QUIZ1 ADEI_20-21Q1: Template for solutions to questions

Júlia Gasull

November 10th, 2020

List of Questions

The data set for this exercise contains 396 observations for the mean daily values of the variables included in the table referred to a wastewater treatment plant. The plant has measures on the quality of the wastewater at the entrance of the plant, they are the variables xxx.e, from here they go to a first decantation process (Primary Treatment) where it is intended that they settle the solids in suspensions. Then it goes into the Biological Reactor. This is the most critical part. Here is a biological mud that literally "lives by eating organic matter." It is activated by temperature and aeration. This process is carried out by many species of microorganisms. These microorganisms work at different temperatures. If they are not balanced with the composition of the water, they eat each other. The control variables are those that graduate aeration, temperature, recirculation and purge of the bioreactor. This is the most difficult part: if things go well, dirty water and biological mud enter the bioreactor and end up with clean water and colonies of microorganisms that have consumed the organic matter. Then the water goes through a second decanting process where the microorganisms settle because if the water is not aerated they fall to the ground. And then clean can be poured into the river. The purified water, at the end of the process, before being poured into the river must have neither DBO, nor DQO, nor SS nor SSV, absolute zeros are impossible and therefore the current legislation has permissible limits that are not dangerous for the life in rivers. Available variables:

- date id from 1 to number of observations
- dateformated dd-mm-yy
- datenorm dd/mm/yyyy
- q.e Input Flow
- qb.b Flow after biological reactor
- qr.g Recirculation Flow
- qp.g Purge Flow
- qa.g Air inflow
- fe.e Iron pretreatment
- ph.e Hydrogen potential
- ss.e Input Solid in Suspension
- ssv.e Input Suspended Volatile Solids
- dqo.e Input Fraction of degradable organic matter
- dbo.e Input BIOdegradable organic matter fraction
- nkt.e Input Hydrogen potential
- nh4.e Input Ammonium concentration
- p.e Input Phosphor concentration
- ph.d Decantation Hydrogen potential at the settler
- ss.d Decantation Solid in Suspension at the settler
- ssv.d Decantation Suspended Volatile Solids at the settler
- dqo.d Decantation Fraction of degradable organic matter at the settler
- dbo.d Decantation BIOdegradable organic matter fraction at the settler
- nkt.d Decantation Hydrogen potential at the settler
- nh4.d Decantation Ammonium concentration at the settler

- p.d Decantation Phosphor concentration at the settler
- ph.s Output Hydrogen potential
- ss.s Output Solid in Suspension
- ssv.s Output Suspended Volatile Solids
- dqo.s Output Fraction of degradable organic matter
- dbo.s Output BIOdegradable organic matter fraction
- nk.s Unknown
- nh4.s Output Ammonium concentration
- p.s Output Phosphor concentration
- v30.b Biological Volumetric Analysis
- mlss.b Biological Mixed Liquor Suspended Solids
- mlvss.b Biological Volatile solids in suspension liquor mixture
- im.b Unknown
- cm1.b Unknown
- cm2.b Unknown
- mcrt.b Biological Cell Age
- trh.c Unknown (non important)
- dbo.dqoe Input Quocient DBO.E into DQO.E
- dbo.dqod Quocient DBO.D into DQO.D at the settler
- dbo.dqos Output Quocient DBO.S into DQO.S
- weekday Day of the week
- season Year season

The data technically correspond to daily measurements and there is a temporal correlation that cannot be dealt with in this subject. You only have to work in this exercise with the data in randomized order. The response variables are considered the fraction of biodegradable organic matter DBO.S, degradable organic matter DQO.S or solids in suspension, either volatile (SSV.S) or not (SS.S) in the OUTPUT of the plant. The response variable DQO.S is initially considered.

Firstly, load dataset and check available variables.

```r
# Clear plots
if(!is.null(dev.list())) dev.off()
```

```
## null device
##           1
```

```r
# Clean workspace
rm(list=ls())
```

Load Required Packages for this deliverable

We load the necessary packages and set working directory

```r
# Load Required Packages
options(contrasts=c("contr.treatment","contr.treatment"))

requiredPackages <- c("missMDA","chemometrics","mvoutlier","effects","FactoMineR","car", "factoextra","
missingPackages <- requiredPackages[!(requiredPackages %in% installed.packages()[,"Package"])]

if(length(missingPackages)) install.packages(missingPackages)
lapply(requiredPackages, require, character.only = TRUE)
```

Some useful functions

```r
calcQ <- function(x) { # Function to calculate the different quartiles
  s.x <- summary(x)
  iqr<-s.x[5]-s.x[2]
  list(souti=s.x[2]-3*iqr, mouti=s.x[2]-1.5*iqr, min=s.x[1], q1=s.x[2], q2=s.x[3],
       q3=s.x[5], max=s.x[6], mouts=s.x[5]+1.5*iqr, souts=s.x[5]+3*iqr )
}

countNA <- function(x) { # Function to count the NA values
  mis_x <- NULL
  for (j in 1:ncol(x)) {mis_x[j] <- sum(is.na(x[,j])) }
  mis_x <- as.data.frame(mis_x)
  rownames(mis_x) <- names(x)
  mis_i <- rep(0,nrow(x))
  for (j in 1:ncol(x)) {mis_i <- mis_i + as.numeric(is.na(x[,j])) }
  list(mis_col=mis_x,mis_ind=mis_i)
}

countX <- function(x,X) { # Function to count a specific number of appearences
  n_x <- NULL
  for (j in 1:ncol(x)) {n_x[j] <- sum(x[,j]==X) }
  n_x <- as.data.frame(n_x)
  rownames(n_x) <- names(x)
  nx_i <- rep(0,nrow(x))
  for (j in 1:ncol(x)) {nx_i <- nx_i + as.numeric(x[,j]==X) }
  list(nx_col=n_x,nx_ind=nx_i)
}
```

```r
setwd("~/Documents/uni/FIB-ADEI/Data and Questions for Quiz 1-20201110")
load("~/Documents/uni/FIB-ADEI/Data and Questions for Quiz 1-20201110/WasteWater.RData")
summary(df)
```

```
##       date         dateformated         datenorm              q.e
##  Min.   :  1.00   Length:396         Length:396         Min.   :20500
##  1st Qu.: 99.75   Class :character   Class :character   1st Qu.:38832
##  Median :198.50   Mode  :character   Mode  :character   Median :42880
##  Mean   :198.50                                         Mean   :41809
##  3rd Qu.:297.25                                         3rd Qu.:45124
##  Max.   :396.00                                         Max.   :54089
##
##       qb.b           qr.g            qp.g              qa.g
##  Min.   :19883   Min.   :17933   Min.   :   0.0   Min.   : 96451
##  1st Qu.:38498   1st Qu.:40592   1st Qu.: 527.8   1st Qu.:193010
##  Median :39000   Median :41920   Median : 653.8   Median :227740
##  Mean   :38903   Mean   :40965   Mean   : 621.6   Mean   :231732
##  3rd Qu.:39000   3rd Qu.:42850   3rd Qu.: 689.0   3rd Qu.:273438
##  Max.   :52245   Max.   :49527   Max.   :1080.0   Max.   :367840
##
##       fe.e            ph.e            ss.e            ssv.e
##  Min.   : 0.00   Min.   :7.200   Min.   : 62.0   Min.   : 19.0
##  1st Qu.:41.67   1st Qu.:7.550   1st Qu.:154.0   1st Qu.:118.9
##  Median :48.25   Median :7.600   Median :187.5   Median :146.0
##  Mean   :45.39   Mean   :7.619   Mean   :209.7   Mean   :157.5
```

```
##  3rd Qu.:54.35   3rd Qu.:7.700   3rd Qu.:245.2   3rd Qu.:183.2
##  Max.   :89.80   Max.   :8.000   Max.   :655.0   Max.   :593.0
##
##      dqo.e           dbo.e           nkt.e           nh4.e
##  Min.   :  27.0   Min.   : 69.0   Min.   :17.90   Min.   :  8.10
##  1st Qu.: 335.8   1st Qu.:155.8   1st Qu.:35.18   1st Qu.: 24.98
##  Median : 422.5   Median :197.0   Median :41.39   Median : 27.56
##  Mean   : 442.4   Mean   :213.6   Mean   :41.95   Mean   : 41.14
##  3rd Qu.: 517.0   3rd Qu.:250.0   3rd Qu.:46.86   3rd Qu.: 31.80
##  Max.   :1579.0   Max.   :987.0   Max.   :82.00   Max.   :347.00
##
##       p.e             ph.d            ss.d            ssv.d
##  Min.   : 2.100   Min.   :7.100   Min.   : 40.00   Min.   : 13.00
##  1st Qu.: 5.825   1st Qu.:7.500   1st Qu.: 76.00   1st Qu.: 56.00
##  Median : 6.800   Median :7.600   Median : 88.00   Median : 64.00
##  Mean   : 9.902   Mean   :7.563   Mean   : 88.72   Mean   : 64.91
##  3rd Qu.:16.000   3rd Qu.:7.700   3rd Qu.: 98.12   3rd Qu.: 73.00
##  Max.   :24.000   Max.   :7.900   Max.   :192.00   Max.   :134.00
##
##      dqo.d           dbo.d           nkt.d           nh4.d
##  Min.   : 27.0   Min.   : 36.0   Min.   :15.10   Min.   : 8.90
##  1st Qu.:217.0   1st Qu.: 95.0   1st Qu.:30.80   1st Qu.:23.80
##  Median :252.5   Median :119.5   Median :35.15   Median :27.03
##  Mean   :249.6   Mean   :120.0   Mean   :36.42   Mean   :25.94
##  3rd Qu.:286.2   3rd Qu.:142.0   3rd Qu.:40.02   3rd Qu.:28.63
##  Max.   :538.0   Max.   :274.0   Max.   :74.00   Max.   :37.10
##
##       p.d             ph.s            ss.s             ssv.s
##  Min.   : 1.500   Min.   :7.000   Min.   :  2.800   Min.   :  1.60
##  1st Qu.: 3.450   1st Qu.:7.400   1st Qu.:  9.275   1st Qu.:  6.50
##  Median : 4.369   Median :7.500   Median : 12.800   Median :  9.20
##  Mean   : 6.071   Mean   :7.533   Mean   : 16.511   Mean   : 12.36
##  3rd Qu.: 9.961   3rd Qu.:7.700   3rd Qu.: 18.000   3rd Qu.: 14.00
##  Max.   :14.700   Max.   :8.000   Max.   :174.800   Max.   :134.80
##
##      dqo.s            dbo.s            nk.s            nh4.s
##  Min.   :  9.00   Min.   : 2.00   Min.   : 2.00   Min.   : 0.500
##  1st Qu.: 34.00   1st Qu.:12.00   1st Qu.: 9.00   1st Qu.: 5.293
##  Median : 45.50   Median :17.00   Median :19.52   Median :11.164
##  Mean   : 51.26   Mean   :18.81   Mean   :19.23   Mean   :12.170
##  3rd Qu.: 63.00   3rd Qu.:22.62   3rd Qu.:26.81   3rd Qu.:18.472
##  Max.   :163.00   Max.   :84.00   Max.   :67.00   Max.   :31.500
##
##       p.s            v30.b           mlss.b         mlvss.b          im.b
##  Min.   :0.600   Min.   : 77.0   Min.   : 754   Min.   : 185   Min.   : 58.50
##  1st Qu.:1.350   1st Qu.:170.0   1st Qu.:1538   1st Qu.:1180   1st Qu.: 94.65
##  Median :1.900   Median :210.0   Median :1760   Median :1343   Median :118.00
##  Mean   :2.746   Mean   :262.7   Mean   :1767   Mean   :1344   Mean   :155.90
##  3rd Qu.:4.757   3rd Qu.:320.0   3rd Qu.:1944   3rd Qu.:1494   3rd Qu.:199.93
##  Max.   :7.000   Max.   :770.0   Max.   :3294   Max.   :2100   Max.   :577.00
##
##      cm1.b            cm2.b            mcrt.b            trh.c
##  Min.   :0.0200   Min.   :0.0500   Min.   :  1.780   Min.   :2.160
##  1st Qu.:0.5000   1st Qu.:0.2300   1st Qu.:  8.658   1st Qu.:2.417
```

```
##  Median :0.6200   Median :0.3200   Median : 10.195   Median :4.845  
##  Mean   :0.6198   Mean   :0.3599   Mean   : 14.290   Mean   :4.196  
##  3rd Qu.:0.7400   3rd Qu.:0.4412   3rd Qu.: 12.582   3rd Qu.:4.910  
##  Max.   :1.4300   Max.   :3.8600   Max.   :341.990   Max.   :9.630  
##                                                                     
##     dbo.dqoe         dbo.dqod         dbo.dqos           weekday      season   
##  Min.   :0.1500   Min.   :0.1500   Min.   :0.0700   Sunday   :57   Autumn:101  
##  1st Qu.:0.4100   1st Qu.:0.4100   1st Qu.:0.2800   Thursday :56   Spring: 92  
##  Median :0.4600   Median :0.4717   Median :0.3817   Monday   :57   Summer:112  
##  Mean   :0.4788   Mean   :0.4832   Mean   :0.4068   Tuesday  :56   Winter: 91  
##  3rd Qu.:0.5400   3rd Qu.:0.5500   3rd Qu.:0.5112   Wednesday:56               
##  Max.   :1.0400   Max.   :1.0000   Max.   :1.0000   Saturday :57               
##                                                     Friday   :57
```

1. Produeix un conjunt de dades aleatori per destruir la correlació en sèrie.

```r
set.seed(10121)
sam<-as.vector(sample(1:nrow(df)))
```

