Evaluation 2 - Rapport

Technologie de l'e-commerce et mobiles - Big Data

Khalladi Mohamed - B32-DA

2024-01-07

Contents

1 ANOVA 2

1.1 Bière et petits maux

L'Administration de la Santé Publique de Bidendumie a recensé le nombre de patients atteints de l'une des 4 maladies bénignes les plus fréquentes et ayant consommé l'une des 3 bières locales les plus répandues. Elle a mesuré un coefficient biochimique représentatif sur 6 patients (si possible) choisis aléatoirement.

Est-il possible d'interpréter de tels résultats ?

Nous allons former notre data-set et vérifier qu'il soit bien formé.

```
setwd("C:\\Users\\amine\\OneDrive\\Bureau\\EcomStat\\Labo\\Evaluation02\\datasets")

dataBiere <- read.csv("bieres_petits_maux.csv", h=TRUE, sep=";", fileEncoding="latin1")
dataBiere
```

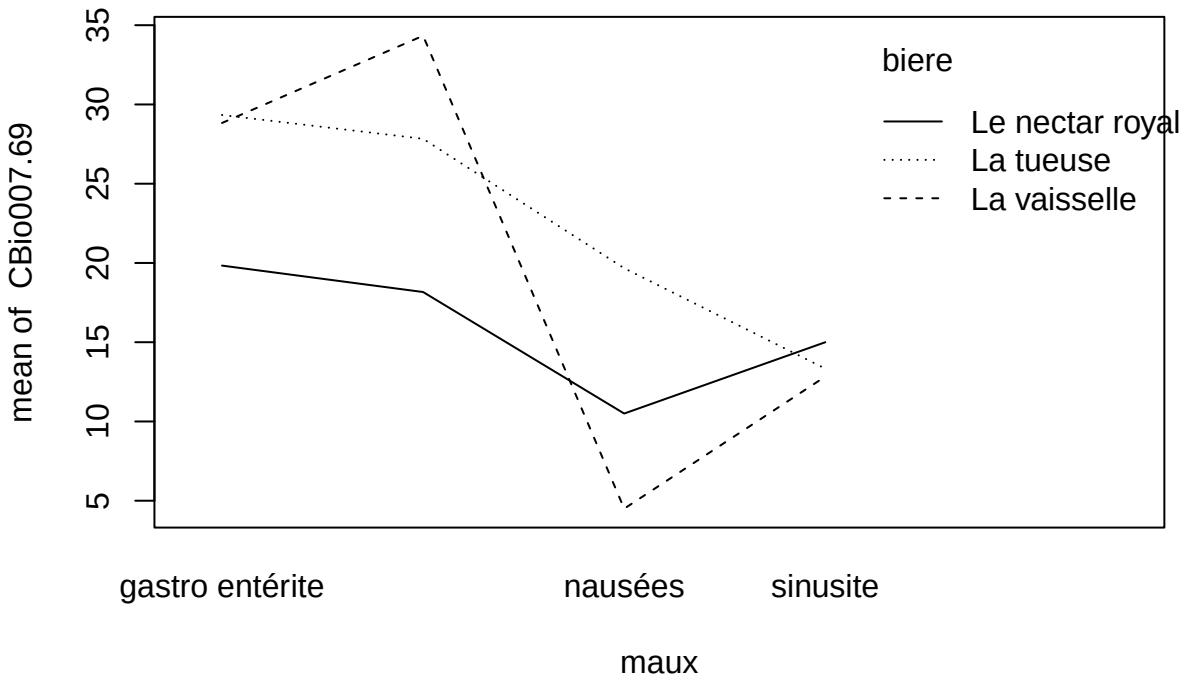
```
##    CBio007.69            maux           biere
## 1          42 gastro entérite       La tueuse
## 2          28     mal de tête       La tueuse
## 3           1         nausées       La tueuse
## 4          24        sinusite       La tueuse
## 5          44 gastro entérite       La tueuse
## 6          23     mal de tête       La tueuse
## 7          29         nausées       La tueuse
## 8           9        sinusite       La tueuse
## 9          36 gastro entérite       La tueuse
## 10         34     mal de tête       La tueuse
## 11         19         nausées       La tueuse
## 12         22        sinusite       La tueuse
## 13         13 gastro entérite       La tueuse
## 14         42     mal de tête       La tueuse
## 15         29         nausées       La tueuse
## 16         -2        sinusite       La tueuse
## 17         19 gastro entérite       La tueuse
## 18         13     mal de tête       La tueuse
## 19         18         nausées       La tueuse
## 20         15        sinusite       La tueuse
## 21         22 gastro entérite       La tueuse
## 22         27     mal de tête       La tueuse
## 23         22         nausées       La tueuse
## 24         12        sinusite       La tueuse
## 25         33 gastro entérite    La vaisselle
## 26         34     mal de tête    La vaisselle
## 27         11         nausées    La vaisselle
## 28         27        sinusite    La vaisselle
## 29         26 gastro entérite    La vaisselle
## 30         33     mal de tête    La vaisselle
## 31          9         nausées    La vaisselle
## 32         12        sinusite    La vaisselle
## 33         33 gastro entérite    La vaisselle
## 34         31     mal de tête    La vaisselle
```

```
## 35          7         nausées    La vaisselle
## 36         12        sinusite    La vaisselle
## 37         21 gastro entérite    La vaisselle
## 38         36     mal de tête    La vaisselle
## 39          1         nausées    La vaisselle
## 40         -5        sinusite    La vaisselle
## 41         29 gastro entérite    La vaisselle
## 42         34     mal de tête    La vaisselle
## 43         -6         nausées    La vaisselle
## 44         16        sinusite    La vaisselle
## 45         31 gastro entérite    La vaisselle
## 46         38     mal de tête    La vaisselle
## 47          5         nausées    La vaisselle
## 48         15        sinusite    La vaisselle
## 49         31 gastro entérite Le nectar royal
## 50          3     mal de tête Le nectar royal
## 51         21         nausées Le nectar royal
## 52         22        sinusite Le nectar royal
## 53         -3 gastro entérite Le nectar royal
## 54         26     mal de tête Le nectar royal
## 55          1         nausées Le nectar royal
## 56          7        sinusite Le nectar royal
## 57         25 gastro entérite Le nectar royal
## 58         28     mal de tête Le nectar royal
## 59          9         nausées Le nectar royal
## 60         25        sinusite Le nectar royal
## 61         25 gastro entérite Le nectar royal
## 62         32     mal de tête Le nectar royal
## 63          3         nausées Le nectar royal
## 64          5        sinusite Le nectar royal
## 65         24 gastro entérite Le nectar royal
## 66          4     mal de tête Le nectar royal
## 67         12         nausées Le nectar royal
## 68         12        sinusite Le nectar royal
## 69         17 gastro entérite Le nectar royal
## 70         16     mal de tête Le nectar royal
## 71         17         nausées Le nectar royal
## 72         19        sinusite Le nectar royal
```

```r
dataBiere$maux <- as.factor(dataBiere$maux)
dataBiere$biere <- as.factor(dataBiere$biere)
summary(dataBiere)
```

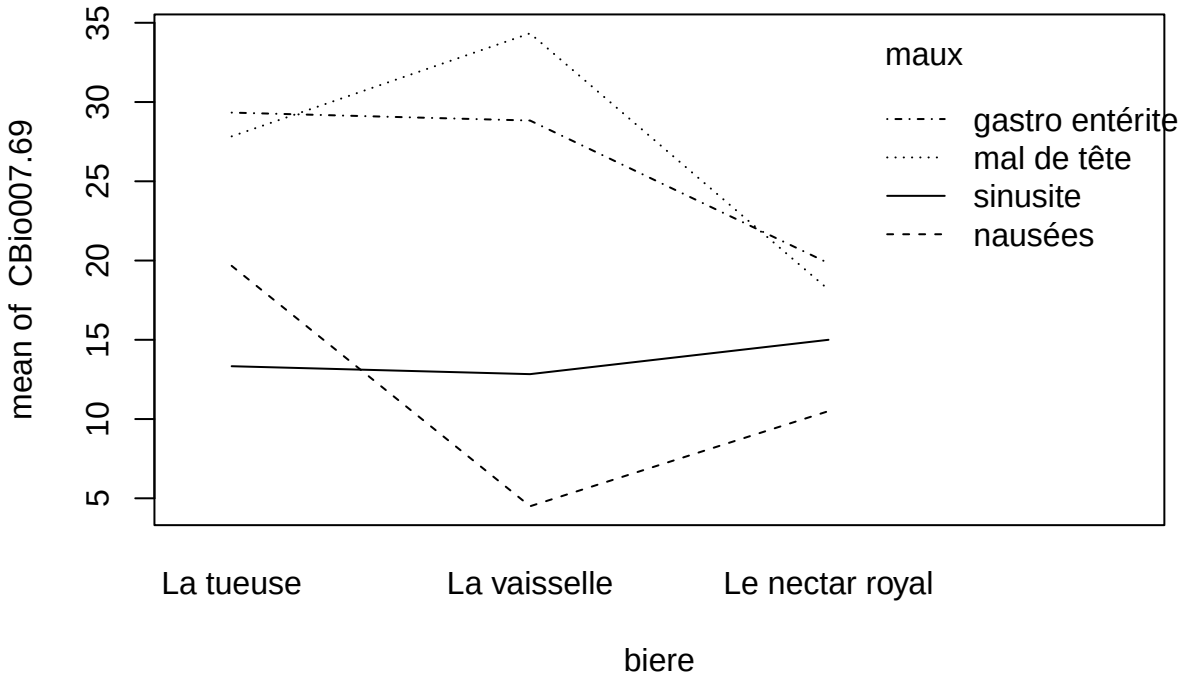
```
##    CBio007.69                maux                biere
##  Min.   :-6.00   gastro entérite:18   La tueuse      :24
##  1st Qu.:11.75   mal de tête    :18   La vaisselle   :24
##  Median :21.00   nausées        :18   Le nectar royal:24
##  Mean   :19.51   sinusite       :18
##  3rd Qu.:29.00
##  Max.   :44.00
```

```r
#Ce graphique montre comment le coefficient biochimique "CBio007-69" varie avec
#le type de maladie ("maux") pour chaque type de bière.
with(dataBiere, interaction.plot(maux, biere, CBio007.69))
```



Chaque ligne du graphique représente un niveau du facteur “biere”, et les points sur les lignes représentent les moyennes du coefficient biochimique pour chaque combinaison de “biere” et “maux”.
La non-parallélité des lignes suggère une interaction entre les types de bières et les maux en termes de coefficient biochimique.

```
#Ce graphique montre comment le coefficient biochimique varie avec
#le type de bière pour chaque type de maladie.
with(dataBiere, interaction.plot(biere, maux, CBio007.69))
```



Chaque ligne du graphique représente un niveau du facteur “maux”, et les points sur les lignes représentent les moyennes du coefficient biochimique pour chaque combinaison de “biere” et “maux”.
Les lignes qui ne sont pas parallèles indiquent également une interaction entre le type de bière et le type de mal.

On va créer le modèle croisé pour pouvoir appliquer l’anova.

```
#y(ijk) = mu + alpha + beta(i) + gamma(j) + epsilon(ijk)

##Modèle avec interaction
modele_croise = lm(CBio007.69 ~ maux * biere, data = dataBiere)
modele_croise
```

```
##
## Call:
## lm(formula = CBio007.69 ~ maux * biere, data = dataBiere)
##
## Coefficients:
##                            (Intercept)                          mauxmal de tête
##                              2.933e+01                               -1.500e+00
##                           mauxnausées                             mauxsinusite
##                              -9.667e+00                               -1.600e+01
##                      biereLa vaisselle                    biereLe nectar royal
##                              -5.000e-01                               -9.500e+00
##    mauxmal de tête:biereLa vaisselle          mauxnausées:biereLa vaisselle
##                              7.000e+00                               -1.467e+01
##       mauxsinusite:biereLa vaisselle  mauxmal de tête:biereLe nectar royal
```

```
##                             -1.305e-14                              -1.667e-01
##     mauxnausées:biereLe nectar royal     mauxsinusite:biereLe nectar royal
##                              3.333e-01                               1.117e+01
```

(Mu): Le terme (Intercept) qui est de 29.33, représente la moyenne estimée du coefficient biochimique pour la catégorie de référence des maux et des bières (alpha1 = 0 et beta1 = 0).

(Alpha): Les coefficients liés à "maux" (par exemple, mauxsinusite de -16.00) représentent l'effet de chaque maladie sur le coefficient biochimique par rapport à la maladie de référence.

(Beta): Les coefficients liés à "biere" (par exemple, biereLe nectar royal de -9.50) indiquent l'effet de chaque type de bière sur le coefficient biochimique par rapport à la bière de référence.

(Gamma): Les coefficients d'interaction (par exemple, mauxmal de tête:biereLa vaisselle de 7.00) montrent l'effet combiné d'un certain mal avec une certaine bière sur le coefficient biochimique.

En d'autres termes, l'Intercept est notre point de départ, les coefficients alpha et beta nous disent comment chaque facteur change ce point de départ individuellement, et les coefficients gamma nous montrent ce qui se passe quand ces facteurs interagissent et se combinent de manière unique.

Effets principaux du facteur "maux":
H0: Il n'y a pas de différence dans les moyennes du coefficient biochimique entre les différents types de "maux".
H1: Il existe au moins une différence dans les moyennes du coefficient biochimique entre les différents types de "maux".

Effets principaux du facteur "biere": H0: Il n'y a pas de différence dans les moyennes du coefficient biochimique entre les différentes bières.
H1: Il existe au moins une différence dans les moyennes du coefficient biochimique entre les différentes bières.

Interaction entre "maux" et "biere":
H0: Il n'y a pas d'interaction entre les "maux" et les "bières", c'est-à-dire que l'effet d'un "mal" sur le coefficient biochimique est le même pour toutes les "bières".
H1: Il existe une interaction entre les "maux" et les "bières", c'est-à-dire que l'effet d'un "mal" sur le coefficient biochimique change selon la "bière" consommée.

```
anova(modele_croise)
```

```
## Analysis of Variance Table
##
## Response: CBio007.69
##            Df Sum Sq Mean Sq F value    Pr(>F)
## maux        3 3450.8 1150.27 12.9620 1.243e-06 ***
## biere       2  546.8  273.39  3.0807   0.05326 .
## maux:biere  6 1305.9  217.65  2.4526   0.03461 *
## Residuals  60 5324.5   88.74
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pour les effets principaux du facteur "maux": Pour un seuil de 1% ou 5% on rejette H0, donc :
H1: Il existe au moins une différence dans les moyennes du coefficient biochimique entre les différents types de "maux".

Pour les effets principaux du facteur "biere":
Pour un seuil de 1% ou 5% on garde H0, donc :
H0: Il n'y a pas de différence dans les moyennes du coefficient biochimique entre les différentes bières.

Pour l'interaction entre "maux" et "biere":
Pour un seuil de 1% on garde H0 mais pour 5% on rejette H0, donc :
H1: Il existe une interaction entre les "maux" et les "bières", c'est-à-dire que l'effet d'un "mal" sur le coefficient biochimique change selon la "bière" consommée.

```
summary(modele_croise)
```

```
##
## Call:
## lm(formula = CBio007.69 ~ maux * biere, data = dataBiere)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.8333  -3.7083   0.3333   6.2500  14.6667
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           2.933e+01  3.846e+00   7.627 2.12e-10 ***
## mauxmal de tête                      -1.500e+00  5.439e+00  -0.276  0.78365
## mauxnausées                          -9.667e+00  5.439e+00  -1.777  0.08058 .
## mauxsinusite                         -1.600e+01  5.439e+00  -2.942  0.00463 **
## biereLa vaisselle                    -5.000e-01  5.439e+00  -0.092  0.92706
## biereLe nectar royal                 -9.500e+00  5.439e+00  -1.747  0.08581 .
## mauxmal de tête:biereLa vaisselle     7.000e+00  7.692e+00   0.910  0.36642
## mauxnausées:biereLa vaisselle        -1.467e+01  7.692e+00  -1.907  0.06133 .
## mauxsinusite:biereLa vaisselle       -1.305e-14  7.692e+00   0.000  1.00000
## mauxmal de tête:biereLe nectar royal -1.667e-01  7.692e+00  -0.022  0.98278
## mauxnausées:biereLe nectar royal      3.333e-01  7.692e+00   0.043  0.96558
## mauxsinusite:biereLe nectar royal     1.117e+01  7.692e+00   1.452  0.15177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.42 on 60 degrees of freedom
## Multiple R-squared:  0.499,  Adjusted R-squared:  0.4072
## F-statistic: 5.433 on 11 and 60 DF,  p-value: 6.31e-06
```

rapport de corrélation = pourcentage de la variance expliquée par le modèle (donc toutes les contributions sauf la variance résiduelle)
R squared : proportion de la SCEf par rapport à la SCEt

La p-value nous indique ici qu'il pourrait y avoir une influence avec la sinusite et le coefficient biochimique vue que la p-value est très faible. Les autres en prenant un seuil de 5% n'auraient pas d'influence.

Pourtant, la p-value du modele est très faible également donc ça veut dire qu'il y a de l'interaction mais il faut en trouver plus. C'est pour cela que l'on va chercher à utiliser le modèle hiérarchisé pour rechercher d'autres interactions.

```
#y(ijk) = mu + alpha + beta(i) + gamma(j) + epsilon(ijk)
#Modèle sans interaction
modele_hierarchise = lm(CBio007.69 ~ maux + biere, data = dataBiere)
modele_hierarchise
```

```
##
## Call:
```

```
## lm(formula = CBio007.69 ~ maux + biere, data = dataBiere)
##
## Coefficients:
##          (Intercept)       mauxmal de tête           mauxnausées
##              29.0278                0.7778             -14.4444
##         mauxsinusite     biereLa vaisselle  biereLe nectar royal
##             -12.2778               -2.4167              -6.6667
```

```
anova(modele_hierarchise)
```

```
## Analysis of Variance Table
##
## Response: CBio007.69
##           Df Sum Sq Mean Sq F value    Pr(>F)
## maux       3 3450.8 1150.27 11.4500 3.895e-06 ***
## biere      2  546.8  273.39  2.7214   0.07317 .
## Residuals 66 6630.4  100.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ici, on voit que la bière a une p-value supérieure à 7% , ce qui vaut dire qu'on pourrait accepter le H0 avec un seuil de 5% , donc que la bière n'aurait pas d'influence sur le coefficient biochimique.

```
summary(modele_hierarchise)
```

```
##
## Call:
## lm(formula = CBio007.69 ~ maux + biere, data = dataBiere)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.361  -5.424   1.792   6.701  14.972
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           29.0278     2.8934  10.032 6.63e-15 ***
## mauxmal de tête        0.7778     3.3410   0.233 0.816639
## mauxnausées          -14.4444     3.3410  -4.323 5.30e-05 ***
## mauxsinusite         -12.2778     3.3410  -3.675 0.000479 ***
## biereLa vaisselle     -2.4167     2.8934  -0.835 0.406598
## biereLe nectar royal  -6.6667     2.8934  -2.304 0.024373 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.02 on 66 degrees of freedom
## Multiple R-squared:  0.3761, Adjusted R-squared:  0.3289
## F-statistic: 7.959 on 5 and 66 DF,  p-value: 6.466e-06
```

Grace à ce modele hierarchise, on voit que les sinusites ont toujours de l'influence mais que les nausées ont également une influence sur le coefficient biochimique ainsi que la bière « Le nectar Royal » (pour un seuil de 5%).

1.2 Les médicaments contre la GCE

Une entreprise pharmaceutique s'intéresse à une maladie tropicale (la Gengivite Cephalopodique Endiablée - GCE) et a mis au point trois molécules susceptibles de soigner cette maladie : AlphaVictoire, BetaTriomphe et GammaSucces. Les tests cliniques ont été pratiqués pour mesurer un coefficient relatif d'amélioration de l'état de patients gravement atteints (plus ce coefficient d'immunité est élevé et plus l'action sera considérée comme efficace). Mais, de plus, on souhaite également tenir compte du mode d'administration des différentes molécules (par voie orale ou par injection intraveineuse).

Observe-t-on une différence significative d'efficacité soit selon la molécule, soit selon le mode d'administration ou encore selon une combinaison des deux facteurs ?

