Test

RASAMOELINA Paulinah

2025-02-25

# Partie 1 : Gestion et nettoyage des bases de données

### Installation et importation des packages

### Dans cette section, nous installons tous les packages qui servirons dans la suite.

packages <- c("readr","cardx","haven","utils","dplyr","gtsummary","gt","labelled","survey","tidyr","ggplot2","naniar")  
  
for (package in packages) {  
 if (!requireNamespace(package, quietly = TRUE)) { # Vérifie si le package n'est pas encore installé  
 install.packages(package)  
 }  
 library(package, character.only = TRUE) # Charge le package  
}  
  
#library(haven) # Pour importer les fichiers .dta  
#library(dplyr) # Pour la manipulation des données  
#library(tidyr) # Pour le nettoyage des données  
#library(gtsummary) # Pour les statistiques descriptives  
#library(ggplot2) # Pour la visualisation des données

### Section I : Importation et analyse des bases

### Importation des bases

## Bases mother\_baseline,mother\_endline,child\_baseline,menage  
  
food\_comp\_mother\_baseline <- haven::read\_dta("../donnees/food\_comp\_mother\_baseline.dta") # Base des mères (Baseline)  
food\_comp\_mother\_endline <- haven::read\_dta("../donnees/food\_comp\_mother\_endline.dta")# Base des mères (Endline)  
food\_comp\_child\_baseline <-haven::read\_dta("../donnees/food\_comp\_child\_baseline.dta") #Base des enfants (Baseline)  
food\_comp\_child\_endline <- haven::read\_dta("../donnees/food\_comp\_child\_endline.dta") # Base des enfants (Endline)  
base\_menage <- haven::read\_dta("../donnees/base\_menage.dta") # Base des ménages

### Structure des bases

utils::str(food\_comp\_mother\_baseline)

## tibble [4,256 × 17] (S3: tbl\_df/tbl/data.frame)  
## $ regionid : num [1:4256] 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "Region ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ communeid : num [1:4256] 25 25 25 25 25 25 25 25 25 25 ...  
## ..- attr(\*, "label")= chr "Commune ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ villageid : num [1:4256] 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 ...  
## ..- attr(\*, "label")= chr "Village ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ hhid : chr [1:4256] "4948484848535052" "4948484848535052" "4948484848535052" "4948484848535052" ...  
## ..- attr(\*, "label")= chr "Household ID"  
## ..- attr(\*, "format.stata")= chr "%45s"  
## $ round : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Survey round : Baseline, Endline"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:2] 1 2  
## .. ..- attr(\*, "names")= chr [1:2] "Baseline" "Endline"  
## $ s1\_q0 : dbl+lbl [1:4256] 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2,...  
## ..@ label : chr "eating occasion"  
## ..@ format.stata: chr "%27.0g"  
## ..@ labels : Named num [1:4] 1 2 3 4  
## .. ..- attr(\*, "names")= chr [1:4] "Breakfast" "Lunch" "Dinner" "Snacks"  
## $ s1\_q1 : dbl+lbl [1:4256] 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Meal consumed? Y/N"  
## ..@ format.stata: chr "%9.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "No" "Yes"  
## $ s1\_q2 : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Who consummed?"  
## ..@ format.stata: chr "%16.0g"  
## ..@ labels : Named num [1:3] 1 2 3  
## .. ..- attr(\*, "names")= chr [1:3] "mother" "child" "mother and child"  
## $ V1 : num [1:4256] 680 NA 634 256 563 ...  
## ..- attr(\*, "label")= chr "Consommation en ernergie (kcal)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ protein\_g : num [1:4256] 23.31 NA 21.62 8.25 11.7 ...  
## ..- attr(\*, "label")= chr "Consommation en proteine (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ lipid\_tot\_g: num [1:4256] 5.3 NA 4.87 8.06 6.49 ...  
## ..- attr(\*, "label")= chr "Consommation en lipide (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ calcium\_mg : num [1:4256] 62.7 NA 57.1 22 116.3 ...  
## ..- attr(\*, "label")= chr "Consommation en calcium (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ iron\_mg : num [1:4256] 10.591 NA 9.897 0.912 2.716 ...  
## ..- attr(\*, "label")= chr "Consommation en fer (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ V9 : num [1:4256] 4.507 NA 4.19 0.456 3.382 ...  
## ..- attr(\*, "label")= chr "Consommation en zinc (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b6\_mg : num [1:4256] 0.3058 NA 0.2835 0.0456 0.2248 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B6 (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b12\_mcg: num [1:4256] 0.00869 NA 0.0078 0 0.00823 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B12 (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_c\_mg : num [1:4256] 0.0441 NA 0.0396 0 0.0002 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine C (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"