Verificació i emmagatzematge de la mostra

```r
head(df)
```

```
##     date dateformated   datenorm     q.e    qb.b    qr.g  qp.g   qa.g fe.e ph.e
## 227  227      14-IV-96 14/04/1996 37144.6 36643.4 42864.9 699.7 209741 35.7  7.8
## 21    21      21-IX-95 21/09/1995 48650.0 39000.0 41766.0 385.9 263090  0.0  7.5
## 291  291      17-VI-96 17/06/1996 44319.4 43731.4 41484.2 840.6 291791 56.9  7.9
## 354  354    19-VIII-96 19/08/1996 36000.0 35416.1 43516.8 676.3 199481 55.2  7.6
## 241  241      28-IV-96 28/04/1996 41040.0 39000.0 34946.2 642.4 267831 15.7  7.6
## 228  228      15-IV-96 15/04/1996 43200.0 42682.0 42477.0 663.0 237011 34.8  7.7
##     ss.e ssv.e dqo.e dbo.e nkt.e    nh4.e    p.e ph.d ss.d ssv.d dqo.d dbo.d
## 227  151   119   588   140 44.64 38.0000  6.560  7.6   63    53   280    70
## 21   154   115   335   257 29.70 18.3000 15.800  7.5   90    64   270   160
## 291  296   228   800   290 56.18 36.6800  7.140  7.6   89    70   329   110
## 354  204   159   498   247 43.10 30.7000  7.800  7.5   87    65   267   118
## 241  410   319   618   294 29.75 17.1375  4.425  7.5   82    62   240   100
## 228  157   126   314   154 42.80 39.2000  6.500  7.6   80    62   305   151
##       nkt.d  nh4.d     p.d ph.s ss.s ssv.s dqo.s dbo.s   nk.s    nh4.s    p.s
## 227 43.4600 33.840  4.4600  7.6  6.8   4.0    30    13 28.460 20.0000 1.5000
## 21  30.8000 18.800 14.7000  7.5 33.2  24.4    74    35  8.400  7.5000 7.0000
## 291 50.2600 35.100  4.1000  7.6 17.4  14.0    65    10 28.660  2.9000 2.0800
## 354 38.0000 29.700  4.5000  7.8 15.6  12.1    46    29 19.600 17.6000 2.1000
## 241 28.5625 18.775  2.4375  7.3 16.4  12.4    65    23  6.275  4.1625 0.9625
## 228 43.1000 35.000  4.4000  7.7  8.0   6.6    43    14 26.800 17.4000 1.5000
##     v30.b mlss.b mlvss.b  im.b cm1.b cm2.b mcrt.b trh.c dbo.dqoe dbo.dqod
## 227   620   1706    1326 269.0  0.67  0.15   9.65  5.23     0.24     0.25
## 21    130   1877    1310  69.3  0.70  0.60   9.48  2.37     0.77     0.59
## 291   160   2296    1757  69.0  0.63  0.24   7.86  4.38     0.36     0.33
## 354   120   1556    1178  77.0  0.63  0.26   9.68  5.41     0.50     0.44
## 241   160   1574    1240  95.0  0.55  0.24   8.01  4.91     0.48     0.42
## 228   610   1872    1438 322.0  0.75  0.39  10.32  4.49     0.49     0.50
##     dbo.dqos  weekday season
## 227     0.43   Sunday Spring
## 21      0.47 Thursday Autumn
## 291     0.15   Monday Spring
## 354     0.63   Monday Summer
```

```
## 241      0.35   Sunday Spring
## 228      0.33   Monday Spring
```

```
df<-df[sam,]
```

2. S'han tractat les dades que falten, però algunes NA codificades com a valors 0 encara romanen a fe.e i qp.g i s'han de tractar aplicant eines d'imputació explicades a la classe.

fe.e

```
summary(df$fe.e)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    0.00   41.67   48.25   45.39   54.35   89.80
```

```
sel_fe.e <- which(df$fe.e == 0)
df[sel_fe.e, "fe.e"] <- NA
summary(df$fe.e)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##    6.50   42.90   48.70   47.43   54.80   89.80      17
```

qp.g

```
summary(df$qp.g)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##     0.0   527.8   653.8   621.6   689.0  1080.0
```

```
sel_qp.g <- which(df$qp.g == 0)
df[sel_qp.g, "qp.g"] <- NA
summary(df$qp.g)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   188.0   530.9   654.4   631.2   689.6  1080.0       6
```

imputació

```
library(missMDA)
vars_to_impute <- names(df)[c(7,9)]
res.imputation<-imputePCA(df[,vars_to_impute],ncp=1)
summary(res.imputation$completeObs)
```

```
##       qp.g             fe.e      
##  Min.   : 188.0   Min.   : 6.50  
##  1st Qu.: 531.5   1st Qu.:43.17  
##  Median : 653.8   Median :48.25  
##  Mean   : 630.7   Mean   :47.37  
##  3rd Qu.: 689.0   3rd Qu.:54.35  
##  Max.   :1080.0   Max.   :89.80
```

```
df[,vars_to_impute] <- res.imputation$completeObs
summary(df$fe.e)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    6.50   43.17   48.25   47.37   54.35   89.80
```

```
summary(df$qp.g)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   188.0   531.5   653.8   630.7   689.0  1080.0
```

3. Els outliers univariants per a la variable de sortida DQO.S també són presents i han de ser tractats. Fes-ho.

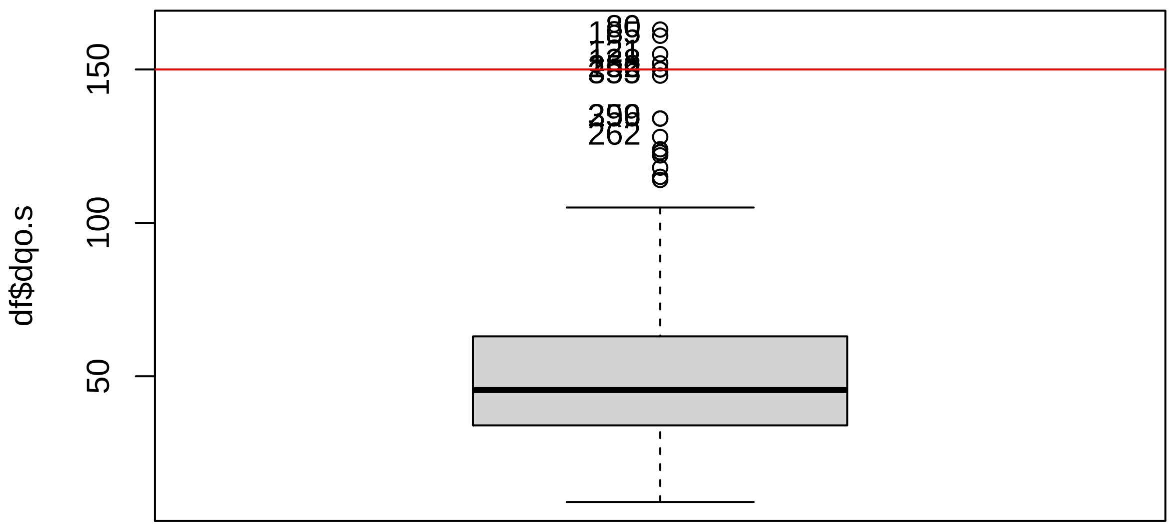
```
summary(df$dqo.s)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.00   34.00   45.50   51.26   63.00  163.00
```

```
mean<-mean(df$dqo.s)
Boxplot(df$dqo.s)
```

```
##  [1]  80 185 121 138 156 252 355 259 396 262
```

```
var_out<-calcQ(df$dqo.s)
abline(h=var_out$souts,col="red")
abline(h=var_out$souti,col="red")
```



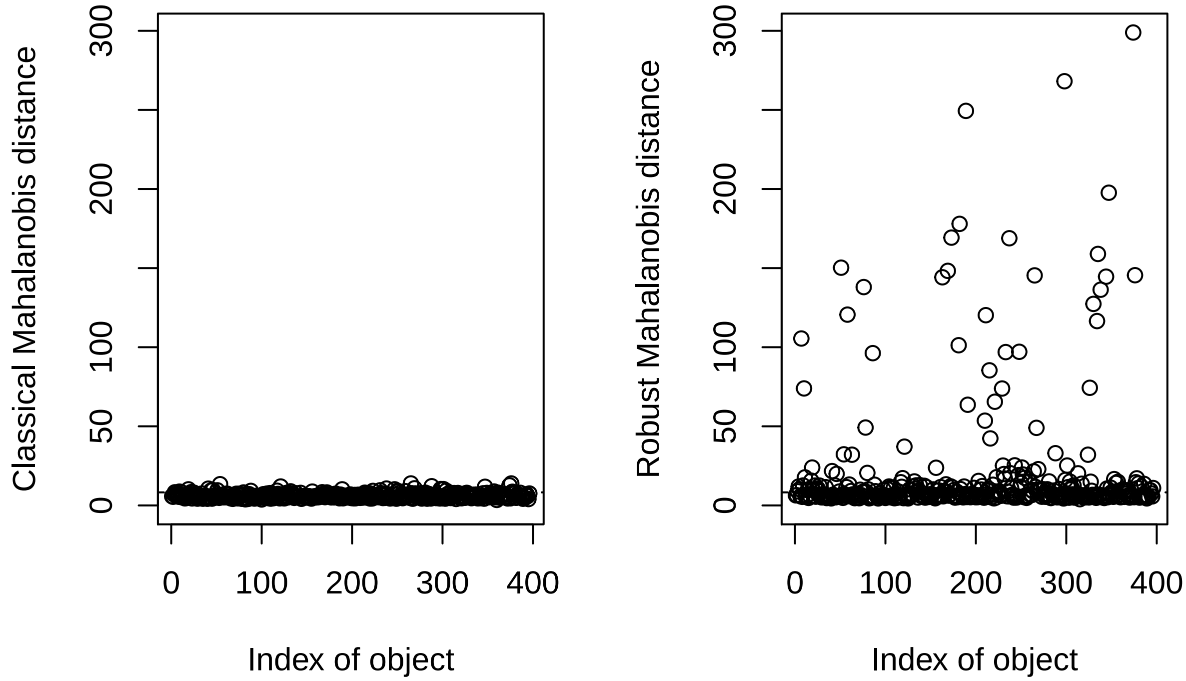

```
llout<-which((df$dqo.s>150))
df[llout,"dqo.s"]<-mean
```

4. Hi ha outliers multivariants? Troba'ls. Intenta explicar la seva singularitat. Els valors atípics multivariants no es tractaran en aquest exercici: mantingueu-los tal com són.

```
library(mvoutlier)
library(chemometrics)
names(df)
```

```
##  [1] "date"         "dateformated" "datenorm"     "q.e"          "qb.b"
##  [6] "qr.g"         "qp.g"         "qa.g"         "fe.e"         "ph.e"
## [11] "ss.e"         "ssv.e"        "dqo.e"        "dbo.e"        "nkt.e"
## [16] "nh4.e"        "p.e"          "ph.d"         "ss.d"         "ssv.d"
## [21] "dqo.d"        "dbo.d"        "nkt.d"        "nh4.d"        "p.d"
## [26] "ph.s"         "ss.s"         "ssv.s"        "dqo.s"        "dbo.s"
## [31] "nk.s"         "nh4.s"        "p.s"          "v30.b"        "mlss.b"
## [36] "mlvss.b"      "im.b"         "cm1.b"        "cm2.b"        "mcrt.b"
## [41] "trh.c"        "dbo.dqoe"     "dbo.dqod"     "dbo.dqos"     "weekday"
## [46] "season"
```

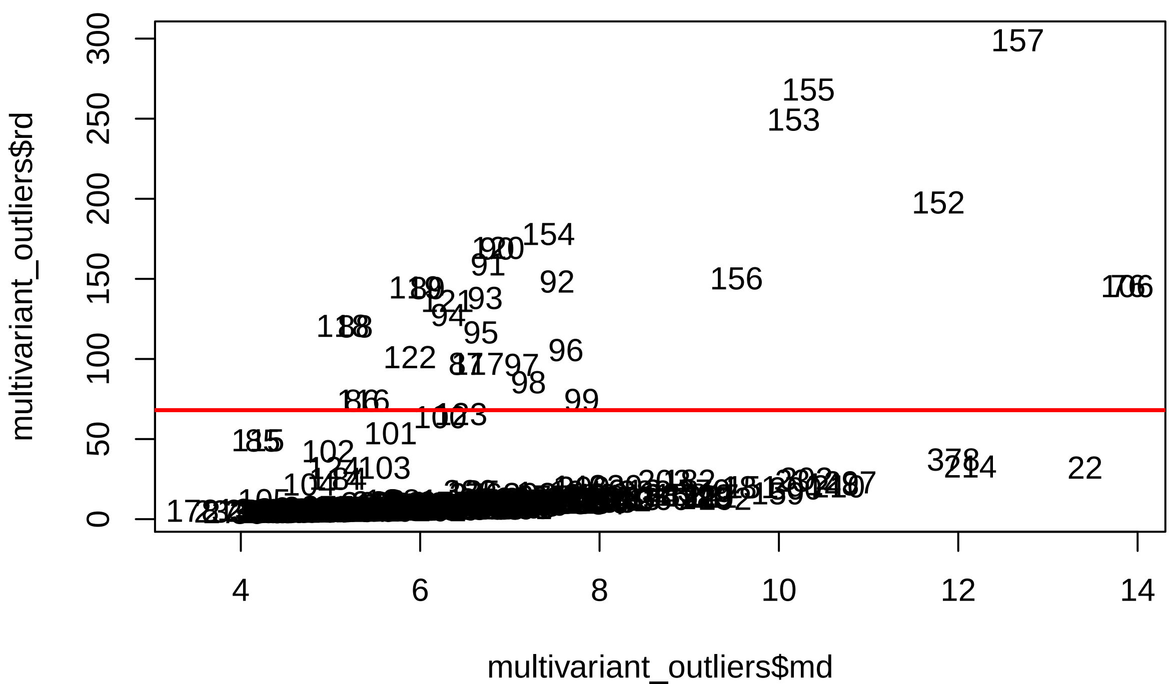
```
multivariant_outliers <- Moutlier(df[, c(4:44)], quantile = 0.995)
```




```r
multivariant_outliers$cutoff
```

```
## [1] 8.249408
```

```r
par(mfrow=c(1,1))
plot(multivariant_outliers$md, multivariant_outliers$rd, type="n")
text(multivariant_outliers$md, multivariant_outliers$rd, labels=rownames(df[, c(4:44)]))
abline(col="red",lwd=2, h=qchisq(0.995, ncol(df[, c(4:44)])))
```



Singularitat:

```r
df[which(row.names(df)=="157"), 1:46]
```

```
##     date dateformated   datenorm     q.e   qb.b    qr.g     qp.g   qa.g    fe.e
## 157  157      4-II-96 04/02/1996 23662.9 22891 18150.3 630.7379 143151 47.3702
##     ph.e ss.e ssv.e dqo.e dbo.e nkt.e nh4.e p.e ph.d ss.d ssv.d dqo.d dbo.d
## 157  7.6   77    51   180    73  17.9   8.1 2.1  7.7   40    30   152    54
##     nkt.d nh4.d p.d ph.s ss.s ssv.s dqo.s dbo.s nk.s nh4.s p.s v30.b mlss.b
## 157  15.1   8.9 1.9  7.6  5.2     4    66     5    2   0.5 1.1   310   2936
##     mlvss.b  im.b cm1.b cm2.b mcrt.b trh.c dbo.dqoe dbo.dqod dbo.dqos weekday
## 157    1878 105.6  0.08  0.05 341.99  8.37     0.41     0.36     0.08  Sunday
##     season
## 157 Winter
```

5. Indiqueu-ho mitjançant eines d'anàlisi de dades exploratòries que aparentment són les variables més associades a la variable de resposta (utilitzeu només les variables indicades). Utilitzeu també eines de perfilat FactoMineR.

```r
vars_res<-names(df)[c(29)]
vars_quantitatives<-names(df)[c(4:28, 30:44)]
vars_categorical<-names(df)[c(45:46)]

res.condes <- condes(df[, c(vars_res,vars_quantitatives, vars_categorical)],1)
res.condes$quanti
```

```
##          correlation      p.value
## dbo.s      0.4132442 9.150194e-18
## ssv.s      0.3789540 5.697798e-15
## ss.s       0.3665570 4.875618e-14
## dqo.d      0.2744302 2.844414e-08
## nk.s       0.2177364 1.232463e-05
## im.b       0.2049816 3.958146e-05
## ph.s       0.1960249 8.612797e-05
## ssv.e      0.1772202 3.947384e-04
## nh4.s      0.1737394 5.148898e-04
## nkt.d      0.1713036 6.182922e-04
## dqo.e      0.1683242 7.708725e-04
## qr.g       0.1663809 8.884228e-04
## ss.e       0.1489591 2.963754e-03
## qp.g       0.1212841 1.574322e-02
## dbo.e      0.1196286 1.723704e-02
## dbo.d      0.1196110 1.725356e-02
## nkt.e      0.1179772 1.884911e-02
## ssv.d      0.1155087 2.150337e-02
## ph.d       0.1148294 2.228833e-02
## v30.b      0.1129367 2.460796e-02
## mcrt.b    -0.1431737 4.305904e-03
## mlvss.b   -0.1628865 1.142315e-03
## mlss.b    -0.1885617 1.604054e-04
## dbo.dqos  -0.3187916 8.357317e-11
```

```r
res.condes$quali
```

```
##                R2     p.value
## season 0.03122189 0.005990612
```

```r
res.condes$category
```

```
##                 Estimate     p.value
## season=Spring  4.77248 0.040570839
## season=Winter -7.22961 0.001142338
```

6. Definiu els factors politòmics f.dbo.s, f.dqo.s, f.sst.s (de SSV.S més SS.S) per a les covariables segons el límit legal (DBO 25 mg/l O2, DQO 125 mg/l O2 i sòlids totals en suspensió 35 mg/l). Perfil factor f.dqo.s.

