Mini Projet Data Mining

DECLARATION DU REPERTOIRE DE TRAVAIL

```
setwd("C:/Desktop/MesCours2021/Cours data mining/ProjetDataMining")

require(readxl)

## Loading required package: readxl
```

CHARGER LE JEUX DE DONNES TEMP.DAT

```
temp <- read.csv("C:/ Desktop/MesCours2021/Cours data mining/ProjetDataMining/temp.
csv", sep = ";", header = TRUE, row.names = 1)


temp2 <- read.csv("C:/Desktop/MesCours2021/Cours data mining/ProjetDataMining/temp.
csv", sep = ";", header = TRUE)


dataM <- read_excel("C:/ Desktop/MesCours2021/Cours data mining/ProjetDataMining/da
taMining.xlsx", sheet = "Feuil1")
```

EXPLORATION DES DONNEES

```
dim(temp) # Afficher la dimmension (nombre de ligne et le nombre de colonne) de la
table du jeu de données


names(temp) # Afficher les noms des variables


str(temp) # Afficher les types des variables afin d'etudier la cohérence des donnée
s


head(temp) # Afficher les 5 premières lignes de mon jeu de donnée


tail(temp) # Afficher les 5 dernières lignes de mon jeu de donnée
```

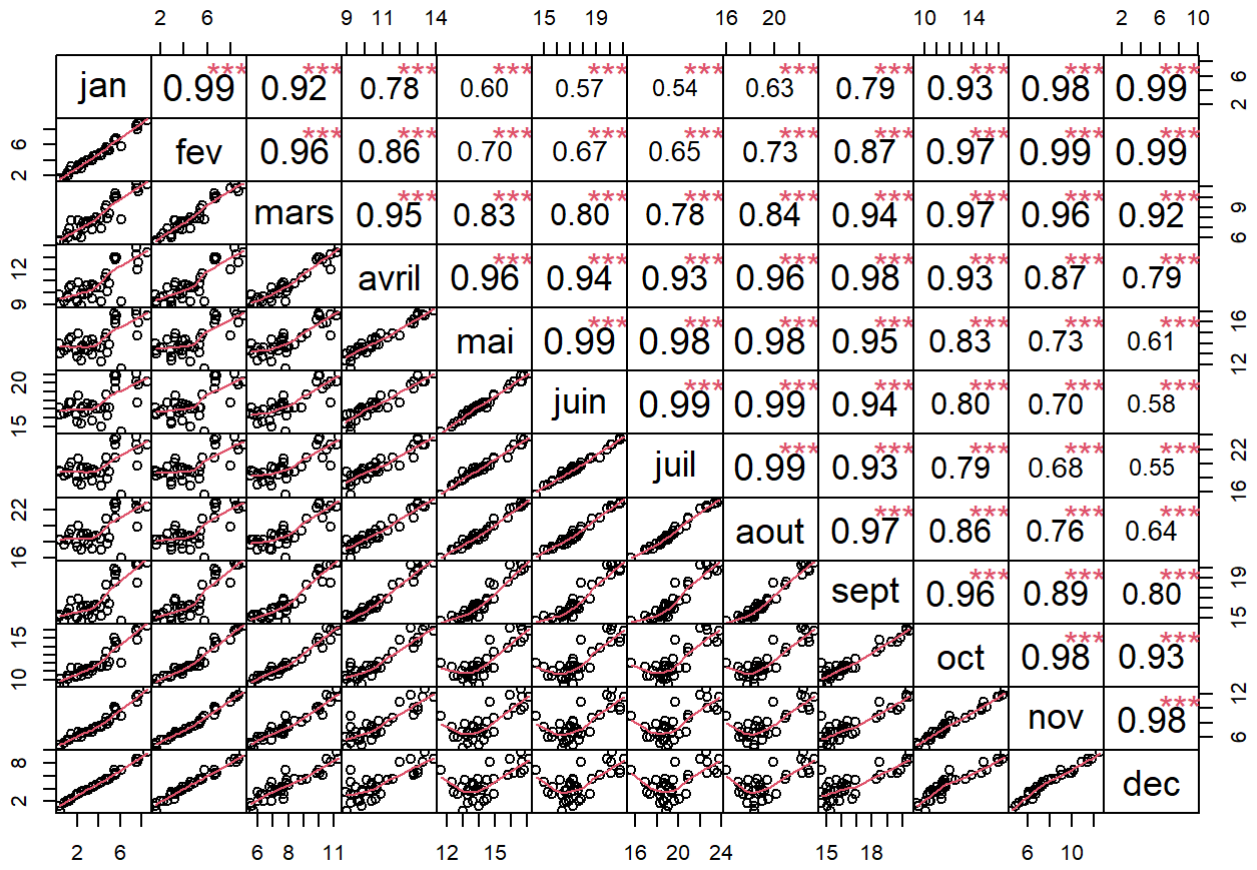
EXPLORATION DES VARIABLES

```
summary(temp)

require(psych)

## Loading required package: psych

require(knitr)

## Loading required package: knitr

expor <- kable(describe(temp, quant = c(.25,.75)))

kable(head(temp))

n <- ncol(temp)
```