Nous allons former notre data-set et vérifier qu'il soit bien formé.

```
#Les médicaments contre la GCE
setwd("C:\\Users\\amine\\OneDrive\\Bureau\\EcomStat\\Labo\\Evaluation02\\datasets")

dataMedicament <- data.frame(
  Amelioration = c(10, 12, 8, 10, 6, 13, 9, 10, 9, 8, 11, 18, 12, 15, 13, 8, 15, 16, 9, 13,
                   7, 14, 10, 11, 9, 10, 11, 7, 9, 9, 8, 9, 10, 9, 11, 13, 7, 14, 15, 12,
                   12, 9, 11, 27, 7, 8, 13, 14, 10, 11, 7, 6, 10, 7, 7, 5, 6, 7, 9, 6),
  Molecule = rep(c("AlphaVictoire", "BetaTriomphe", "GammaSucces"), each = 20),
  Administration = rep(c("Oral", "Injection"), each = 10, times = 3)
)
dataMedicament
```

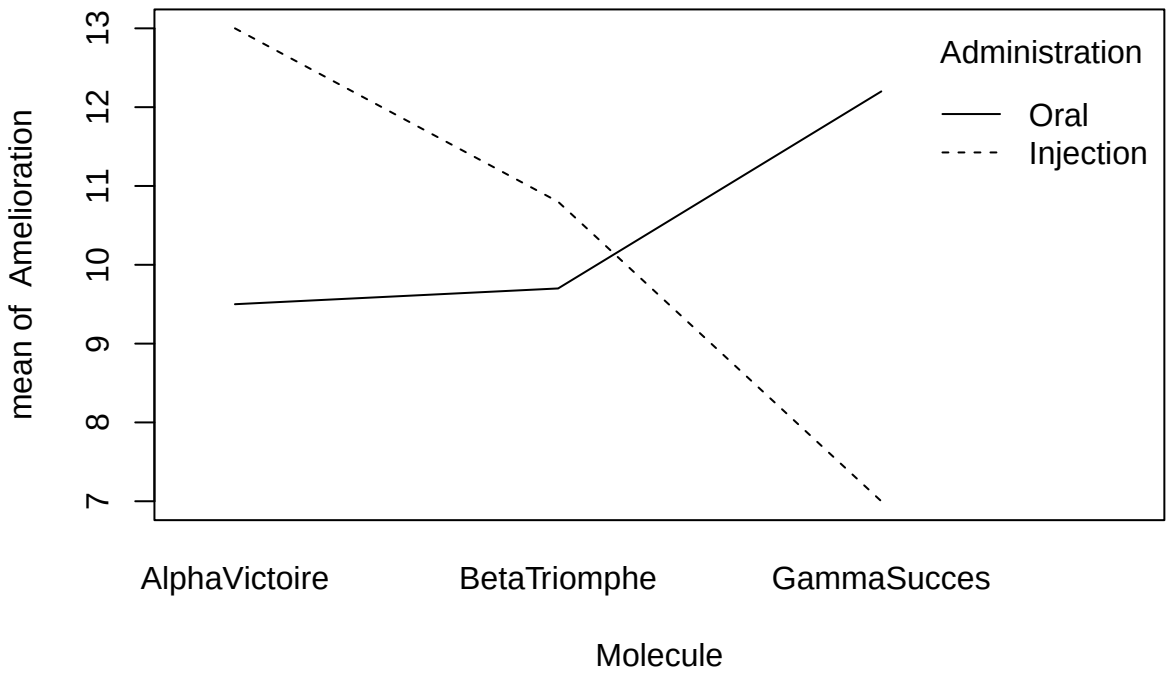
```
##    Amelioration      Molecule Administration
## 1            10 AlphaVictoire           Oral
## 2            12 AlphaVictoire           Oral
## 3             8 AlphaVictoire           Oral
## 4            10 AlphaVictoire           Oral
## 5             6 AlphaVictoire           Oral
## 6            13 AlphaVictoire           Oral
## 7             9 AlphaVictoire           Oral
## 8            10 AlphaVictoire           Oral
## 9             9 AlphaVictoire           Oral
## 10            8 AlphaVictoire           Oral
## 11           11 AlphaVictoire      Injection
## 12           18 AlphaVictoire      Injection
## 13           12 AlphaVictoire      Injection
## 14           15 AlphaVictoire      Injection
## 15           13 AlphaVictoire      Injection
## 16            8 AlphaVictoire      Injection
## 17           15 AlphaVictoire      Injection
## 18           16 AlphaVictoire      Injection
## 19            9 AlphaVictoire      Injection
## 20           13 AlphaVictoire      Injection
## 21            7  BetaTriomphe           Oral
## 22           14  BetaTriomphe           Oral
## 23           10  BetaTriomphe           Oral
## 24           11  BetaTriomphe           Oral
## 25            9  BetaTriomphe           Oral
## 26           10  BetaTriomphe           Oral
```

```
## 27           11  BetaTriomphe           Oral
## 28            7  BetaTriomphe           Oral
## 29            9  BetaTriomphe           Oral
## 30            9  BetaTriomphe           Oral
## 31            8  BetaTriomphe      Injection
## 32            9  BetaTriomphe      Injection
## 33           10  BetaTriomphe      Injection
## 34            9  BetaTriomphe      Injection
## 35           11  BetaTriomphe      Injection
## 36           13  BetaTriomphe      Injection
## 37            7  BetaTriomphe      Injection
## 38           14  BetaTriomphe      Injection
## 39           15  BetaTriomphe      Injection
## 40           12  BetaTriomphe      Injection
## 41           12   GammaSucces           Oral
## 42            9   GammaSucces           Oral
## 43           11   GammaSucces           Oral
## 44           27   GammaSucces           Oral
## 45            7   GammaSucces           Oral
## 46            8   GammaSucces           Oral
## 47           13   GammaSucces           Oral
## 48           14   GammaSucces           Oral
## 49           10   GammaSucces           Oral
## 50           11   GammaSucces           Oral
## 51            7   GammaSucces      Injection
## 52            6   GammaSucces      Injection
## 53           10   GammaSucces      Injection
## 54            7   GammaSucces      Injection
## 55            7   GammaSucces      Injection
## 56            5   GammaSucces      Injection
## 57            6   GammaSucces      Injection
## 58            7   GammaSucces      Injection
## 59            9   GammaSucces      Injection
## 60            6   GammaSucces      Injection
```

```
dataMedicament$Molecule <- as.factor(dataMedicament$Molecule)
dataMedicament$Administration <- as.factor(dataMedicament$Administration)
summary(dataMedicament)
```

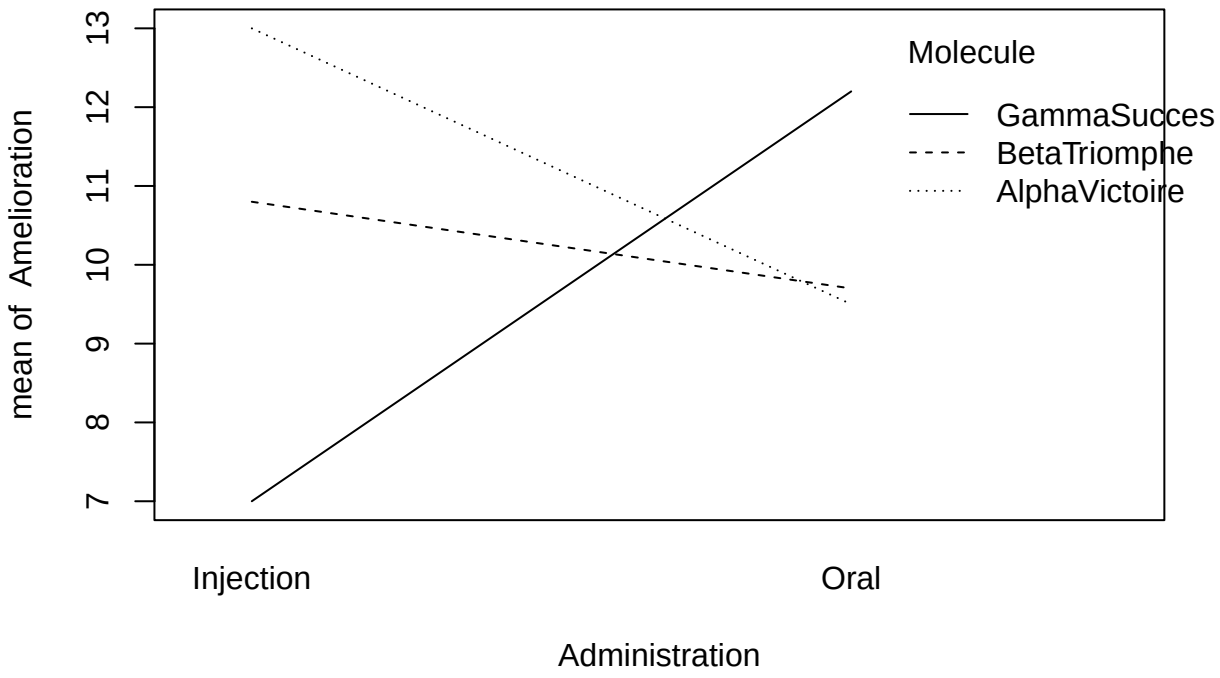
```
##   Amelioration            Molecule    Administration
##  Min.   : 5.00   AlphaVictoire:20   Injection:30
##  1st Qu.: 8.00   BetaTriomphe :20   Oral     :30
##  Median :10.00   GammaSucces  :20
##  Mean   :10.37
##  3rd Qu.:12.00
##  Max.   :27.00
```

```
#Ce graphique montre comment le coefficient d'immunité varie avec
#le type de molecule pour chaque type d'administration.
with(dataMedicament, interaction.plot(Molecule, Administration, Amelioration))
```



Chaque ligne du graphique représente un niveau du facteur “Administration”, et les points sur les lignes représentent les moyennes du coefficient d’immunité pour chaque combinaison de “Administration” et “Molecule”.
La croisement des lignes suggère une interaction entre les types des administrations et les molecules en termes de coefficient d’immunité.

```
#Ce graphique montre comment le coefficient d'immunité varie avec
#le type d'administration pour chaque type de molecule.
with(dataMedicament, interaction.plot(Administration, Molecule, Amelioration))
```



Chaque ligne du graphique représente un niveau du facteur “Molecule”, et les points sur les lignes représentent les moyennes du coefficient d’immunité pour chaque combinaison de “Molecule” et “Administration”. Les lignes qui se croisent indiquent également une interaction entre le type de molecule et le type d’administration.

```
modele_croise <- lm(Amelioration ~ Molecule * Administration, data = dataMedicament)
modele_croise
```

```
##
## Call:
## lm(formula = Amelioration ~ Molecule * Administration, data = dataMedicament)
##
## Coefficients:
##                                 (Intercept)
##                                        13.0
##                         MoleculeBetaTriomphe
##                                        -2.2
##                          MoleculeGammaSucces
##                                        -6.0
##                           AdministrationOral
##                                        -3.5
## MoleculeBetaTriomphe:AdministrationOral
##                                         2.4
##  MoleculeGammaSucces:AdministrationOral
##                                         8.7
```

Effets principaux du facteur “Molecule”:

H0: Il n'y a pas de différence dans les moyennes du coefficient d'immunité entre les différents types de "Molecule".
H1: Il existe au moins une différence dans les moyennes du coefficient d'immunité entre les différents types de "Molecule".

Effets principaux du facteur "Administration": H0: Il n'y a pas de différence dans les moyennes du coefficient d'immunité entre les différentes types d'administrations.
H1: Il existe au moins une différence dans les moyennes du coefficient d'immunité entre les différentes types d'administrations.

Interaction entre "Molecule" et "Administration":
H0: Il n'y a pas d'interaction entre les "Molecule" et les "Administration", c'est-à-dire que l'effet d'un "Molecule" sur le coefficient d'immunité est le même pour toutes les "Administration".
H1: Il existe une interaction entre les "Molecule" et les "Administration", c'est-à-dire que l'effet d'un "Molecule" sur le coefficient d'immunité change selon l' "Administration".

```
anova(modele_croise)
```

```
## Analysis of Variance Table
##
## Response: Amelioration
##                          Df Sum Sq Mean Sq F value    Pr(>F)
## Molecule                  2  27.63  13.817  1.4030 0.2546829
## Administration            1   0.60   0.600  0.0609 0.8059756
## Molecule:Administration   2 201.90 100.950 10.2507 0.0001683 ***
## Residuals                54 531.80   9.848
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pour les effets principaux du facteur "Molecule":
Pour un seuil de 1% ou 5% on garde H0, donc :
H0: Il n'y a pas de différence dans les moyennes du coefficient d'immunité entre les différents types de "Molecule".

Pour les effets principaux du facteur "Administration":
Pour un seuil de 1% ou 5% on garde H0, donc :
H0: Il n'y a pas de différence dans les moyennes du coefficient d'immunité entre les différentes types d'administrations.

Pour l'interaction entre "Molecule" et "Administration":
Pour un seuil de 1% on garde H0 mais pour 5% on rejette H0, donc :
H1: Il existe une interaction entre les "Molecule" et les "Administration", c'est-à-dire que l'effet d'un "Molecule" sur le coefficient d'immunité change selon l' "Administration".

```
summary(modele_croise)
```

```
##
## Call:
## lm(formula = Amelioration ~ Molecule * Administration, data = dataMedicament)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.200 -1.575 -0.100  1.300 14.800
##
## Coefficients:
```

```
##                                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)                              13.0000     0.9924  13.100  < 2e-16
## MoleculeBetaTriomphe                     -2.2000     1.4034  -1.568   0.1228
## MoleculeGammaSucces                      -6.0000     1.4034  -4.275 7.82e-05
## AdministrationOral                       -3.5000     1.4034  -2.494   0.0157
## MoleculeBetaTriomphe:AdministrationOral   2.4000     1.9848   1.209   0.2318
## MoleculeGammaSucces:AdministrationOral    8.7000     1.9848   4.383 5.43e-05
##
## (Intercept)                              ***
## MoleculeBetaTriomphe
## MoleculeGammaSucces                      ***
## AdministrationOral                       *
## MoleculeBetaTriomphe:AdministrationOral
## MoleculeGammaSucces:AdministrationOral   ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.138 on 54 degrees of freedom
## Multiple R-squared:  0.302,  Adjusted R-squared:  0.2374
## F-statistic: 4.674 on 5 and 54 DF,  p-value: 0.001285
```

Les p-values nous indiquent ici que les molecules BetaTriomphe et GammaSucces on de l'influence sur le coefficient d'immunité, et du côté des types des administrations "Oral" a bien un influence sur le coefficient d'immunité.

Cependant, on va quand même analyser l'anova pour chacune des variables qualitatives séparément pour essayer de voir si il n'y aurait pas encore plus d'interaction.

```
modele_hierarchise <- lm(Amelioration ~ Molecule + Administration, data = dataMedicament)
modele_hierarchise
```

```
##
## Call:
## lm(formula = Amelioration ~ Molecule + Administration, data = dataMedicament)
##
## Coefficients:
##          (Intercept)  MoleculeBetaTriomphe   MoleculeGammaSucces
##                11.15                 -1.00                 -1.65
##   AdministrationOral
##                 0.20
```

```
anova(modele_hierarchise)
```

```
## Analysis of Variance Table
##
## Response: Amelioration
##                Df Sum Sq Mean Sq F value Pr(>F)
## Molecule        2  27.63  13.817  1.0546 0.3552
## Administration  1   0.60   0.600  0.0458 0.8313
## Residuals      56 733.70  13.102
```

Ici, on voit que les deux p-value sont à nouveau très élevées. Donc on peut déjà en conclure qu'il n'y aura aucune influence ... Mais nous allons le confirmer en analysant le summary.

```
summary(modele_hierarchise)
```

```
##
## Call:
## lm(formula = Amelioration ~ Molecule + Administration, data = dataMedicament)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -5.35  -2.50  -0.60   1.70  17.30
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           11.1500     0.9346  11.930   <2e-16 ***
## MoleculeBetaTriomphe  -1.0000     1.1446  -0.874    0.386
## MoleculeGammaSucces   -1.6500     1.1446  -1.442    0.155
## AdministrationOral     0.2000     0.9346   0.214    0.831
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.62 on 56 degrees of freedom
## Multiple R-squared:  0.03705,    Adjusted R-squared:  -0.01453
## F-statistic: 0.7183 on 3 and 56 DF,  p-value: 0.5452
```

Grace à ce summary, on ne constate aucune différence qu'avec l'anova au système croisé. Toutes les p-values sont plus élevées que le seuil « logique » qui est de 5%. Donc aucune des deux variables qualitatives influencent le coefficient d'immunité.

D'ailleurs cela se confirme en regardant la p-value du modele qui est très élevée et qui montre qu'il n'y aucune influence !

2 ACP et ACM

2.1 Eaux minérales (ACP)

> Le fichier Eaux1.txt contient des données sur la teneur en divers éléments chimiques pour quelques eaux minérales commercialisées en France.

Quelles relations peut-on détecter ?

Peut-on donner une signification claire aux axes principaux ?

Pour pourvoir réaliser nos ACP, il faut inclure la librairie FactoMineR.
Nous allons aussi inclure la librairie FactoExtra afin d'obtenir des graphiques plus présentables.

```
library(FactoMineR)
```

```
## Warning: le package 'FactoMineR' a été compilé avec la version R 4.2.3
```

```
library(factoextra)
```

```
## Warning: le package 'factoextra' a été compilé avec la version R 4.2.3
```

```
## Le chargement a nécessité le package : ggplot2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
remplaceNAparMOY<-function(x)
{
  return ( ifelse(is.na(x), mean(x,na.rm = TRUE), x)  )
}
```

Tout d'abord, on lit le fichier dans le but d'avoir un data set correctement formé.

```
setwd("C:\\Users\\amine\\OneDrive\\Bureau\\EcomStat\\Labo\\Evaluation02\\datasets")

donnees <- read.table("Eaux1.txt", sep="\t", header=TRUE, row.names=7)

summary(donnees)
```

```
##       HCO3             SO4               Cl              Ca
##  Min.   : 59.0   Min.   :  3.00   Min.   : 2.00   Min.   :  4.00
##  1st Qu.:185.2   1st Qu.:  8.50   1st Qu.: 6.00   1st Qu.: 53.25
##  Median :259.5   Median : 14.50   Median : 8.50   Median : 72.00
##  Mean   :250.4   Mean   : 42.40   Mean   :13.65   Mean   : 77.50
##  3rd Qu.:334.2   3rd Qu.: 24.75   3rd Qu.:18.50   3rd Qu.: 92.25
##  Max.   :402.0   Max.   :306.00   Max.   :44.00   Max.   :202.00
##        Mg              Na
##  Min.   : 1.00   Min.   : 2.00
##  1st Qu.: 4.00   1st Qu.: 4.75
##  Median : 6.00   Median : 9.00
##  Mean   :11.85   Mean   :10.10
##  3rd Qu.:19.25   3rd Qu.:13.00
##  Max.   :36.00   Max.   :31.00
```

On peut dès maintenant effectuer la fonction PCA sur le data set formé qui va nous donner ce qu'il nous faut et on va pouvoir analyser sur base de ces résultats.

```
resultat_acp <- PCA(donnees, graph = FALSE)
```

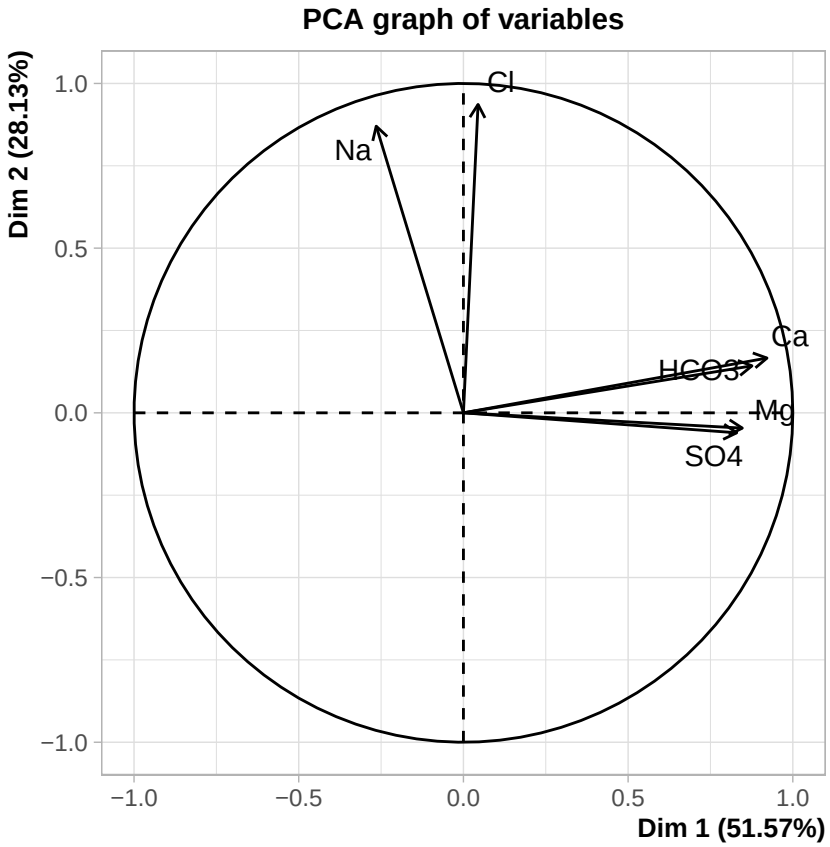
Ensuite, on va pouvoir analyser les différentes valeurs propres afin de voir quels sont les axes qui conservent le maximum d'Inertie.

```
resultat_acp$eig
```

```
##        eigenvalue percentage of variance cumulative percentage of variance
## comp 1 3.09408747             51.5681245                           51.56812
## comp 2 1.68756032             28.1260054                           79.69413
## comp 3 0.59651319              9.9418865                           89.63602
## comp 4 0.50284416              8.3807361                           98.01675
## comp 5 0.09323922              1.5539871                           99.57074
## comp 6 0.02575563              0.4292605                          100.00000
```

Les deux premiers axes, seraient suffisants pour notre etude puisque ils expliquant presque 80% de l'inertie.

```
plot(resultat_acp, choix="var")
```



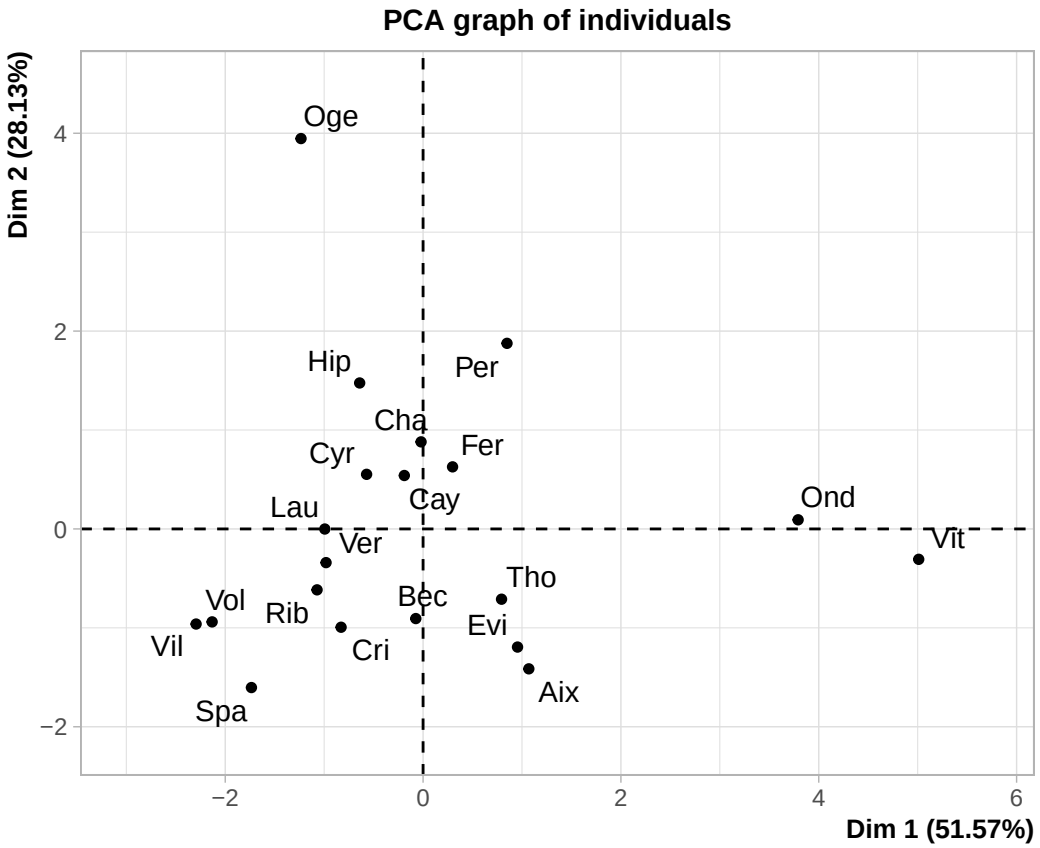
Graphique des Variables

Axe de la première composante principale (Dim 1) : Il explique 51,57% de la variance. Cet axe est dominé par des variables telles que Ca, HCO3, Mg, et SO4. Ces variables sont fortement corrélées avec la première composante principale, ce qui signifie que cette composante pourrait être interprétée comme un facteur lié à la "minéralité" de l'eau, car ces éléments sont des indicateurs classiques de ce caractéristique dans l'eau.

Axe de la deuxième composante principale (Dim 2) : Il explique 28,13% de la variance. Sur cet axe, Na et Cl sont les plus éloignés de l'origine, suggérant qu'ils contribuent significativement à cette composante. Cette composante pourrait être liée à la "salinité" de l'eau, étant donné que le sodium (Na) et le chlore (Cl) sont des composants majeurs du sel.

Les angles entre les vecteurs indiquent le niveau de corrélation entre les variables. Par exemple, Ca et Mg sont assez proches, indiquant une corrélation positive. En revanche, Cl et SO4 sont presque perpendiculaires, suggérant peu ou pas de corrélation directe entre eux.

