



# La poule qui chante

**Projet-9 : Produire une étude de marché avec R**



# Introduction

La société, « la Poule qui chante », entreprise française d'agroalimentaire, souhaite se développer à l'international.

🐔 Problématique : Proposer les pays les plus favorables à l'export de Poulet français vivants ou entier.

🐔 Plan

1. Les données et les variables
2. Le clustering pour réduire le nombre de pays favorables
3. L'ACP pour restreindre aux pays les plus favorables

🐔 Conclusion et recommandations





# Plan

## 1. Les données et les variables

1. Les données sélectionnées
2. Les données supprimées
3. Nettoyage des données
4. Sélection des données

- [notebook R\Moty Philippe 1.1 notebook nettoyageDonnees 0622.Rmd](#)
- [notebook R\Moty Philippe 1.1 notebook nettoyageDonnees 0622.nb.html](#)

## 1. Les clustering pour réduire le nombre de pays favorables

1. Classification Ascendante Hiérarchique
  2. K-Means
- ## 2. L'ACP pour restreindre aux pays les plus favorables



## Conclusion et recommandations





# 1. Les données et les variables sélectionnées

- **Analyse PESTEL**

- **Politique** : stabilité politique
- **Economie** : PIB, consommation volailles
- **Social** : Population
- **Technologie** : distance à la France
- **Environnement** : Business score, Score Starting a business
- **Légal** : législation et crise Grippe aviaire

- 13 fichiers csv compilés et nettoyés :

- [notebook R\Moty Philippe 1.1 notebook nettoyageDonnees 0622.Rmd](#)
- [notebook R\Moty Philippe 1.1 notebook nettoyageDonnees 0622.nb.html](#)

En raison de l'années définie, 2017, une législation particulière s'applique.

Les 8 variables quantitatives retenues : Production, Importation, PIB, Stabilité politique, inflation moyenne, population en 2017, Distance de la France, Easy of doing business score





# Les données et les variables supprimées

- Pays où l'import de volailles françaises est interdit
  - [http://classe-export.com/wp-content/uploads/2019/03/Etatdeslieux\\_filiaireaviaire.pdf](http://classe-export.com/wp-content/uploads/2019/03/Etatdeslieux_filiaireaviaire.pdf)

Bangladesh  
Belize  
Burkina Faso  
Cameroun  
Équateur  
Guyana  
Inde  
Israël  
Kenya  
Madagascar

Malawi  
Népal  
Nigéria  
Ouganda  
République populaire démocratique de Corée  
Rwanda  
Sénégal  
Sri Lanka  
Tunisie





## Plan

1. Les données et les variables
    1. Les données sélectionnées
    2. Les données supprimées
    3. Nettoyage des données
    4. Sélection des données
  2. Les clustering pour réduire le nombre de pays favorables
    1. Classification Ascendante Hiérarchique
    2. K-Means
  3. L'ACP pour restreindre aux pays les plus favorables
- [notebook R\Moty Philippe 1.2 notebook Analyse 0622 1.nb.html](#)
- [notebook R\Moty Philippe 1.2 notebook Analyse 0622 1.Rmd](#)

