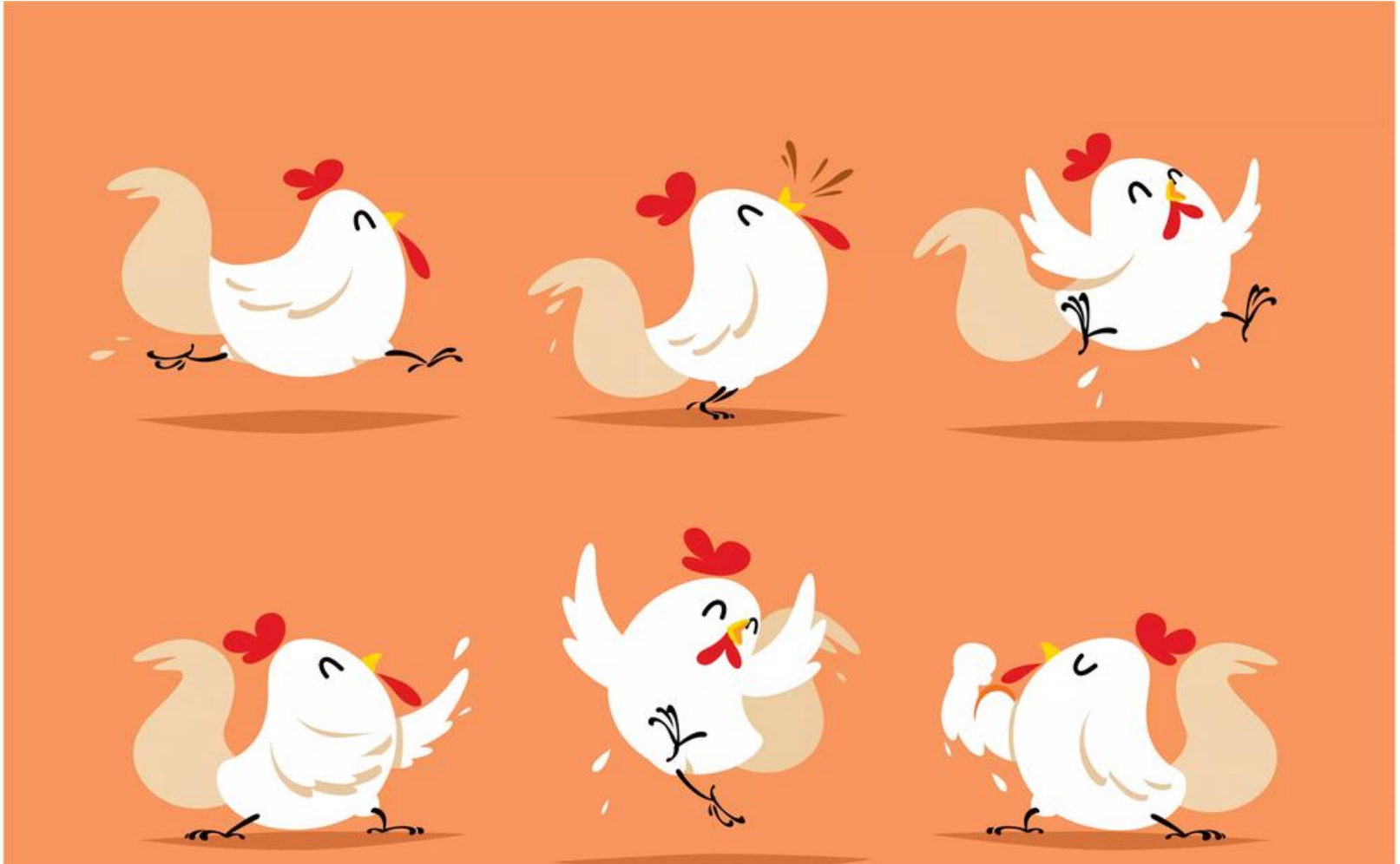


# Produisez une étude de marché



Pinos Bruno

# SOMMAIRE

- Exucative summary
- Source des données
- Classification
- Analyse
- Prise de décision
- Vérification
- Bilan

# Exucative Summary

L'entreprise souhaite se développer à l'international.

Mais quels pays sont les plus propices pour exporter nos poulets ?

Insight et recommandation:

Les pays occidentaux ont un régime alimentaire très riche en calories, en protéine et un Pib très élevé.

- Malgré une augmentation faible de leur population il semble que se soit le groupe à cibler.

Certain pays de ce groupe se détachent particulièrement au niveau du PIB et de la démographie.

- *Il faut cibler en priorité ces pays là pour exporter nos produits.*

# Source des données

*Objectif:*

*Récupérer les données qui nous intéressent et les regrouper le plus clairement possible*

# Importation et traitement des données

```
pop_2008<-read.csv("pop_2008.csv", encoding = "UTF-8")%>%
  select("Zone", "pop2008" = "Valeur")

pop_2018<-read.csv("pop_2018.csv", encoding = "UTF-8")%>%
  select("Zone", "pop2018" = "Valeur")

prot_anh_hab<-read.csv("prot_anh_habitant.csv", encoding = "UTF-8")%>%
  select("Zone", "prot_anh" = "Valeur")

prot_hab<-read.csv("prot_habitant.csv", encoding = "UTF-8")%>%
  select("Zone", "prot" = "Valeur")

kcal_hab<-read.csv("kcal_habitant.csv", encoding = "UTF-8")%>%
  select("Zone", "kcal" = "Valeur")

pib_hab<-read.csv("pib_habitant.csv", encoding = "UTF-8")%>%
  select("Zone", "pib" = "Valeur")%>%group_by(Zone)%>%
  summarise_if(is.numeric, round)
```

Zone	pop2008	Zone	pop2018	Zone	prot_anh	Zone	kcal
Afghanistan	27722.276	Afghanistan	37171.921	Afghanistan	10.55	Afghanistan	2000
Afrique du Sud	49779.471	Afrique du Sud	57792.518	Afrique du Sud	34.56	Afrique du Sud	2994
Albanie	3002.678	Albanie	2882.740	Albanie	66.34	Albanie	3400
Algérie	34730.608	Algérie	42228.408	Algérie	25.71	Algérie	3349
Allemagne	81065.752	Allemagne	83124.418	Allemagne	63.08	Allemagne	3556
Andorre	83.862	Andorre	77.006	Angola	16.48	Angola	2270