utils::str(food\_comp\_mother\_endline)

## tibble [4,256 × 17] (S3: tbl\_df/tbl/data.frame)  
## $ regionid : num [1:4256] 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "Region ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ communeid : num [1:4256] 25 25 25 25 25 25 25 25 25 25 ...  
## ..- attr(\*, "label")= chr "Commune ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ villageid : num [1:4256] 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 ...  
## ..- attr(\*, "label")= chr "Village ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ hhid : chr [1:4256] "4948484848535052" "4948484848535052" "4948484848535052" "4948484848535052" ...  
## ..- attr(\*, "label")= chr "Household ID"  
## ..- attr(\*, "format.stata")= chr "%45s"  
## $ round : dbl+lbl [1:4256] 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,...  
## ..@ label : chr "Survey round : Baseline, Endline"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:2] 1 2  
## .. ..- attr(\*, "names")= chr [1:2] "Baseline" "Endline"  
## $ s1\_q0 : dbl+lbl [1:4256] 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2,...  
## ..@ label : chr "eating occasion"  
## ..@ format.stata: chr "%27.0g"  
## ..@ labels : Named num [1:4] 1 2 3 4  
## .. ..- attr(\*, "names")= chr [1:4] "Breakfast" "Lunch" "Dinner" "Snacks"  
## $ s1\_q1 : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1,...  
## ..@ label : chr "Meal consumed? Y/N"  
## ..@ format.stata: chr "%9.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "No" "Yes"  
## $ s1\_q2 : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Who consummed?"  
## ..@ format.stata: chr "%16.0g"  
## ..@ labels : Named num [1:3] 1 2 3  
## .. ..- attr(\*, "names")= chr [1:3] "mother" "child" "mother and child"  
## $ energ\_kcal : num [1:4256] 2217 1021 3038 716 618 ...  
## ..- attr(\*, "label")= chr "Consommation en ernergie (kcal)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ protein\_g : num [1:4256] 48.7 34.8 92 19.1 16.5 ...  
## ..- attr(\*, "label")= chr "Consommation en proteine (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ lipid\_tot\_g: num [1:4256] 89.6 5.94 183.6 4.73 1.63 ...  
## ..- attr(\*, "label")= chr "Consommation en lipide (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ calcium\_mg : num [1:4256] 818.8 72.1 1781.3 131.6 10.8 ...  
## ..- attr(\*, "label")= chr "Consommation en calcium (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ iron\_mg : num [1:4256] 22.659 16.558 11.345 8.877 0.105 ...  
## ..- attr(\*, "label")= chr "Consommation en fer (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ zinc\_mg : num [1:4256] 9.5691 6.4618 2.2278 3.5867 0.0334 ...  
## ..- attr(\*, "label")= chr "Consommation en zinc (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b6\_mg : num [1:4256] 0.8748 0.4327 0.314 0.5277 0.0059 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B6 (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b12\_mcg: num [1:4256] 0.00948 0.02232 0 0.01281 0.01016 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B12 (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_c\_mg : num [1:4256] 11.547 0.183 26.738 0.133 0 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine C (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"

utils::str(food\_comp\_child\_baseline)