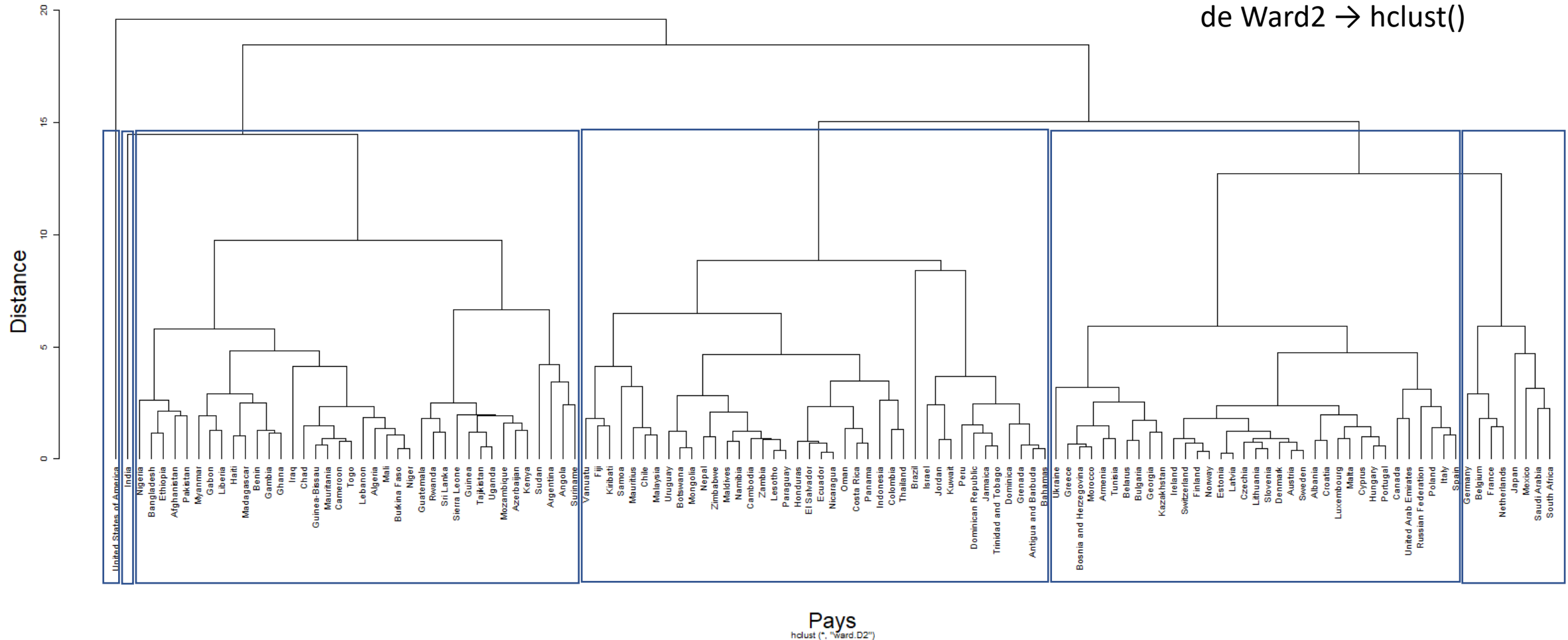
## Conclusion et recommandations



# Clustering CAH - Dendrogramme

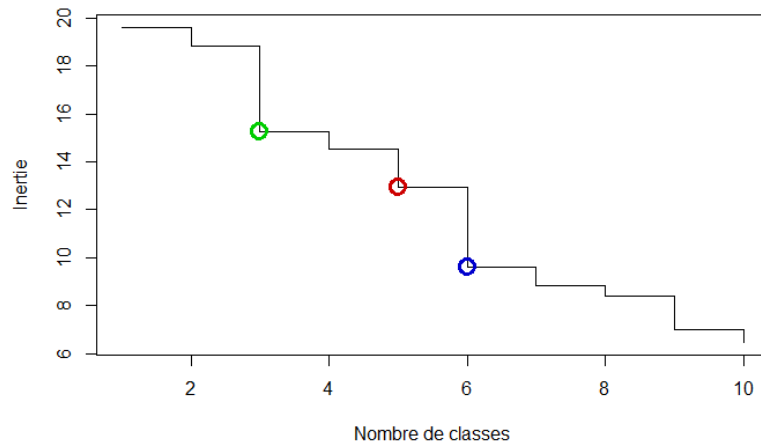
## Classification ascendante hiérarchique

- 9 variables quantitatives
  - Normalisation
  - Calcul des distances euclidiennes entre les pays
  - regroupements méthode de Ward2 → `hclust()`



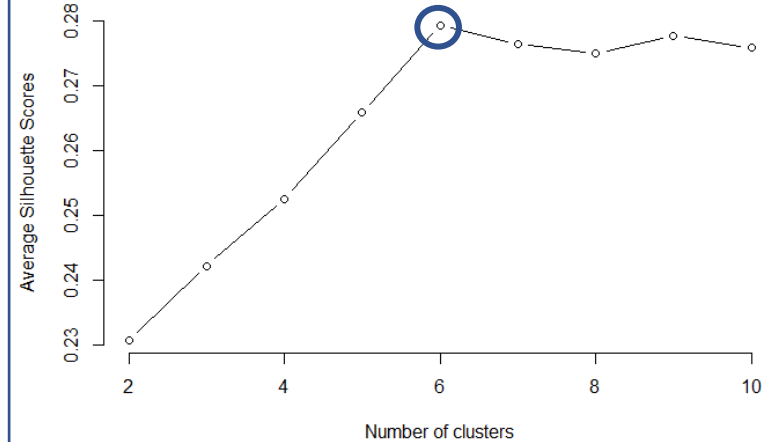


# Bien choisir le nombre de clusters : 6

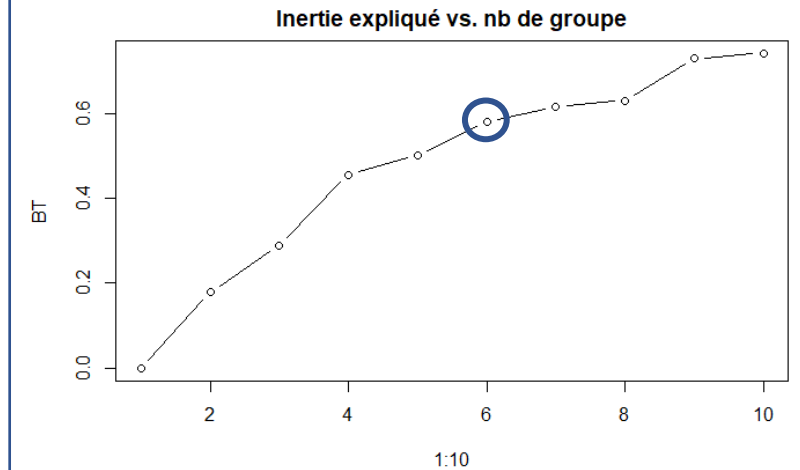


Pour obtenir une partition de la population, nous pouvons représenter les sauts d'inertie du dendrogramme selon le nombre de classes retenues.

On voit 3 saut assez nets à 3,5,6 représenté en vert, rouge, bleu



Utiliser le coefficient de Silhouette qui montre que le meilleur nombre de cluster est 6

