

# Faire des statistiques de regroupement (clustering) avec R

Michaël Val D.O.F. M.Sc.

04 juin, 2021

## Installation

### Installation de R et RStudio

Lien pour installer R-base

Lien pour installer RStudio

### Installation des paquets supplémentaires

```
install.packages("psy", repos = "https://cloud.r-project.org/")  
install.packages("corrplot", repos = "https://cloud.r-project.org/")
```

## Chargement des données

### Fixer son environnement de travail

```
setwd("~/votredossier")
```

~ indique le répertoire courant, donc /home/votrenom sur Linux /Users/votrenom sur Mac et /c:/Users/votrenom/Documents sur Windows.

### Importer des données

```
mesDonnees <- read.csv2("data.csv")
```

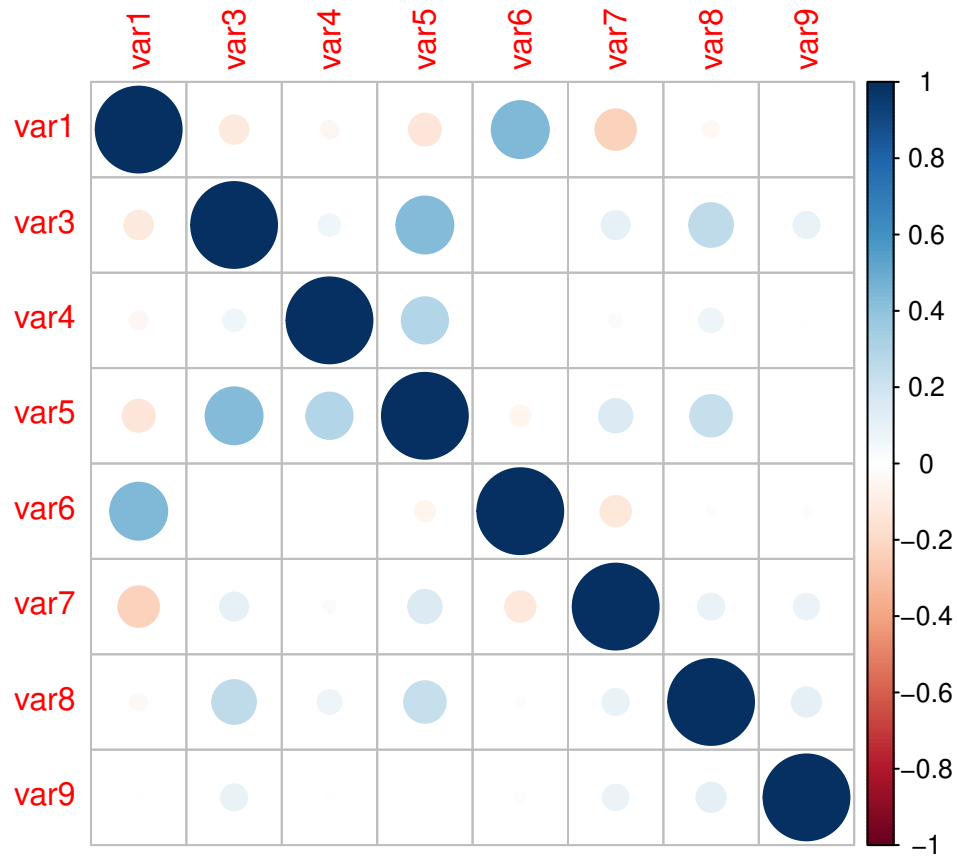
## Matrice de corrélations

```
tab <- c("var1", "var3", "var4", "var5", "var6", "var7", "var8", "var9")  
round(cor(mesDonnees[, tab], use="complete.obs"), digits=3)
```

```
##      var1  var3  var4  var5  var6  var7  var8  var9  
## var1  1.000 -0.110 -0.044 -0.139  0.441 -0.223 -0.038  0.003  
## var3 -0.110  1.000  0.064  0.439  0.002  0.107  0.259  0.093  
## var4 -0.044  0.064  1.000  0.290  0.003  0.021  0.077 -0.009  
## var5 -0.139  0.439  0.290  1.000 -0.055  0.151  0.234  0.001  
## var6  0.441  0.002  0.003 -0.055  1.000 -0.126  0.011  0.015  
## var7 -0.223  0.107  0.021  0.151 -0.126  1.000  0.093  0.088  
## var8 -0.038  0.259  0.077  0.234  0.011  0.093  1.000  0.115  
## var9  0.003  0.093 -0.009  0.001  0.015  0.088  0.115  1.000
```

