

Statistique descriptive avec R

TP 7

Nous allons travailler sur les données de population des communes de France (millésime 2022, France hors Mayotte).

1. Rendez-vous sur la page officielle de l'Insee : <https://www.insee.fr/fr/statistiques/8581696> Téléchargez le fichier suivant : *Évolution et structure de la population en 2022 - Commune - France hors Mayotte*. Prenez les deux formats : **.xlsx** (pour explorer les noms de colonnes dans Excel/LibreOffice) et **.csv** (celui qu'on utilisera dans R).

, téléchargez les données *Évolution et structure de la population en 2022 - Commune - France hors Mayotte* en format **csv** et au format **xlsx** pour mieux les visualiser. Avec le format **xlsx**, détectez le nom de la colonne du code géographique, et celles des effectifs de la population en 2022, au total et par tranche d'âge (ce qui fait 9 colonnes). Nous créons le vecteur avec le nom des colonnes :

```
noms_col <- c("CODGEO", "P22_POP", "P22_POP0014", "P22_POP1529", "P22_POP3044", "P22_POP4559", "P22_POP6074", "P22_POP7589", "P22_POP90P")
```

2. Grâce à **data.table**, importez dans **R** les données du fichier **csv**. Vous sélectionnez à l'import les 9 colonnes choisies, ce qui évite de tout importer. Vous nommez la variable **Pop**.

3. Rendez-vous sur :

<https://www.data.gouv.fr/datasets/communes-de-france-base-des-codes-postaux/> et téléchargez les données **communes-département-région.csv** puis importez-les dans **R**. Vous nommez la variable **Info**.

4. Les colonnes **CODGEO** et **code_commune_INSEE** représente la même chose, détectez une anomalie, expliquez-là et corrigez-là.

5. On veut enrichir **Pop** avec uniquement trois colonnes d'**Info**. Faites une jointure à gauche où vous ajoutez à **Pop** les colonnes **code_département**, **latitude**, **longitude** de **Info** que vous nommez respectivement **"DEP"**, **"LAT"**, **"LONG"**.

6. Pour quelques communes, le département n'a pas été trouvé, il vaut **NA**. Pour celles-ci, affichez la taille de la population et le code géo : sont-elles négligeables ?

7. Pour Paris, Marseille et Lyon, ajoutez à la main le code du département dans **Pop**.

8. Calculez le nombre d'habitants par département, on supprimera les **NA** résiduels, et on l'enregistre dans une variable **Pop_Dep**.

9. Après avoir installé le package si nécessaire, exécutez le code suivant.

```

# Package qui permettra d'afficher la carte
library(geodata)
# Récupération de la carte au niveau des départements
fr ← gadm(country = "FRA", level = 2, path = tempdir())
# Les numéros de département sont ici
fr$CC_2
# Test d'affichage
couleurs ← rainbow(length(fr))
plot(fr, col = couleurs, border = "white", lwd = 0.6)

```

10. Une couleur peut être décomposée en quantité de rouge, de vert et de bleu entre 0 et 255. La fonction `rgb` pour une quantité de rouge `r`, de vert `v` et de bleu renvoie `b` renvoie le code couleur correspondant `rgb(r, v, b)`, et elle est vectorielle.

Construisez un vecteur de couleurs où, pour chaque taille de population `Pop_Dep$POP_DEP`, la couleur va linéairement de `col_0 ← c(236, 240, 241)` (population nulle) à `col_max ← c(44, 62, 80)` (population la plus élevée de l'ensemble des départements).

10. Affichez la carte des départements avec les couleurs correspondantes.