Projet d'apprentissage non supervisé

Marie Guibert - Clémence Chesnais

2023-04-06

Environnement de travail

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(stargazer)
```

```
##
## Please cite as:
##
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library(gridExtra)
```

```
##
## Attachement du package : 'gridExtra'
##
## L'objet suivant est masqué depuis 'package:dplyr':
##
##     combine
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(cluster)
library(NbClust)
```

Question 1

Importation des données

```
d <- read.csv("Pays_donnees.csv",sep=",",dec=".",stringsAsFactors = T,row.names="pays")
str(d)
```

```
## 'data.frame':    167 obs. of  9 variables:
##  $ enfant_mort: num  90.2 16.6 27.3 119 10.3 14.5 18.1 4.8 4.3 39.2 ...
##  $ exports    : num  10 28 38.4 62.3 45.5 18.9 20.8 19.8 51.3 54.3 ...
##  $ sante      : num  7.58 6.55 4.17 2.85 6.03 8.1 4.4 8.73 11 5.88 ...
##  $ imports    : num  44.9 48.6 31.4 42.9 58.9 16 45.3 20.9 47.8 20.7 ...
##  $ revenu     : int  1610 9930 12900 5900 19100 18700 6700 41400 43200 16000 ...
##  $ inflation  : num  9.44 4.49 16.1 22.4 1.44 20.9 7.77 1.16 0.873 13.8 ...
##  $ esper_vie  : num  56.2 76.3 76.5 60.1 76.8 75.8 73.3 82 80.5 69.1 ...
##  $ fert       : num  5.82 1.65 2.89 6.16 2.13 2.37 1.69 1.93 1.44 1.92 ...
##  $ pib_h      : int  553 4090 4460 3530 12200 10300 3220 51900 46900 5840 ...
```

```
# summary(d)
```

Dans ce jeu de données, nous pouvons observer 10 variables dont 9 numériques et 1 facteur comprenant les différents pays (individus). Nous avons choisi de transformer la variable pays en facteur pour simplifier nos traitement des données.

Prétraitement des données

Données manquantes

```
sum(is.na(d))
```

```
## [1] 0
```

Le jeu de données ne présentent pas de valeur manquante. Nous n'avons pas besoin de faire de modification de ce point de vue.

Standardisation des données

Nous pouvons remarquer que les données sont dans des unités différentes et les ordres de grandeur sont très variables. Nous avons donc choisi de standardiser les données.

```
data <- scale(d)
```

Afin de pouvoir analyser ces données, nous allons réaliser des statistiques descriptives de base.

Statistiques descriptives

On effectue les statistiques descriptives sur les valeurs avant standardisation. # DEMANDER AU PROF SI CELA A DU SENS ?

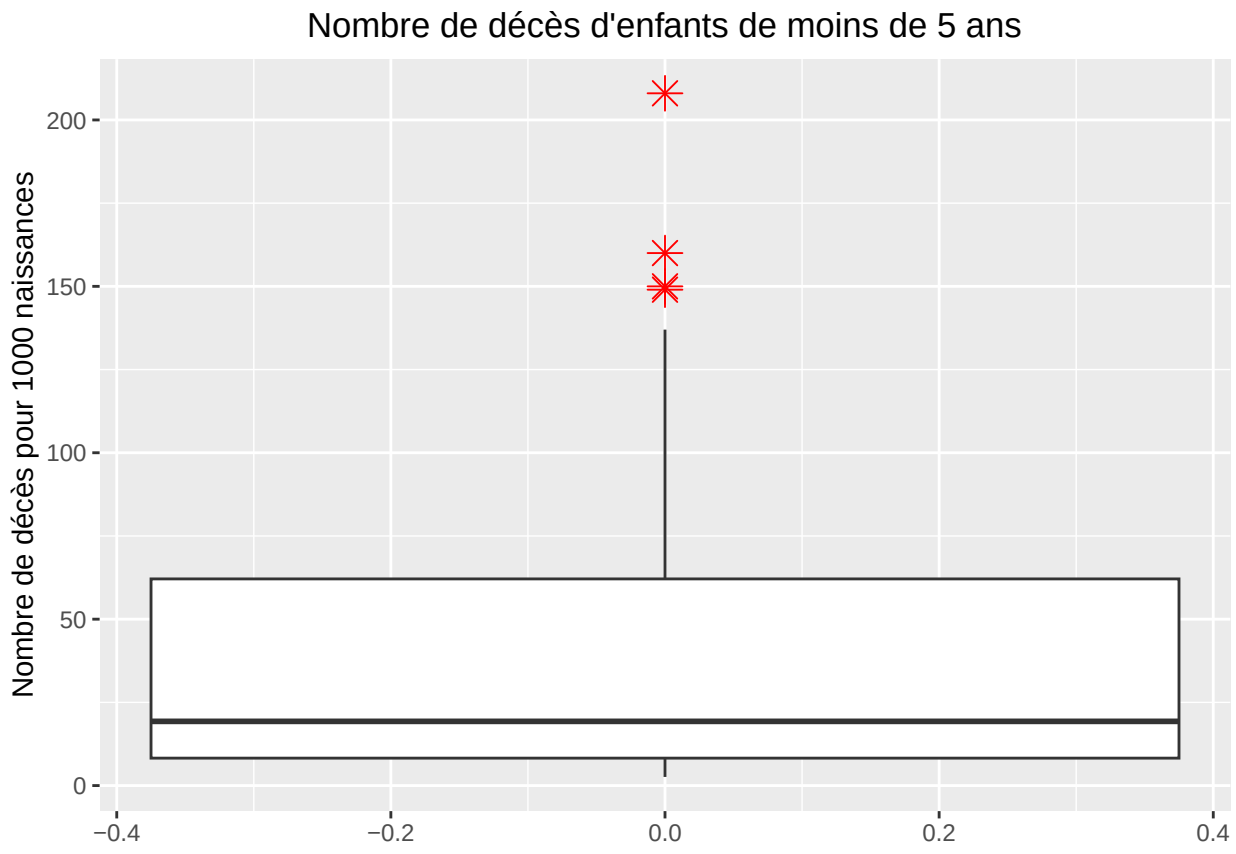
Résumé des données :

```
# stargazer(d,type="text",title="Résumé des données",out="resume_donnnees.txt")
```

Ce résumé statistique nous permet d'avoir une vue d'ensemble sur les données.
Notre jeu de données est composé de 167 pays très hétérogènes. En effet, nous pouvons observer une assez grande différence entre le minimum et le maximum de chaque variable, ce qui prouve la diversité de notre échantillon.

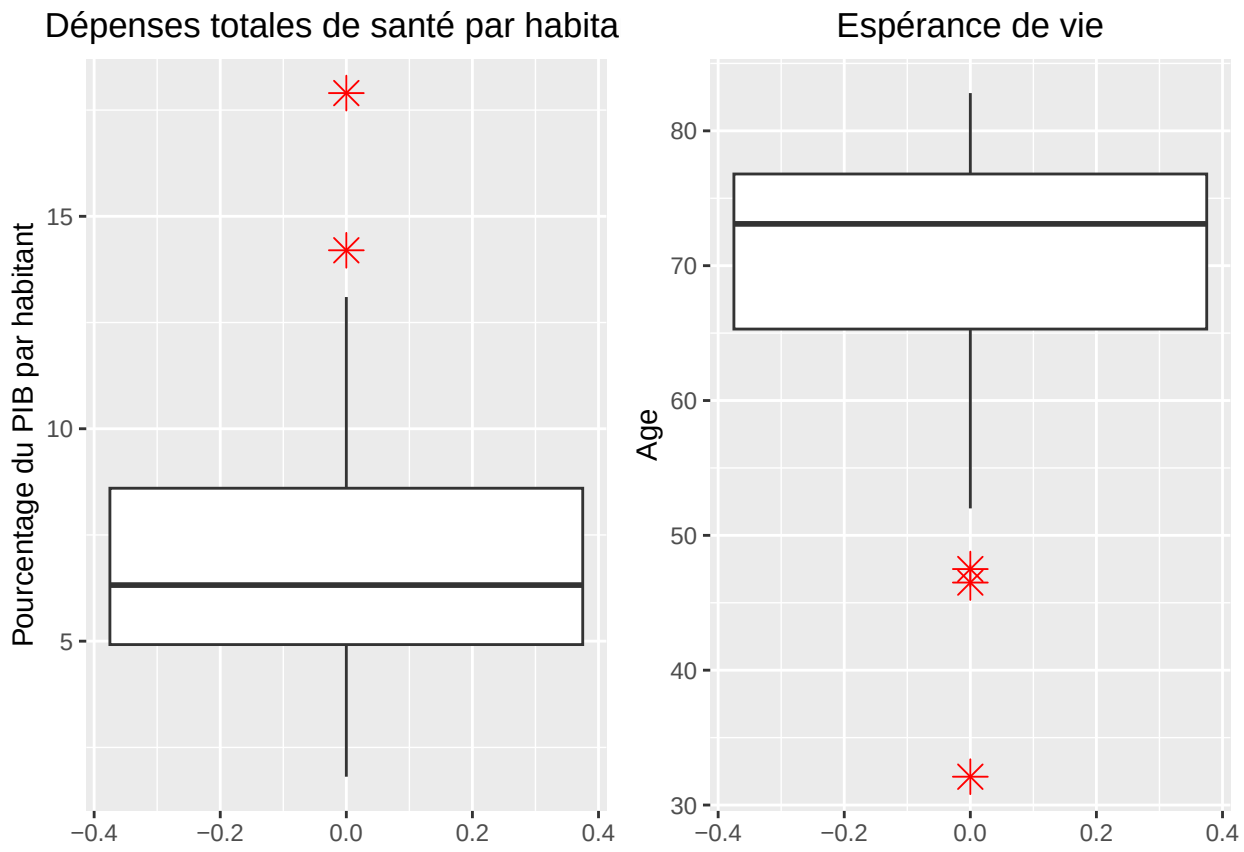
Graphiques :