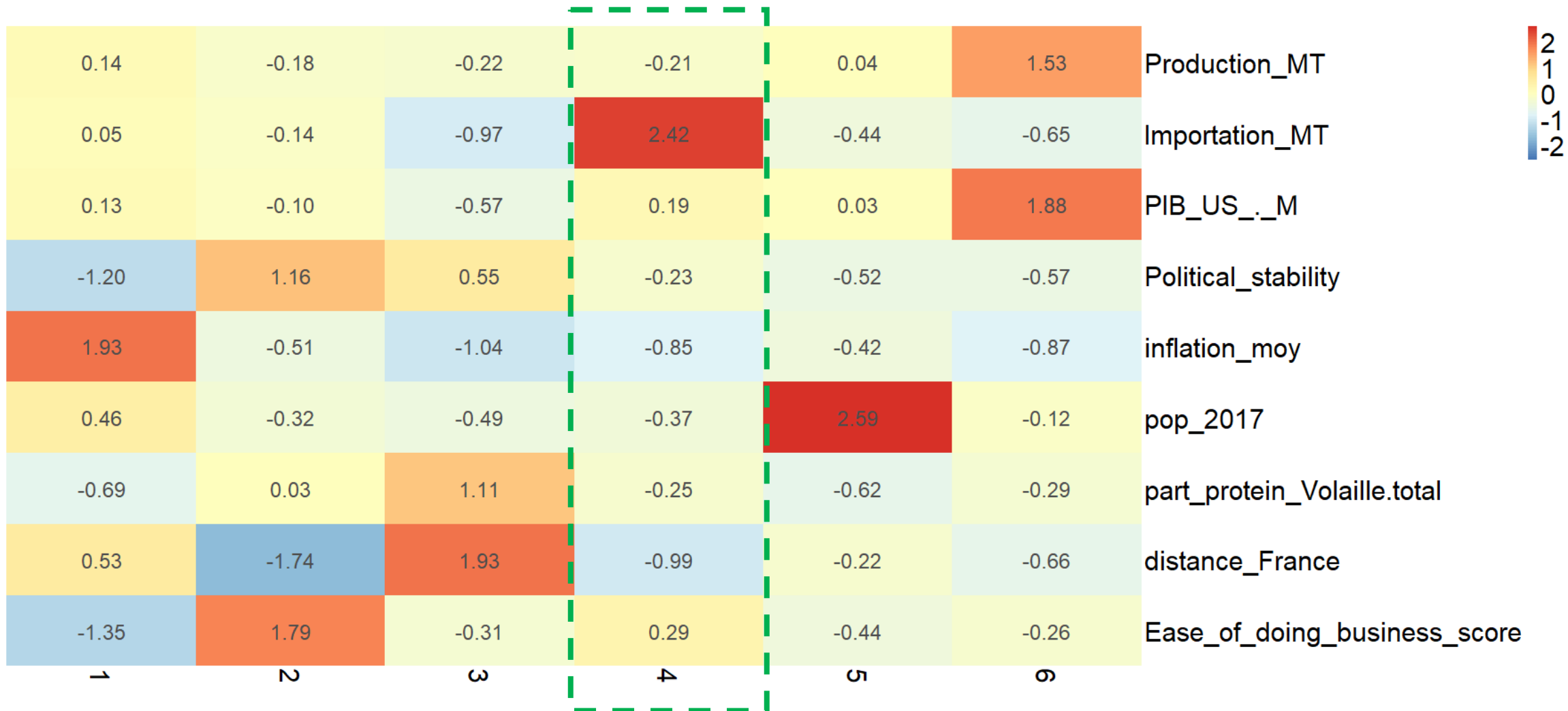


Choisir un nombre de cluster ou groupe qui explique le maximum les données. Plus de la moitié de la variances des données s'explique avec 6 clusters ce qui confirme le choix de 6 clusters

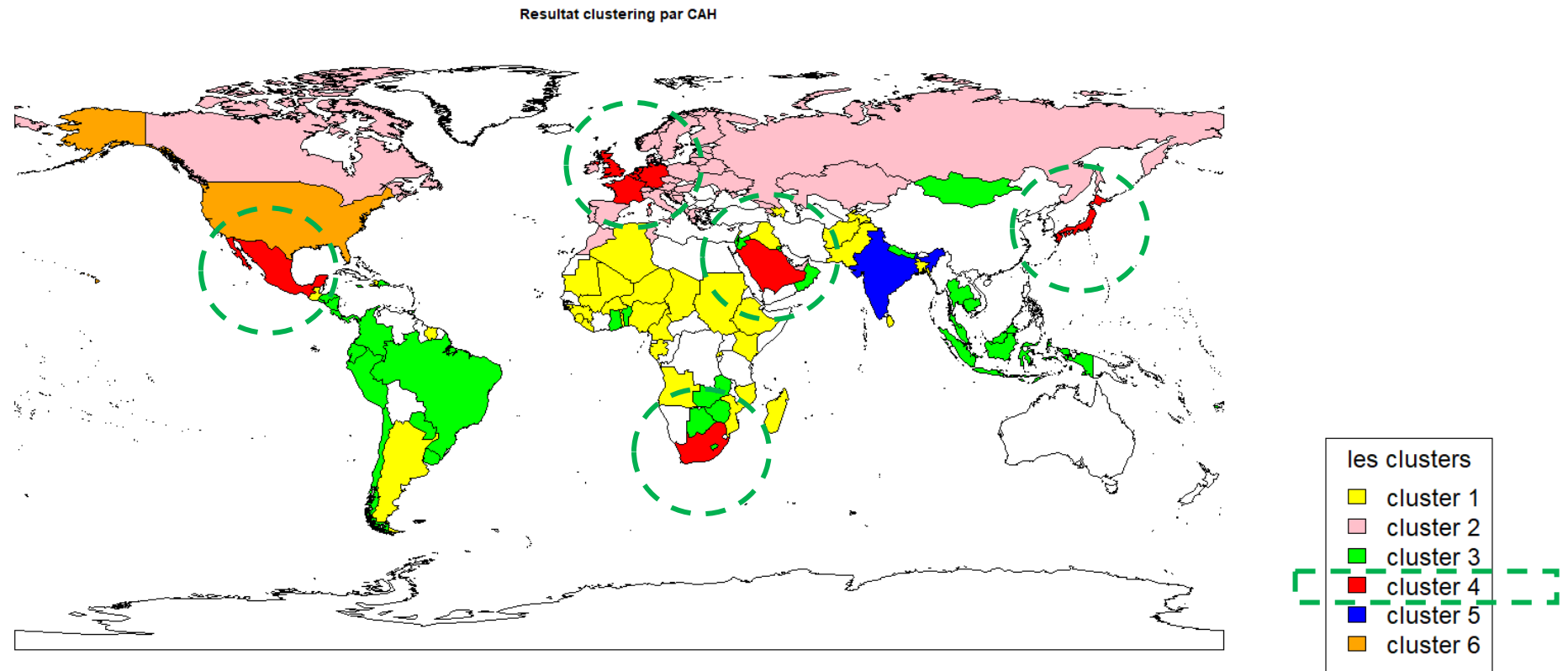


# Clustering Ascendant Hiérarchique – centroïdes

Centroïdes = moyenne pour chaque cluster et chaque variable, des valeurs centré réduites



# Evaluer la cohérence de la répartition



Les clusters ne sont pas géographiquement regroupés et les pays qui les composent sont très différents.

