

# Evaluation n1 : Programmation Statistique avec R

## Prérequis :

- Charger les packages tidy, tibble, readr, dplyr, ggplot2 et scales. Si ces packages ne sont pas présents, vous devrez les installer.

## Rendu :

- 2 fichiers sont attendus pour chaque groupe : Un fichier HTML dans lequel il est possible de voir le code et les résultats + Le fichier Rmd correspondant.
- Les noms des fichiers doivent être formatés de la manière suivante :
  - NomPrenom.html
  - NomPrenom.Rmd
- Si possible, mettre les 2 fichiers dans un seul zip avec nom : NomPrenom.zip
- Le rendu doit être déposé dans moodle avant la fin de la séance (Si problème, envoyez les fichiers à l'adresse : sabrine.bendimerad1@gmail.com)
- Les données utilisées sont les données mtcars et iris disponibles déjà dans R + Données de la banque mondiale **WGI** déposés sur moofdle.

**Notation :** La notation prendra en compte la qualité du code (clarté, commentaires quand nécessaire et rendu compréhensible).

## Exercice 1 : Manipulation

1. Créer un objet `mtcars2` qui contient les données `mtcars` auquel vous ajouterez la variable `nom_vehicule` qui correspond au label des lignes.
2. Trouver le nombre d'observation et le nombre de variable de `mtcars2` (Indication: `glimpse()`).
3. Trouvez le type de chaque variable.
4. Afficher les dix premières lignes de `mtcars2`.
5. Sélectionner les variables `gear`, `carb` et `nom_vehicule` de `mtcars2`, les enregistrer dans une nouvelle table `mtcars_select1`.
6. Sélectionner toutes les variables sauf `gear`, `carb` de `mtcars2`, les enregistrer la nouvelle table `mtcars_select2`.
7. Créer `mtcars_fusion` en fusionnant `mtcars_select1` et `mtcars_select2` par la variable `nom_vehicule`.
8. Créer une variable `rapport` dans `mtcars_fusion` qui calcule le rapport entre `carb` et `gear`.
9. Renommer la variable `rapport` en `rapport_carb_gear`, puis créer une variable `mpg_plus_cyl` contenant la somme entre `mpg` et `cyl`.
10. Faire la moyenne de toutes les variables `mtcars_fusion`.

## Exercice 2 : Visualisation

1. Créer un histogramme de la longueur des pétales.  
⇒ Changer l'effectif en fréquence, la fréquence en pourcentage.
2. Tracer le nuage de points de la variable *hp* en fonction de *mpg*. Ajouter la droite de régression linéaire sur le graphique.
3. Tracer les histogrammes en densité de la variable *mpg* suivant les modalités de la variable *cyl*. Colorer chaque histogramme.

## Exercice 3 : Manipulation + Visualisation

1. Lire les données de la banque mondiale (4 fichiers déposés dans moodle, dont le nom commence par **WGI**).
2. Faire une jointure afin de concaténer l'information des 4 tables.
3. Filtrer pour garder que les valeurs des indicateurs estimés (**EST** dans la colonne **TypeCode**)
4. Créer une colonne pour les données manquantes dans la Table longue (La colonne affiche **True** si il existe des valeurs manquantes, sinon **False**).

5. Représenter sur un graphique l'évolution du pourcentage de données manquantes en fonction de l'année.

⇒ Indication : changer les étiquettes de la variable année en valeurs numériques **as.numeric(substring(annee, 2))**

6. Créer la table longue restreinte aux valeurs du contrôle de la corruption, la nommer **evolution\_corruption**.

7. Représenter l'évolution de la corruption en fonction de l'année.

### Quelques informations sur les fichiers :

- Country : Nom du pays et le code associé
- Indice : Le type de méthode de calcul l'indicateur existant et le code associé
- Serie : Le type de l'indicateur et son code
- Value : Les valeurs des indicateurs pour chaque pays et pour chaque année