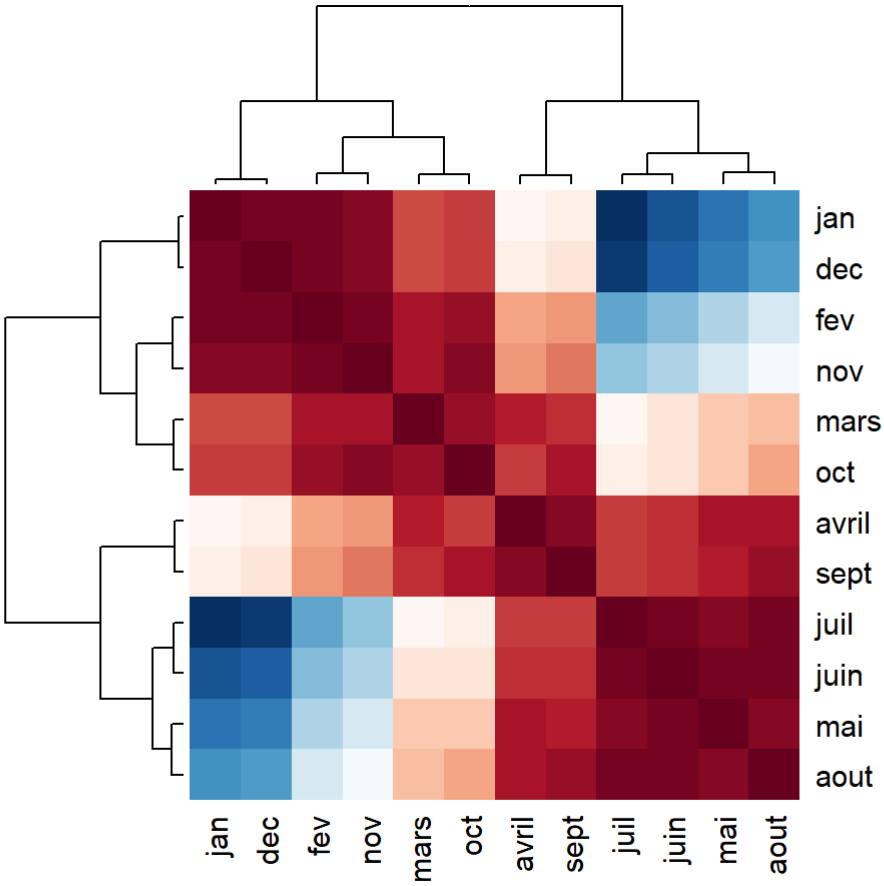
```
for (i in 1:n) {
  x11()
  hist(temp[,i], main = paste("Distribution des températures
  des villes en",names(temp[i])),
       xlab = paste("Les températures en", names(temp[i])), ylab = "Nombre de ville
s", col = "#CCCCFF", cex.axis=1.5,cex.main=2, cex.lab=1.7, font.lab=2,font.axis=2)
  grid()
}
```

ETUDIONS LA CORRELATION

```
library("PerformanceAnalytics")
## Loading required package: xts
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
##
## Attaching package: 'PerformanceAnalytics'
## The following object is masked from 'package:graphics':
##
##     legend
require(corrplot)
## Loading required package: corrplot
## corrplot 0.84 loaded
source("http://www.sthda.com/upload/rquery_cormat.r")
x11()
chart.Correlation(temp, histogram=FALSE, pch=19)
```