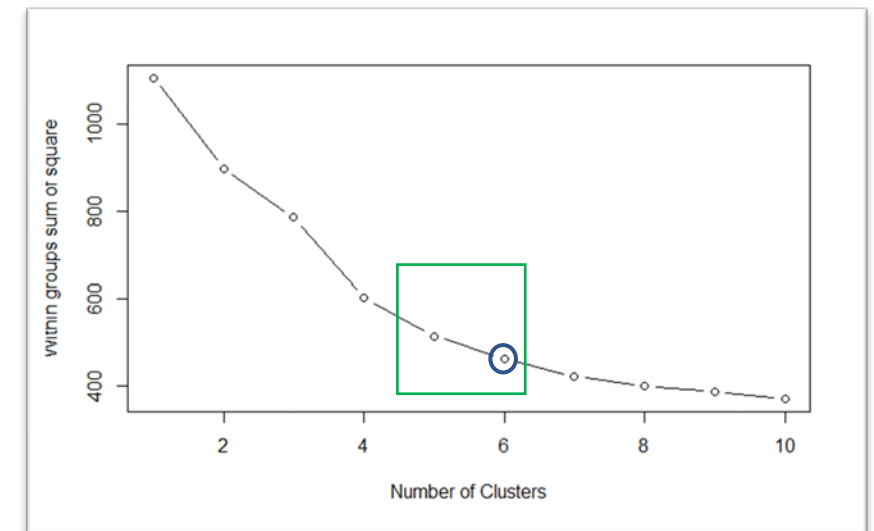
La méthode du CAH n'est donc pas la plus pertinente, nous optons donc pour la méthode des K-Means.



# K-MEANS

- On détermine le k du K-Means
  - Le coude nous indique le bon nombre k pour le kmeans
  - On en déduit que c'est entre 5 et 6 mais 5 n'apporte rien car il englobe Inde et USA dans le même cluster
  - Après analyse des clusters on choisit un partitionnement en 6 clusters qui permet de faire des groupes plus pertinents pour répondre à notre problématique.