```
ggplot(data=d, aes(y=enfant_mort)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,outlier.size=4)+
  labs(title="Nombre de décès d'enfants de moins de 5 ans",y="Nombre de décès pour 1000 naissances")+
  theme(plot.title = element_text(hjust=0.5))
```

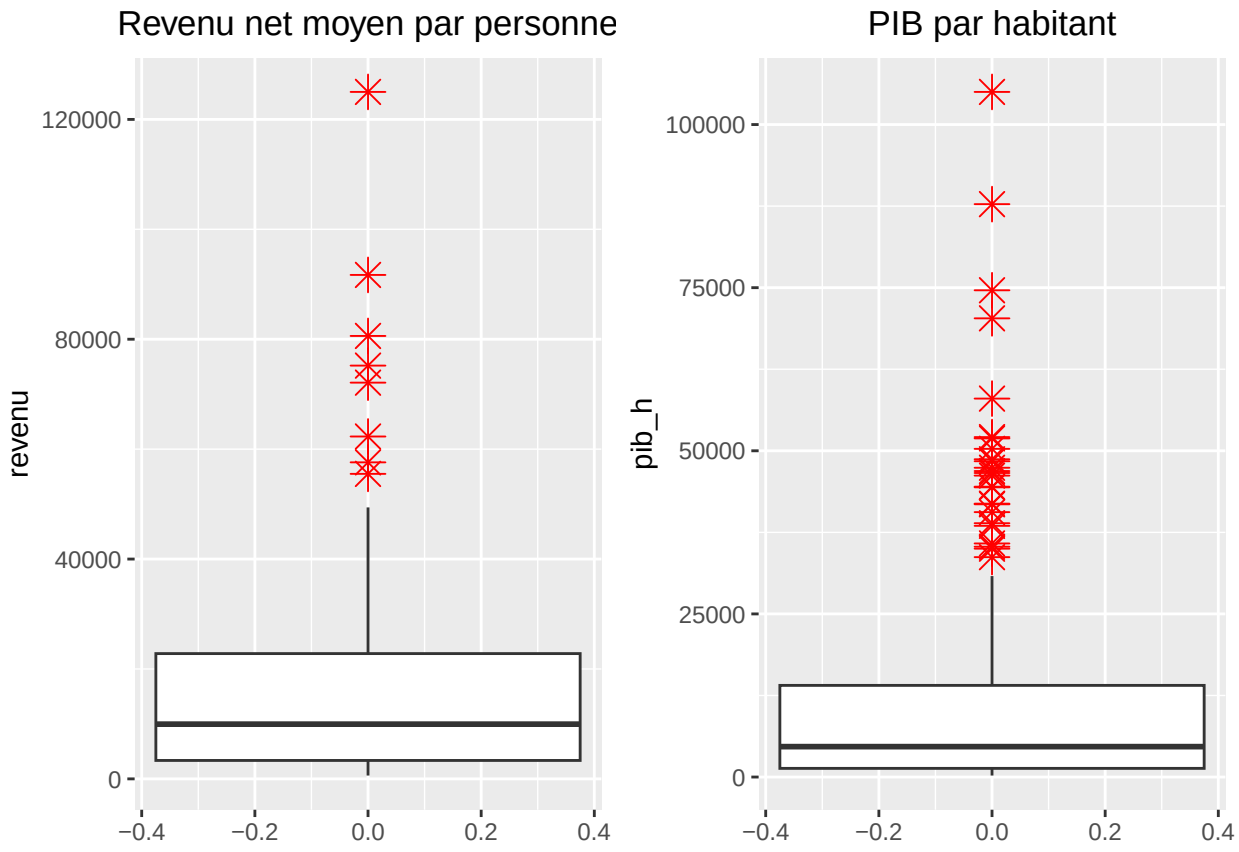



```
sante <- ggplot(data=d, aes(y=sante)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,outlier.size=4)+
  labs(title="Dépenses totales de santé par habitant",y="Pourcentage du PIB par habitant")+
  theme(plot.title = element_text(hjust=0.5))
```

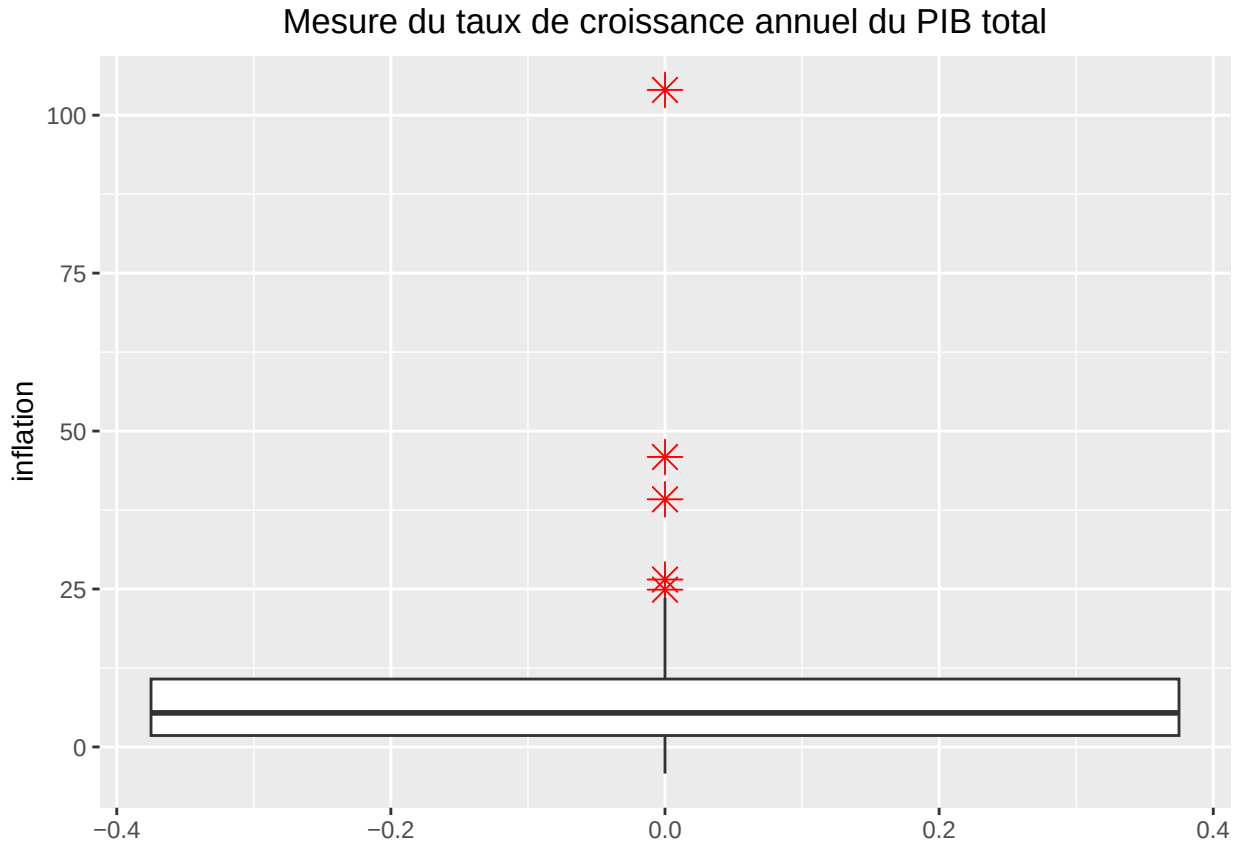
```
esperance <- ggplot(data=d, aes(y=esper_vie)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,outlier.size=4)+
  labs(title="Espérance de vie",y="Age")+
  theme(plot.title = element_text(hjust=0.5))

grid.arrange(sante,esperance,ncol=2)
```



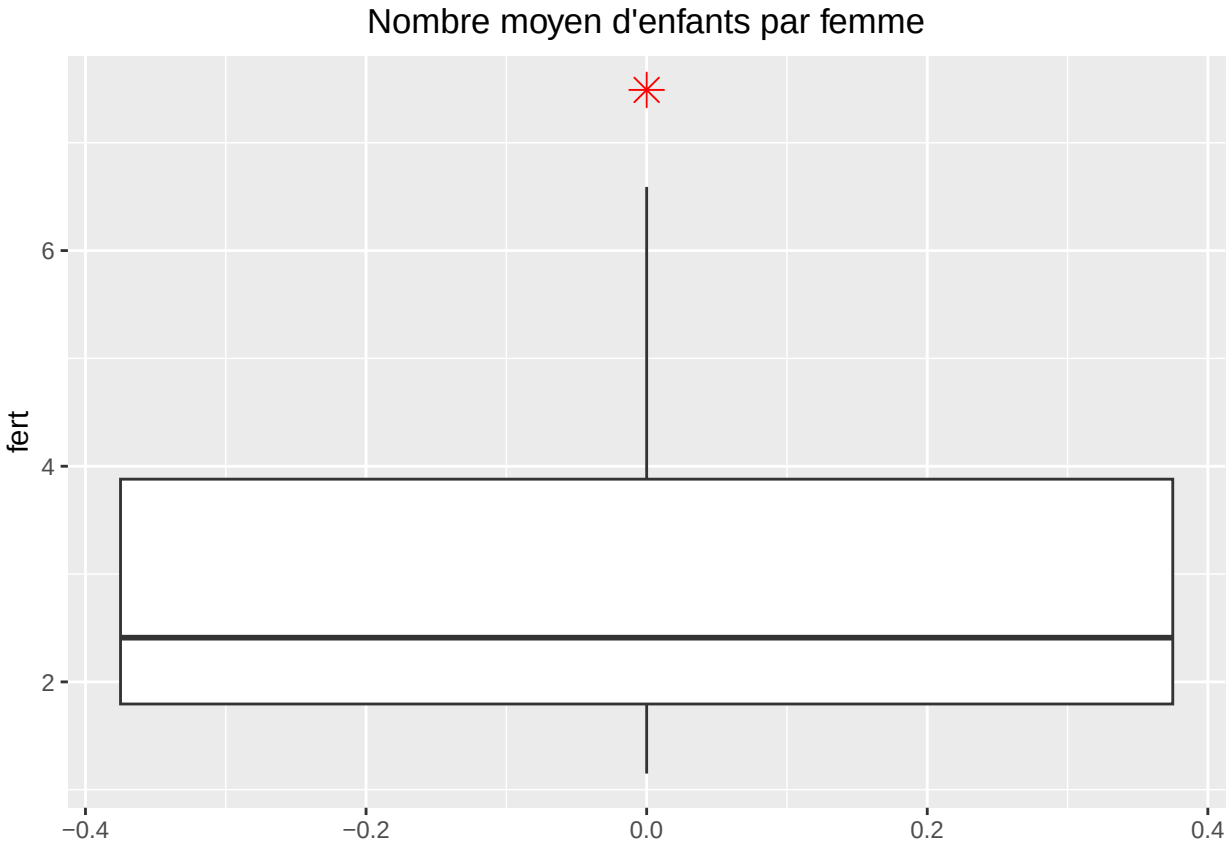

```
revenu_net <- ggplot(data=d, aes(y=revenu)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,outlier.size=4)+
  labs(title="Revenu net moyen par personne")+
  theme(plot.title = element_text(hjust=0.5))

pib_hab <- ggplot(data=d, aes(y=pib_h)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,outlier.size=4)+
  labs(title="PIB par habitant")+
  theme(plot.title = element_text(hjust=0.5))

grid.arrange(revenu_net,pib_hab,ncol=2)
```




```
ggplot(data=d, aes(y=inflation)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,outlier.size=4)+
  labs(title="Mesure du taux de croissance annuel du PIB total")+
  theme(plot.title = element_text(hjust=0.5))
```

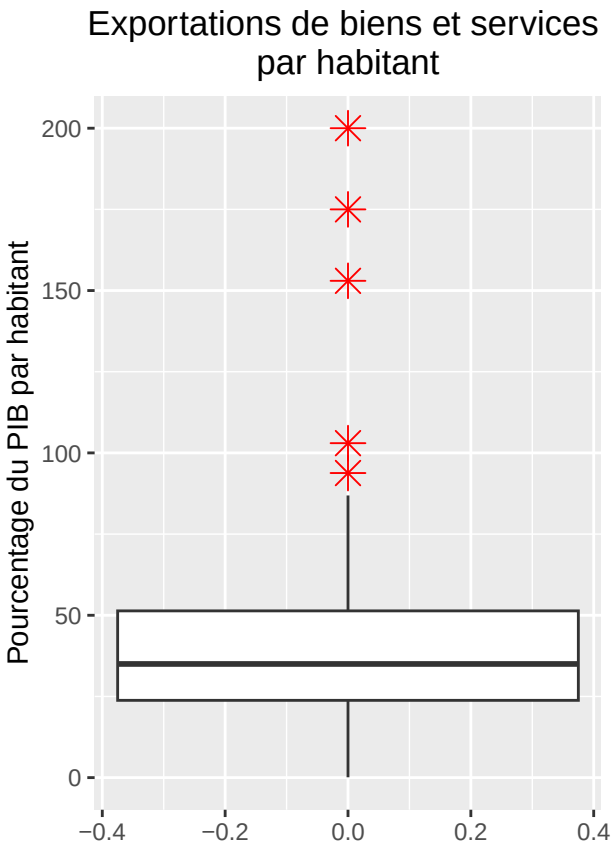


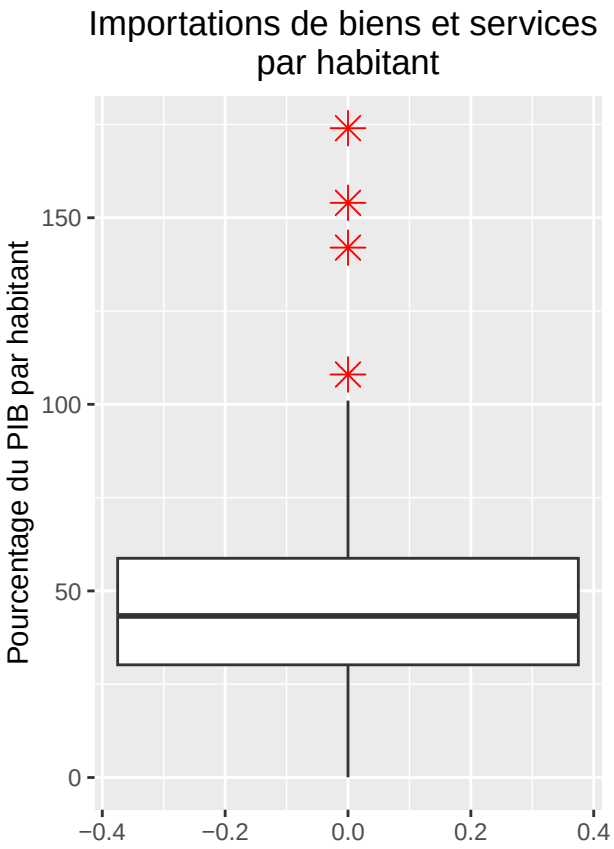