Factoritzar:

```r
# f.dbo.s DBO 25 mg/l O2
table(df$dbo.s)
```

```
## 
##                2                4                5 5.33333333333333 
##                2                3                8                1 
## 5.66666666666667                6 6.33333333333333              6.5 
##                1                7                1                1 
## 6.66666666666667                7              7.5                8 
##                1                8                1                5 
## 8.33333333333333                9              9.7               10 
##                1                4                1               23 
##             10.5               11             11.5               12 
##                2               14                2               18 
## 12.3333333333333             12.5               13 13.6666666666667 
##                2                3               20                1 
##             13.7               14             14.5 14.6666666666667 
##                3               15                5                2 
##            14.75               15 15.3333333333333             15.5 
##                1               13                1                2 
##               16             16.5 16.6666666666667               17 
##               22                1                1               13 
##             17.3             17.5 17.6666666666667               18 
##                2                2                1               17 
##             18.5               19             19.5               20 
##                4               14                1               13 
## 20.3333333333333               21 21.3333333333333               22 
##                1               16                1               13 
##             22.5               23             23.5               24 
##                3               11                2                6 
##            24.25               25 25.3333333333333             25.5 
##                1                7                1                2 
##               26 26.6666666666667               27               28 
##                2                1                3                8 
##               29               30               32               33 
##                7                5                5                1 
##               34 34.6666666666667               35               36 
##                5                1                8                1 
##               37               38               39               40 
##                1                3                1                4 
##               41               42               43               44 
##                1                1                1                1 
## 44.6666666666667               45               50               51 
```

```
##                  1                 2                 1                 1
## 60.3333333333333                76                79                84
##                  1                 1                 1                 1
```

```r
df$dbo.s[df$dbo.s <= 25] = 0
df$dbo.s[df$dbo.s > 25] = 1
df$f.dbo.s <- factor(df$dbo.s, labels =c("Legal","Ilegal"))

# f.dqo.s DQO 125 mg/l O2
table(df$dqo.s)
```

```
##
##                  9                10                17                18
##                  4                 4                 1                10
##                 19                20                24                26
##                  3                14                 1                10
##                 27              27.5                28                29
##                 24                 1                 8                 1
##               29.3                30                31                32
##                  1                10                 3                 2
##                 34                35              35.3                36
##                  3                14                 2                20
##                 37                38                39                40
##                 11                 5                 2                 3
##               40.5                41                42              42.5
##                  1                 2                 3                 1
##                 43                44                45                46
##                  3                11                20                11
##                 47                48              49.5                50
##                  7                 6                 1                 1
##                 51  51.2573232323232                52                53
##                  1                 4                 2                 9
##                 54              54.7                55                56
##                 15                 1                 8                 6
##                 57                58                59                60
##                  5                 4                 3                 2
##                 61                62                63                64
##                  1                10                10                 6
##                 65              65.5                66                67
##                  5                 2                 3                 2
##                 68                69                70                71
##                  1                 2                 4                 1
##                 72                74                75                76
##                  3                 3                 3                 5
##               77.5                78                79                80
##                  2                 4                 4                 2
##                 81              81.3                82                84
##                  2                 1                 2                 2
##                 85                87                89                91
##                  2                 2                 2                 1
##                 93                94                95                96
##                  2                 1                 2                 1
##                 98               100               103               105
##                  2                 2                 1                 1
##                114               115               118               122
```

```
##               1              1              1              1
##             123            124            128            134
##               2              1              1              2
##             148            150
##               1              2
```

```
df$dqo.s[df$dqo.s <= 125] = 0
df$dqo.s[df$dqo.s > 125] = 1
df$f.dqo.s <- factor(df$dqo.s, labels =c("Legal","Ilegal"))

# f.sst.s 35 mg/l
df$sst.s=rowSums(cbind(df$ssv.s,df$ss.s),na.rm=TRUE)
table(df$sst.s)
```

```
##
##             4.8              5            5.2            5.4
##               1              1              1              1
##               6            6.2            6.4            6.6
##               1              1              1              1
##             7.6              8            8.4            8.6
##               1              1              2              1
##             8.8 8.86666666666666            9.2            9.5
##               4              1              5              1
##             9.6            9.8             10           10.8
##               4              2              1              4
##            11.2 11.5333333333333           11.6             12
##               7              1              2              3
##            12.4           12.8             13           13.2
##               2              1              1              5
##            13.4           13.6           13.8 13.8666666666667
##               1              2              1              1
##              14           14.2           14.3           14.4
##               6              1              1              5
##            14.6 14.7333333333333           14.8             15
##               1              1              9              1
##            15.2           15.4           15.6          15.65
##               4              1              4              1
##            15.7             16           16.2           16.4
##               1              5              2              2
##            16.5           16.8           16.9             17
##               1              6              1              4
##            17.2           17.4           17.6           17.7
##               1              2              6              1
##            17.8 17.8666666666666           17.9             18
##               2              1              1              5
##            18.4           18.6           18.8 18.9333333333334
##               3              1             10              1
##              19           19.2           19.3           19.4
##               1             10              2              1
##            19.6           19.8             20           20.4
##               3              3              7              6
##            20.6             21           21.2           21.5
##               3              2              6              1
##           21.55           21.6             22           22.2
##               1              4              3              1
```

```
##             22.4             22.8             22.9             23.2 
##                3                1                1                1 
##             23.4 23.4666666666667             23.6             23.7 
##                1                1                5                2 
##             23.8               24             24.4             24.6 
##                2                3                3                2 
##             24.8             24.9               25             25.2 
##                5                1                1                4 
##             25.6               26             26.4             26.8 
##                2                4                3                1 
##             27.2             27.6             27.7             27.8 
##                3                5                1                1 
##               28 28.1333333333333             28.4             28.6 
##                2                1                5                1 
##             28.8               29             29.2             29.5 
##                2                2                2                1 
##             29.6             30.4             30.6 30.6666666666666 
##                1                2                1                1 
##             30.8               31             31.1             31.2 
##                3                1                1                3 
##             31.4             31.6             31.7               32 
##                1                3                1                2 
## 32.3333333333334             32.4             32.6             32.8 
##                1                2                1                2 
##               33             33.2             33.6               34 
##                1                1                1                7 
##             34.2             34.4             34.6             34.8 
##                1                5                1                1 
##             35.4             35.6               36             36.2 
##                1                3                5                1 
##               37             37.2             37.6               38 
##                1                4                3                3 
##             38.4             39.6             39.8             40.7 
##                1                1                1                1 
##             41.2               42             42.4               44 
##                1                2                2                1 
##             44.8               46             46.4             47.2 
##                2                3                1                2 
##               48             48.4             49.2               50 
##                2                1                1                3 
##               56             57.6               62               67 
##                2                1                2                1 
##             72.4             74.4             74.8             77.2 
##                1                2                1                1 
##               79             82.8              101            105.5 
##                1                1                1                1 
##              107            107.2              118            126.4 
##                1                1                1                1 
##              129              132              142            154.4 
##                1                1                1                1 
##            162.8            196.4            233.2            309.6 
##                1                1                1                1 
```

```
df$sst.s[df$sst.s <= 35] = 0
df$sst.s[df$sst.s > 35] = 1
df$f.sst.s <- factor(df$sst.s, labels =c("Legal","Ilegal"))
```

Profiling de f.dqo.s:

```
vars_res<-names(df)[c(48)]
vars_quantitatives<-names(df)[c(4:44)]
vars_categorical<-names(df)[c(45:47, 50)]

res.catdes <- catdes(df[, c(vars_res,vars_quantitatives, vars_categorical)],1)
res.catdes$test.chi2
```

```
##              p.value df
## f.sst.s 3.610966e-05  1
## f.dbo.s 1.917010e-03  1
```

```
res.catdes$quanti.var
```

```
##              Eta2      P-value
## dqo.s  1.00000000 0.000000e+00
## ssv.s  0.16606401 2.828654e-17
## ss.s   0.16279494 6.173060e-17
## im.b   0.02463311 1.731371e-03
## dbo.s  0.02431149 1.857973e-03
## mlss.b 0.01089606 3.786359e-02
```

```
res.catdes$category
```

```
## $Legal
##                   Cla/Mod Mod/Cla   Global     p.value    v.test
## f.sst.s=Legal  99.69040 82.5641 81.56566 0.001006991  3.288566
## f.dbo.s=Legal  99.38272 82.5641 81.81818 0.012259831  2.504576
## f.dbo.s=Ilegal 94.44444 17.4359 18.18182 0.012259831 -2.504576
## f.sst.s=Ilegal 93.15068 17.4359 18.43434 0.001006991 -3.288566
##
## $Ilegal
##                  Cla/Mod  Mod/Cla   Global     p.value    v.test
## f.sst.s=Ilegal 6.8493151 83.33333 18.43434 0.001006991  3.288566
## f.dbo.s=Ilegal 5.5555556 66.66667 18.18182 0.012259831  2.504576
## f.dbo.s=Legal  0.6172840 33.33333 81.81818 0.012259831 -2.504576
## f.sst.s=Legal  0.3095975 16.66667 81.56566 0.001006991 -3.288566
```

7. Una anàlisi normalitzada de components principals s'aborda utilitzant com a variables addicionals disponibles el factor disponible i les variables de sortida de xxxx.s.

```
names(df)
```

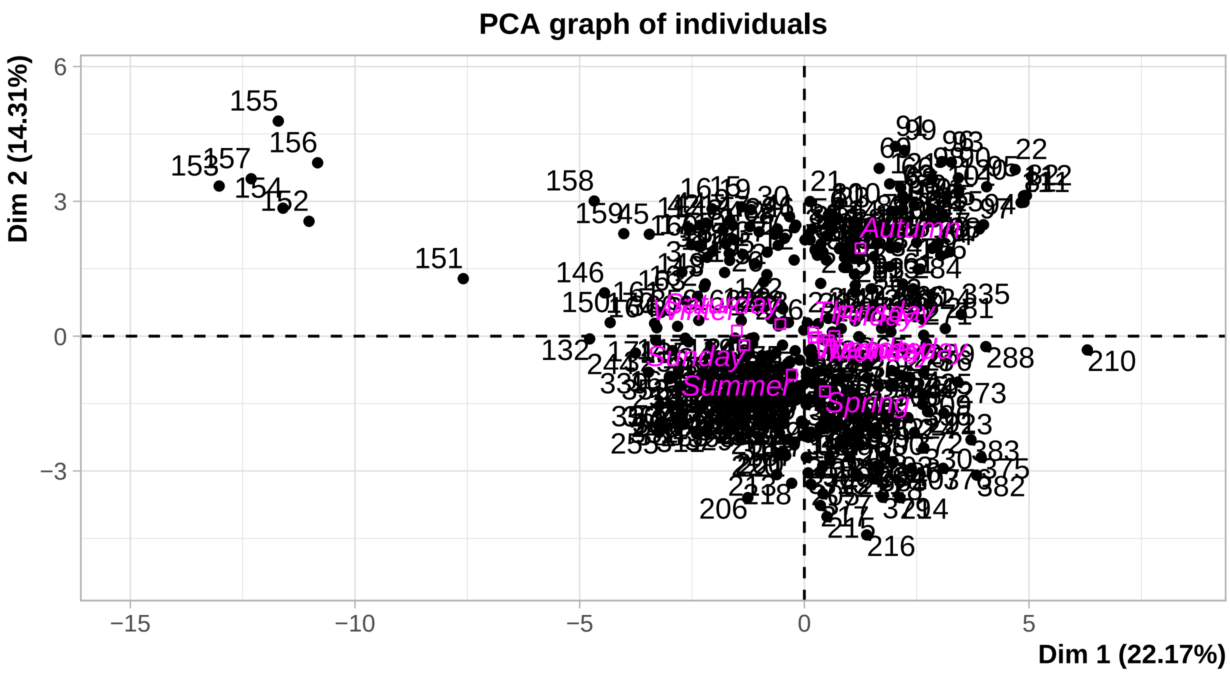
```
##  [1] "date"          "dateformated" "datenorm"      "q.e"           "qb.b"
##  [6] "qr.g"          "qp.g"         "qa.g"          "fe.e"          "ph.e"
## [11] "ss.e"          "ssv.e"        "dqo.e"         "dbo.e"         "nkt.e"
## [16] "nh4.e"         "p.e"          "ph.d"          "ss.d"          "ssv.d"
## [21] "dqo.d"         "dbo.d"        "nkt.d"         "nh4.d"         "p.d"
## [26] "ph.s"          "ss.s"         "ssv.s"         "dqo.s"         "dbo.s"
## [31] "nk.s"          "nh4.s"        "p.s"           "v30.b"         "mlss.b"
## [36] "mlvss.b"       "im.b"         "cm1.b"         "cm2.b"         "mcrt.b"
## [41] "trh.c"         "dbo.dqoe"     "dbo.dqod"      "dbo.dqos"      "weekday"
```