Méthode du coude



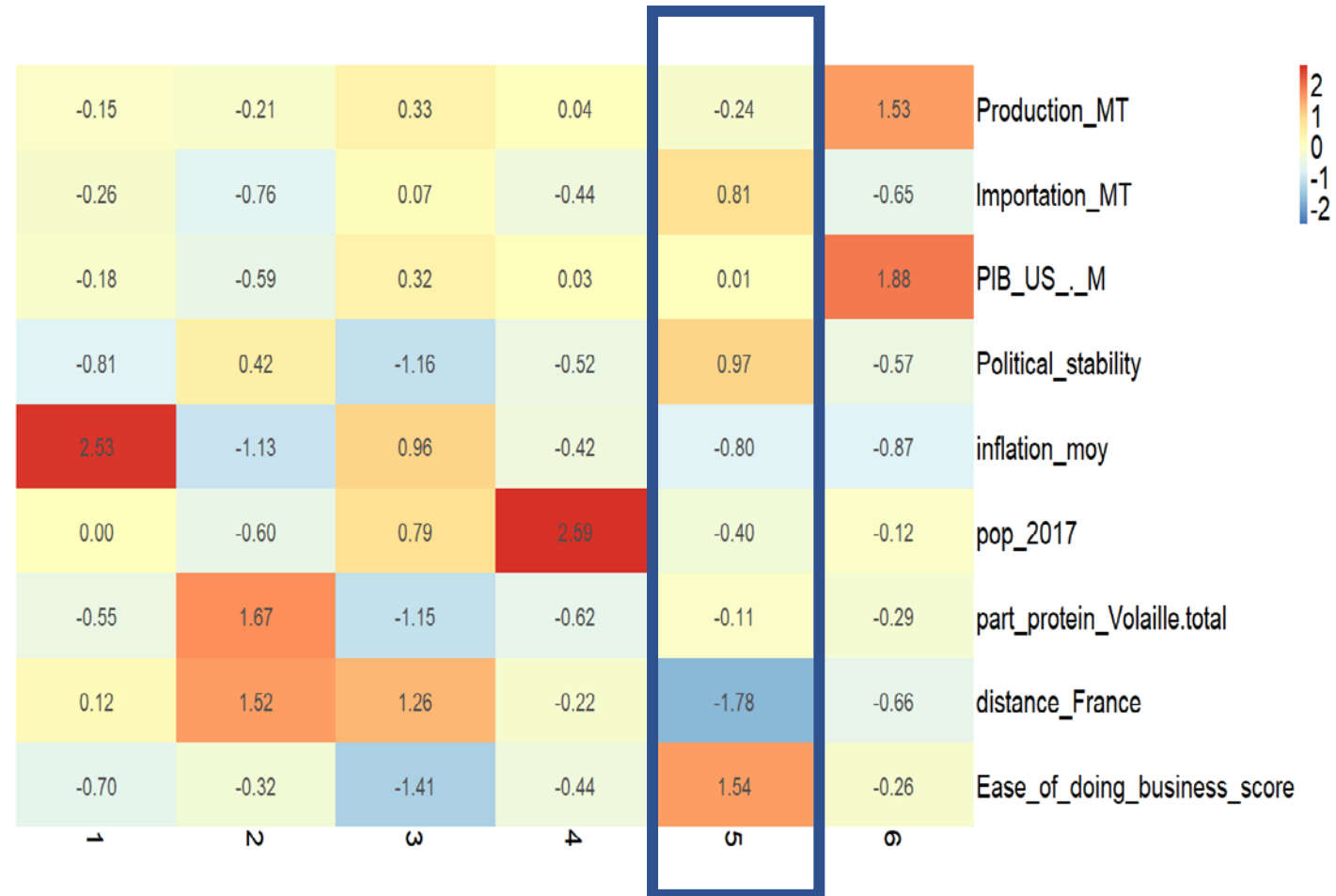
# Le cluster 5 est le plus pertinent

## Cluster 5 : Les points forts

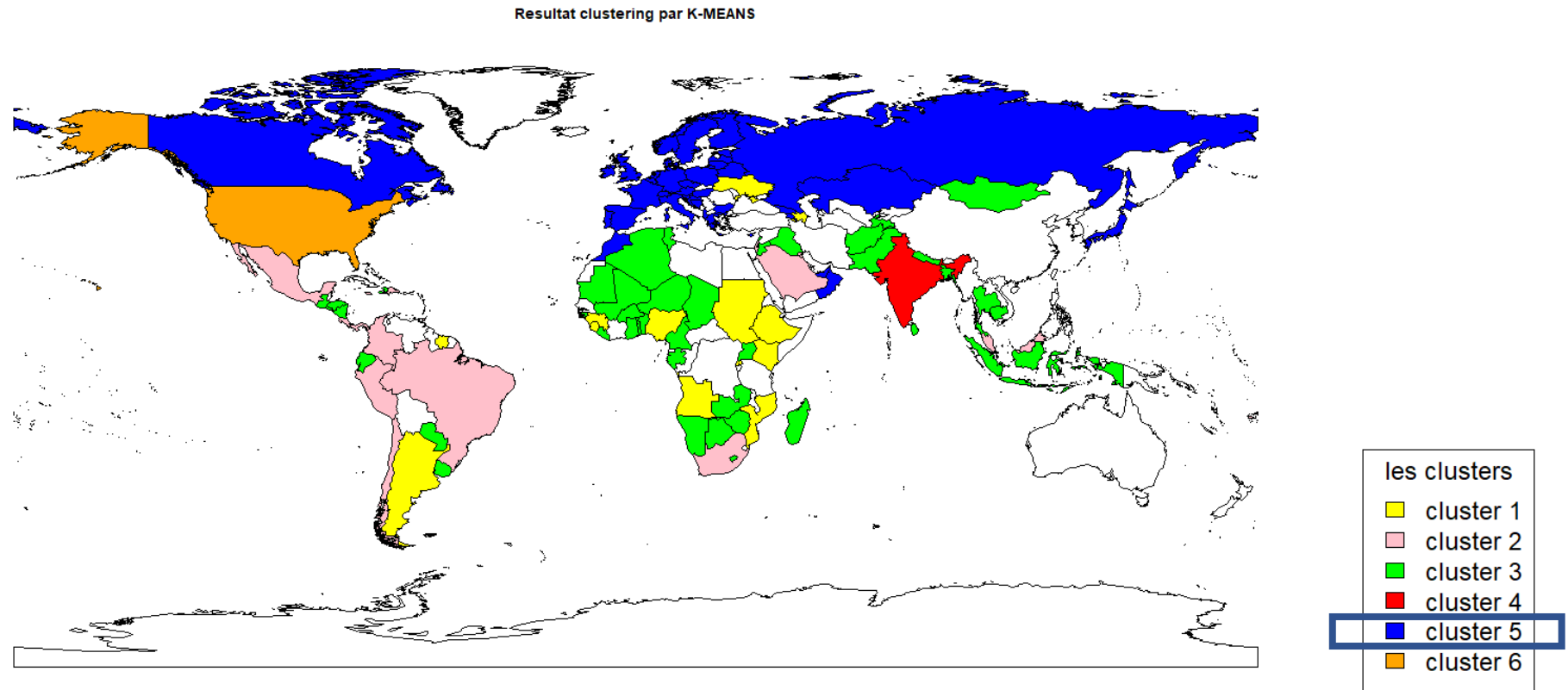
- Importation
- PIB faible mais positif
- La stabilité politique : meilleur score
- Faible inflation
- Consommation de Volailles
- Distance : proche
- Score Easy Doing Business : meilleur score

La carte (slide suivante) confirme la cohérence géographique de se choix, renforcé par l'appartenance des pays du cluster à des union monétaire (UE), économique, UE et autre.

La distance et les moyens de transports sont des variables importantes dans l'export de produits vivants ou frais