```
plot(resultat_acp, choix="ind")
```



Graphique des Individus

Axe de la première composante principale (Dim 1) : Les eaux situées à droite sur l'axe (avec des valeurs positives) sont probablement plus riches en Ca, HCO3, Mg, et SO4, tandis que celles à gauche sont moins riches en ces minéraux.

Axe de la deuxième composante principale (Dim 2) : Les eaux en haut du graphique (avec des valeurs positives sur Dim 2) sont probablement plus riches en Na et Cl, indiquant une teneur plus élevée en sel.

Les positions relatives des eaux minérales peuvent suggérer des similitudes ou des différences dans leur composition chimique. Par exemple, "Oge" est distinctement différent des autres eaux minérales, suggérant une composition chimique unique par rapport aux autres échantillons.

```
dimdesc(resultat_acp)
```

```
## $Dim.1
##
## Link between the variable and the continuous variables (R-square)
## ================================================================================
##      correlation      p.value
## Ca     0.9216272 7.951555e-09
## HCO3   0.8759693 4.170667e-07
## Mg     0.8466723 2.514351e-06
## SO4    0.8297245 6.050759e-06
##
## $Dim.2
##
## Link between the variable and the continuous variables (R-square)
```

```
## ========================================================================================
##    correlation      p.value
## Cl   0.9362016 1.317448e-09
## Na   0.8702075 6.140146e-07
##
## $Dim.3
##
## Link between the variable and the continuous variables (R-square)
## ========================================================================================
##    correlation    p.value
## Mg   0.4644415 0.03910773
```

2.2 Etude de maïs (ACP)

e ministère de l'Agriculture du Malabarland a commandité une étude sur les plants de maïs afin d'optimiser les techniques de culture. Un échantillon de 100 pieds de maïs a été constitué (sur 50000 pieds possibles) et les résultats ont été compilés dans le fichier etude- agro-mais.csv.

Certaines variables s'interprètent par leur nom et pour les autres :
Masse: masse de l'ensemble des grains du plant
Germination.epi: le grain est-il germé sur épi ?
Verse: le pied est-il penché ou tombé ?
Attaque: attaqué par des insectes ?
Hauteur.J7: hauteur 7 jours après la récolte
Verse.Traitement: verse après traitement ?
Nb.jours.attaque: nombre de jours entre la pousse jusqu'à l'attaque
Censure.droite: non utilisée dans la suite

Le Ministère commandite une étude globale d'exploration des données (il ne le sait pas vu sa formation, mais ceci implique donc une ACM et une ACP de l'ensemble des données). Que peut-on observer ?

Tout d'abord, nous lisons notre fichier dans le but d'avoir un data Set correctement formé.
Ensuite, nous faisons un summary pour nous assurer que R interprète correctement le mode des colonnes.

```r
setwd("C:\\Users\\amine\\OneDrive\\Bureau\\EcomStat\\Labo\\Evaluation02\\datasets")

dataMais <- read.table(file = "etude-agro-mais.csv", header=TRUE, sep=";", row.names=1)

summary(dataMais)
```

```
##     Hauteur          Masse        Nb.grains      Masse.grains
##  Min.   :155.0   Min.   :1104   Min.   : 73.0   Min.   : 21.9
##  1st Qu.:228.0   1st Qu.:1525   1st Qu.:203.0   1st Qu.: 60.9
##  Median :263.0   Median :1830   Median :298.0   Median : 89.4
##  Mean   :259.4   Mean   :1812   Mean   :292.6   Mean   : 88.0
##  3rd Qu.:291.0   3rd Qu.:2022   3rd Qu.:369.0   3rd Qu.:110.7
##  Max.   :359.0   Max.   :2752   Max.   :509.0   Max.   :152.7
##  NA's   :3       NA's   :3      NA's   :3       NA's   :3
##    Couleur          Germination.epi    Enracinement          Verse
##  Length:100         Length:100         Length:100         Length:100
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

```
##
##
##
##
##    Attaque             Parcelle           Hauteur.J7    Verse.Traitement
##  Length:100         Length:100         Min.   :163.0   Length:100
##  Class :character   Class :character   1st Qu.:224.2   Class :character
##  Mode  :character   Mode  :character   Median :265.0   Mode  :character
##                                        Mean   :257.4
##                                        3rd Qu.:291.0
##                                        Max.   :347.0
##
##  Nb.jours.attaque Censure.droite
##  Min.   : 12.00   Min.   :0.0000
##  1st Qu.: 47.50   1st Qu.:0.0000
##  Median : 79.00   Median :1.0000
##  Mean   : 83.82   Mean   :0.5735
##  3rd Qu.:133.00   3rd Qu.:1.0000
##  Max.   :133.00   Max.   :1.0000
##  NA's   :33       NA's   :32
```

La censure étant une donnée non utilisée dans la suite, il est préférable de l'enlever pour ne pas avoir des données qui peuvent les fausser.

```
dataMais$Censure.droite <- as.factor(dataMais$Censure.droite)
```

Etant donné que nous réalisons une ACP, il faut retirer du data-set, toutes les valeurs qualitatives car une ACP s'effectue entre variables quantitatives.

```
dataMais <- dataMais[which(sapply(dataMais, is.numeric))]

summary(dataMais)
```

```
##     Hauteur          Masse        Nb.grains      Masse.grains     Hauteur.J7
##  Min.   :155.0   Min.   :1104   Min.   : 73.0   Min.   : 21.9   Min.   :163.0
##  1st Qu.:228.0   1st Qu.:1525   1st Qu.:203.0   1st Qu.: 60.9   1st Qu.:224.2
##  Median :263.0   Median :1830   Median :298.0   Median : 89.4   Median :265.0
##  Mean   :259.4   Mean   :1812   Mean   :292.6   Mean   : 88.0   Mean   :257.4
##  3rd Qu.:291.0   3rd Qu.:2022   3rd Qu.:369.0   3rd Qu.:110.7   3rd Qu.:291.0
##  Max.   :359.0   Max.   :2752   Max.   :509.0   Max.   :152.7   Max.   :347.0
##  NA's   :3       NA's   :3      NA's   :3       NA's   :3
##  Nb.jours.attaque
##  Min.   : 12.00
##  1st Qu.: 47.50
##  Median : 79.00
##  Mean   : 83.82
##  3rd Qu.:133.00
##  Max.   :133.00
##  NA's   :33
```

Malheureusement, certains individus ne disposent pas de données sur certaines variables. C'est pourquoi, afin d'avoir une représentation la plus fidèle de la réalité, nous allons changer ces valeurs inconnues par la moyenne de la colonne grâce à une fonction fournie :

```
remplaceNAparMOY<-function(x)
{
  return ( ifelse(is.na(x), mean(x,na.rm = TRUE), x)  )
}
```

Voici l'exécution de la fonction :

```
dataMais <- apply(dataMais, 2, remplaceNAparMOY)
dataMais
```

```
##      Hauteur    Masse Nb.grains Masse.grains Hauteur.J7 Nb.jours.attaque
## 1   259.3608 1811.619  292.6392     88.00206        171           83.8209
## 2   199.0000 1431.000  320.0000     92.10000        196           83.8209
## 3   205.0000 1468.000  290.0000     89.40000        198           83.8209
## 4   173.0000 1398.000  147.0000     42.60000        176           83.8209
## 5   233.0000 1622.000  138.0000     43.20000        230           83.8209
## 6   206.0000 1428.000  166.0000     44.10000        200           83.8209
## 7   261.0000 1574.000  151.0000     50.70000        266           83.8209
## 8   155.0000 1215.000  293.0000     88.20000        176           83.8209
## 9   214.0000 1457.000  345.0000    108.60000        220           83.8209
## 10  174.0000 1368.000  234.0000     78.80000        182           83.8209
## 11  207.0000 1355.000  142.0000     42.90000        202           83.8209
## 12  196.0000 1469.000  301.0000     90.40000        188           83.8209
## 13  224.0000 1474.000  311.0000     99.90000        222           83.8209
## 14  237.0000 1599.000  273.0000     84.60000        220           83.8209
## 15  210.0000 1437.000  307.0000     99.60000        205           83.8209
## 16  187.0000 1282.000  303.0000     88.80000        187           83.8209
## 17  211.0000 1505.000  156.0000     46.80000        208           83.8209
## 18  215.0000 1491.000  278.0000     82.80000        214           83.8209
## 19  242.0000 1830.000  288.0000     90.60000        231           83.8209
## 20  197.0000 1448.000  263.0000     56.40000        201           83.8209
## 21  265.0000 1881.000  252.0000     63.30000        250           83.8209
## 22  227.0000 1376.000  276.0000     59.70000        217           83.8209
## 23  244.0000 1667.000  283.0000     86.40000        260           83.8209
## 24  222.0000 1518.000  271.0000     83.40000        231           83.8209
## 25  238.0000 1729.000  296.0000     91.20000        231           83.8209
## 26  235.0000 1596.000  304.0000    101.10000        230           83.8209
## 27  217.0000 1462.000  287.0000     84.90000        229           83.8209
## 28  184.0000 1272.000  262.0000     78.60000        186           83.8209
## 29  166.0000 1104.000  256.0000     81.30000        163           83.8209
## 30  231.0000 1614.000  298.0000     85.50000        237           83.8209
## 31  209.0000 1309.000  280.0000     81.90000        217           83.8209
## 32  259.3608 1811.619  292.6392     88.00206        193           83.8209
## 33  278.0000 2004.000  294.0000     86.10000        277           79.0000
## 34  260.0000 1853.000  378.0000    102.90000        269           76.0000
## 35  217.0000 1623.000  139.0000     43.80000        201           41.0000
## 36  247.0000 1660.000  391.0000    126.30000        254          133.0000
## 37  358.0000 2206.000  339.0000    102.60000        339           17.0000
## 38  251.0000 1838.000  282.0000     87.90000        264           20.0000
## 39  324.0000 2314.000  261.0000     78.90000        312           48.0000
## 40  328.0000 2119.000  467.0000    140.10000        314           17.0000
## 41  300.0000 2067.000  337.0000    108.30000        296          133.0000
## 42  266.0000 1821.000  338.0000     84.00000        273          133.0000
```

```
## 43  231.0000 1771.000  332.0000    103.50000        235          69.0000
## 44  313.0000 1967.000  382.0000    110.70000        299         133.0000
## 45  279.0000 1925.000  331.0000     98.70000        282          68.0000
## 46  236.0000 1792.000  272.0000     81.60000        240          30.0000
## 47  256.0000 1818.000  450.0000    135.00000        267         133.0000
## 48  293.0000 2147.000  152.0000     42.90000        291          12.0000
## 49  228.0000 1706.000  369.0000    113.40000        225          47.0000
## 50  269.0000 1914.000  329.0000     98.40000        274          83.0000
## 51  287.0000 1860.000  362.0000    102.40000        286         133.0000
## 52  196.0000 1525.000  132.0000     39.60000        194         133.0000
## 53  218.0000 1456.000  421.0000    116.70000        216          34.0000
## 54  249.0000 1800.000  341.0000    101.40000        261         133.0000
## 55  321.0000 2293.000  374.0000    112.20000        305          70.0000
## 56  319.0000 2045.000  144.0000     39.00000        301         133.0000
## 57  280.0000 1964.000  385.0000    120.90000        283          59.0000
## 58  265.0000 1794.000  145.0000     51.10000        270          94.0000
## 59  296.0000 2098.000  425.0000    124.80000        294          58.0000
## 60  240.0000 1812.000  302.0000     90.90000        243          40.0000
## 61  311.0000 2011.000  403.0000    114.60000        298         109.0000
## 62  283.0000 1971.000  430.0000    116.90000        285          42.0000
## 63  272.0000 1966.000  361.0000     96.00000        276          17.0000
## 64  246.0000 1757.000  416.0000    136.80000        252          16.0000
## 65  259.3608 1811.619  292.6392     88.00206        225          83.8209
## 66  286.0000 2097.000  449.0000    134.70000        282         133.0000
## 67  299.0000 2207.000  154.0000     45.30000        292          55.0000
## 68  281.0000 1956.000  220.0000     76.80000        280          94.0000
## 69  278.0000 1954.000  364.0000    110.00000        275         133.0000
## 70  270.0000 1970.000  328.0000    101.40000        271          46.0000
## 71  269.0000 1742.000  342.0000    105.60000        271          18.0000
## 72  359.0000 2752.000  389.0000    114.60000        347         133.0000
## 73  250.0000 1685.000  458.0000    129.00000        255          70.0000
## 74  280.0000 1937.000   73.0000     21.90000        275         133.0000
## 75  307.0000 2180.000  203.0000     60.90000        309          68.0000
## 76  252.0000 1910.000  285.0000     87.00000        258          48.0000
## 77  288.0000 2078.000  343.0000    101.40000        292         133.0000
## 78  291.0000 2022.000  383.0000    121.80000        281         104.0000
## 79  299.0000 2190.000  146.0000     45.60000        297          18.0000
## 80  311.0000 2301.000  169.0000     48.90000        319          48.0000
## 81  283.0000 1934.000  211.0000     57.60000        277         122.0000
## 82  329.0000 2422.000  406.0000    137.10000        340          66.0000
## 83  237.0000 1574.000  380.0000    117.30000        248         133.0000
## 84  304.0000 2177.000  456.0000    127.50000        309         113.0000
## 85  315.0000 2351.000  188.0000     75.60000        329          40.0000
## 86  282.0000 1840.000  163.0000     49.80000        280          72.0000
## 87  314.0000 2102.000  192.0000     70.20000        320          92.0000
## 88  287.0000 2079.000  143.0000     46.20000        291         112.0000
## 89  235.0000 1479.000  427.0000    126.60000        234          23.0000
## 90  341.0000 2473.000  173.0000     66.00000        341          60.0000
## 91  249.0000 1713.000  100.0000     30.00000        252         133.0000
## 92  299.0000 1967.000  509.0000    152.70000        290         133.0000
## 93  310.0000 2282.000  457.0000    137.10000        312         133.0000
## 94  299.0000 2174.000  352.0000    114.60000        305         133.0000
## 95  286.0000 2036.000  390.0000    115.50000        295         133.0000
## 96  263.0000 1924.000  130.0000     41.70000        263          61.0000
```

```
## 97  308.0000 1994.000  194.0000     58.20000        310          81.0000
## 98  259.0000 1324.000  199.0000     51.90000        262         133.0000
## 99  268.0000 1903.000  422.0000    128.10000        266         133.0000
## 100 269.0000 1722.000  333.0000    101.40000        270         133.0000
```

Nous pouvons dès maintenant effectuer la fonction PCA sur notre nouveau data-set :

```
resultat_acp_mais <- PCA(dataMais, graph = FALSE)
```

Affichons les valeurs propres afin de savoir si nos axes conservent le maximum d'Inertie.

```
resultat_acp_mais$eig
```

```
##        eigenvalue percentage of variance cumulative percentage of variance
## comp 1 3.03627237             50.6045396                           50.60454
## comp 2 1.80172092             30.0286820                           80.63322
## comp 3 0.99223967             16.5373278                           97.17055
## comp 4 0.10262515              1.7104191                           98.88097
## comp 5 0.04404064              0.7340107                           99.61498
## comp 6 0.02310125              0.3850209                          100.00000
```

Les 2 premiers axes permettent d'expliquer 80% de l'Inertie, ce qui est vraiment très important ! C'est pour cela que l'étude va se porter uniquement sur ces deux axes.

Avant d'afficher les graphiques, il est peut-être plus prudent de voir les corrélations entre les individus et les axes, de même pour les variables (c'est-à-dire les colonnes).

```
resultat_acp_mais$var$cos2
```

```
##                       Dim.1      Dim.2        Dim.3        Dim.4        Dim.5
## Hauteur          0.880648110 0.08930040 0.0009505018 0.0017013698 2.639322e-02
## Masse            0.838724364 0.09730452 0.0009177140 0.0594613332 3.209788e-03
## Nb.grains        0.210504985 0.77262026 0.0045777529 0.0001161875 1.514899e-03
## Masse.grains     0.238593328 0.74176049 0.0071410522 0.0002511237 1.475512e-03
## Hauteur.J7       0.866064352 0.08036184 0.0010336735 0.0408408789 1.143167e-02
## Nb.jours.attaque 0.001737234 0.02037342 0.9776189738 0.0002542558 1.555869e-05
```

Nous constatons que les variables des Hauteurs (Hauteur et Hauteur.J7) et de la Masse ont une très forte corrélation entre elles sur la Dimension 1.
Dans la 2ème dimension, il s'agît du nombre de grains (Nb.grains) ainsi que leurs masses (Masse.grains) qui ont une bonne corrélation.

