

## classif\_non\_supervisee

2024-05-18

```
sub_sante_mentale=read.csv("sante_mentale_regroupee.csv")
sub_sante_mentale <- sub_sante_mentale[sub_sante_mentale$depressif %in% c("Oui", "Non"), ]

# Convertir toutes les variables du dataframe en facteurs
for (column in names(sub_sante_mentale)) {
  sub_sante_mentale[[column]] <- as.factor(sub_sante_mentale[[column]])
}

# Subsetting the dataframe to keep only rows where depressif is "Oui" or "Non"
#sub_sante_mentale

# Assuming your data frame is named 'data'
sub_sante_mentale <- sub_sante_mentale[c("ville_oyer", "mange_a_ta_faim", "alcool", "Conflits.Familiaux",
  "tiers_temps", "niveau_etude", "anxiété", "suivi_psycho",
  "diagnostic_episode_depressif", "depressif", "suivi_medicale_depressif")]

#sub_sante_mentale
```

On convertit tout en numérique et on centre et réduit

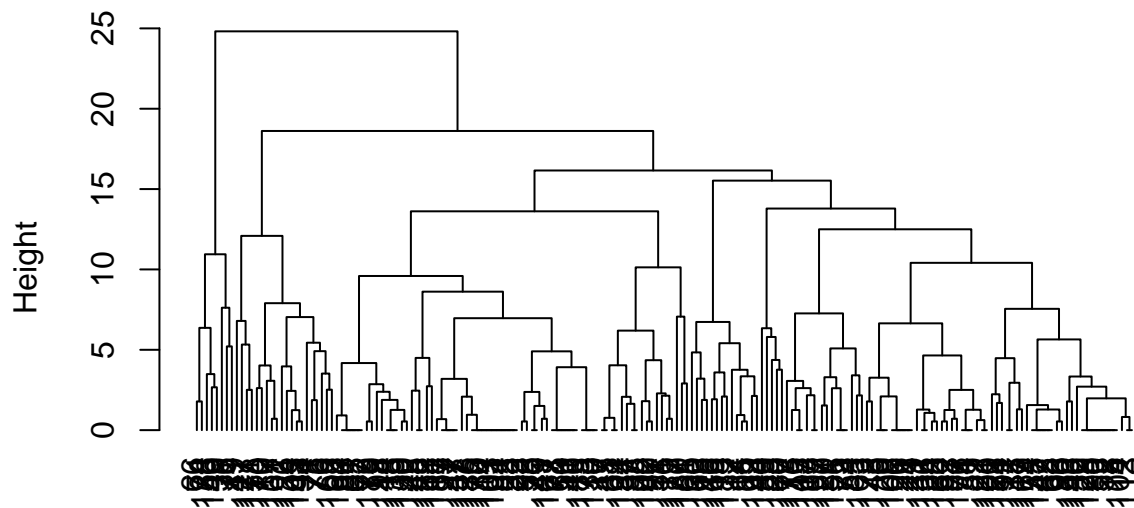
```
sub_sante_mentale_numeric <- as.data.frame(lapply(sub_sante_mentale, function(x) {
  if(is.factor(x)) as.numeric(x) else x
}))

sub_sante_mentale_numeric.cr <- scale(sub_sante_mentale_numeric, center = TRUE, scale = TRUE)
```

Faisons d'abord le CAH

```
d.sub_sante_mentale_numeric = dist(sub_sante_mentale_numeric.cr)
cah.ward <- hclust(d.sub_sante_mentale_numeric, method="ward.D2")
plot(cah.ward, hang=-1)
```

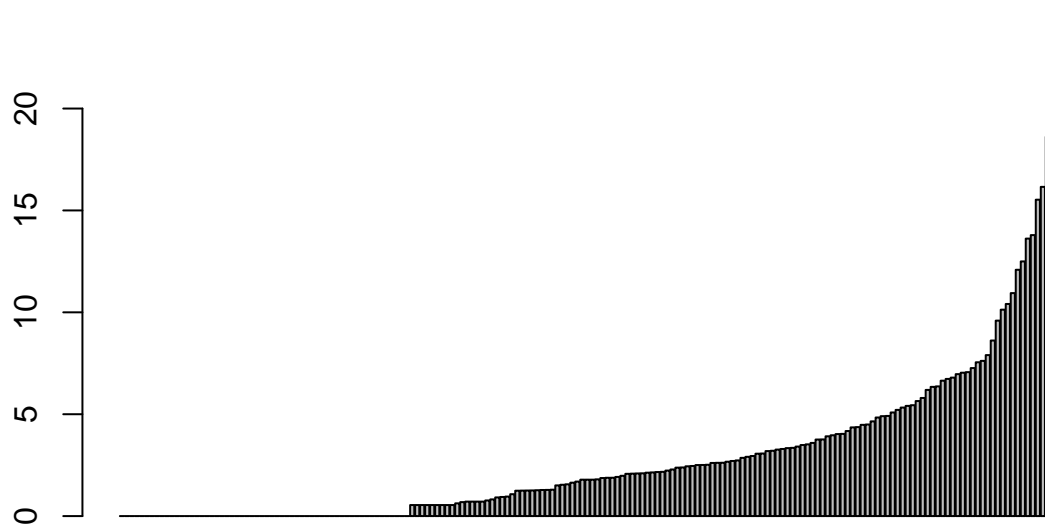
## Cluster Dendrogram



```
d.sub_sante_mentale_numeric  
hclust (*, "ward.D2")
```