# Pertinence du cluster 5 - MAP



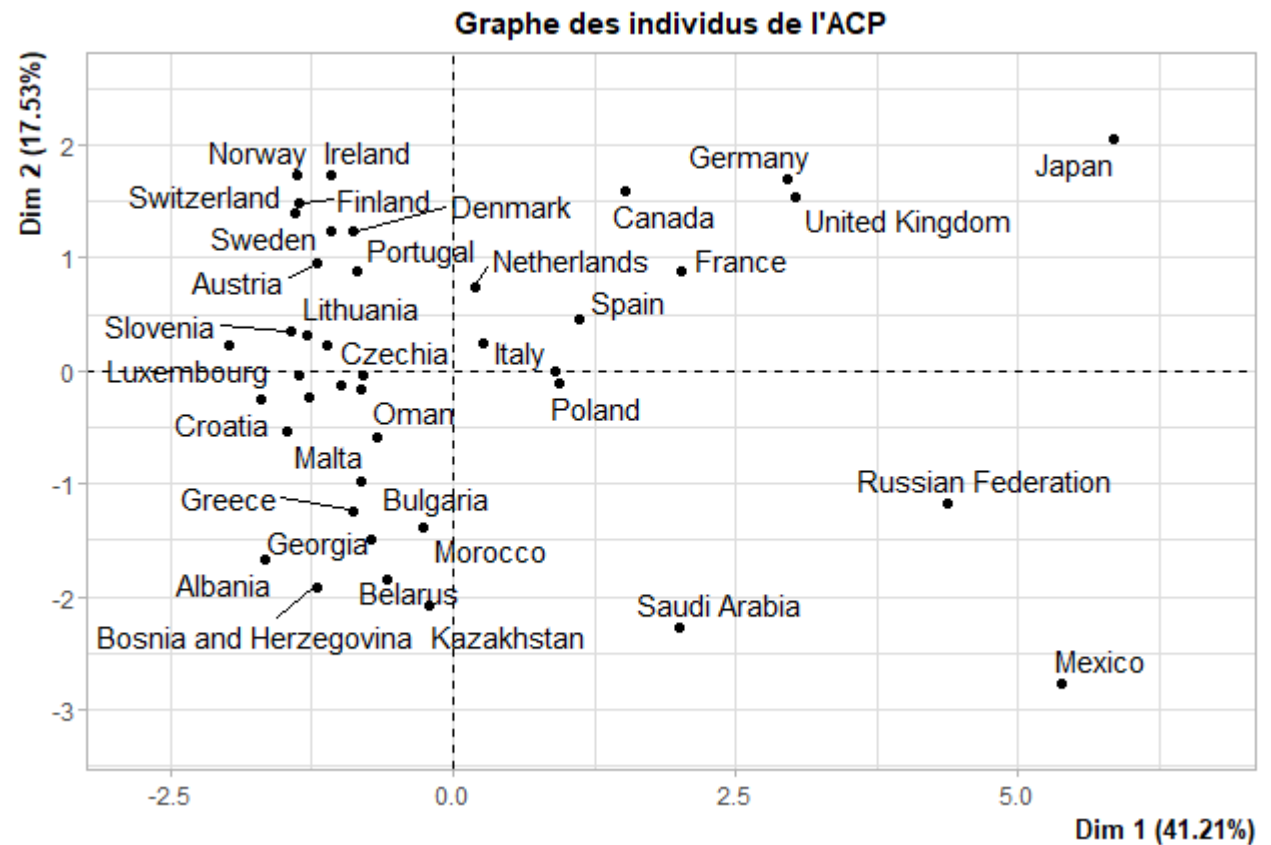
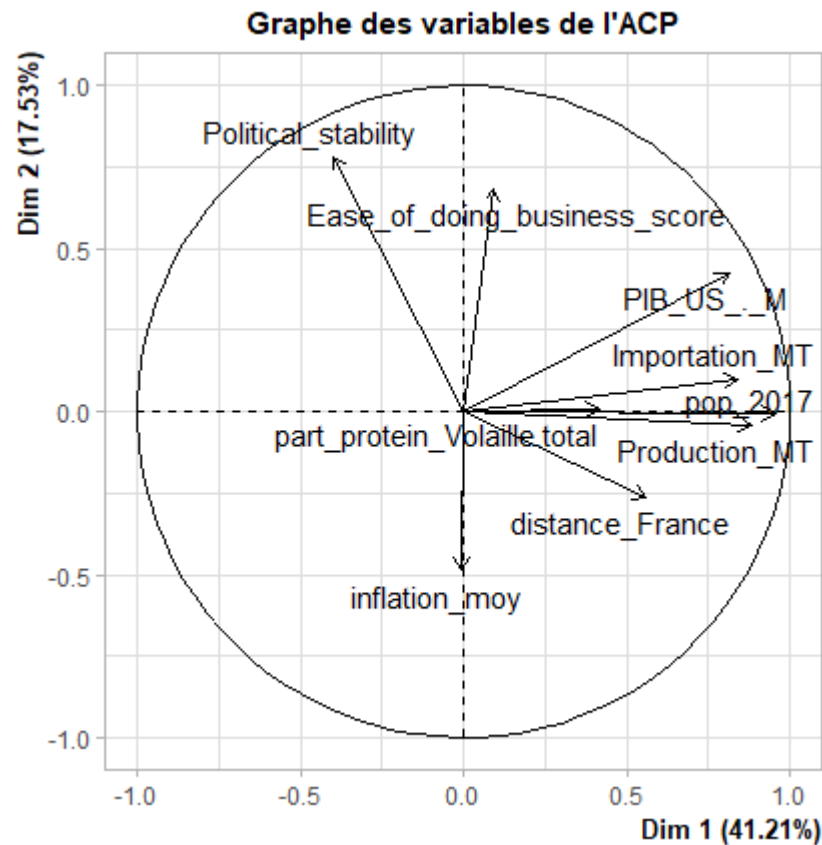
## Plan

1. Les données et les variables
  1. Les données sélectionnées
  2. Les données supprimées
  3. Nettoyage des données
  4. Sélection des données
2. Les clustering pour réduire le nombre de pays favorables
  1. Classification Ascendante Hiérarchique
  2. K-Means
3. L'ACP sur le cluster extrait pour restreindre aux pays les plus favorables

