Projet ADD L3

Kanthakumar, Nesho

Décembre 2023

Question 1.

On charge en mémoire les données et on en donne un aperçu:

```r
# setwd("chemin_vers_répertoire_de_travail") # A modifier! Par exemple:
setwd("C:\\Users\\Nesho\\Desktop\\projet" )
# En Windows ne pas oublier de remplacer \ par \\

x <- read.table("climats.txt", sep = ";",
                header = TRUE, row.names = 1)
# View(x)
head(x) # pour afficher les six premières colonnes
```

```
##             January February March April  May June July August September
## Amsterdam       2.9      2.5   5.7   8.2 12.5 14.8 17.1   17.1      14.5
## Athens          9.1      9.7  11.7  15.4 20.1 24.5 27.4   27.2      23.8
## Berlin         -0.2      0.1   4.4   8.2 13.8 16.0 18.3   18.0      14.4
## Brussels        3.3      3.3   6.7   8.9 12.8 15.6 17.8   17.8      15.0
## Budapest       -1.1      0.8   5.5  11.6 17.0 20.2 22.0   21.3      16.9
## Copenhagen     -0.4     -0.4   1.3   5.8 11.1 15.4 17.1   16.6      13.3
##             October November December Annual Amplitude Latitude Longitude  Area
## Amsterdam      11.4      7.0      4.4    9.9      14.6     52.2       4.5  West
## Athens         19.2     14.6     11.0   17.8      18.3     37.6      23.5 South
## Berlin         10.0      4.2      1.2    9.1      18.5     52.3      13.2  West
## Brussels       11.1      6.7      4.4   10.3      14.4     50.5       4.2  West
## Budapest       11.3      5.1      0.7   10.9      23.1     47.3      19.0  East
## Copenhagen      8.8      4.1      1.3    7.8      17.5     55.4      12.3 North
```

Question 2.

```r
dim(x)
```

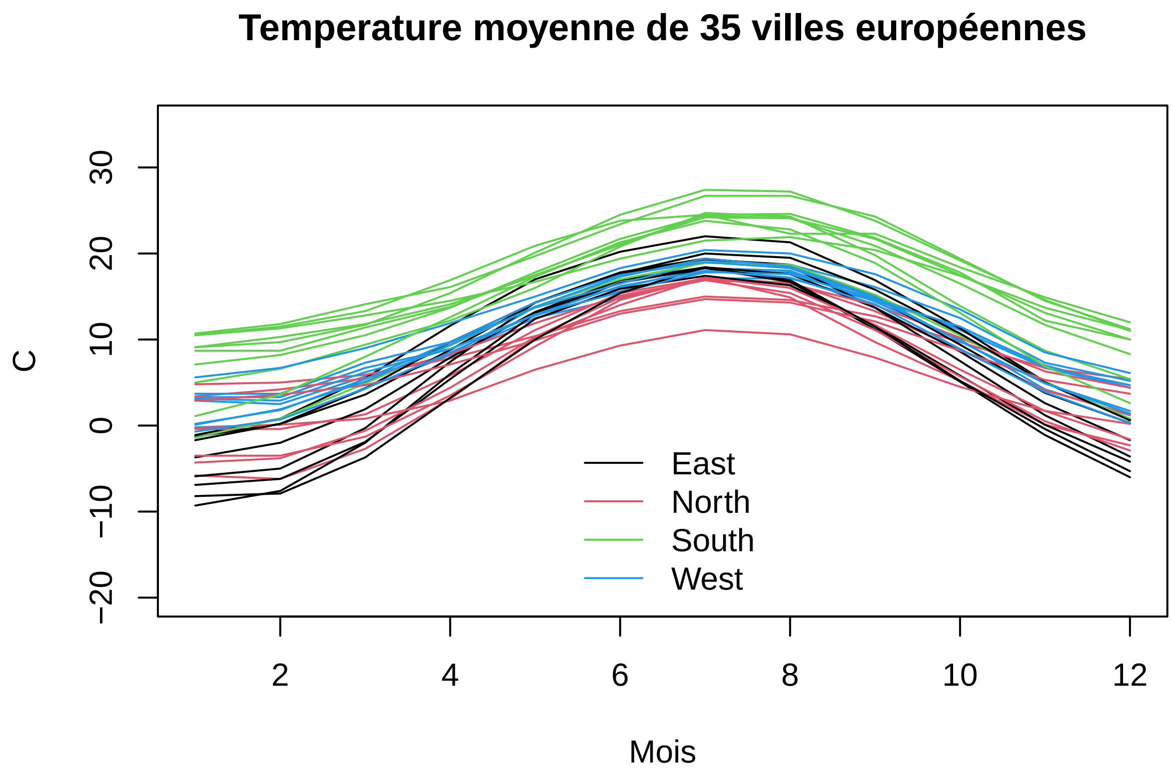
```
## [1] 35 17
```

```r
str(x)
```

```
## 'data.frame':    35 obs. of  17 variables:
##  $ January  : num  2.9 9.1 -0.2 3.3 -1.1 -0.4 4.8 -5.8 -5.9 -3.7 ...
##  $ February : num  2.5 9.7 0.1 3.3 0.8 -0.4 5 -6.2 -5 -2 ...
##  $ March    : num  5.7 11.7 4.4 6.7 5.5 1.3 5.9 -2.7 -0.3 1.9 ...
##  $ April    : num  8.2 15.4 8.2 8.9 11.6 5.8 7.8 3.1 7.4 7.9 ...
##  $ May      : num  12.5 20.1 13.8 12.8 17 11.1 10.4 10.2 14.3 13.2 ...
##  $ June     : num  14.8 24.5 16 15.6 20.2 15.4 13.3 14 17.8 16.9 ...
##  $ July     : num  17.1 27.4 18.3 17.8 22 17.1 15 17.2 19.4 18.4 ...
##  $ August   : num  17.1 27.2 18 17.8 21.3 16.6 14.6 14.9 18.5 17.6 ...
##  $ September: num  14.5 23.8 14.4 15 16.9 13.3 12.7 9.7 13.7 13.7 ...
##  $ October  : num  11.4 19.2 10 11.1 11.3 8.8 9.7 5.2 7.5 8.6 ...
##  $ November : num  7 14.6 4.2 6.7 5.1 4.1 6.7 0.1 1.2 2.6 ...
##  $ December : num  4.4 11 1.2 4.4 0.7 1.3 5.4 -2.3 -3.6 -1.7 ...
##  $ Annual   : num  9.9 17.8 9.1 10.3 10.9 7.8 9.3 4.8 7.1 7.7 ...
##  $ Amplitude: num  14.6 18.3 18.5 14.4 23.1 17.5 10.2 23.4 25.3 22.1 ...
##  $ Latitude : num  52.2 37.6 52.3 50.5 47.3 55.4 53.2 60.1 50.3 50 ...
##  $ Longitude: num  4.5 23.5 13.2 4.2 19 12.3 6.1 25 30.3 19.6 ...
##  $ Area     : chr  "West" "South" "West" "West" ...
```

Il y a $p = 12$ variables quantitatives primaires, chacune correspondante à une temperature mensuelle moyenne. L'espace des individus est donc $\mathbb{R}^{12}$. Chaque individu correspond à une ville européene. Au total, il y a $n = 35$ villes.

```r
plot(as.numeric(x[1,1:12]), type = 'l',
     col = as.factor(x$Area)[1], ylim = c(-20,35),
     ylab = "C", xlab = "Mois",
     main = "Temperature moyenne de 35 villes européennes")
for(i in 2:nrow(x)) points(as.numeric(x[i,1:12]),type='l',col=as.factor(x$Area)[i])
legend(x = 'bottom', lty = 1,
       col = 1:4, legend = c("East", "North", "South", "West"),
       bty = 'n')
```



Question 3.

Calculons la moyenne de chaque variable :

```
apply(x[,1:12], 2, mean)
```

```
##   January  February     March     April       May      June      July    August
##  1.345714  2.217143  5.228571  9.282857 13.911429 17.414286 19.622857 18.980000
## September   October  November  December
## 15.631429 11.002857  6.065714  2.880000
```

Calculons désormais l'écart-type de chaque variable :

```
sd2 <- sqrt(apply(x[,1:12], 2, var))*11/12
sd2
```

```
##   January  February     March     April       May      June      July    August
##  5.043644  5.040710  4.457787  3.489252  3.000783  3.043582  3.276783  3.417277
## September   October  November  December
##  3.767251  3.962957  4.186252  4.553460
```

Enfin, déterminons la variance totale du nuage de points dans l'espace des individus :

```
var_tot <- sum(apply(x[,1:12], 2, var))
var_tot
```

```
## [1] 228.1781
```

Les écarts types se trouvent entre 3 et 5, donc les variations de température sont plutôt grandes entre les différentes villes.

