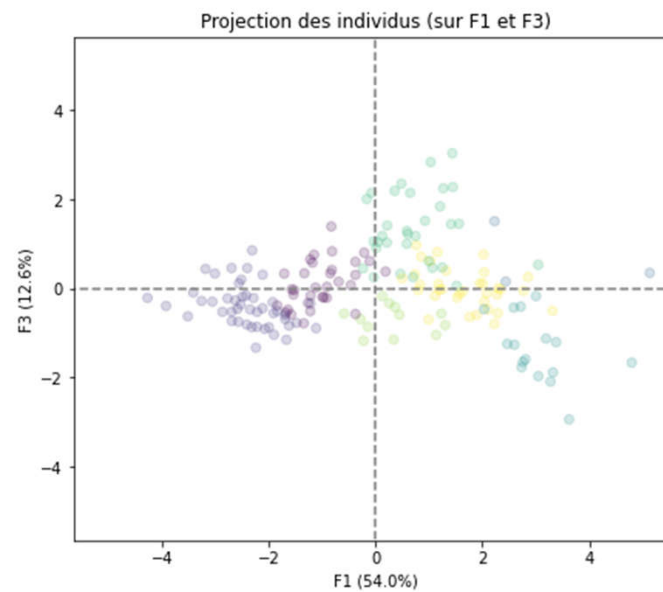
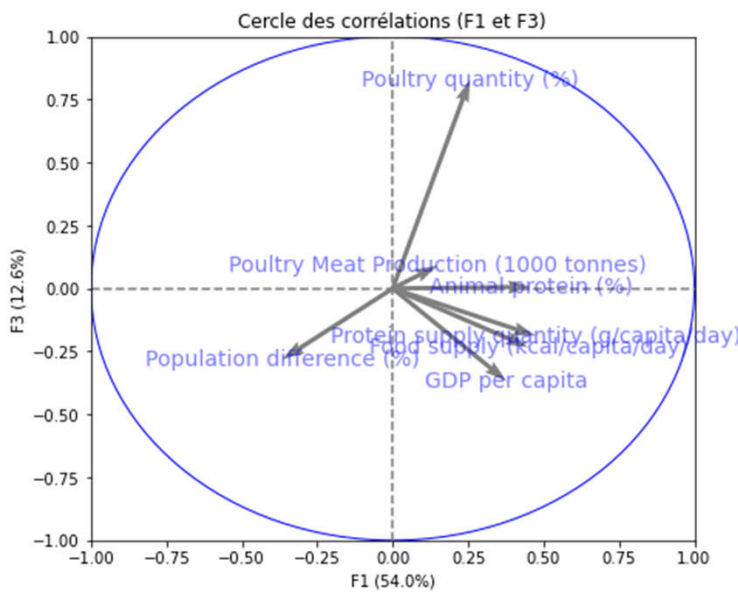
- Etudier la variabilité des individus (leurs ressemblances et différences) ;
- Etudier les liaisons entre les variables (et au besoin, regrouper les variables liées en nouvelles variables synthétiques pour réduire le nombre de colonnes de nos données). Ces nouvelles variables sont nommées « composantes principales » ou axes principaux