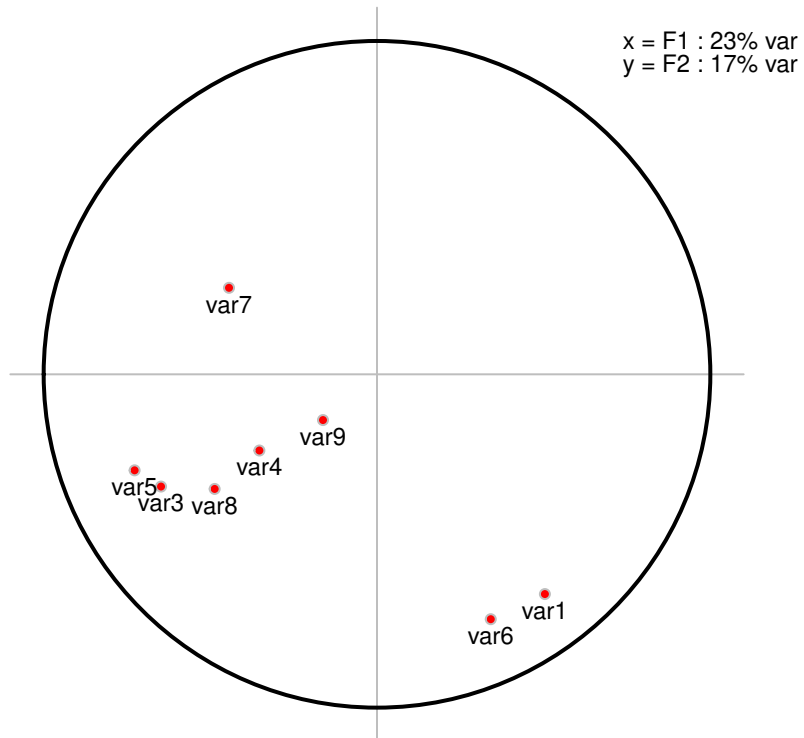
*# use="pairwise.complete.obs" si pas besoin d'éliminer les données manquantes.*

```
library(corrplot)
corrplot(cor(mesDonnees[,tab],use="complete.obs"),method="circle")
```