## tibble [4,256 × 17] (S3: tbl\_df/tbl/data.frame)  
## $ regionid : num [1:4256] 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "Region ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ communeid : num [1:4256] 25 25 25 25 25 25 25 25 25 25 ...  
## ..- attr(\*, "label")= chr "Commune ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ villageid : num [1:4256] 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 ...  
## ..- attr(\*, "label")= chr "Village ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ hhid : chr [1:4256] "4948484848535052" "4948484848535052" "4948484848535052" "4948484848535052" ...  
## ..- attr(\*, "label")= chr "Household ID"  
## ..- attr(\*, "format.stata")= chr "%45s"  
## $ round : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Survey round : Baseline, Endline"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:2] 1 2  
## .. ..- attr(\*, "names")= chr [1:2] "Baseline" "Endline"  
## $ s1\_q0 : dbl+lbl [1:4256] 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2,...  
## ..@ label : chr "eating occasion"  
## ..@ format.stata: chr "%27.0g"  
## ..@ labels : Named num [1:4] 1 2 3 4  
## .. ..- attr(\*, "names")= chr [1:4] "Breakfast" "Lunch" "Dinner" "Snacks"  
## $ s1\_q1 : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Meal consumed? Y/N"  
## ..@ format.stata: chr "%9.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "No" "Yes"  
## $ s1\_q2 : dbl+lbl [1:4256] 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,...  
## ..@ label : chr "Who consummed?"  
## ..@ format.stata: chr "%16.0g"  
## ..@ labels : Named num [1:3] 1 2 3  
## .. ..- attr(\*, "names")= chr [1:3] "mother" "child" "mother and child"  
## $ energ\_kcal : num [1:4256] 355 224 334 494 235 ...  
## ..- attr(\*, "label")= chr "Consommation en ernergie (kcal)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ protein\_g : num [1:4256] 12.74 8.01 12.72 16.65 4.45 ...  
## ..- attr(\*, "label")= chr "Consommation en proteine (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ lipid\_tot\_g: num [1:4256] 3.14 1.96 3.42 8.84 4.76 ...  
## ..- attr(\*, "label")= chr "Consommation en lipide (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ calcium\_mg : num [1:4256] 40.1 24.8 47 60.1 93.3 ...  
## ..- attr(\*, "label")= chr "Consommation en calcium (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ iron\_mg : num [1:4256] 5.39 3.42 4.9 6.83 1.54 ...  
## ..- attr(\*, "label")= chr "Consommation en fer (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ zinc\_mg : num [1:4256] 2.41 1.52 2.35 3.13 1.63 ...  
## ..- attr(\*, "label")= chr "Consommation en zinc (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b6\_mg : num [1:4256] 0.167 0.105 0.168 0.207 0.116 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B6 (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b12\_mcg: num [1:4256] 0.00616 0.00377 0.00785 0.00959 0.00688 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B12 (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_c\_mg : num [1:4256] 0.031282 0.019173 0.03986 0.048689 0.000167 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine C (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"

utils::str(food\_comp\_child\_endline)

