# TP2: les détails en ACP

## 1 Exercice 1

Récupérer les données où $n = 8$ eaux minérales sont décrites sur $p = 13$ variables.

```
load("../data/eaux.rda")
X <- data # Matrice des données brutes
dim(X)
```

1. Calculer la distance entre les eaux St Yorre et Badoit.

```
sqrt(sum((X[1,]-X[2,])^2))
```

2. Constuire la matrice des distances avec la fonction `dist`. Quelle est la classe de l'objet en sortie de cette fonction ? Comment transformer cet objet en une matrice $8 \times 8$ ?

```
D <- dist(X)
class(D)
as.matrix(D)
```

3. Calculer la moyenne et l'écart-type des 13 variables avec la fonction `apply`.

```
m <-  apply(X,2,mean) #mean
s <- apply(X,2,sd) #standard deviation
```

4. Utiliser ces moyennes pour construire la matrice `Y` des données centrées sans utiliser la fonction `scale` mais en utilisant la fonction `sweep`. Vérifier que les colonnes de `Y` sont bien de moyenne nulle.

```
Y <- sweep(X,2,STATS=m,FUN="-")
apply(Y,2,mean)
```

5. Construire la matrice de covariance à partir de la formule $\frac{1}{n}Y^TY$ (le produit matriciel dans R se fait avec (`%*%`). Que faut-il faire pour retrouver les mêmes résultats avec la fonction `cov`.

```
n <- nrow(X)
t(as.matrix(Y))%*%as.matrix(Y)/
cov(X)
#Attention division par n-1=7
all.equal(cov(X),t(as.matrix(Y))%*%as.matrix(Y)/(n-1))
```

6. Construire la matrice `Z` des données centrées-réduites sans utiliser la fonction `scale` mais en utilisant la fonction `sweep`. Vérifier que les colonnes de `Z` sont bien de variance unité.

```
Z <- sweep(Y,2,s,"/")
apply(Z,2,sd)
```

5. Construire la matrice des corrélations à partir de la formule $\frac{1}{n}Z^TZ$. Constatez en utilisant la fonction `all.equal` que vous ne trouvez pas les mêmes résultats qu'avec la fonction `cor`.

```
(t(as.matrix(Z))%*%as.matrix(Z)/n)
cor(X)
all.equal(cor(X),t(as.matrix(Z))%*%as.matrix(Z)/n)
```

6. Construire maintenant la matrice `Z` des données centréees-réduites en utilisant l'écart-type **non corrigé**.

```
s <- apply(X,2,sd)*sqrt((n-1)/n) #non corrected standard deviation
print(s,digits=3)
```

```
Z <- sweep(Y,2,s,"/")
apply(Z,2,mean)
apply(Z,2,sd)*sqrt((n-1)/(n))
```

7. Vérifier maintenant que la formule $\frac{1}{n}Z^TZ$ donne bien les mêmes résultats que la fonction `cor`.

```
(t(as.matrix(Z))%*%as.matrix(Z)/n)
cor(X)
all.equal(cor(X),t(as.matrix(Z))%*%as.matrix(Z)/n)
```

# 2 Exercice 2

On va utiliser la fonction `PCA` du package `FactoMineR` pour faire l'analyse en composantes principales des données sur les eaux minérales.

```
library(FactoMineR)
?PCA
```

