

PROJET à rendre : *Etude des données* EauxFM.txt

**A rendre lors de la dernière séance de dernier TP
(à préciser selon le groupe) !**

Les données à traiter (**EauxFM.txt**) ont déjà été partiellement analysées (voir fichier **eaux1.txt** (20*7)) dans Bellanger & Tomassone (2014). Le fichier a été mis à jour et complété par de nouvelles données.

DESCRIPTION

Le fichier comprend les mesures de 12 variables pour 95 eaux. Il est constitué de 95 lignes (sans compter la 1ere ligne correspondant au nom de chaque colonne) et 12 colonnes correspondant aux variables suivantes :

- Nom
- Nature : plat (eau plate), gaz (eau gazeuse)
- Ca : Calcium (en mg/l)
- Mg : Magnesium (en mg/l)
- Na : Sodium (en mg/l)
- K : Potassium (en mg/l)
- Cl : Chlorures (en mg/l)
- NO3 : Nitrates (en mg/l)
- SO4 : Sulfates (en mg/l)
- HCO3 : bicarbonates (en mg/l)
- PH
- Pays : France ou Maroc

Les objectifs sont de :

- construire une fonction sous R permettant de réaliser une ACP.
- décrire et synthétiser les informations contenues dans fichier **EauxFM.txt** en effectuant :
 - une description univariée et bivariée
 - une description multivariée à l'aide :
 - d'une ACP (à l'aide de votre fonction ; mais aussi des fonctions disponibles sous R et en comparant les deux) en considérant les eaux marocaines comme supplémentaires.
 - et si vous avez le temps à l'aide
 - d'une classification non supervisée de type kmeans.

INDICATIONS (pour la rédaction du travail à rendre **par 2 au plus**)

- Décrire les données : représentation graphique, résumé numérique.... puis commenter et interpréter les résultats ! Attention la table peut contenir des données manquantes ou des données aberrantes, à décrire aussi.
- Décrire la fonction R construite permettant de réaliser une ACP (points forts et points faibles en comparant à l'existant, comment l'utiliser ?)
- Rédiger un rapport construit :
 - titre, auteur, sommaire, pagination, bibliographie, introduction, conclusion, figures et tableaux numérotés ...
- Faire figurer en annexe vos programmes R permettant d'aboutir aux résultats graphiques et/ou numériques.
- 20 pages au plus sans compter les annexes.

Remarque : L'utilisation dans votre rapport des sorties issues de la library FactoInvestigate (Automatic Description of Factorial Analysis) ou de toute autre library similaire, sans aucun apport de votre part, sera évidemment sanctionnée.

Rendu semaine 42 ou 43 selon groupe : un rapport version papier et un dossier (informatique) contenant une version pdf de votre rapport, votre code R ainsi que le fichier de données utilisé pour faire fonctionner votre fonction R.

Contacts : Lise Bellanger (lise.bellanger@univ-nantes.fr) et Klervi Le Gall (Klervi.Legall@univ-nantes.fr)