

TP1

Exercice n°1 : Tortues

- (a) Charger le jeu de données `PaintedTurtles.txt` contenu dans le dossier `data` et le décrire.
- (b) Décrire chaque variable du jeu de données et proposer un modèle probabiliste possible pour chaque variable (sans aller jusqu'à tester l'adéquation des données à une loi donnée).
- (c) Réaliser une ACP et interpréter les résultats de cette visualisation.

Exercice n°2 : Visualisations des données transcriptomiques T-cell

Ces données sont tirées du livre "Modern Statistics for Modern Biology", Susan Holmes et Wolfgang Huber. Il s'agit de données transcriptomiques issus de l'article suivant Holmes, Susan, Michael He, Tong Xu, and Peter P Lee. 2005. "Memory T Cells Have Gene Expression Patterns Intermediate Between Naive and Effector." PNAS 102 (15) : 5519–23. Les profils d'expression de gènes ont été mesurés chez 10 individus pour trois types de cellules.

```
load("data/Msig3transp.RData")
```

- (a) Décrire le jeu de données.
- (b) Décrire chaque variable du jeu de données et proposer un modèle probabiliste possible pour chaque variable (sans aller jusqu'à tester l'adéquation des données à une loi donnée).
- (c) Réaliser une ACP et interpréter les résultats de cette visualisation.

Exercice n°3 : Données d'expression de gènes

Les données suivantes sont extraites de cet article <https://pubmed.ncbi.nlm.nih.gov/21455293/>. L'étude visait à comparer l'expression des gènes à l'aide de la technologie RNA-seq chez deux lignées de souris. Les données sont téléchargeables depuis les adresses https://bowtie-bio.sourceforge.net/recount/countTables/bottomly_count_table.txt et https://bowtie-bio.sourceforge.net/recount/phenotypeTables/bottomly_phenodata.txt.

- (a) Charger les données suivantes. Décrire le jeu de données.
- (b) Décrire chaque variable du jeu de données et proposer un modèle probabiliste possible pour chaque variable (sans aller jusqu'à tester l'adéquation des données à une loi donnée).
- (c) Réaliser une ACP et interpréter les résultats de cette visualisation.

Exercice n°4 : Abondance de bactéries

- (a) Télécharger et installer la package `phyloseq`.

```
if (!require("BiocManager", quietly = TRUE))  
  install.packages("BiocManager")  
BiocManager::install("phyloseq")
```

- (b) Extraire les données suivantes et comprendre la structure des données à l'aide des commandes suivantes.

```
library(phyloseq)  
data("GlobalPatterns", package = "phyloseq")  
GPOTUs = as.matrix(t(phyloseq::otu_table(GlobalPatterns)))  
#GPOTUs[1:4, 6:13]  
#help(GlobalPatterns)
```

- (c) Décrire chaque variable du jeu de données et proposer un modèle probabiliste possible pour chaque variable (sans aller jusqu'à tester l'adéquation des données à une loi donnée).
- (d) Réaliser une ACP et interpréter les résultats de cette visualisation.