



LA POULE QUI CHANTE

SOURCE DE DONNÉES : POP

	Zone	Taux_evol_pop
0	Afghanistan	10.23
1	Afrique du sud	5.62
2	Albanie	-0.47
3	Algérie	7.83
4	Allemagne	2.01
...
233	Îles salomon	10.08
234	Îles turques-et-caïques	6.10
235	Îles vierges américaines	-0.41
236	Îles vierges britanniques	2.73
237	Îles wallis-et-futuna	-5.82

238 rows × 2 columns

Calcul du taux l'évolution de la population entre 2014 et 2018

SOURCE DE DONNÉES : DISPO

	Zone	Production	Disponibilité alimentaire (Kcal/personne/jour)	Disponibilité de protéines en quantité (g/personne/jour)	Exportations - Quantité	Importations - Quantité
0	Afghanistan	28.0	5.0	0.54	NaN	29.0
1	Afrique du sud	1667.0	143.0	14.11	63.0	514.0
2	Albanie	13.0	85.0	6.26	0.0	38.0
3	Algérie	275.0	22.0	1.97	0.0	2.0
4	Allemagne	1514.0	71.0	7.96	646.0	842.0
...
167	Émirats arabes unis	48.0	147.0	14.80	94.0	433.0
168	Équateur	340.0	83.0	6.15	0.0	0.0
169	États-unis d'amérique	21914.0	219.0	19.93	3692.0	123.0
170	Éthiopie	14.0	0.0	0.04	NaN	1.0
171	Îles salomon	0.0	18.0	1.51	0.0	6.0

172 rows x 6 columns

Disponibilité alimentaire concentrée sur la viande de volaille sur l'année 2017

SOURCE DE DONNÉES : CONTINENT ET DISTANCE

	Zone	Continent	distance en km
0	Afghanistan	Asie	5457
1	Afrique du sud	Afrique	8921
2	Albanie	Europe	1635
3	Algérie	Afrique	2320
4	Allemagne	Europe	694

Distance au départ de Paris en avion,
avec continent associé

SOURCE DE DONNÉES : STABILITÉ POLITIQUE

	Zone	Indice stabilite
0	Afghanistan	-2.80
1	Afrique du sud	-0.28
2	Albanie	0.38
3	Algérie	-0.92
4	Allemagne	0.59
...
192	Venezuela (république bolivarienne du)	-1.25
193	Viet nam	0.29
194	Yémen	-2.94
195	Zambie	0.15
196	Zimbabwe	-0.71

197 rows x 2 columns

Indice de stabilité politique par pays
sur l'année 2017

SOURCE DE DONNÉES : PIB PAR HABITANTS

	Zone	pib2017
15	Afghanistan	2058.4
36	Afrique du sud	12701.3
57	Albanie	12771.0
78	Algérie	11737.4
99	Allemagne	52952.9
...
3795	Uruguay	23009.9
3816	Vanuatu	3081.8
3837	Viet nam	7155.4
3858	Zambie	3485.0
3879	Zimbabwe	3274.6

PIB par habitant sur l'année 2017

SOURCE DE DONNÉES : TABLE FINALE

	Zone	Taux_evol_pop	pib2017	Production	Disponibilité alimentaire (Kcal/personne /jour)	Disponibilité de protéines en quantité (g/personne/jour)	Exportations - Quantité	Importations - Quantité	Continent	distance en km	Indice stabilite
0	Afghanistan	10.23	2058.4	28.0	5.0	0.54	NaN	29.0	Asie	5457	-2.80
1	Afrique du sud	5.62	12701.3	1667.0	143.0	14.11	63.0	514.0	Afrique	8921	-0.28
2	Albanie	-0.47	12771.0	13.0	85.0	6.26	0.0	38.0	Europe	1635	0.38
3	Algérie	7.83	11737.4	275.0	22.0	1.97	0.0	2.0	Afrique	2320	-0.92
4	Allemagne	2.01	52952.9	1514.0	71.0	7.96	646.0	842.0	Europe	694	0.59
...
233	Îles salomon	10.08	2663.9	0.0	18.0	1.51	0.0	6.0	Océanie	14974	0.20
234	Îles turques- et-caïques	6.10	NaN	NaN	NaN	NaN	NaN	NaN	Amérique	7048	NaN
235	Îles vierges américaines	-0.41	NaN	NaN	NaN	NaN	NaN	NaN	Amérique	6958	NaN
236	Îles vierges britanniques	2.73	NaN	NaN	NaN	NaN	NaN	NaN	Amérique	6965	NaN
237	Îles wallis-et- futuna	-5.82	NaN	NaN	NaN	NaN	NaN	NaN	Océanie	16173	NaN