Question 4

Pour chaque région géographique, calculons la température moyenne de chaque mois :

```
apply(x[x$Area=="West",1:12], 2, mean)
```

```
##   January  February     March     April       May      June      July    August
##  2.000000  2.622222  6.011111  9.266667 13.522222 16.411111 18.544444 18.133333
## September   October  November  December
## 15.133333 10.944444  6.022222  3.266667
```

```
apply(x[x$Area=="East",1:12], 2, mean)
```

```
##   January  February     March     April       May      June      July    August
##   -4.7625   -3.4375    0.9250    7.5000   13.5625   17.2625   19.1500   18.1875
## September   October  November  December
##   13.6625    7.9500    2.0375   -2.4000
```

```
apply(x[x$Area=="South",1:12], 2, mean)
```

```
##   January  February     March     April       May      June      July    August
##      7.04      8.25     10.79     13.86     17.72     21.41     24.06     23.67
## September   October  November  December
##     20.91     16.37     11.54      8.24
```

```
apply(x[x$Area=="North",1:12], 2, mean)
```

```
##   January  February     March     April       May      June      July    August
##   -0.4000   -0.1250    1.7000    5.3625    9.9375   13.7000   15.7625   14.8625
## September   October  November  December
##   11.5625    7.4125    3.3000    1.0250
```

Les températures sont plutôt élevées au Sud et à l'Ouest contrairement à l'Est et au Nord où les températures sont les plus faibles.

Pour chaque région géographique, calculons désormais la variance de chaque mois :

```
apply(x[x$Area=="West",1:12], 2, var)
```

```
##   January  February     March     April       May      June      July    August
## 4.8175000 3.6994444 2.2411111 1.3075000 0.6644444 1.2311111 1.0002778 0.7900000
## September   October  November  December
## 1.1700000 2.1477778 2.5169444 4.4250000
```

```r
apply(x[x$Area=="East",1:12], 2, var)
```

```
##   January  February     March     April       May      June      July    August
## 10.596964 13.399821 11.407857  6.951429  3.991250  2.159821  1.977143  2.792679
## September   October  November  December
##  4.511250  5.974286  5.974107  7.491429
```

```r
apply(x[x$Area=="South",1:12], 2, var)
```

```
##   January  February     March     April       May      June      July    August
## 17.696000 13.202778  7.556556  4.696000  4.377333  4.821000  5.811556  5.971222
## September   October  November  December
##  6.814333  8.551222 12.004889 15.569333
```

```r
apply(x[x$Area=="North",1:12], 2, var)
```

```
##   January  February     March     April       May      June      July    August
## 15.194286 17.253571 10.814286  4.765536  2.568393  3.885714  4.565536  3.691250
## September   October  November  December
##  3.951250  4.818393  6.914286 10.256429
```

Enfin pour chaque région géographique, déterminons la variance totale du sous-nuage de points :

```r
var_1 <- sum(apply(x[x$Area=="West",1:12], 2, var))
var_2 <- sum(apply(x[x$Area=="East",1:12], 2, var))
var_3 <- sum(apply(x[x$Area=="South",1:12], 2, var))
var_4 <- sum(apply(x[x$Area=="North",1:12], 2, var))
```

Question 5

Calculons les variances inter et intra de la classification Cr donnée par les régions :

```r
n1 <- sum(x$Area == "West")
n2 <- sum(x$Area == "East")
n3 <- sum(x$Area == "South")
n4 <- sum(x$Area == "North")
n1+n2+n3+n4
```

```
## [1] 35
```

```r
var_intra <- n1/35* var_1 + n2/35* var_2 + n3/35 * var_3 + n4/35 * var_4
var_intra
```

```
## [1] 75.20223
```

```
var_inter <- var_tot - var_intra
var_inter
```

```
## [1] 152.9758
```

On obtient une variance intra de 75.2 et une variance inter de 152.98.

Question 6

```
apply(x[,1:12],2,var)
```

```
##   January  February     March     April       May      June      July    August
##  30.27373  30.23852  23.64916  14.48911  10.71634  11.02420  12.77829  13.89753
## September   October  November  December
##  16.88987  18.69029  20.85585  24.67518
```

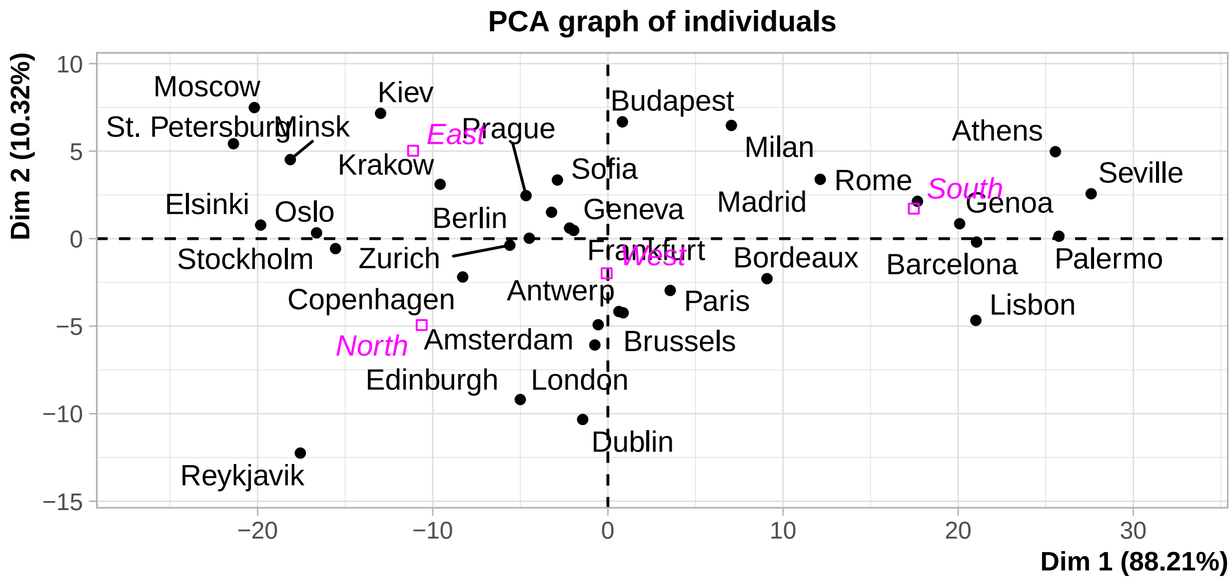
Les variances sont plutôt proches donc on va utiliser l'ACP simple

```
library(FactoMineR)
```