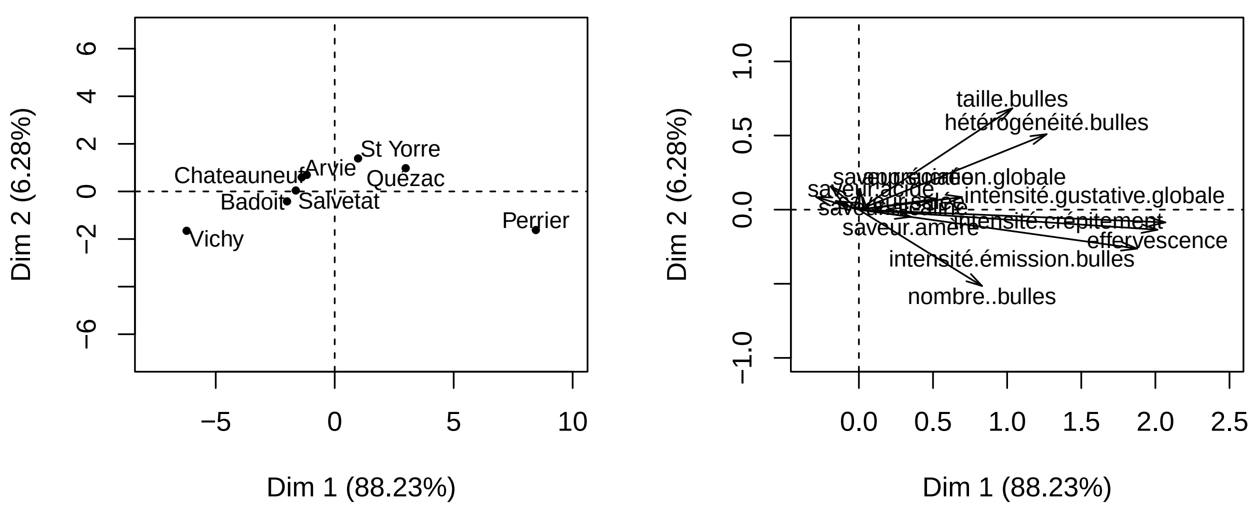
## 2.1 ACP sur matrice de covariance

L'ACP sur matrice de covariance analyse les lignes et les colonnes de la matrice $Y$ des données centrées via une décomposition en valeurs propores de la matrice de covariances $C = \frac{1}{n}Y^tY$.

1. Quel est l'argument de la fonction `PCA` qui permet de réaliser une ACP non normée (sur matrice des covariances) ?

2. Faire une ACP non normées des données sur les eaux minérales pour retrouver les graphiques ci-dessous et les interpréter.

```
res <- PCA(X, graph=FALSE, scale.unit=FALSE)
par(mfrow=c(1,2))
plot(res,choix="ind", cex=0.8, title="", graph.type="classic")
plot(res,choix="var", cex=0.8, title="", graph.type="classic")
```

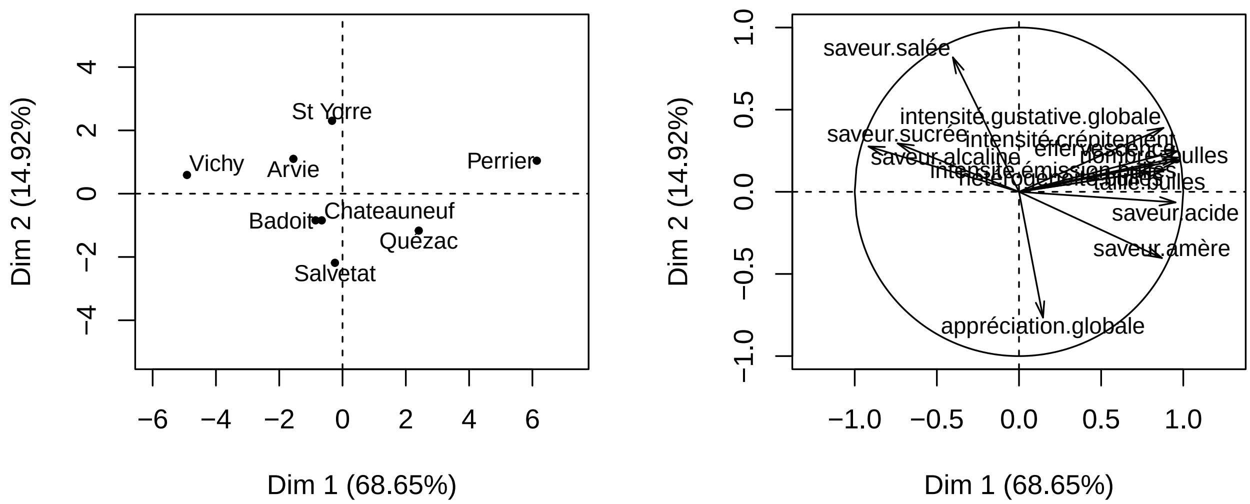


## 2.2 ACP sur matrice des corrélations

L'ACP sur matrice de corrélations analyse les lignes et les colonnes de la matrice $Z$ des données centrées-réduites via une décomposition en valeurs propores de la matrice de corrélations $R = \frac{1}{n}Y^tY$.