```
ggplot(data=d, aes(y=fert)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,outlier.size=4)+
  labs(title="Nombre moyen d'enfants par femme")+
  theme(plot.title = element_text(hjust=0.5))
```



Imports et Exports :

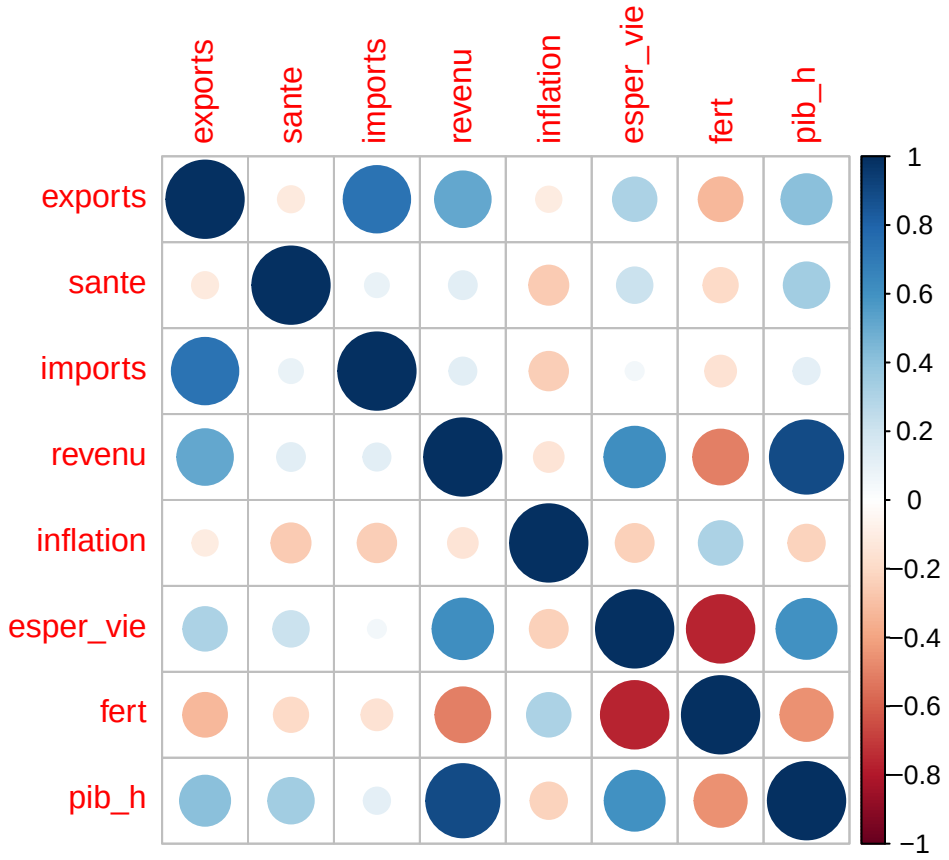
```
imports <- ggplot(data=d, aes(y=imports)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,outlier.size=4)+
  labs(title="Importations de biens et services \npar habitant",y="Pourcentage du PIB par habitant")+
  theme(plot.title = element_text(hjust=0.5))

exports <- ggplot(data=d, aes(y=exports)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,outlier.size=4)+
  labs(title="Exportations de biens et services \npar habitant",y="Pourcentage du PIB par habitant")+
  theme(plot.title = element_text(hjust=0.5))

grid.arrange(exports, imports, ncol=2)
```





Matrice de corrélation :

```
corrplot(cor(d[-1]),method="circle")
```



Grâce à cette matrice de corrélation, nous pouvons observer une corrélation négative, entre l'espérance de vie et le nombre d'enfants par femme. Par ailleurs, une corrélation positive,proche de 1, apparaît entre le PIB par habitant et le revenu net moyen par personne.

Question 2

Matrice de dissimilarité :

Afin de chercher les individus similaires, on peut calculer une matrice de distance / dissimilarité.

```
MD <- as.matrix(dist(data,method = "euclidean"))
# MD <- as.matrix(dist(data,method = "minkowski"))
# MD <- as.matrix(dist(data,method = "manhattan"))
which(MD == min(MD[row(MD)!=col(MD)]),arr.ind=TRUE)
```

```
##         row col
## Poland  122  42
## Croatia  42 122
```

Avec la méthode de la distance euclidienne, de manhattan et minkowski, les deux pays les plus proches / similaires sont la Pologne et la Croatie.

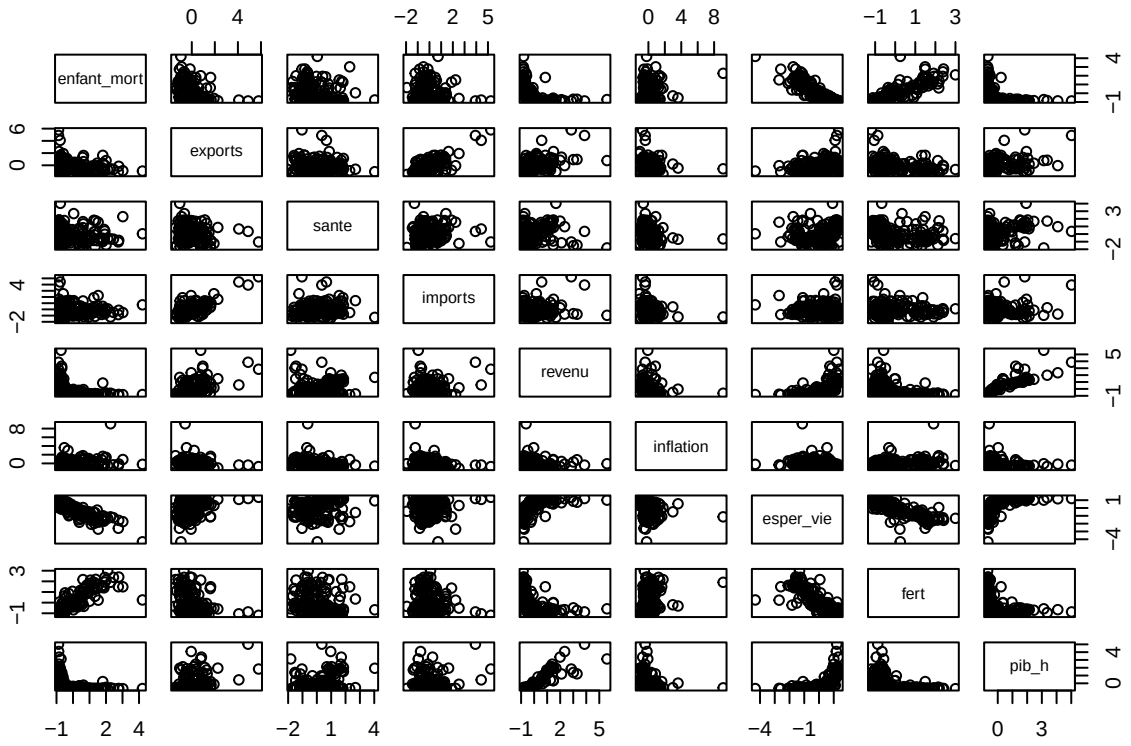
Dans notre situation, les variables sont quantitatives, nous pouvons donc utiliser une approche en termes de distances. On cherche à partitionner les pays en groupes distincts et homogènes afin de déterminer leur besoin de d'aide. L'objectif est de former des groupes compacts avec une faible variabilité au sein des groupes.

Première approche : CAH

Classification Ascendante Hiérarchique :

Cette première représentation nous permet d'observer de potentiels groupes de pays. Ayant beaucoup de variables, cette analyse est un peu plus compliquée et aucune partition ne semble se démarquer.

```
pairs(data)
```

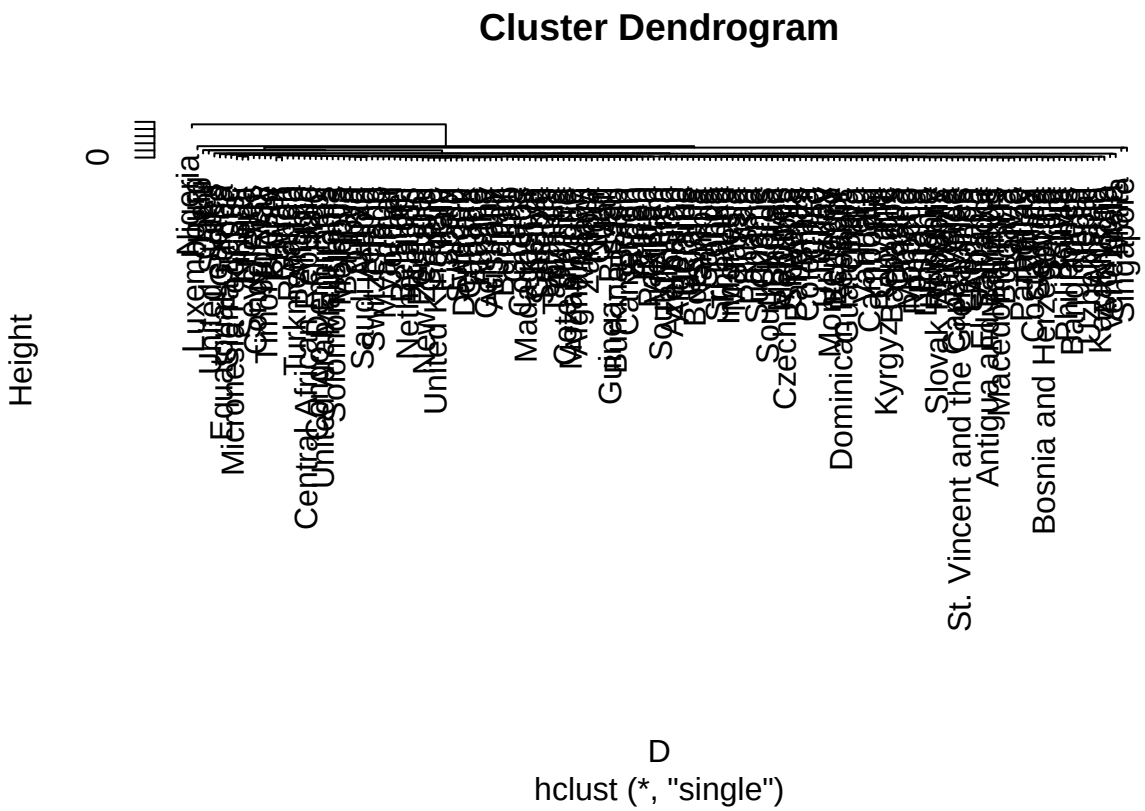


On calcule d'abord la distance euclidienne au carré. Le calcul de la distance euclidienne nous permettra par la suite d'être optimal lors de l'utilisation de la stratégie de Ward.

