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.
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 8×8 ?
3. Calculer la moyenne et l'écart-type des 13 variables avec la fonction `apply`.
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.
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`.
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.
7. 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`.
8. 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)
```

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

2 Exercice 2

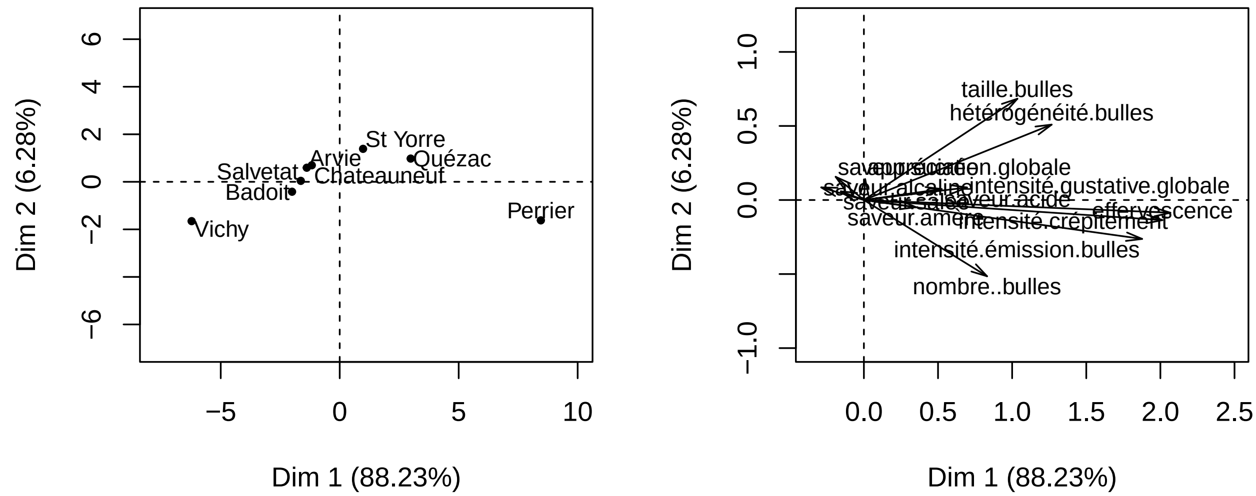
On va utiliser la fonction `PCA` du package `FactoMineR` pour faire l'analyse en composantes principales des données sur les eaux minérales.

```
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 $C = \frac{1}{n}Y^tY$.

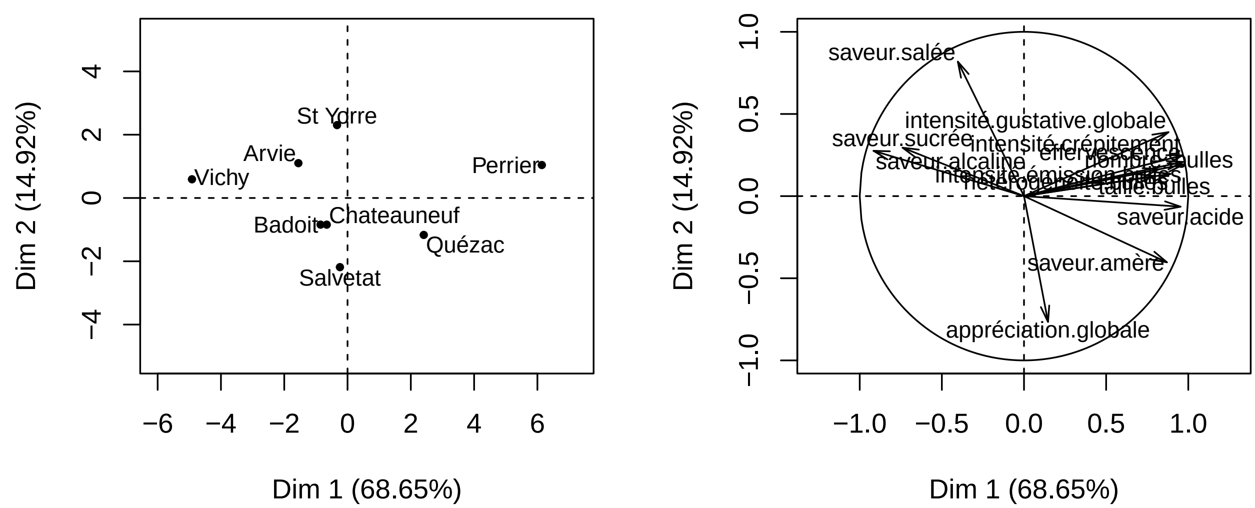
1. Quel est l'argument de la fonction `PCA` qui permet de réaliser une ACP non normée (sur matrice des covariances) ?
2. Faire une ACP non normées des données sur les eaux minérales pour retrouver les graphiques ci-dessous et les interpréter.



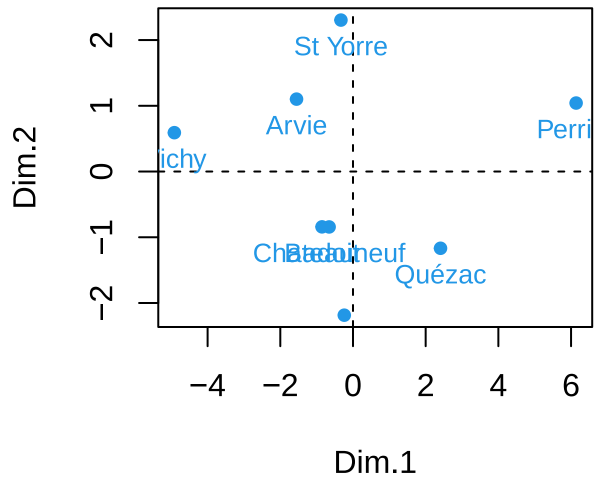
2.2 ACP sur matrice des corrélations

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 $R = \frac{1}{n}Y^tY$.

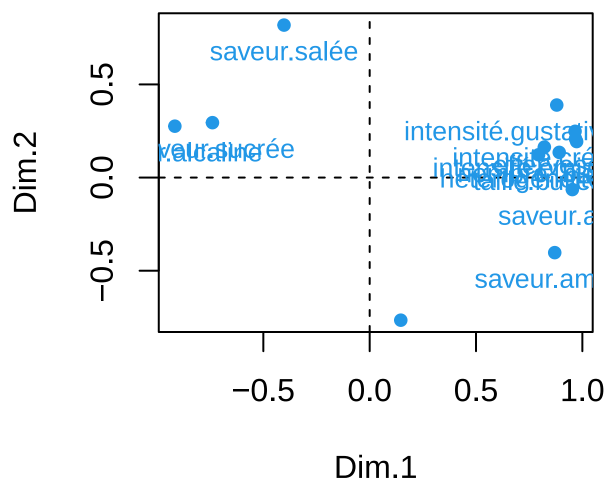
1. Quel est l'argument de la fonction `PCA` qui permet de réaliser une ACP normée (sur matrice des corrélations) ?
2. Faire une ACP normées des données sur les eaux minérales pour retrouver les graphiques ci-dessous et les interpréter.



3. Combien de composantes principale sont constuites par défaut ? Vérifier que ces nouvelles variables sont centrées et que leurs variances correspondent aux valeurs propres.
4. Utiliser la fonction `plot` pour représenter les individus sur le premier plan factoriel et retrouver le graphique ci-dessous.



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



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 ?
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.

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.