## tibble [4,256 × 17] (S3: tbl\_df/tbl/data.frame)  
## $ regionid : num [1:4256] 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "Region ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ communeid : num [1:4256] 25 25 25 25 25 25 25 25 25 25 ...  
## ..- attr(\*, "label")= chr "Commune ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ villageid : num [1:4256] 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 ...  
## ..- attr(\*, "label")= chr "Village ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ hhid : chr [1:4256] "4948484848535052" "4948484848535052" "4948484848535052" "4948484848535052" ...  
## ..- attr(\*, "label")= chr "Household ID"  
## ..- attr(\*, "format.stata")= chr "%45s"  
## $ round : dbl+lbl [1:4256] 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,...  
## ..@ label : chr "Survey round : Baseline, Endline"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:2] 1 2  
## .. ..- attr(\*, "names")= chr [1:2] "Baseline" "Endline"  
## $ s1\_q0 : dbl+lbl [1:4256] 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2,...  
## ..@ label : chr "eating occasion"  
## ..@ format.stata: chr "%27.0g"  
## ..@ labels : Named num [1:4] 1 2 3 4  
## .. ..- attr(\*, "names")= chr [1:4] "Breakfast" "Lunch" "Dinner" "Snacks"  
## $ s1\_q1 : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Meal consumed? Y/N"  
## ..@ format.stata: chr "%9.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "No" "Yes"  
## $ s1\_q2 : dbl+lbl [1:4256] 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,...  
## ..@ label : chr "Who consummed?"  
## ..@ format.stata: chr "%16.0g"  
## ..@ labels : Named num [1:3] 1 2 3  
## .. ..- attr(\*, "names")= chr [1:3] "mother" "child" "mother and child"  
## $ energ\_kcal : num [1:4256] 1193 535 1701 531 188 ...  
## ..- attr(\*, "label")= chr "Consommation en ernergie (kcal)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ protein\_g : num [1:4256] 24.94 18.38 58.84 13.89 5.04 ...  
## ..- attr(\*, "label")= chr "Consommation en proteine (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ lipid\_tot\_g: num [1:4256] 51.915 3.205 90.268 3.917 0.496 ...  
## ..- attr(\*, "label")= chr "Consommation en lipide (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ calcium\_mg : num [1:4256] 479.73 41.32 1122.63 74.44 3.29 ...  
## ..- attr(\*, "label")= chr "Consommation en calcium (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ iron\_mg : num [1:4256] 11.5857 8.709 7.2277 6.4328 0.0321 ...  
## ..- attr(\*, "label")= chr "Consommation en fer (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ zinc\_mg : num [1:4256] 4.9181 3.3562 1.4184 2.7147 0.0102 ...  
## ..- attr(\*, "label")= chr "Consommation en zinc (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b6\_mg : num [1:4256] 0.4678 0.2251 0.2028 0.3523 0.0018 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B6 (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b12\_mcg: num [1:4256] 0.00396 0.01364 0 0.00736 0.00309 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B12 (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_c\_mg : num [1:4256] 6.8299 0.1201 16.904 0.0764 0 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine C (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"

utils::str(base\_menage)

## tibble [1,065 × 21] (S3: tbl\_df/tbl/data.frame)  
## $ regionid : num [1:1065] 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "Region ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ communeid : num [1:1065] 25 25 25 25 25 25 25 25 25 25 ...  
## ..- attr(\*, "label")= chr "Commune ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ villageid : num [1:1065] 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 ...  
## ..- attr(\*, "label")= chr "Village ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ hhid : chr [1:1065] "4948484848535052" "4948484848535053" "4948484848535055" "4948484848535056" ...  
## ..- attr(\*, "label")= chr "Household ID"  
## ..- attr(\*, "format.stata")= chr "%45s"  
## $ hhsize : num [1:1065] 4 8 11 9 16 6 31 8 23 5 ...  
## ..- attr(\*, "label")= chr "Household size"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ poly : dbl+lbl [1:1065] 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0,...  
## ..@ label : chr "Polygamous household?"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ hh\_primary : dbl+lbl [1:1065] 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...  
## ..@ label : chr "Household head completed primary education"  
## ..@ format.stata: chr "%9.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ s1\_q2 : dbl+lbl [1:1065] 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Male head of household"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Female" "Male"  
## $ s1\_q4a : num [1:1065] 37 52 67 38 75 41 56 52 63 26 ...  
## ..- attr(\*, "label")= chr "Age head of household"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ s2\_q1 : dbl+lbl [1:1065] 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0,...  
## ..@ label : chr "Is head of household literate in local language"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ s2\_q2 : dbl+lbl [1:1065] 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...  
## ..@ label : chr "Is head of household literate in French?"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ s2\_q4 : dbl+lbl [1:1065] 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0,...  
## ..@ label : chr "Has head of household been to school (formal or informal)?"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ s29\_q1 : dbl+lbl [1:1065] 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0,...  
## ..@ label : chr "Est-ce qu'un membre de votre ménage a pris un prêt ou fait un emprunt en argent"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ demgrp1 : num [1:1065] 1 0 0 0 1 0 5 1 1 1 ...  
## ..- attr(\*, "label")= chr "Number of children 0-36 months"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ demgrp2 : num [1:1065] 1 2 2 1 1 1 5 0 2 1 ...  
## ..- attr(\*, "label")= chr "Number of children 36-72 months"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ demgrp3 : num [1:1065] 0 3 3 3 7 2 11 4 9 1 ...  
## ..- attr(\*, "label")= chr "Number of adults 6-14 years"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ demgrp4 : num [1:1065] 1 2 5 5 6 3 10 3 11 2 ...  
## ..- attr(\*, "label")= chr "Number of adults 14-65 years"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ demgrp5 : num [1:1065] 0 1 1 0 1 0 0 0 0 0 ...  
## ..- attr(\*, "label")= chr "Number of elders 65+ years"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ dependencyratio: num [1:1065] 2 3 1.2 0.8 1.67 ...  
## ..- attr(\*, "label")= chr "Dependency ratio"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ hfias\_score : num [1:1065] 18 21 0 4 1 3 12 0 14 0 ...  
## ..- attr(\*, "label")= chr "HFIAS Score (0-27)"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ T1 : dbl+lbl [1:1065] 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Treatment: 1st level comparison"  
## ..@ format.stata: chr "%12.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Controle" "Intervention"  
## - attr(\*, "label")= chr "One row per household"