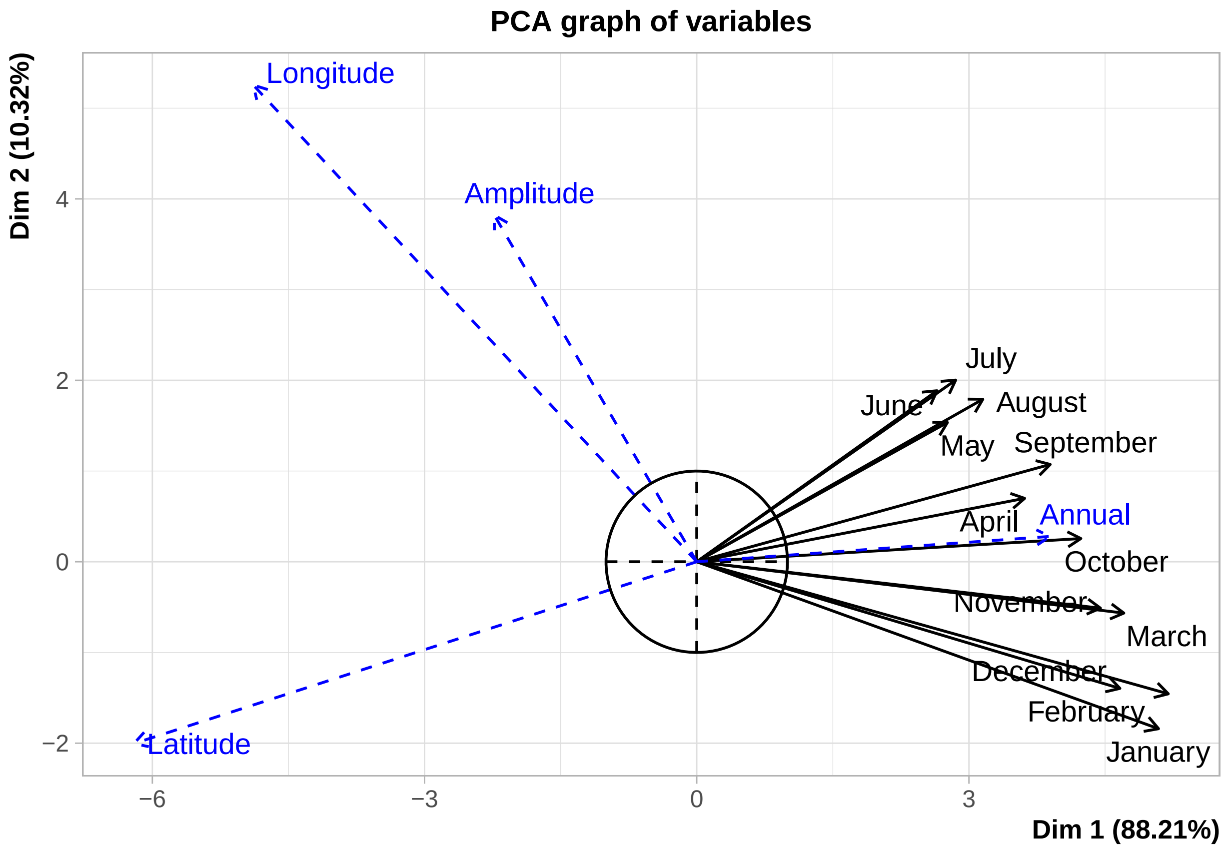
```
## Warning: le package 'FactoMineR' a été compilé avec la version R 4.3.2
```

```
acp <- PCA(x,
           quanti.sup=c(13:16),
           quali.sup=17, #17e colonne est qualitative
           scale.unit = FALSE, # ACP simple
           ncp=Inf)
```

```
## Warning: ggrepel: 1 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



PCA graph of individuals
Dim 2 (10.32%)
Dim 1 (88.21%)
10
5
0
−5
−10
−15
−20
−10
0
10
20
30
Moscow
Kiev
Budapest
St. Petersburg
Minsk
East
Prague
Athens
Milan
Krakow
Sofia
Seville
Rome
South
Elsinki
Oslo
Madrid
Genoa
Berlin
Geneva
Frankfurt
West
Stockholm
Zurich
Bordeaux
Barcelona
Palermo
Copenhagen
Antwerp
Paris
Lisbon
North
Amsterdam
Brussels
Edinburgh
London
Dublin
Reykjavik

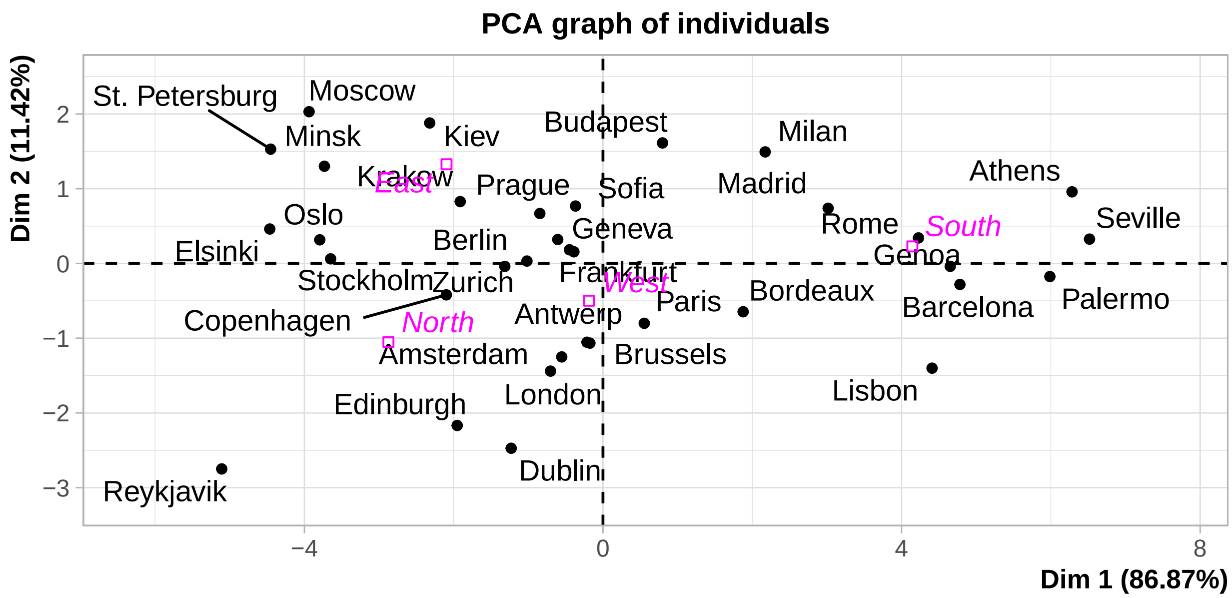


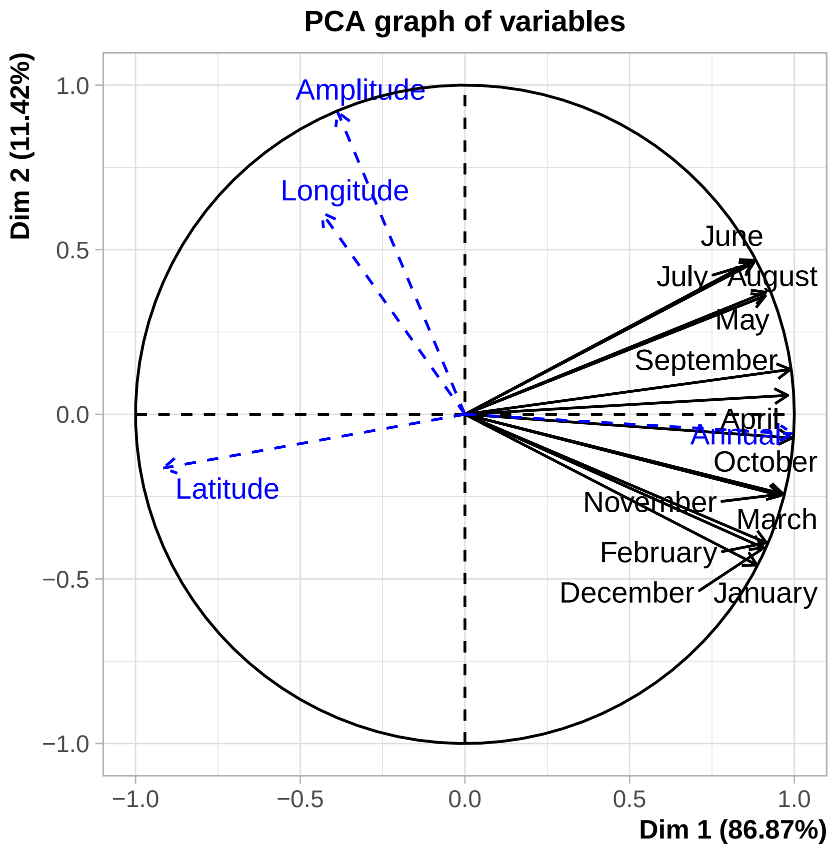
Après avoir effectué une ACP simple, on remarque qu'on ne peut pas en tirer une conclusion. De ce fait, effectuons une ACP standard.

```r
library(FactoMineR)

acp <- PCA(x,
           quanti.sup=c(13:16),
           quali.sup=17, #17e colonne est qualitative
           scale.unit = TRUE, # ACP standard
           ncp=Inf)
```

```
## Warning: ggrepel: 1 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```


PCA graph of individuals
Dim 2 (11.42%)
Dim 1 (86.87%)
2
1
0
−1
−2
−3
−4
0
4
8
St. Petersburg
Moscow
Minsk
Kiev
Budapest
Milan
Krakow
East
Prague
Sofia
Madrid
Athens
Oslo
Rome
South
Seville
Berlin
Geneva
Elsinki
Genoa
Frankfurt
West
Stockholm
Zurich
Paris
Bordeaux
Barcelona
Palermo
Copenhagen
North
Antwerp
Amsterdam
Brussels
Edinburgh
London
Lisbon
Dublin
Reykjavik




```r
acp$eig
```

