

## **PROJET à rendre : Données d'expression génétique** *Projet M1 AD 1920.csv*

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

### **DESCRIPTION**

Les données à traiter (*Projet M1 AD 1920.csv*) se composent de 40 échantillons de tissus associés à des mesures sur 1000 gènes. Les 20 premiers échantillons proviennent de patients en bonne santé, tandis que les 20 autres proviennent d'un groupe de malades. Le jeu de données est donc de 1000 lignes (gènes) et 40 colonnes (individus).

**Les objectifs** sont de :

- construire une fonction sous *R* permettant de réaliser une ACP.
- décrire et synthétiser les informations contenues dans fichier *Projet M1 AD 1920.csv* en effectuant :
  - une description univariée et bivariée
  - une description multivariée à l'aide :
    - d'une ACP sur le jeu de données transposé 40\*1000 (à l'aide de votre fonction ; mais aussi des fonctions disponibles sous *R* et en comparant les deux). Examiner quels sont les gènes qui diffèrent le plus entre les individus en bonne santé et les patients malades.  
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 !
- 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.

### **Contacts :**

Lise Bellanger ([lise.bellanger@univ-nantes.fr](mailto:lise.bellanger@univ-nantes.fr))