## 1.Vérification et suppression des doublons

cat("Nombre de doublons dans mother\_baseline:", nrow(food\_comp\_mother\_baseline) - nrow(distinct(food\_comp\_mother\_baseline)), "\n")

## Nombre de doublons dans mother\_baseline: 0

food\_comp\_mother\_baseline <- distinct(food\_comp\_mother\_baseline)   
  
cat("Nombre de doublons dans mother\_endline:", nrow(food\_comp\_mother\_endline) - nrow(distinct(food\_comp\_mother\_endline)), "\n")

## Nombre de doublons dans mother\_endline: 0

food\_comp\_mother\_endline <- distinct(food\_comp\_mother\_endline)   
  
cat("Nombre de doublons dans child\_baseline:", nrow(food\_comp\_child\_baseline) - nrow(distinct(food\_comp\_child\_baseline)), "\n")

## Nombre de doublons dans child\_baseline: 0

food\_comp\_child\_baseline <- distinct(food\_comp\_child\_baseline)   
  
cat("Nombre de doublons dans child\_endline:", nrow(food\_comp\_child\_endline) - nrow(distinct(food\_comp\_child\_endline)), "\n")

## Nombre de doublons dans child\_endline: 0

food\_comp\_child\_endline <- distinct(food\_comp\_child\_endline)   
  
cat("Nombre de doublons dans menage:", nrow(base\_menage) - nrow(distinct(base\_menage)), "\n")

## Nombre de doublons dans menage: 1

base\_menage <- distinct(base\_menage)

base\_menage <- base\_menage[!duplicated(base\_menage), ]  
cat("Nombre de doublons dans menage:", nrow(base\_menage) - nrow(distinct(base\_menage)), "\n")

## Nombre de doublons dans menage: 0

base\_menage <- distinct(base\_menage) # Suppression des doublons

## 2.Harmonisation des noms des variables

### 2.a. Vérification des noms de variables pour garantir la cohérence entre baseline et endline

colnames(food\_comp\_mother\_baseline)

## [1] "regionid" "communeid" "villageid" "hhid" "round"   
## [6] "s1\_q0" "s1\_q1" "s1\_q2" "V1" "protein\_g"   
## [11] "lipid\_tot\_g" "calcium\_mg" "iron\_mg" "V9" "vit\_b6\_mg"   
## [16] "vit\_b12\_mcg" "vit\_c\_mg"

colnames(food\_comp\_mother\_endline)