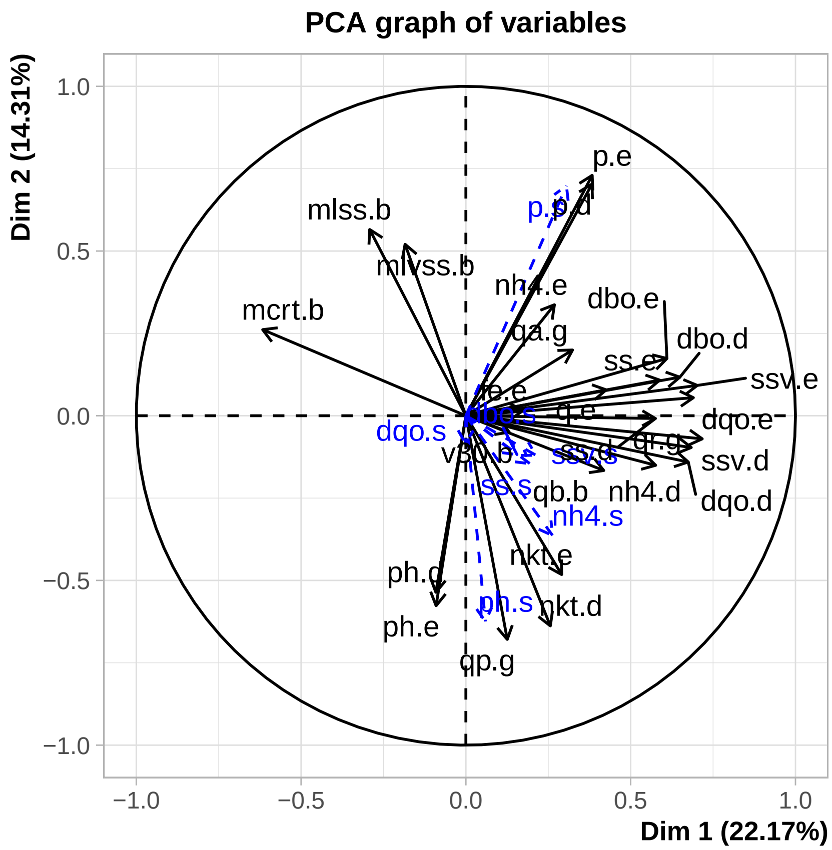
```
## [46] "season"       "f.dbo.s"      "f.dqo.s"      "sst.s"        "f.sst.s"
```

```r
names(df[, c(4:30,32:36,40,45:46)])
```

```
##  [1] "q.e"     "qb.b"    "qr.g"    "qp.g"    "qa.g"    "fe.e"    "ph.e"
##  [8] "ss.e"    "ssv.e"   "dqo.e"   "dbo.e"   "nkt.e"   "nh4.e"   "p.e"
## [15] "ph.d"    "ss.d"    "ssv.d"   "dqo.d"   "dbo.d"   "nkt.d"   "nh4.d"
## [22] "p.d"     "ph.s"    "ss.s"    "ssv.s"   "dqo.s"   "dbo.s"   "nh4.s"
## [29] "p.s"     "v30.b"   "mlss.b"  "mlvss.b" "mcrt.b"  "weekday" "season"
```

```r
res.pca <- PCA(df[, c(4:30,32:36,40,45:46)], quanti.sup=c(23:29), quali.sup=c(34:35))
```





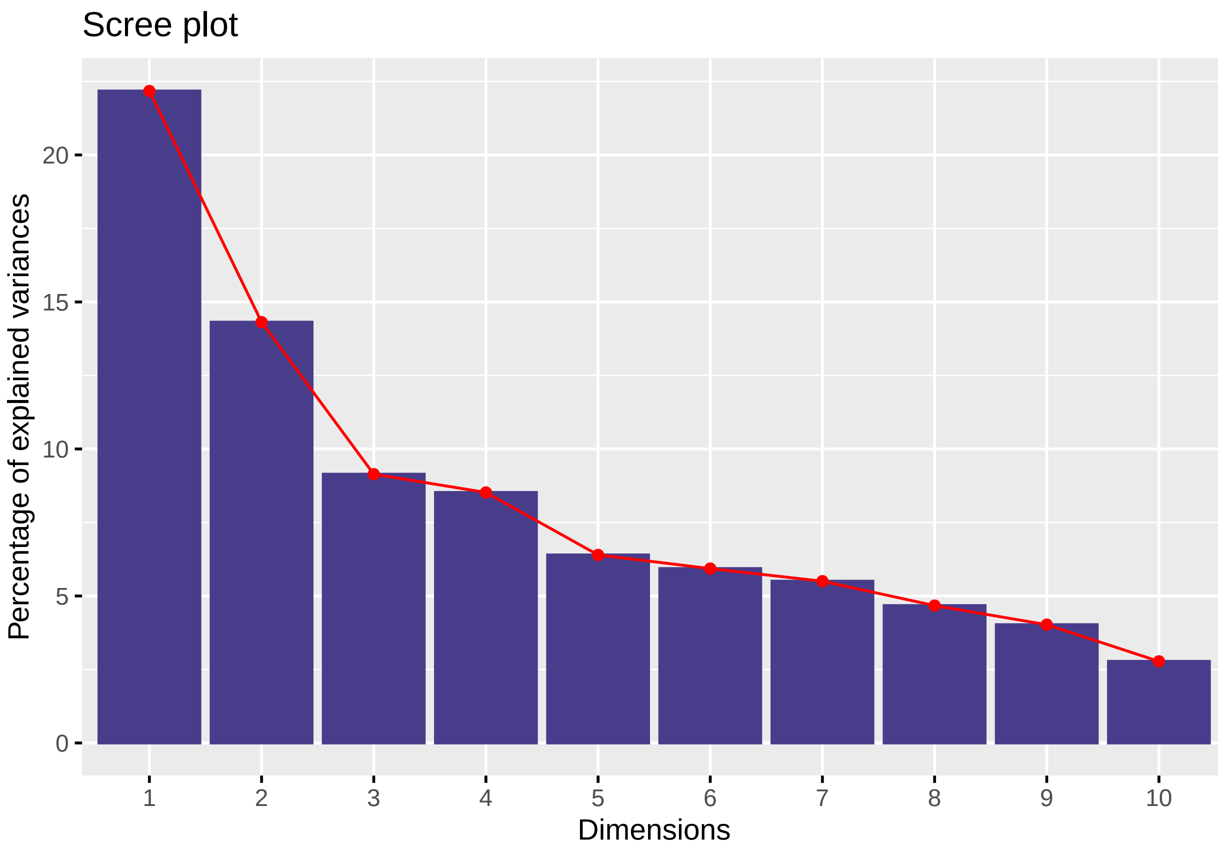

```r
summary(res.pca,nbind=0,nbelements = 25)
```

```
##
## Call:
## PCA(X = df[, c(4:30, 32:36, 40, 45:46)], quanti.sup = c(23:29),
##      quali.sup = c(34:35))
##
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6   Dim.7
## Variance               5.765   3.721   2.377   2.215   1.663   1.542   1.431
## % of var.             22.175  14.312   9.142   8.520   6.395   5.932   5.503
## Cumulative % of var.  22.175  36.487  45.629  54.149  60.545  66.476  71.979
##                        Dim.8   Dim.9  Dim.10  Dim.11  Dim.12  Dim.13  Dim.14
## Variance               1.215   1.046   0.722   0.654   0.597   0.499   0.453
## % of var.              4.673   4.024   2.776   2.515   2.295   1.919   1.743
## Cumulative % of var.  76.652  80.675  83.451  85.967  88.262  90.181  91.924
##                       Dim.15  Dim.16  Dim.17  Dim.18  Dim.19  Dim.20  Dim.21
## Variance               0.414   0.348   0.324   0.214   0.196   0.175   0.131
## % of var.              1.592   1.338   1.244   0.825   0.754   0.674   0.504
## Cumulative % of var.  93.516  94.854  96.099  96.924  97.678  98.352  98.856
##                       Dim.22  Dim.23  Dim.24  Dim.25  Dim.26
## Variance               0.105   0.078   0.048   0.041   0.025
## % of var.              0.402   0.302   0.185   0.158   0.097
## Cumulative % of var.  99.258  99.560  99.745  99.903 100.000
##
## Variables (the 25 first)
```

```
##              Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr   cos2
## q.e       |  0.426  3.144  0.181 |  0.079  0.166  0.006 |  0.134  0.750  0.018
## qb.b      |  0.418  3.036  0.175 | -0.166  0.743  0.028 |  0.108  0.495  0.012
## qr.g      |  0.684  8.114  0.468 | -0.096  0.250  0.009 | -0.403  6.839  0.163
## qp.g      |  0.126  0.274  0.016 | -0.679 12.391  0.461 |  0.108  0.487  0.012
## qa.g      |  0.324  1.817  0.105 |  0.200  1.070  0.040 |  0.220  2.028  0.048
## fe.e      |  0.175  0.532  0.031 |  0.017  0.008  0.000 | -0.214  1.919  0.046
## ph.e      | -0.090  0.141  0.008 | -0.577  8.943  0.333 |  0.201  1.694  0.040
## ss.e      |  0.588  5.996  0.346 |  0.108  0.314  0.012 |  0.287  3.477  0.083
## ssv.e     |  0.702  8.554  0.493 |  0.092  0.228  0.008 |  0.264  2.923  0.069
## dqo.e     |  0.690  8.267  0.477 |  0.055  0.081  0.003 |  0.328  4.524  0.108
## dbo.e     |  0.610  6.453  0.372 |  0.174  0.812  0.030 |  0.278  3.254  0.077
## nkt.e     |  0.291  1.466  0.084 | -0.482  6.248  0.232 | -0.442  8.222  0.195
## nh4.e     |  0.269  1.253  0.072 |  0.337  3.053  0.114 | -0.499 10.455  0.249
## p.e       |  0.383  2.548  0.147 |  0.730 14.314  0.533 | -0.407  6.954  0.165
## ph.d      | -0.091  0.144  0.008 | -0.536  7.724  0.287 |  0.266  2.966  0.071
## ss.d      |  0.575  5.736  0.331 | -0.008  0.002  0.000 |  0.322  4.349  0.103
## ssv.d     |  0.717  8.923  0.514 | -0.070  0.132  0.005 |  0.286  3.437  0.082
## dqo.d     |  0.675  7.897  0.455 | -0.140  0.530  0.020 |  0.245  2.529  0.060
## dbo.d     |  0.649  7.311  0.422 |  0.117  0.368  0.014 |  0.131  0.722  0.017
## nkt.d     |  0.256  1.140  0.066 | -0.638 10.939  0.407 | -0.192  1.552  0.037
## nh4.d     |  0.576  5.753  0.332 | -0.150  0.609  0.023 | -0.396  6.610  0.157
## p.d       |  0.383  2.542  0.147 |  0.704 13.301  0.495 | -0.379  6.032  0.143
## v30.b     |  0.132  0.301  0.017 | -0.050  0.067  0.002 | -0.163  1.112  0.026
## mlss.b    | -0.292  1.477  0.085 |  0.565  8.590  0.320 |  0.413  7.186  0.171
## mlvss.b   | -0.185  0.590  0.034 |  0.521  7.282  0.271 |  0.390  6.402  0.152
##
## q.e       |
## qb.b      |
## qr.g      |
## qp.g      |
## qa.g      |
## fe.e      |
## ph.e      |
## ss.e      |
## ssv.e     |
## dqo.e     |
## dbo.e     |
## nkt.e     |
## nh4.e     |
## p.e       |
## ph.d      |
## ss.d      |
## ssv.d     |
## dqo.d     |
## dbo.d     |
## nkt.d     |
## nh4.d     |
## p.d       |
## v30.b     |
## mlss.b    |
## mlvss.b   |
##
## Supplementary continuous variables
```

```
##              Dim.1   cos2    Dim.2   cos2    Dim.3   cos2
## ph.s       |  0.059  0.004 | -0.624  0.390 |  0.041  0.002 |
## ss.s       |  0.194  0.038 | -0.145  0.021 | -0.093  0.009 |
## ssv.s      |  0.213  0.045 | -0.120  0.014 | -0.098  0.010 |
## dqo.s      | -0.001  0.000 | -0.084  0.007 | -0.027  0.001 |
## dbo.s      |  0.155  0.024 | -0.118  0.014 |  0.013  0.000 |
## nh4.s      |  0.263  0.069 | -0.363  0.132 | -0.307  0.094 |
## p.s        |  0.305  0.093 |  0.698  0.488 | -0.356  0.127 |
##
## Supplementary categories
##               Dist     Dim.1    cos2 v.test    Dim.2    cos2 v.test    Dim.3
## Sunday     |  1.974 | -1.333  0.456 -4.524 | -0.199  0.010 -0.840 | -0.928
## Thursday   |  0.457 |  0.196  0.185  0.660 |  0.087  0.036  0.362 |  0.082
## Monday     |  0.674 |  0.540  0.643  1.833 | -0.095  0.020 -0.400 |  0.219
## Tuesday    |  0.452 |  0.266  0.347  0.894 | -0.022  0.002 -0.094 |  0.161
## Wednesday  |  0.577 |  0.209  0.131  0.702 | -0.038  0.004 -0.159 |  0.260
## Saturday   |  0.851 | -0.535  0.396 -1.817 |  0.266  0.098  1.125 | -0.081
## Friday     |  1.012 |  0.668  0.436  2.269 |  0.002  0.000  0.007 |  0.296
## Autumn     |  2.471 |  1.249  0.255  6.049 |  1.955  0.626 11.787 | -0.338
## Spring     |  1.919 |  0.455  0.056  2.073 | -1.227  0.409 -6.956 |  0.794
## Summer     |  1.694 | -0.279  0.027 -1.453 | -0.856  0.255 -5.537 | -0.592
## Winter     |  1.898 | -1.503  0.626 -6.793 |  0.124  0.004  0.698 |  0.301
##             cos2 v.test
## Sunday     0.221 -4.903 |
## Thursday   0.032  0.428 |
## Monday     0.105  1.155 |
## Tuesday    0.127  0.840 |
## Wednesday  0.203  1.360 |
## Saturday   0.009 -0.426 |
## Friday     0.086  1.565 |
## Autumn     0.019 -2.546 |
## Spring     0.171  5.628 |
## Summer     0.122 -4.795 |
## Winter     0.025  2.122 |
```

```
fviz_screeplot(
  res.pca,
  barfill = "darkslateblue",
  barcolor = "darkslateblue",
  linecolor = "red",
  ggtheme = theme_gray())
```