Nous pouvons
remarquer qu'il
y a pas mal de
valeur NaN

SOURCE DE DONNÉES : FINALITÉ

```
Taux_evol_pop 2
pib2017 51
Production 70
Disponibilité alimentaire (Kcal/personne/jour) 66
Disponibilité de protéines en quantité (g/personne/jour) 66
Exportations - Quantité 103
Importations - Quantité 68
Indice stabilite 41
```

Nombre de NaN par colonne

Au vu du nombre important de NaN pour l'analyse, et pour cibler au mieux les pays intéressants, je supprime les pays où je n'ai pas toutes les informations

238 rows x 11 columns



129 rows x 11 columns



Il me reste 129
pays à analyser

DENDROGRAMME OU CAH

La Classification Ascendante Hiérarchique est un algorithme de clustering. C'est à dire qu'à partir de données, l'algorithme va chercher à créer des groupes d'individus homogènes : Des groupes dans lesquels les individus se ressemblent. Des groupes qui se distinguent le plus possible les uns des autres.

DONNÉES STANDARDISÉES

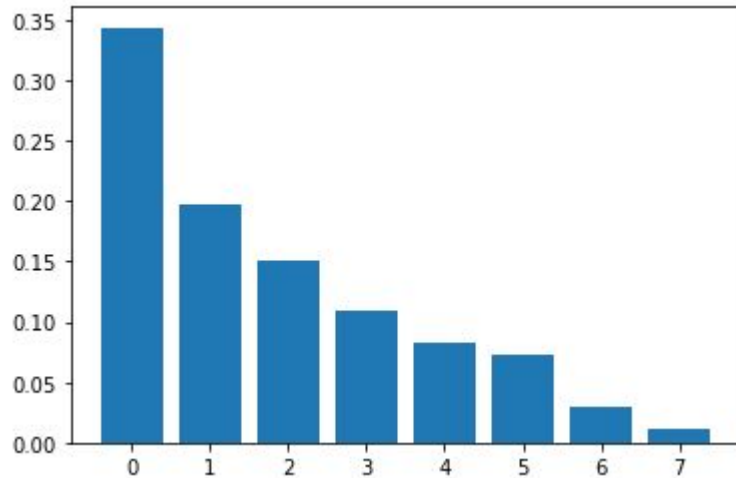
La standardisation des données a pour objectif d'assurer une compatibilité optimale des données, en vue de leur réutilisation. L'application d'une « commune mesure » permet d'améliorer la qualité des données. Cette signification fait référence à la transformation de données en soustrayant à chaque valeur une valeur de référence (classiquement une moyenne d'échantillon) et en la divisant par l'écart-type. Cette importante transformation rendra toutes les valeurs en unités compatibles avec distribution de moyenne 0 et d'écart-type 1.

PCA

L'**analyse en composantes principales** (ACP ou PCA en anglais pour *principal component analysis*), est une méthode de la famille de l'analyse des données qui consiste à transformer des variables liées entre elles (dites « corrélées » en statistique) en nouvelles variables décorrélées les unes des autres. Ces nouvelles variables sont nommées « composantes principales ».

Elle permet de réduire le nombre de variables et de rendre l'information moins redondante.

