welfare merger

Dior MBENGUE

2025-03-20

# vérification et installation des packages

#vérifier et installer les packages nécessaires  
packages <- c("haven", "utils", "dplyr", "gtsummary", "survey","labelled","compareDF")  
for (pkg in packages) {  
 if (!require(pkg, character.only = TRUE)) install.packages(pkg, dependencies = TRUE)  
 library(pkg, character.only = TRUE)  
}

## Le chargement a nécessité le package : haven

## Le chargement a nécessité le package : dplyr

##   
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':  
##   
## filter, lag

## Les objets suivants sont masqués depuis 'package:base':  
##   
## intersect, setdiff, setequal, union

## Le chargement a nécessité le package : gtsummary

## Le chargement a nécessité le package : survey

## Le chargement a nécessité le package : grid

## Le chargement a nécessité le package : Matrix

## Le chargement a nécessité le package : survival

##   
## Attachement du package : 'survey'

## L'objet suivant est masqué depuis 'package:graphics':  
##   
## dotchart

## Le chargement a nécessité le package : labelled

## Le chargement a nécessité le package : compareDF

## Warning: le package 'compareDF' a été compilé avec la version R 4.4.3

#chargement des bases welfare 2021 et 2018

#chargement de la base welfare 2021  
welfare21<-haven::read\_dta("../données/ehcvm\_welfare\_sen2021.dta")  
#chargement de la base welfare 2018  
welfare18<-haven::read\_dta("../données/ehcvm\_welfare\_sen2018.dta")

# pour faire une bonne merge, il est important d’analyser les données de chaque base, de vérifier s’il ya similitudes et de divergences entre les variables

##Vérification s'il ya des doublons  
doublon18<-welfare18[duplicated(welfare18[, c("grappe","menage")]), ]  
doublon21<-welfare21[duplicated(welfare21[, c("grappe","menage")]), ]  
  
cat("Nombre de doublons présents dans la base welfare 2018 :", nrow(doublon18), "\n")

## Nombre de doublons présents dans la base welfare 2018 : 0

cat("Nombre de doublons présents dans la base welfare 2021 :", nrow(doublon21))

## Nombre de doublons présents dans la base welfare 2021 : 0

Il n’y a pas de doublons dans les deux bases

##Vérification des valeurs manquantes  
print("Valeurs manquante base 2018")

## [1] "Valeurs manquante base 2018"

NA18 <- colSums(is.na(welfare18))  
NA18

## country year hhid grappe menage vague zae region   
## 0 0 0 0 0 0 0 0   
## milieu hhweight hhsize eqadu1 eqadu2 hgender hage hmstat   
## 0 0 0 0 0 0 0 2   
## hreligion hnation halfab heduc hdiploma hhandig hactiv7j hactiv12m   
## 0 0 0 0 0 0 0 0   
## hbranch hsectins hcsp dali dnal dtot pcexp zzae   
## 1722 1722 1722 0 0 0 0 0   
## zref def\_spa def\_temp   
## 0 0 0

print("Valeurs manquante base 2021")

## [1] "Valeurs manquante base 2021"

NA21 <- colSums(is.na(welfare21))  
NA21

## grappe menage country   
## 0 0 0   
## year hhid vague   
## 0 0 0   
## month zae region   
## 0 0 0   
## milieu hhweight hhsize   
## 0 0 0   
## eqadu1 eqadu2 hgender   
## 0 0 0   
## hage hmstat hreligion   
## 0 0 0   
## hnation hethnie halfa   
## 0 82 0   
## halfa2 heduc hdiploma   
## 0 0 0   
## hhandig hactiv7j hactiv12m   
## 0 0 0   
## hbranch hsectins hcsp   
## 1838 1359 1326   
## dali dnal dtot   
## 0 0 0   
## pcexp zzae zref   
## 0 0 0   
## def\_spa def\_temp def\_temp\_prix2021m11   
## 0 0 0   
## def\_temp\_cpi def\_temp\_adj zali0   
## 0 0 0   
## dtet monthly\_cpi cpi2017   
## 0 0 0   
## icp2017 dollars   
## 0 0

le nombre de variables ayant valeurs manquantes dans les deux bases n’est pas trés grand Nous en avons 2 variables qui présentent de valeurs manquantes en 2021 contre une seule en 2018

##Présence des variables dans les deux bases ou dans une seule  
vars18 = colnames(welfare18)  
  
vars21 = colnames(welfare21)   
  
# Identifions les variables communes  
vars\_communes <- intersect(vars18, vars21)   
  
# Variables présentes uniquement en 2018  
vars\_seulement\_18 <- setdiff(vars18, vars21)   
  
# Variables présentes uniquement en 2021  
vars\_seulement\_21 <- setdiff(vars21, vars18)

Sachant que halpha et halphab désigne la meme variable, renommons celle de 2021 en halfab.