Eblouis des valeurs propres

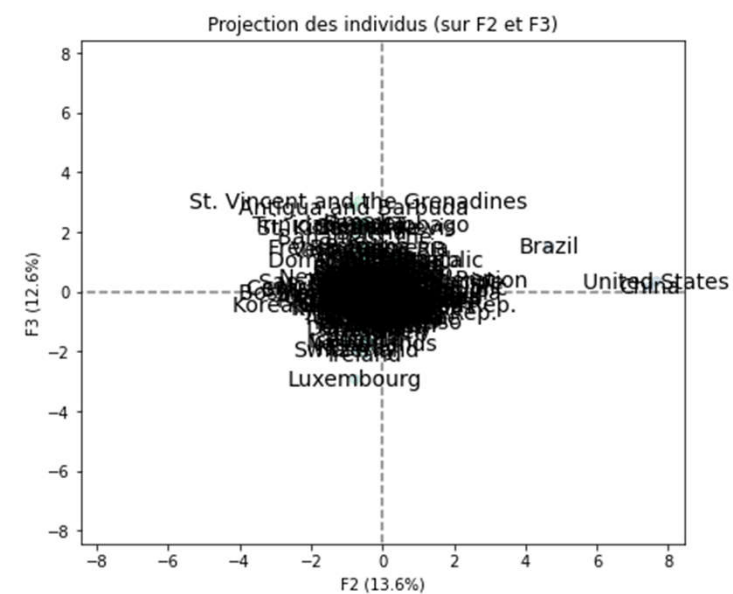
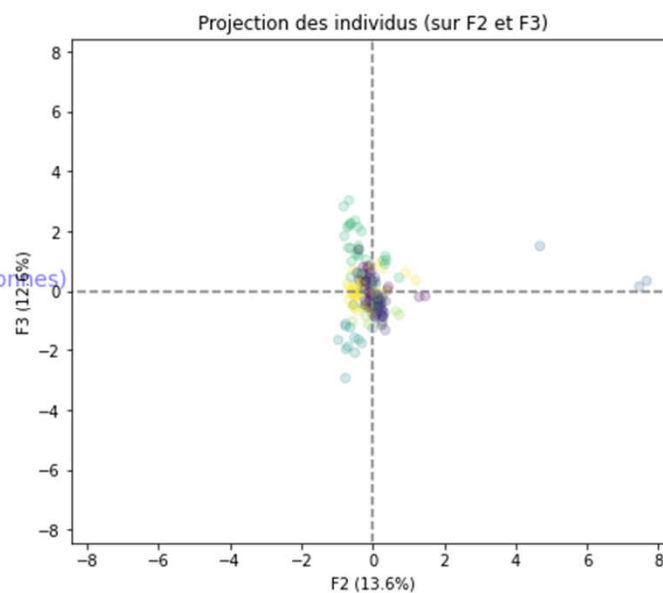
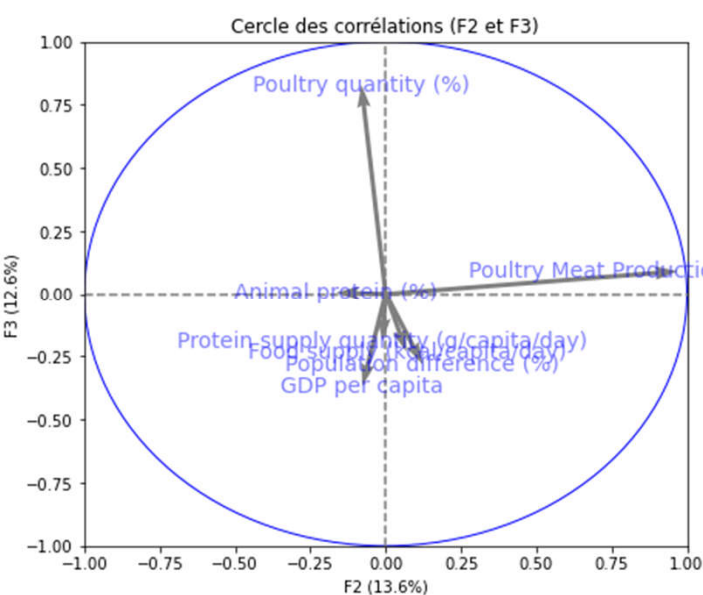
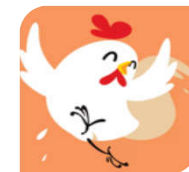




Projection sur le plan factoriel (F1,F3)



Projection sur le plan factoriel (F2,F3)



Partie VI

Statistique inférentielle

Formalisez un problème de test



➤ Les hypothèses d'un test

- Hypothèse nulle $H_0 \rightarrow$ l'hypothèse qu'on ne souhaite pas rejeter trop facilement
- L'hypothèse alternative $H_1 \rightarrow$ indique dans quelles conditions on rejette H_0

➤ La p-valeur du test :

- la plus petite valeur du niveau de test conduisant au rejet de H_0
- $\{\text{Rejet de } H_0 \text{ au niveau de test } \alpha\} \Leftrightarrow \{p\text{-valeur} < \alpha\}$

Normalité d'un jeu de donnée



- Il s'agit donc de s'assurer que les variables continues sont distribuées selon la loi normale.
 - Si cela est le cas, les tests d'hypothèse classiques sont applicables.
 - Sinon, il faudra trouver une alternative dite "non paramétrique" au test d'hypothèse à réaliser.