# Construction du dataframe principal

```
df<-left_join(prot_ani_hab, pop_2008)%>%  
  left_join(pop_2018)%>%left_join( prot_hab)%>%  
  left_join(kcal_hab)%>%left_join(pib_hab)
```

Zone	prot_ani	pop2008	pop2018	prot	kcal	pib
Afghanistan	10.55	27722.276	37171.92	54.09	2000	564
Afrique du Sud	34.56	49779.471	57792.52	83.41	2994	6413
Albanie	66.34	3002.678	2882.74	119.55	3400	5132
Algérie	25.71	34730.608	42228.41	92.91	3349	4136
Allemagne	63.08	81065.752	83124.42	104.20	3556	47993
Angola	16.48	21695.634	30809.79	54.11	2270	3441

```
sapply(df,function(x) sum(is.na(x)))
```

```
na_table<-filter(df, is.na(pop_diff)|is.na(pib))
```

```
df["150","pop_2008"]<- 33060
```

```
df["35","pib"]<-24971
```

# Construction du dataframe principal

```
df<-df%>%  
  mutate(pop_diff = round((pop2018 -pop2008)*100/pop2008,1),  
         prot_anl_prct = round(prot_anl*100/prot,0))%>%  
  select(-"prot_anl", -"pop2018", -"pop2008")
```

Zone	prot	kcal	pib	pop_diff	prot_anl_prct
Afghanistan	54.09	2000	564	34.1	20
Afrique du Sud	83.41	2994	6413	16.1	41
Albanie	119.55	3400	5132	-4.0	55
Algérie	92.91	3349	4136	21.6	28
Allemagne	104.20	3556	47993	2.5	61
Angola	54.11	2270	3441	42.0	30

# Classification

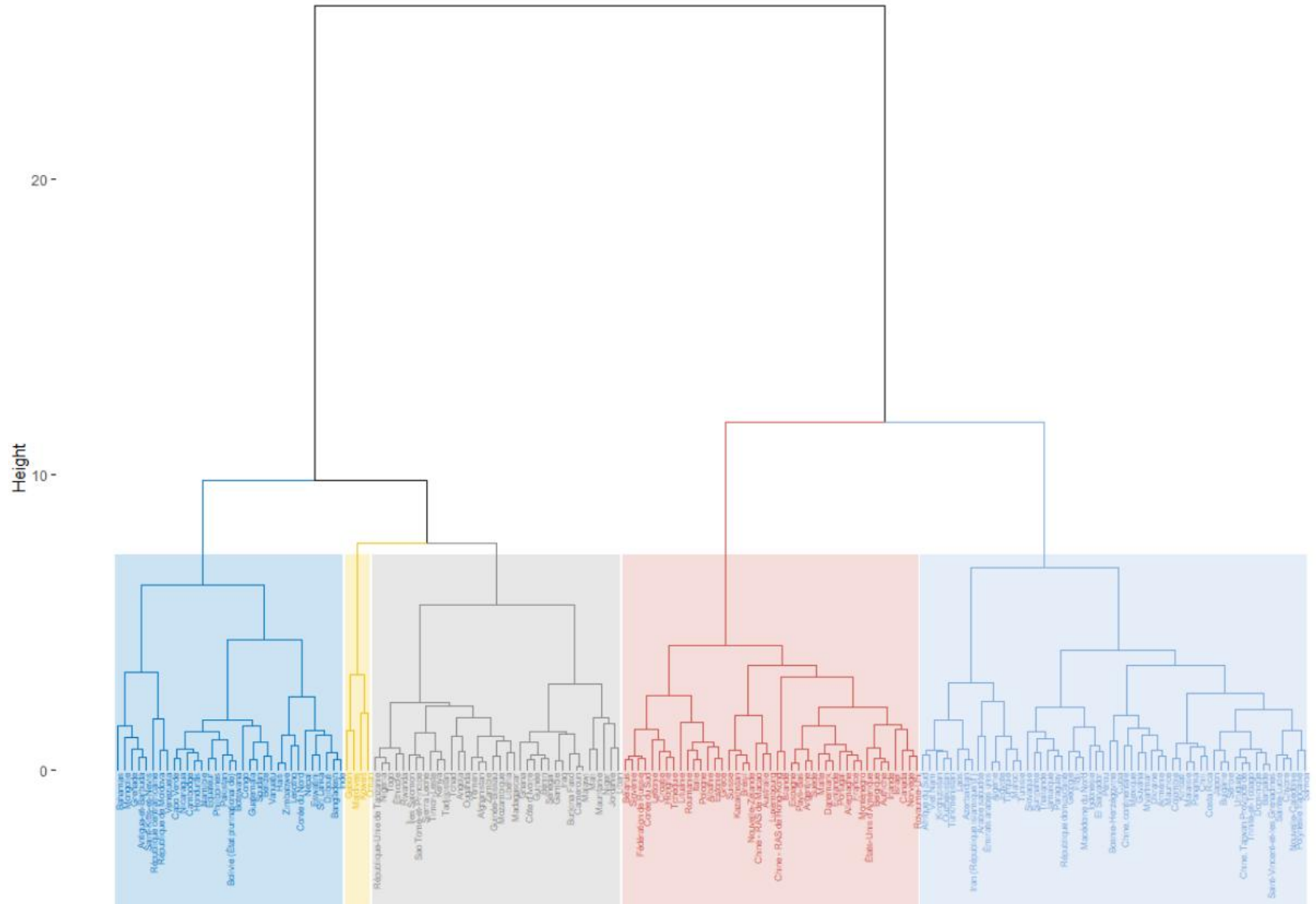
*Objectif:*

*Faire 5 groupes les plus cohérents  
possibles selon les critères choisis*



# CAH

### Cluster Dendrogram



# Cluster 1

Afghanistan
Angola
Bénin
Burkina Faso
Cameroun
Côte d'Ivoire
Éthiopie
Gambie
Ghana
Guinée
Guinée-Bissau
Îles Salomon
Iraq
Jordanie
Kenya
Liban
Libéria