A présent regardons le cos^2 des individus afin de voir la corrélation de chaque individu par rapport aux dimensions.

```
resultat_acp_mais$ind$cos2
```

```
##            Dim.1        Dim.2        Dim.3        Dim.4        Dim.5
## 1   0.2852393480 4.460282e-02 1.041758e-03 3.979617e-01 2.595708e-01
## 2   0.7494920701 2.447102e-01 1.225307e-03 3.498997e-03 1.257342e-05
## 3   0.8402472846 1.503395e-01 5.738892e-04 7.589875e-03 3.782163e-06
## 4   0.9353580763 4.684761e-02 1.601343e-03 1.172739e-02 3.829873e-03
```

```
## 5   0.5632898822 4.285329e-01 7.798524e-03 1.512198e-05 1.457529e-07
## 6   0.8891123456 1.055093e-01 3.339765e-03 7.209850e-05 7.633556e-04
## 7   0.2365382580 6.165377e-01 1.295832e-02 1.226535e-01 2.289561e-03
## 8   0.8393058659 1.474147e-01 4.664487e-04 1.006844e-04 1.247800e-02
## 9   0.4244239029 5.571367e-01 3.745545e-03 7.513513e-03 2.477036e-03
## 10  0.9472155920 3.196561e-02 5.389656e-06 7.773196e-03 1.072628e-02
## 11  0.8689374534 1.223511e-01 4.006250e-03 3.236779e-03 1.035230e-03
## 12  0.7989339325 1.780117e-01 9.800013e-04 2.195278e-02 1.182481e-04
## 13  0.6089121101 3.599982e-01 1.589304e-03 1.359345e-02 4.020695e-04
## 14  0.9148797141 4.132579e-02 4.103897e-08 5.426337e-03 2.753649e-02
## 15  0.6933361574 2.945906e-01 1.497753e-03 1.243645e-05 1.248740e-05
## 16  0.8273801280 1.717454e-01 4.605121e-04 4.012831e-04 5.337129e-08
## 17  0.8354058497 1.594506e-01 3.964985e-03 1.093224e-03 8.483575e-05
## 18  0.9304364134 6.933434e-02 1.885846e-05 1.917869e-04 9.341141e-06
## 19  0.4553784393 9.711205e-02 2.672041e-03 4.334222e-01 3.449265e-03
## 20  0.9528166706 4.642531e-05 5.677063e-04 8.524011e-04 1.671185e-06
## 21  0.0627758178 7.515057e-01 8.538802e-03 5.564564e-02 4.319456e-02
## 22  0.8768913782 3.350053e-04 1.559142e-03 4.132307e-02 3.959618e-02
## 23  0.5506153412 1.588292e-02 7.407628e-04 3.322140e-01 1.003636e-01
## 24  0.9258417614 4.219470e-02 5.191721e-05 2.393671e-02 6.859165e-03
## 25  0.6849046999 2.204501e-01 2.112788e-03 8.899180e-02 3.819970e-04
## 26  0.5135524300 4.329803e-01 2.760231e-03 1.052304e-03 7.642510e-05
## 27  0.8580061967 9.853593e-02 7.791693e-06 3.815393e-02 5.089296e-03
## 28  0.9396086423 5.961360e-02 4.452433e-06 4.282833e-04 2.568609e-04
## 29  0.9185743347 7.897328e-02 3.763622e-05 4.112498e-04 1.760954e-05
## 30  0.8429561533 1.313563e-01 1.514469e-04 1.102149e-02 5.090357e-03
## 31  0.8721337171 7.722236e-02 1.718886e-05 5.053735e-02 7.117199e-05
## 32  0.2852393480 4.460282e-02 1.041758e-03 3.979617e-01 2.595708e-01
## 33  0.7664933189 1.835739e-01 1.908190e-02 1.763732e-02 2.588396e-03
## 34  0.2991453157 5.268350e-01 8.880361e-02 6.338049e-03 6.860982e-03
## 35  0.6266387877 2.379593e-01 1.171010e-01 1.754737e-02 5.680779e-04
## 36  0.0087520534 6.709647e-01 2.997413e-01 6.778997e-03 3.406372e-03
## 37  0.6707786300 4.179495e-02 2.437164e-01 2.513104e-02 1.755768e-02
## 38  0.0007725622 2.228059e-02 9.580355e-01 2.613327e-03 1.628785e-02
## 39  0.5914301806 2.896253e-01 1.130647e-01 4.456871e-03 1.323047e-03
## 40  0.5904649323 9.994235e-02 2.937576e-01 7.834515e-03 7.755817e-03
## 41  0.6173729606 1.769880e-02 3.591418e-01 1.821825e-04 3.779544e-05
## 42  0.0703002287 2.569037e-02 8.201884e-01 1.312119e-02 5.432151e-05
## 43  0.1559330896 5.291155e-01 2.078288e-01 6.841960e-02 3.838670e-02
## 44  0.6109266528 6.738109e-02 2.871640e-01 1.550155e-02 1.901272e-02
## 45  0.7568471542 1.552380e-02 2.048935e-01 2.150915e-02 3.596817e-04
## 46  0.1635240010 1.664355e-02 7.952707e-01 1.282983e-02 9.773279e-03
## 47  0.1370053477 6.689607e-01 1.881631e-01 5.163671e-05 5.678706e-03
## 48  0.0271880260 6.752938e-01 2.946293e-01 9.151057e-04 6.673549e-05
## 49  0.0821428894 4.959676e-01 3.996959e-01 1.992839e-02 2.037412e-03
## 50  0.8493306122 1.178654e-01 8.415685e-03 4.825260e-04 2.005520e-02
## 51  0.3592321750 1.022907e-01 5.042659e-01 2.677088e-02 4.940571e-03
## 52  0.6791448069 1.008755e-01 2.009510e-01 1.646844e-02 2.440479e-03
## 53  0.1395339480 5.151423e-01 3.330304e-01 5.707520e-03 2.534830e-03
## 54  0.0138539128 2.609534e-01 7.003649e-01 2.654750e-04 2.257759e-02
## 55  0.9502398538 1.878273e-03 3.301605e-02 1.201453e-02 2.792621e-03
## 56  0.0681965259 6.428009e-01 2.730311e-01 3.935828e-03 1.199926e-02
## 57  0.5470291035 2.398259e-01 2.071006e-01 1.569499e-03 1.825165e-03
## 58  0.0734598012 8.375939e-01 6.563213e-02 1.446561e-02 1.844490e-03
```

```
## 59  0.6974965026 1.683065e-01 1.324722e-01 2.300781e-04 3.389343e-05
## 60  0.0820912514 1.368187e-02 8.669504e-01 2.241954e-02 1.284502e-02
## 61  0.7889435168 1.133144e-01 7.029202e-02 8.682969e-03 1.773838e-02
## 62  0.4340272508 2.034035e-01 3.407556e-01 3.734970e-03 7.650646e-04
## 63  0.1565345079 4.990095e-03 8.131432e-01 2.815857e-04 2.352027e-05
## 64  0.0216724636 3.990087e-01 5.682840e-01 6.404311e-04 3.277819e-03
## 65  0.2852393480 4.460282e-02 1.041758e-03 3.979617e-01 2.595708e-01
## 66  0.4737703677 3.518208e-01 1.587601e-01 1.472161e-02 7.510965e-04
## 67  0.0799683223 8.653245e-01 4.360561e-02 9.914648e-03 3.811886e-05
## 68  0.1646975541 6.952410e-01 1.005824e-01 2.722306e-04 3.691042e-03
## 69  0.3482830261 1.829353e-01 4.646964e-01 3.878772e-03 1.636057e-04
## 70  0.3029596500 1.534475e-02 6.694501e-01 6.694149e-03 5.454570e-03
## 71  0.0414266515 4.571615e-02 8.561315e-01 5.247768e-02 1.961774e-03
## 72  0.8924303434 3.574151e-03 7.534121e-02 2.704426e-02 3.392541e-04
## 73  0.0417717820 8.638192e-01 7.527973e-02 1.264785e-02 2.165445e-04
## 74  0.0195653157 7.345676e-01 2.457704e-01 4.062062e-05 5.303869e-05
## 75  0.3055335418 6.804053e-01 1.262993e-02 1.511227e-04 6.363284e-04
## 76  0.0006321017 4.077185e-02 8.830351e-01 3.466504e-02 3.848916e-02
## 77  0.5488192985 1.596416e-02 4.245515e-01 4.434294e-03 4.456359e-03
## 78  0.6922768435 2.448258e-01 4.977072e-02 4.519278e-03 8.873593e-04
## 79  0.0555517844 7.019606e-01 2.407969e-01 1.210067e-03 3.086492e-04
## 80  0.2191017325 7.144451e-01 5.984247e-02 4.961118e-04 3.517884e-03
## 81  0.0191063962 5.412744e-01 4.324470e-01 1.827832e-04 2.927153e-03
## 82  0.9354762187 2.099853e-02 3.148755e-02 7.073559e-04 7.534457e-03
## 83  0.0134100664 6.254878e-01 3.375985e-01 1.787155e-02 3.626226e-03
## 84  0.7664662764 1.798400e-01 4.469559e-02 4.123797e-04 6.068466e-04
## 85  0.4130046015 4.447871e-01 1.172556e-01 3.909183e-04 1.597330e-02
## 86  0.0028801888 9.580041e-01 2.774036e-03 3.246163e-02 3.300449e-03
## 87  0.3791296713 5.592211e-01 2.799399e-02 2.007802e-02 1.299167e-03
## 88  0.0278390365 8.043801e-01 1.620293e-01 1.583441e-03 4.160872e-03
## 89  0.0262892657 4.957710e-01 4.480345e-01 2.625657e-02 3.486372e-03
## 90  0.4798550061 4.977764e-01 1.577819e-02 1.126755e-03 1.876518e-03
## 91  0.1974137533 5.092617e-01 2.919819e-01 8.901463e-04 4.496910e-04
## 92  0.4287125507 4.767219e-01 9.113799e-02 8.813600e-04 2.261641e-03
## 93  0.7221487910 1.604498e-01 1.078826e-01 7.069376e-03 2.033952e-03
## 94  0.6964552040 2.639532e-02 2.610189e-01 4.370140e-03 8.846763e-03
## 95  0.5458478872 1.500862e-01 2.974004e-01 4.232722e-06 5.373779e-03
## 96  0.0594918215 9.070481e-01 2.876992e-02 2.798743e-03 1.890499e-03
## 97  0.2010415443 7.477397e-01 3.293832e-03 4.594337e-02 1.912518e-03
## 98  0.2583061157 1.222981e-01 3.818667e-01 2.160848e-01 2.144190e-02
## 99  0.2227682464 5.213550e-01 2.504048e-01 4.768200e-03 5.764270e-04
## 100 0.0601794317 1.717973e-01 7.146370e-01 4.829588e-02 1.837354e-03
```

Nous pouvons observer que les individus 4, 72, 82 ont une très forte corrélation avec la dimension 1. Alors que 74, 73 ont une très forte corrélation avec la dimension 2.

Maintenant, il est peut-être intéressant de regarder la contribution des variables.

```
resultat_acp_mais$var$contrib
```

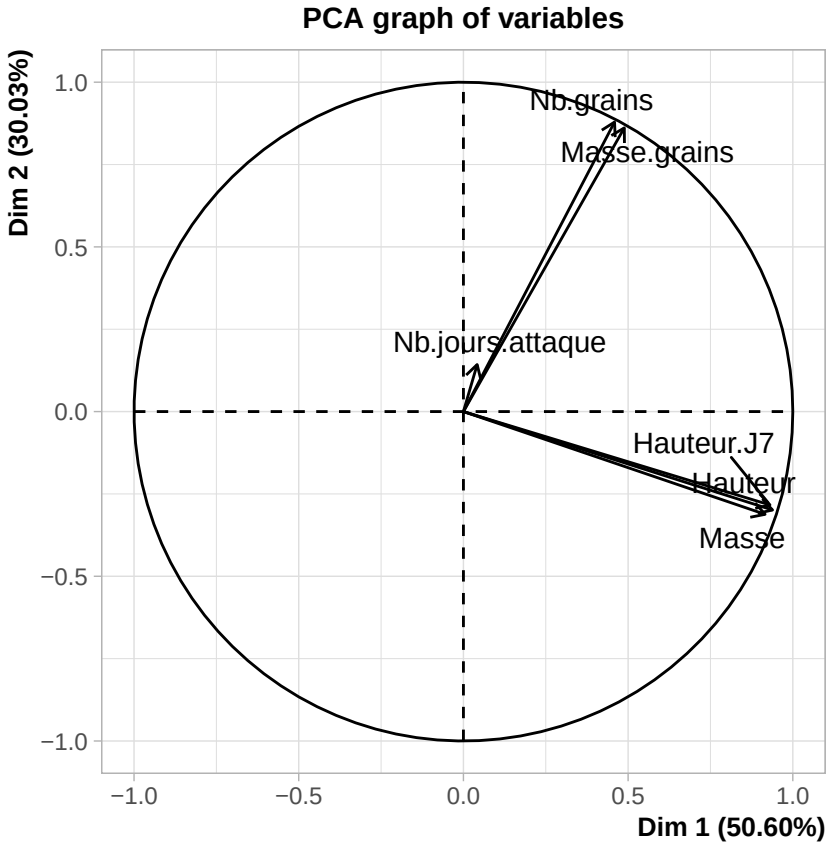
```
##                      Dim.1     Dim.2       Dim.3      Dim.4       Dim.5
## Hauteur          29.00425265  4.956395  0.09579357  1.6578488 59.92922945
## Masse            27.62348896  5.400643  0.09248914 57.9403137  7.28824207
## Nb.grains         6.93300730 42.882349  0.46135557  0.1132154  3.43977547
```

```
## Masse.grains      7.85810027 41.169555  0.71969025  0.2446999  3.35034190
## Hauteur.J7       28.52393480  4.460282  0.10417579 39.7961702 25.95708309
## Nb.jours.attaque  0.05721601  1.130775 98.52649568  0.2477520  0.03532802
```

Nous constatons que la masse et les hauteurs ont les plus grandes contributions pour l'axe des x alors que pour l'axe y, c'est le nombre de grains et la masse des grains.

Jusqu'à maintenant, nous avons regardé les cos^2 des individus et des variables et nous pouvons en déduire que l'axe 1 pourrait représenter la masse par rapport à la hauteur. L'axe 2 serait le nombre de grains par rapport à la masse des grains. C'est ce que l'on va chercher à prouver en analysant les graphiques.

```
plot(resultat_acp_mais, choix="var")
```



Interprétation du Graphique des Variables

Le graphique des variables montre les relations entre les variables quantitatives de l'étude sur les plants de maïs. Les axes Dim 1 et Dim 2 montrent les deux premières composantes principales qui expliquent la plus grande part de la variance dans l'ensemble des données.

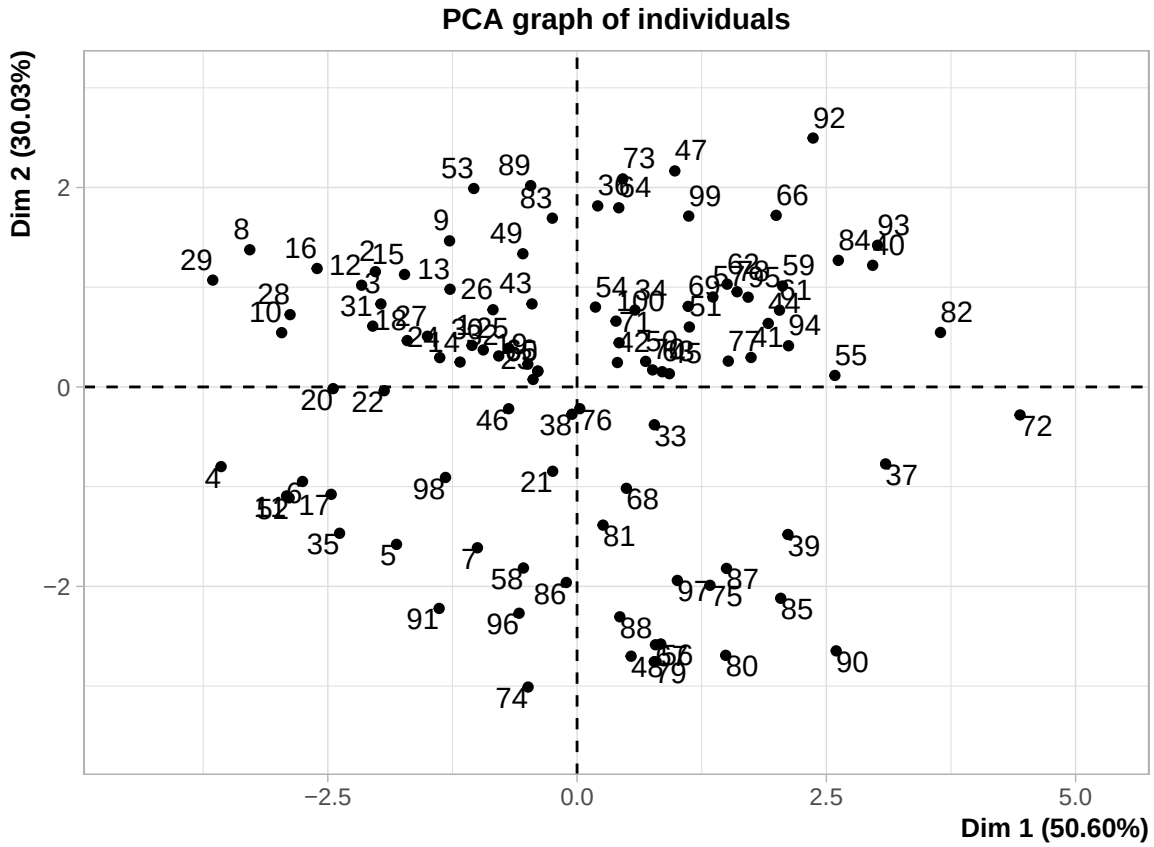
Observations :

Nous voyons 2 groupes de projections, le premier est constitué de la masse et des hauteurs. Il tend vers l'axe des X et qui se projette très bien sur l'axe des abscisses.

Le second est constitué de la masse et de la hauteur des grains se projette plutôt bien sur l'axe des ordonnées (la dimension 2).

La variable Nb.jours.attaque ne va pas être étudiée ici car elle se trouve trop proche du centre du graphique. Elle ne donnerait aucun résultat intéressant sur les 2 axes d'études.

```
plot(resultat_acp_mais, choix="ind")
```



Interprétation du Graphique des Individus

Le graphique des individus montre la distribution des 100 pieds de maïs étudiés dans le plan des deux premières composantes principales.

Observations :

Lorsqu'on se focalise sur l'axe des X, on remarque que les individus 4,72,82 (que l'on a mis en évidence en début de réflexion) ont une très bonne projection sur la dimension 1.

Conclusion

Grâce aux analyses des graphiques, nous pouvons conclure que la dimension 1 est la représentation de la masse par rapport aux différentes hauteurs. La dimension 2 représente la masse des grains par rapport au nombre de grains.

```
dimdesc(resultat_acp_mais)
```

```
## $Dim.1
##
## Link between the variable and the continuous variables (R-square)
## ================================================================================
##               correlation      p.value
## Hauteur         0.9384285 4.975846e-47
## Hauteur.J7      0.9306258 1.424584e-44
## Masse           0.9158190 1.298517e-40
## Masse.grains    0.4884602 2.527508e-07
```

```
## Nb.grains      0.4588082 1.580149e-06
##
## $Dim.2
##
## Link between the variable and the continuous variables (R-square)
## ================================================================================
##              correlation      p.value
## Nb.grains      0.8789882 2.759331e-33
## Masse.grains   0.8612552 1.437804e-30
## Hauteur.J7    -0.2834816 4.262506e-03
## Hauteur       -0.2988317 2.527129e-03
## Masse         -0.3119367 1.581085e-03
##
## $Dim.3
##
## Link between the variable and the continuous variables (R-square)
## ================================================================================
##                   correlation      p.value
## Nb.jours.attaque   0.9887462 1.132804e-82
```

2.3 Etude de maïs (ACM)

```
setwd("C:\\Users\\amine\\OneDrive\\Bureau\\EcomStat\\Labo\\Evaluation02\\datasets")

dataMais <- read.table(file = "etude-agro-mais.csv", header=TRUE, sep=";", row.names=1)

summary(dataMais)
```

```
##     Hauteur          Masse        Nb.grains      Masse.grains
##  Min.   :155.0   Min.   :1104   Min.   : 73.0   Min.   : 21.9
##  1st Qu.:228.0   1st Qu.:1525   1st Qu.:203.0   1st Qu.: 60.9
##  Median :263.0   Median :1830   Median :298.0   Median : 89.4
##  Mean   :259.4   Mean   :1812   Mean   :292.6   Mean   : 88.0
##  3rd Qu.:291.0   3rd Qu.:2022   3rd Qu.:369.0   3rd Qu.:110.7
##  Max.   :359.0   Max.   :2752   Max.   :509.0   Max.   :152.7
##  NA's   :3       NA's   :3      NA's   :3       NA's   :3
##    Couleur          Germination.epi    Enracinement          Verse
##  Length:100         Length:100         Length:100         Length:100
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    Attaque            Parcelle           Hauteur.J7    Verse.Traitement
##  Length:100         Length:100         Min.   :163.0   Length:100
##  Class :character   Class :character   1st Qu.:224.2   Class :character
##  Mode  :character   Mode  :character   Median :265.0   Mode  :character
##                                        Mean   :257.4
##                                        3rd Qu.:291.0
##                                        Max.   :347.0
##
```

```
##  Nb.jours.attaque Censure.droite
##  Min.   : 12.00   Min.   :0.0000
##  1st Qu.: 47.50   1st Qu.:0.0000
##  Median : 79.00   Median :1.0000
##  Mean   : 83.82   Mean   :0.5735
##  3rd Qu.:133.00   3rd Qu.:1.0000
##  Max.   :133.00   Max.   :1.0000
##  NA's   :33       NA's   :32
```

Etant donné que nous réalisons une ACM, il faut retirer du data-set, toutes les valeurs quantitatives car une ACM s'effectue entre variables qualitatives.

```
dataMaisACM <- dataMais[, c("Couleur", "Germination.epi", "Enracinement",
                            "Verse", "Attaque", "Parcelle", "Verse.Traitement")]

summary(dataMaisACM)
```

```
##    Couleur          Germination.epi    Enracinement          Verse
##  Length:100         Length:100         Length:100         Length:100
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##    Attaque            Parcelle         Verse.Traitement
##  Length:100         Length:100         Length:100
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
```

Puis on va enlever les NA.

```
dataMaisACM <- na.omit(dataMaisACM)
```

Nos données sont bien des valeurs interprétées comme étant qualitatives et nous pouvons donc procéder à l'ACM.

```
resultat_acm <- MCA(dataMaisACM, graph = FALSE)
```

```
resultat_acm$eig
```

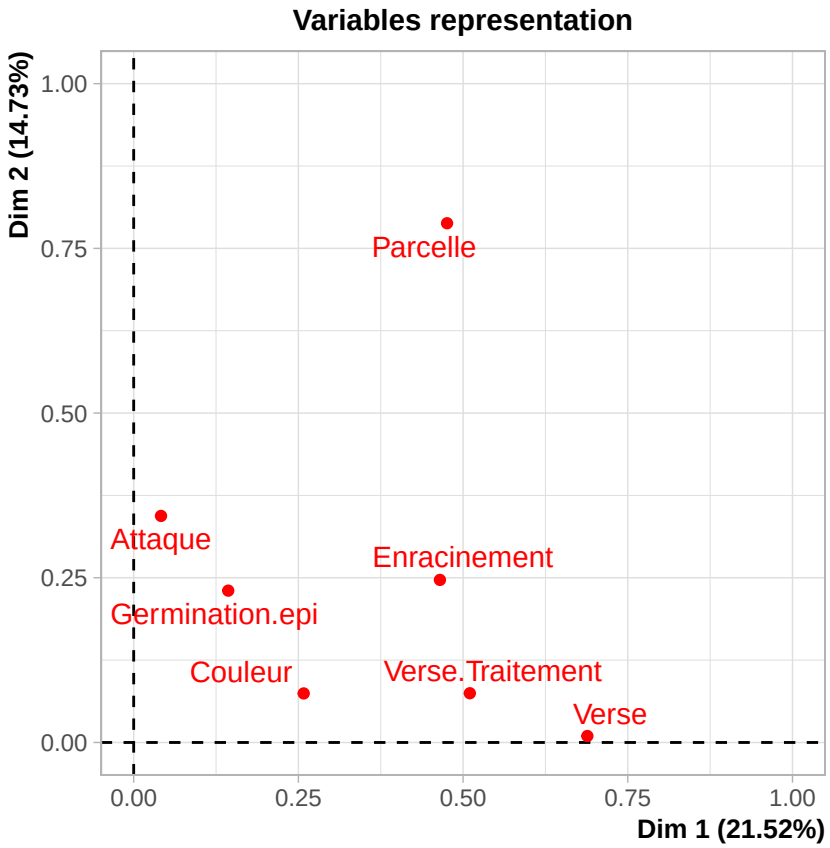
```
##        eigenvalue percentage of variance cumulative percentage of variance
## dim 1  0.36892825              21.520814                           21.52081
## dim 2  0.25259706              14.734829                           36.25564
## dim 3  0.20614283              12.024999                           48.28064
## dim 4  0.18324288              10.689168                           58.96981
## dim 5  0.17038309               9.939014                           68.90882
## dim 6  0.12799402               7.466318                           76.37514
## dim 7  0.12300380               7.175222                           83.55036
## dim 8  0.08631427               5.034999                           88.58536
## dim 9  0.07374111               4.301565                           92.88693
## dim 10 0.05695401               3.322317                           96.20924
## dim 11 0.03930146               2.292585                           98.50183
## dim 12 0.02568294               1.498171                          100.00000
```

