

```
## Loading required package: FactoMineR
```

Introduction à R

Clustering

Pascal Bessonneau

06/2015

Cluster

Le principe

Le principe est à partir des variables de calculer la distance entre individus et de grouper les individus les plus proches.

Température

Nous utilisons toujours les données sur la température.

```
temp <- read.csv2("data/temp.csv")  
colnames(temp)
```

```
## [1] "Ville"      "Janvier"    "Fevrier"  
## [4] "Mars"       "Avril"      "Mai"  
## [7] "Juin"       "Juillet"    "Aout"  
## [10] "Septembre" "Octobre"    "Novembre"  
## [13] "Decembre"  "lati"       "long"
```

Préparation des données

Dans ce cas il faut centrer et réduire les données pour éviter les problèmes de différence d'unités.

```
numerics <- sapply(temp,is.numeric)
for (ii in which(numerics))
  temp[[ii]] <- scale(temp[[ii]])
```

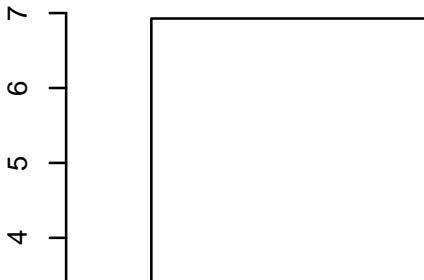
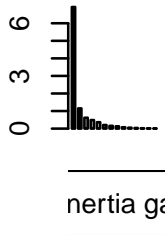
Peut-on résumer les informations ?

```
hcpc <- HCPC(temp[,1:12],nb.clust = 3)
```

```
## Error in FUN(data[x, , drop = FALSE], ...): 'x' doit être  
numérique
```

Hierarchical Clustering

Hierarchical Classification



Résultats

Nous pouvons voir à la longueur des branches de l'arbre quelles sont les villes les plus proches les unes des autres.
Après l'algorithme nous propose une coupure optimale à 3 groupes.

Résultats

Nous avons 1 groupe qui réunit les villes les plus au sud, un cluster qui réunit les villes de Bretagne au climat peu continental avec peu de variations entre les températures extrêmes et les villes au climat plus continental et situé au nord de la Loire.

Résultats

Le clustering fait partie des méthodes de *Machine Learning* qui permettent d'analyser les comportements consommateurs et du profilage des individus sur internet.