```
D <- dist(data,method="euclidean")^2
```

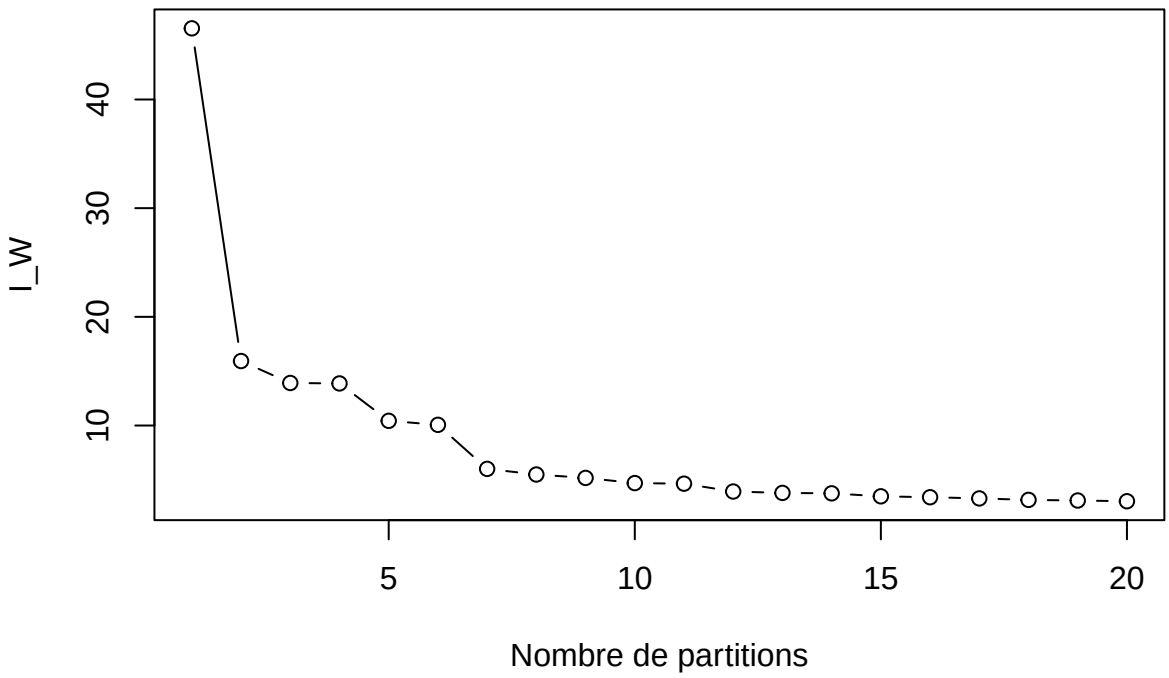
Méthode CAH avec le saut minimal (single linkage) :

```
CAH_min <- hclust(d= D,method="single")
plot(CAH_min)
```



Ce premier dendrogramme n'est pas très explicite et ne nous permet pas de faire un choix de partition clair.

```
plot(rev(CAH_min$height)[1:20],type="b",xlab="Nombre de partitions",ylab="I_W")
```

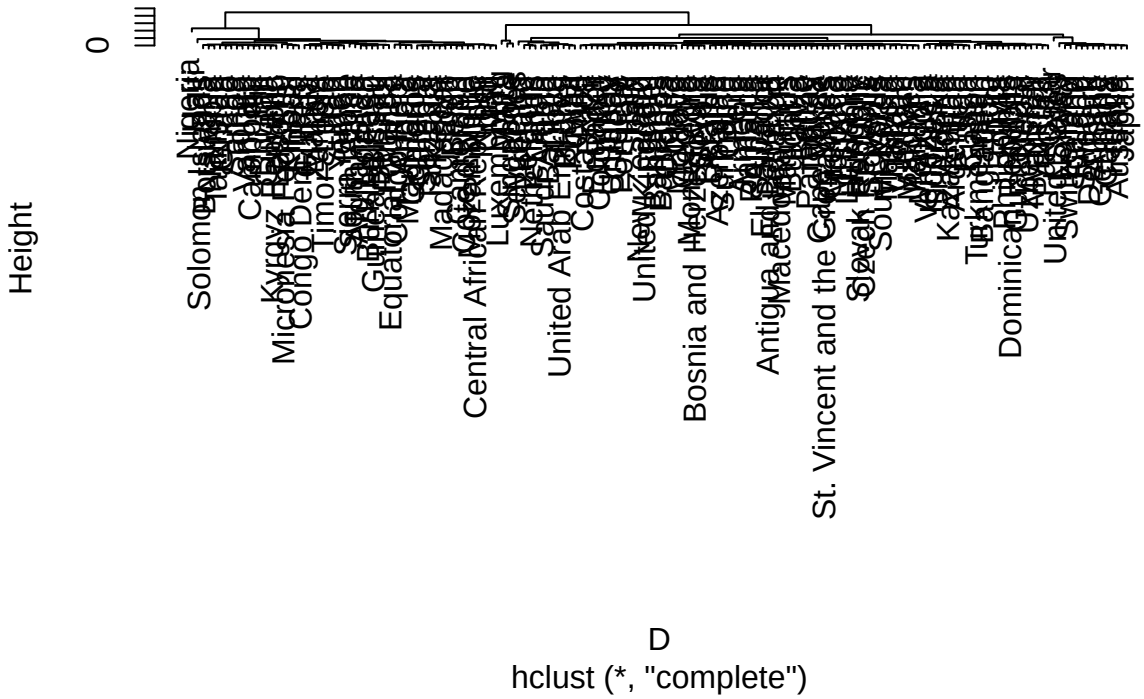


Cependant, le tracé de la pertie d'inertie nous suggère de choisir une partition en 2 groupes. Nous avons choisi de représenter seulement les 20 premières valeurs pour ne pas "noyer" l'information importante. Chaque coupure correspond à un saut important d'inertie intra-classes.

Faisons maintant les mêmes graphiques avec la méthode de distance de saut maximal (complet linkage).

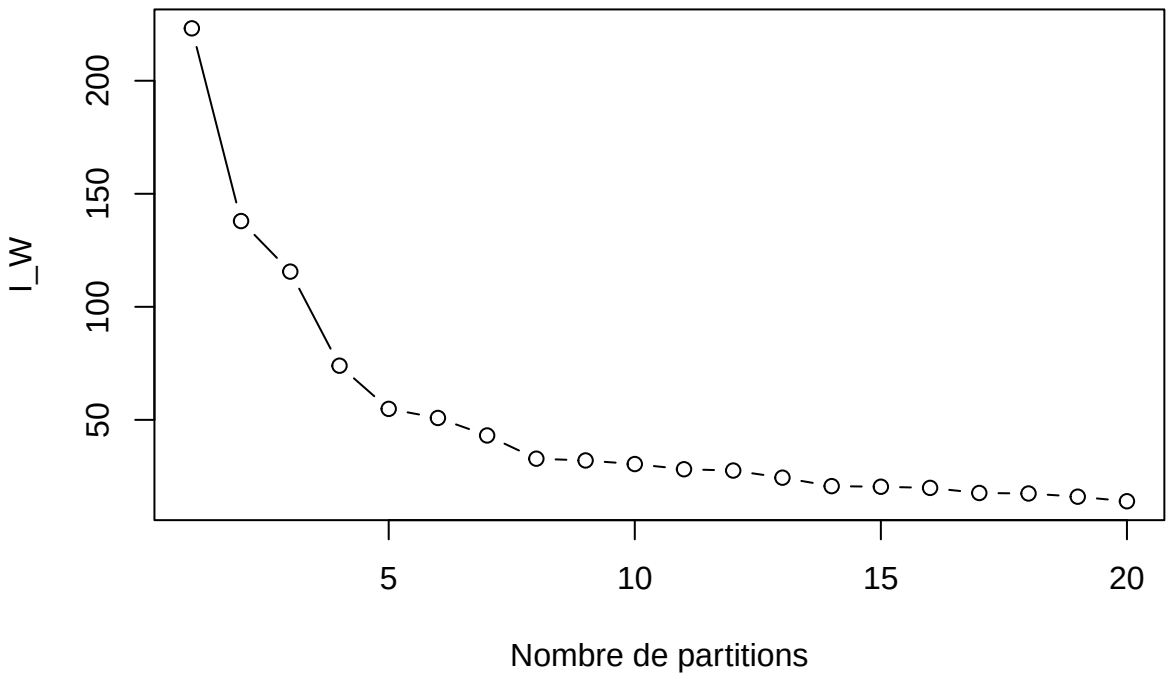
Méthode CAH avec le saut maximal :