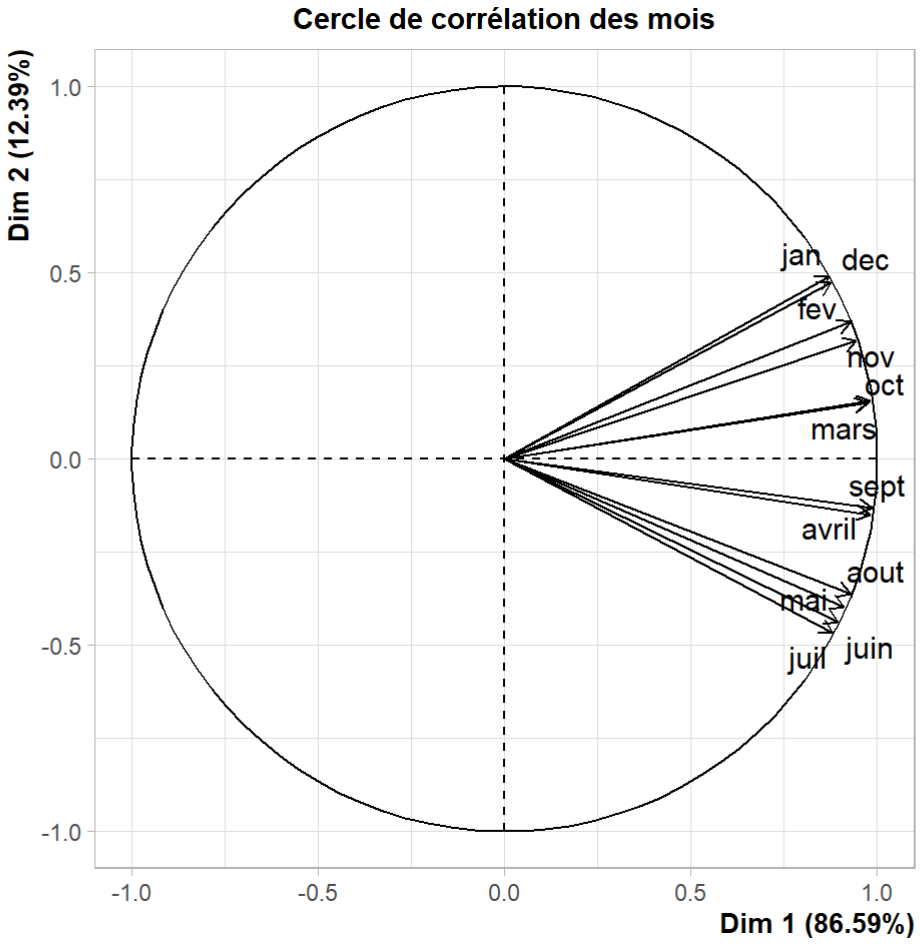



```
x11()
rquery.cormat(temp, graphType="heatmap")
```

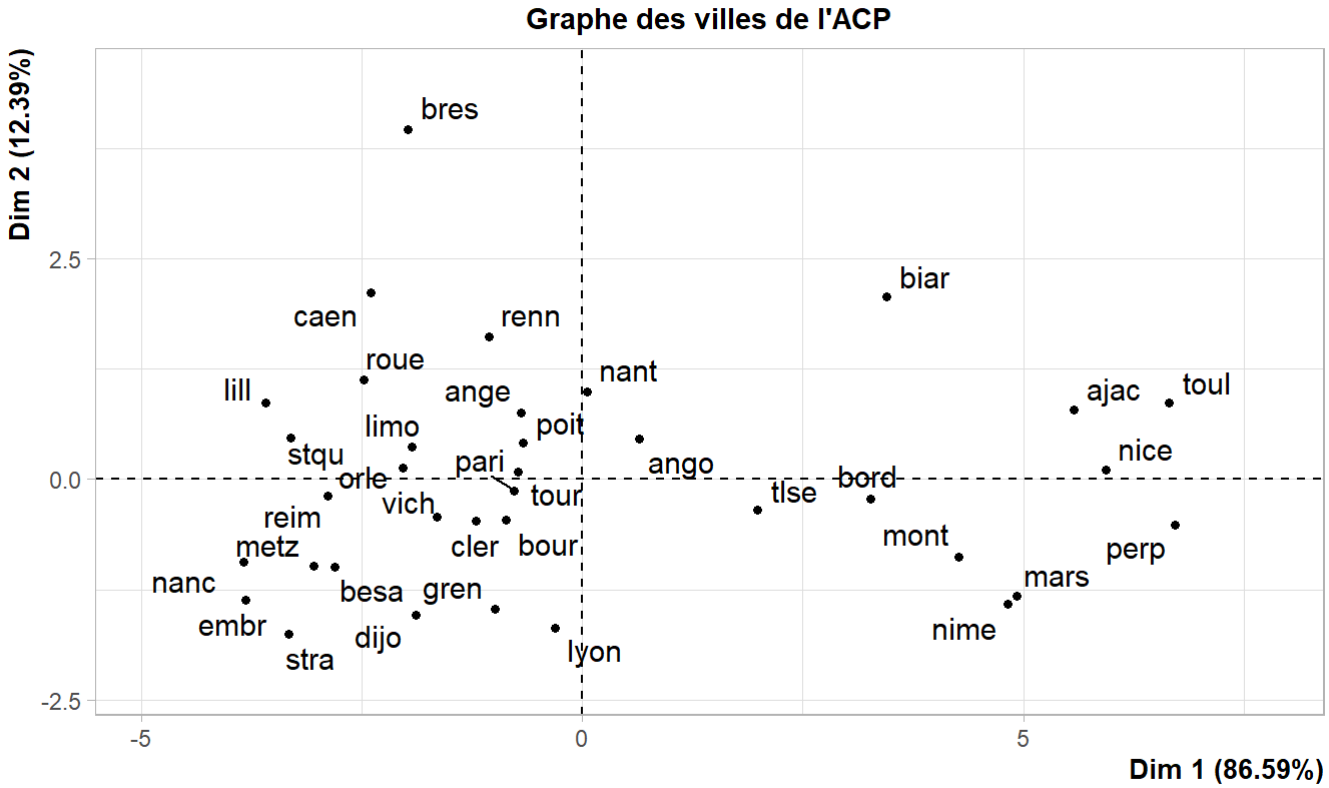


FAIRE ANALYSE DES COMPOSANTS PRINCIPALES

```
require(FactoMineR)
## Loading required package: FactoMineR
res.PCA<-PCA(temp,graph=FALSE)
x11()
plot.PCA(res.PCA,choix='var',title="Cercle de corrélation des mois")
```