Approches graphiques



➤ Histogramme de fréquence

- Représenter les données à l'aide de l'histogramme de fréquence et regarder si elles semblent s'ajuster à une distribution normale.

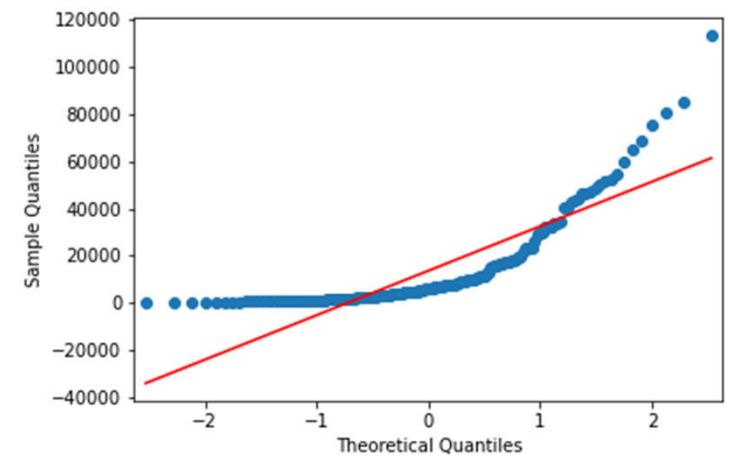
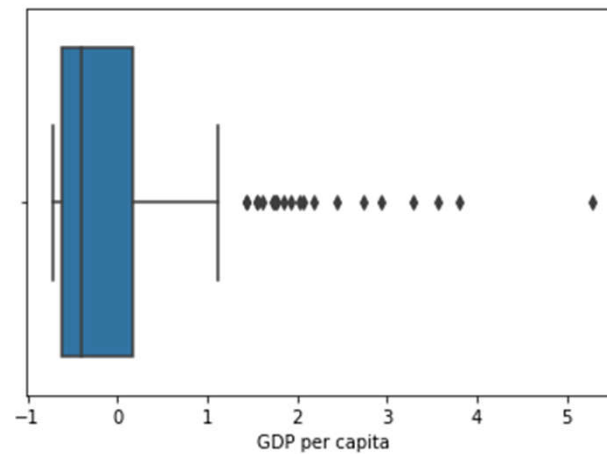
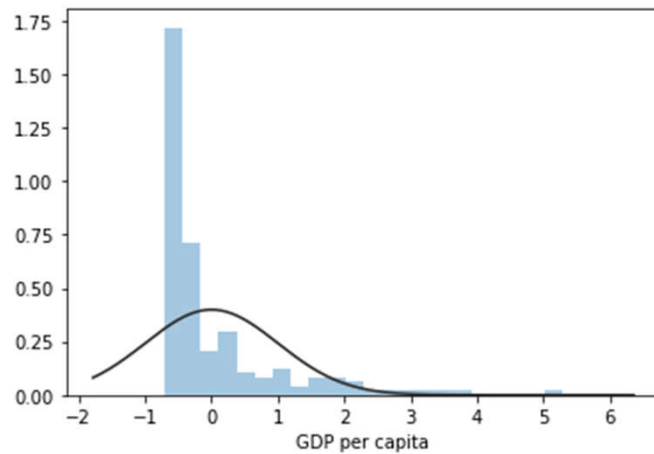
➤ Boîte à moustache

- Permet d'observer les valeurs extrêmes (outliers) mais également d'avoir une idée sur la symétrie de la distribution.
- La symétrie d'une distribution n'affirme pas la normalité, mais une distribution normale est forcément symétrique.