# Renommons halfa en halfab  
  
colnames(welfare21)[colnames(welfare21)=="halfa"] <- "halfab"  
  
# Ajoutons cette variable à la liste des variables communes.  
  
vars\_communes <- append(vars\_communes,"halfab")  
utils::View(welfare21)

# Maintenant, nous passons à la labellisation des bases

nous allons vérifier si les variables communes ont la meme labellisation puis nous allons vérifier s’il ya incohérence ou pas

variable\_label\_diff <- c()   
  
for (variable in vars\_communes) {   
   
 if(labelled::is.labelled(welfare18[[variable]])){   
   
 value\_label18 <- labelled::val\_labels(welfare18[[variable]])   
  
 }else{  
 value\_label18 <- NULL   
 }  
   
 if(labelled::is.labelled(welfare21[[variable]])){   
   
 value\_label21 <- labelled::val\_labels(welfare21[[variable]])   
   
 }else{  
   
 value\_label21 <- NULL   
 }  
   
 if(!identical(value\_label18, value\_label21)){   
 variable\_label\_diff <- append(variable\_label\_diff,variable)   
 print(variable)  
   
 }  
}

## [1] "zae"  
## [1] "hnation"  
## [1] "hdiploma"  
## [1] "hactiv7j"  
## [1] "hbranch"  
## [1] "hcsp"

Aprés avoir détecter les variables avec label différents dans les deux bases, nous allons voir si nous allons détecter une incohérence aprés les avoir afficher

nous allons faire la visualiation de la distribution puis la détection de l’incohérence,puis la correction de l’incohérence et enfin la revisualisation pour la confirmation de la correction

#nous allons commencer par la variable”hnation”

#   
labelled::val\_labels(welfare18$hnation)

## Benin Burkina Faso Côte d'Ivoire   
## 1 2 3   
## Guinée Bissau Mali Niger   
## 4 5 6   
## Sénégal Togo Nigéria   
## 7 8 9   
## Autre CEDEAO Autre Afrique Autre pays hors Afrique   
## 10 11 12

labelled::val\_labels(welfare21$hnation)

## Bénin Burkina Faso Cape-vert   
## 1 2 3   
## Cote d'ivoire Gambie Ghana   
## 4 5 6   
## Guinee Guinée Bissau Liberia   
## 7 8 9   
## Mali Niger Nigeria   
## 10 11 12   
## Sénégal Serra-Leonne Togo   
## 13 14 15   
## Autre Afrique Autre pays hors Afrique   
## 17 18

welfare21 %>%  
 to\_factor() %>%   
 select(hnation) %>%   
 tbl\_summary(missing = "always",   
 missing\_text = "Valeurs manquantes")%>%  
modify\_header(label="Les caractéristiques de la nation du chef de ménage")

| Les caractéristiques de la nation du chef de ménage | **N = 7,120***1* |
| --- | --- |
| Nationalite du CM |  |
| Bénin | 0 (0%) |
| Burkina Faso | 0 (0%) |
| Cape-vert | 0 (0%) |
| Cote d'ivoire | 1 (<0.1%) |
| Gambie | 2 (<0.1%) |
| Ghana | 1 (<0.1%) |
| Guinee | 39 (0.5%) |
| Guinée Bissau | 8 (0.1%) |
| Liberia | 0 (0%) |
| Mali | 18 (0.3%) |
| Niger | 2 (<0.1%) |
| Nigeria | 1 (<0.1%) |
| Sénégal | 7,038 (99%) |
| Serra-Leonne | 0 (0%) |
| Togo | 1 (<0.1%) |
| Autre Afrique | 7 (<0.1%) |
| Autre pays hors Afrique | 2 (<0.1%) |
| Valeurs manquantes | 0 |
| *1*n (%) | |