## Conclusion et recommandations

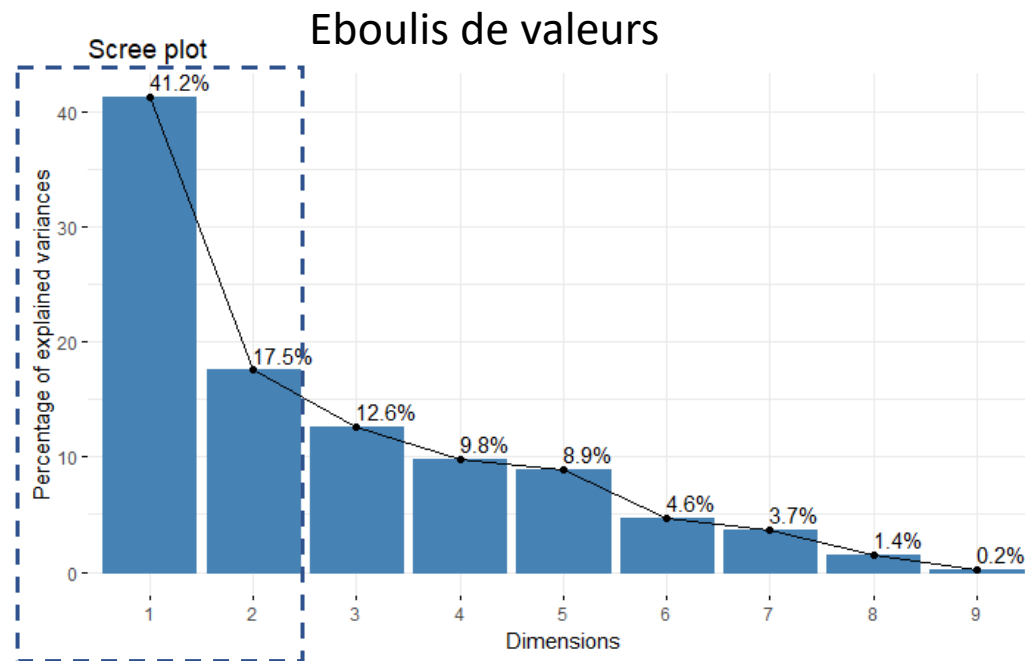


# Analyse en composante Principale (ACP)



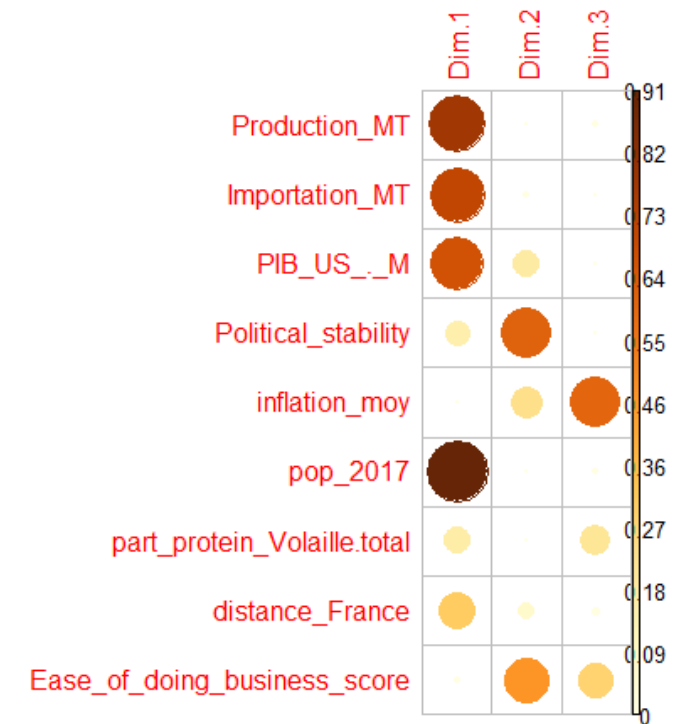


# Evaluation de la pertinences des indicateurs

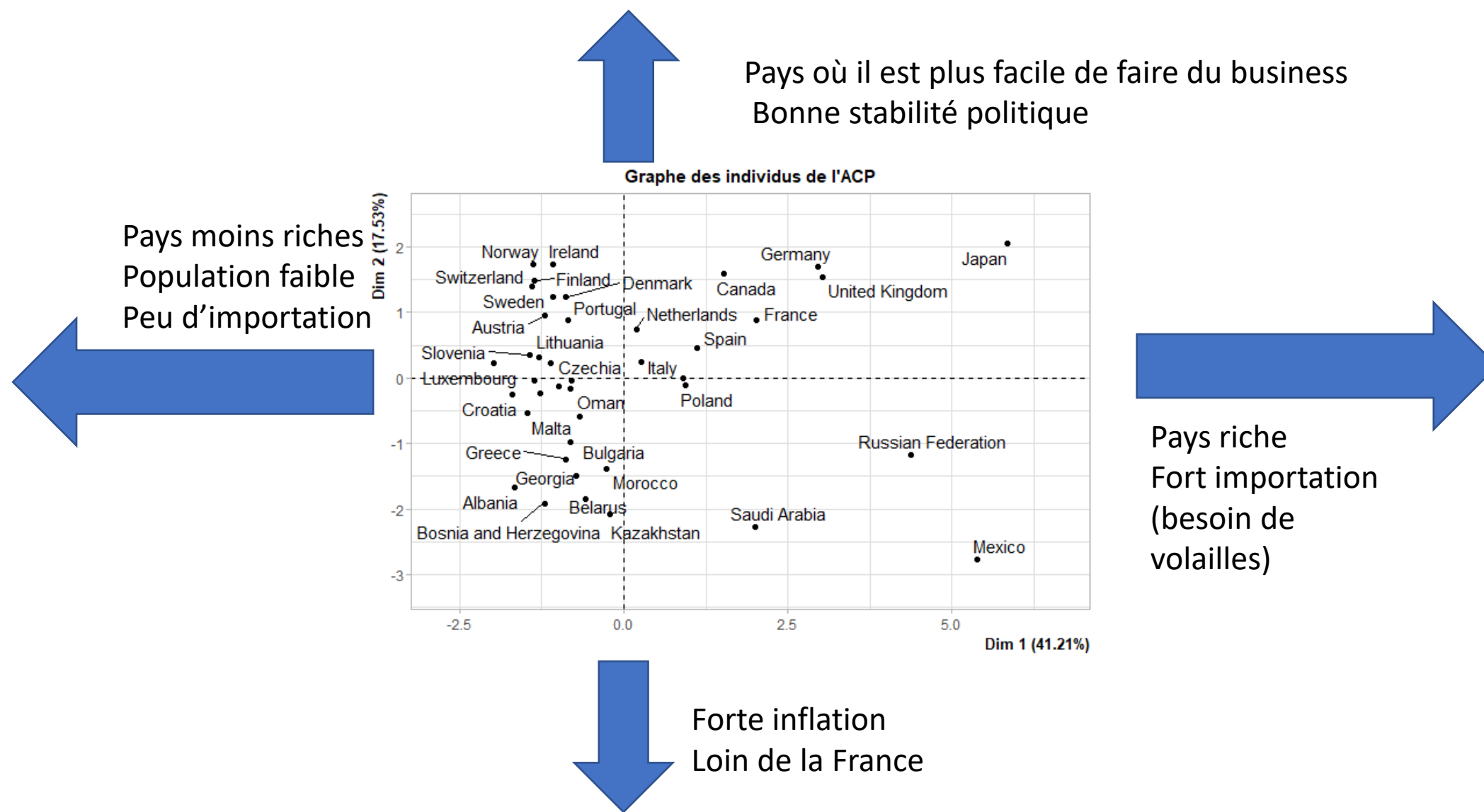


Les dimensions 1 et 2 expliquent 58,74% des données

Pertinences des dimensions



# Graphe des individus de l'ACP





# Retour aux données et sélection

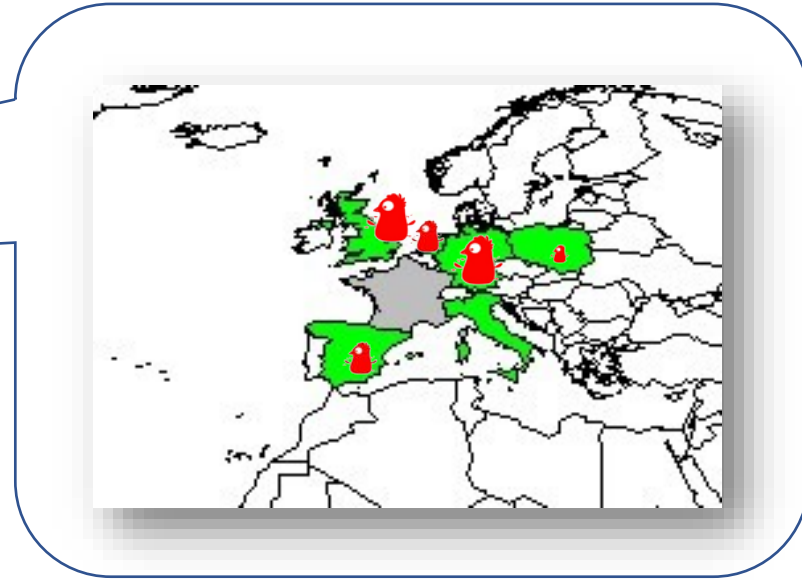
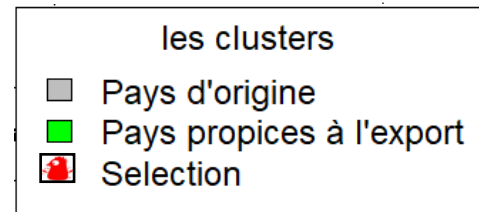
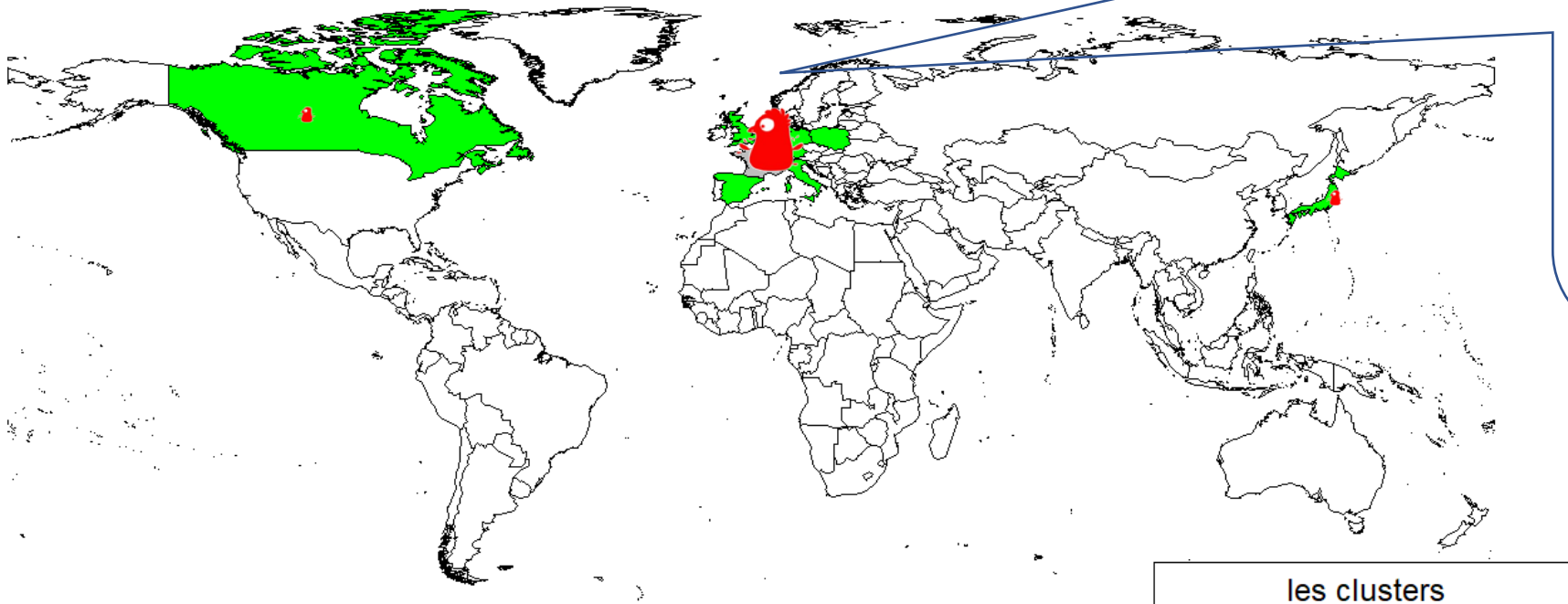
Country	Production MT	Importation MT	PIB US \$ M	Political stability	inflation moyenne	population en 2017	part de la consommation de protein de Volaille/total %	distance France	Ease of doing business_score (%)
France	1750	506	2595151.045	0.28	1.088588333	64 842 509	7.93	278.1876	76.20155734
<hr/>									
United Kingdom	1814	779	2699016.715	0.33	2.261151417	66 727 461	13.4	342.9475	83.34036985