En faisant le barplot on obtient :

```
barplot(cah.ward$height)
```

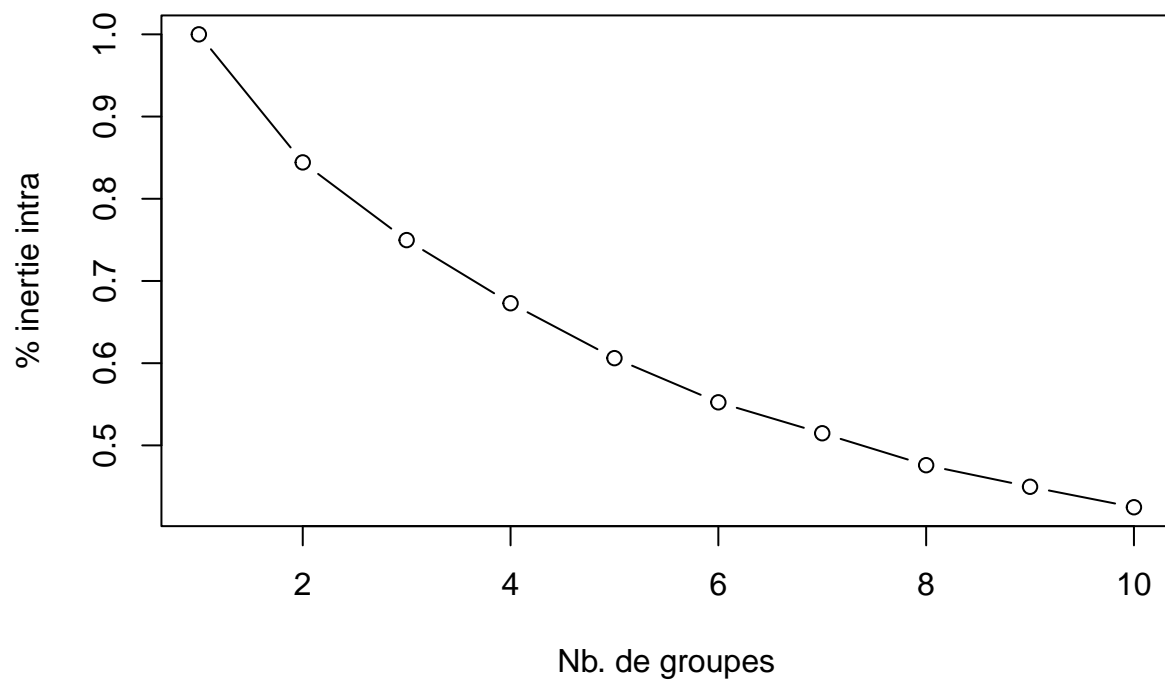


Nous allons donc considéré 4 classes

SI imaginions que nous n'avions pas cette information, alors il faudrait faire dans une boucle while un nombre de fois fini kmeans afin de déterminet quelle est le meilleur nombre de centre tout en réduisant l'erreur.

KMEANS sans K choisi :

```
inertie.intra <- rep(0,times=10)
for (k in 1:10){
  kmeans.result <- kmeans(sub_sante_mentale_numeric.cr,centers=k,nstart=100)
  inertie.intra[k] <- kmeans.result$tot.withinss/kmeans.result$totss
}
# graphique
plot(1:10,inertie.intra,type="b",xlab="Nb. de groupes",ylab="% inertie intra")
```



NOus allons donc choisir  $K = 4$

```
kmeans.result <- kmeans(sub_sante_mentale_numeric,centers = 3)
kmeans.result$size
```

```
## [1] 38 69 81
```

On peut alors faire une ACP pour l'interprétation

```
library(FactoMineR)
res = PCA(sub_sante_mentale_numeric,scale.unit = TRUE)
```