welfare21 <- welfare21 %>%  
 mutate(hnation = dplyr::recode(hnation,  
 `4` = 3,   
 `8` = 4,   
 `10` = 5,   
 `11` = 6,   
 `13` = 7,   
 `15` = 8,  
 `12` = 9,   
 `17` = 11,   
 `18` = 12,   
 `3` = 10, `5` = 10, `6` = 10, `7` = 10, `9` = 10, `14` = 10   
 ))  
# nous avons détecté l'erreur qui concernait la labellisation, on a décidé de recoder les modalités et on a 10 pour les pays de l'UEMOA et on a recodé les autres pays aussi pour enlever l'incohérence

Affectons les labels de 2018 à ceux de 2021

labelled::val\_labels(welfare21$hnation) <- labelled::val\_labels(welfare18$hnation)  
  
#Nous avons affecté les labels de 2018 à ceux de 2021, possible grace au recodage de variable qui donne les memes codes

#Revisualisation

#Vérification  
welfare21 %>%  
 to\_factor() %>%   
 select(hnation) %>%   
 tbl\_summary(missing = "always",   
 missing\_text = "Valeurs manquantes")%>%  
modify\_header(label="Les caractéristiques de la nation du chef de ménage")

| Les caractéristiques de la nation du chef de ménage | **N = 7,120***1* |
| --- | --- |
| Nationalite du CM |  |
| Benin | 0 (0%) |
| Burkina Faso | 0 (0%) |
| Côte d'Ivoire | 1 (<0.1%) |
| Guinée Bissau | 8 (0.1%) |
| Mali | 18 (0.3%) |
| Niger | 2 (<0.1%) |
| Sénégal | 7,038 (99%) |
| Togo | 1 (<0.1%) |
| Nigéria | 1 (<0.1%) |
| Autre CEDEAO | 42 (0.6%) |
| Autre Afrique | 7 (<0.1%) |
| Autre pays hors Afrique | 2 (<0.1%) |
| Valeurs manquantes | 0 |
| *1*n (%) | |

# la variable “hdiploma”

val\_labels(welfare18$hdiploma)

## Aucun CEP/CFEE BEPC/BFEM cap bt   
## 0 1 2 3 4   
## bac DEUG, DUT, BTS Licence Maitrise Master/DEA/DESS   
## 5 6 7 8 9   
## Doctorat/Phd   
## 10

val\_labels(welfare21$hdiploma)

## Aucun cepe bepc cap bt   
## 0 1 2 3 4   
## bac DEUG, DUT, BTS Licence Maitrise Master/DEA/DESS   
## 5 6 7 8 9   
## Doctorat/Phd   
## 10

l’incohérence vient du codage des différents diplomes

welfare18 %>%   
 to\_factor() %>%   
 select(hdiploma) %>%   
 tbl\_summary(missing = "always",  
 missing\_text = "NA")%>%  
modify\_header(label="Les caractéristiques du diplome du chef de ménage")

| Les caractéristiques du diplome du chef de ménage | **N = 7,156***1* |
| --- | --- |
| Diplome du CM |  |
| Aucun | 5,697 (80%) |
| CEP/CFEE | 587 (8.2%) |
| BEPC/BFEM | 359 (5.0%) |
| cap | 52 (0.7%) |
| bt | 16 (0.2%) |
| bac | 154 (2.2%) |
| DEUG, DUT, BTS | 49 (0.7%) |
| Licence | 83 (1.2%) |
| Maitrise | 63 (0.9%) |
| Master/DEA/DESS | 65 (0.9%) |
| Doctorat/Phd | 31 (0.4%) |
| NA | 0 |
| *1*n (%) | |

val\_labels(welfare18$hdiploma) <- val\_labels(welfare21$hdiploma)

nous avons affecté les labels de 2018 à ceux de 2021

#Vérification  
welfare18 %>%   
 to\_factor() %>%   
 select(hdiploma) %>%   
 tbl\_summary(missing = "always",  
 missing\_text = "NA")%>%  
modify\_header(label="Les caractéristiques du diplome du chef de ménage")

| Les caractéristiques du diplome du chef de ménage | **N = 7,156***1* |
| --- | --- |
| Diplome du CM |  |
| Aucun | 5,697 (80%) |
| cepe | 587 (8.2%) |
| bepc | 359 (5.0%) |
| cap | 52 (0.7%) |
| bt | 16 (0.2%) |
| bac | 154 (2.2%) |
| DEUG, DUT, BTS | 49 (0.7%) |
| Licence | 83 (1.2%) |
| Maitrise | 63 (0.9%) |
| Master/DEA/DESS | 65 (0.9%) |
| Doctorat/Phd | 31 (0.4%) |
| NA | 0 |
| *1*n (%) | |