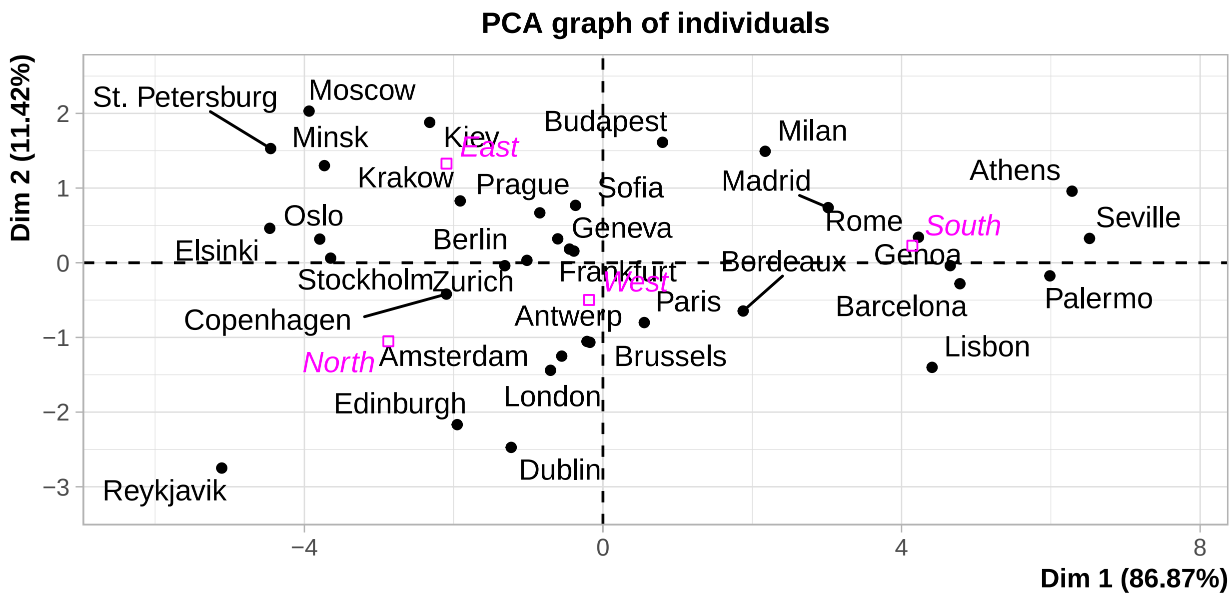
```
##         eigenvalue percentage of variance cumulative percentage of variance
## comp 1  1.042445e+01          86.870441346                          86.87044
## comp 2  1.370499e+00          11.420823117                          98.29126
## comp 3  1.205076e-01           1.004230241                          99.29549
## comp 4  4.233298e-02           0.352774838                          99.64827
## comp 5  2.292280e-02           0.191023370                          99.83929
## comp 6  8.684234e-03           0.072368614                          99.91166
## comp 7  4.178064e-03           0.034817200                          99.94648
## comp 8  2.930325e-03           0.024419371                          99.97090
## comp 9  1.475750e-03           0.012297915                          99.98320
## comp 10 8.529732e-04           0.007108110                          99.99030
## comp 11 7.862929e-04           0.006552441                          99.99686
## comp 12 3.772122e-04           0.003143435                         100.00000
```

Les 2 premiers axes conservent 95% de la variance et les 3 premiers axes conservent 98% de la variance. Il est donc raisonnable de prendre dans un premier temps les 2 premiers axes.

La projection des individus sur le plan principal :

```r
plot(acp, choix="ind", axes=c(1,2))
```

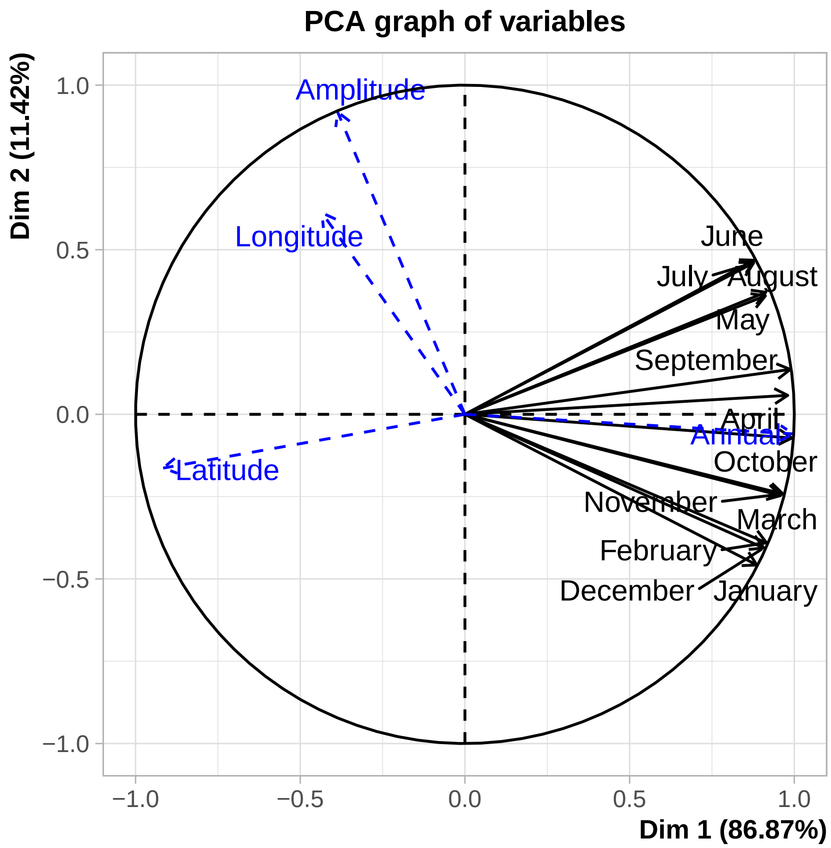
```
## Warning: ggrepel: 1 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



En général, les villes situées au Nord et à l'Est se situent à gauche du graphe alors que les villes du Sud se situent droite. En revanche, les villes de l'Ouest se situent au centre. Ainsi plus les températures sont froides, plus on se trouve à gauche et inversement.

Le cercle de corrélation :

```
plot(acp, choix="var", axes=c(1,2))
```



Toutes les variables sont proches du cercle de rayon 1 et sont donc bien représentées par les variables principales. On voit que toutes les variables sont corrélées positivement avec le premier axe en particulier le mois d'octobre et d'avril qui sont fortement corrélés positivement avec le 1er axe, mais toutes les variables sont peu corrélées avec le deuxième axe. Néanmoins, les mois d'octobre à avril ont une corrélation négative avec le 2e axe tandis que les mois de mai à septembre ont une corrélation positive avec ce dernier. Ainsi, le 1er axe distingue les mois où la moyenne des températures est proche de la moyenne annuelle. Le 2e axe quant à lui, est spécifique des mois où les températures moyennes sont élevées.

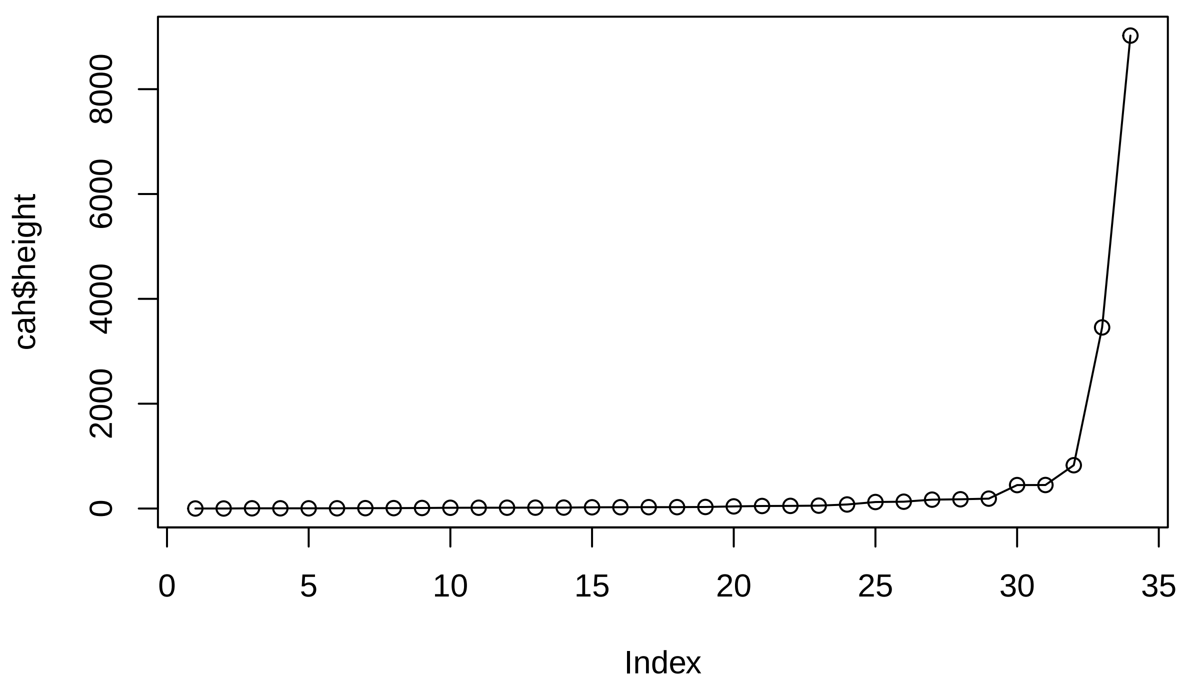
Question 7

Appliquons l'algorithme de Classification Ascendante Hiérarchique (CAH) de Ward aux projections des points sur le plan principal :

```r
d = dist(x[,1:12], method='euclidian')
cah <- hclust(d=d^2, method="ward.D")
cah$height
```