Madagascar
Malawi
Mali
Mauritanie
Mozambique
Niger
Nigéria
Ouganda
République-Unie de Tanzanie
Rwanda
Sao Tomé-et-Principe
Sénégal
Sierra Leone
Tadjikistan
Tchad
Timor-Leste
Togo
Yémen
Zambie

# Cluster 2

Afrique du Sud
Algérie
Arabie saoudite
Arménie
Azerbaïdjan
Barbade
Bosnie-Herzégovine
Bulgarie
Chili
Chine, continentale
Chine, Taiwan Province de
Chypre
Colombie
Costa Rica
Cuba
Dominique
Égypte
El Salvador
Émirats arabes unis
Fidji
Géorgie

Géorgie
Guyana
Indonésie
Iran (République islamique d')
Jamaïque
Japon
Kirghizistan
Kiribati
Macédoine du Nord
Malaisie
Maroc
Maurice
Mexique
Myanmar
Nouvelle-Calédonie
Ouzbékistan
Panama
Paraguay
Pérou

Polynésie française
Laos
République dominicaine
Sainte-Lucie
Saint-Vincent-et-les Grenadines
Samoa
Serbie
Slovaquie
Suriname
Thaïlande
Trinité-et-Tobago
Tunisie
Turkménistan
Turquie
Ukraine
Uruguay
Viet Nam

# Cluster 3

Albanie
Allemagne
Argentine
Australie
Autriche
Bélarus
Belgique
Brésil
Canada
Chine - RAS de Hong-Kong
Chine - RAS de Macao
Croatie
Danemark
Espagne
Estonie
États-Unis d'Amérique
Fédération de Russie
Finlande
France
Grèce

Hongrie
Irlande
Islande
Israël
Italie
Kazakhstan
Lettonie
Lituanie
Luxembourg
Malte
Monténégro
Norvège
Nouvelle-Zélande
Pays-Bas
Pologne
Portugal
Corée du Sud
Roumanie
Royaume-Uni
Slovénie
Suède
Suisse
Tchéquie

# Cluster 4

Antigua-et-Barbuda
Bahamas
Bangladesh
Belize
Bolivie (État plurinational de)
Botswana
Cabo Verde
Cambodge
Congo
Djibouti
Équateur
Eswatini
Grenade
Guatemala
Haïti
Honduras

Inde
Lesotho
Mongolie
Namibie
Népal
Nicaragua
Pakistan
Philippines
République centrafricaine
République de Moldova
Corée du Nord
Saint-Kitts-et-Nevis
Soudan
Sri Lanka
Vanuatu
Venezuela
Zimbabwe