#la variable hactivty7j

val\_labels(welfare18$hactiv7j)

## Occupe Chomeur TF cherchant emploi TF cherchant pas   
## 1 2 3 4   
## Inactif Moins de 5 ans   
## 5 6

val\_labels(welfare21$hactiv7j)

## Occupe TF cherchant emploi TF cherchant pas Chomeur   
## 1 2 3 4   
## Inactif Moins de 5 ans   
## 5 6

Erreur de codification

welfare18 %>%   
 to\_factor() %>%   
 select(hactiv7j) %>%   
 tbl\_summary(missing = "always",  
 missing\_text = "NA")%>%  
modify\_header(label="Les caractéristiques de l'activité des 7 derniers jours du chef de ménage")

| Les caractéristiques de l'activité des 7 derniers jours du chef de ménage | **N = 7,156***1* |
| --- | --- |
| Activite 7 jours du CM |  |
| Occupe | 5,362 (75%) |
| Chomeur | 44 (0.6%) |
| TF cherchant emploi | 3 (<0.1%) |
| TF cherchant pas | 60 (0.8%) |
| Inactif | 1,687 (24%) |
| Moins de 5 ans | 0 (0%) |
| NA | 0 |
| *1*n (%) | |

welfare18 <- welfare18 %>%  
 mutate(hactiv7j = dplyr::recode(hactiv7j,  
 `2` = 4,   
 `3` = 2,   
 `4` = 3))  
#on recode avec ceux de 2021

val\_labels(welfare18$hactiv7j) <- val\_labels(welfare21$hactiv7j)

#Visualisation et vérification  
welfare18 %>%   
 to\_factor() %>%   
 select(hactiv7j) %>%   
 tbl\_summary(missing = "always",  
 missing\_text = "NA")%>%  
modify\_header(label="Les caractéristiques de l'activité des 7 derniers jours du chef de ménage")

| Les caractéristiques de l'activité des 7 derniers jours du chef de ménage | **N = 7,156***1* |
| --- | --- |
| Activite 7 jours du CM |  |
| Occupe | 5,362 (75%) |
| TF cherchant emploi | 3 (<0.1%) |
| TF cherchant pas | 60 (0.8%) |
| Chomeur | 44 (0.6%) |
| Inactif | 1,687 (24%) |
| Moins de 5 ans | 0 (0%) |
| NA | 0 |
| *1*n (%) | |

# La variable “hbranch”

val\_labels(welfare18$hbranch)

## Agriculture Elevage/peche Indust. extr. Autr. indust.   
## 1 2 3 4   
## btp Commerce Restaurant/Hotel Trans./Comm.   
## 5 6 7 8   
## Education/Sante Services perso. Aut. services   
## 9 10 11

val\_labels(welfare21$hbranch)

## Agriculture Elevage/syl./peche Indust. extr. Autr. indust.   
## 1 2 3 4   
## btp Commerce Restaurant/Hotel Trans./Comm.   
## 5 6 7 8   
## Education/Sante Services perso. Aut. services   
## 9 10 11

#Erreur de labellisation

welfare18 %>%   
 to\_factor() %>%   
 select(hbranch) %>%   
 tbl\_summary(missing = "always",  
 missing\_text = "NA")%>%  
modify\_header(label="Les caractéristiques de la branche activité du chef de ménage")

| Les caractéristiques de la branche activité du chef de ménage | **N = 7,156***1* |
| --- | --- |
| Branche activite du CM |  |
| Agriculture | 1,366 (25%) |
| Elevage/peche | 374 (6.9%) |
| Indust. extr. | 58 (1.1%) |
| Autr. indust. | 497 (9.1%) |
| btp | 313 (5.8%) |
| Commerce | 1,094 (20%) |
| Restaurant/Hotel | 63 (1.2%) |
| Trans./Comm. | 251 (4.6%) |
| Education/Sante | 379 (7.0%) |
| Services perso. | 761 (14%) |
| Aut. services | 278 (5.1%) |
| NA | 1,722 |
| *1*n (%) | |

l’incohérence concernait la labellisation

val\_labels(welfare18$hbranch) <- val\_labels(welfare21$hbranch)