Le faible pourcentage correspondant à chaque dimension est faible : mais dans ACM on travaille dans des espaces avec un nombre important de dimensions, puisqu'elles correspondent à toutes les modalités des différentes variables qualitatives.

```
resultat_acm$var$cos2[,1:2]
```

```
##                             Dim 1       Dim 2
## Jaune                 0.253465651 0.021482619
## Jaune.rouge           0.041260955 0.014578484
## Rouge                 0.134913126 0.073577170
## Germination.epi_Non   0.143520940 0.230450635
## Germination.epi_Oui   0.143520940 0.230450635
## Faible                0.305460776 0.011327063
## Fort                  0.041839889 0.038802058
## Moyen                 0.035487802 0.244963293
## Tres.fort             0.218115034 0.045667545
## Verse_Non             0.688637567 0.009806367
## Verse_Oui             0.688637567 0.009806367
## Attaque_Non           0.041484901 0.343868243
## Attaque_Oui           0.041484901 0.343868243
## Est                   0.259569159 0.362438369
## Nord                  0.019234390 0.455935956
## Ouest                 0.441319041 0.045826732
## Sud                   0.006617623 0.157363440
## Verse.Traitement_Non 0.510248989 0.074704165
## Verse.Traitement_Oui 0.510248989 0.074704165
```

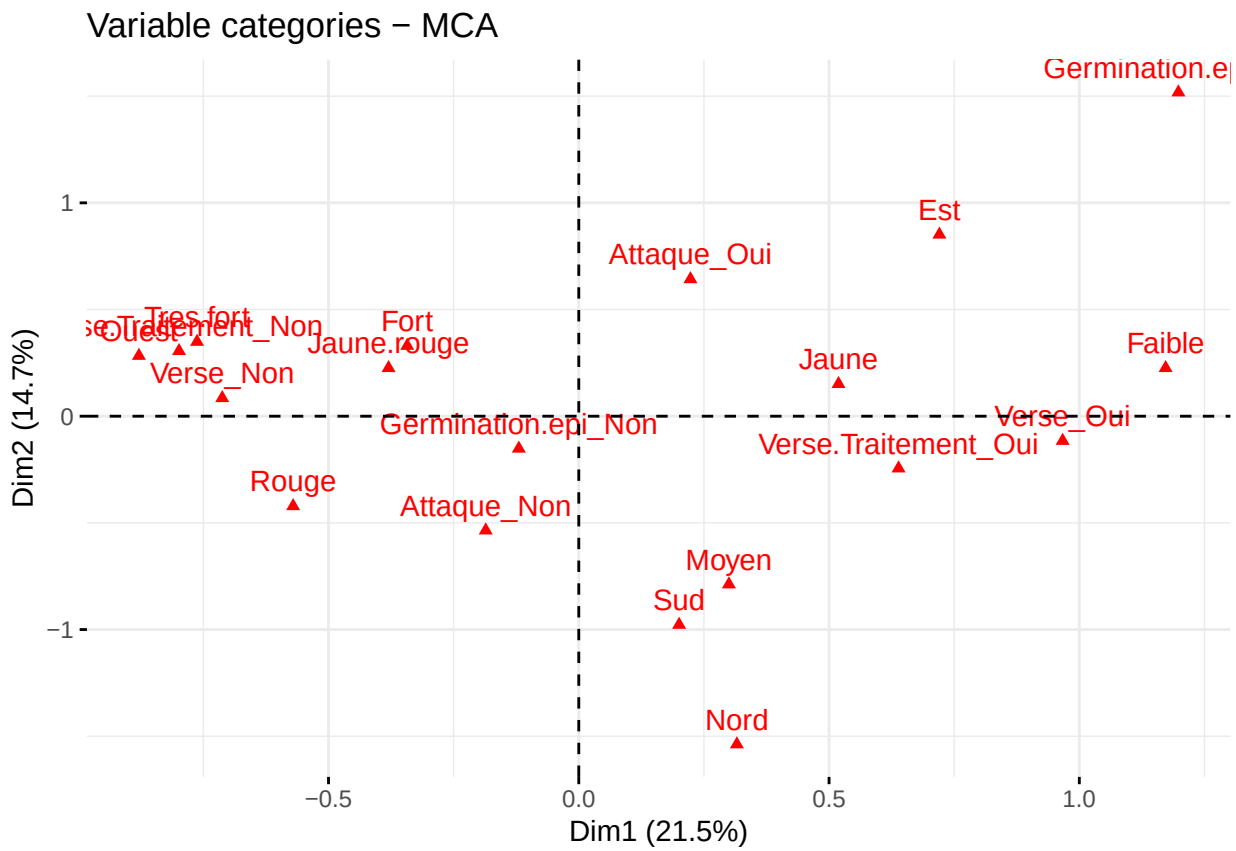
```
plot(resultat_acm, choix="var")
```



Ce graphique représente la corrélation entre les différentes variables et les dimensions principales étudiées. Il faut mettre en évidence que ces 2 dimensions ne regroupent qu'approximativement 35% de l'Inertie totale. Les conclusions que l'on va faire sur ces données sont donc à prendre avec des pincettes.
On remarque tout de même que la variables Verse est la mieux projetée sur la première dimension. Pour la seconde dimension, on peut dire que la Parcelle est pas mal projetée.

Nous allons vérifier cela avec d'autres graphiques qui nous donneront plus d'informations visuelles sur l'analyse de ces données.

```
fviz_mca_var(resultat_acm)
```



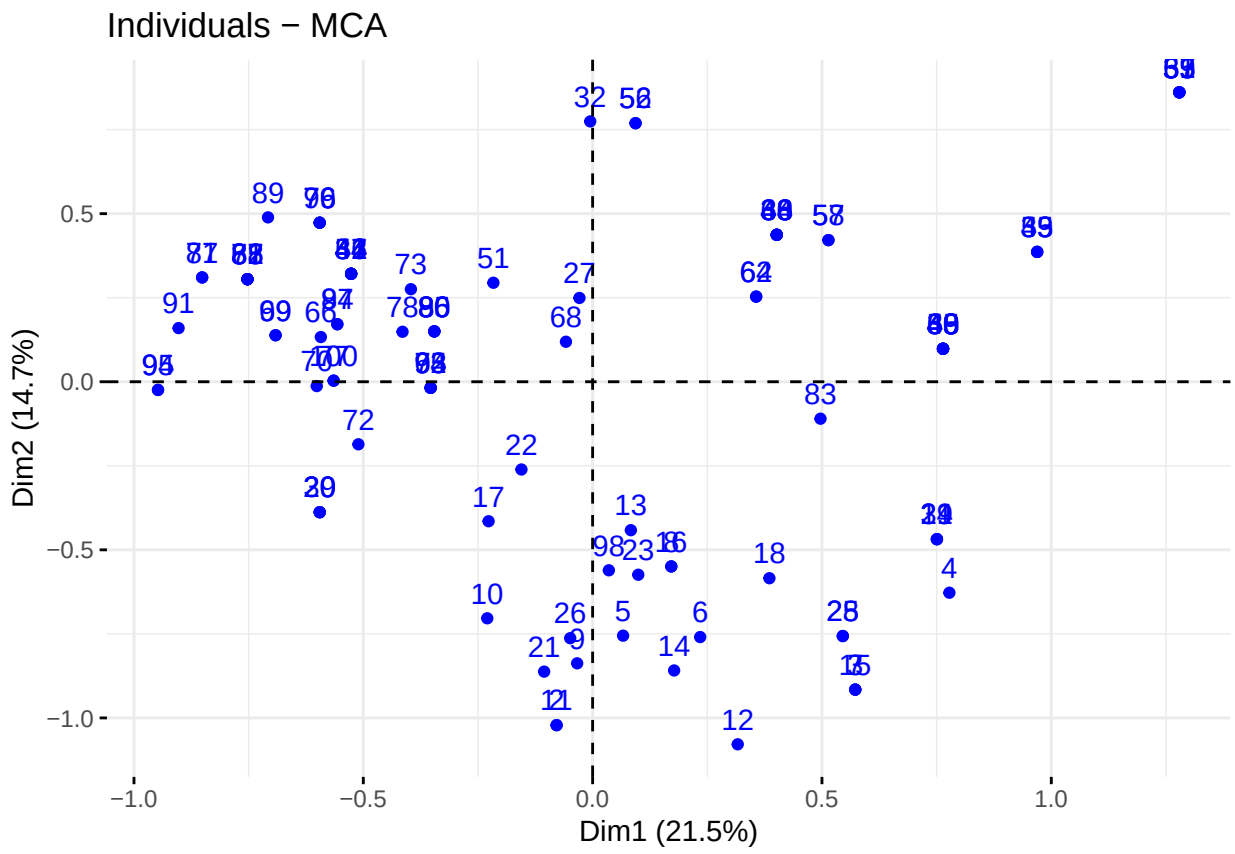
Graphique des Catégories de Variables
Ce graphique montre la projection des catégories des variables qualitatives sur les deux premières dimensions de l'ACM.

Observations :

Dim 1 (21.5% de la variance) : Cette dimension oppose les catégories relatives à la germination sur épi (Germination.epi_Non), à la présence d'attaque d'insectes (Attaque_Non), et à la non-inclinaison du plant (Verse_Non) d'un côté, aux catégories telles que Verse_Oui (plant incliné ou tombé) et Verse.Traitement_Oui (inclinaison après traitement) de l'autre côté. Cela suggère que cette dimension pourrait être associée à la robustesse et à la santé des plants de maïs.

Dim 2 (14.73% de la variance) : Cette dimension semble séparer les couleurs des plants (Jaune, Rouge), avec Jaune.rouge proche de l'origine, suggérant que cette catégorie peut être moins distincte ou moins informative.

```
fviz_mca_ind(resultat_acm)
```



Graphique des Individus
Le deuxième graphique montre la projection des 100 pieds de maïs sur les mêmes dimensions.

Observations :

Les points sont dispersés le long des deux axes, indiquant la variabilité dans les caractéristiques qualitatives des plants de maïs. Les individus à droite ou à gauche sur l'axe de Dim 1 peuvent être différenciés par leur santé et ceux plus haut ou plus bas sur Dim 2 pourraient être différenciés par la couleur de leur plant.

2.4 Le retour du Titanic (ACM)

Un historien a réalisé une étude des données en rapport avec le naufrage du Titanic. Les résultats sont dans le fichier de données titanic.csv.
Que peut-on en déduire ?

La colonne 1 est l'identificateur des personnes.
La colonne 2 correspond à la classe de cabine, selon les codes
0 = équipage, 1 = première classe, 2 = seconde classe, 3 = troisième classe.
La colonne 3 est la catégorie d'age : 0 = enfant, 1 = adulte.
La colonne 4 est le sexe de la personne : 0 = femme, 1 = homme.
La colonne 5 indique si la personne a survécu : 0 = non, 1 = oui.

```
setwd("C:\\Users\\amine\\OneDrive\\Bureau\\EcomStat\\Labo\\Evaluation02\\datasets")

dataTitanic <- read.table(file = "titanic.csv", header=TRUE, sep=";", row.names=1)

summary(dataTitanic)
```

```
##      CLASS            AGE              SEX              SURV
##  Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.000   1st Qu.:1.0000   1st Qu.:1.0000   1st Qu.:0.000
##  Median :1.000   Median :1.0000   Median :1.0000   Median :0.000
##  Mean   :1.369   Mean   :0.9505   Mean   :0.7865   Mean   :0.323
##  3rd Qu.:3.000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :3.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
```

Etant donné que nous réalisons une ACM, il faut tronsformer toutes les valeurs quantitatives en valeurs qualitatives car une ACM s'effectue entre variables qualitatives.

```
dataTitanic$CLASS <- factor(dataTitanic$CLASS, levels = c(0, 1, 2, 3), labels =
                      c("Equipage", "Premiere Classe", "Seconde Classe", "Troisieme Classe"))
dataTitanic$AGE <- factor(dataTitanic$AGE, levels = c(0, 1), labels = c("Enfant", "Adulte"))
dataTitanic$SEX <- factor(dataTitanic$SEX, levels = c(0, 1), labels = c("Femme", "Homme"))
dataTitanic$SURV <- factor(dataTitanic$SURV, levels = c(0, 1), labels = c("Non", "Oui"))

summary(dataTitanic)
```

```
##               CLASS         AGE          SEX        SURV
##  Equipage        :885   Enfant: 109   Femme: 470   Non:1490
##  Premiere Classe :325   Adulte:2092   Homme:1731   Oui: 711
##  Seconde Classe  :285
##  Troisieme Classe:706
```

Nos données sont bien des valeurs interprétées comme étant qualitatives et nous pouvons donc procéder à l'ACM.

```
resultat_acm <- MCA(dataTitanic, graph = FALSE)
```

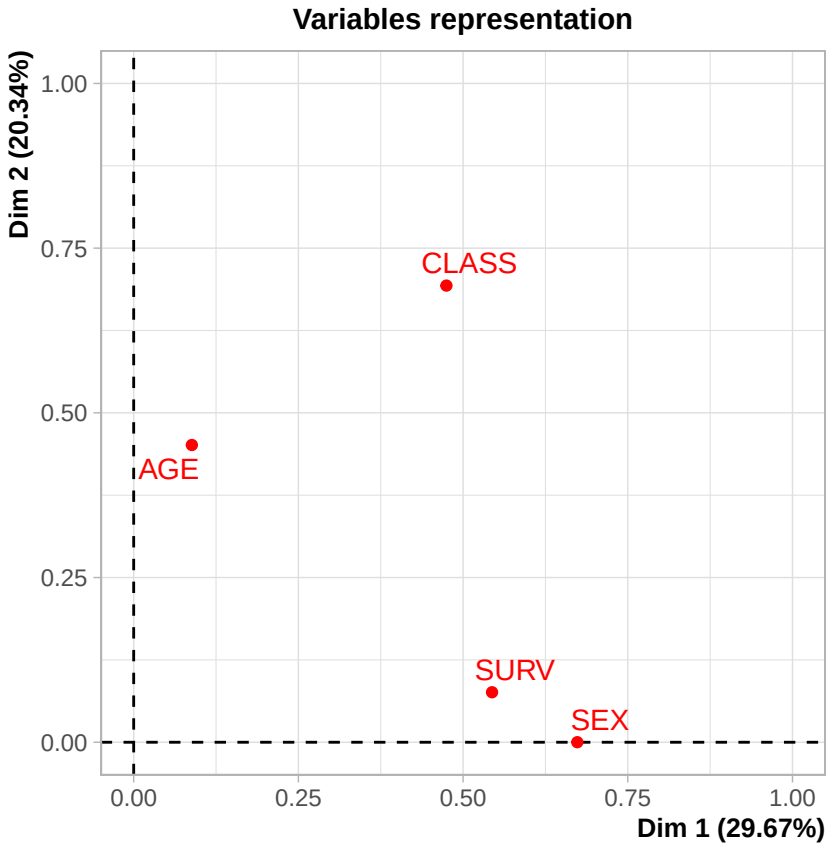
```
resultat_acm$eig
```

```
##       eigenvalue percentage of variance cumulative percentage of variance
## dim 1  0.4450795              29.671965                           29.67196
## dim 2  0.3050437              20.336249                           50.00821
## dim 3  0.2500060              16.667067                           66.67528
## dim 4  0.2050373              13.669154                           80.34443
## dim 5  0.1785152              11.901011                           92.24544
## dim 6  0.1163183               7.754555                          100.00000
```

```
resultat_acm$var$cos2[,1:2]
```

```
##                        Dim 1        Dim 2
## Equipage          0.365218157 1.567075e-01
## Premiere Classe   0.229885143 2.627010e-01
## Seconde Classe    0.063089405 9.485207e-03
## Troisieme Classe  0.008054591 5.407190e-01
## Enfant            0.088298776 4.511706e-01
## Adulte            0.088298776 4.511706e-01
## Femme             0.673361430 2.163951e-05
## Homme             0.673361430 2.163951e-05
## Non               0.543958670 7.584192e-02
## Oui               0.543958670 7.584192e-02
```

```
plot(resultat_acm, choix="var")
```

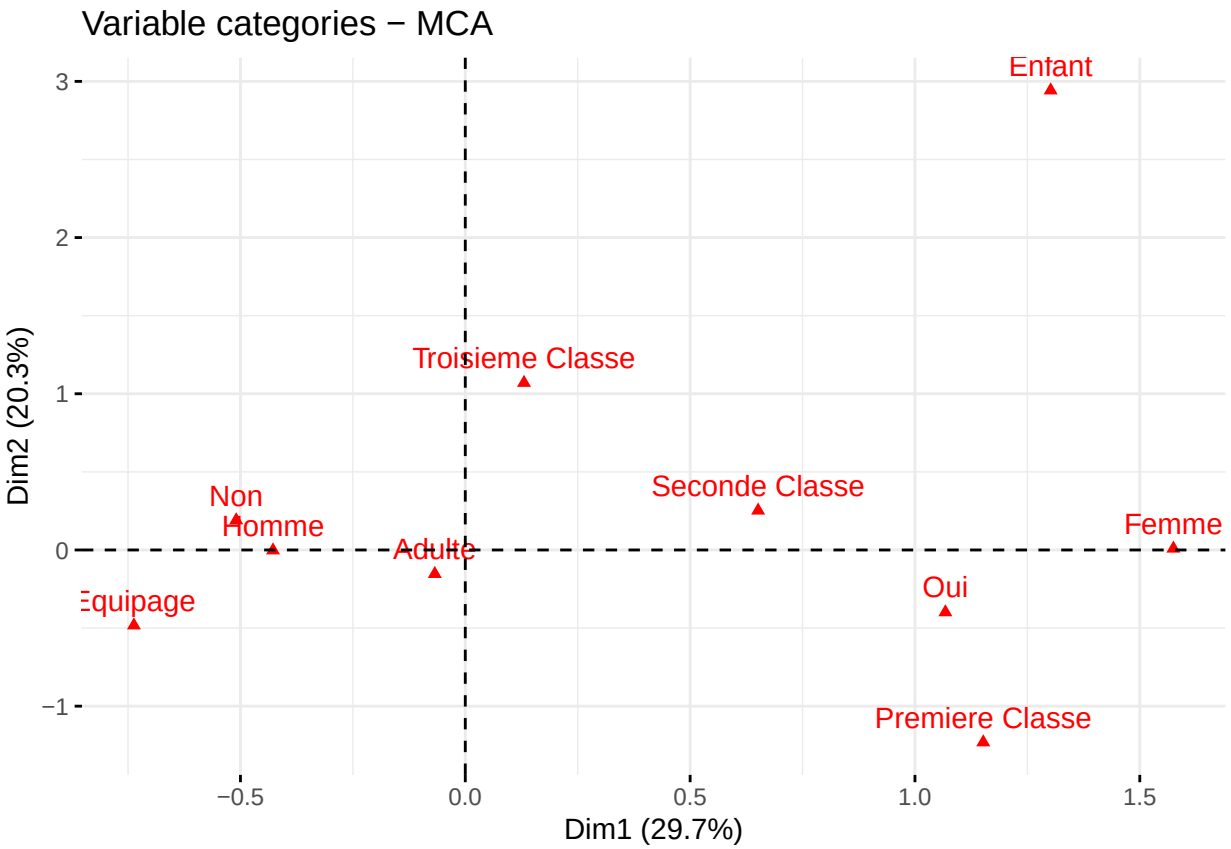


Ce graphique représente la corrélation entre les différentes variables et les dimensions principales étudiées.

On remarque tout de même que les variables de la SURV et du SEX sont les mieux projetées sur la première dimension. Pour la seconde dimension, on peut dire que AGE et CLASS sont pas mal projetées également.

Nous allons vérifier cela avec d'autres graphiques qui nous donneront plus d'informations visuelles sur l'analyse de ces données.

```
fviz_mca_var(resultat_acm)
```



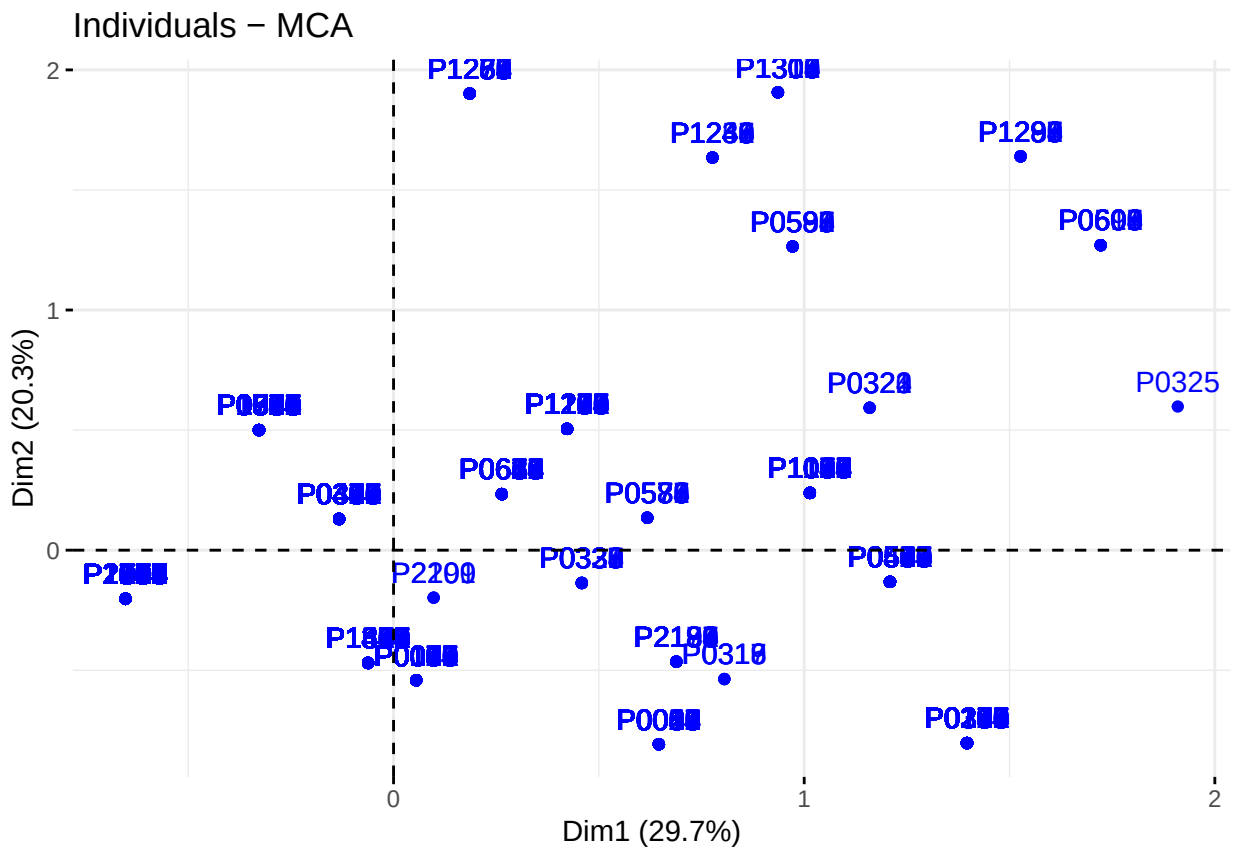
Graphique des Catégories de Variables
Ce graphique représente les différentes catégories pour chaque variable étudiée.