## [1] "regionid" "communeid" "villageid" "hhid" "round"   
## [6] "s1\_q0" "s1\_q1" "s1\_q2" "energ\_kcal" "protein\_g"   
## [11] "lipid\_tot\_g" "calcium\_mg" "iron\_mg" "zinc\_mg" "vit\_b6\_mg"   
## [16] "vit\_b12\_mcg" "vit\_c\_mg"

colnames(food\_comp\_child\_baseline)

## [1] "regionid" "communeid" "villageid" "hhid" "round"   
## [6] "s1\_q0" "s1\_q1" "s1\_q2" "energ\_kcal" "protein\_g"   
## [11] "lipid\_tot\_g" "calcium\_mg" "iron\_mg" "zinc\_mg" "vit\_b6\_mg"   
## [16] "vit\_b12\_mcg" "vit\_c\_mg"

colnames(food\_comp\_child\_endline)

## [1] "regionid" "communeid" "villageid" "hhid" "round"   
## [6] "s1\_q0" "s1\_q1" "s1\_q2" "energ\_kcal" "protein\_g"   
## [11] "lipid\_tot\_g" "calcium\_mg" "iron\_mg" "zinc\_mg" "vit\_b6\_mg"   
## [16] "vit\_b12\_mcg" "vit\_c\_mg"

names(food\_comp\_mother\_endline) <- names(food\_comp\_mother\_baseline) # Harmonisation des noms des colonnes  
names(food\_comp\_child\_endline) <- names(food\_comp\_child\_baseline) # Harmonisation des noms des colonnes

## 2.b. Nom des colonnes dans la base menage

colnames(base\_menage)

## [1] "regionid" "communeid" "villageid" "hhid"   
## [5] "hhsize" "poly" "hh\_primary" "s1\_q2"   
## [9] "s1\_q4a" "s2\_q1" "s2\_q2" "s2\_q4"   
## [13] "s29\_q1" "demgrp1" "demgrp2" "demgrp3"   
## [17] "demgrp4" "demgrp5" "dependencyratio" "hfias\_score"   
## [21] "T1"

## 3. Gestion des valeurs manquantes

val\_manq1<-sum(is.na(food\_comp\_mother\_baseline))  
print(val\_manq1)

## [1] 6069

val\_manq2<-sum(is.na(food\_comp\_mother\_endline))  
print(val\_manq2)

## [1] 5976

val\_manq3<-sum(is.na(food\_comp\_child\_baseline))  
print(val\_manq3)

## [1] 3474

val\_manq4<-sum(is.na(food\_comp\_child\_endline))  
print(val\_manq4)

## [1] 4788

val\_manq5<-sum(is.na(base\_menage))  
print(val\_manq5)

## [1] 1

# Correction des données manquantes si possible  
# Noms remplacerons les NA par la moyenne pour les variables numériques dans chaquebase de donnée  
food\_comp\_mother\_baseline <- food\_comp\_mother\_baseline %>%  
 mutate(across(where(is.numeric), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))  
  
food\_comp\_mother\_endline <- food\_comp\_mother\_endline %>%  
 mutate(across(where(is.numeric), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))  
  
food\_comp\_child\_baseline <- food\_comp\_child\_baseline %>%  
 mutate(across(where(is.numeric), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))  
  
food\_comp\_child\_endline <- food\_comp\_child\_endline %>%  
 mutate(across(where(is.numeric), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))  
  
base\_menage <- base\_menage %>%  
 mutate(across(where(is.numeric), ~ ifelse(is.na(.), mean(., na.rm=TRUE),.)))

## 4.Consommation d’énergie moyenne pour chaque repas

# Calculer la moyenne de energ\_kcal par groupe de s1\_q0  
consommation\_moyenne\_par\_groupe <- food\_comp\_child\_baseline %>%  
 group\_by(s1\_q0) %>%  
 summarise(consommation\_moyenne = mean(energ\_kcal, na.rm = TRUE))  
  
# Afficher le résultat  
print(consommation\_moyenne\_par\_groupe)

## # A tibble: 4 × 2  
## s1\_q0 consommation\_moyenne  
## <dbl> <dbl>  
## 1 1 354.  
## 2 2 415.  
## 3 3 398.  
## 4 4 483.