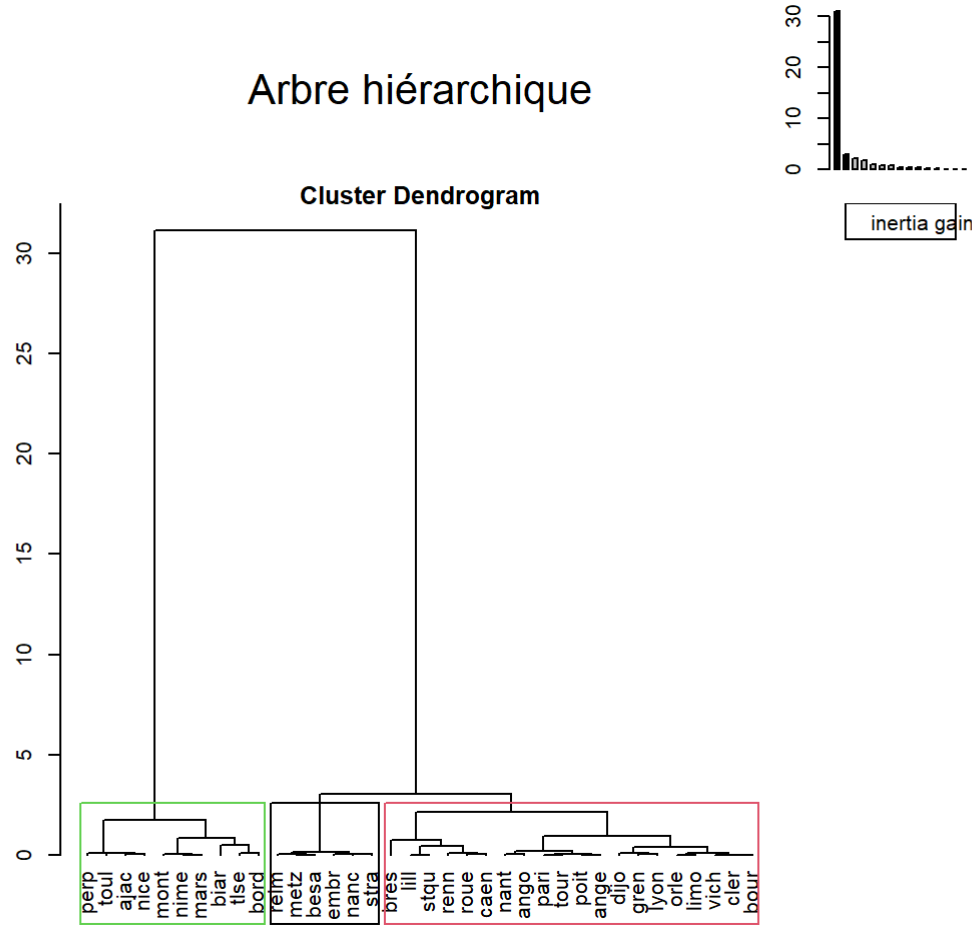



```
x11()
plot.PCA(res.PCA,title="Graphe des villes de l'ACP")
```