```
##  [1]    0.820000    0.870000    4.010000    4.270000    4.436667    4.490000
##  [7]    7.296667    7.910000   10.118333   14.230000   15.240000   15.670000
## [13]   16.783333   17.977333   23.611714   25.226667   26.095000   26.703333
## [19]   29.540000   40.810000   48.476667   50.380000   54.996667   77.694286
## [25]  124.266667  130.318571  169.305000  176.660476  189.103333  447.472222
## [31]  448.644762  826.068509 3454.783793 9021.827429
```

```r
plot(cah$height, type = 'o')
```



On observe qu'à partir de 3 classes le gain de variance intra (c'est-à-dire la perte de variance inter) est stable (il y a un coude). Ainsi, il est raisonnable de retenir 3 classes.

Question 8

Présentation du dendrogramme avec les trois classes

```r
c3 <- cutree(tree=cah, k=3)
(J1 <- x[c3==1,1:12])
```

```
##              January February March April  May June July August September
## Amsterdam        2.9      2.5   5.7   8.2 12.5 14.8 17.1   17.1      14.5
## Berlin          -0.2      0.1   4.4   8.2 13.8 16.0 18.3   18.0      14.4
## Brussels         3.3      3.3   6.7   8.9 12.8 15.6 17.8   17.8      15.0
## Budapest        -1.1      0.8   5.5  11.6 17.0 20.2 22.0   21.3      16.9
## Copenhagen      -0.4     -0.4   1.3   5.8 11.1 15.4 17.1   16.6      13.3
## Dublin           4.8      5.0   5.9   7.8 10.4 13.3 15.0   14.6      12.7
## London           3.4      4.2   5.5   8.3 11.9 15.1 16.9   16.5      14.0
## Madrid           5.0      6.6   9.4  12.2 16.0 20.8 24.7   24.3      19.8
## Paris            3.7      3.7   7.3   9.7 13.7 16.5 19.0   18.7      16.1
## Prague          -1.3      0.2   3.6   8.8 14.3 17.6 19.3   18.7      14.9
```

```
## Sarajevo       -1.4       0.8   4.9   9.3 13.8 17.0 18.9   18.7      15.2
## Sofia          -1.7       0.2   4.3   9.7 14.3 17.7 20.0   19.5      15.8
## Antwerp         3.1       2.9   6.2   8.9 12.9 15.5 17.9   17.6      14.7
## Bordeaux        5.6       6.7   9.0  11.9 15.0 18.3 20.4   20.0      17.6
## Edinburgh       2.9       3.6   4.7   7.1  9.9 13.0 14.7   14.3      12.1
## Frankfurt       0.2       1.8   5.4   9.7 14.3 17.5 19.0   18.3      14.8
## Geneva          0.1       1.9   5.1   9.4 13.8 17.3 19.4   18.5      15.0
## Milan           1.1       3.6   8.0  12.6 17.3 21.3 23.8   22.8      18.9
## Zurich         -0.7       0.7   4.3   8.5 12.9 16.2 18.0   17.2      14.1
##             October November December
## Amsterdam      11.4      7.0      4.4
## Berlin         10.0      4.2      1.2
## Brussels       11.1      6.7      4.4
## Budapest       11.3      5.1      0.7
## Copenhagen      8.8      4.1      1.3
## Dublin          9.7      6.7      5.4
## London         10.2      6.3      4.4
## Madrid         13.9      8.7      5.4
## Paris          12.5      7.3      5.2
## Prague          9.4      3.8      0.3
## Sarajevo       10.5      5.1      0.8
## Sofia          10.7      5.0      0.6
## Antwerp        11.5      6.8      4.7
## Bordeaux       13.5      8.5      6.1
## Edinburgh       8.7      5.3      3.7
## Frankfurt       9.8      4.9      1.7
## Geneva          9.8      4.9      1.4
## Milan          13.1      6.9      2.6
## Zurich          8.9      3.9      0.3
```

```r
nrow(J1)
```

```
## [1] 19
```

```r
(J2 <- x[c3==2,1:12])
```

```
##            January February March April  May June July August September October
## Athens         9.1      9.7  11.7  15.4 20.1 24.5 27.4   27.2      23.8    19.2
## Lisbon        10.5     11.3  12.8  14.5 16.7 19.4 21.5   21.9      20.4    17.4
## Rome           7.1      8.2  10.5  13.7 17.8 21.7 24.4   24.1      20.9    16.5
## Barcelona      9.1     10.3  11.8  14.1 17.4 21.2 24.2   24.1      21.7    17.5
## Genoa          8.7      8.7  11.4  13.8 17.5 21.0 24.5   24.6      21.8    17.8
## Palermo       10.5     11.5  13.3  16.9 20.9 23.8 24.5   22.3      22.3    18.4
## Seville       10.7     11.8  14.1  16.1 19.7 23.4 26.7   26.7      24.3    19.4
##            November December
## Athens         14.6     11.0
## Lisbon         13.7     11.1
## Rome           11.7      8.3
## Barcelona      13.1     10.0
## Genoa          12.2     10.0
## Palermo        14.9     12.0
## Seville        14.5     11.2
```

```r
nrow(J2)
```

```
## [1] 7
```

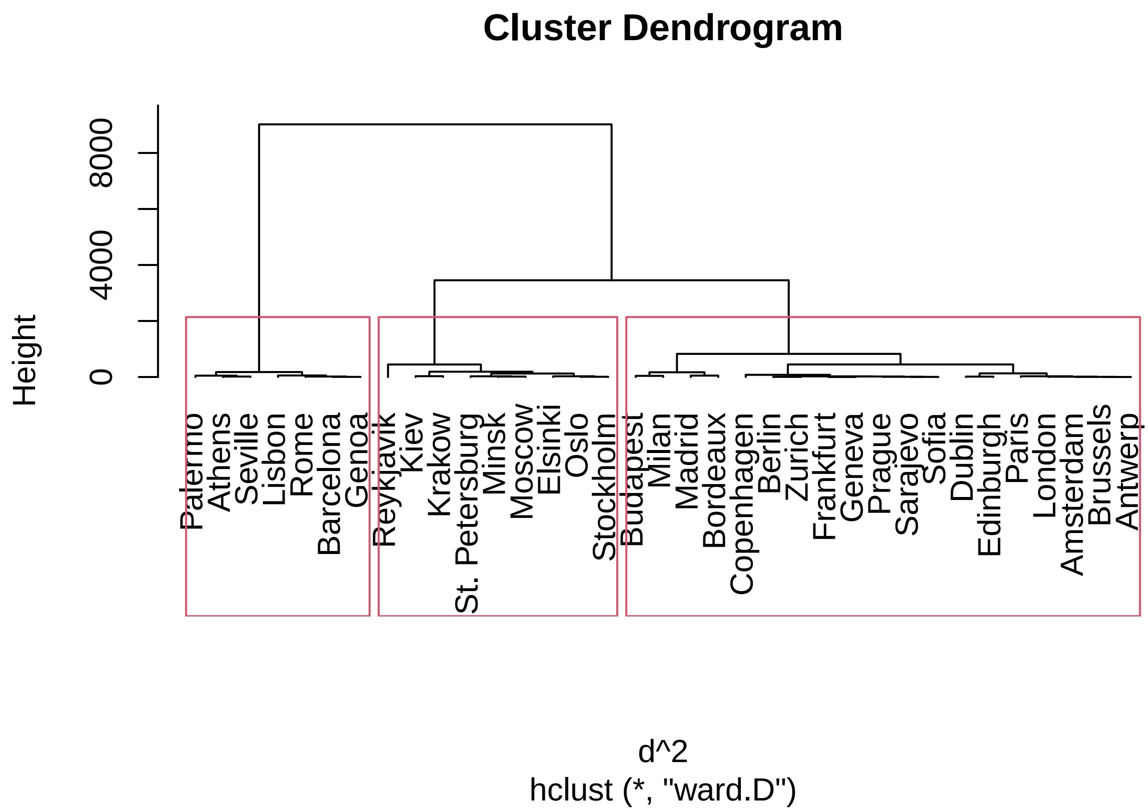
```r
(J3 <- x[c3==3,1:12])
```

```
##                 January February March April  May June July August September
## Elsinki            -5.8     -6.2  -2.7   3.1 10.2 14.0 17.2   14.9       9.7
## Kiev               -5.9     -5.0  -0.3   7.4 14.3 17.8 19.4   18.5      13.7
## Krakow             -3.7     -2.0   1.9   7.9 13.2 16.9 18.4   17.6      13.7
## Minsk              -6.9     -6.2  -1.9   5.4 12.4 15.9 17.4   16.3      11.6
## Moscow             -9.3     -7.6  -2.0   6.0 13.0 16.6 18.3   16.7      11.2
## Oslo               -4.3     -3.8  -0.6   4.4 10.3 14.9 16.9   15.4      11.1
## Reykjavik          -0.3      0.1   0.8   2.9  6.5  9.3 11.1   10.6       7.9
## Stockholm          -3.5     -3.5  -1.3   3.5  9.2 14.6 17.2   16.0      11.7
## St. Petersburg     -8.2     -7.9  -3.7   3.2 10.0 15.4 18.4   16.9      11.5
##                 October November December
## Elsinki             5.2      0.1     -2.3
## Kiev                7.5      1.2     -3.6
## Krakow              8.6      2.6     -1.7
## Minsk               5.8      0.1     -4.2
## Moscow              5.1     -1.1     -6.0
## Oslo                5.7      0.5     -2.9
## Reykjavik           4.5      1.7      0.2
## Stockholm           6.5      1.7     -1.6
## St. Petersburg      5.2     -0.4     -5.3
```