```
CAH_max <- hclust(d= D,method="complete")
plot(CAH_max)
```



En analysant ce dendrogramme, nous pouvons distinguer 2 ou 3 groupes de pays différents. Le graphique ci-dessous n'est pas très concluant quant à cette hypothèse. Nous avons donc besoin de continuer nos analyses.

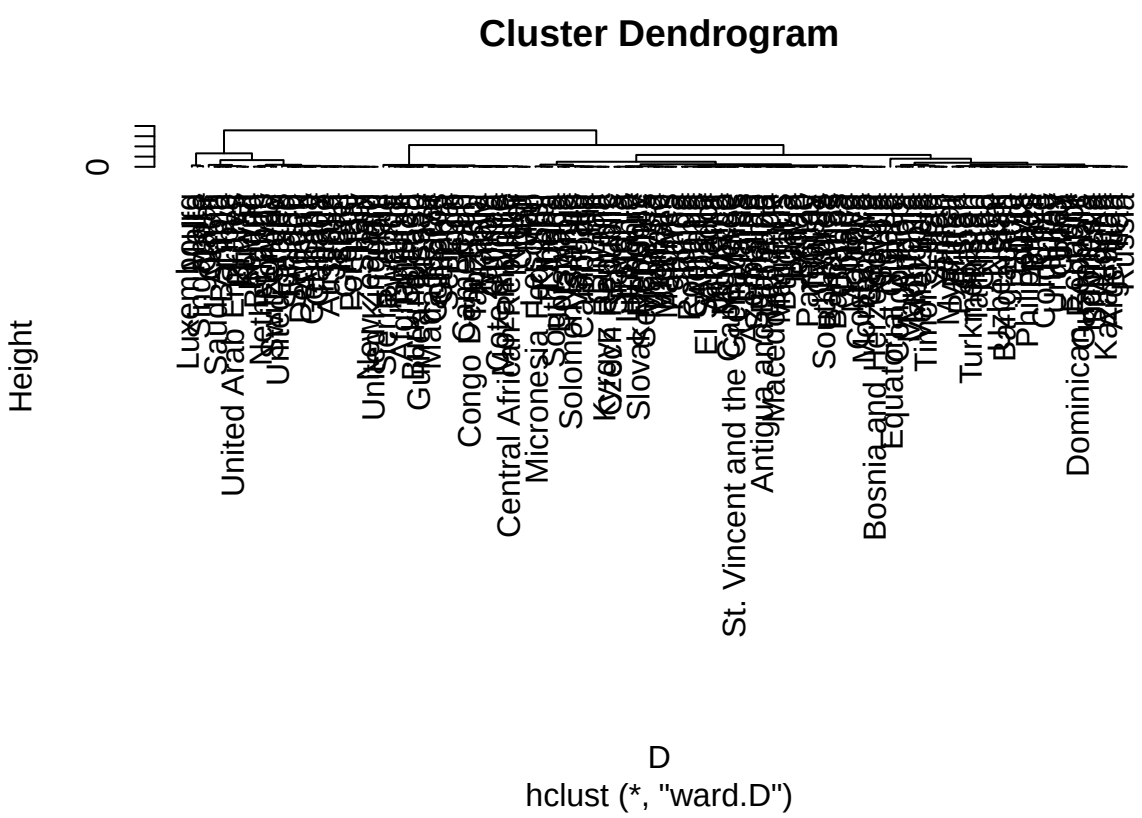
```
plot(rev(CAH_max$height)[1:20],type="b",xlab="Nombre de partitions",ylab="I_W")
```



CAH avec la distance de Ward :

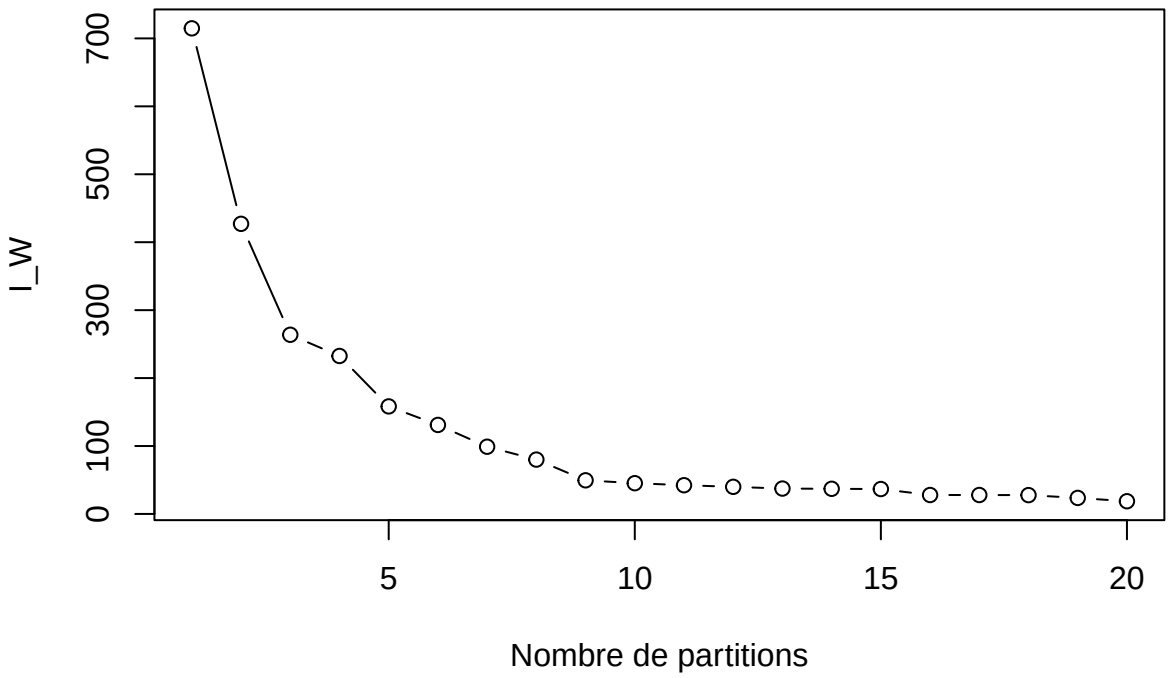
Enfin, avec la distance de ward, on obtient les résultats suivants :

```
CAH_ward <- hclust( d = D,method="ward.D")
plot(CAH_ward,hang=-1)
```



Le dendrogramme nous permet aussi de supposer l'existence de 2 voire 3 groupes.
Le tracé de la perte d'inertie nous permet de nous pencher vers le choix de 2 groupes.

```
plot(rev(CAH_ward$height)[1:20],type="b",xlab="Nombre de partitions",ylab="I_W")
```

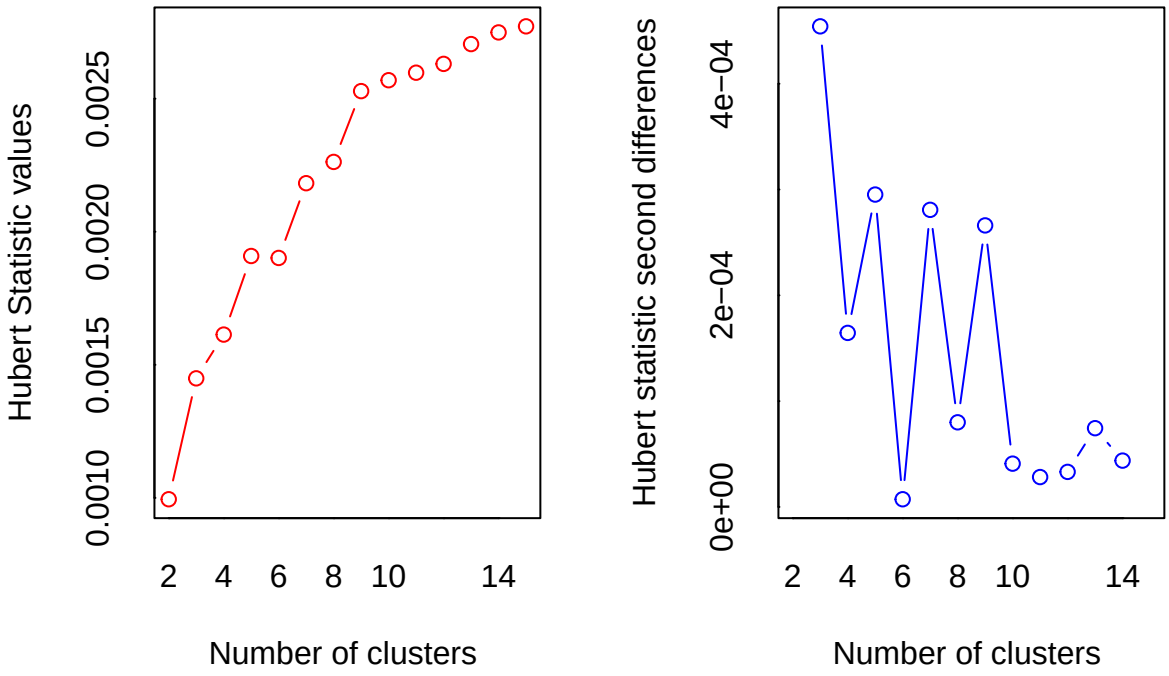


Ainsi, cette dernière classification ascendante hiérarchique nous permet de confirmer notre hypothèse de partition.

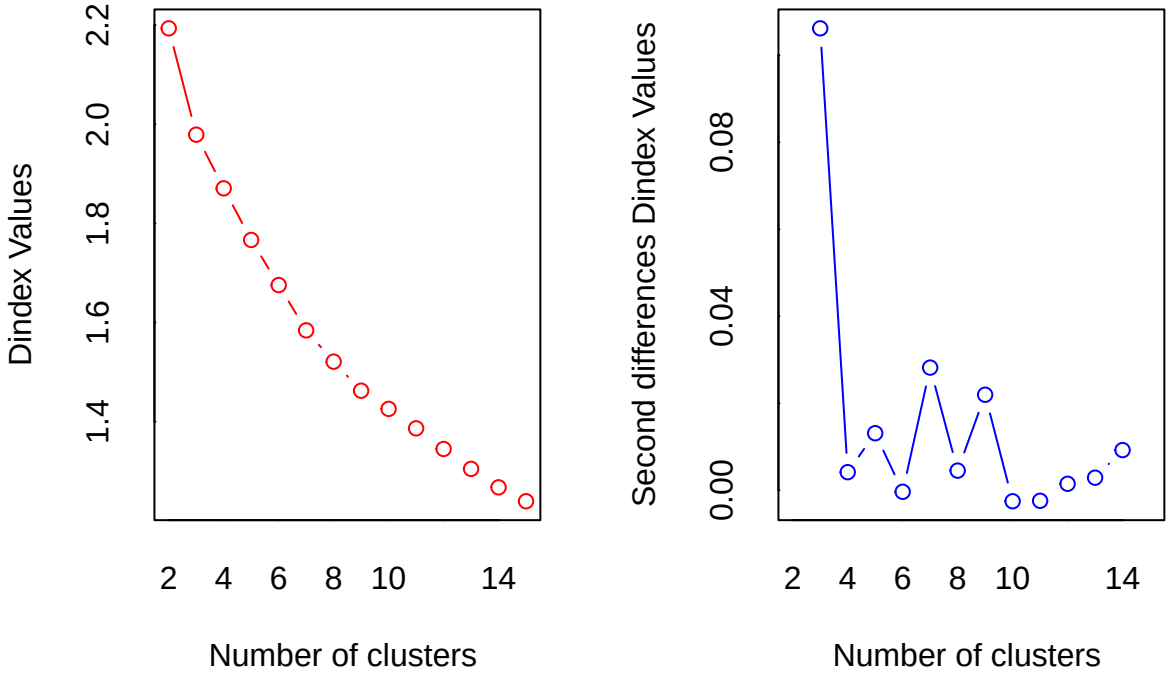
A REVOIR : POURQUOI ON A PAS DE COUDE ???

Critère automatique à partir du package 'NbClust' :

```
NbClust(data,min.nc = 2,max.nc = 15,method="ward.D",index="all")
```




```
## *** : The Hubert index is a graphical method of determining the number of clusters.
##                 In the plot of Hubert index, we seek a significant knee that corresponds to a
##                 significant increase of the value of the measure i.e the significant peak in Hubert
##                 index second differences plot.
##
```




```
## *** : The D index is a graphical method of determining the number of clusters.
##                 In the plot of D index, we seek a significant knee (the significant peak in Dindex
##                 second differences plot) that corresponds to a significant increase of the value of
##                 the measure.
##
## *******************************************************************
## * Among all indices:
## * 5 proposed 2 as the best number of clusters
## * 4 proposed 3 as the best number of clusters
## * 5 proposed 4 as the best number of clusters
## * 1 proposed 5 as the best number of clusters
## * 1 proposed 8 as the best number of clusters
## * 4 proposed 9 as the best number of clusters
## * 1 proposed 12 as the best number of clusters
## * 1 proposed 14 as the best number of clusters
## * 1 proposed 15 as the best number of clusters
##
##                     ***** Conclusion *****
##
## * According to the majority rule, the best number of clusters is  2
##
##
## *******************************************************************