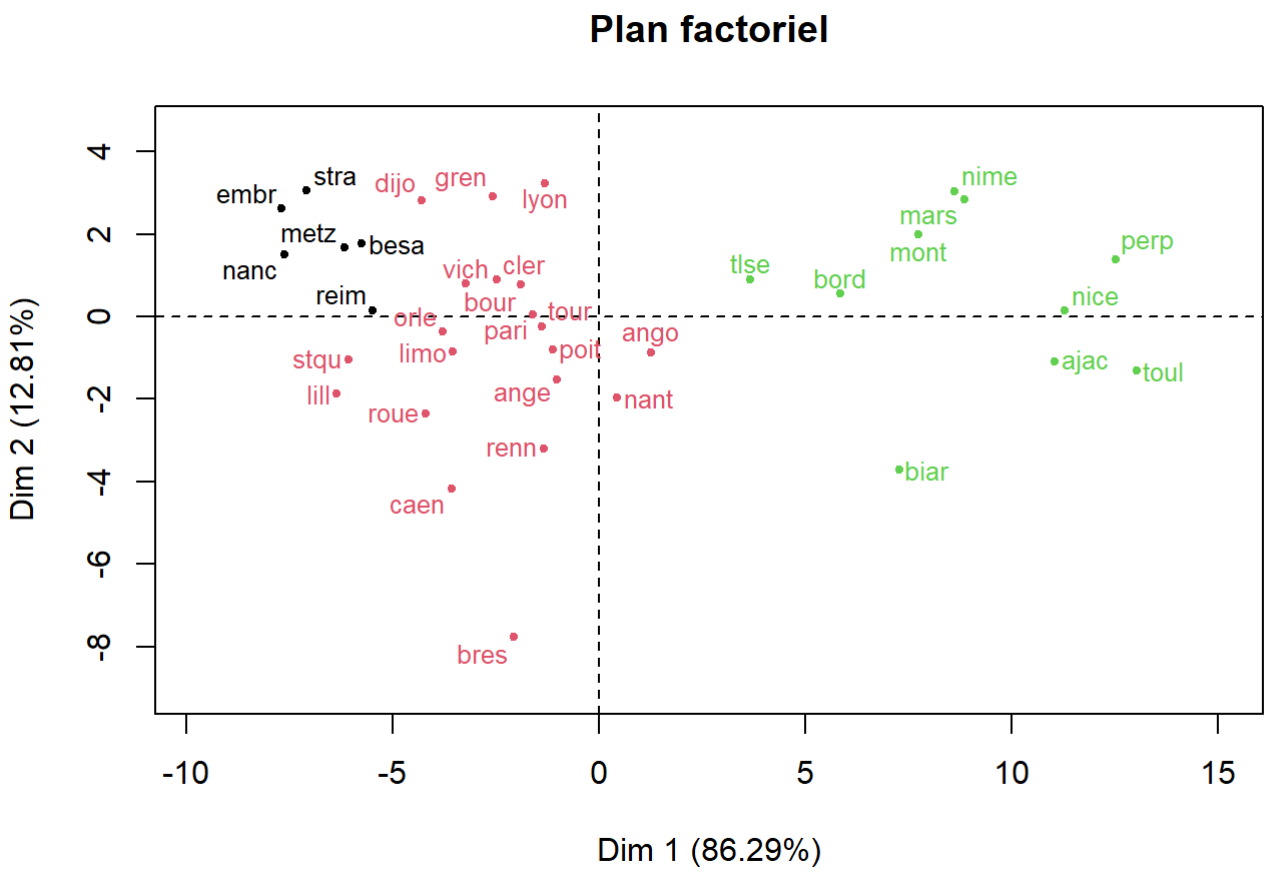


CLASSIFICATION HIERARCHIQUE EN UTILISANT L'ACP

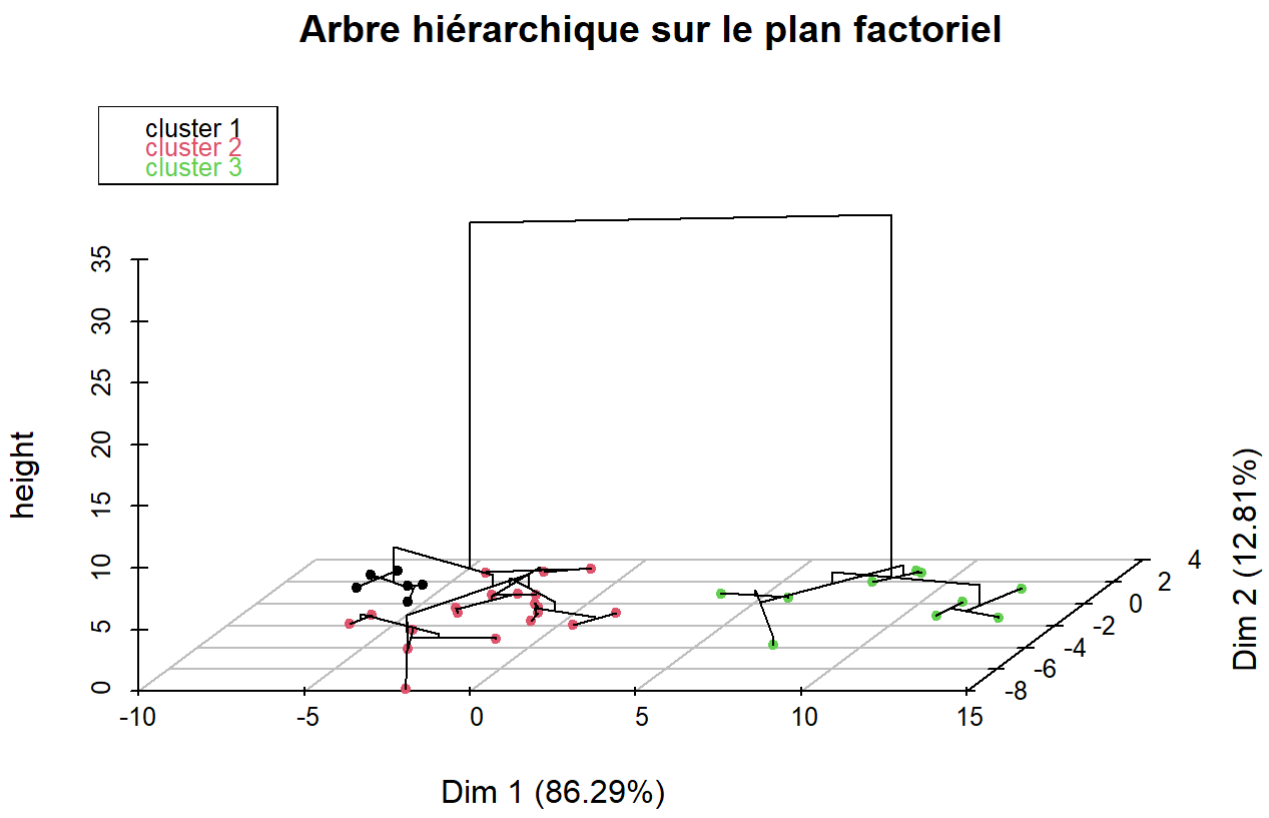
```
res.PCA <- PCA(temp,ncp=Inf, scale.unit=FALSE,graph=FALSE)
res.HCPC <- HCPC(res.PCA,nb.clust=3,consol=FALSE,graph=FALSE)

x11()
plot.HCPC(res.HCPC,choice='tree',title='Arbre hiérarchique')
```