# Cluster 5

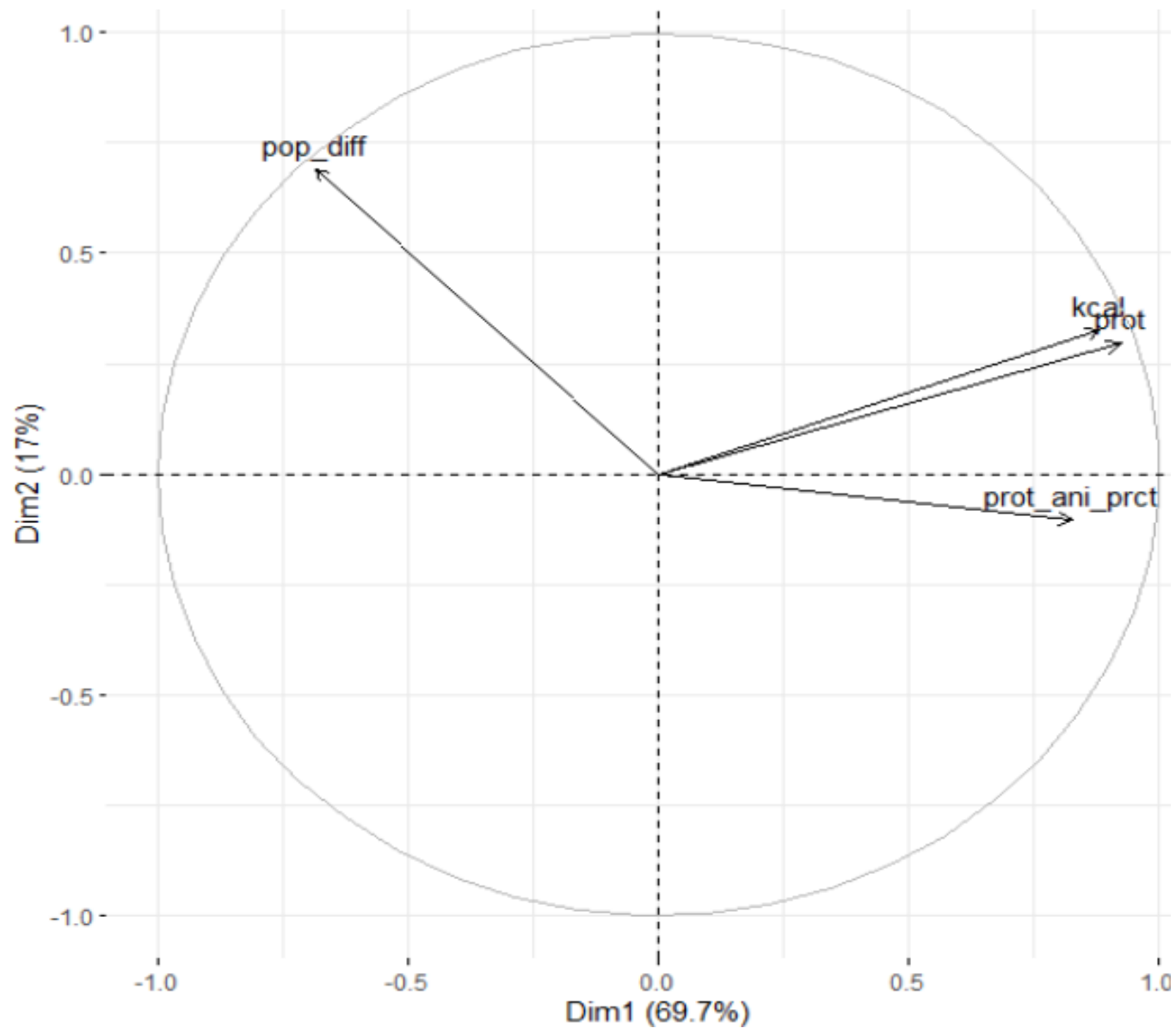
Gabon
Koweit
Maldives
Oman

# Analyse

*Objectif:*

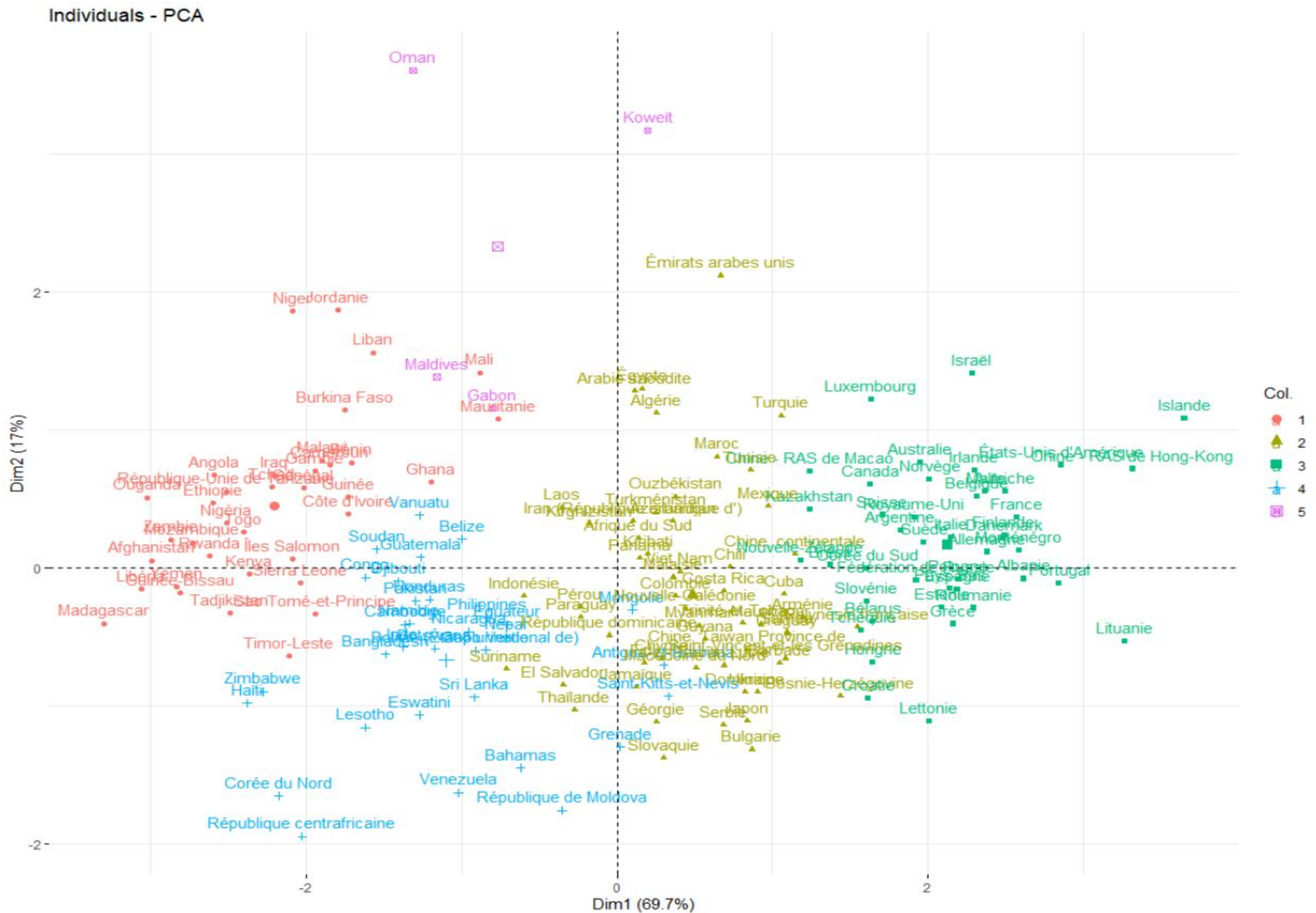
*Bien comprendre la composition de  
nos groupes*