Quants eixos heu de conservar segons els criteris de Kaiser? 9 dimensions. I segons la regla de Elbow? 5 dimensions. Quina és la inèrcia que expliquen els components principals conservats basats en Kaiser? 80.675%

7. Intenta explicar el significat dels eixos en el primer pla factorial. Quines són les 3 variables amb més correlació amb el primer pla factorial?

```
summary(res.pca,nb.dec=2,nbind=0,nbelements = 25,ncp=2)
```

```
##
## Call:
## PCA(X = df[, c(4:30, 32:36, 40, 45:46)], quanti.sup = c(23:29),
##      quali.sup = c(34:35))
##
##
## Eigenvalues
##                        Dim.1  Dim.2  Dim.3  Dim.4  Dim.5  Dim.6  Dim.7  Dim.8
## Variance                5.77   3.72   2.38   2.22   1.66   1.54   1.43   1.21
## % of var.              22.17  14.31   9.14   8.52   6.40   5.93   5.50   4.67
## Cumulative % of var.   22.17  36.49  45.63  54.15  60.54  66.48  71.98  76.65
##                        Dim.9 Dim.10 Dim.11 Dim.12 Dim.13 Dim.14 Dim.15 Dim.16
## Variance                1.05   0.72   0.65   0.60   0.50   0.45   0.41   0.35
## % of var.               4.02   2.78   2.52   2.29   1.92   1.74   1.59   1.34
## Cumulative % of var.   80.68  83.45  85.97  88.26  90.18  91.92  93.52  94.85
##                       Dim.17 Dim.18 Dim.19 Dim.20 Dim.21 Dim.22 Dim.23 Dim.24
## Variance                0.32   0.21   0.20   0.18   0.13   0.10   0.08   0.05
## % of var.               1.24   0.82   0.75   0.67   0.50   0.40   0.30   0.19
## Cumulative % of var.   96.10  96.92  97.68  98.35  98.86  99.26  99.56  99.75
##                       Dim.25 Dim.26
```

```
## Variance                0.04   0.03
## % of var.               0.16   0.10
## Cumulative % of var.  99.90 100.00
##
## Variables (the 25 first)
##             Dim.1   ctr  cos2   Dim.2   ctr  cos2
## q.e       |  0.43  3.14  0.18 |  0.08  0.17  0.01 |
## qb.b      |  0.42  3.04  0.18 | -0.17  0.74  0.03 |
## qr.g      |  0.68  8.11  0.47 | -0.10  0.25  0.01 |
## qp.g      |  0.13  0.27  0.02 | -0.68 12.39  0.46 |
## qa.g      |  0.32  1.82  0.10 |  0.20  1.07  0.04 |
## fe.e      |  0.18  0.53  0.03 |  0.02  0.01  0.00 |
## ph.e      | -0.09  0.14  0.01 | -0.58  8.94  0.33 |
## ss.e      |  0.59  6.00  0.35 |  0.11  0.31  0.01 |
## ssv.e     |  0.70  8.55  0.49 |  0.09  0.23  0.01 |
## dqo.e     |  0.69  8.27  0.48 |  0.05  0.08  0.00 |
## dbo.e     |  0.61  6.45  0.37 |  0.17  0.81  0.03 |
## nkt.e     |  0.29  1.47  0.08 | -0.48  6.25  0.23 |
## nh4.e     |  0.27  1.25  0.07 |  0.34  3.05  0.11 |
## p.e       |  0.38  2.55  0.15 |  0.73 14.31  0.53 |
## ph.d      | -0.09  0.14  0.01 | -0.54  7.72  0.29 |
## ss.d      |  0.58  5.74  0.33 | -0.01  0.00  0.00 |
## ssv.d     |  0.72  8.92  0.51 | -0.07  0.13  0.00 |
## dqo.d     |  0.67  7.90  0.46 | -0.14  0.53  0.02 |
## dbo.d     |  0.65  7.31  0.42 |  0.12  0.37  0.01 |
## nkt.d     |  0.26  1.14  0.07 | -0.64 10.94  0.41 |
## nh4.d     |  0.58  5.75  0.33 | -0.15  0.61  0.02 |
## p.d       |  0.38  2.54  0.15 |  0.70 13.30  0.49 |
## v30.b     |  0.13  0.30  0.02 | -0.05  0.07  0.00 |
## mlss.b    | -0.29  1.48  0.09 |  0.57  8.59  0.32 |
## mlvss.b   | -0.18  0.59  0.03 |  0.52  7.28  0.27 |
##
## Supplementary continuous variables
##             Dim.1  cos2   Dim.2  cos2
## ph.s      |  0.06  0.00 | -0.62  0.39 |
## ss.s      |  0.19  0.04 | -0.15  0.02 |
## ssv.s     |  0.21  0.05 | -0.12  0.01 |
## dqo.s     |  0.00  0.00 | -0.08  0.01 |
## dbo.s     |  0.16  0.02 | -0.12  0.01 |
## nh4.s     |  0.26  0.07 | -0.36  0.13 |
## p.s       |  0.31  0.09 |  0.70  0.49 |
##
## Supplementary categories
##              Dist   Dim.1  cos2 v.test   Dim.2  cos2 v.test
## Sunday    |  1.97 | -1.33  0.46  -4.52 | -0.20  0.01  -0.84 |
## Thursday  |  0.46 |  0.20  0.18   0.66 |  0.09  0.04   0.36 |
## Monday    |  0.67 |  0.54  0.64   1.83 | -0.09  0.02  -0.40 |
## Tuesday   |  0.45 |  0.27  0.35   0.89 | -0.02  0.00  -0.09 |
## Wednesday |  0.58 |  0.21  0.13   0.70 | -0.04  0.00  -0.16 |
## Saturday  |  0.85 | -0.54  0.40  -1.82 |  0.27  0.10   1.12 |
## Friday    |  1.01 |  0.67  0.44   2.27 |  0.00  0.00   0.01 |
## Autumn    |  2.47 |  1.25  0.26   6.05 |  1.96  0.63  11.79 |
## Spring    |  1.92 |  0.46  0.06   2.07 | -1.23  0.41  -6.96 |
## Summer    |  1.69 | -0.28  0.03  -1.45 | -0.86  0.26  -5.54 |
```

```
## Winter    |  1.90 | -1.50  0.63  -6.79 |  0.12  0.00   0.70 |
```

```r
ddd<-dimdesc(res.pca,axes=1:2)
ddd$Dim.1
```

```
## $quanti
##         correlation      p.value
## ssv.d     0.7172345 8.702867e-64
## ssv.e     0.7022531 4.130053e-60
## dqo.e     0.6903675 2.358761e-57
## qr.g      0.6839592 6.373242e-56
## dqo.d     0.6747353 6.332099e-54
## dbo.d     0.6492347 9.182255e-49
## dbo.e     0.6099533 1.020883e-41
## ss.e      0.5879781 3.452770e-38
## nh4.d     0.5759342 2.294667e-36
## ss.d      0.5750476 3.103914e-36
## q.e       0.4257444 7.242552e-19
## qb.b      0.4183785 3.269946e-18
## p.e       0.3832906 2.630581e-15
## p.d       0.3828651 2.839262e-15
## qa.g      0.3236407 4.155347e-11
## p.s       0.3051301 5.599140e-10
## nkt.e     0.2906850 3.769062e-09
## nh4.e     0.2687437 5.595631e-08
## nh4.s     0.2628777 1.106464e-07
## nkt.d     0.2563545 2.316727e-07
## ssv.s     0.2127715 1.957327e-05
## ss.s      0.1943145 9.952683e-05
## fe.e      0.1750958 4.645130e-04
## dbo.s     0.1550750 1.968511e-03
## v30.b     0.1316599 8.712224e-03
## qp.g      0.1255977 1.237126e-02
## mlvss.b  -0.1845031 2.227597e-04
## mlss.b   -0.2918414 3.248056e-09
## mcrt.b   -0.6165374 7.888993e-43
##
## $quali
##                 R2      p.value
## season  0.17117557 6.918856e-16
## weekday 0.07369764 4.050303e-05
##
## $category
##                  Estimate      p.value
## season=Autumn   1.2684654 6.204938e-10
## weekday=Friday  0.6667511 2.309695e-02
## season=Spring   0.4746323 3.804156e-02
## weekday=Sunday -1.3345613 4.751263e-06
## season=Winter  -1.4830732 2.711014e-12
##
## attr(,"class")
## [1] "condes" "list "
```

```r
ddd$Dim.2
```

```
## $quanti
```

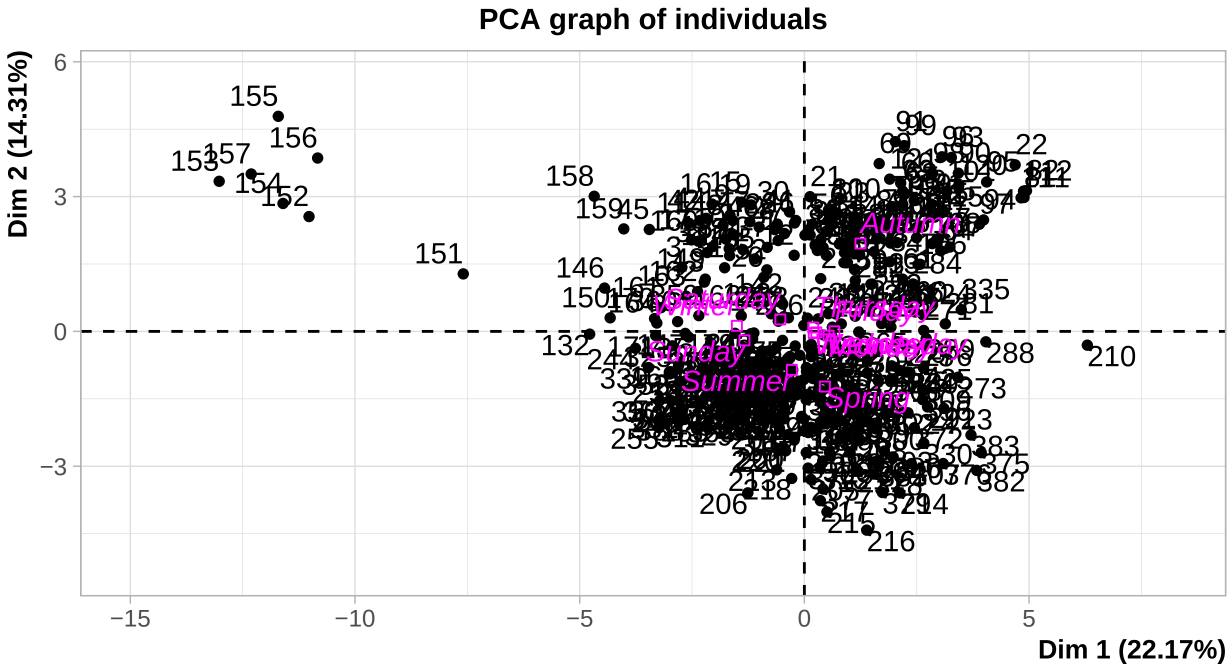
```
##           correlation      p.value
## p.e         0.7298391 4.533983e-67
## p.d         0.7035411 2.036918e-60
## p.s         0.6983003 3.530509e-59
## mlss.b      0.5653911 7.859456e-35
## mlvss.b     0.5205493 7.001826e-29
## nh4.e       0.3370715 5.615938e-12
## mcrt.b      0.2613343 1.320249e-07
## qa.g        0.1995670 6.358950e-05
## dbo.e       0.1738470 5.107139e-04
## dbo.d       0.1170809 1.977781e-02
## ss.e        0.1081635 3.140211e-02
## dbo.s      -0.1179087 1.891867e-02
## ssv.s      -0.1199365 1.695015e-02
## dqo.d      -0.1404357 5.115187e-03
## ss.s       -0.1454264 3.728858e-03
## nh4.d      -0.1504904 2.678851e-03
## qb.b       -0.1662918 8.941924e-04
## nh4.s      -0.3632852 8.461935e-14
## nkt.e      -0.4821699 1.900945e-24
## ph.d       -0.5361076 7.607530e-31
## ph.e       -0.5768834 1.658854e-36
## ph.s       -0.6243612 3.475558e-44
## nkt.d      -0.6380109 1.204814e-46
## qp.g       -0.6790493 7.526410e-55
##
## $quali
##                R2      p.value
## season 0.4126682 5.12791e-45
##
## $category
##                   Estimate      p.value
## season=Autumn  1.9561567 5.582386e-39
## season=Summer -0.8548954 1.710921e-08
## season=Spring -1.2263131 7.458427e-13
##
## attr(,"class")
## [1] "condes" "list "
```

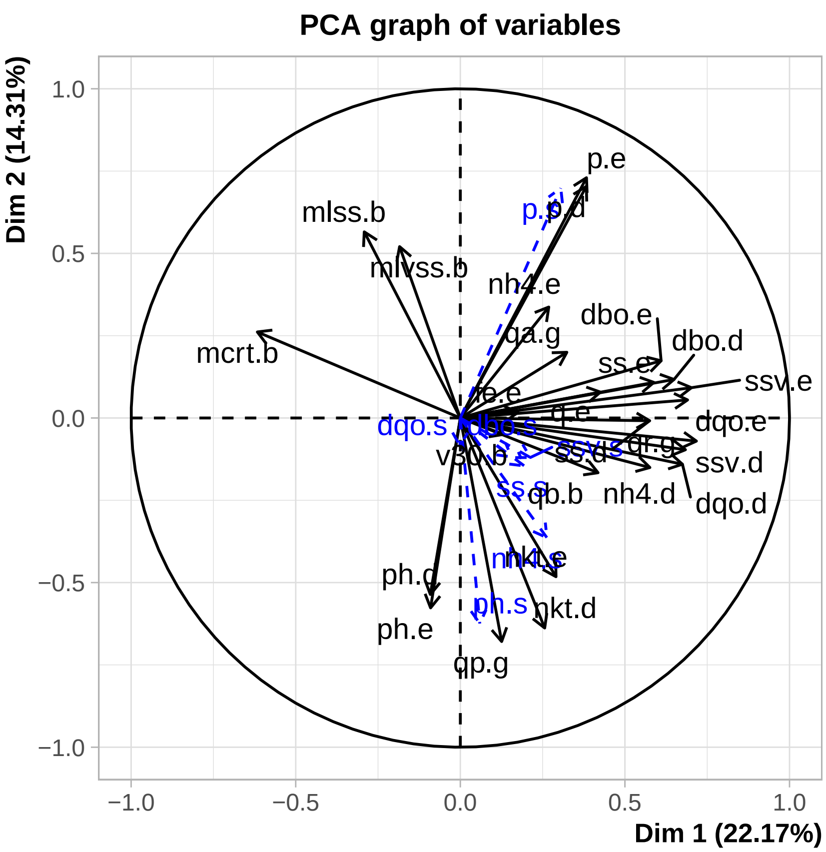
Eix 1: * És difícil resumir-lo, però la correlació positiva amb l'eix 1 és + ssv.d + ssv.e * i negativa correlacionada amb * mcrt.b * mlss.b