## 5. Sauvegarde des bases

# 1. Charger et sauvegarder la base food\_comp\_mother\_baseline.dta en mother\_baseline\_v1.dta  
# Charger la base de données food\_comp\_mother\_baseline.dta  
mother\_baseline <- food\_comp\_mother\_baseline  
  
# Effectuer toutes les corrections et vérifications nécessaires dans la base (assurées précédemment)  
  
# Sauvegarder la base corrigée sous un nouveau nom "mother\_baseline\_v1.dta"  
write\_dta(mother\_baseline, "mother\_baseline\_v1.dta")  
# Message pour indiquer que le fichier a été sauvegardé  
cat("La base mother\_baseline a été sauvegardée sous le nom 'mother\_baseline\_v1.dta'.\n")

## La base mother\_baseline a été sauvegardée sous le nom 'mother\_baseline\_v1.dta'.

# 2. Charger et sauvegarder la base food\_comp\_mother\_endline.dta en mother\_endline\_v1.dta  
# Charger la base de données food\_comp\_mother\_endline.dta  
mother\_endline <- food\_comp\_mother\_endline  
  
# Effectuer toutes les corrections et vérifications nécessaires dans la base (assurées précédemment)  
  
# Sauvegarder la base corrigée sous un nouveau nom "mother\_endline\_v1.dta"  
write\_dta(mother\_endline, "mother\_endline\_v1.dta")  
# Message pour indiquer que le fichier a été sauvegardé  
cat("La base mother\_endline a été sauvegardée sous le nom 'mother\_endline\_v1.dta'.\n")

## La base mother\_endline a été sauvegardée sous le nom 'mother\_endline\_v1.dta'.

# 3. Charger et sauvegarder la base food\_comp\_child\_baseline.dta en child\_baseline\_v1.dta  
# Charger la base de données food\_comp\_child\_baseline.dta  
child\_baseline <- food\_comp\_child\_baseline  
  
# Effectuer toutes les corrections et vérifications nécessaires dans la base (assurées précédemment)  
  
# Sauvegarder la base corrigée sous un nouveau nom "child\_baseline\_v1.dta"  
write\_dta(child\_baseline, "child\_baseline\_v1.dta")  
# Message pour indiquer que le fichier a été sauvegardé  
cat("La base child\_baseline a été sauvegardée sous le nom 'child\_baseline\_v1.dta'.\n")

## La base child\_baseline a été sauvegardée sous le nom 'child\_baseline\_v1.dta'.

# 4. Charger et sauvegarder la base food\_comp\_child\_endline.dta en child\_endline\_v1.dta  
# Charger la base de données food\_comp\_child\_endline.dta  
child\_endline <- food\_comp\_child\_endline  
  
# Effectuer toutes les corrections et vérifications nécessaires dans la base (assurées précédemment)  
  
# Sauvegarder la base corrigée sous un nouveau nom "child\_endline\_v1.dta"  
write\_dta(child\_endline, "child\_endline\_v1.dta")  
# Message pour indiquer que le fichier a été sauvegardé  
cat("La base child\_endline a été sauvegardée sous le nom 'child\_endline\_v1.dta'.\n")

## La base child\_endline a été sauvegardée sous le nom 'child\_endline\_v1.dta'.

# 5. Charger et sauvegarder la base base\_menage.dta en base\_menage\_final.dta  
# Charger la base de données base\_menage.dta  
base\_menage <- base\_menage  
  
# Effectuer toutes les corrections et vérifications nécessaires dans la base (assurées précédemment)  
  
# Sauvegarder la base corrigée sous un nouveau nom "base\_menage\_final.dta"  
write\_dta(base\_menage, "base\_menage\_final.dta")  
# Message pour indiquer que le fichier a été sauvegardé  
cat("La base base\_menage a été sauvegardée sous le nom 'base\_menage\_final.dta'.\n")