```r
nrow(J3)
```

```
## [1] 9
```

```r
plot(cah, hang = -1)
rect.hclust(cah, k=3)
```

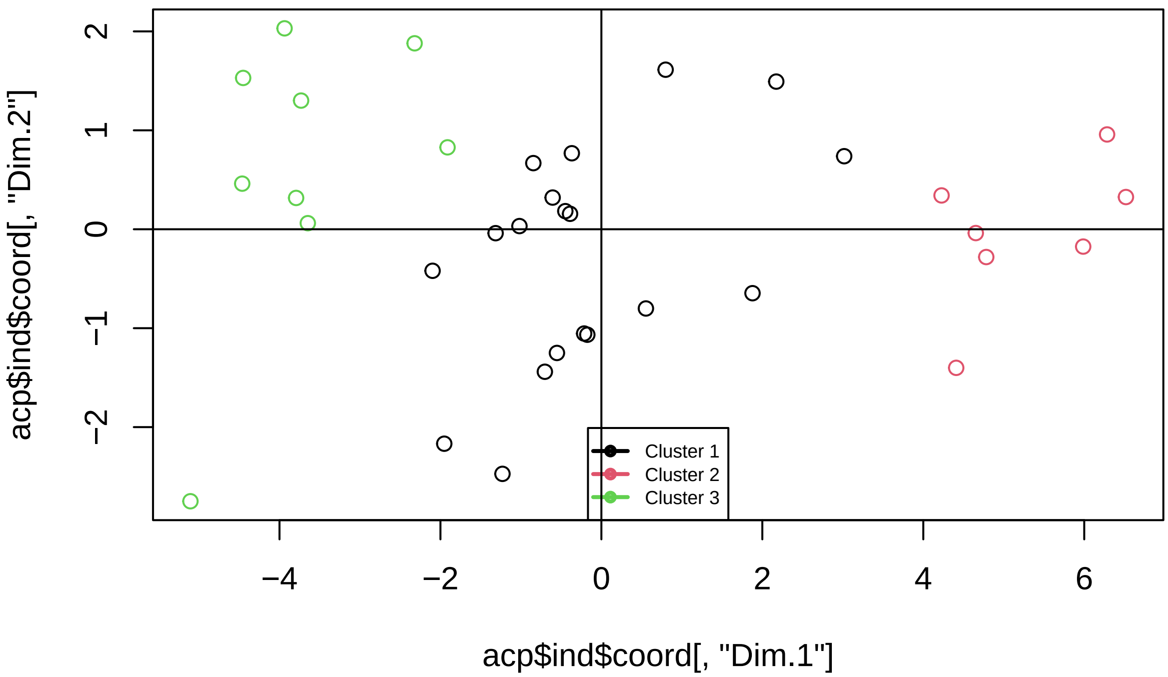


Présentation du nuage de points dans le plan principal

```r
plot(x=acp$ind$coord[,"Dim.1"], y=acp$ind$coord[,"Dim.2"],
     col =cutree(tree = cah, k=3))

legend(x= "bottom",
       col = 1:3, pch = 1,
       legend = c("Cluster 1", "Cluster 2", "Cluster 3"), cex=0.6,lwd=2)

abline(h=0, col = "black")
abline(v=0, col = "black")
```



Calculs des variances inter et intra correspondantes à CCAH,3

Variance intra de la classification :

```
v1=sum(apply(J1, 2, var))*nrow(J1)/35
v2=sum(apply(J2, 2, var))*nrow(J2)/35
v3=sum(apply(J3, 2, var))*nrow(J3)/35
v1
```

```
## [1] 28.00505
```

```
v2
```

```
## [1] 5.280952
```

```
v3
```

```
## [1] 13.90921
```

```
var_intra2 <- v1+v2+v3
var_intra2
```

```
## [1] 47.19521
```

Variance inter de la classification :

```
var_inter2 <- var_tot-var_intra2
var_inter2
```

```
## [1] 180.9828
```

La variance intra correspondante à CCAH,3 est plus faible que la variance intra de Cr. Elle est de 47.20 pour CCAH,3 contre 75.20 pour Cr. La variance totale étant constante la variance inter correspondante à CCAH,3 est plus élevé que celle de Cr, 180.98 contre 152.98. Ainsi, la classification CCAH,3 est meilleure que la classification Cr.

Question 9

Présentation du dendrogramme avec les quatre classes

```
d = dist(x[,1:12], method='euclidian')
cah <- hclust(d=d^2, method="ward.D")
sum(cah$height)/(2*35)
```

```
## [1] 221.6587
```

```
plot(cah, hang = -1)


c4 <- cutree(tree=cah, k=4)
(J1 <- x[c4==1,1:12])
```

```
##            January February March April  May June July August September
## Amsterdam      2.9      2.5   5.7   8.2 12.5 14.8 17.1   17.1      14.5
## Berlin        -0.2      0.1   4.4   8.2 13.8 16.0 18.3   18.0      14.4
## Brussels       3.3      3.3   6.7   8.9 12.8 15.6 17.8   17.8      15.0
## Copenhagen    -0.4     -0.4   1.3   5.8 11.1 15.4 17.1   16.6      13.3
## Dublin         4.8      5.0   5.9   7.8 10.4 13.3 15.0   14.6      12.7
## London         3.4      4.2   5.5   8.3 11.9 15.1 16.9   16.5      14.0
## Paris          3.7      3.7   7.3   9.7 13.7 16.5 19.0   18.7      16.1
## Prague        -1.3      0.2   3.6   8.8 14.3 17.6 19.3   18.7      14.9
## Sarajevo      -1.4      0.8   4.9   9.3 13.8 17.0 18.9   18.7      15.2
## Sofia         -1.7      0.2   4.3   9.7 14.3 17.7 20.0   19.5      15.8
## Antwerp        3.1      2.9   6.2   8.9 12.9 15.5 17.9   17.6      14.7
## Edinburgh      2.9      3.6   4.7   7.1  9.9 13.0 14.7   14.3      12.1
## Frankfurt      0.2      1.8   5.4   9.7 14.3 17.5 19.0   18.3      14.8
## Geneva         0.1      1.9   5.1   9.4 13.8 17.3 19.4   18.5      15.0
## Zurich        -0.7      0.7   4.3   8.5 12.9 16.2 18.0   17.2      14.1
##            October November December
## Amsterdam     11.4      7.0      4.4
## Berlin        10.0      4.2      1.2
## Brussels      11.1      6.7      4.4
## Copenhagen     8.8      4.1      1.3
```

```
## Dublin          9.7       6.7      5.4
## London         10.2       6.3      4.4
## Paris          12.5       7.3      5.2
## Prague          9.4       3.8      0.3
## Sarajevo       10.5       5.1      0.8
## Sofia          10.7       5.0      0.6
## Antwerp        11.5       6.8      4.7
## Edinburgh       8.7       5.3      3.7
## Frankfurt       9.8       4.9      1.7
## Geneva          9.8       4.9      1.4
## Zurich          8.9       3.9      0.3
```

```r
nrow(J1)
```

```
## [1] 15
```

```r
(J2 <- x[c4==2,1:12])
```

```
##             January February March April  May June July August September October
## Athens          9.1      9.7  11.7  15.4 20.1 24.5 27.4   27.2      23.8    19.2
## Lisbon         10.5     11.3  12.8  14.5 16.7 19.4 21.5   21.9      20.4    17.4
## Rome            7.1      8.2  10.5  13.7 17.8 21.7 24.4   24.1      20.9    16.5
## Barcelona       9.1     10.3  11.8  14.1 17.4 21.2 24.2   24.1      21.7    17.5
## Genoa           8.7      8.7  11.4  13.8 17.5 21.0 24.5   24.6      21.8    17.8
## Palermo        10.5     11.5  13.3  16.9 20.9 23.8 24.5   22.3      22.3    18.4
## Seville        10.7     11.8  14.1  16.1 19.7 23.4 26.7   26.7      24.3    19.4
##             November December
## Athens          14.6     11.0
## Lisbon          13.7     11.1
## Rome            11.7      8.3
## Barcelona       13.1     10.0
## Genoa           12.2     10.0
## Palermo         14.9     12.0
## Seville         14.5     11.2
```