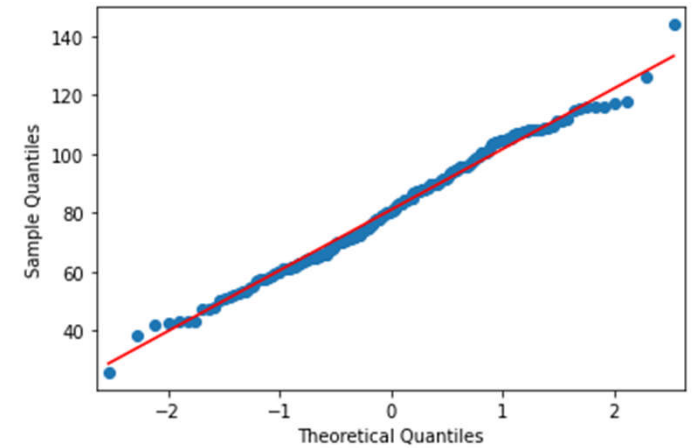
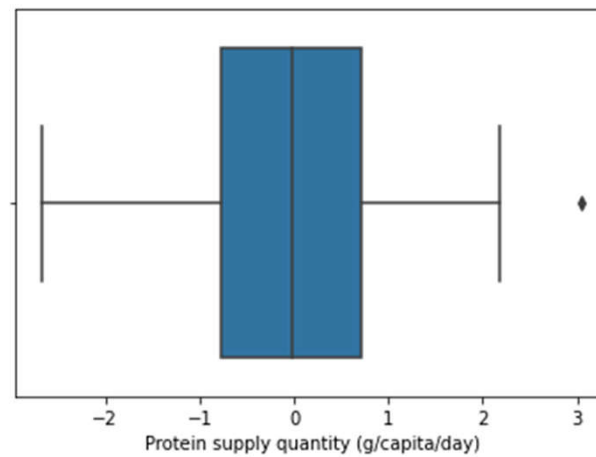
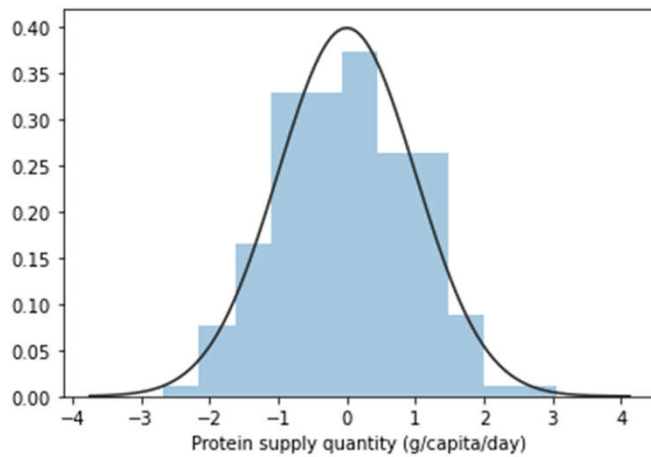
➤ QQ-plot

- "diagramme Quantile-Quantile" ou "Q-Q plot" est un outil graphique permettant d'évaluer la pertinence de l'ajustement d'une distribution donnée à un modèle théorique

PIB par habitant



Disponibilité de protéines en quantité (g/personne/jour)





➤ Test de D'Agostino - Pearson

- Test statistique qui utilise la forme de la distribution de X , telle qu'elle est mesurée par l'asymétrie (skewness) et l'aplatissement (kurtosis).
- L'hypothèse nulle H_0 est la normalité de la distribution parente,
- l'hypothèse alternative H_1 est le défaut de normalité.
- **α 0.05**

Test d'Agostino - Pearson



➤ PIB par habitant

- statistic=101.41045982379198
- pvalue=9.527927539016695e-23
- **p-value < 0.05, donc on rejette l'hypothèse nulle**

➤ Disponibilité de protéines en quantité (g/personne/jour)