Observations :

Dim 1 (29.7% de la variance) : Cette dimension semble opposer les membres de l'équipage (Equipage) et les passagers en troisième classe (Troisieme Classe) à ceux en première (Premiere Classe) et seconde classe (Seconde Classe). Cela pourrait refléter la différence socio-économique entre les passagers et l'équipage du Titanic.

Dim 2 (20.3% de la variance) : Cette dimension sépare les enfants (Enfant) des adultes (Adulte), et peut-être les hommes des femmes, bien que ces derniers ne soient pas aussi clairement opposés sur cette dimension. Cela peut indiquer des différences dans les taux de survie en fonction de l'âge et du sexe.

```
fviz_mca_ind(resultat_acm)
```



Graphique des Individus
Ce graphique montre la projection des individus (les passagers et membres de l'équipage) sur les mêmes dimensions.

Observations :

Les individus sont dispersés le long des deux dimensions, ce qui suggère une variation dans les caractéristiques et les taux de survie des personnes à bord du Titanic.

Certains points, sont proches des catégories Premiere Classe et Seconde Classe sur la première dimension, ce qui peut indiquer qu'ils appartenaient à ces classes et avaient potentiellement un taux de survie plus élevé.

Conclusion

L'ACM indique des disparités potentielles dans le taux de survie basées sur la classe socio-économique (première, seconde et troisième classe, ou membre de l'équipage), l'âge (enfant ou adulte) et le sexe (homme ou femme). Historiquement, nous savons que les femmes et les enfants ont eu la priorité pour les canots de sauvetage et que les passagers des classes supérieures avaient un meilleur accès aux ressources de survie. Ces graphiques semblent refléter ces faits historiques.

3 Les classifications : CAH et HCPC

3.1 Histoire d'eaux (minérales)

On revient sur des données sur la teneur en divers éléments chimiques des eaux minérales de diverses provenances (plates ou gazeuses) commercialisées en France en utilisant cette fois le fichier Eaux2010.txt.

On demande de chercher à établir une classification de ces eaux minérales. On demande notamment de comparer les 4 méthodes classiques de regroupement. Une étude HCPC apporte-t-elle des informations complémentaires ?

3.1.1 CAH

```
library(cluster)
```

```
## Warning: le package 'cluster' a été compilé avec la version R 4.2.3
```

```
setwd("C:\\Users\\amine\\OneDrive\\Bureau\\EcomStat\\Labo\\Evaluation02\\datasets")

dataEaux <- read.table("Eaux2010.txt", sep="\t", header=TRUE, row.names=7)

summary(dataEaux)
```

```
##       HCO3             SO4               Cl               Ca
##  Min.   :   5.0   Min.   :   0.0   Min.   :  0.26   Min.   :  1.00
##  1st Qu.: 133.0   1st Qu.:   7.0   1st Qu.:  4.50   1st Qu.: 27.00
##  Median : 280.0   Median :  17.0   Median :  9.00   Median : 55.00
##  Mean   : 487.3   Mean   : 102.5   Mean   : 40.29   Mean   : 92.91
##  3rd Qu.: 401.5   3rd Qu.:  46.0   3rd Qu.: 24.00   3rd Qu.:109.50
##  Max.   :4263.0   Max.   :1479.0   Max.   :649.00   Max.   :555.00
##  NA's   :6        NA's   :10       NA's   :12       NA's   :2
##        Mg               Na              Pays              Nature
##  Min.   :  0.10   Min.   :   1.00   Length:113         Length:113
##  1st Qu.:  5.00   1st Qu.:   5.00   Class :character   Class :character
##  Median : 10.00   Median :  12.00   Mode  :character   Mode  :character
##  Mean   : 22.92   Mean   :  84.78
##  3rd Qu.: 31.75   3rd Qu.:  29.50
##  Max.   :160.00   Max.   :1744.00
##  NA's   :3        NA's   :6
```

```
dataEauxCAH <- dataEaux[, c("HCO3", "SO4",  "Cl",   "Ca",   "Mg",   "Na")]

remplaceNAparMOY<-function(x)
{
  return ( ifelse(is.na(x), mean(x,na.rm = TRUE), x)  )
}

dataEauxCAH <- apply(dataEauxCAH, 2, remplaceNAparMOY)
dataEauxCAH
```

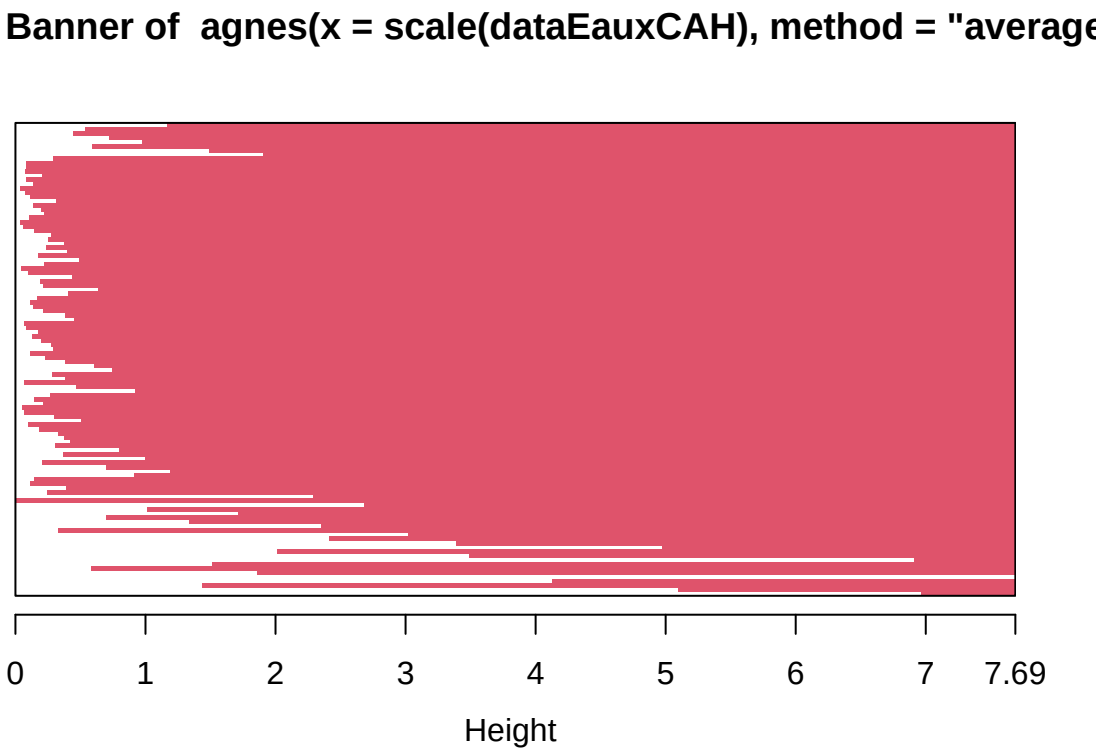
```
##          HCO3       SO4        Cl        Ca        Mg         Na
## Rom  410.0000  293.0000   8.00000 144.00000  66.00000   14.00000
## Voa  256.0000  224.0000  19.00000 115.00000  39.00000   13.00000
## Spa  110.0000   65.0000   5.00000   4.00000   1.00000    3.00000
## Kam  487.3271    5.0000   4.50000   5.00000   9.00000   10.00000
## Cub  354.0000   16.0000  22.00000 112.00000   3.00000   84.78364
## Lou  367.0000    7.0000  35.00000  12.00000  73.00000   21.00000
```

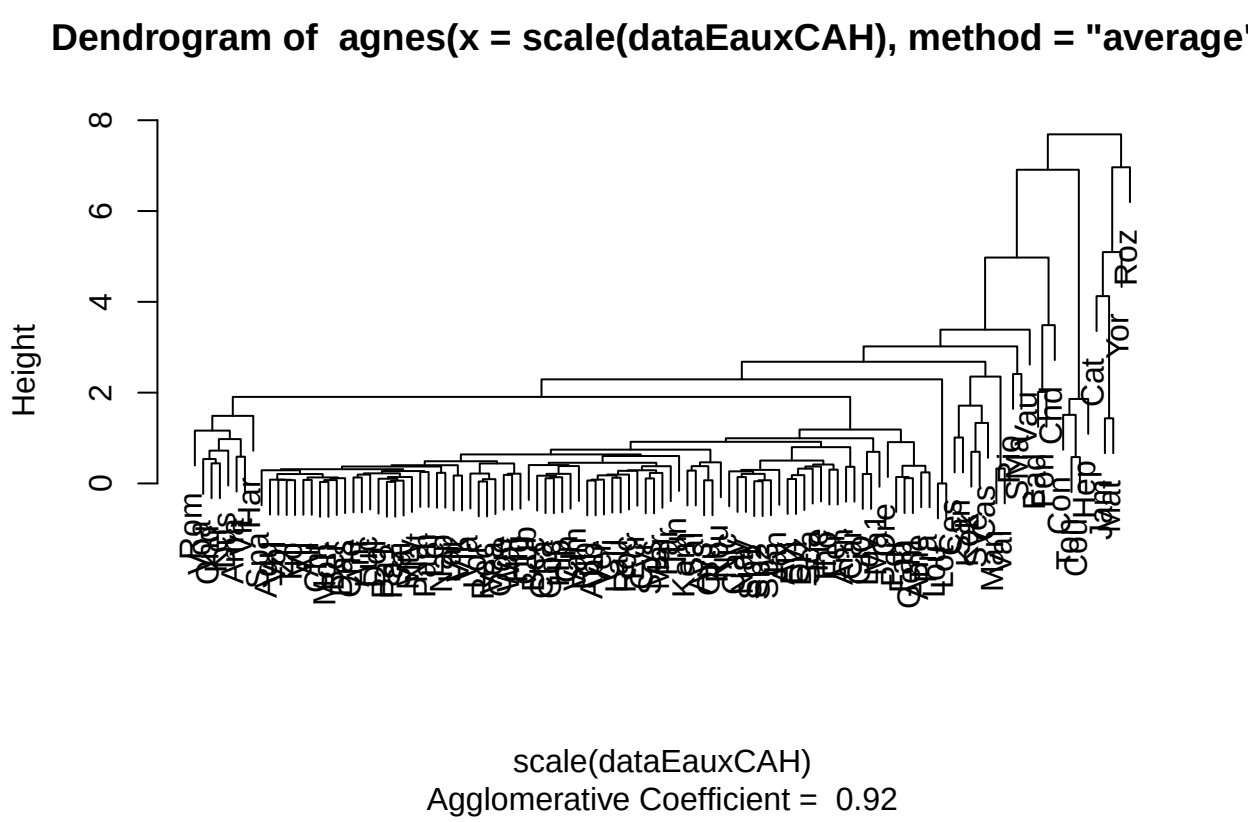
```
## Nic  304.0000   13.0000  21.00000  47.00000  32.00000   18.00000
## Via  135.0000   24.0000  19.00000  27.00000  15.00000   18.00000
## Aqu   78.0000    8.0000  14.00000  14.00000   7.00000   12.00000
## Rib  168.0000   24.0000   8.00000  55.00000   5.00000    9.00000
## Ver  196.0000   18.0000   6.00000  58.00000   6.00000   13.00000
## Vil   59.0000    7.0000   6.00000  16.00000   2.00000    9.00000
## Abb  160.0000   19.0000   7.00000  50.00000   6.00000    5.00000
## Aix  341.0000   27.0000   3.00000  84.00000  23.00000    2.00000
## Ama  312.0000  372.0000  40.29267 176.00000  46.00000   28.00000
## Bec  263.0000   23.0000   9.00000  91.00000   5.00000    3.00000
## Cay  287.0000    3.0000   5.00000  44.00000  24.00000   23.00000
## Ces 1200.0000    8.0000  20.00000 180.00000  50.00000  250.00000
## Cha  298.0000    9.0000  23.00000  96.00000   6.00000   11.00000
## Che  151.0000    3.0000   5.00000  45.00000   3.00000    5.00000
## Cht   13.0000    1.0000   1.00000   3.00000   1.00000    2.00000
## Con  386.0000 1058.0000   6.00000 451.00000  66.00000    8.00000
## Cri  200.0000   15.0000   8.00000  70.00000   2.00000    4.00000
## Cyr  250.0000    5.0000  20.00000  71.00000   6.00000   11.00000
## Ecr  173.0000   51.0000   1.00000  63.00000  10.00000    1.00000
## Evi  357.0000   10.0000   2.00000  78.00000  24.00000    5.00000
## Fie  360.0000   47.0000  28.00000  89.00000  31.00000   17.00000
## Hep  403.0000 1479.0000  40.29267 555.00000 110.00000   84.78364
## Hip  256.0000    6.0000  23.00000  86.00000   3.00000   18.00000
## Lau  186.0000   10.0000  16.00000  64.00000   4.00000    9.00000
## Luc   78.0000    8.0000   2.00000  27.00000   1.00000    1.00000
## Mar  219.0000   39.0000  28.00000  68.00000  11.00000   21.00000
## Mon    5.0000    9.0000   1.00000   3.00000   1.00000    2.00000
## Oge  183.0000   16.0000  44.00000  48.00000  11.00000   31.00000
## Ond  398.0000  218.0000  15.00000 157.00000  35.00000    8.00000
## Ore  731.0000   12.0000   8.00000 200.00000  20.00000    6.00000
## Par  223.0000  102.5437  19.00000  65.00000   5.00000   20.00000
## Per  348.0000   51.0000  31.00000 140.00000   4.00000   14.00000
## Rou  630.0000    3.0000   3.00000   1.70000   0.50000    3.10000
## Tho  332.0000   14.0000   8.00000 103.00000  16.00000    5.00000
## Ton  329.0000 1342.0000   9.00000 528.00000  78.00000    9.00000
## Vau  280.0000  620.0000  58.00000 230.00000  66.00000   40.00000
## Vit  402.0000  306.0000  15.00000 202.00000  36.00000    3.00000
## Vol   64.0000    7.0000   8.00000  10.00000   6.00000    8.00000
## Wat  135.0000   24.0000  40.29267  35.00000  11.00000    3.00000
## Zil   68.0000    5.0000  15.00000  11.00000   5.00000   15.00000
## Lot  367.0000    7.0000  35.00000  12.00000  73.00000   21.00000
## Dus  329.0000  110.0000  40.29267  80.00000  40.00000   18.00000
## Ker  487.0000  144.0000  40.29267 144.00000  34.00000   32.00000
## Drz  416.0000    6.0000   4.00000  86.00000  27.00000   12.00000
## Jan  354.0000    7.0000   2.00000  63.00000  34.00000    2.00000
## Ali   43.0000    8.0000   1.00000  13.00000   3.00000    2.00000
## Ame  429.0000  102.5437  18.00000 157.00000   8.00000   14.00000
## Ann   29.0000    8.0000  40.29267  12.00000  22.92364    1.00000
## Ant  133.0000    2.0000   1.00000  34.00000   5.00000    4.00000
## Ber   30.0000    3.0000   1.00000  10.00000   1.00000    1.00000
## Cou  176.0000 1420.0000   1.00000 533.00000  66.00000    1.00000
## Dol   88.0000   20.1000   0.90000  23.70000   8.00000    1.40000
## Fab  361.0000   28.0000  26.00000 134.00000   5.00000   14.00000
## Fer  311.0000   14.0000  18.00000  73.00000  18.00000   13.00000
```

```
## Gem  401.0000   42.0000  24.00000 137.00000   7.00000   20.00000
## Gio  113.0000  102.5437  86.00000  29.00000  16.00000   45.00000
## Goc  281.0000   26.0000  20.00000  93.00000   4.00000   19.00000
## Lar   31.0000    1.4000   0.40000   1.00000   0.30000   84.78364
## Lei   59.0000   14.0000  40.29267  21.00000   2.00000    2.00000
## Lev   74.0000  102.5437  73.00000  15.00000  14.00000   39.00000
## Lor  154.0000   21.0000   1.00000  35.00000  15.00000    1.00000
## Mtv  139.0000   27.0000   7.00000  33.00000  10.00000   14.00000
## Nap   16.0000    5.0000  26.00000   4.00000   2.00000   16.00000
## Nat  213.0000    3.0000  16.00000  35.00000   4.00000   32.00000
## Pam  108.0000   21.0000   9.00000  34.00000   7.00000   63.00000
## Pa1  100.0000   21.0000   7.00000  30.00000   7.00000    7.00000
## Pa2  106.0000   21.0000   9.00000  33.00000   7.00000    6.00000
## Pra  476.0000  102.5437   9.00000 140.00000  12.00000    4.00000
## Rec  161.0000   24.0000   1.00000  37.00000  16.00000    1.00000
## Roc   78.0000    8.0000   8.00000  57.00000   3.00000    5.00000
## Sb1  293.0000    5.0000   3.00000  46.00000  30.00000    7.00000
## Sb2  293.0000   17.0000   2.00000  45.00000  30.00000    7.00000
## Sca  378.0000   67.0000   4.00000 107.00000  27.00000    6.00000
## Tio   90.0000    7.0000  10.00000  15.00000   5.00000   13.00000
## Vam  487.3271    5.0000   0.26000  92.90991   2.70000    1.35000
## Vea  150.0000   18.0000   3.00000  36.00000  13.00000    2.00000
## Ma1  487.3271  102.5437  65.00000  35.00000  22.92364   15.00000
## Car   36.0000   26.0000   7.00000   3.00000  22.92364   15.00000
## Das    5.0000   14.0000   7.00000   1.00000   3.00000    2.00000
## Nay  245.0000   27.0000   2.00000  45.00000  25.00000    7.00000
## Hig  133.0000    7.0000  40.29267  32.00000   8.00000   45.00000
## Sai  103.0000   11.0000   6.00000  27.00000   7.00000    7.00000
## Ays  487.3271    0.0000   7.00000  27.00000   5.00000   84.78364
## Oru  487.3271   14.0000   9.00000  26.00000   6.00000   84.78364
## Sos  207.0000   10.0000  28.00000  68.00000   6.00000   10.00000
## Tun  244.0000   24.0000  64.00000  80.00000  10.00000   37.00000
## Sal  487.3271    6.0000   7.00000  36.00000   9.00000    7.00000
## Vik  105.0000    6.5000   4.50000   1.60000   0.10000   55.00000
## Cat 2147.0000   48.0000 610.00000  50.00000   8.00000 1136.00000
## Bad 1700.0000  102.5437  40.29267 272.00000 102.00000  180.00000
## Chd 2975.0000   20.0000   7.00000 383.00000  49.00000  240.00000
## Man 1567.0000   45.0000  27.00000  49.00000  26.00000  482.00000
## Rio 1850.0000  180.0000 180.00000  16.00000   4.00000   62.00000
## Roz 1837.0000  230.0000 649.00000 301.00000 160.00000  493.00000
## SMa  812.0000   59.0000 230.00000  71.00000  40.00000  302.00000
## Val 1403.0000   39.0000  27.00000  45.00000  21.00000  453.00000
## Yor 4263.0000  182.0000 329.00000  78.00000   9.00000 1744.00000
## Fon 2150.0000  210.0000  51.00000 169.00000 138.00000  405.00000
## Har  540.0000  333.0000 133.00000 186.00000  48.00000  129.00000
## Kek 1150.0000  102.5437  40.29267 246.00000  56.00000   36.00000
## Jam 2352.0000  112.0000 251.00000 105.00000  34.00000  900.00000
## Cas 1159.0000    8.0000  18.00000 304.00000  17.00000   47.00000
## Cin  339.0000   11.0000  16.00000 107.00000   7.00000    9.00000
## Mat 2206.0000  281.0000 298.00000 180.00000  56.00000  780.00000
## San 1360.0000  102.5437  40.29267  92.90991  60.00000  290.00000
## Sb3  311.0000    4.0000   2.00000  50.00000  29.00000    6.00000
## Sve 1220.0000  102.5437  40.29267 228.00000  38.00000   84.78364
```

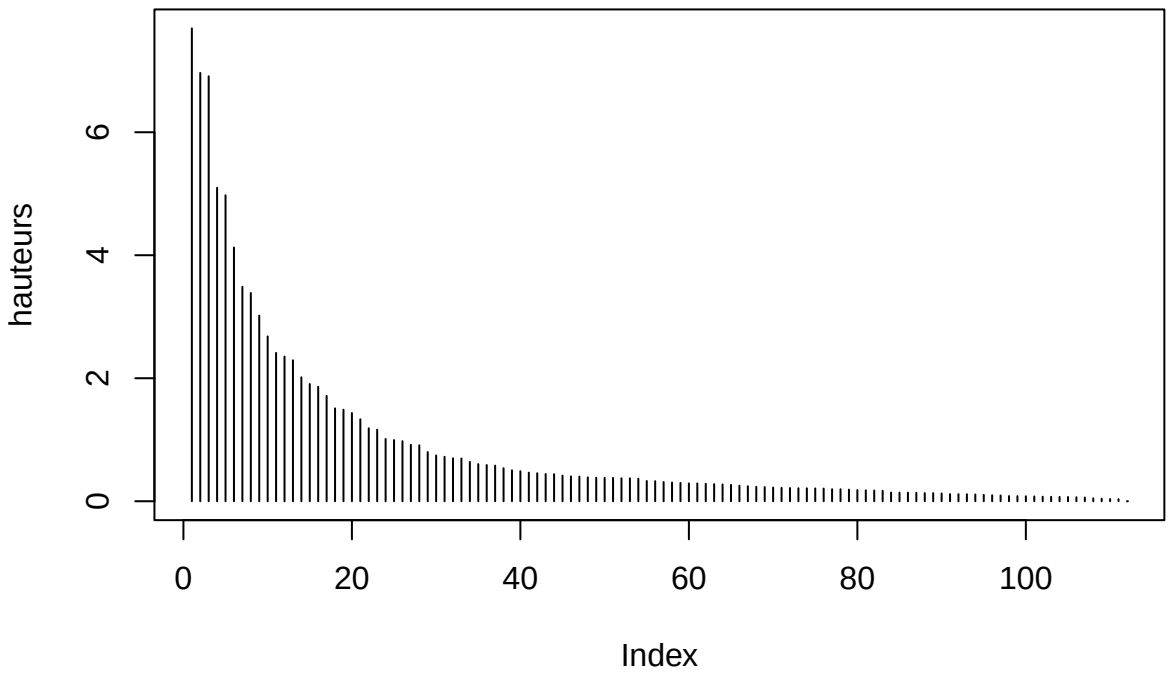
Pour pouvoir effectuer notre comparaisons, nous allons utiliser la méthode Agnes qui permet de construire une hiérarchie arborescente. On va commencer avec la méthode « average », la méthode de la moyenne non pondérée.

```
classificationAverage <- agnes(scale(dataEauxCAH), method = "average")
plot(classificationAverage)
```






```
classificationAverage.h <- as.hclust(classificationAverage)
plot(rev(classificationAverage.h$height), type="h", ylab="hauteurs")
```



On obtient un très bon coefficient d'agglomération à 0,92 ce qui fait une partition déjà très discriminante. Cependant, on remarque quand même que plusieurs eaux se retrouvent seules très haut dans la répartition.