DÉTECTION DES COMPOSANTES

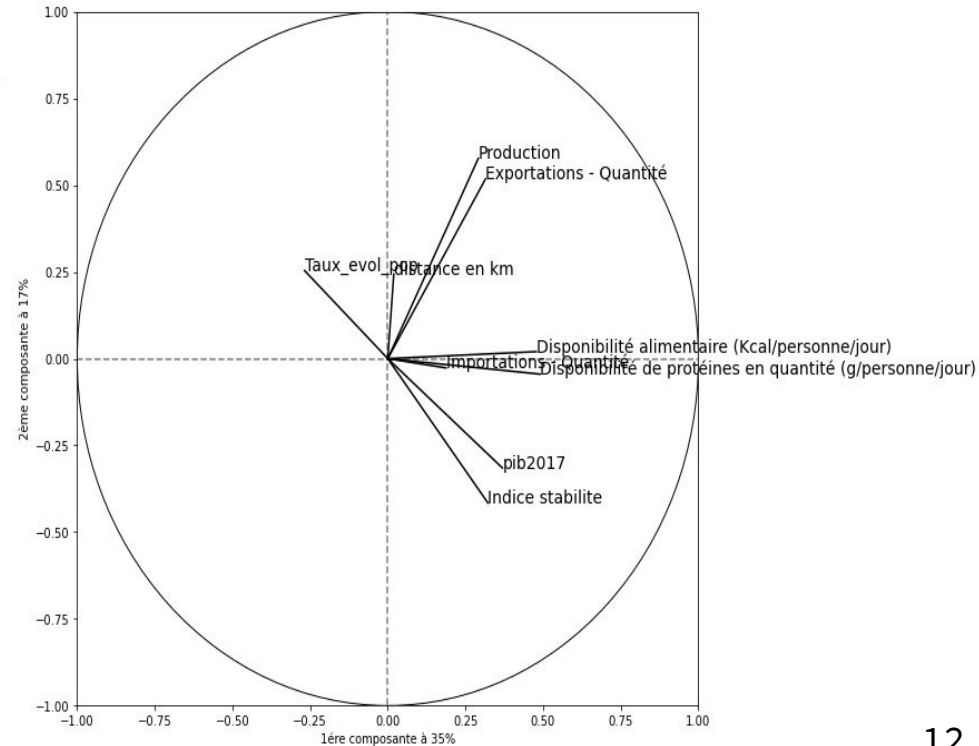
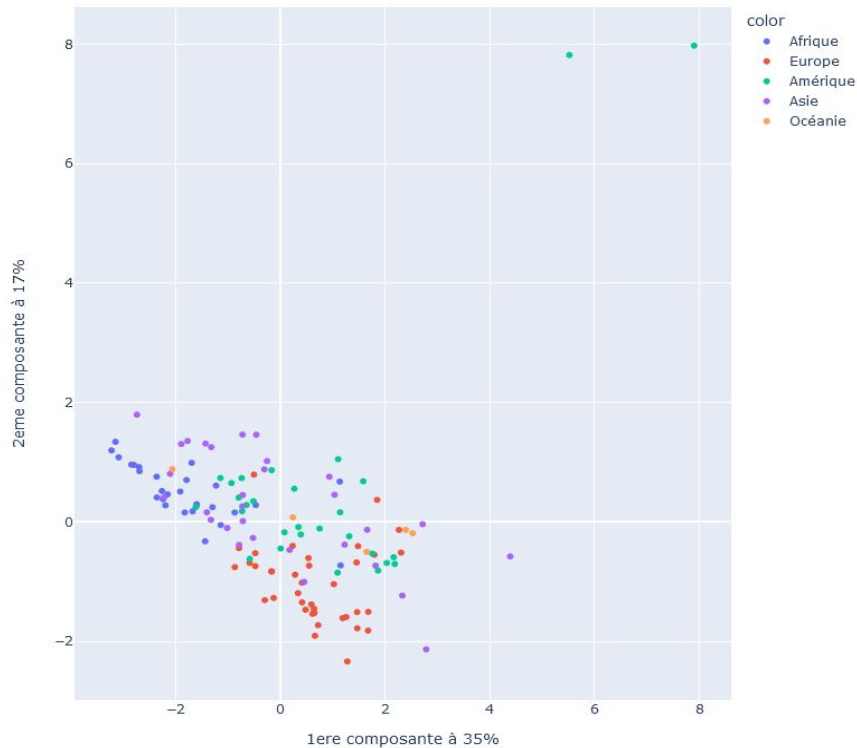


La cassure veut que je garde 2 composantes afin de les représenter graphiquement.

Avec 3 composantes, je garde 68% de nos données.