# Observation

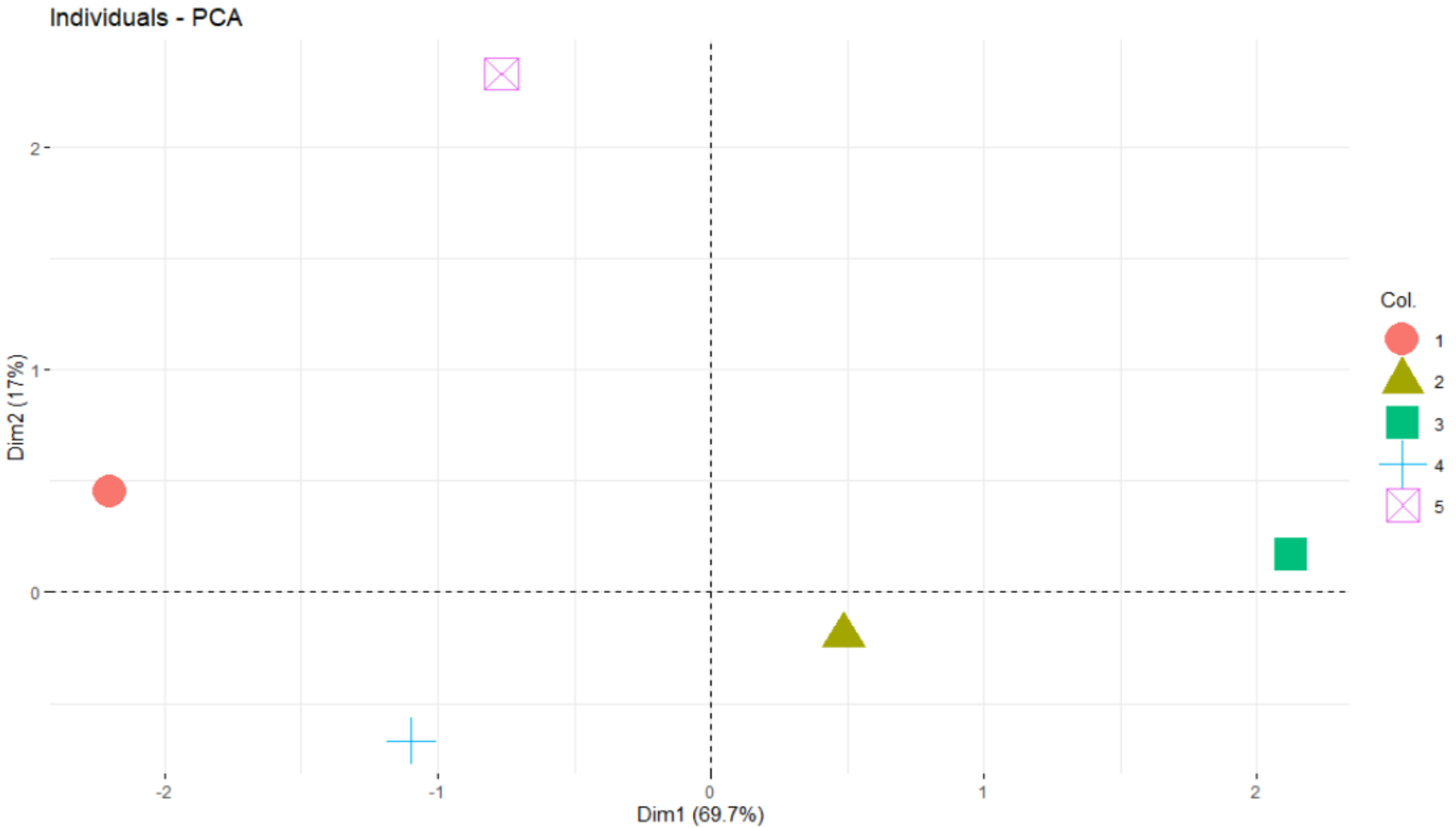




# Observation








# Observation



# Observation

clust	prot	kcal	pib	pop_diff	prot_ani_prct
1	60.7	2432.2	1590.2	32.8	22.3
2	84.7	2989.4	10272.1	9.3	44.8
3	106.4	3387.7	39269.1	5.0	59.5
4	64.5	2423.7	4944.1	14.8	38.2
5	86.6	2818.0	18260.8	55.3	51.8

# Caratérisation

- Cluster 1  « Pays sous-développés »
- Cluster 2  « Autres »
- Cluster 3  « Pays occidentalisés »
- Cluster4  « Pays en transition »
- Cluster 5  « Pays à forte démographie»

# Prise de décision

*Objectif :*

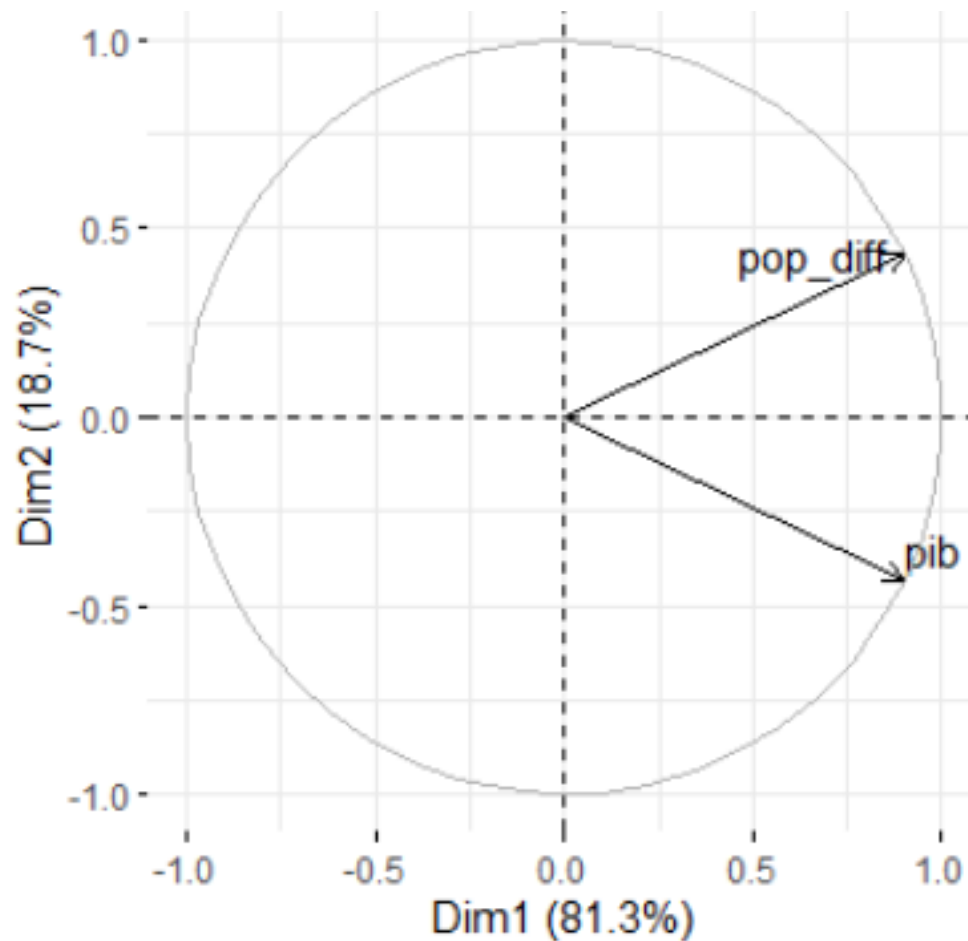
*Choisir les pays les plus intéressants  
pour notre développement*