## $All.index
```

```
##        KL      CH Hartigan     CCC     Scott      Marriot    TrCovW    TraceW
## 2  2.3386 68.6210  33.7214 -2.6434  225.4314 7.313743e+16 23244.901 1055.1703
## 3  1.8427 57.8307  20.9221 -2.4313  418.9626 5.164539e+16 17615.022  876.1166
## 4  1.5564 50.1387  14.8863 -2.6786  609.3518 2.936225e+16 13591.412  776.9927
## 5  1.1442 44.4853  12.5772 -3.1997  758.7814 1.875028e+16 12060.060  711.9705
## 6  0.2926 40.6143  28.1333 -3.4454  808.6017 2.003593e+16  9447.050  660.6776
## 7  2.0767 44.1722  16.2945  0.1436  962.9977 1.081897e+16  6363.185  562.4029
## 8  0.5419 43.7703  28.2797  1.4605 1070.1097 7.440776e+15  4995.833  510.4213
## 9  4.3721 48.3399   9.3875  5.1119 1220.3724 3.829616e+15  3394.298  433.3464
## 10 1.1997 46.2701   8.1650  5.2609 1293.2347 3.056237e+15  3031.433  409.0434
## 11 0.8340 44.3410   8.9717  5.3083 1376.7840 2.242316e+15  2823.262  388.8221
## 12 1.2044 43.1654   7.8186  5.6184 1447.0356 1.752186e+15  2505.380  367.6767
## 13 0.9742 41.9434   7.8612  5.7962 1506.8698 1.437148e+15  2361.741  350.0208
## 14 1.5395 41.0299   5.8597  6.0549 1560.7539 1.207096e+15  2068.108  333.0212
## 15 1.2204 39.7156   5.0920  6.0050 1614.6635 1.003397e+15  1975.666  320.7374
##    Friedman  Rubin Cindex     DB Silhouette   Duda Pseudot2   Beale Ratkowsky
## 2   16.5255 1.4159 0.2743 1.5019     0.2817 0.7340  36.2418  2.1548    0.3334
## 3   25.3844 1.7053 0.2347 1.5929     0.2289 0.6755  15.3735  2.7975    0.3307
## 4   36.3323 1.9228 0.2246 1.4508     0.2470 0.8298  12.9207  1.2123    0.3286
## 5   40.4147 2.0984 0.2063 1.7295     0.2079 0.7280  24.6545  2.2097    0.3070
## 6   42.7143 2.2613 0.2032 1.7717     0.1599 0.3414  17.3592 10.4241    0.2951
## 7   51.2007 2.6565 0.1987 1.5066     0.1827 0.7968  10.4568  1.4951    0.2928
## 8   54.2108 2.9270 0.1882 1.4706     0.2036 1.0665  -1.6830 -0.3609    0.2832
## 9   56.8841 3.4476 0.3108 1.2326     0.2160 0.7129  10.4715  2.3289    0.2800
## 10  59.8853 3.6524 0.2997 1.2020     0.2206 0.7251  12.1343  2.2081    0.2688
## 11  62.3662 3.8424 0.2914 1.2141     0.2056 0.4977  12.1126  5.5951    0.2587
## 12  64.6251 4.0634 0.2862 1.1860     0.2105 0.6908   9.4000  2.5657    0.2503
## 13  68.3495 4.2683 0.2798 1.2286     0.1915 0.7334  11.6328  2.1168    0.2424
## 14  70.8060 4.4862 0.2721 1.2700     0.1875 0.7291   7.4314  2.1250    0.2353
## 15  72.5606 4.6580 0.2654 1.3223     0.1767 0.5385  13.7138  4.8442    0.2286
##        Ball Ptbiserial    Frey McClain   Dunn Hubert SDindex Dindex   SDbw
## 2  527.5852     0.3422  0.2595  0.6490 0.0751 0.0010  2.7240 2.1934 1.0880
## 3  292.0389     0.4053 -0.1502  1.1448 0.0751 0.0014  2.8973 1.9790 0.9576
## 4  194.2482     0.4357  0.4909  1.1797 0.0757 0.0016  3.0800 1.8707 1.0002
## 5  142.3941     0.4312  5.5517  1.5461 0.0757 0.0019  3.1602 1.7666 0.9660
## 6  110.1129     0.3474 -0.1479  2.5747 0.0685 0.0019  3.1581 1.6756 0.7597
## 7   80.3433     0.3571  0.1256  2.5488 0.0685 0.0022  3.1016 1.5843 0.6454
## 8   63.8027     0.3659 -0.1291  2.8953 0.0717 0.0023  3.1233 1.5211 0.6410
## 9   48.1496     0.3835  0.1872  2.8069 0.1221 0.0025  2.9451 1.4625 0.4967
## 10  40.9043     0.3829  0.8481  2.9427 0.1221 0.0026  2.9247 1.4258 0.4672
## 11  35.3475     0.3608  0.0927  3.4443 0.1221 0.0026  3.1190 1.3866 0.4425
## 12  30.6397     0.3617  0.4867  3.4868 0.1221 0.0026  3.0365 1.3450 0.4064
## 13  26.9247     0.3521  1.0162  3.7714 0.1154 0.0027  2.9761 1.3048 0.3864
## 14  23.7872     0.3208  0.4286  4.6963 0.1154 0.0027  3.4028 1.2676 0.3676
## 15  21.3825     0.3095  1.4422  5.1586 0.1154 0.0028  3.4239 1.2396 0.3519
##
## $All.CriticalValues
##    CritValue_Duda CritValue_PseudoT2 Fvalue_Beale
## 2          0.7868            27.0994       0.0231
## 3          0.6825            14.8875       0.0037
## 4          0.7508            20.9124       0.2845
## 5          0.7548            21.4445       0.0200
## 6          0.4954             9.1671       0.0000
## 7          0.7098            16.7607       0.1478
```

```
## 8          0.6621            13.7821       1.0000
## 9          0.6573            13.5542       0.0158
## 10         0.6825            14.8875       0.0216
## 11         0.5447            10.0311       0.0000
## 12         0.6292            12.3746       0.0083
## 13         0.6825            14.8875       0.0282
## 14         0.6225            12.1297       0.0296
## 15         0.5901            11.1141       0.0000
##
## $Best.nc
##                       KL     CH Hartigan     CCC    Scott      Marriot   TrCovW
## Number_clusters 9.0000  2.000   9.0000 14.0000   3.0000 5.000000e+00    3.000
## Value_Index     4.3721 68.621  18.8923  6.0549 193.5312 1.189761e+16 5629.879
##                  TraceW Friedman   Rubin Cindex     DB Silhouette   Duda
## Number_clusters   3.00    4.000  9.0000 8.0000 12.000     2.0000 4.0000
## Value_Index      79.93   10.948 -0.3158 0.1882  1.186     0.2817 0.8298
##                 PseudoT2  Beale Ratkowsky     Ball PtBiserial Frey McClain
## Number_clusters   4.0000 4.0000    2.0000   3.0000     4.0000    1   2.000
## Value_Index      12.9207 1.2123    0.3334 235.5463     0.4357   NA   0.649
##                    Dunn Hubert SDindex Dindex    SDbw
## Number_clusters 9.0000      0   2.000      0 15.0000
## Value_Index     0.1221      0   2.724      0  0.3519
##
## $Best.partition
##                     Afghanistan                         Albania
##                               1                               2
##                         Algeria                          Angola
##                               2                               1
##             Antigua and Barbuda                       Argentina
##                               2                               2
##                         Armenia                       Australia
##                               2                               2
##                         Austria                      Azerbaijan
##                               2                               2
##                         Bahamas                         Bahrain
##                               2                               2
##                      Bangladesh                        Barbados
##                               1                               2
##                         Belarus                         Belgium
##                               2                               2
##                          Belize                           Benin
##                               2                               1
##                          Bhutan                         Bolivia
##                               1                               1
##          Bosnia and Herzegovina                        Botswana
##                               2                               1
##                          Brazil                          Brunei
##                               2                               2
##                        Bulgaria                    Burkina Faso
##                               2                               1
##                         Burundi                        Cambodia
##                               1                               1
##                        Cameroon                          Canada
##                               1                               2
```

```
##                  Cape Verde       Central African Republic
##                           2                              1
##                        Chad                          Chile
##                           1                              2
##                       China                       Colombia
##                           2                              2
##                     Comoros                 Congo Dem. Rep.
##                           1                              1
##                  Congo Rep.                     Costa Rica
##                           1                              2
##               Cote d'Ivoire                        Croatia
##                           1                              2
##                      Cyprus                 Czech Republic
##                           2                              2
##                     Denmark              Dominican Republic
##                           2                              2
##                     Ecuador                          Egypt
##                           2                              1
##                 El Salvador              Equatorial Guinea
##                           2                              1
##                     Eritrea                        Estonia
##                           1                              2
##                        Fiji                        Finland
##                           1                              2
##                      France                          Gabon
##                           2                              1
##                      Gambia                        Georgia
##                           1                              2
##                     Germany                          Ghana
##                           2                              1
##                      Greece                        Grenada
##                           2                              2
##                   Guatemala                         Guinea
##                           2                              1
##               Guinea-Bissau                         Guyana
##                           1                              1
##                       Haiti                        Hungary
##                           1                              2
##                     Iceland                          India
##                           2                              1
##                   Indonesia                           Iran
##                           2                              2
##                        Iraq                        Ireland
##                           1                              2
##                      Israel                          Italy
##                           2                              2
##                     Jamaica                          Japan
##                           2                              2
##                      Jordan                     Kazakhstan
##                           2                              2
##                       Kenya                       Kiribati
##                           1                              1
##                      Kuwait                  Kyrgyz Republic
##                           2                              1
```

```
##                     Lao                  Latvia
##                       1                       2
##                 Lebanon                Lesotho
##                       2                       1
##                 Liberia                   Libya
##                       1                       2
##               Lithuania              Luxembourg
##                       2                       2
##            Macedonia FYR              Madagascar
##                       2                       1
##                  Malawi                Malaysia
##                       1                       2
##                 Maldives                    Mali
##                       2                       1
##                   Malta              Mauritania
##                       2                       1
##                Mauritius    Micronesia Fed. Sts.
##                       2                       1
##                 Moldova                Mongolia
##                       2                       2
##              Montenegro                Morocco
##                       2                       2
##              Mozambique                Myanmar
##                       1                       1
##                 Namibia                   Nepal
##                       1                       1
##             Netherlands             New Zealand
##                       2                       2
##                   Niger                 Nigeria
##                       1                       1
##                  Norway                    Oman
##                       2                       2
##                Pakistan                  Panama
##                       1                       2
##                Paraguay                    Peru
##                       2                       2
##             Philippines                  Poland
##                       1                       2
##                Portugal                   Qatar
##                       2                       2
##                 Romania                  Russia
##                       2                       2
##                  Rwanda                   Samoa
##                       1                       2
##            Saudi Arabia                 Senegal
##                       2                       1
##                  Serbia              Seychelles
##                       2                       2
##            Sierra Leone               Singapore
##                       1                       2
##          Slovak Republic                Slovenia
##                       2                       2
##          Solomon Islands            South Africa
##                       1                       1
```

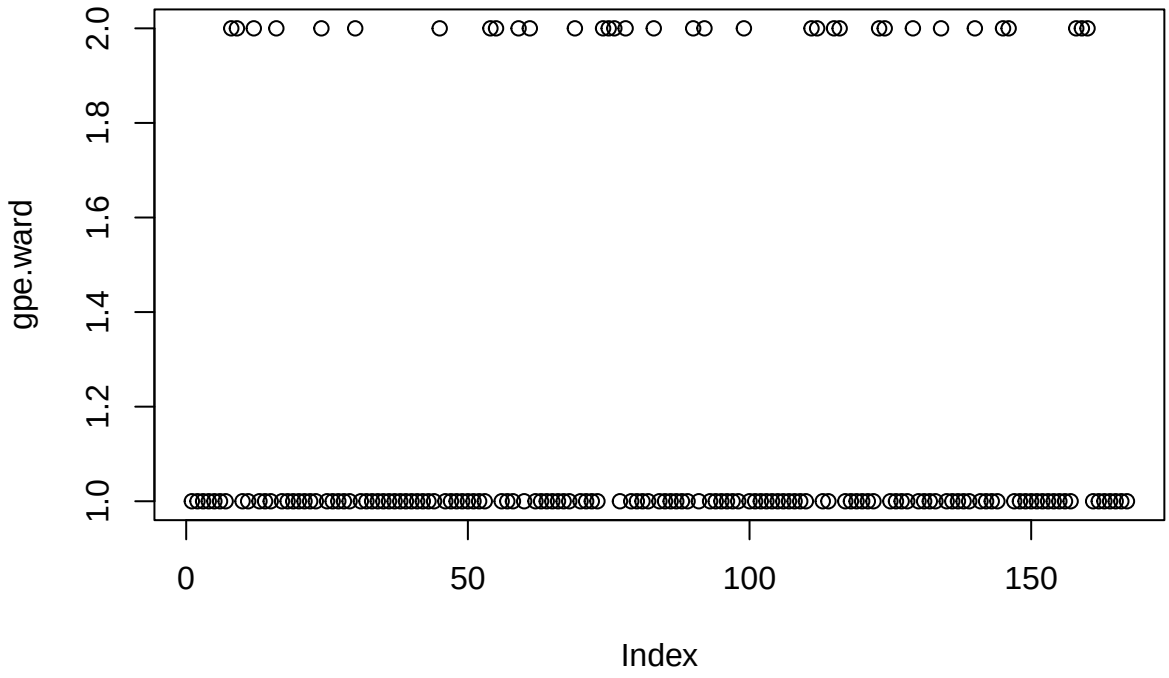
```
##                South Korea                       Spain
##                          2                           2
##                  Sri Lanka St. Vincent and the Grenadines
##                          2                           2
##                      Sudan                    Suriname
##                          1                           2
##                     Sweden                 Switzerland
##                          2                           2
##                 Tajikistan                    Tanzania
##                          1                           1
##                   Thailand                 Timor-Leste
##                          2                           1
##                       Togo                       Tonga
##                          1                           2
##                    Tunisia                      Turkey
##                          2                           2
##                Turkmenistan                      Uganda
##                          1                           1
##                    Ukraine        United Arab Emirates
##                          2                           2
##             United Kingdom               United States
##                          2                           2
##                    Uruguay                  Uzbekistan
##                          2                           1
##                    Vanuatu                   Venezuela
##                          1                           2
##                    Vietnam                       Yemen
##                          2                           1
##                     Zambia
##                          1
```

C'est aussi une partition en 2 groupes que l'on obtient.

Représentation graphique des clusters avec la fonction cutree :

La fonction cutree permet de faire apparaitre visuellement les groupes. Dans notre cas, on fixe K = 2 car nous avons choisi de réaliser une partition en 2 groupes.