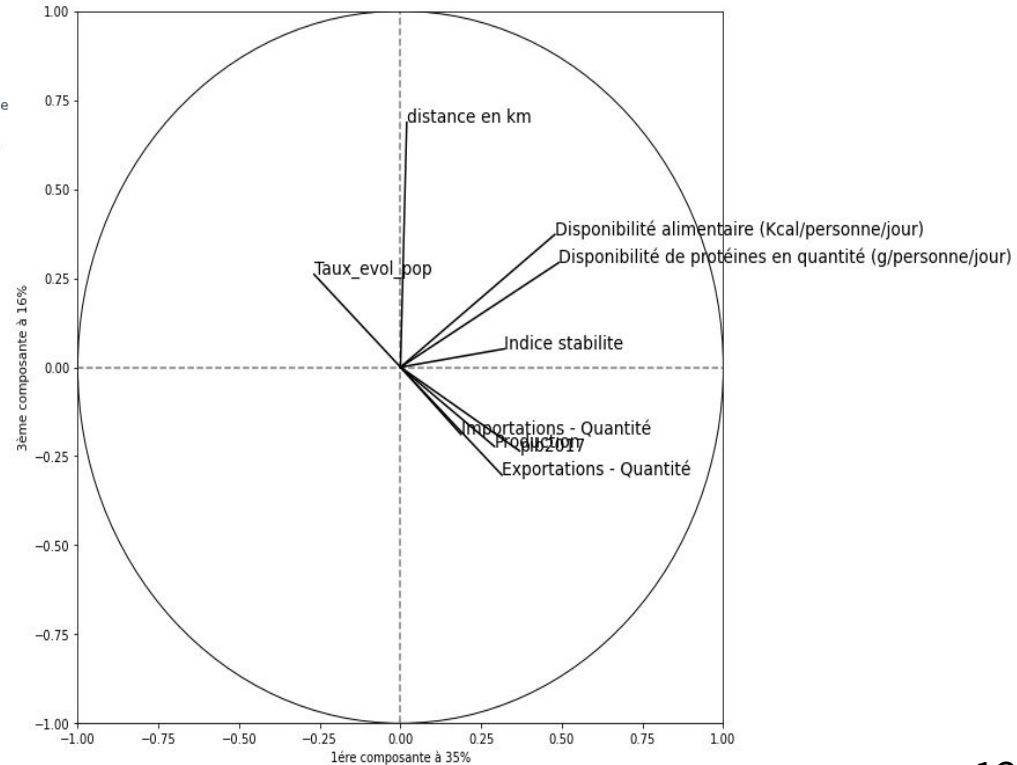
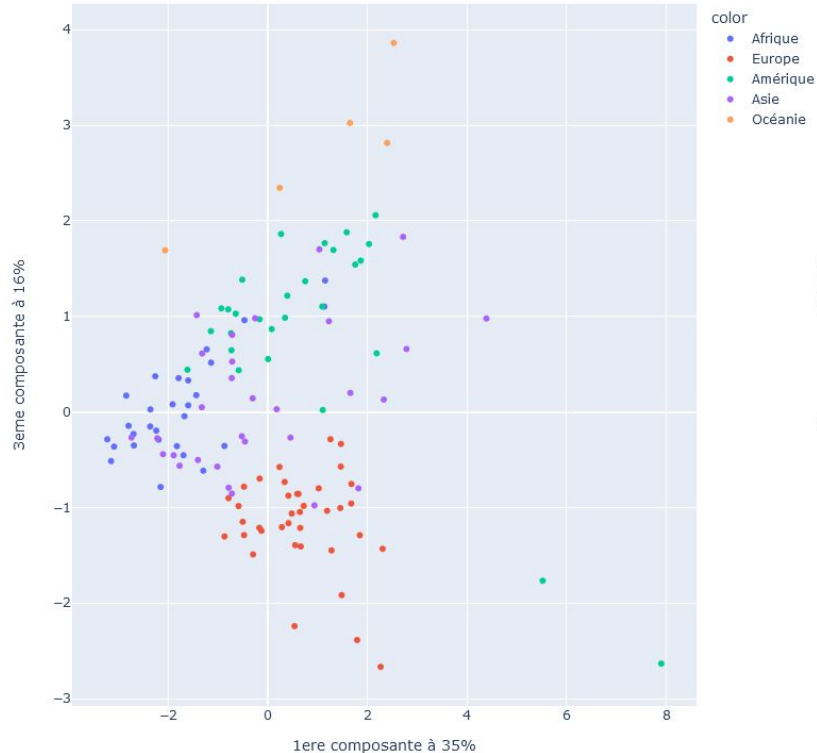
REPRÉSENTATION DES COMPOSANTES 1 ET 2