# Choix

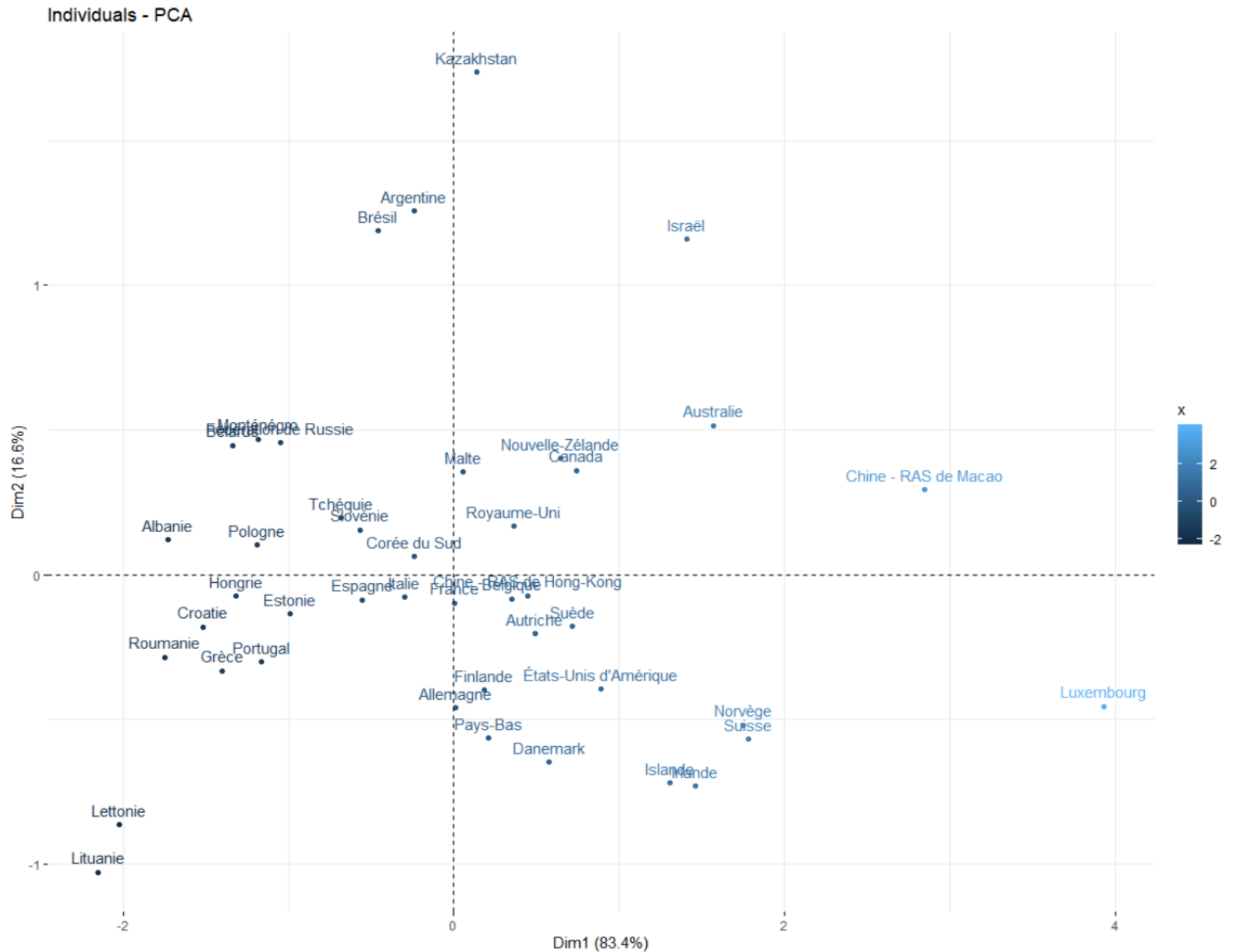
Cluster n°3  
«Pays occidentalisés »



# Affinage du groupe



# Affinage du groupe





# Liste des pays à cibler

 Zone 	Dim.1 
<b>1</b> Luxembourg	3.930617102
<b>2</b> Chine - RAS de Macao	2.842400609
<b>3</b> Suisse	1.778446933
<b>4</b> Norvège	1.748649719
<b>5</b> Australie	1.569058544
<b>6</b> Irlande	1.459909881
<b>7</b> Israël	1.407317487
<b>8</b> Islande	1.306202806
<b>9</b> États-Unis d'Amérique	0.891733052
<b>10</b> Canada	0.743860375
<b>11</b> Suède	0.712306321
<b>12</b> Nouvelle-Zélande	0.647511948
<b>13</b> Danemark	0.574614825
<b>14</b> Autriche	0.488283165
<b>15</b> Chine - RAS de Hong-Kong	0.445411038

# Vérification

***Objectif :***

*Confirmer que nos groupes soient  
significativement différents*

# Test d'adéquation

**CLUSTERS** ✓



**VARIANCE** ✓



**MOYENNE** ✓

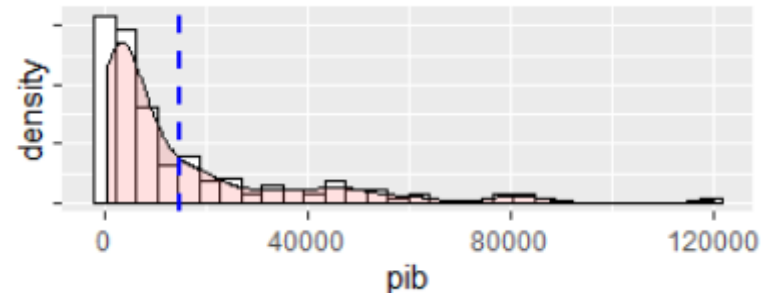
1. Sélection des deux clusters les plus proches graphiquement
2. Test d'adéquation sur les variances : test de Fisher
3. Si les variances sont similaires, on peut tester la moyenne dans de bonnes conditions avec un test de Student