#Visualisation et vérification  
welfare18 %>%   
 to\_factor() %>%   
 select(hbranch) %>%   
 tbl\_summary(missing = "always",  
 missing\_text = "NA")%>%  
modify\_header(label="Les caractéristiques de la branche activité du chef de ménage")

| Les caractéristiques de la branche activité du chef de ménage | **N = 7,156***1* |
| --- | --- |
| Branche activite du CM |  |
| Agriculture | 1,366 (25%) |
| Elevage/syl./peche | 374 (6.9%) |
| Indust. extr. | 58 (1.1%) |
| Autr. indust. | 497 (9.1%) |
| btp | 313 (5.8%) |
| Commerce | 1,094 (20%) |
| Restaurant/Hotel | 63 (1.2%) |
| Trans./Comm. | 251 (4.6%) |
| Education/Sante | 379 (7.0%) |
| Services perso. | 761 (14%) |
| Aut. services | 278 (5.1%) |
| NA | 1,722 |
| *1*n (%) | |

# La variable “hcsp”

val\_labels(welfare18$hcsp)

## Cadre supérieur   
## 1   
## Cadre moyen/agent de maîtrise   
## 2   
## Ouvrier ou employé qualifié   
## 3   
## Ouvrier ou employé non qualifié   
## 4   
## Manœuvre, aide ménagère   
## 5   
## Stagiaire ou Apprenti rénuméré   
## 6   
## Stagiaire ou Apprenti non rénuméré   
## 7   
## Travailleur familial contribuant à une entreprise familiale   
## 8   
## Travailleur pour compte propre   
## 9   
## Patron   
## 10

val\_labels(welfare21$hcsp)

## Cadre supérieur   
## 1   
## Cadre moyen/agent de maîtrise   
## 2   
## Ouvrier ou employé qualifié   
## 3   
## Ouvrier ou employé non qualifié   
## 4   
## Manœuvre, aide ménagère   
## 5   
## Stagiaire ou Apprenti rénuméré   
## 6   
## Stagiaire ou Apprenti non rénuméré   
## 7   
## Travailleur Familial contribuant pour une entreprise familial   
## 8   
## Travailleur pour compte propre   
## 9   
## Patron   
## 10

welfare21 %>%   
 to\_factor() %>%   
 select(hcsp) %>%   
 tbl\_summary(missing = "always",  
 missing\_text = "NA")%>%  
modify\_header(label="Les caractéristiques du CSP du chéf de ménage")

| Les caractéristiques du CSP du chéf de ménage | **N = 7,120***1* |
| --- | --- |
| CSP du CM |  |
| Cadre supérieur | 57 (1.0%) |
| Cadre moyen/agent de maîtrise | 280 (4.8%) |
| Ouvrier ou employé qualifié | 450 (7.8%) |
| Ouvrier ou employé non qualifié | 332 (5.7%) |
| Manœuvre, aide ménagère | 151 (2.6%) |
| Stagiaire ou Apprenti rénuméré | 34 (0.6%) |
| Stagiaire ou Apprenti non rénuméré | 3 (<0.1%) |
| Travailleur Familial contribuant pour une entreprise familial | 66 (1.1%) |
| Travailleur pour compte propre | 4,302 (74%) |
| Patron | 119 (2.1%) |
| NA | 1,326 |
| *1*n (%) | |

l’incohérence concerne la labellisation de “travailleur Familial contribuant pour une entreprise familial”

val\_labels(welfare21$hcsp) <- val\_labels(welfare18$hcsp)

welfare21 %>%   
 to\_factor() %>%   
 select(hcsp) %>%   
 tbl\_summary(missing = "always",  
 missing\_text = "NA")%>%  
modify\_header(label="Les caractéristiques du CSP du chéf de ménage")