représentation des 2 premières composantes



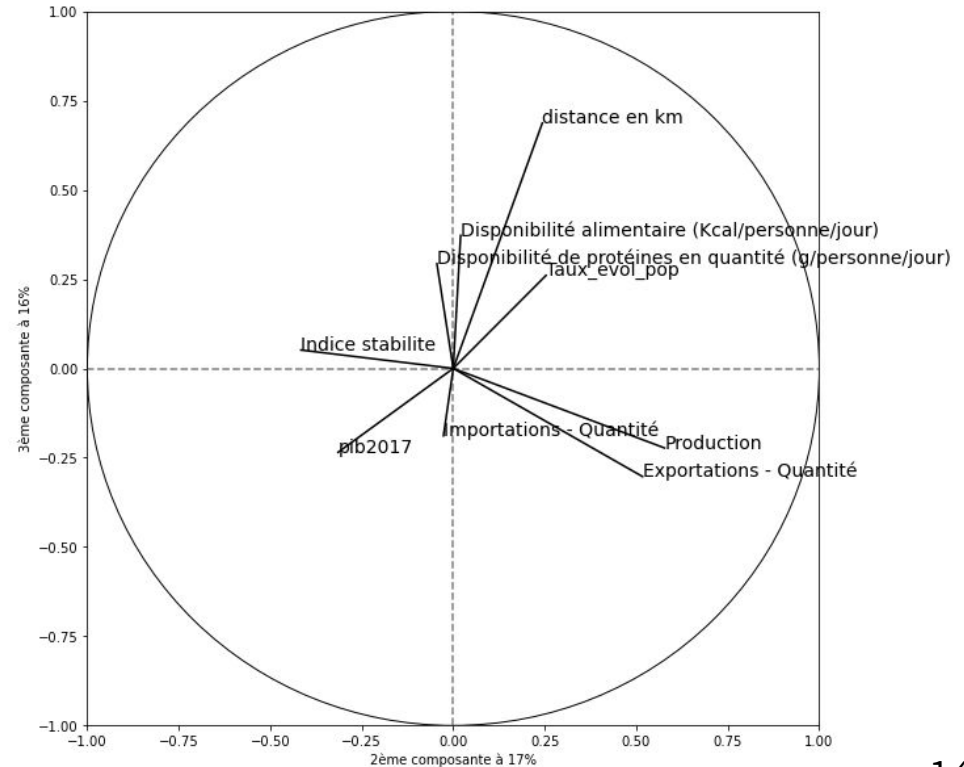
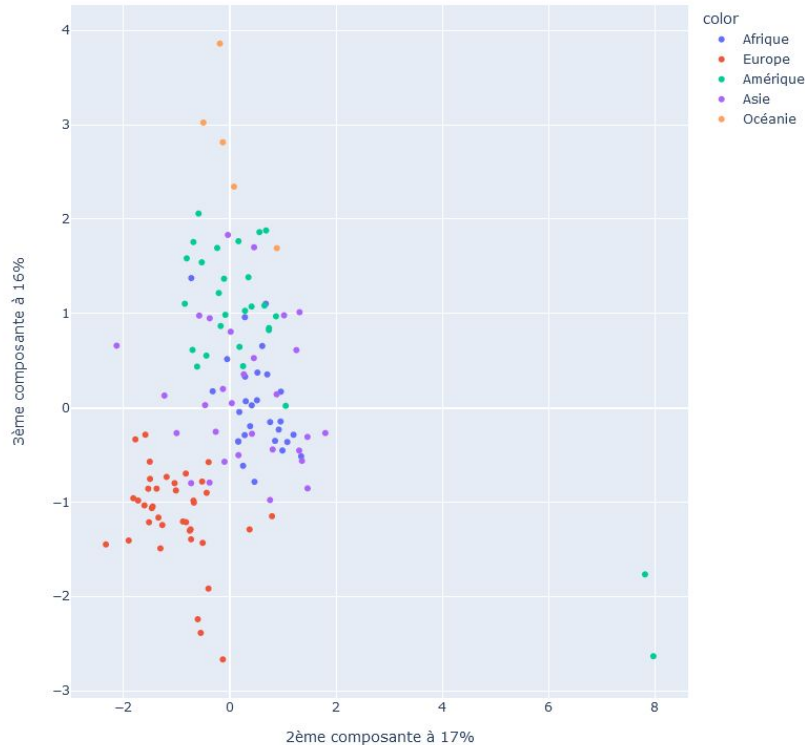
REPRÉSENTATION DES COMPOSANTES 1 ET 3