```r
nrow(J2)
```

```
## [1] 7
```

```r
(J3 <- x[c4==3,1:12])
```

```
##            January February March April  May June July August September October
## Budapest      -1.1      0.8   5.5  11.6 17.0 20.2 22.0   21.3      16.9    11.3
## Madrid         5.0      6.6   9.4  12.2 16.0 20.8 24.7   24.3      19.8    13.9
## Bordeaux       5.6      6.7   9.0  11.9 15.0 18.3 20.4   20.0      17.6    13.5
## Milan          1.1      3.6   8.0  12.6 17.3 21.3 23.8   22.8      18.9    13.1
##            November December
## Budapest       5.1      0.7
## Madrid         8.7      5.4
## Bordeaux       8.5      6.1
## Milan          6.9      2.6
```

```r
nrow(J3)
```

```
## [1] 4
```

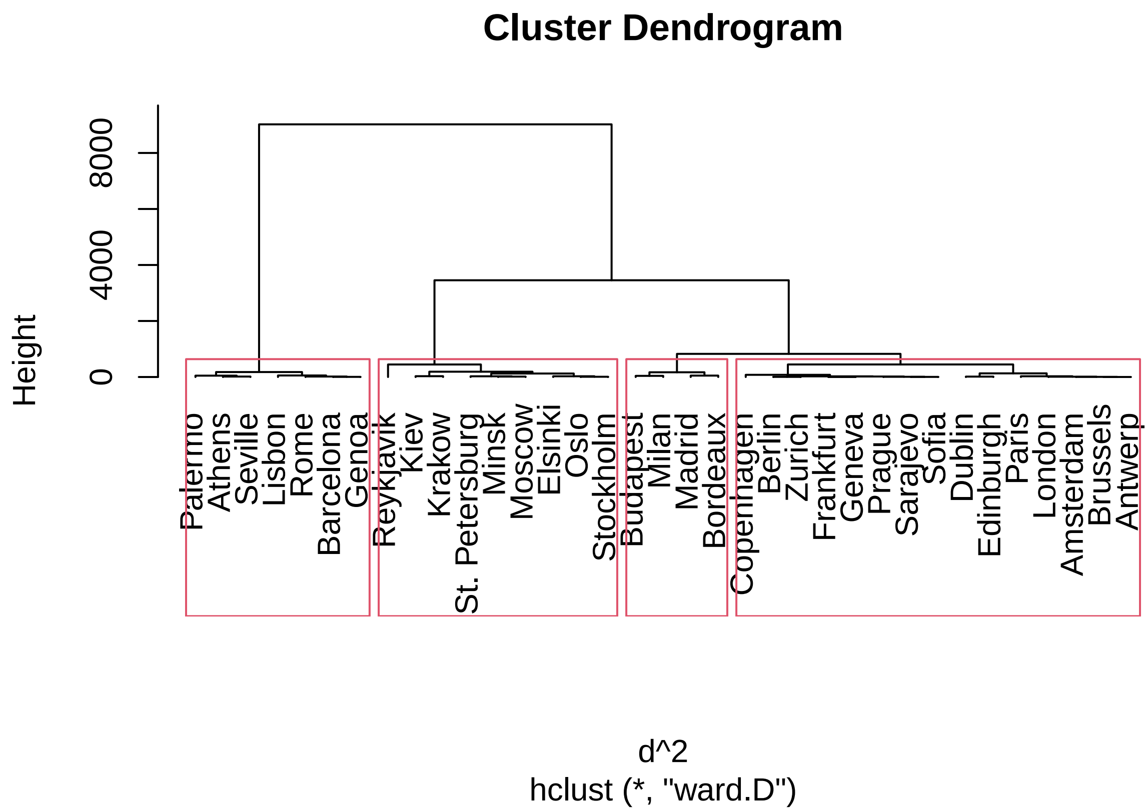
```r
(J4 <- x[c4==4,1:12])
```

```
##                 January February March April  May June July August September
## Elsinki            -5.8     -6.2  -2.7   3.1 10.2 14.0 17.2   14.9       9.7
## Kiev               -5.9     -5.0  -0.3   7.4 14.3 17.8 19.4   18.5      13.7
## Krakow             -3.7     -2.0   1.9   7.9 13.2 16.9 18.4   17.6      13.7
## Minsk              -6.9     -6.2  -1.9   5.4 12.4 15.9 17.4   16.3      11.6
## Moscow             -9.3     -7.6  -2.0   6.0 13.0 16.6 18.3   16.7      11.2
## Oslo               -4.3     -3.8  -0.6   4.4 10.3 14.9 16.9   15.4      11.1
## Reykjavik          -0.3      0.1   0.8   2.9  6.5  9.3 11.1   10.6       7.9
## Stockholm          -3.5     -3.5  -1.3   3.5  9.2 14.6 17.2   16.0      11.7
## St. Petersburg     -8.2     -7.9  -3.7   3.2 10.0 15.4 18.4   16.9      11.5
##                 October November December
## Elsinki             5.2      0.1     -2.3
## Kiev                7.5      1.2     -3.6
## Krakow              8.6      2.6     -1.7
## Minsk               5.8      0.1     -4.2
## Moscow              5.1     -1.1     -6.0
## Oslo                5.7      0.5     -2.9
## Reykjavik           4.5      1.7      0.2
## Stockholm           6.5      1.7     -1.6
## St. Petersburg      5.2     -0.4     -5.3
```

```r
nrow(J4)
```

```
## [1] 9
```

```r
plot(cah, hang = -1)
rect.hclust(cah, k=4)
```

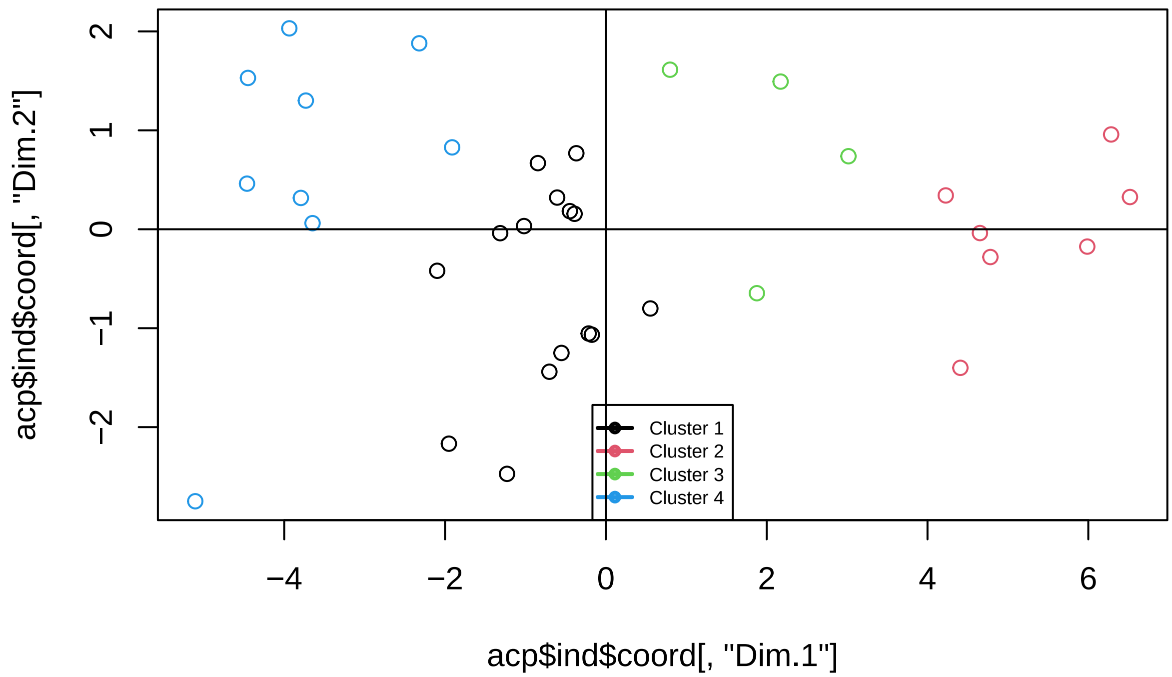


Présentation du nuage de points dans le plan principal

```
plot(x=acp$ind$coord[,"Dim.1"], y=acp$ind$coord[,"Dim.2"],
     col =cutree(tree = cah, k=4))

legend(x= "bottom",
       col = 1:4, pch = 1,
       legend = c("Cluster 1", "Cluster 2", "Cluster 3", "Cluster 4"), cex=0.6,lwd=2)

abline(h=0, col = "black")
abline(v=0, col = "black")
```



Calculs des variances inter et intra correspondantes à CCAH,4

Variance intra de la classification :

```
v1=sum(apply(J1, 2, var))*nrow(J1)/35
v2=sum(apply(J2, 2, var))*nrow(J2)/35
v3=sum(apply(J3, 2, var))*nrow(J3)/35
v4=sum(apply(J4, 2, var))*nrow(J4)/35

v1
```

```
## [1] 11.7951
```

```
v2
```

```
## [1] 5.280952
```

```
v3
```

```
## [1] 4.96181
```

```r
v4
```

```
## [1] 13.90921
```

```r
var_intra3 <- v1+v2+v3+v4
var_intra3
```

```
## [1] 35.94708
```

Variance inter de la classification :

```r
var_inter3 <- var_tot-var_intra3
var_inter3
```

```
## [1] 192.231
```

La variance intra correspondante à CCAH,4 est toujours plus faible que la variance intra de Cr. Elle est de 35,95 pour CCAH,3 contre 75.20 pour Cr. De plus, la variance inter correspondantes à CCAH,4 est plus élevé que celle de Cr, 192.23 contre 152.98. La classification CCAH,4 est donc meilleure que la classification Cr. Ainsi, la classification CCAH,4 est pour l'instant la meilleure de toutes les classifications confondues.

Question 10

Appliquons l'algorithme des centres mobiles au nuage de points initial pour obtenir une classification Ckmeans,4 avec 4 centres.