Germany	1514	842	3690849.153	0.59	2.762496833	82 658 409	7.1	439.8984	79.42867278
Netherlands	1100	608	833869.6417	0.92	2.6643255	17 021 347	6.58	427.9169	75.29031801
Spain	1515	205	1312539.279	0.28	1.231401667	46 647 428	11.61	1054.656	77.63075376
Poland	2351	55	526504.2336	0.52	4.314900333	37 953 180	13.34	1368.178	77.64919003
Japan	2215	1069	4930837.369	1.11	0.698263333	127 502 725	8.34	9725.629	78.27679314
Canada	1417	182	1649518.967	1.1	-0.975202333	36 732 095	13.21	6004.645	80.04017419
Italy	1315	97	1961796.197	0.31	1.880168583	60 673 701	7.47	1109.901	71.82262903





# Conclusion

Les pays à fort potentiel



Liste des pays pertinents :

- Allemagne
- Royaume-Uni
- Espagne
- Puis : Pologne
- Et enfin : Canada, Japon



# Les sources

 <http://www.cepii.fr/%5C/francgraph/bdd/distances.htm>

 <https://www.fao.org/faostat/en/#data/FS>

 <https://sql.sh/514-liste-pays-csv-xml>

 <https://archive.doingbusiness.org/en/data>

 <http://www.agridata.tn/dataset/>