## La base base\_menage a été sauvegardée sous le nom 'base\_menage\_final.dta'.

# Fin du processus de sauvegarde  
cat("Tous les fichiers ont été sauvegardés avec succès.\n")

## Tous les fichiers ont été sauvegardés avec succès.

# Partie 2

mother\_endline\_v1 <- haven::read\_dta("mother\_endline\_v1.dta")  
child\_endline\_v1 <- haven::read\_dta("child\_endline\_v1.dta")

## 1. Empilement et Fusion des données (Baseline)

mother\_chil\_endline\_v1 <- bind\_rows(mother\_endline\_v1,child\_endline\_v1) %>% # empiler les deux bases   
 arrange(hhid,s1\_q2) # trié suivant le numéro du ménage et suivant la personne qui a consommé  
  
# Renommer les colonnes de consommation d'énergie et nutriments en ajoutant "\_b"  
  
mother\_chil\_endline\_v1 <- mother\_chil\_endline\_v1 %>%  
 rename\_with(~paste0(., "\_e"), # Renommer le nom de la variable par le nom concatener avec \_b  
 .cols = energ\_kcal:vit\_c\_mg) # spécifie les variables concernées par l'opération

### Créez une base de données qui résume les consommations journalières totales par individu (somme des 4 repas) pour l’énergie et tous les nutriments en utilisant la commande collapse.

# Calculer la somme pour chaque variable entre energ\_kcal\_b et vit\_c\_mg\_b  
base <- mother\_chil\_endline\_v1 %>%  
 group\_by(hhid, s1\_q2) %>% ## Grouper suivant le numér du ménage et suivant le fait que ça soit mere ou enfant  
 summarise(across(energ\_kcal\_e:vit\_c\_mg\_e, \(x) sum(x, na.rm = TRUE)), .groups = "drop") # calculons la somme des consommation par individu et en ignorant les valeurs manquantes  
  
base # Afficher la base

## # A tibble: 2,128 × 4  
## hhid s1\_q2 energ\_kcal\_e vit\_c\_mg\_e  
## <chr> <dbl> <dbl> <dbl>  
## 1 4948484848535052 1 0 38.6   
## 2 4948484848535052 2 3960. 23.9   
## 3 4948484848535053 1 0 16.1   
## 4 4948484848535053 2 1237. 12.9   
## 5 4948484848535055 1 0 38.8   
## 6 4948484848535055 2 3149. 29.7   
## 7 4948484848535056 1 0 9.73  
## 8 4948484848535056 2 3873. 1.25  
## 9 4948484848535151 1 0 218.   
## 10 4948484848535151 2 1208. 87.0   
## # ℹ 2,118 more rows

# Renommer les colonnes en supprimant le suffixe "\_e"  
names(base) <- gsub("\_e$", "", names(base))  
write\_dta(base, "base.dta")  
# Afficher les nouveaux noms des colonnes  
names(base)

## [1] "hhid" "s1\_q2" "energ\_kcal" "vit\_c\_mg"

colnames(base)

## [1] "hhid" "s1\_q2" "energ\_kcal" "vit\_c\_mg"

### Sauvegarder la base de données finale contenant seulement les hhid s1\_q2 energ\_kcal protein\_g lipid\_tot\_g calcium\_mg iron\_mg zinc\_mg sous le nom « endline\_final.dta ». Veuillez noter que pour cette base de données, pour chaque ménage, nous avons une ligne pour la mère et une ligne pour l’enfant.

# Charger les bases de données baseline et endline  
endline\_final <- read\_dta("endline\_final.dta")  
# Vérifier les colonnes avant la fusion  
  
colnames(endline\_final)

## [1] "hhid" "s1\_q2" "energ\_kcal\_tot\_e" "protein\_tot\_e"   
## [5] "lipid\_tot\_e" "calcium\_tot\_e" "iron\_tot\_e" "zinc\_tot\_e"