| Les caractéristiques du CSP du chéf de ménage | **N = 7,120***1* |
| --- | --- |
| CSP du CM |  |
| Cadre supérieur | 57 (1.0%) |
| Cadre moyen/agent de maîtrise | 280 (4.8%) |
| Ouvrier ou employé qualifié | 450 (7.8%) |
| Ouvrier ou employé non qualifié | 332 (5.7%) |
| Manœuvre, aide ménagère | 151 (2.6%) |
| Stagiaire ou Apprenti rénuméré | 34 (0.6%) |
| Stagiaire ou Apprenti non rénuméré | 3 (<0.1%) |
| Travailleur familial contribuant à une entreprise familiale | 66 (1.1%) |
| Travailleur pour compte propre | 4,302 (74%) |
| Patron | 119 (2.1%) |
| NA | 1,326 |
| *1*n (%) | |

# La variable “zae”

val\_labels(welfare18$zae)

## NULL

la variable “zae” n’est pas labellisée pour la base 2018

val\_labels(welfare21$zae)

## Kédougou Saint-Louis-Matam   
## 1 3   
## Thies-Diourbel-Louga Kaolack-Fatick-Kaffrine   
## 5 7   
## Ziguinchor-Tamba-Kolda-Sédhiou Dakar   
## 9 11

welfare18 %>%   
 to\_factor() %>%   
 select(zae) %>%   
 tbl\_summary(missing = "always",  
 missing\_text = "NA")%>%  
modify\_header(label="Les caractéristiques de la zone agroécologique")

| Les caractéristiques de la zone agroécologique | **N = 7,156***1* |
| --- | --- |
| Zone agroecologique |  |
| 1 | 1,020 (14%) |
| 2 | 912 (13%) |
| 3 | 1,602 (22%) |
| 4 | 1,414 (20%) |
| 5 | 1,752 (24%) |
| 6 | 456 (6.4%) |
| NA | 0 |
| *1*n (%) | |

La codification n’est pas la meme pour les deux bases

welfare21 %>%   
 to\_factor() %>%   
 select(zae) %>%   
 tbl\_summary(missing = "always",  
 missing\_text = "NA")%>%  
modify\_header(label="Les caractéristiques de la zone agroécologique")

| Les caractéristiques de la zone agroécologique | **N = 7,120***1* |
| --- | --- |
| Zone agroecologique |  |
| Kédougou | 452 (6.3%) |
| Saint-Louis-Matam | 911 (13%) |
| Thies-Diourbel-Louga | 1,599 (22%) |
| Kaolack-Fatick-Kaffrine | 1,413 (20%) |
| Ziguinchor-Tamba-Kolda-Sédhiou | 1,740 (24%) |
| Dakar | 1,005 (14%) |
| NA | 0 |
| *1*n (%) | |

welfare18 <- welfare18 %>%  
 mutate(zae = dplyr::recode(zae,  
 `1` = 11,   
 `2` = 3,   
 `3` = 5,  
 `4` = 7,  
 `5` = 9,  
 `6` = 1))

val\_labels(welfare18$zae) <- val\_labels(welfare21$zae)

# Visualisation et vérification  
welfare18 %>%   
 to\_factor() %>%   
 select(zae) %>%   
 tbl\_summary(missing = "always",  
 missing\_text = "NA")%>%  
modify\_header(label="Les caractéristiques de la zone agroécologique")

| Les caractéristiques de la zone agroécologique | **N = 7,156***1* |
| --- | --- |
| zae |  |
| Kédougou | 456 (6.4%) |
| Saint-Louis-Matam | 912 (13%) |
| Thies-Diourbel-Louga | 1,602 (22%) |
| Kaolack-Fatick-Kaffrine | 1,414 (20%) |
| Ziguinchor-Tamba-Kolda-Sédhiou | 1,752 (24%) |
| Dakar | 1,020 (14%) |
| NA | 0 |
| *1*n (%) | |

variable\_label\_diff <- c()   
  
for (variable in vars\_communes) {   
   
 if(labelled::is.labelled(welfare18[[variable]])){   
   
 value\_label18 <- labelled::val\_labels(welfare18[[variable]])   
  