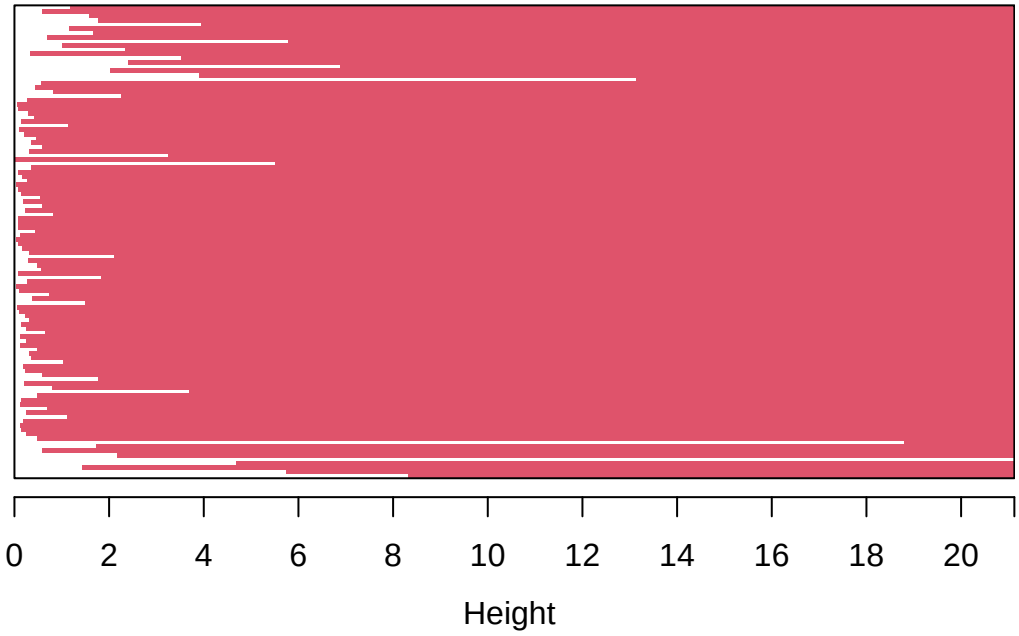
Pour essayer de réparer cela, on va executer la fonction agnes avec la méthode « ward » qui a pour but d'agréger en faisant perdre le moins d'inertie inter-classe en se basant sur un critère de regroupement en plus, le poids. En plus de la distance, la méthode se préoccupe du poids qui permet d'agréger de manière plus précises les classes entre elles.

Graphique des hauteurs des fusions des classes

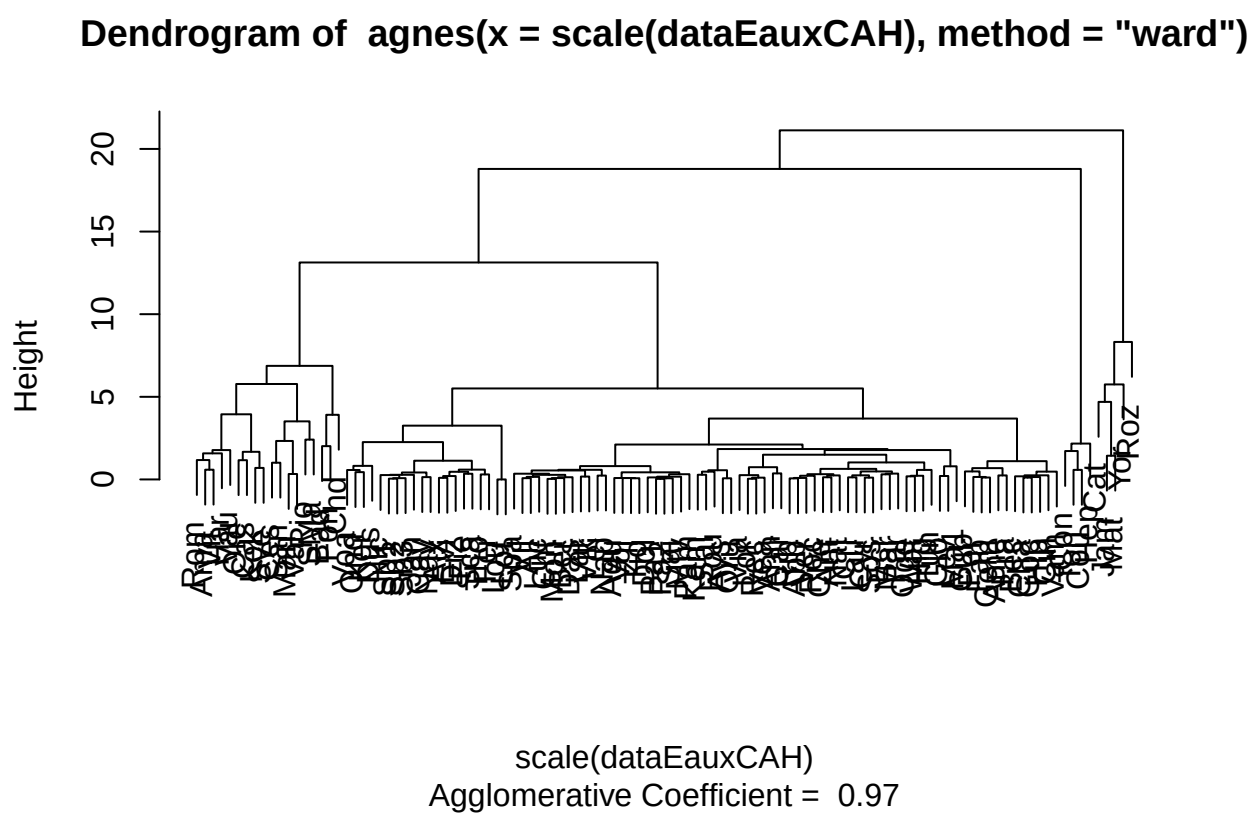
L'objectif de ce graphique est de nous montrer la ou les hauteurs de coupe raisonnable pour une hiérarchisation. Pour être une bonne hauteur, il faut qu'il y ait une différence significative entre la hauteur x et sa hauteur x-1. Dans ce cadre-ci, la coupe à la hauteur 3 et 4 est faisable car il y a une réelle différence de hauteurs entre les deux barres. Les hauteurs 4 et 5 ne seraient pas 2 coupes assez signifiantes pour remarquer un changement, leurs hauteurs sont quasiment égales.

```
classificationWard <- agnes(scale(dataEauxCAH), method = "ward")
plot(classificationWard)
```

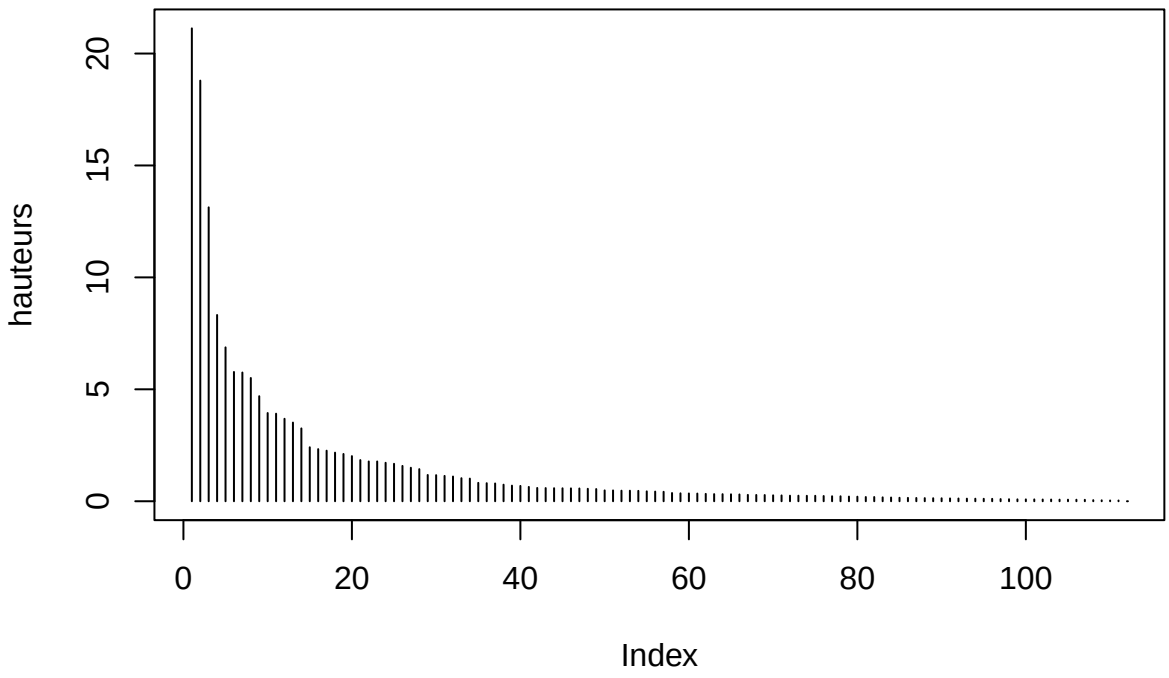
Banner of agnes(x = scale(dataEauxCAH), method = "ward")



Agglomerative Coefficient = 0.97

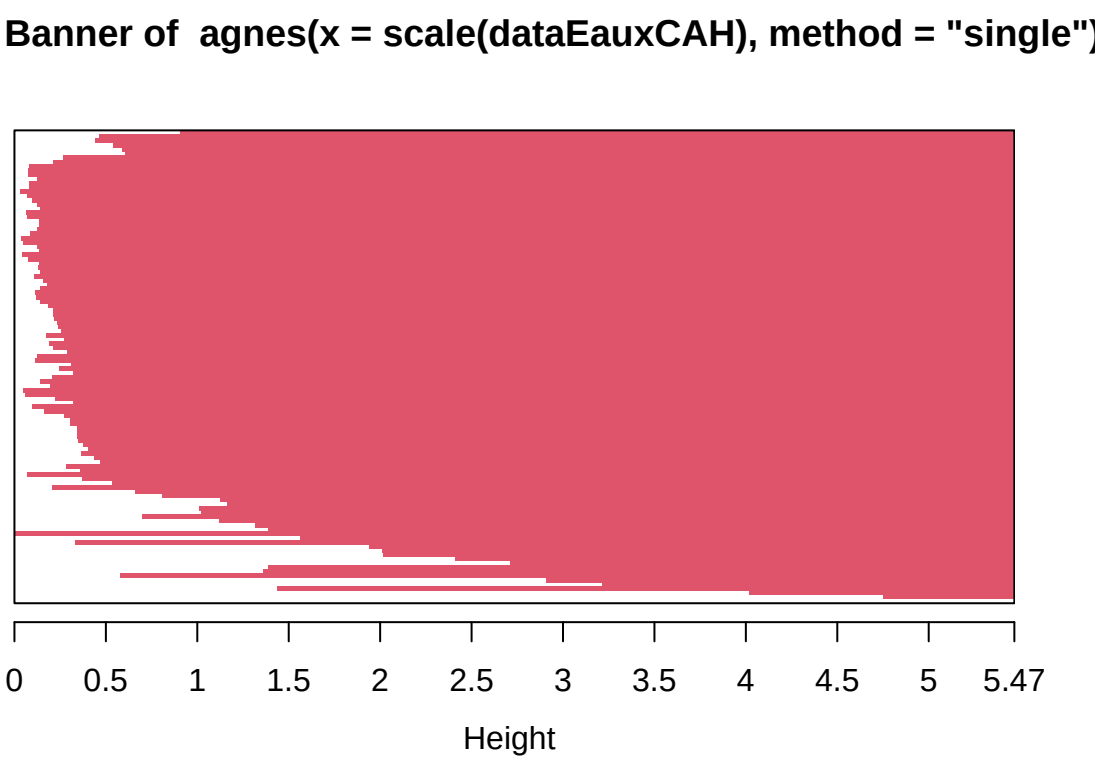


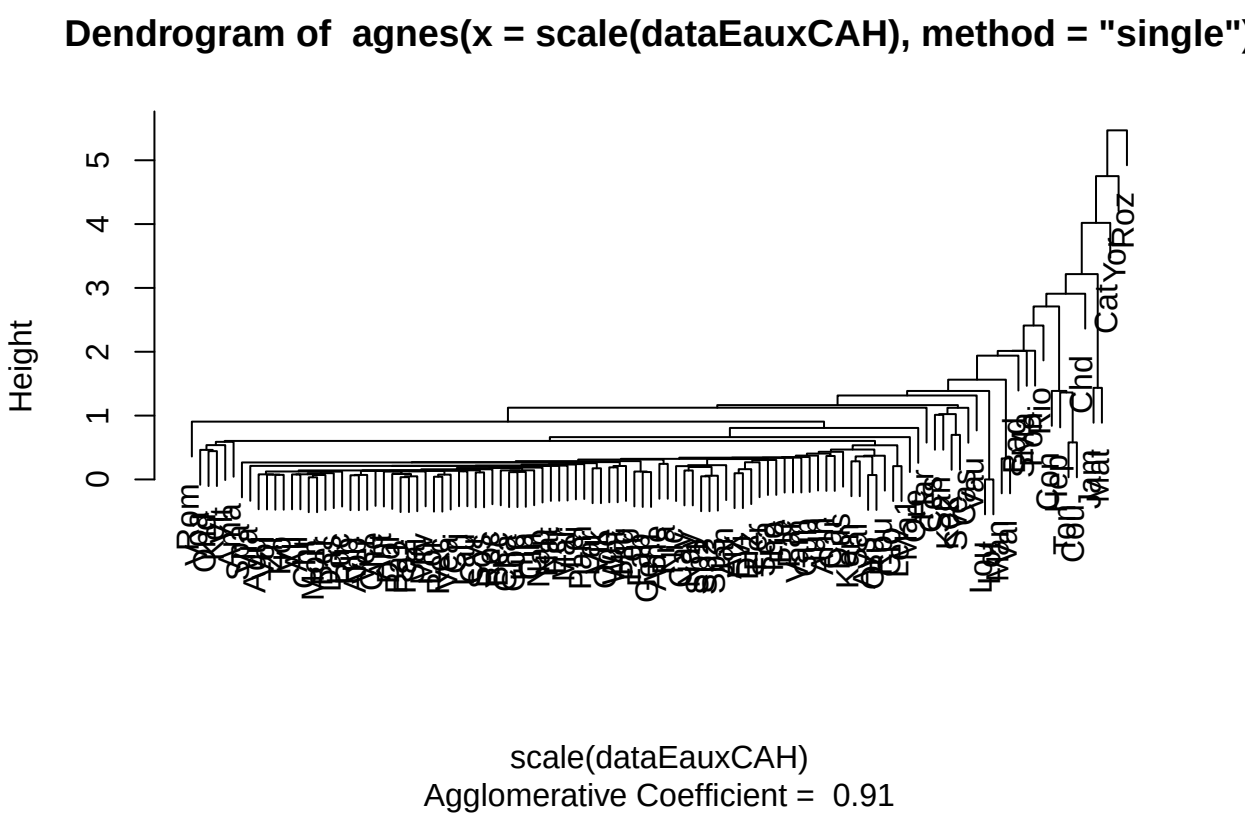

```
classificationWard.h <- as.hclust(classificationWard)
plot(rev(classificationWard.h$height), type="h", ylab="hauteurs")
```



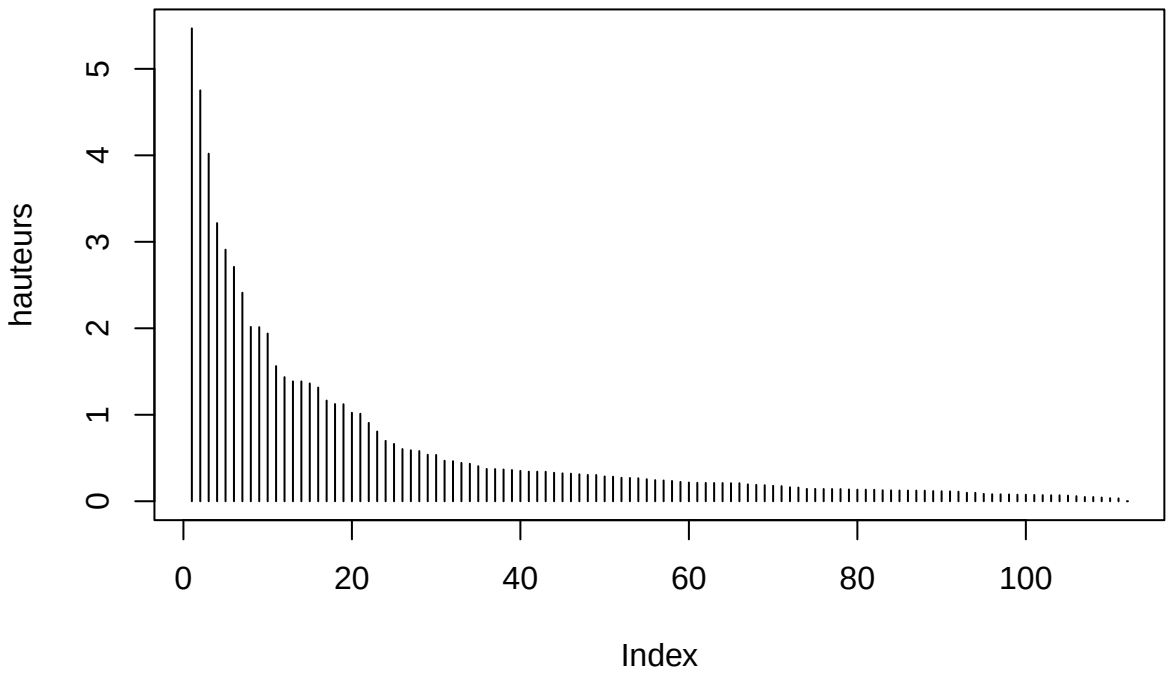
On obtient un bien meilleur arbre, sans eaux qui se trouvent seules très hautes dans la hiérarchie et avec une classification plus précise également. De plus, on remarque que le coefficient d'agglomération est égal à 0,97 ce qui est mieux que celui de notre graphique précédent, ce qui prouve qu'il est encore meilleur.

```
classificationSingle <- agnes(scale(dataEauxCAH), method = "single")
plot(classificationSingle)
```


Banner of agnes(x = scale(dataEauxCAH), method = "single")
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.47
Height
Agglomerative Coefficient = 0.91

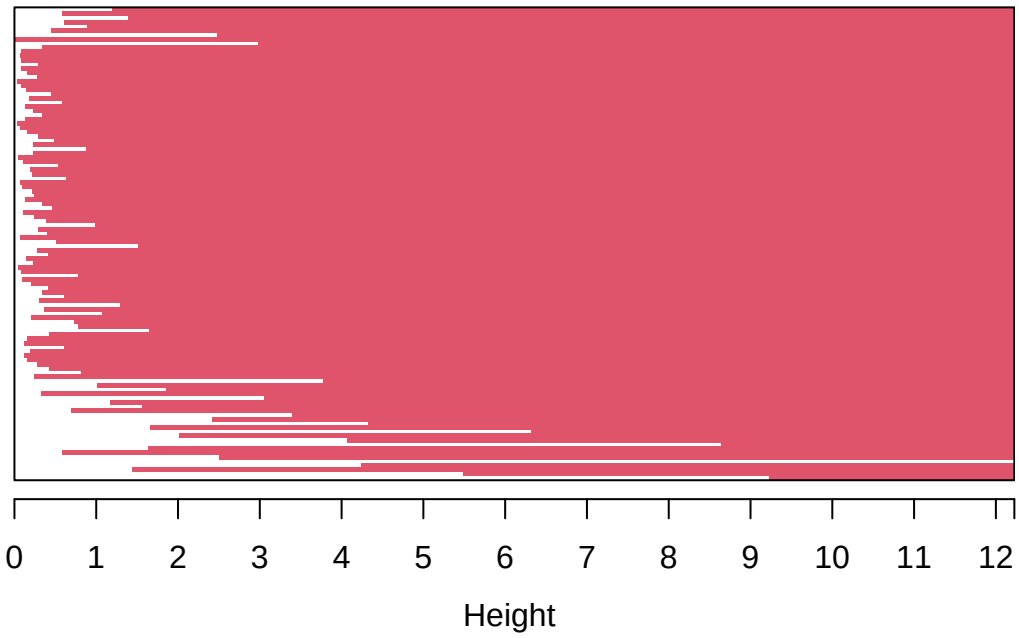



```
classificationSingle.h <- as.hclust(classificationSingle)
plot(rev(classificationSingle.h$height), type="h", ylab="hauteurs")
```