# Shapiro test

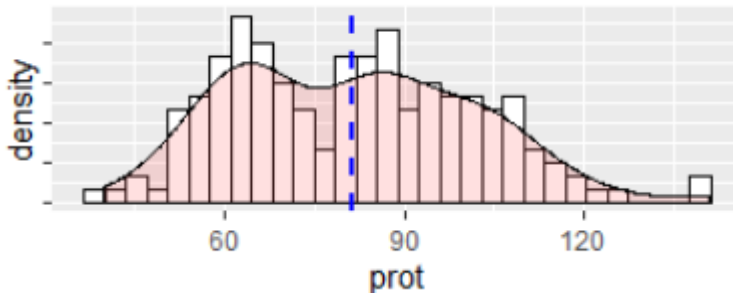
H0 : La distribution est normale

H1 : La distribution n'est pas normale

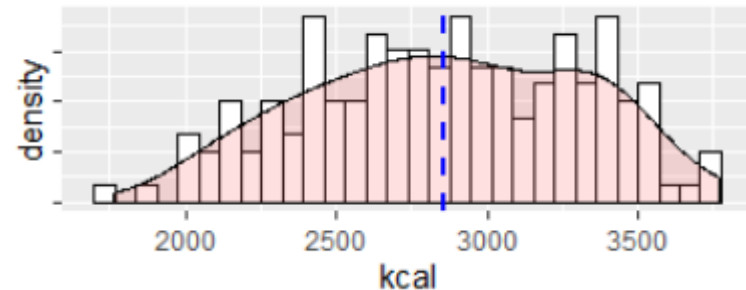
On rejette l'hypothèse nulle au seuil de 5%



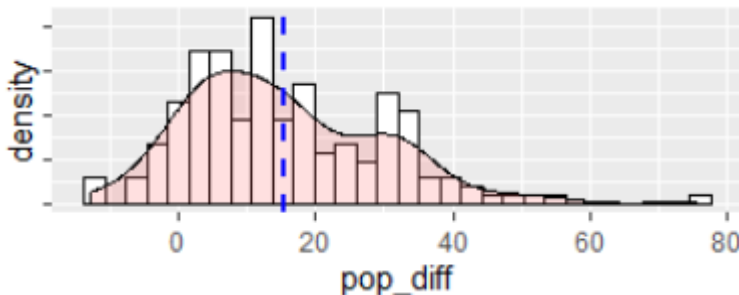
$W = 0.70032$ ,  $p\text{-value} < 2.2e-16$



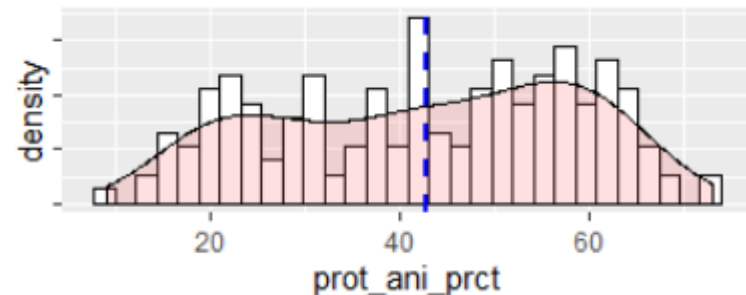
$W = 0.97924$ ,  $p\text{-value} = 0.01123$



$W = 0.98133$ ,  $p\text{-value} = 0.02066$



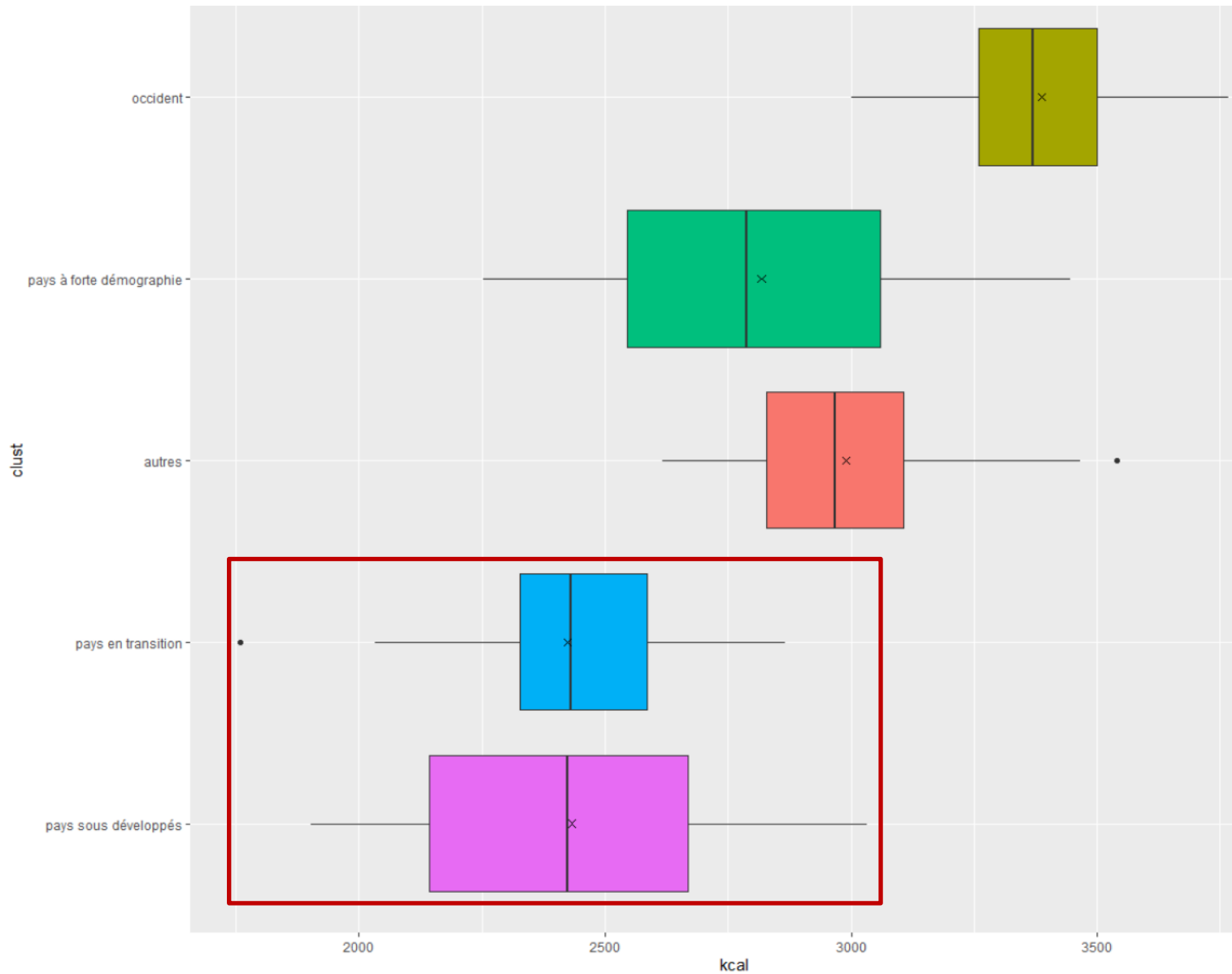
$W = 0.95572$ ,  $p\text{-value} = 3.012e-05$