Eix 2: * Per a l'eix 2, apareix una correlació positiva per a + p.e + p.d * s'associa inversament a + qp.g + nkt.d

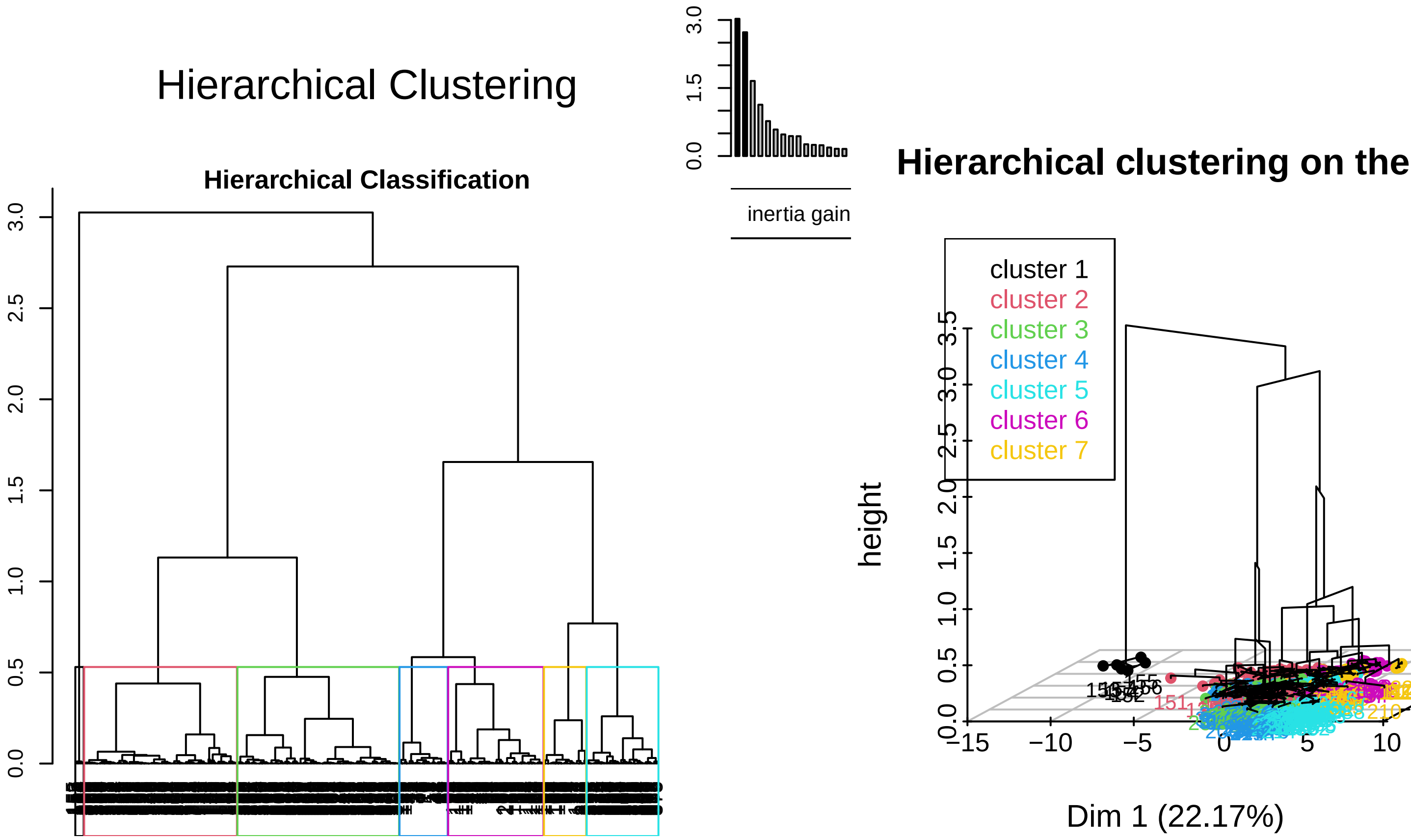
8. A Hierarchical Clustering is addressed. How many clusters are needed to represent 60% of the total inertia.