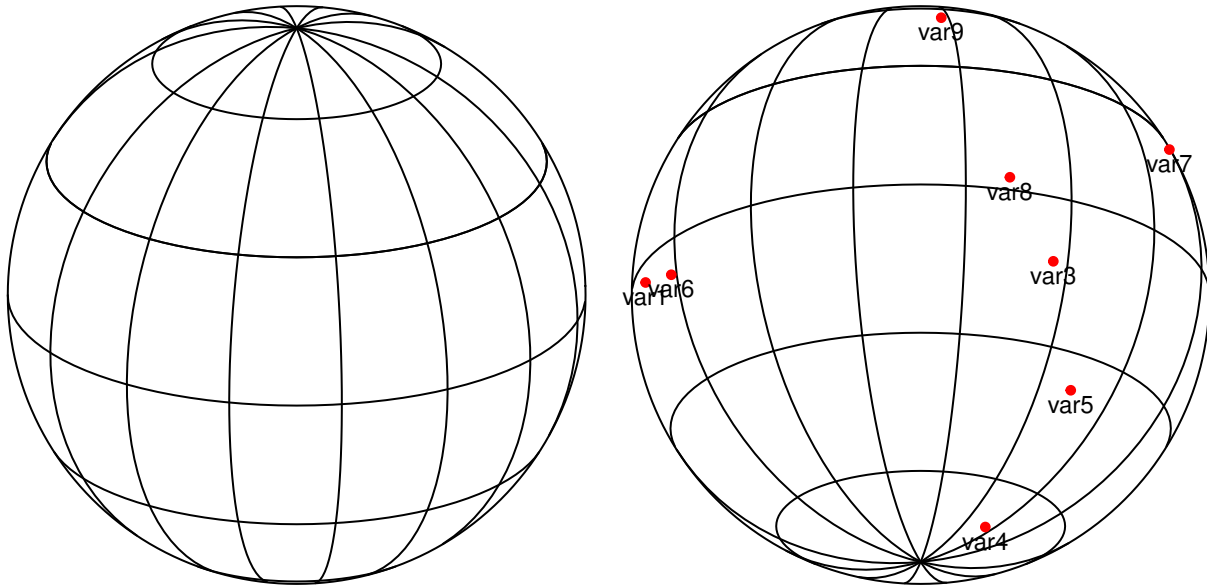


# Analyse en Composante Principale

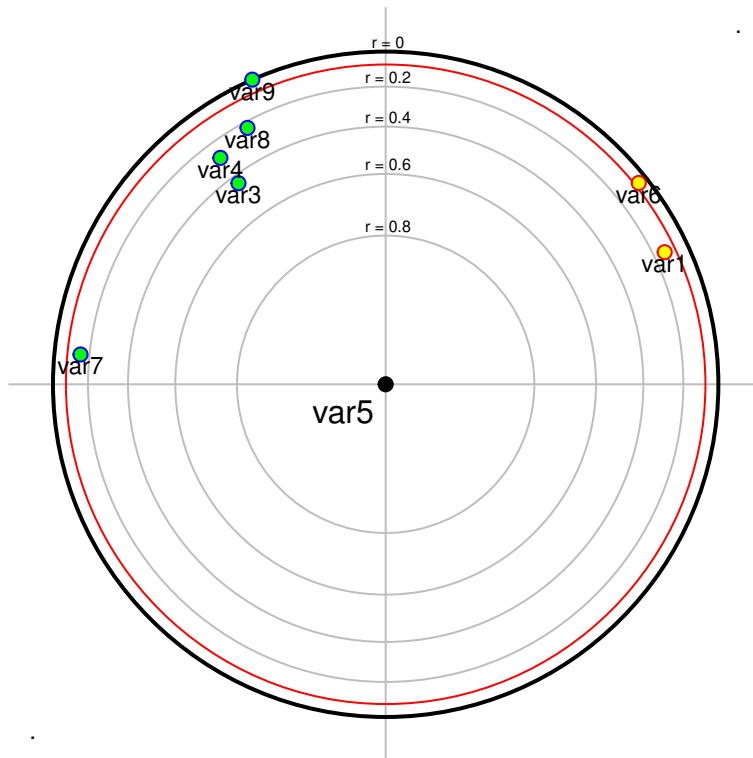
```
tab2 <- c("var1", "var3", "var4", "var5", "var6", "var7", "var8", "var9")  
library(psy)  
mdspca(mesDonnees[, tab])
```



```
sphpca(mesDonnees[,tab],v=55)
```



```
# Ou si une variable à expliquer par rapport à des variables explicatives
expliquer <- "var5"
explicatives <- c("var1","var3","var4", "var6","var7","var8","var9")
fPCA(data=mesDonnees,y=expliquer,x=explicatives,partial="No")
```



## Classification ascendante hiérarchique

```
tab <- c("var1","var3","var4","var5","var6","var7","var8","var9")
cah <- hclust(dist(t(scale(mesDonnees[,tab]))),method="ward.D")
plot(cah,xlab="",ylab="",main="Classification hiérarchique")
```

### Classification hiérarchique



hclust (\*, "ward.D")

```
obj <- cor(mesDonnees[,tab], use="pairwise.complete.obs")  
heatmap(obj, col=gray(seq(1,0,length=16)))
```