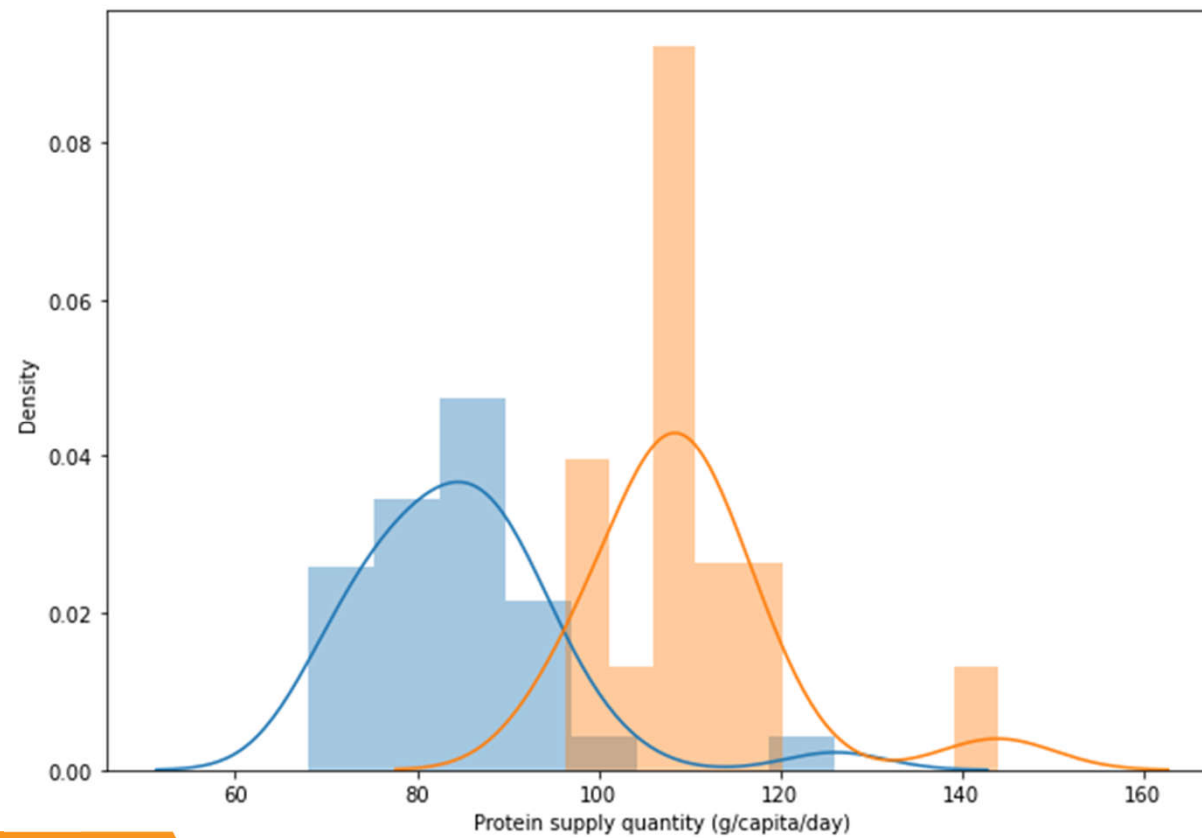
- statistic=1.6463574949110016
- pvalue=0.4390338563056493
- **p-value > 0.05, donc on ne rejette pas l'hypothèse nulle**

Comparaison de population



- Comparaison entre Cluster 5 et Cluster 1 pour la variable :
 - Disponibilité de protéines en quantité (g/personne/jour)
- **Test d'égalité des variances**
 - BartlettResult(statistic=5.292680889634976, pvalue=0.021415228227561426)
 - La p-valeur $> 0,05$, on rejette l'égalité des variances (niveau de test 5%)
- **Test d'égalité des moyennes**
 - Ttest_indResult(statistic=-6.942699317003766, pvalue=5.385888153806907e-09)
 - On constate que l'hypothèse d'égalité des moyennes est (très facilement) rejetée à un niveau de test de 5%

Comparaison de population



17/05/2022

Conclusion et perspectives

Des question ?