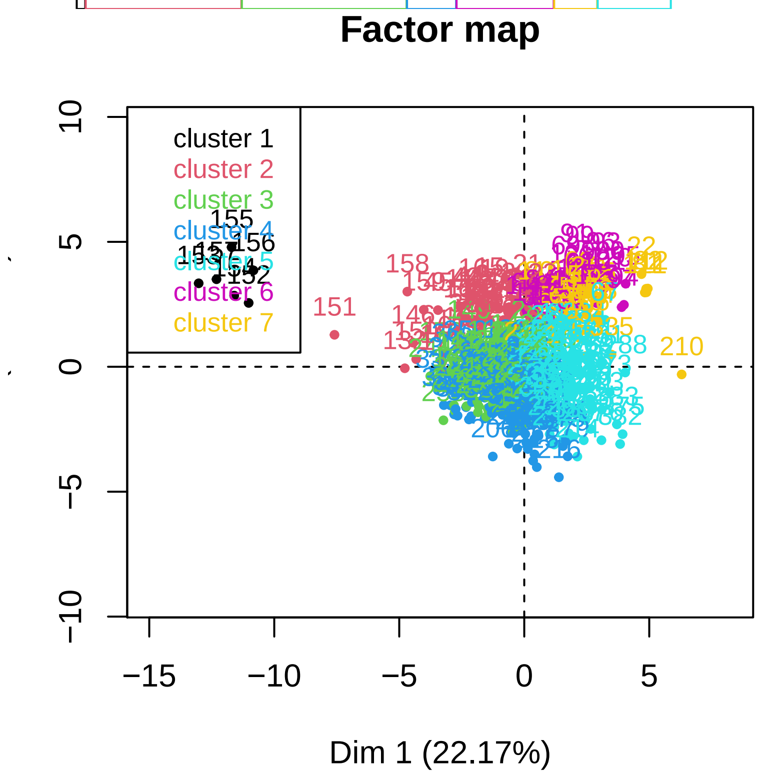
```r
# 6 dimensions have to be selected according to Kaiser's criteria
res.pca <- PCA(df[, c(4:30,32:36,40,45:46)], quanti.sup=c(23:29), quali.sup=c(34:35))
```






```r
res.hcpc <- HCPC(res.pca,nb.clust=7, graph=T)
```






```
res.hcpc$desc.var
```

```
##
```

```
## Link between the cluster variable and the categorical variables (chi-square test)
## =================================================================================
##             p.value df
## season 1.348225e-67 18
##
## Description of each cluster by the categories
## =============================================
## $`1`
##                Cla/Mod Mod/Cla  Global      p.value   v.test
## season=Winter 6.593407     100 22.9798 0.0001292797 3.827827
##
## $`2`
##                Cla/Mod  Mod/Cla   Global      p.value    v.test
## season=Autumn 26.73267 45.76271 25.50505 2.550525e-04  3.657133
## season=Spring  0.00000  0.00000 23.23232 3.834159e-08 -5.498325
##
## $`3`
##                 Cla/Mod   Mod/Cla   Global      p.value    v.test
## season=Winter 45.054945 47.674419 22.97980 7.055466e-09  5.789599
## season=Spring 42.391304 45.348837 23.23232 2.146518e-07  5.186180
## season=Summer  4.464286  5.813953 28.28283 9.833661e-09 -5.733573
## season=Autumn  0.990099  1.162791 25.50505 9.561075e-12 -6.812960
##
## $`4`
##                  Cla/Mod   Mod/Cla   Global      p.value    v.test
## season=Summer  60.714286 70.833333 28.28283 1.679165e-24 10.216127
## weekday=Sunday 40.350877 23.958333 14.39394 3.613743e-03  2.910047
## season=Spring  15.217391 14.583333 23.23232 1.871663e-02 -2.351127
## season=Winter   8.791209  8.333333 22.97980 3.097796e-05 -4.166155
## season=Autumn   5.940594  6.250000 25.50505 6.432604e-08 -5.406339
##
## $`5`
##                  Cla/Mod   Mod/Cla   Global      p.value    v.test
## season=Spring  40.217391 51.388889 23.23232 5.736828e-09  5.824254
## weekday=Sunday  3.508772  2.777778 14.39394 5.760943e-04 -3.442627
## season=Autumn   4.950495  6.944444 25.50505 1.480660e-05 -4.331546
##
## $`6`
##                  Cla/Mod Mod/Cla   Global      p.value    v.test
## season=Autumn 30.6930693    77.5 25.50505 3.501513e-13  7.273554
## season=Summer  0.8928571     2.5 28.28283 1.462911e-05 -4.334200
## season=Spring  0.0000000     0.0 23.23232 1.338267e-05 -4.353753
##
## $`7`
##                 Cla/Mod   Mod/Cla   Global      p.value    v.test
## season=Autumn 30.693069 83.783784 25.50505 5.460010e-15  7.815834
## season=Winter  2.197802  5.405405 22.97980 3.792140e-03 -2.894954
## season=Spring  2.173913  5.405405 23.23232 3.417305e-03 -2.927471
## season=Summer  1.785714  5.405405 28.28283 3.767414e-04 -3.555862
##
##
## Link between the cluster variable and the quantitative variables
## ================================================================
##               Eta2      P-value
```

```
## mcrt.b  0.93898517 9.914575e-233
## p.e     0.71293982 3.690727e-102
## p.d     0.67450573  1.359173e-91
## p.s     0.67337209  2.663855e-91
## qr.g    0.60623525  1.334174e-75
## qp.g    0.54775435  5.427995e-64
## ssv.e   0.51169712  1.431562e-57
## nh4.e   0.49402239  1.346049e-54
## q.e     0.49086892  4.450176e-54
## qb.b    0.46430530  7.882009e-50
## dqo.e   0.44245360  1.707933e-46
## ss.e    0.42937147  1.464882e-44
## ssv.d   0.42473874  6.910469e-44
## nh4.d   0.37902450  1.588827e-37
## dbo.e   0.37873285  1.738118e-37
## mlss.b  0.37787682  2.261778e-37
## ph.s    0.35355001  3.447846e-34
## dbo.d   0.35326769  3.747544e-34
## dqo.d   0.34878244  1.401556e-33
## nkt.d   0.33062051  2.657144e-31
## ss.d    0.32462719  1.451161e-30
## nkt.e   0.31395131  2.870430e-29
## nh4.s   0.29191383  1.165500e-26
## mlvss.b 0.29071902  1.604821e-26
## ph.e    0.28824569  3.105839e-26
## ph.d    0.22100966  7.779382e-19
## qa.g    0.19084698  9.454592e-16
## v30.b   0.17144284  7.707267e-14
## fe.e    0.12533723  1.579860e-09
## ssv.s   0.03910414  1.609294e-02
## ss.s    0.03420714  3.432497e-02
##
## Description of each cluster by quantitative variables
## =====================================================
## $`1`
##             v.test Mean in category Overall mean sd in category   Overall sd
## mcrt.b   19.241642     2.582033e+02 1.429049e+01   5.794879e+01 3.124888e+01
## mlss.b    7.934118     2.929833e+03 1.766515e+03   2.365928e+02 3.614446e+02
## mlvss.b   5.512139     1.937167e+03 1.344251e+03   9.663232e+01 2.651639e+02
## ph.s     -2.202465     7.358333e+00 7.533333e+00   1.426437e-01 1.958714e-01
## ph.e     -2.229630     7.491667e+00 7.618813e+00   1.017213e-01 1.405767e-01
## ss.e     -2.253882     1.325000e+02 2.096944e+02   8.287692e+01 8.442999e+01
## p.s      -2.459619     9.750000e-01 2.746086e+00   8.539126e-02 1.775063e+00
## p.d      -2.920306     1.800000e+00 6.071465e+00   6.831301e-02 3.605706e+00
## fe.e     -3.017039     3.483013e+01 4.737019e+01   1.783332e+01 1.024615e+01
## p.e      -3.149663     2.675000e+00 9.902146e+00   3.927998e-01 5.656459e+00
## nh4.s    -3.279402     1.800000e+00 1.216982e+01   8.880691e-01 7.795045e+00
## dbo.e    -3.333698     8.341667e+01 2.136061e+02   7.115938e+00 9.627007e+01
## ss.d     -3.730008     5.891667e+01 8.872096e+01   1.934429e+01 1.969749e+01
## dbo.d    -4.501487     5.066667e+01 1.199558e+02   2.277100e+00 3.794470e+01
## ssv.e    -4.701110     3.716667e+01 1.574583e+02   9.770989e+00 6.307787e+01
## dqo.e    -4.951190     1.101667e+02 4.423813e+02   6.359092e+01 1.654059e+02
## nkt.d    -5.170245     1.973462e+01 3.641604e+01   3.166045e+00 7.953600e+00
## nkt.e    -5.292623     2.210000e+01 4.195316e+01   2.869146e+00 9.246992e+00
```

```
## qa.g     -5.361405     1.228293e+05 2.317322e+05   1.400397e+04 5.007290e+04
## ssv.d    -6.371728     2.583333e+01 6.490783e+01   1.096079e+01 1.511743e+01
## dqo.d    -6.629825     8.833333e+01 2.496414e+02   5.250608e+01 5.997854e+01
## nh4.d    -8.168626     1.065000e+01 2.594192e+01   1.195478e+00 4.614821e+00
## q.e      -9.284755     2.256380e+04 4.180854e+04   1.066996e+03 5.109559e+03
## qb.b    -10.336716     2.179937e+04 3.890254e+04   1.013377e+03 4.078833e+03
## qr.g    -14.106104     1.813152e+04 4.096489e+04   1.267343e+02 3.990294e+03
##              p.value
## mcrt.b  1.658705e-82
## mlss.b  2.119973e-15
## mlvss.b 3.544987e-08
## ph.s    2.763251e-02
## ph.e    2.577200e-02
## ss.e    2.420360e-02
## p.s     1.390846e-02
## p.d     3.496882e-03
## fe.e    2.552569e-03
## p.e     1.634590e-03
## nh4.s   1.040274e-03
## dbo.e   8.569976e-04
## ss.d    1.914739e-04
## dbo.d   6.747975e-06
## ssv.e   2.587515e-06
## dqo.e   7.376090e-07
## nkt.d   2.337869e-07
## nkt.e   1.205745e-07
## qa.g    8.257733e-08
## ssv.d   1.869103e-10
## dqo.d   3.360847e-11
## nh4.d   3.119218e-16
## q.e     1.620793e-20
## qb.b    4.807277e-25
## qr.g    3.482973e-45
##
## $`2`
##            v.test Mean in category Overall mean sd in category   Overall sd
## p.s      8.265379         4.510363     2.746086      1.9182178    1.7750626
## p.d      6.214796         8.766147     6.071465      3.9459771    3.6057065
## p.e      5.855474        13.885026     9.902146      5.3318790    5.6564592
## mlss.b   4.716562      1971.516949  1766.515152    299.4233861  361.4446128
## mlvss.b  4.021792      1472.491525  1344.251263    192.7648407  265.1639251
## q.e      2.752603     43499.827119 41808.539773   4897.0590615 5109.5594150
## ph.d    -3.778421         7.499153     7.562879      0.1703664    0.1402548
## dbo.e   -4.394829       162.728814   213.606061     66.4209306   96.2700653
## dbo.d   -4.410633        99.830508   119.955808     28.5098274   37.9446951
## ph.e    -4.521873         7.542373     7.618813      0.1653998    0.1405767
## nkt.e   -4.560575        36.881959    41.953157      7.0592318    9.2469924
## nkt.d   -4.647924        31.970612    36.416035      4.4723368    7.9535998
## v30.b   -4.714438       187.169492   262.714646     51.3918653  133.2558559
## ss.d    -4.733698        77.508475    88.720960     14.8622179   19.6974930
## nh4.d   -4.790340        23.283573    25.941919      4.6200681    4.6148210
## ss.e    -4.860417       160.347458   209.694444     46.6544505   84.4299924
## dqo.e   -5.321985       336.525424   442.381313    104.5651781  165.4059089
## dqo.d   -5.875723       207.262712   249.641414     52.5820066   59.9785437
```

```
## ssv.e    -5.895824       112.737288   157.458333     36.7601428   63.0778719
## ssv.d    -5.995596        54.008475    64.907828     10.6335301   15.1174319
## ph.s     -6.632244         7.377119     7.533333      0.1690758    0.1958714
## nh4.s    -7.866669         4.795878    12.169823      3.1739990    7.7950446
## qp.g    -11.260392       430.684746   630.737857     82.4954006  147.7411209
##              p.value
## p.s     1.392512e-16
## p.d     5.139134e-10
## p.e     4.756510e-09
## mlss.b  2.398637e-06
## mlvss.b 5.775715e-05
## q.e     5.912358e-03
## ph.d    1.578261e-04
## dbo.e   1.108602e-05
## dbo.d   1.030690e-05
## ph.e    6.129478e-06
## nkt.e   5.101363e-06
## nkt.d   3.352931e-06
## v30.b   2.423789e-06
## ss.d    2.204660e-06
## nh4.d   1.664991e-06
## ss.e    1.171390e-06
## dqo.e   1.026413e-07
## dqo.d   4.210009e-09
## ssv.e   3.728160e-09
## ssv.d   2.027411e-09
## ph.s    3.306222e-11
## nh4.s   3.642084e-15
## qp.g    2.058401e-29
##
## $`3`
##          v.test Mean in category Overall mean sd in category   Overall sd
## qb.b   7.596793     4.186259e+04 3.890254e+04   3.736642e+03 4.078833e+03
## ph.e   6.998062     7.712791e+00 7.618813e+00   1.151809e-01 1.405767e-01
## q.e    6.213987     4.484164e+04 4.180854e+04   3.097591e+03 5.109559e+03
## ph.d   3.725233     7.612791e+00 7.562879e+00   1.154609e-01 1.402548e-01
## qa.g   3.441173     2.481927e+05 2.317322e+05   4.119189e+04 5.007290e+04
## ssv.d -2.267158     6.163372e+01 6.490783e+01   9.795975e+00 1.511743e+01
## ss.s  -2.565305     1.252674e+01 1.651111e+01   6.968606e+00 1.625876e+01
## ssv.s -2.792663     8.992442e+00 1.235758e+01   5.524150e+00 1.261394e+01
## dqo.e -2.890822     3.967035e+02 4.423813e+02   8.052204e+01 1.654059e+02
## nh4.s -2.992351     9.941572e+00 1.216982e+01   6.502827e+00 7.795045e+00
## nh4.e -3.196035     2.482396e+01 4.114419e+01   4.905343e+00 5.345419e+01
## qr.g  -3.712500     3.954973e+04 4.096489e+04   2.921590e+03 3.990294e+03
## dbo.e -4.043637     1.764186e+02 2.136061e+02   5.374166e+01 9.627007e+01
## ssv.e -4.196953     1.321686e+02 1.574583e+02   2.550039e+01 6.307787e+01
## fe.e  -4.241511     4.321860e+01 4.737019e+01   1.216351e+01 1.024615e+01
## dbo.d -4.318439     1.043023e+02 1.199558e+02   2.595910e+01 3.794470e+01
## ss.e  -4.468116     1.736570e+02 2.096944e+02   3.473013e+01 8.442999e+01
## nh4.d -4.814362     2.381952e+01 2.594192e+01   4.253499e+00 4.614821e+00
## p.e   -7.462399     5.869815e+00 9.902146e+00   1.131790e+00 5.656459e+00
## p.s   -7.636893     1.451105e+00 2.746086e+00   8.383211e-01 1.775063e+00
## p.d   -8.365265     3.190067e+00 6.071465e+00   8.400500e-01 3.605706e+00
##            p.value
```

```
## qb.b  3.035599e-14
## ph.e  2.595269e-12
## q.e   5.165683e-10
## ph.d  1.951345e-04
## qa.g  5.791976e-04
## ssv.d 2.338057e-02
## ss.s  1.030853e-02
## ssv.s 5.227609e-03
## dqo.e 3.842359e-03
## nh4.s 2.768377e-03
## nh4.e 1.393300e-03
## qr.g  2.052224e-04
## dbo.e 5.262845e-05
## ssv.e 2.705305e-05
## fe.e  2.220198e-05
## dbo.d 1.571363e-05
## ss.e  7.891150e-06
## nh4.d 1.476709e-06
## p.e   8.496134e-14
## p.s   2.225257e-14
## p.d   5.997998e-17
##
## $`4`
##            v.test Mean in category Overall mean sd in category   Overall sd
## qp.g     8.369860     7.407260e+02 6.307379e+02   1.019273e+02 1.477411e+02
## ph.s     7.882365     7.670660e+00 7.533333e+00   1.511864e-01 1.958714e-01
## nh4.s    7.260424     1.720375e+01 1.216982e+01   6.991434e+00 7.795045e+00
## nkt.d    6.091541     4.072544e+01 3.641604e+01   9.888748e+00 7.953600e+00
## nkt.e    5.665565     4.661299e+01 4.195316e+01   9.802301e+00 9.246992e+00
## fe.e     5.361318     5.225625e+01 4.737019e+01   7.045681e+00 1.024615e+01
## nh4.d    3.641203     2.743652e+01 2.594192e+01   2.427227e+00 4.614821e+00
## qr.g     3.372423     4.216183e+04 4.096489e+04   1.259576e+03 3.990294e+03
## ph.d     3.226123     7.603125e+00 7.562879e+00   9.236829e-02 1.402548e-01
## ph.e     2.577513     7.651042e+00 7.618813e+00   8.798847e-02 1.405767e-01
## nh4.e   -2.093786     3.118921e+01 4.114419e+01   1.358853e+01 5.345419e+01
## dbo.e   -2.830604     1.893681e+02 2.136061e+02   5.007340e+01 9.627007e+01
## ssv.e   -2.939045     1.409688e+02 1.574583e+02   3.744897e+01 6.307787e+01
## ssv.d   -3.081832     6.076389e+01 6.490783e+01   1.041396e+01 1.511743e+01
## dqo.e   -3.511547     3.907188e+02 4.423813e+02   1.128559e+02 1.654059e+02
## dbo.d   -3.532657     1.080330e+02 1.199558e+02   3.089277e+01 3.794470e+01
## ss.e    -3.808495     1.810938e+02 2.096944e+02   5.155376e+01 8.442999e+01
## ss.d    -4.637792     8.059549e+01 8.872096e+01   1.477450e+01 1.969749e+01
## p.d     -4.783936     4.537193e+00 6.071465e+00   1.504983e+00 3.605706e+00
## p.e     -5.541238     7.114242e+00 9.902146e+00   2.566078e+00 5.656459e+00
## p.s     -5.587146     1.863961e+00 2.746086e+00   4.651973e-01 1.775063e+00
## qa.g    -6.122437     2.044642e+05 2.317322e+05   4.097892e+04 5.007290e+04
## qb.b    -6.244545     3.663704e+04 3.890254e+04   3.339490e+03 4.078833e+03
## mlvss.b -8.274550     1.149094e+03 1.344251e+03   2.057226e+02 2.651639e+02
## mlss.b  -8.750404     1.485198e+03 1.766515e+03   2.811038e+02 3.614446e+02
## q.e     -9.262379     3.759902e+04 4.180854e+04   3.839056e+03 5.109559e+03
##              p.value
## qp.g    5.768686e-17
## ph.s    3.212422e-15
## nh4.s   3.858797e-13
```

```
## nkt.d   1.118292e-09
## nkt.e   1.465409e-08
## fe.e    8.261702e-08
## nh4.d   2.713665e-04
## qr.g    7.450997e-04
## ph.d    1.254794e-03
## ph.e    9.951428e-03
## nh4.e   3.627905e-02
## dbo.e   4.646018e-03
## ssv.e   3.292256e-03
## ssv.d   2.057312e-03
## dqo.e   4.455069e-04
## dbo.d   4.114051e-04
## ss.e    1.398153e-04
## ss.d    3.521510e-06
## p.d     1.718956e-06
## p.e     3.003411e-08
## p.s     2.308316e-08
## qa.g    9.215473e-10
## qb.b    4.250360e-10
## mlvss.b 1.289419e-16
## mlss.b  2.125897e-18
## q.e     1.999270e-20
##
## $`5`
##          v.test Mean in category Overall mean sd in category   Overall sd
## ssv.d  9.548430     8.031481e+01 6.490783e+01   1.530824e+01   15.1174319
## ss.d   9.089570     1.078310e+02 8.872096e+01   2.239636e+01   19.6974930
## dqo.d  8.397822     3.034028e+02 2.496414e+02   4.792838e+01   59.9785437
## dbo.d  6.937578     1.480532e+02 1.199558e+02   3.612720e+01   37.9446951
## ss.e   6.023836     2.639792e+02 2.096944e+02   9.162323e+01   84.4299924
## dqo.e  5.909328     5.467083e+02 4.423813e+02   1.245167e+02  165.4059089
## nkt.d  5.803392     4.134270e+01 3.641604e+01   6.830697e+00    7.9535998
## qp.g   5.415340     7.161333e+02 6.307379e+02   1.215539e+02  147.7411209
## ssv.e  5.237758     1.927222e+02 1.574583e+02   5.045831e+01   63.0778719
## nh4.d  4.761835     2.828743e+01 2.594192e+01   4.223998e+00    4.6148210
## ph.s   4.362483     7.624537e+00 7.533333e+00   1.669205e-01    0.1958714
## nkt.e  4.315820     4.621278e+01 4.195316e+01   8.274843e+00    9.2469924
## ph.d   3.500238     7.615278e+00 7.562879e+00   9.952627e-02    0.1402548
## dbo.e  3.427859     2.488287e+02 2.136061e+02   6.646388e+01   96.2700653
## ssv.s  2.674456     1.595833e+01 1.235758e+01   1.877528e+01   12.6139385
## ss.s   2.653115     2.111528e+01 1.651111e+01   2.430286e+01   16.2587559
## dbo.s  2.330972     2.777778e-01 1.818182e-01   4.479032e-01    0.3856946
## ph.e   2.124786     7.650694e+00 7.618813e+00   9.446231e-02    0.1405767
## qb.b   2.041321     3.979124e+04 3.890254e+04   3.070174e+03 4078.8329019
## v30.b  1.975021     2.908056e+02 2.627146e+02   1.457519e+02  133.2558559
## p.d   -3.956847     4.548649e+00 6.071465e+00   1.547965e+00    3.6057065
## p.e   -4.380875     7.257222e+00 9.902146e+00   2.222683e+00    5.6564592
## p.s   -4.653603     1.864406e+00 2.746086e+00   7.213007e-01    1.7750626
##            p.value
## ssv.d 1.316765e-21
## ss.d  9.943106e-20
## dqo.d 4.548371e-17
## dbo.d 3.988798e-12
```

```
## ss.e  1.703307e-09
## dqo.e 3.435068e-09
## nkt.d 6.498681e-09
## qp.g  6.117220e-08
## ssv.e 1.625386e-07
## nh4.d 1.918403e-06
## ph.s  1.285945e-05
## nkt.e 1.590116e-05
## ph.d  4.648427e-04
## dbo.e 6.083610e-04
## ssv.s 7.485059e-03
## ss.s  7.975278e-03
## dbo.s 1.975481e-02
## ph.e  3.360449e-02
## qb.b  4.121893e-02
## v30.b 4.826573e-02
## p.d   7.594566e-05
## p.e   1.182036e-05
## p.s   3.261844e-06
##
## $`6`
##             v.test Mean in category Overall mean sd in category   Overall sd
## nh4.e   13.908195       152.740286    41.144192     109.4581539   53.4541918
## p.e     11.428445        19.605643     9.902146       2.5264603    5.6564592
## p.d     10.136979        11.557964     6.071465       0.9721012    3.6057065
## p.s      8.983253         5.139643     2.746086       0.4133082    1.7750626
## v30.b    6.640961       395.550000   262.714646     140.5862636  133.2558559
## dbo.d    5.457040       151.037500   119.955808      27.8837738   37.9446951
## nh4.d    4.438122        29.016250    25.941919       2.3294040    4.6148210
## qr.g     3.914289     43309.410000 40964.888636    1542.8587806 3990.2935927
## mlvss.b  3.412438      1480.075000  1344.251263     205.1073362  265.1639251
## ss.d     2.956518        97.462500    88.720960      13.3099528   19.6974930
## ssv.d    2.679203        70.987500    64.907828      11.0405432   15.1174319
## mlss.b   2.589814      1907.025000  1766.515152     227.4090464  361.4446128
## nkt.e    2.299765        45.145286    41.953157       8.8360203    9.2469924
## dbo.e    2.105886       244.037500   213.606061      62.5530163   96.2700653
## nkt.d   -3.726221        31.967380    36.416035       3.4349947    7.9535998
## ph.e    -4.564316         7.522500     7.618813       0.1254741    0.1405767
## qp.g    -4.785547       524.610000   630.737857      55.0190322  147.7411209
## ph.s    -5.427753         7.373750     7.533333       0.1710217    0.1958714
## ph.d    -5.836639         7.440000     7.562879       0.1475635    0.1402548
##               p.value
## nh4.e   5.648705e-44
## p.e     3.014595e-30
## p.d     3.786292e-24
## p.s     2.628762e-19
## v30.b   3.116449e-11
## dbo.d   4.841376e-08
## nh4.d   9.074698e-06
## qr.g    9.067098e-05
## mlvss.b 6.438455e-04
## ss.d    3.111339e-03
## ssv.d   7.379760e-03
## mlss.b  9.602784e-03
```

```
## nkt.e   2.146153e-02
## dbo.e   3.521429e-02
## nkt.d   1.943720e-04
## ph.e    5.011250e-06
## qp.g    1.705226e-06
## ph.s    5.706794e-08
## ph.d    5.326427e-09
##
## $`7`
##             v.test Mean in category Overall mean sd in category   Overall sd
## ssv.e   10.786488     2.640946e+02 1.574583e+02   7.926426e+01 6.307787e+01
## dbo.e    9.949702     3.637297e+02 2.136061e+02   1.689280e+02 9.627007e+01
## ss.e     9.812605     3.395405e+02 2.096944e+02   9.487997e+01 8.442999e+01
## dqo.e    9.501222     6.886892e+02 4.423813e+02   2.377629e+02 1.654059e+02
## p.d      7.265442     1.017729e+01 6.071465e+00   2.694517e+00 3.605706e+00
## p.e      6.862087     1.598558e+01 9.902146e+00   4.045566e+00 5.656459e+00
## p.s      6.831997     4.646766e+00 2.746086e+00   1.324410e+00 1.775063e+00
## dbo.d    3.761303     1.423243e+02 1.199558e+02   3.530164e+01 3.794470e+01
## qr.g     2.719729     4.266579e+04 4.096489e+04   2.030601e+03 3.990294e+03
## qa.g     2.553264     2.517699e+05 2.317322e+05   5.099712e+04 5.007290e+04
## ssv.d    2.331714     7.043243e+01 6.490783e+01   1.072643e+01 1.511743e+01
## mlvss.b -2.481215     1.241135e+03 1.344251e+03   2.933543e+02 2.651639e+02
## qp.g    -3.020391     5.608000e+02 6.307379e+02   1.123057e+02 1.477411e+02
## ph.s    -3.022703     7.440541e+00 7.533333e+00   1.240302e-01 1.958714e-01
## ph.d    -4.151467     7.471622e+00 7.562879e+00   1.353377e-01 1.402548e-01
## nkt.d   -4.408933     3.092006e+01 3.641604e+01   3.258221e+00 7.953600e+00
## ph.e    -5.331328     7.501351e+00 7.618813e+00   1.411758e-01 1.405767e-01
## nkt.e   -5.890872     3.341571e+01 4.195316e+01   3.093561e+00 9.246992e+00
##              p.value
## ssv.e   3.987375e-27
## dbo.e   2.529386e-23
## ss.e    9.936922e-23
## dqo.e   2.074415e-21
## p.d     3.718204e-13
## p.e     6.786160e-12
## p.s     8.374078e-12
## dbo.d   1.690307e-04
## qr.g    6.533552e-03
## qa.g    1.067184e-02
## ssv.d   1.971573e-02
## mlvss.b 1.309355e-02
## qp.g    2.524488e-03
## ph.s    2.505282e-03
## ph.d    3.303515e-05
## nkt.d   1.038810e-05
## ph.e    9.749690e-08
## nkt.e   3.841633e-09
```