représentation de la 1ère et 3ème composantes



REPRÉSENTATION DES COMPOSANTES 2 ET 3

représentation de la 2ème et 3ème composantes



ANALYSE



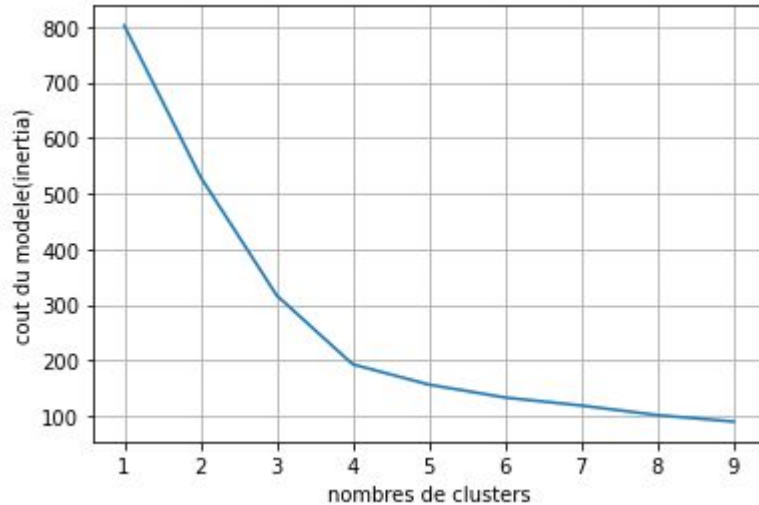
 publicdomainvectors.org

L'Europe a le plus fort PIB par personne, se trouve dans a
une distance honorable du point de vue écologique et a un
indice de stabilité politique satisfaisant !

K-MEANS

Le clustering est une discipline particulière du Machine Learning ayant pour objectif de séparer vos données en groupes homogènes ayant des caractéristiques communes.

REPRÉSENTATION DU NOMBRE DE CLUSTERS

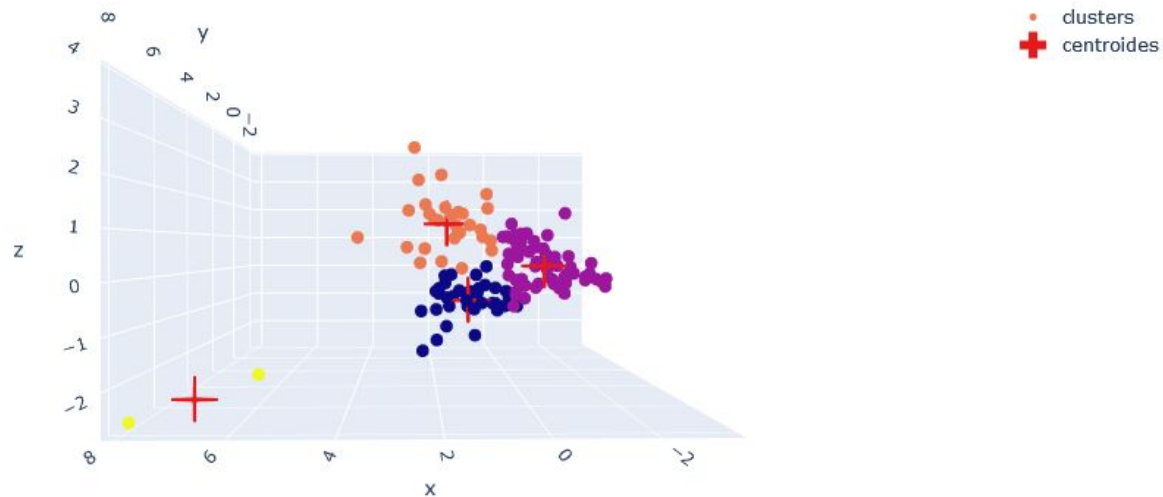


Nous utilisons la technique du coude pour pouvoir déterminer le bon nombre de cluster.

