**Evaluation n 2 : Programmation Statistique avec R**

**Prérequis :**

* Charger les packages tidyr, tibble, readr, dplyr, ggplot2 et scales. Si ces packages ne sont pas présents, vous devez les installer.

**Rendu :**

* Fichier zip nommé NomPrenom.zip contenant 2 fichiers : **Un fichier HTML** dans lequel il est possible de voir le code et les résultats + **Un fichier Rmd**.
* Le rendu doit être envoyé à l’adresse : [sabrine.bendimerad1@gmail.com](mailto:sabrine.bendimerad1@gmail.com) (Maximum 16h30, prendre en compte 15 min pour l’enregistrement, le nommage et surtout la création du zip : aucun rendu ne sera noté après 16h30)
* Les données utilisées sont les données **mtcars** disponibles dans R + Données de la banque mondiale **WGI** + Données **COMMERCEPARIS.**

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

**Exercice 1 : Visualisation**

1. A partir des données mtcars, tracer le nuage de points de la variable hp en fonction de mpg. Ajouter la droite de régression linéaire sur le graphique.
2. Tracer les histogrammes en densité de la variable *mpg* suivant les modalités de la variable *cyl*. Colorer chaque histogramme.

**Exercice 2 : Manipulation**

Cet exercice porte sur les données de commerces à Paris. Ces données, datant de 2014, recensent tous les commerces de la ville de Paris, leurs types, leurs arrondissements, leurs adresses, leurs longitudes et latitudes, ainsi que d’autres données géographiques.

1. Charger les données au format .RData avec la commande load
2. Extraire de la table les éléments suivants :
   1. le commerce de numéro d’ordre 86181
   2. tous les commerces situés au “137 avenue de Versailles”
   3. le nombre de commerces situés avenue de Versailles (voir variable : LIBELLE VOIE)
3. Créer une table au format tibble contenant les informations suivantes : ordre, code activité, libellé activité, adresse complète, numéro, type voie, libellé voie, iris, ilot, quartier, arrondissement, longitude et latitude.
4. Créer une table contenant les informations suivantes pour chaque arrondissement : nombre de commerces et nombre de types de commerces différents présents dans l’arrondissement (voir variable : CODE ACTIVITE pour le type de commerce)
5. Faire deux graphiques représentant :
   1. Le nombre de commerces par arrondissement,
   2. Le nombre de types de commerces différents par arrondissement

**Exercice 3 : Manipulation + Visualisation**

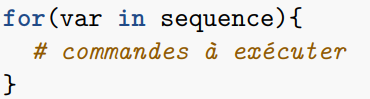
1. Lire les données de la banque mondiale (4 fichiers déposés dans moodle, dont le nom commence par **WGI**).
2. Faire une jointure afin de concaténer l’information des 4 tables.
3. Filtrer pour garder que les valeurs des indicateurs estimés (**EST** dans la colonne **TypeCode**)
4. Etudier le pourcentage de données manquantes pour chaque année de la manière que vous estimez la plus optimale.
5. Représenter sur un graphique l’évolution du pourcentage de données manquantes en fonction de l’année.

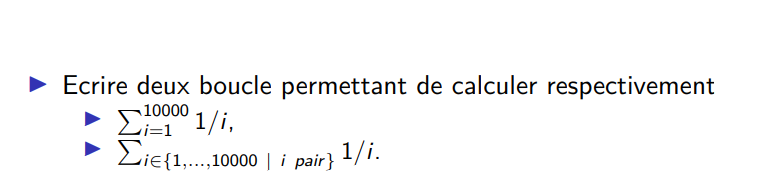
**Quelques informations sur les fichiers :**

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

**Exercice 4 : Boucles**

Une boucle en programmation est une commande qui permet de réaliser un traitement itératif. La boucle la plus utilsiée est nommée for et suit la syntaxe suivante :



****