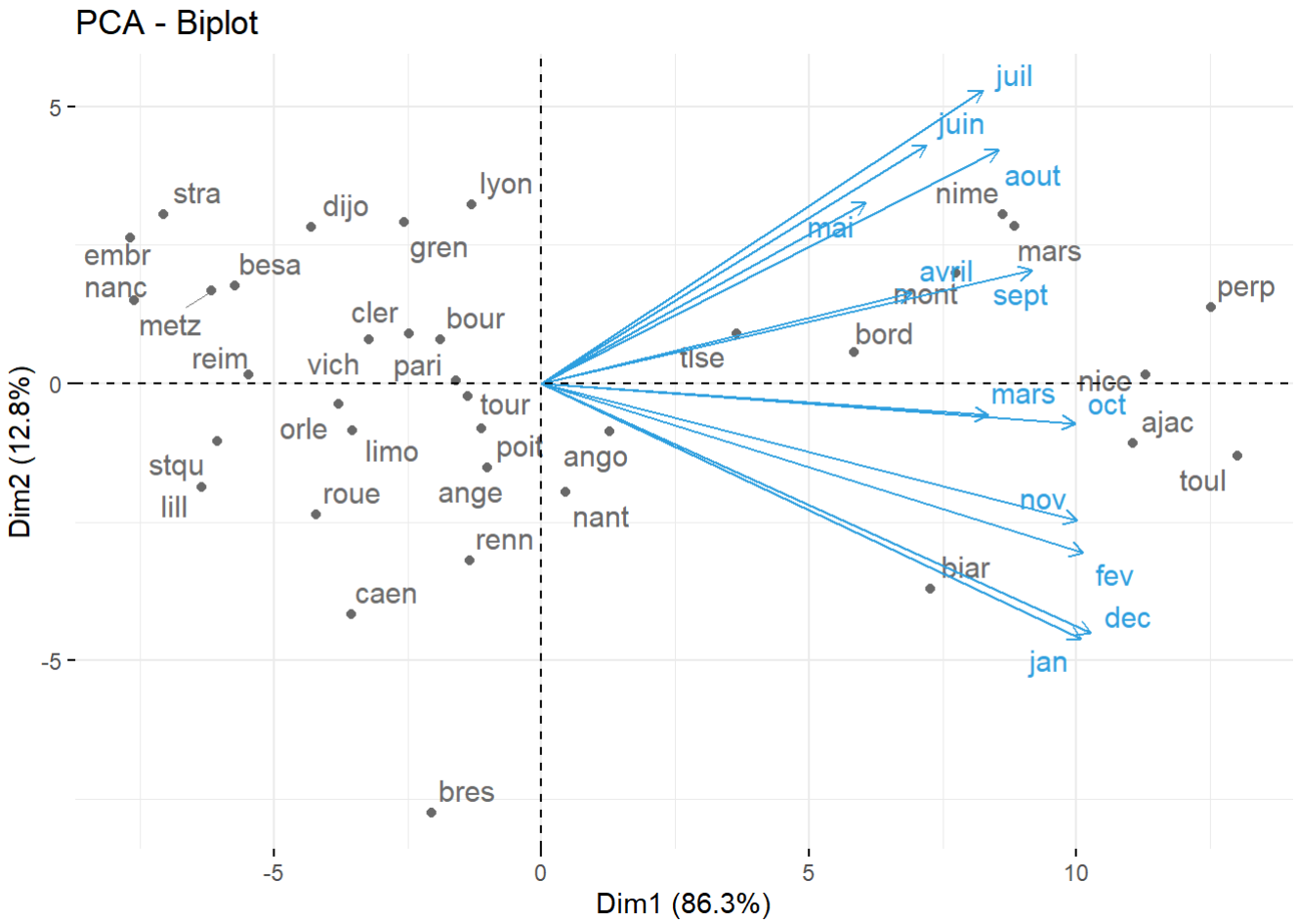

```
x11()
plot.HCPC(res.HCPC,choice='map',draw.tree=FALSE,title='Plan factoriel')
```




```
x11()

plot.HCPC(res.HCPC,choice='3D.map',ind.names=FALSE,centers.plot=FALSE,angle=60,titl
e='Arbre hiérarchique sur le plan factoriel')
```