```
K = 2
gpe.ward = cutree(CAH_ward,k=K)
plot(gpe.ward)
```

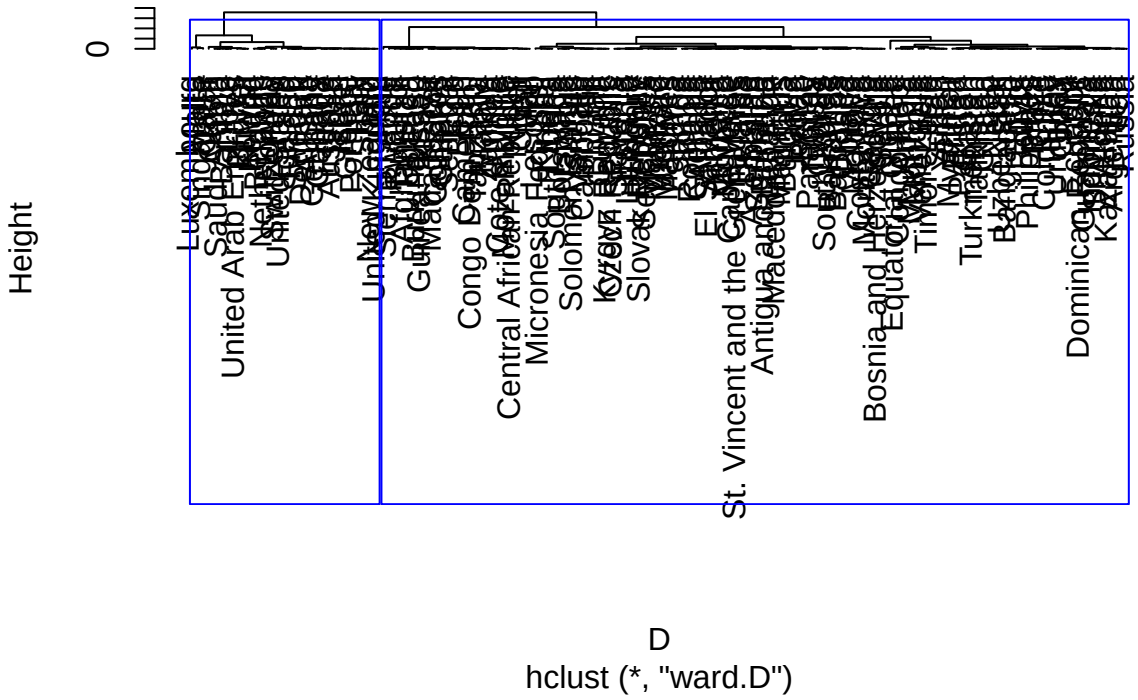



```
# gpe.ward
```

Sur ce graphique, on remarque correctement 2 groupes distincts.

Représentation graphique des clusters avec un dendrogramme :

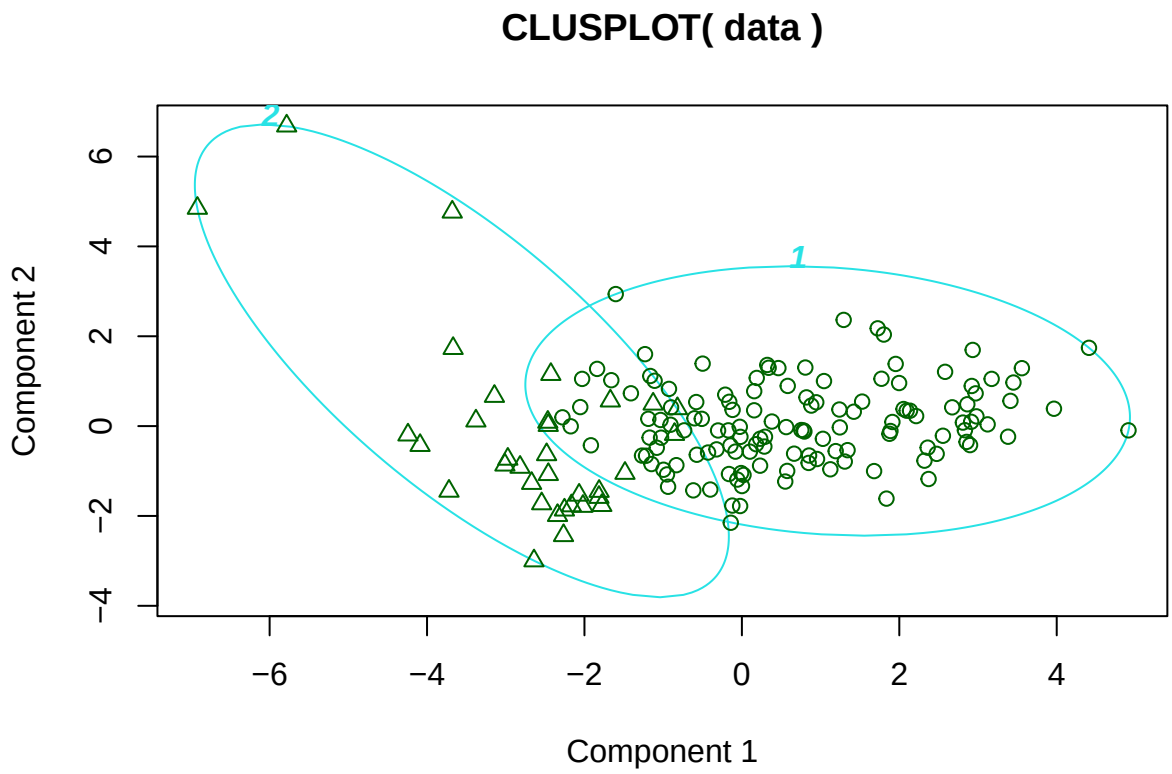
```
K=2
plot(CAH_ward,hang=-1)
rect.hclust(CAH_ward,K,border="blue")
```



Représentation des groupes avec la fonction clusplot :

PROBLEME : LA LIGNE EN # POSE PROBLEME PQ ??

```
# data$gpe.ward <- gpe.ward
clusplot(data,cutree(CAH_ward,K),labels=4)
```



Ce graphe correspond à la représentation des groupes sur les deux premiers axes principaux d'une ACP. De plus, des ellipses de contour autour des groupes sont tracées. Ici, nous pouvons voir 2 groupes.

Deuxième approche : Agrégation autour de centres mobiles

La fonction kmeans donne le résultat de l'algorithme d'agrégation autour des centres mobiles.

```
K = 2 # 2 groupes
c <- kmeans(data,K,nstart=50)
c
```

```
## K-means clustering with 2 clusters of sizes 68, 99
##
## Cluster means:
##   enfant_mort    exports      sante     imports     revenu  inflation
## 1   0.9425200 -0.3980176 -0.2641371 -0.13416458 -0.6700512  0.3137840
## 2  -0.6473874  0.2733858  0.1814275  0.09215345  0.4602372 -0.2155284
##    esper_vie       fert      pib_h
## 1 -0.9721385  0.9683006 -0.5992197
## 2  0.6677315 -0.6650953  0.4115852
##
## Clustering vector:
##                     Afghanistan                         Albania
##                               1                               2
##                         Algeria                          Angola
```