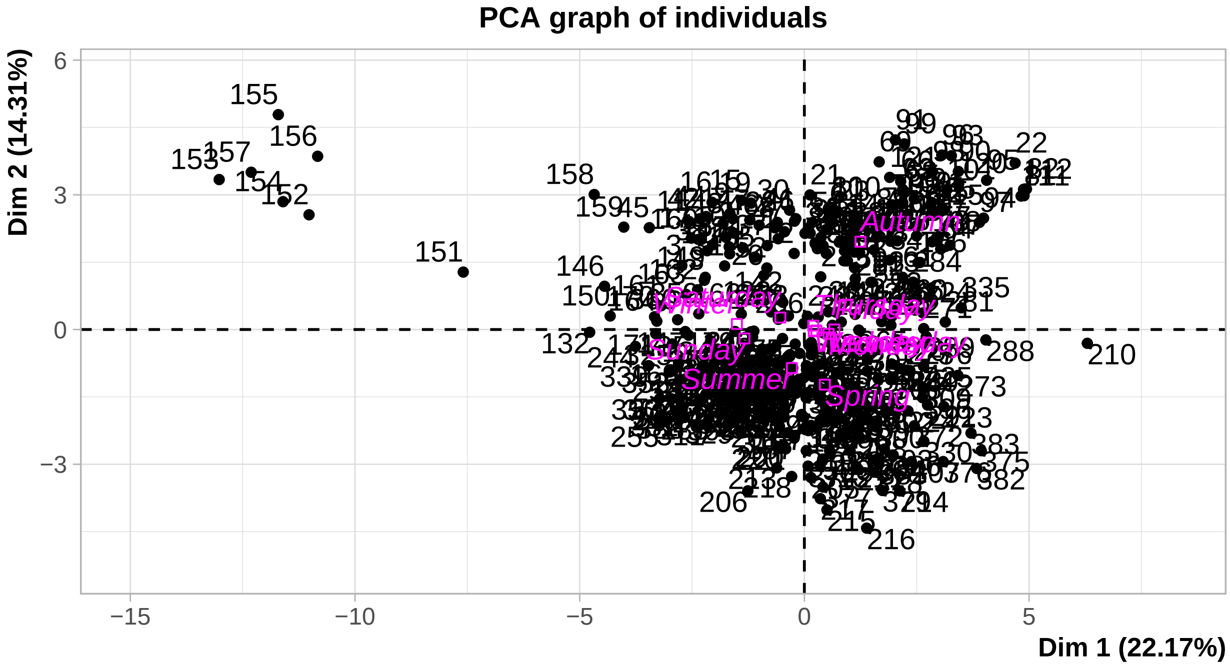
```
(res.hcpc$call$t$within[1]-res.hcpc$call$t$within[7])/res.hcpc$call$t$within[1] # representació de la i
```

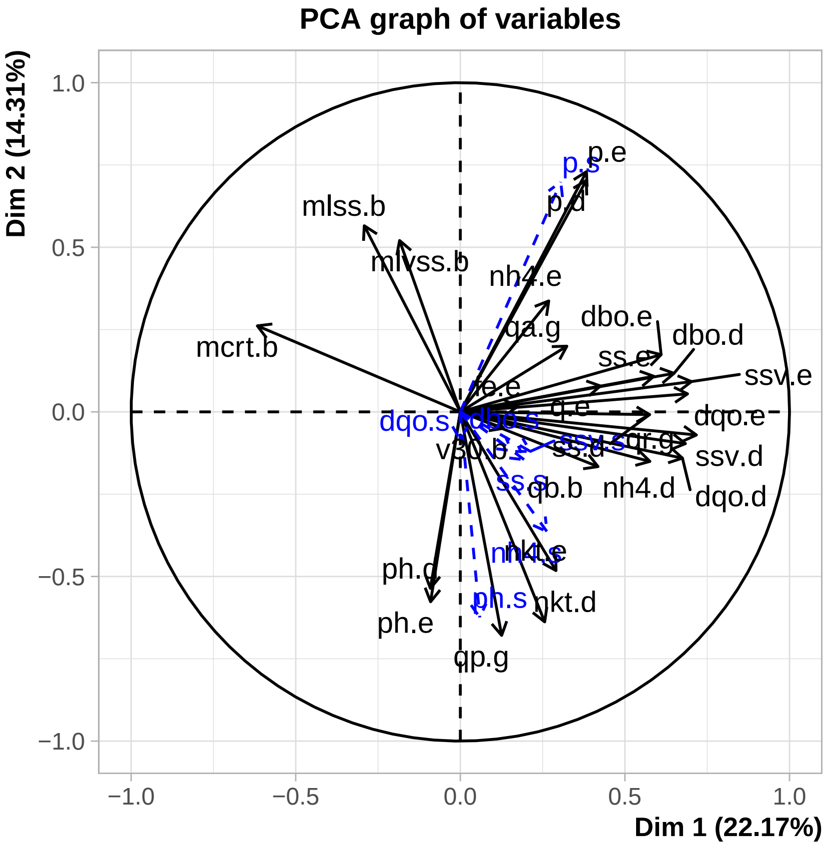
```
## [1] 0.6286321
```

Clusters: 7. Inèrcia: 62.86%

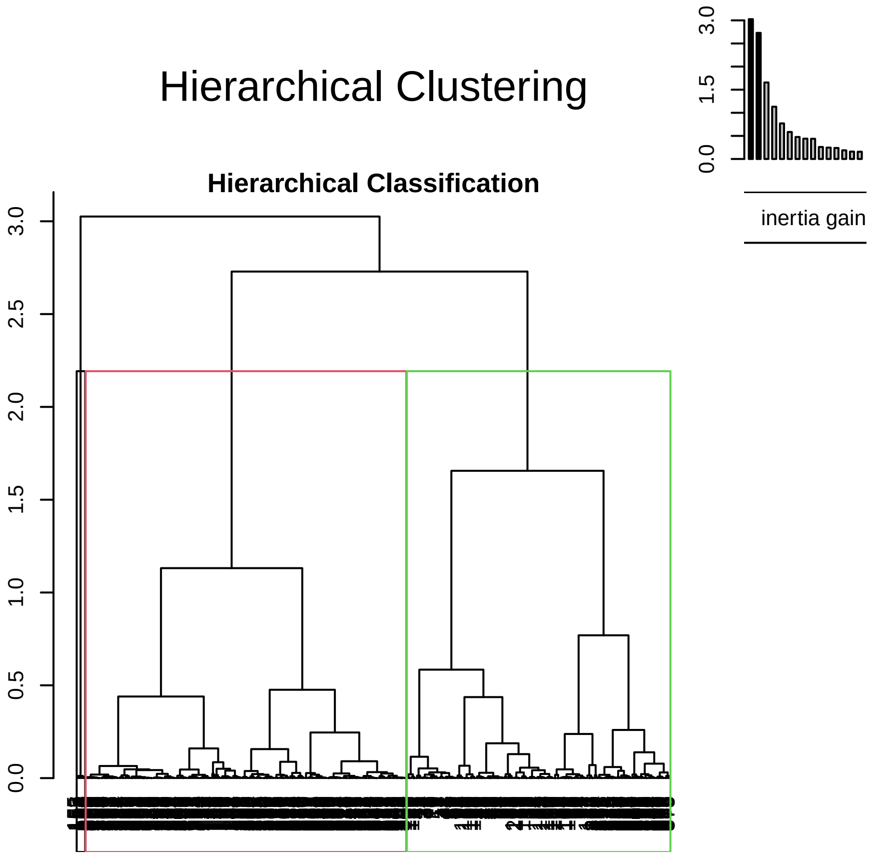
9. A nondefault criteria for selecting the number of clusters to 3 has to be set. Explain the characteristics of cluster number 3.

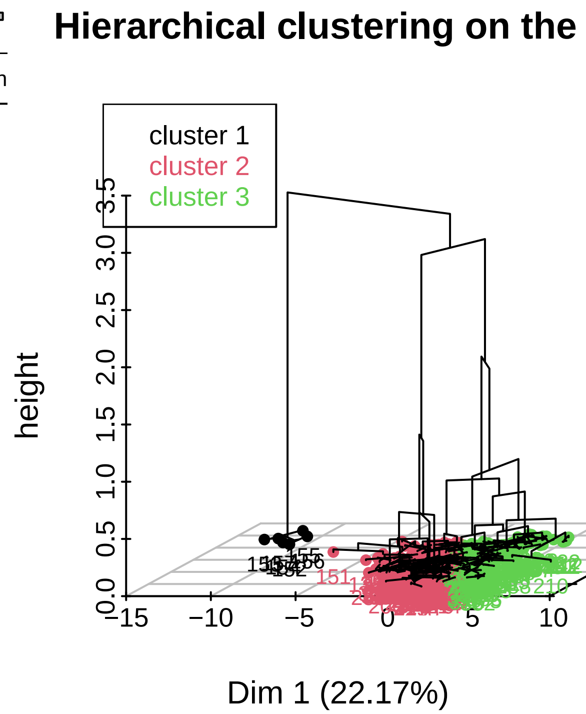
```r
# 6 dimensions have to be selected according to Kaiser's criteria
res.pca <- PCA(df[, c(4:30,32:36,40,45:46)], quanti.sup=c(23:29), quali.sup=c(34:35))
```

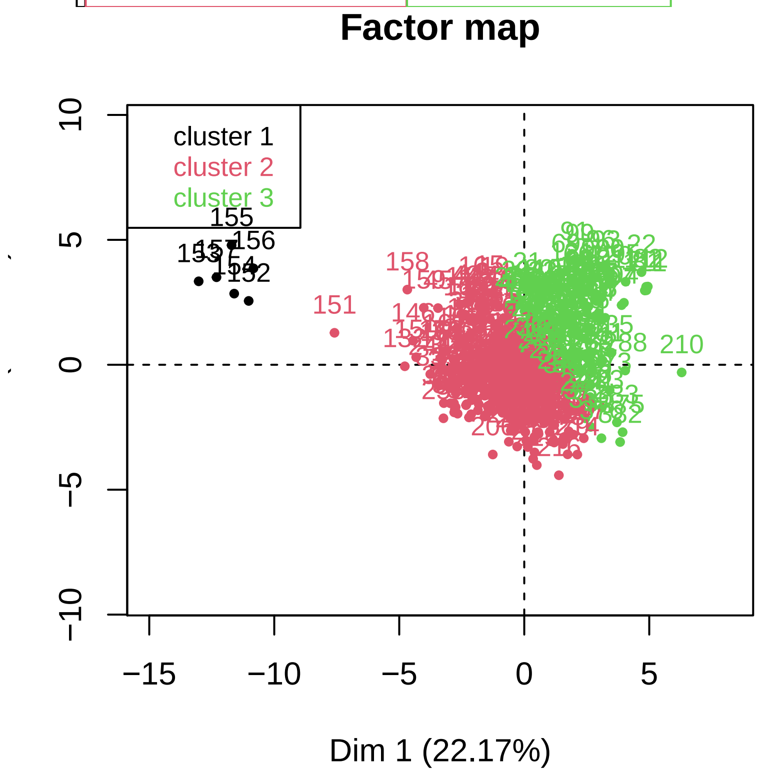





```r
res.hcpc <- HCPC(res.pca,nb.clust=3, graph=T)
```








```
(res.hcpc$call$t$within[1]-res.hcpc$call$t$within[3])/res.hcpc$call$t$within[1] # representació de la i
```

```
## [1] 0.3655823
```

10. Use a partition method to group available data into the selected number of clusters found in Question 7. Determine the quality of the partition and plot the resulting partition in the first factorial plane.

```
# Ja no m'ha donat temps.
```

Do not forget to Knit to .pdf before posting your answers in ATENEA.