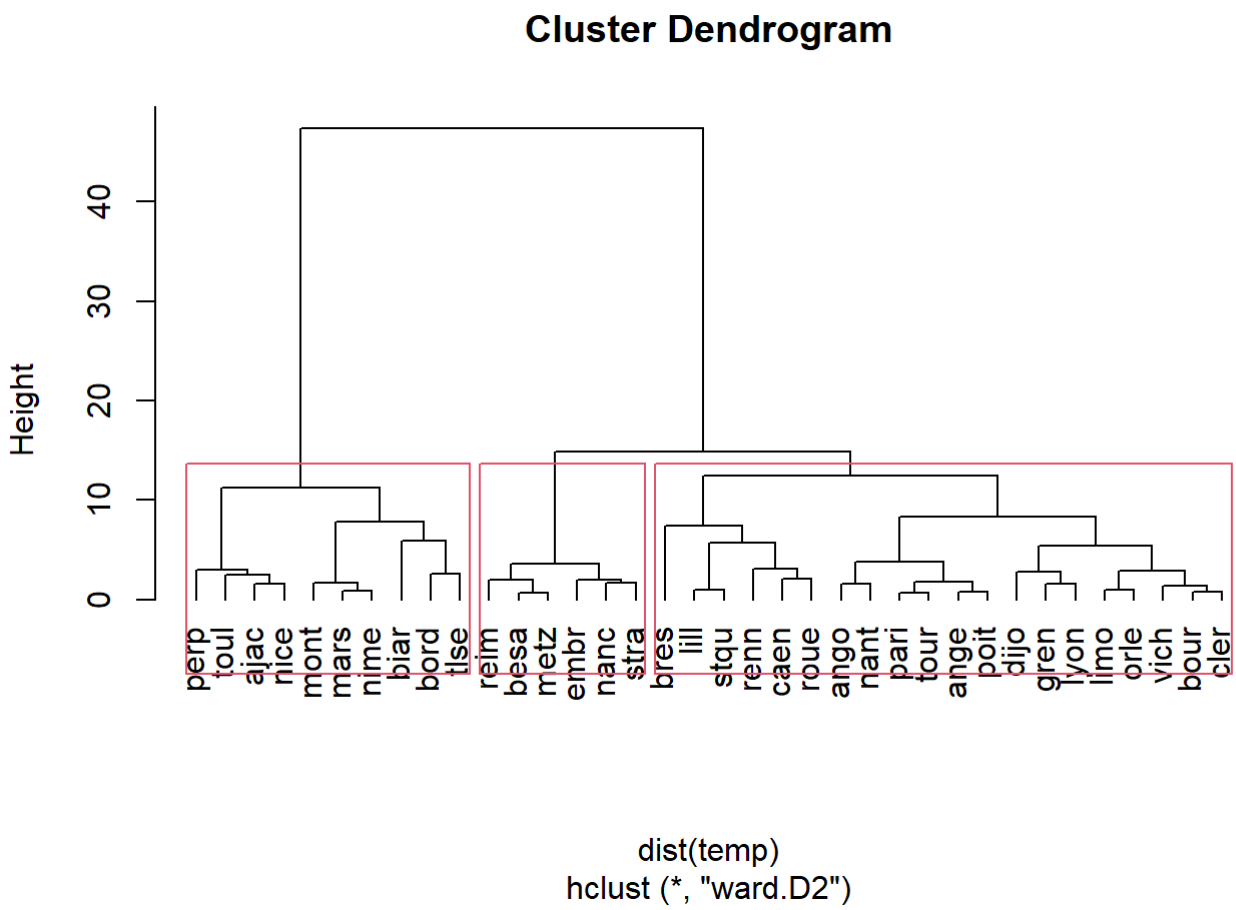



```
#########################FAIRE UN BIPLOT#############################
require(ggplot2)
## Loading required package: ggplot2
##
## Attaching package: 'ggplot2'
## The following objects are masked from 'package:psych':
##
##     %+%, alpha
require(factoextra)
## Loading required package: factoextra
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
x11()
fviz_pca_biplot(res.PCA, repel = TRUE,
                col.var = "#2E9FDF", # Variables color
                col.ind = "#696969"  # Individuals color
                )
```