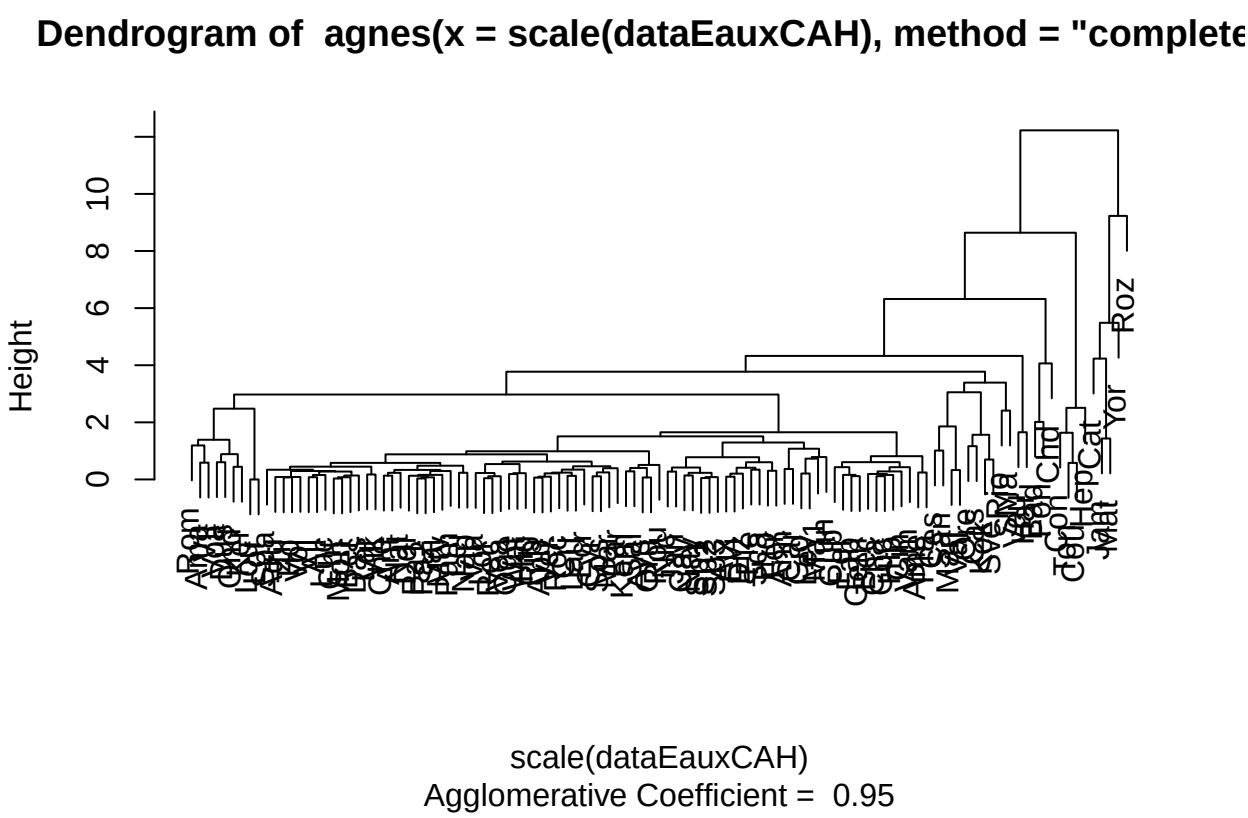



```
classificationComplete <- agnes(scale(dataEauxCAH), method = "complete")
plot(classificationComplete)
```

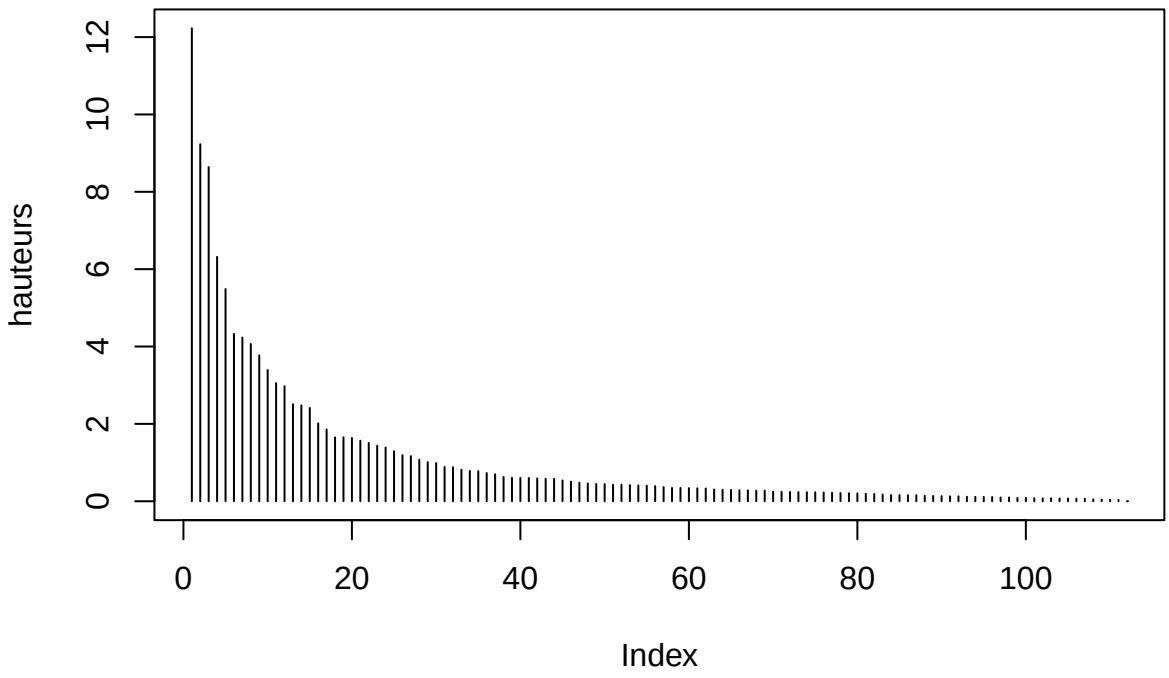
Banner of agnes(x = scale(dataEauxCAH), method = "comple



Agglomerative Coefficient = 0.95




```
classificationComplete.h <- as.hclust(classificationComplete)
plot(rev(classificationComplete.h$height), type="h", ylab="hauteurs")
```

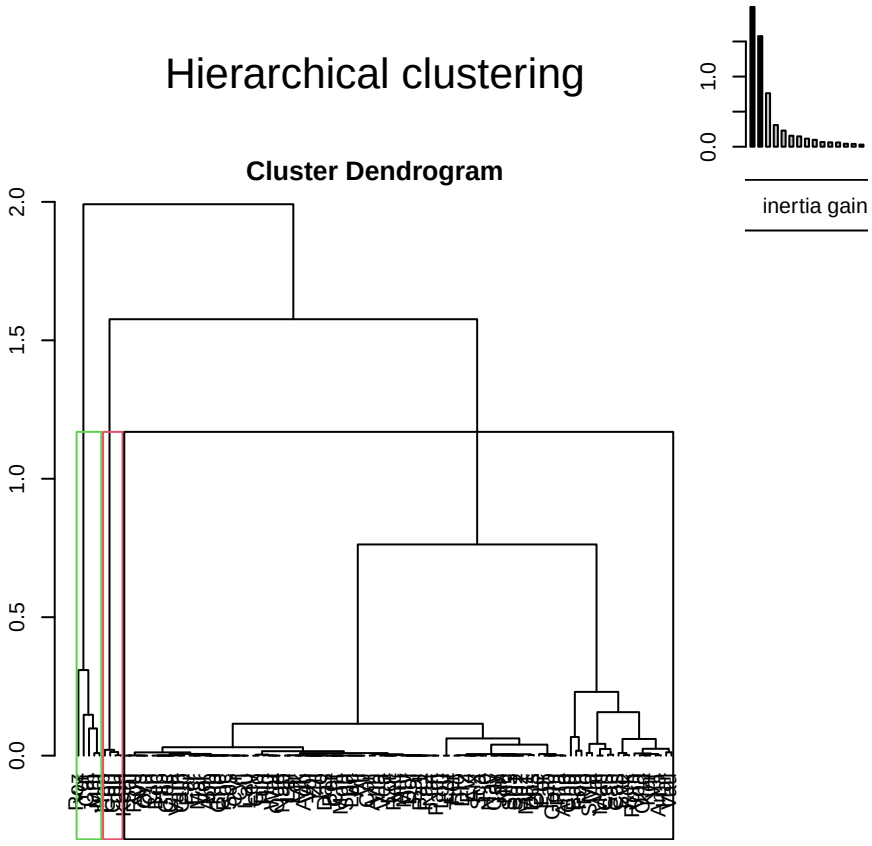



```
#la decoupe et order
```

3.1.2 HCPC

Maintenant nous allons effectuer une HCPC qui permet de faire une classification en se basant sur une analyse factorielle qui est, dans le cas présent, une ACP.

```
classification.acp <- PCA(dataEauxCAH, graph = FALSE)

classification.hcpc <- HCPC(classification.acp, consol = F, graph = FALSE)

plot(classification.hcpc, choice = "tree")
```



Le programme nous a construit cette classification qui nous propose de couper pour avoir le nombre de clusters que l'on souhaite. Pour savoir comme découper ce cluster, on regarder le graphique situé en haut à droite qui nous permet de savoir quelle est la meilleure répartition (moins de perte inter classe) et à partir de quand cela devient inutile de découper en plus de clusters. Dans notre cas de figure, on voit qu'après 3, on se rend compte que cela devient très dur de découper.

Après avoir découpé au bon endroit, on obtient ce graphique :

```
#image avec les clusters apres la decoupe
```

Maintenant, il est temps de chercher quels sont les variables qui différencient ces 3 groupes. Pour cela, on utiliser cette fonction :

```
classification.hcpc$desc.var$quanti.var
```

```
##           Eta2      P-value
## SO4  0.8596980 1.226062e-47
## Cl   0.7547079 2.707669e-34
## Na   0.7226910 2.306842e-31
## Ca   0.5693836 7.498594e-21
## HCO3 0.4482367 6.257541e-15
## Mg   0.2137150 1.806758e-06
```

On obtient des p-values qui nous permettent d'établir un rapport de corrélation (c'est comme si on faisait une anova à un facteur) . Cela nous permet donc de définir la ou les variables qui expliquent le mieux la différence entre les différents clusters.

On peut également obtenir le parangon de chaque classe (c'est-à-dire l'individu le plus proche du centre de gravité de chaque cluster). Cela nous permet d'avoir le meilleur individu potentiel pour pouvoir comparer les différences entre individus de chaque classe. Pour cela, on doit faire cette méthode :

```
classification.hcpc$desc.ind$para
```

```
## Cluster: 1
##       Fer       Aix       Evi       Tho       Mar
## 0.2359312 0.3315191 0.3580195 0.3803324 0.4121528
## ------------------------------------------------------------
## Cluster: 2
##       Ton       Cou       Con       Hep
## 0.1654119 0.6978545 1.3174169 1.3345781
## ------------------------------------------------------------
## Cluster: 3
##      Mat      Jam      Cat      Yor      Roz
## 1.837931 2.086248 2.803589 4.495310 5.285264
```

3.2 Les vins italiens

> Une étude internationale porte sur des vins italiens afin de déterminer si les classifications établies par les viticulteurs reposent sur des données objectives ou relèvent plutôt de traditions et autres arguments subjectifs. Cette étude dit essentiellement :

These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.
The attributes are
1) Alcohol
2) Malic acid
3) Ash
4) Alcalinity of ash
5) Magnesium
6) Total phenols
7) Flavanoids
8) Nonflavanoid phenols
9) Proanthocyanins
10)Color intensity
11)Hue
12)OD280/OD315 of diluted wines
13)Proline"

Le fichier "winedata.csv" contient ces données. On demande une étude des classifications envisageables pour ces vins.

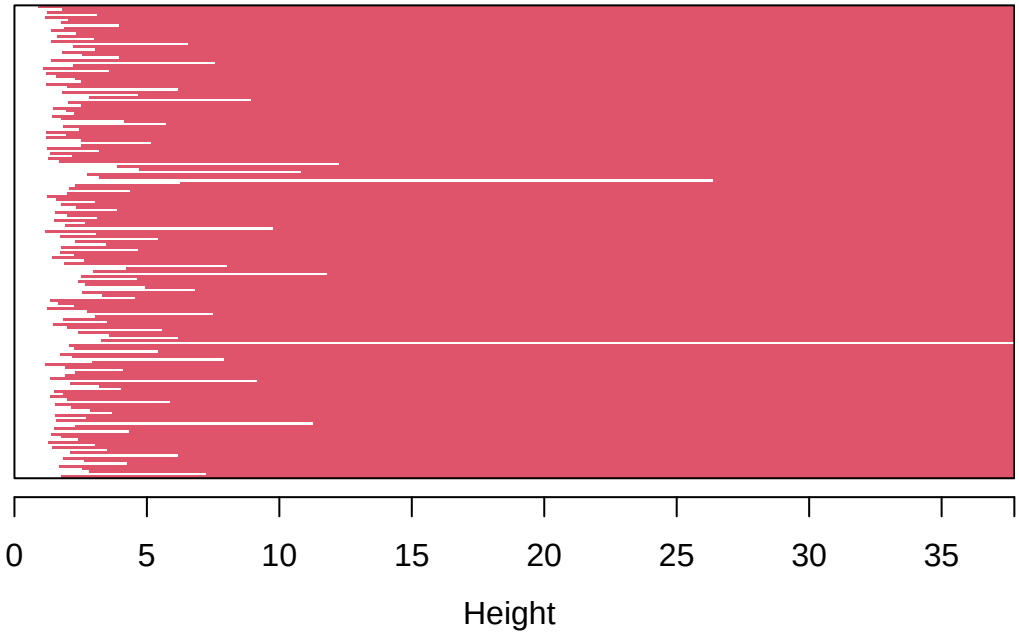
```
setwd("C:\\Users\\amine\\OneDrive\\Bureau\\EcomStat\\Labo\\Evaluation02\\datasets")

dataVin <- read.table(file = "winedata.csv", header=TRUE, sep=";", row.names=1)

summary(dataVin)
```

```
##     Alcohol        Malic.acid         Ash        Alcalinity.of.ash
```

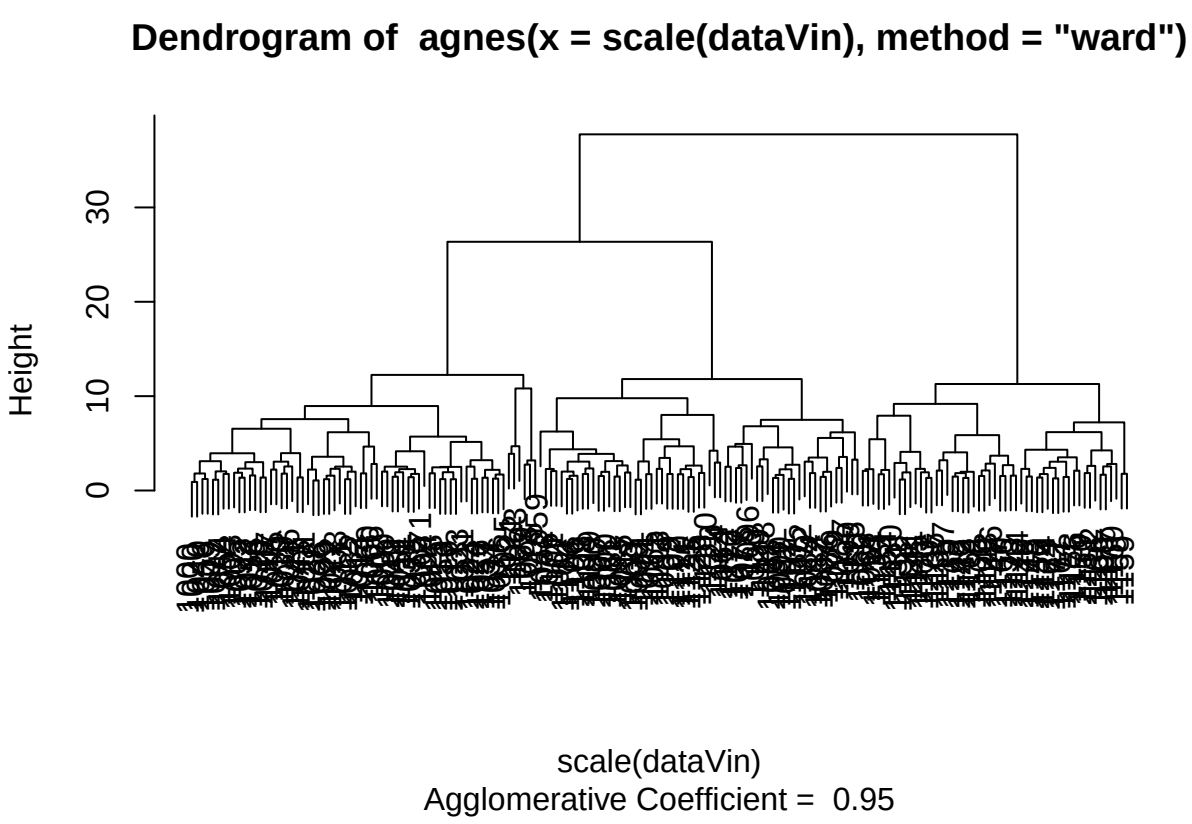
```
##  Min.   :1.000   Min.   :11.03   Min.   :0.740   Min.   :1.360
##  1st Qu.:1.000   1st Qu.:12.36   1st Qu.:1.603   1st Qu.:2.210
##  Median :2.000   Median :13.05   Median :1.865   Median :2.360
##  Mean   :1.938   Mean   :13.00   Mean   :2.336   Mean   :2.367
##  3rd Qu.:3.000   3rd Qu.:13.68   3rd Qu.:3.083   3rd Qu.:2.558
##  Max.   :3.000   Max.   :14.83   Max.   :5.800   Max.   :3.230
##    Magnesium     Total.phenols      Flavanoids    Nonflavanoid.phenols
##  Min.   :10.60   Min.   : 70.00   Min.   :0.980   Min.   :0.340
##  1st Qu.:17.20   1st Qu.: 88.00   1st Qu.:1.742   1st Qu.:1.205
##  Median :19.50   Median : 98.00   Median :2.355   Median :2.135
##  Mean   :19.49   Mean   : 99.74   Mean   :2.295   Mean   :2.029
##  3rd Qu.:21.50   3rd Qu.:107.00   3rd Qu.:2.800   3rd Qu.:2.875
##  Max.   :30.00   Max.   :162.00   Max.   :3.880   Max.   :5.080
##  Proanthocyanins  Color.intensity      Hue          OD280.OD315.of.diluted.wines
##  Min.   :0.1300   Min.   :0.410   Min.   : 1.280   Min.   :0.4800
##  1st Qu.:0.2700   1st Qu.:1.250   1st Qu.: 3.220   1st Qu.:0.7825
##  Median :0.3400   Median :1.555   Median : 4.690   Median :0.9650
##  Mean   :0.3619   Mean   :1.591   Mean   : 5.058   Mean   :0.9574
##  3rd Qu.:0.4375   3rd Qu.:1.950   3rd Qu.: 6.200   3rd Qu.:1.1200
##  Max.   :0.6600   Max.   :3.580   Max.   :13.000   Max.   :1.7100
##     Proline
##  Min.   :1.270
##  1st Qu.:1.938
##  Median :2.780
##  Mean   :2.612
##  3rd Qu.:3.170
##  Max.   :4.000
```

```
classification <- agnes(scale(dataVin), method = "ward")
plot(classification)
```

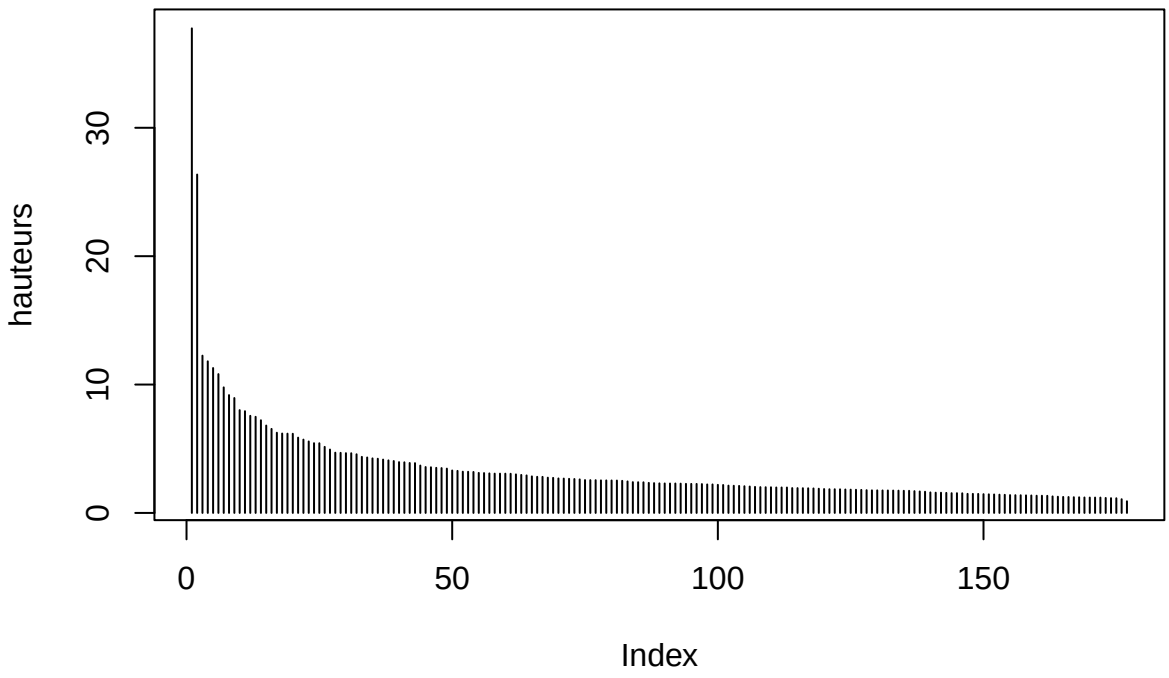
Banner of agnes(x = scale(dataVin), method = "ward")



Agglomerative Coefficient = 0.95




```
classification.h <- as.hclust(classification)
plot(rev(classification.h$height), type="h", ylab="hauteurs")
```




```
#la decoupe et order
```