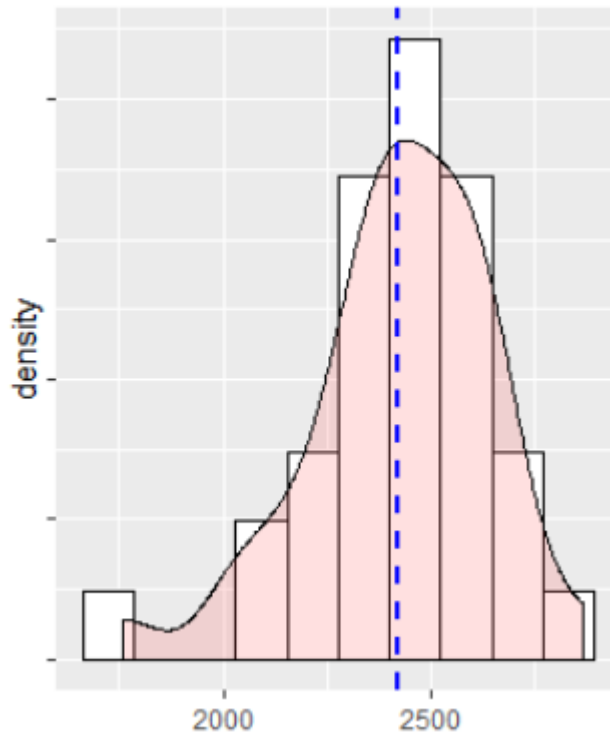
$W = 0.95562$ ,  $p\text{-value} = 2.944e-05$



# Choix des tests à effectuer

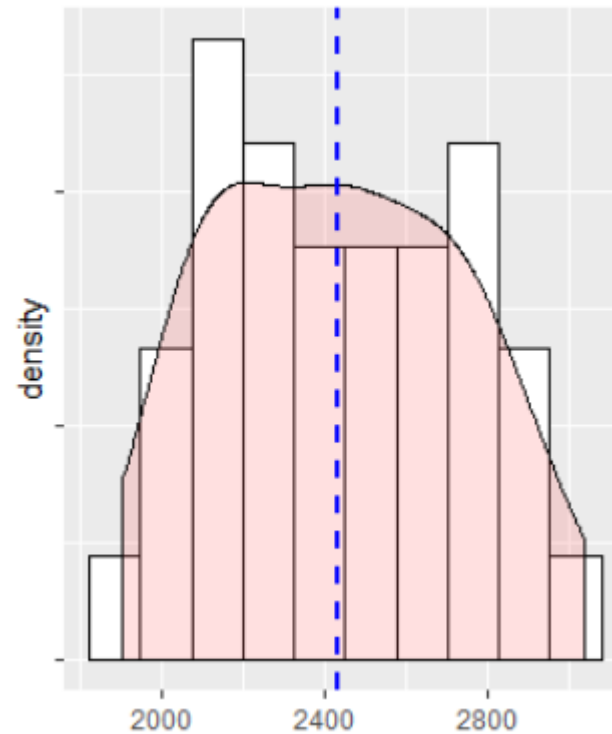


# Choix des tests à effectuer



Pays en transitions

$W = 0.95841$ ,  $p\text{-value} = 0.2329$



Pays sous-développés

$W = 0.9698$ ,  $p\text{-value} = 0.4202$

# Kcal

**H0** : Les variances sont similaires

**On rejette l'hypothèse nulle au seuil de 10%**

**H1**: Les variances sont différentes

F test to compare two variances

```
data: df_1$kcal and df_2$kcal
F = 1.7916, num df = 35, denom df = 32, p-value = 0.09905
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.8941111 3.5515242
sample estimates:
ratio of variances
      1.791575
```



**Les variances sont différentes pour cette variable**

# Conclusion

Les tests d'adéquations valident mes groupes et donc mes choix.

Les 15 pays les plus propices pour notre expansion à l'international sont donc bien ceux de la liste.

	Zone	Dim.1
1	Luxembourg	3.930617102
2	Chine - RAS de Macao	2.842400609
3	Suisse	1.778446933
4	Norvège	1.748649719
5	Australie	1.569058544
6	Irlande	1.459909881
7	Israël	1.407317487
8	Islande	1.306202806
9	États-Unis d'Amérique	0.891733052
10	Canada	0.743860375
11	Suède	0.712306321
12	Nouvelle-Zélande	0.647511948
13	Danemark	0.574614825
14	Autriche	0.488283165
15	Chine - RAS de Hong-Kong	0.445411038



# Bilan

Ce projet m'a permis de me familiariser avec les algorithmes de tri et les tests statistiques.

Mais j'ai trouvé certains points de la démarche trop guidées.