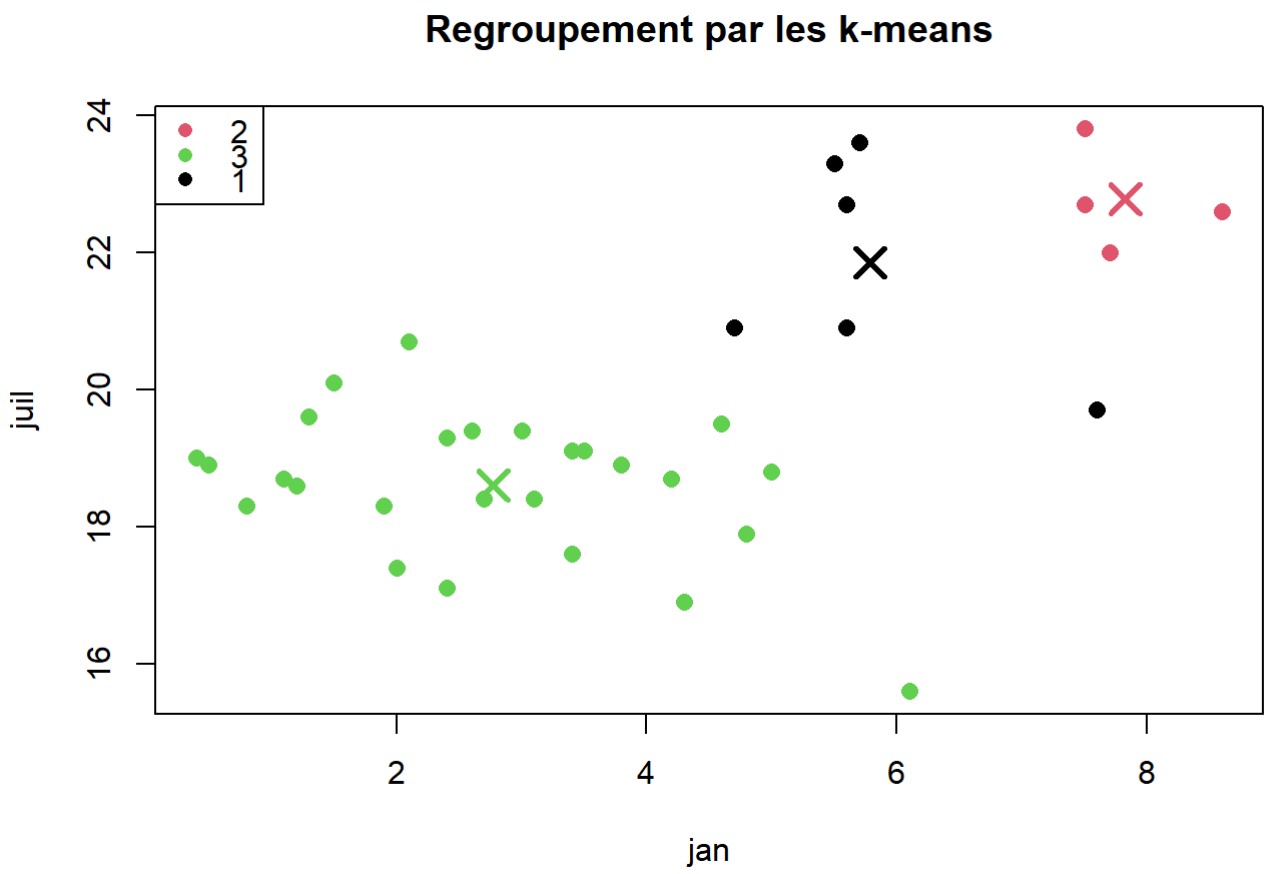
CLASSIFICATION HIERARCHIQUE SANS PASSER PAR L'ACP

```
hc <- hclust(dist(temp),method = "ward.D2")


x11()
plot(hc, hang = -1, labels=temp2$ville)
# cut tree into 3 clusters
rect.hclust(hc, k=3)
```



CLASSIFICATION AVEC LE KMEANS

```
kmeans.res <- kmeans(temp, 3)

summary(kmeans.res) # pour obtenir une description de l'objet ainsi créé


# Afficher les résultats

library(fpc)

library(cluster)


x11()

plot(temp[c("jan","juil")], col = kmeans.res$cluster, pch=16, cex=1.2, main="Regrou
pement par les k-means")


points(kmeans.res$centers[,c("jan","juil")], col = 1:3, pch = 4,cex=2,lwd=3)


legend(x="topleft", legend=unique(kmeans.res$cluster), col=unique(kmeans.res$cluste
r), pch=16)
```




```
library(wordcloud2)

# have a look to the example dataset
head(demoFreq)

# Basic plot

wordcloud2(data=dataM, size=0.5)
```