

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

Introduction à R

ACP

Pascal Bessonneau

06/2015

Analyse en composante principale

Le principe

Les détails mathématiques ne seront pas présentés. Il s'agit juste de montrer comment on peut synthétiser un problème avec des variables artificielles dont le nombre est inférieur ou très inférieure au nombre de variables initiales qui décrivent les individus.

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"
```

Température

```
head(temp)
```

```
##           Ville Janvier Fevrier Mars  Avril
## 1      Bordeaux      5.6      6.6 10.3  12.8
## 2        Brest      6.1      5.8  7.8   9.2
## 3 Clermont-Ferrand      2.6      3.7  7.5  10.3
## 4      Grenoble      1.5      3.2  7.7  10.6
## 5        Lille      2.4      2.9  6.0   8.9
## 6         Lyon      2.1      3.3  7.7  10.9
##      Mai Juin Juillet Aout  Septembre  Octobre
## 1 15.8 19.3    20.9 21.0      18.6    13.8
## 2 11.6 14.4    15.6 16.0      14.7    12.0
## 3 13.8 17.3    19.4 19.1      16.2    11.2
## 4 14.5 17.8    20.1 19.5      16.7    11.4
## 5 12.4 15.3    17.1 17.1      14.7    10.4
## 6 14.9 18.5    20.7 20.1      16.9    11.4
##  Novembre  Decembre  lati  long
## 1      9.1      6.2 44.50 -0.34
## 2      9.0      7.0 48.24 -4.29
## 3      6.6      3.6 45.47  3.05
## 4      6.5      2.3 45.10  5.43
```

Les données

Ce sont donc les données qui correspondent aux températures moyennes tout au long de l'année pour des villes de France.

Peut-on résumer les informations ?

Le principe de l'ACP est de chercher et de simplifier les corrélations qui existe entre les variables pour créer des variables synthétiques qui avec peu de nouvelles variables résumeront le maximum d'informations possibles.

Pour réaliser l'analyse on utilise le paquet FactoMineR.

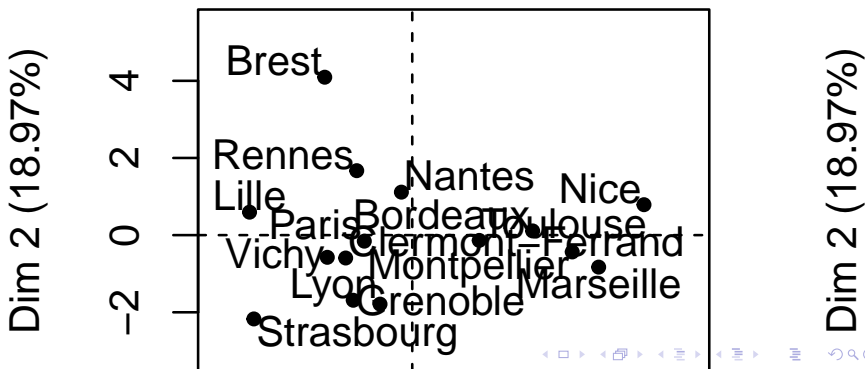
Préparation des données

```
rownames(temp) <- temp$Ville  
temp <- temp[,-1]
```


Graphiques des variables

```
pca <- PCA(temp, quanti.sup = c(13,14))
```

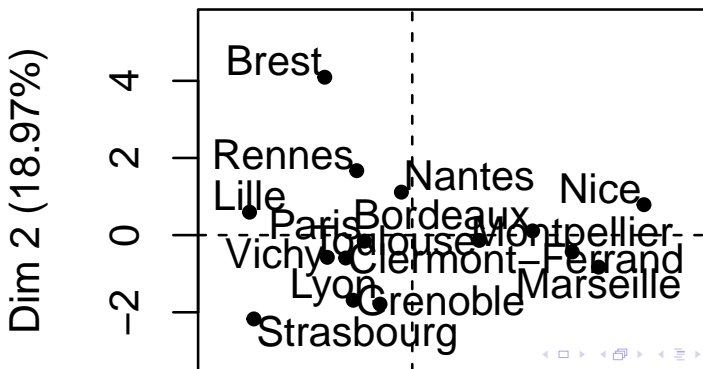
Individuals factor map (PCA)



Graphique des individus sur les deux premiers axes

```
plot(pca, choix = "ind")
```

Individuals factor map (PCA)



Graphique des individus sur les deux premiers axes

```
cor(pca$ind$coord[,1:2],temp[,c("lati","long")])
```

```
##           lati           long  
## Dim.1 -0.8389348  0.1714839  
## Dim.2  0.3064996 -0.7922192
```

Graphique des individus sur les deux premiers axes

A partir des données et des résultats de l'ACP, nous avons pu retrouver la latitude et la longitude approximative.