

TP2

LAWA FOUMSOU Prosper

2025-01-23

Verification puis installation des packages

```
## Chargement des packages après vérification
packages <- c("ggplot2", "dplyr", "haven", "utils", "knitr")
for (i in packages) {
  if (!require(i, character.only = TRUE)) {
    # Si le package n'est pas installé, l'installer
    install.packages(i, dependencies = TRUE)

    # Charger le package après l'installation
    if (!require(i, character.only = TRUE)) {
      stop(paste("Le package", i, "n'a pas pu être installé."))
    }
  } else {
    message(paste("Le package", i, "est déjà installé."))
  }
}
```

Chargement des bases ménage et individus

Après installation des packages nécessaires pour la suite de notre travail, importons les bases.

```
# Importation des bases
base_ind <- haven::read_dta("../Données/ehcvm_individu_SEN2018.dta")
base_men <- haven::read_dta("../Données/ehcvm_ménage_SEN2018.dta")

# Gestion des labels
base_ind <- base_ind %>%
  mutate(across(where(is.labelled), haven::as_factor))
base_men <- base_men %>%
  mutate(across(where(is.labelled), haven::as_factor))
```

Informtions sur les bases

```
# Nombre d'observations et de variables dans chaque bases
```

```
base::dim(base_ind)
```

```
## [1] 66120    51
```

```
base::dim(base_men)
```

```
## [1] 7156    34
```

Jointures

Les différents types de jointures avec chacune sa fonctionnalité. On distingue:

1. La jointure qui conserve les lignes communes aux deux tables. Elle est appelée inner join
2. La jointure qui conserve toutes les lignes de la table de gauche. Elle est appelée left join
3. La jointure qui conserve toutes les lignes de la table de droite. Elle est appelée right join
4. La jointure qui conserve toutes les lignes des deux tables. Elle est appelée full join
5. La jointure qui conserve toutes les lignes de la table de gauche qui ont une correspondance. Elle est appelée semi join
6. La jointure qui conserve toutes les lignes de la table de gauche sans correspondance. Elle est appelée anti join.

Effectuons une left join pour récupérer les informations du chef de ménage

```
base_ind_unique <- base_ind %>%
  dplyr::group_by(hhid) %>%
  dplyr::slice(1) %>%
  dplyr::ungroup()
```

```
base_CM <- base_men %>%
  dplyr::left_join(base_ind_unique, by = "hhid")
```

```
# Informations de la base
```

```
base::dim(base_CM)
```

```
## [1] 7156    84
```

Informations sur le chef de ménage

```
# Répartition des CM selon le sexe
base_CM %>%
  count(sexe) %>%
  kable(col.names = c("Sexe", "Fréquence"), caption = "Tableau des fréquences de sexe")
```

Table 1: Tableau des fréquences de sexe

Sexe	Fréquence
Masculin	5256
Féminin	1900

```
# Répartition des CM selon le milieu de résidence
base_CM %>%
  count(milieu) %>%
  kable(col.names = c("Milieu de résidence", "Fréquence"), caption = "Tableau des fréquences du milieu de résidence")
```

Table 2: Tableau des fréquences du milieu de résidence

Milieu de résidence	Fréquence
Urbain	3941
Rural	3215

```
# Répartition des CM selon la religion
base_CM %>%
  count(religion) %>%
  kable(col.names = c("Religion", "Fréquence"), caption = "Tableau des fréquences de la religion")
```

Table 3: Tableau des fréquences de la religion

Religion	Fréquence
Musulman	6743
Chrétien	398
Animiste	8
Autre Religion	4
Sans Religion	3

```
# Répartition des CM par elon le diplome
base_CM %>%
  count(diplome) %>%
  kable(col.names = c("Diplome", "Fréquence"), caption = "Tableau des fréquences du diplôme")
```

Table 4: Tableau des fréquences du diplôme

Diplome	Fréquence
Aucun	5695
CEP/CFEE	586
BEPC/BFEM	361
CAP	53
BT	16
BAC	154
DEUG, DUT, BTS	49
Licence	83
Maitrise	63
Master/DEA/DESS	65
Doctorat/Phd	31