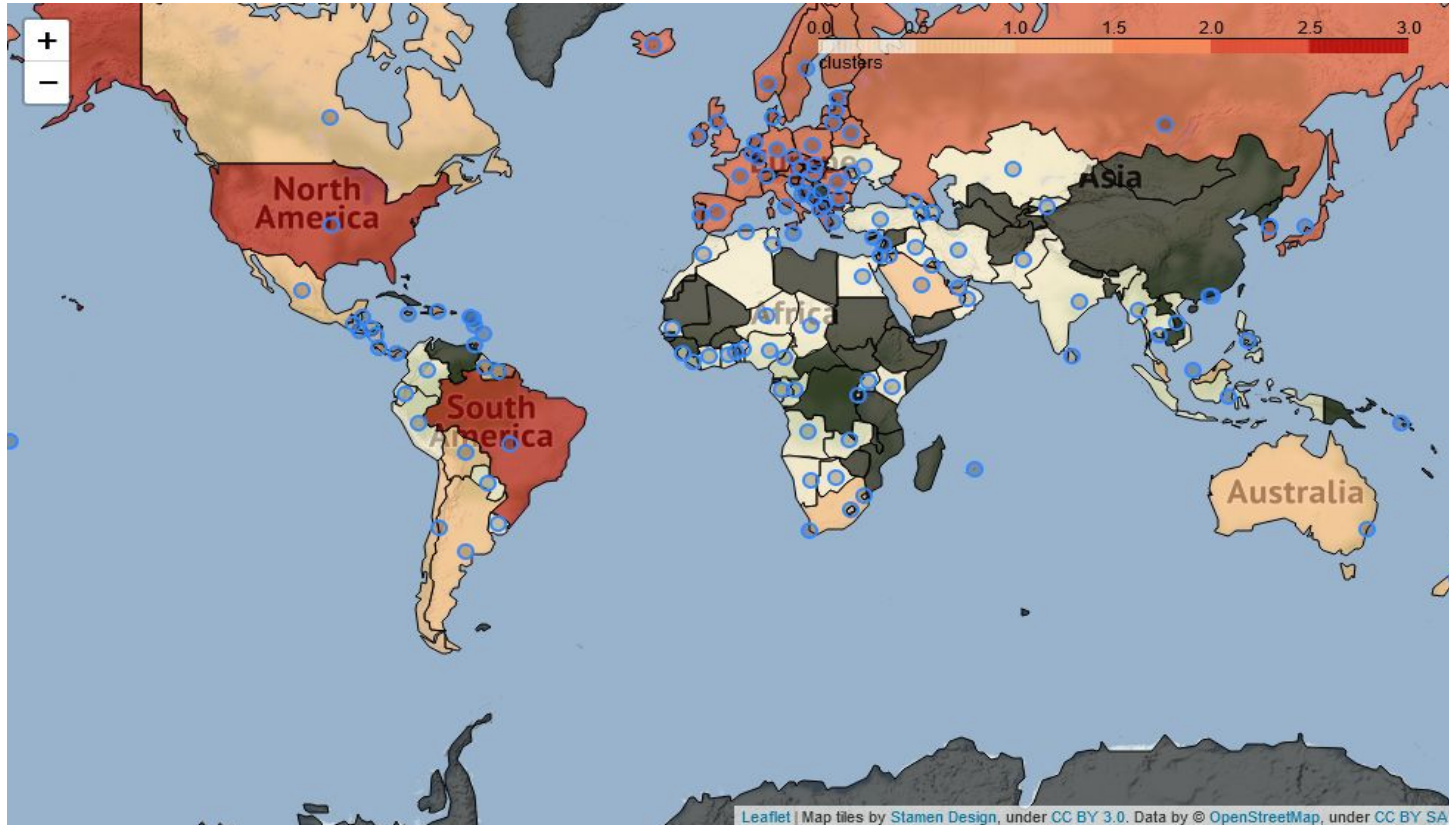
Ici il serait de 4

NUAGE DE POINTS 3D CLUSTERS/CENTROÏDES

clusters avec centroides



VISUALISATION DES CLUSTERS VIA UNE MAPMONDE



Les pays en noirs sont ceux où je n'ai pas assez de données.