1. Quel est l'argument de la fonction `PCA` qui permet de réaliser une ACP normée (sur matrice des corrélations) ?
2. Faire une ACP normées des données sur les eaux minérales pour retrouver les graphiques ci-dessous et les interpréter.

```
res <- PCA(X, graph=FALSE)
par(mfrow=c(1,2))
plot(res,choix="ind", cex=0.8, title="", graph.type="classic")
plot(res,choix="var", cex=0.8, title="", graph.type="classic")
```

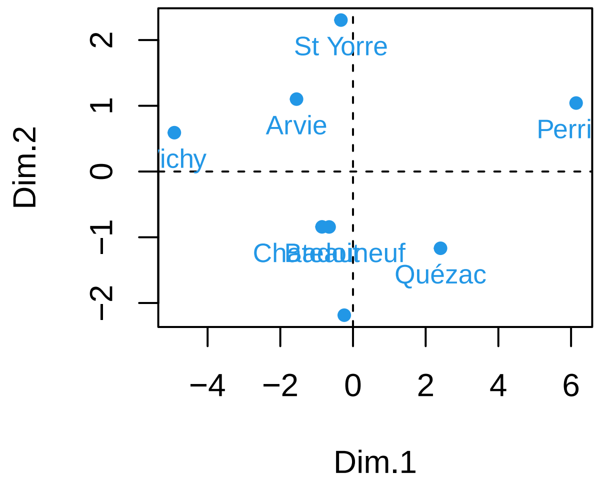



3. Combien de composantes principale sont constuites par défaut ? Vérifier que ces nouvelles variables sont centrées et que leurs variances correspondent aux valeurs propres.

```
F <- res$ind$coord
apply(F,2,mean)
n <- nrow(X)
apply(F,2,var)*(n-1)/n
res$eig[,1]
```

4. Utiliser la fonction `plot` pour représenter les individus sur le premier plan factoriel et retrouver le graphique ci-dessous.

```
axes <- c(1,2)
plot(F[,axes],pch=19,col=4,cex=0.8)
abline(h=0,lty=2)
abline(v=0,lty=2)
text(F[,axes],labels=rownames(Z),pos=1,col=4,cex=0.8)
```

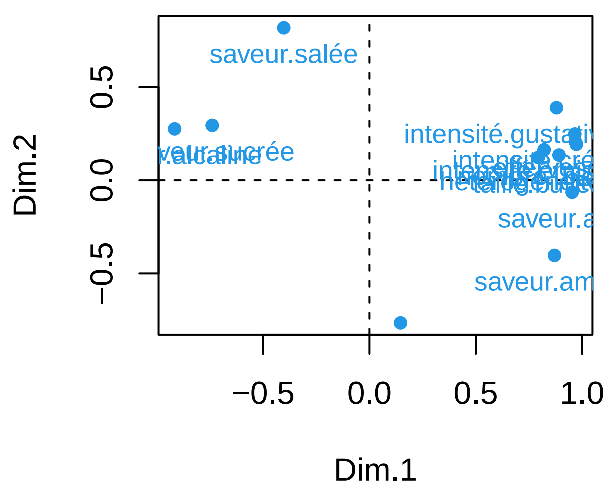



5. Calculer les correlations entre la première composantes principales et les 13 variables.Vérifier que les coordonnées factorielles des variables sur la première dimension correspondent bien à ces corrélations.

```r
A <- res$var$coord
cor(X[,1],F[,1])
cor(X,F[,1])
A[,1, drop=FALSE]
```

4. Utiliser la fonction `plot` pour représenter les variables sur le premier plan factoriel et retrouver le graphique ci-dessous.

```r
axes <- c(1,2)
plot(A[,axes],pch=19,col=4,cex=0.8)
abline(h=0,lty=2)
abline(v=0,lty=2)
text(A[,axes],labels=colnames(Z),pos=1,col=4,cex=0.8)
```




## 3 Exercice 3.

Il exsite au moins deux fonctions qui font de l'ACP dans le paquet de base `stats` : `princomp` et `prcomp`.

1. Regarder l'aide de ces deux fonctions pour savoir si par défaut elles font une ACP ou non normée ? Quel argument pour chacune d'entre elle permet de controler le type d'ACP réalisée ?

```
?princomp
# argument cor=FALSE par défaut donc ACP non normée
?prcomp
# argument scale=FAMSE par défaut donc ACP non normée
```

2. Essayer de retrouver avec ces deux fonctions les résultats numériques (composantes principales, coordonnées des variables, valeurs propres) que vous avez obtenus à l'exercice précédent pour l'ACP normée.

```
#pr <- princomp(X,cor=T) # on covariance matrix
?prcomp
pr <- prcomp(X,scale=TRUE) # on covariance matrix
pr$sdev^2 # eigenvalues
```

# 4 Exercice 4 (à rendre).

Retrouver maintenant les résultats numériques (composantes principales, coordonnées factorielles des variables et valeurs propres) de l'ACP normée en effectuant avec la fonction `eigen` une décomposition en valeurs propres de la matrice de corrélations.