```r
km <- kmeans(x=x[,1:12], centers=4, algorithm = "Lloyd") #variance intra / variance CAH + kmeans

# Classe 1
n_1 <- km$siz[1]
n_1 #nombre individu dans J1
```

```
## [1] 14
```

```r
which(km$cluster == 1) # indice des individu dans cluster 1
```

```
##   Amsterdam      Berlin     Brussels Copenhagen      Dublin      London
##           1           3           4           6           7          12
##      Prague    Sarajevo       Sofia     Antwerp   Edinburgh   Frankfurt
##          18          21          22          24          27          28
##      Geneva      Zurich
##          29          35
```

```r
km$centers[1,] #point moyen de la classe 1
```

```
##   January  February     March     April       May      June      July    August
##  1.071429  1.914286  4.857143  8.471429 12.764286 15.857143 17.814286 17.385714
## September   October  November  December
## 14.321429 10.035714  5.335714  2.471429
```

```r
#somme des carrés des distances entre chaque point de la classe 1
#et le point moyen de la classe:
#res$withinss{1}
```

Variance du sous-nuage donné par la classe 1:

```r
var1 <- km$withinss[1]/n_1
var1
```

```
## [1] 24.68612
```

```r
#calcul alternatif:
X1 <- x[km$cluster == 1,1:12]
X1
```

```
##            January February March April  May June July August September
## Amsterdam      2.9      2.5   5.7   8.2 12.5 14.8 17.1   17.1      14.5
## Berlin        -0.2      0.1   4.4   8.2 13.8 16.0 18.3   18.0      14.4
## Brussels       3.3      3.3   6.7   8.9 12.8 15.6 17.8   17.8      15.0
## Copenhagen    -0.4     -0.4   1.3   5.8 11.1 15.4 17.1   16.6      13.3
## Dublin         4.8      5.0   5.9   7.8 10.4 13.3 15.0   14.6      12.7
## London         3.4      4.2   5.5   8.3 11.9 15.1 16.9   16.5      14.0
## Prague        -1.3      0.2   3.6   8.8 14.3 17.6 19.3   18.7      14.9
## Sarajevo      -1.4      0.8   4.9   9.3 13.8 17.0 18.9   18.7      15.2
## Sofia         -1.7      0.2   4.3   9.7 14.3 17.7 20.0   19.5      15.8
## Antwerp        3.1      2.9   6.2   8.9 12.9 15.5 17.9   17.6      14.7
## Edinburgh      2.9      3.6   4.7   7.1  9.9 13.0 14.7   14.3      12.1
## Frankfurt      0.2      1.8   5.4   9.7 14.3 17.5 19.0   18.3      14.8
## Geneva         0.1      1.9   5.1   9.4 13.8 17.3 19.4   18.5      15.0
## Zurich        -0.7      0.7   4.3   8.5 12.9 16.2 18.0   17.2      14.1
##            October November December
## Amsterdam     11.4      7.0      4.4
## Berlin        10.0      4.2      1.2
## Brussels      11.1      6.7      4.4
## Copenhagen     8.8      4.1      1.3
## Dublin         9.7      6.7      5.4
## London        10.2      6.3      4.4
## Prague         9.4      3.8      0.3
## Sarajevo      10.5      5.1      0.8
## Sofia         10.7      5.0      0.6
## Antwerp       11.5      6.8      4.7
## Edinburgh      8.7      5.3      3.7
## Frankfurt      9.8      4.9      1.7
## Geneva         9.8      4.9      1.4
## Zurich         8.9      3.9      0.3
```

Sous-nuage des individus de la classe 2:

```r
X2 <- x[km$cluster == 2,1:12]
X2
```

```
##           January February March April  May June July August September October
```

```
## Athens          9.1      9.7  11.7  15.4 20.1 24.5 27.4   27.2      23.8    19.2
## Lisbon         10.5     11.3  12.8  14.5 16.7 19.4 21.5   21.9      20.4    17.4
## Rome            7.1      8.2  10.5  13.7 17.8 21.7 24.4   24.1      20.9    16.5
## Barcelona       9.1     10.3  11.8  14.1 17.4 21.2 24.2   24.1      21.7    17.5
## Genoa           8.7      8.7  11.4  13.8 17.5 21.0 24.5   24.6      21.8    17.8
## Palermo        10.5     11.5  13.3  16.9 20.9 23.8 24.5   22.3      22.3    18.4
## Seville        10.7     11.8  14.1  16.1 19.7 23.4 26.7   26.7      24.3    19.4
##             November December
## Athens          14.6     11.0
## Lisbon          13.7     11.1
## Rome            11.7      8.3
## Barcelona       13.1     10.0
## Genoa           12.2     10.0
## Palermo         14.9     12.0
## Seville         14.5     11.2
```

```
n_2 <- nrow(X2)
var2 <- km$withinss[2]/n_2
```

Sous-nuage des individus de la classe 3:

```
X3 <- x[km$cluster == 3,1:12]
X2
```

```
##             January February March April  May June July August September October
## Athens          9.1      9.7  11.7  15.4 20.1 24.5 27.4   27.2      23.8    19.2
## Lisbon         10.5     11.3  12.8  14.5 16.7 19.4 21.5   21.9      20.4    17.4
## Rome            7.1      8.2  10.5  13.7 17.8 21.7 24.4   24.1      20.9    16.5
## Barcelona       9.1     10.3  11.8  14.1 17.4 21.2 24.2   24.1      21.7    17.5
## Genoa           8.7      8.7  11.4  13.8 17.5 21.0 24.5   24.6      21.8    17.8
## Palermo        10.5     11.5  13.3  16.9 20.9 23.8 24.5   22.3      22.3    18.4
## Seville        10.7     11.8  14.1  16.1 19.7 23.4 26.7   26.7      24.3    19.4
##             November December
## Athens          14.6     11.0
## Lisbon          13.7     11.1
## Rome            11.7      8.3
## Barcelona       13.1     10.0
## Genoa           12.2     10.0
## Palermo         14.9     12.0
## Seville         14.5     11.2
```

```
n_3 <- nrow(X3)
var3 <- km$withinss[3]/n_3
```

Sous-nuage des individus de la classe 4:

```
X4 <- x[km$cluster == 4,1:12]
X4
```

```
##            January February March April  May June July August September October
## Budapest      -1.1      0.8   5.5  11.6 17.0 20.2 22.0   21.3      16.9    11.3
## Madrid         5.0      6.6   9.4  12.2 16.0 20.8 24.7   24.3      19.8    13.9
```

```
## Paris          3.7      3.7   7.3   9.7 13.7 16.5 19.0   18.7      16.1    12.5
## Bordeaux       5.6      6.7   9.0  11.9 15.0 18.3 20.4   20.0      17.6    13.5
## Milan          1.1      3.6   8.0  12.6 17.3 21.3 23.8   22.8      18.9    13.1
##            November December
## Budapest        5.1      0.7
## Madrid          8.7      5.4
## Paris           7.3      5.2
## Bordeaux        8.5      6.1
## Milan           6.9      2.6
```

```
n_4 <- nrow(X4)
var4 <- km$withinss[4]/n_4
```

Calculs des variances inter et intra correspondantes à Ckmeans,4

Calcul de la variance intra :

```
var_intra4 <- n_1/35 * var1 + n_2/35 * var2 + n_3/35 * var3 + n_4/35 * var4
var_intra4
```

```
## [1] 31.87273
```

Calcul de la variance inter :

```
var_inter4 <- var_tot - var_intra4
var_inter4
```

```
## [1] 196.3053
```

La variance intra correspondante à Ckmeans,4 est de 31.99. La variance inter elle, est de 196.19. En conclusion, la classification Ckmeans,4 est la meilleure de toutes les classifications confondues puisqu'elle possède la variance intra la plus faible (mais également la variance inter la plus forte).