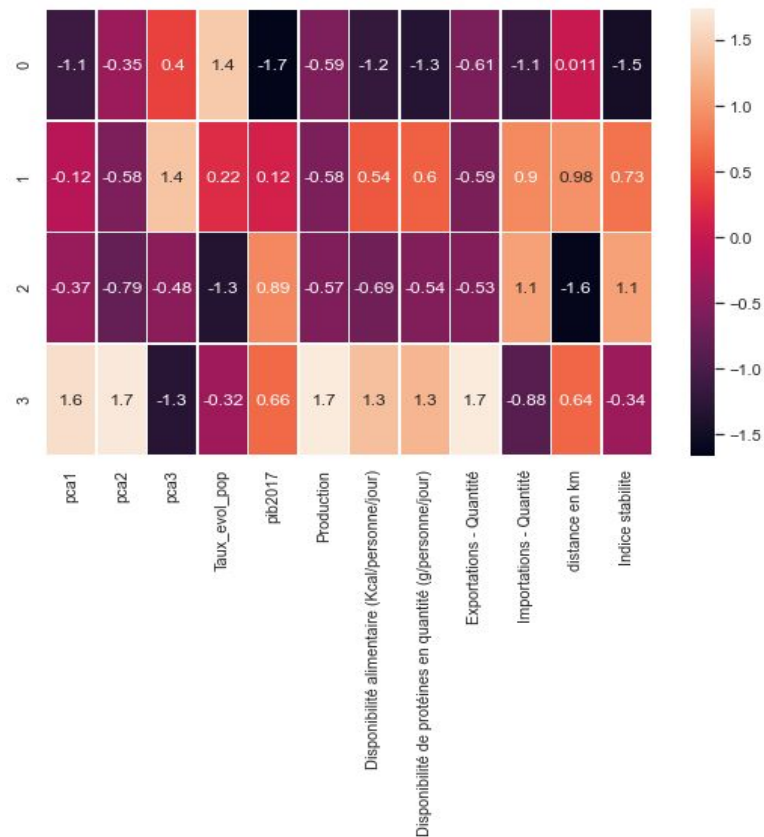
Ils sont alors écartés de l'analyse

VISUALISATION DES CENTROÏDES

	pca1	pca2	pca3	Taux_evol_pop	pib2017	Production	Disponibilité alimentaire (Kcal/personne /jour)	Disponibilité de protéines en quantité (g/personne /jour)	Exportations - Quantité	Importations - Quantité	distance en km	Indice stabilité
Clusters												
0	-1.486573	0.561470	0.105447	6.997586	9230.084483	457.741379	44.482759	4.247069	28.206897	51.482759	6308.586207	-0.542069
1	1.492880	-0.267847	1.442229	4.158276	30767.337931	528.172414	159.103448	14.329310	50.206897	149.275862	9004.000000	0.433448
2	0.737752	-1.014753	-1.088499	0.476750	40055.890000	575.800000	77.800000	8.375500	156.350000	158.150000	1740.700000	0.592500
3	6.708817	7.896210	-2.200302	2.890000	37317.150000	18057.500000	211.500000	17.805000	3957.500000	63.000000	8061.000000	-0.040000

Nous pouvons constater que le groupe 2 est le groupe qui a le + fort PIB. La distance de celui ci est le + proche de la France ce qui nous fait penser à un 'groupe' Europe. Ce groupe serait en cohérence avec le résultat de l'acp

HEATMAP



Nous pouvons constater une forte corrélation sur l'indice de stabilité, l'importation, le pib du groupe 2, (groupe qui nous intéresse).

MAIS ALORS OÙ FAUT IL INVESTIR ??

Zone	Taux_evol_pop	pib2017	Production	Disponibilité alimentaire (Kcal/personne /jour)	Disponibilité de protéines en quantité (g/personne /jour)	Exportations - Quantité	Importations - Quantité	Continent	distance en km	Indice stabilite
Luxembourg	8.23	112308.0	0.0	69.0	7.19	1.0	11.0	Europe	294	1.33
Irlande	3.98	78655.6	110.0	100.0	11.60	93.0	99.0	Europe	850	1.00
Suisse	3.75	69103.6	91.0	55.0	6.49	4.0	51.0	Europe	470	1.26
Norvège	3.67	64050.8	101.0	66.0	7.81	0.0	2.0	Europe	1537	1.17
Autriche	3.11	54170.0	148.0	65.0	7.49	78.0	110.0	Europe	832	1.05
Suède	2.80	51948.0	157.0	61.0	7.09	23.0	84.0	Europe	1658	0.98
Islande	2.41	55638.5	10.0	104.0	12.33	0.0	2.0	Europe	2157	1.35
Allemagne	2.01	52952.9	1514.0	71.0	7.96	646.0	842.0	Europe	694	0.59
Danemark	1.53	55356.7	173.0	112.0	12.97	139.0	133.0	Europe	947	0.87
Pays-bas	0.98	55088.6	1100.0	70.0	8.48	1418.0	608.0	Europe	447	0.92



Le Luxembourg, l'Irlande et la suisse me paraissent de bon choix concernant l'exportation de nos poulets, avec un pib assez fort, une évolution de population et un indice de stabilité politique satisfaisant.