 }else{  
 value\_label18 <- NULL   
 }  
   
 if(labelled::is.labelled(welfare21[[variable]])){   
   
 value\_label21 <- labelled::val\_labels(welfare21[[variable]])   
   
 }else{  
   
 value\_label21 <- NULL   
 }  
   
 if(!identical(value\_label18, value\_label21)){   
 variable\_label\_diff <- append(variable\_label\_diff,variable)   
 print(variable)  
   
 }  
}

Nous avons vu toutes les incohérences et les avons rectifiées Maintenant, nous allons vérifié que les variables communes ont le meme type et sinon, corrigé

variable\_type\_diff <- c()   
  
for (variable in vars\_communes) {   
  
 type\_var\_18 <- class(welfare18[[variable]])   
 type\_var\_21 <- class(welfare21[[variable]])   
  
 if (!identical(type\_var\_18, type\_var\_21)) {   
 variable\_type\_diff <- append(variable\_type\_diff, variable)   
 }  
}  
  
print(variable\_type\_diff)

## NULL

## Il n’y a pas de variables communes avec des types différents

Aprés avoir vérifier les différentes choses à faire pour une bonne merge, nous allons maintenant merger les deux bases welfare avec l’aide de bind-rows

welfare\_merge <- bind\_rows(welfare18,welfare21) %>%   
 arrange(grappe,menage)

Dans la nouvelle base, nous avons de nouveaux manquants du à la fusion et qui sont propres à 2021, on va donc faire les derniers réglements

welfare\_merge <- welfare\_merge %>%   
 labelled::to\_factor()

# nous allons faire un extrait de la base  
welfare\_merge[1:20,1:20]

## # A tibble: 20 × 20  
## country year hhid grappe menage vague zae region milieu hhweight hhsize  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <fct> <fct> <fct> <dbl> <dbl>  
## 1 SEN 2018 1001 1 1 1 Dakar dakar Urbain 1750. 2  
## 2 SEN 2018 1002 1 2 1 Dakar dakar Urbain 1750. 2  
## 3 SEN 2018 1003 1 3 1 Dakar dakar Urbain 1750. 1  
## 4 SEN 2018 2001 2 1 2 Dakar dakar Urbain 266. 10  
## 5 SEN 2021 201 2 1 2 Dakar dakar Urbain 290. 9  
## 6 SEN 2018 2002 2 2 2 Dakar dakar Urbain 266. 6  
## 7 SEN 2018 2003 2 3 2 Dakar dakar Urbain 266. 4  
## 8 SEN 2021 203 2 3 2 Dakar dakar Urbain 386. 3  
## 9 SEN 2018 2004 2 4 2 Dakar dakar Urbain 266. 3  
## 10 SEN 2021 204 2 4 2 Dakar dakar Urbain 290. 3  
## 11 SEN 2018 2005 2 5 2 Dakar dakar Urbain 266. 2  
## 12 SEN 2021 205 2 5 2 Dakar dakar Urbain 386. 3  
## 13 SEN 2018 2006 2 6 2 Dakar dakar Urbain 266. 3  
## 14 SEN 2021 206 2 6 2 Dakar dakar Urbain 290. 3  
## 15 SEN 2018 2007 2 7 2 Dakar dakar Urbain 266. 1  
## 16 SEN 2021 207 2 7 2 Dakar dakar Urbain 290. 1  
## 17 SEN 2018 2008 2 8 2 Dakar dakar Urbain 266. 4  
## 18 SEN 2021 208 2 8 2 Dakar dakar Urbain 290. 1  
## 19 SEN 2018 2009 2 9 2 Dakar dakar Urbain 266. 1  
## 20 SEN 2018 2010 2 10 2 Dakar dakar Urbain 266. 3  
## # ℹ 9 more variables: eqadu1 <dbl>, eqadu2 <dbl>, hgender <fct>, hage <dbl>,  
## # hmstat <fct>, hreligion <fct>, hnation <fct>, halfab <fct>, heduc <fct>

# Nous avons afficher les vingt premiéres lignes et colonnes

# Télécharger la nouvelle base fusionnée

welfare\_merge <- as.data.frame(welfare\_merge)  
write.csv(welfare\_merge, "welfare\_merge.csv", row.names = FALSE)  
write.csv(welfare\_merge, "welfare\_merge.dta", row.names = FALSE)