```
##                         2                         1
##          Antigua and Barbuda                 Argentina
##                         2                         2
##                   Armenia                 Australia
##                         2                         2
##                   Austria                Azerbaijan
##                         2                         2
##                   Bahamas                   Bahrain
##                         2                         2
##                Bangladesh                  Barbados
##                         1                         2
##                   Belarus                   Belgium
##                         2                         2
##                    Belize                     Benin
##                         2                         1
##                    Bhutan                   Bolivia
##                         2                         1
##       Bosnia and Herzegovina                  Botswana
##                         2                         1
##                    Brazil                    Brunei
##                         2                         2
##                  Bulgaria              Burkina Faso
##                         2                         1
##                   Burundi                  Cambodia
##                         1                         1
##                  Cameroon                    Canada
##                         1                         2
##                Cape Verde  Central African Republic
##                         2                         1
##                      Chad                     Chile
##                         1                         2
##                     China                  Colombia
##                         2                         2
##                   Comoros            Congo Dem. Rep.
##                         1                         1
##                Congo Rep.                Costa Rica
##                         1                         2
##             Cote d'Ivoire                   Croatia
##                         1                         2
##                    Cyprus             Czech Republic
##                         2                         2
##                   Denmark          Dominican Republic
##                         2                         2
##                   Ecuador                     Egypt
##                         2                         1
##               El Salvador           Equatorial Guinea
##                         2                         1
##                   Eritrea                   Estonia
##                         1                         2
##                      Fiji                   Finland
##                         2                         2
##                    France                     Gabon
##                         2                         1
##                    Gambia                   Georgia
```

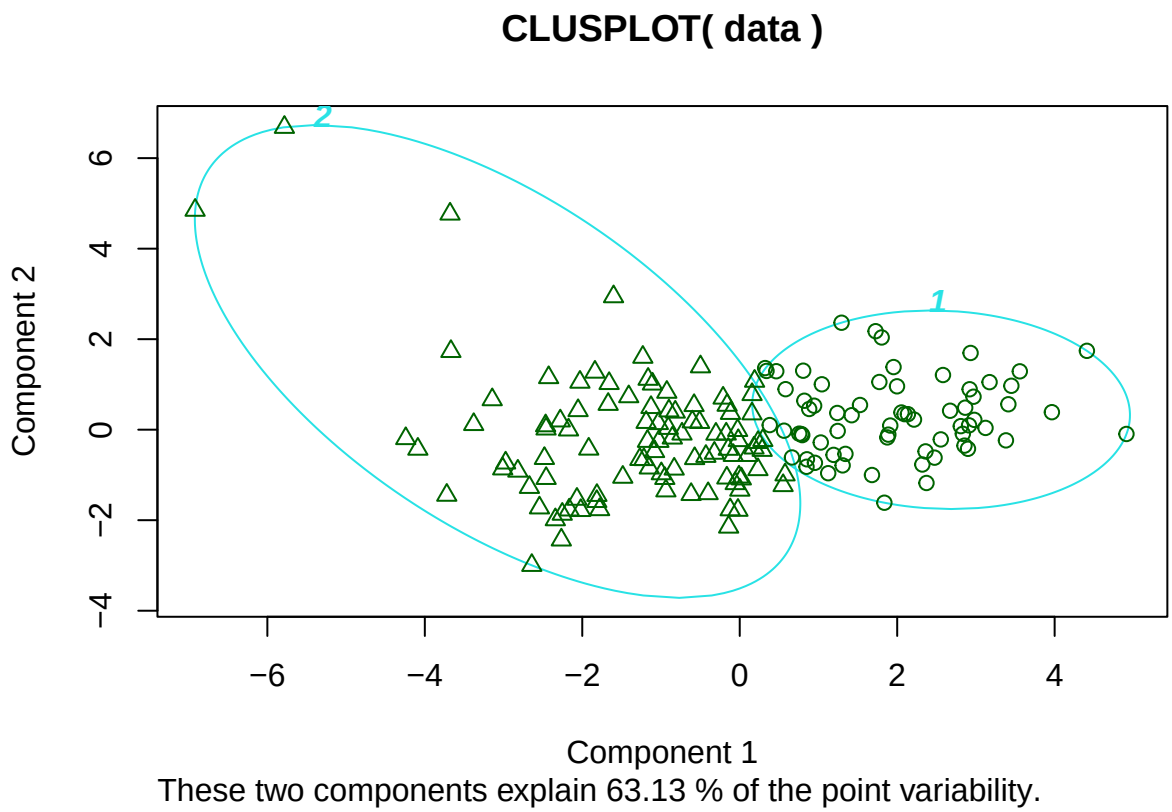
```
##                      1                      2
##                Germany                  Ghana
##                      2                      1
##                 Greece                Grenada
##                      2                      2
##              Guatemala                 Guinea
##                      1                      1
##          Guinea-Bissau                 Guyana
##                      1                      1
##                  Haiti                Hungary
##                      1                      2
##                Iceland                  India
##                      2                      1
##              Indonesia                   Iran
##                      1                      2
##                   Iraq                Ireland
##                      1                      2
##                 Israel                  Italy
##                      2                      2
##                Jamaica                  Japan
##                      2                      2
##                 Jordan             Kazakhstan
##                      2                      2
##                  Kenya               Kiribati
##                      1                      1
##                 Kuwait         Kyrgyz Republic
##                      2                      1
##                    Lao                 Latvia
##                      1                      2
##                Lebanon                Lesotho
##                      2                      1
##                Liberia                  Libya
##                      1                      2
##              Lithuania             Luxembourg
##                      2                      2
##          Macedonia FYR             Madagascar
##                      2                      1
##                 Malawi               Malaysia
##                      1                      2
##                Maldives                   Mali
##                      2                      1
##                  Malta             Mauritania
##                      2                      1
##              Mauritius    Micronesia Fed. Sts.
##                      2                      1
##                Moldova               Mongolia
##                      2                      1
##             Montenegro                Morocco
##                      2                      2
##             Mozambique                Myanmar
##                      1                      1
##                Namibia                  Nepal
##                      1                      1
##            Netherlands            New Zealand
```

```
##                         2                         2
##                     Niger                   Nigeria
##                         1                         1
##                    Norway                      Oman
##                         2                         2
##                  Pakistan                    Panama
##                         1                         2
##                  Paraguay                      Peru
##                         2                         2
##               Philippines                    Poland
##                         1                         2
##                  Portugal                     Qatar
##                         2                         2
##                   Romania                    Russia
##                         2                         2
##                    Rwanda                     Samoa
##                         1                         1
##              Saudi Arabia                   Senegal
##                         2                         1
##                    Serbia                  Seychelles
##                         2                         2
##              Sierra Leone                 Singapore
##                         1                         2
##            Slovak Republic                  Slovenia
##                         2                         2
##            Solomon Islands              South Africa
##                         1                         1
##                South Korea                     Spain
##                         2                         2
##                 Sri Lanka St. Vincent and the Grenadines
##                         2                         2
##                     Sudan                  Suriname
##                         1                         2
##                    Sweden                Switzerland
##                         2                         2
##                Tajikistan                   Tanzania
##                         1                         1
##                  Thailand                Timor-Leste
##                         2                         1
##                      Togo                     Tonga
##                         1                         1
##                   Tunisia                    Turkey
##                         2                         2
##              Turkmenistan                    Uganda
##                         1                         1
##                   Ukraine      United Arab Emirates
##                         2                         2
##            United Kingdom             United States
##                         2                         2
##                   Uruguay                Uzbekistan
##                         2                         1
##                   Vanuatu                 Venezuela
##                         1                         2
##                   Vietnam                     Yemen
```

```
##                              2                              1
##                         Zambia
##                              1
##
## Within cluster sum of squares by cluster:
## [1] 400.5512 643.3747
##  (between_SS / total_SS =  30.1 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

La fonction **kmeans** nous permet d'obtenir le partitionnement final. Dans notre cas, elle nous rend 2 clusters composés de 135 et 32 pays. Ici, nous avons initialisé nstart à 50 pour répéter la procédure plusieurs fois et garder la partition avec la plus faible inertie intra classes. On peut en déduire qu'on a un groupe nécessiteux et l'autre plus aisé.

```
clusplot(data,c$cluster,labels=4)
```



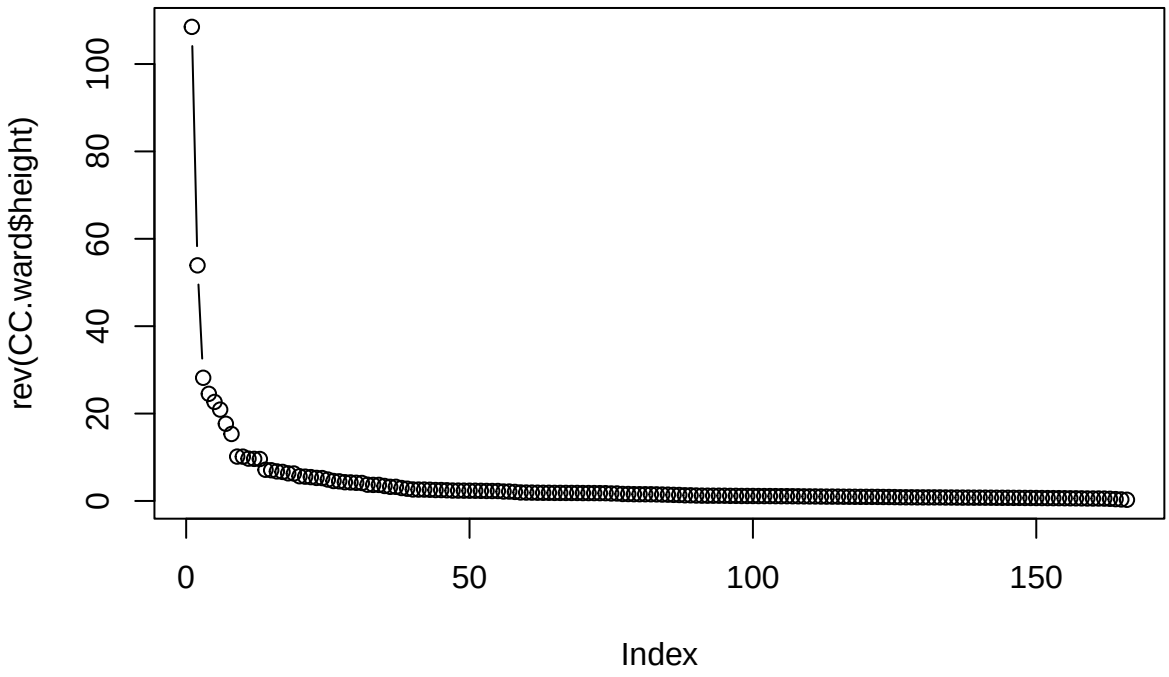
Autre approche

QUELLE EST LA DIFFERENCE AVEC NOTRE 1ERE APPROCHE ??

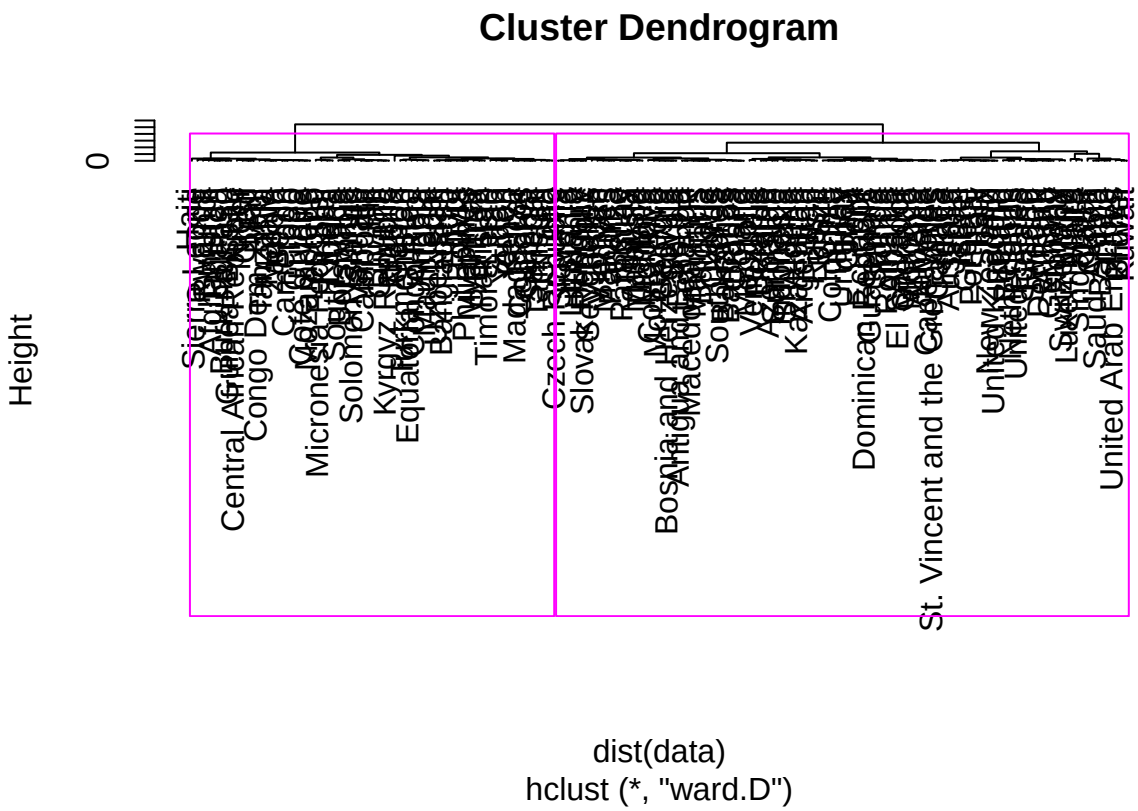
Calcul de la matrice distance avec la dissimilarité basée sur la corrélation.

```
CorDist = as.dist((1-cor(t(data)))/2)
```

```
CC.ward = hclust(dist(data),method="ward.D")
plot(rev(CC.ward$height),type="b")
```




```
plot(CC.ward,hang=-1)
rect.hclust(CC.ward, 2, border ="magenta")
```




```
# gpe = cutree(CC.ward,k=2)
# data$gpecah = as.factor(gpe)
# interpcah = catdes(data,num.var = 8)
# interpcah
```