# Maintenant, nous allons faire quelques statistiques sur la nouvelle base

Table <- welfare\_merge %>%  
 select(year, hgender, hage, hhsize,halfab, zae, pcexp) %>%   
 tbl\_summary(by = year,   
   
 label = list(   
 hgender ~ "Sexe du chef de ménage",  
 hage ~ "Âge (moyenne)",  
 hhsize ~ "Taille du ménage",  
 halfab~ "Niveau d'alphabétisation du chef de ménage",  
 zae ~ "Zone agroécologiques",  
 pcexp ~ "Dépense per capital (FCFA)"  
 ),  
   
 missing = "no",   
   
 statistic = list(  
 all\_continuous() ~ "{mean} ± {sd}",   
 all\_categorical() ~ "{n} ({p}%)"   
 ),  
   
 digits = all\_continuous() ~ 1,   
 ) %>%  
modify\_header(label="Les caractéristiques du ménage")%>%  
   
 bold\_labels()  
  
Table

| Les caractéristiques du ménage | **2018** N = 7,156*1* | **2021** N = 7,120*1* |
| --- | --- | --- |
| **Sexe du chef de ménage** |  |  |
| Masculin | 5,280 (74%) | 5,095 (72%) |
| Féminin | 1,876 (26%) | 2,025 (28%) |
| **Âge (moyenne)** | 51.5 ± 14.1 | 54.1 ± 13.7 |
| **Taille du ménage** | 9.2 ± 6.0 | 8.7 ± 5.3 |
| **Niveau d'alphabétisation du chef de ménage** |  |  |
| Non | 3,788 (53%) | 3,477 (49%) |
| Oui | 3,368 (47%) | 3,643 (51%) |
| **Zone agroécologiques** |  |  |
| Kédougou | 456 (6.4%) | 452 (6.3%) |
| Saint-Louis-Matam | 912 (13%) | 911 (13%) |
| Thies-Diourbel-Louga | 1,602 (22%) | 1,599 (22%) |
| Kaolack-Fatick-Kaffrine | 1,414 (20%) | 1,413 (20%) |
| Ziguinchor-Tamba-Kolda-Sédhiou | 1,752 (24%) | 1,740 (24%) |
| Dakar | 1,020 (14%) | 1,005 (14%) |
| **Dépense per capital (FCFA)** | 615,630.2 ± 624,819.1 | 621,198.4 ± 536,958.0 |
| *1*n (%); Mean ± SD | | |

#Interprétation des résultats

## Sexe du chef de ménage

La proportion de ménages dirigés par des femmes augmente légèrement (de 26% à 28%).

Cela peut être dû à des changements sociétaux (ex. autonomisation des femmes), économiques (ex. migration des hommes pour le travail) ou démographiques (ex. veuvage).

## Age moyen du chef de ménage

La taille moyenne des ménages diminue légèrement (9.2 → 8.7 personnes).

Possible explication : baisse du taux de fécondité, indépendance accrue des jeunes adultes ou migration vers les villes.

## Alphabétisation du chef de ménage

Amélioration de l’alphabétisation (+4 points).

Possible impact des politiques d’éducation ou de l’accès accru à l’information (ex. internet, campagnes de sensibilisation).

## Répartition géographique

Distribution stable des ménages dans les différentes zones.

Aucune migration importante observée entre ces deux années.

## Dépense per capita

Légère augmentation des dépenses par habitant (615,630 → 621,198 FCFA).

Écart-type diminue (624,819 → 536,958 FCFA) → moins d’inégalités dans les dépenses.

Cela peut suggérer une stabilisation économique ou un meilleur accès aux ressources.

# Conclusion générale

Les résultats montrent des évolutions significatives entre les deux années. On observe une légère augmentation de la proportion de femmes chefs de ménage, une amélioration du niveau d’alphabétisation et une hausse des dépenses par habitant, ce qui suggère des progrès en matière d’éducation et de conditions de vie. Toutefois, le vieillissement des chefs de ménage et la diminution de la taille des ménages pourraient refléter des changements démographiques notables, tels qu’une baisse du renouvellement générationnel ou une évolution des structures familiales. Ces tendances méritent d’être suivies pour mieux comprendre leur impact sur le bien-être des ménages.