

RÉPUBLIQUE DU SÉNÉGAL



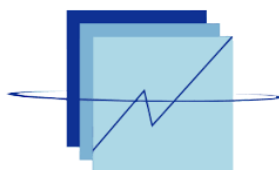
Un Peuple - Un But - Une Foi

Ministère de l'Économie, du Plan et de la Coopération

RÉPUBLIQUE DU SÉNÉGAL
Un Peuple - Un But - Une Foi

MINISTÈRE DE L'ÉCONOMIE
DU PLAN ET DE LA COOPÉRATION

Agence Nationale de la Statistique et de la Démographie (ANSD)



ANSD

École Nationale de la Statistique et de l'Analyse Économique Pierre Ndiaye (ENSAE)



PROJET STATISTIQUES SOUS R

TP9 : Merge des bases de données EHCVM 2018 et 2021

Rédigé par :

Mame Balla BOUSSO

Ameth FAYE

EDIMA Biyenda Hildégarde

Papa Amadou NIANG

Elèves ingénieurs statisticiens économistes

Sous la supervision de :

M. Aboubacar HEMA

ANALYSTE DE RECHERCHE CHEZ

IFPRI

Année scolaire : 2024/2025

Introduction

Dans le cadre de ce projet, nous combinons les observations de la base de données EHCVM 2021 avec celles de 2018 via une opération d'append. L'objectif principal est d'harmoniser les variables afin d'assurer la cohérence de l'ensemble des données pour des analyses ultérieures.

Pour ce faire, nous suivrons les étapes suivantes :

1. **Extraction et listing des variables labelisées (Base 2018)**
 - Extraire et vérifier la liste des variables labelisées dans la base 2018;
 - Documenter les codes et modalités associées à chaque variable.
2. **Extraction et listing des variables correspondantes (Base 2021)**
 - Réaliser la même extraction pour la base 2021.
 - Identifier les différences éventuelles dans les codes ou modalités par rapport à 2018.
3. **Analyse comparative et plan d'harmonisation**
 - Comparer les listes de variables des deux bases.
 - Identifier les écarts et proposer des solutions (recodage, ajout de modalités, renommage avec suffixe, etc.) pour harmoniser les variables fixes.
 - Documenter le plan d'action détaillé.
4. **Mise en œuvre et contrôle qualité**
 - Appliquer le plan d'harmonisation sur les données.
 - Effectuer l'append des observations de 2021 à la base 2018.
 - Réaliser des tests pour vérifier la cohérence globale des données.

Les sections suivantes du document illustrent la démarche avec le chargement et la description des bases, l'identification des variables communes et distinctes (avec ajout de suffixes pour distinguer d'éventuelles différences de codification), le recodage des variables nécessitant un ajustement, puis la combinaison finale des bases.

Chargement des Données et Description

```
# Chargement des bases de données
base2018 <- haven::read_dta("data/ehcvm_welfare_SEN2018.dta")
base2021 <- haven::read_dta("data/ehcvm_welfare_sen2021.dta")

# Affichage du nombre d'observations et de variables pour chaque base
cat("Dimensions de la base 2018 :", dim(base2018), "\n")

## Dimensions de la base 2018 : 7156 35
cat("Dimensions de la base 2021 :", dim(base2021), "\n")

## Dimensions de la base 2021 : 7120 47

# Aperçu rapide des données
summary(base2018)
```

```

##      country      year      hhid      grappe
## Length:7156      Min.   :2018      Min.   : 1001      Min.   : 1.0
## Class :character  1st Qu.:2018      1st Qu.:151002      1st Qu.:151.0
## Mode  :character  Median :2018      Median :300003      Median :300.0
##                               Mean  :2018      Mean  :299934      Mean  :299.9
##                               3rd Qu.:2018      3rd Qu.:449010      3rd Qu.:449.0
##                               Max.   :2018      Max.   :598012      Max.   :598.0
##
##      menage      vague      zae      region
## Min.   : 1.000      Min.   :1.000      Min.   :1.000      Min.   : 1.000
## 1st Qu.: 3.000      1st Qu.:1.000      1st Qu.:2.000      1st Qu.: 3.000
## Median : 6.000      Median :2.000      Median :4.000      Median : 7.000
## Mean   : 6.491      Mean   :1.501      Mean   :3.466      Mean   : 6.781
## 3rd Qu.: 9.000      3rd Qu.:2.000      3rd Qu.:5.000      3rd Qu.:10.000
## Max.   :12.000      Max.   :2.000      Max.   :6.000      Max.   :14.000
##
##      milieu      hhweight      hhsize      eqadul
## Min.   :1.000      Min.   : 15.29      Min.   : 1.00      Min.   : 0.660
## 1st Qu.:1.000      1st Qu.:120.57      1st Qu.: 5.00      1st Qu.: 3.960
## Median :1.000      Median :203.88      Median : 8.00      Median : 5.982
## Mean   :1.449      Mean   :250.38      Mean   : 9.24      Mean   : 6.867
## 3rd Qu.:2.000      3rd Qu.:322.74      3rd Qu.:12.00      3rd Qu.: 8.720
## Max.   :2.000      Max.   :2808.65      Max.   :56.00      Max.   :41.240
##
##      eqadu2      hgender      hage      hmstat
## Min.   : 1.000      Min.   :1.000      Min.   :17.00      Min.   :1.000
## 1st Qu.: 3.167      1st Qu.:1.000      1st Qu.:41.00      1st Qu.:2.000
## Median : 4.333      Median :1.000      Median :51.00      Median :2.000
## Mean   : 4.811      Mean   :1.262      Mean   :51.49      Mean   :2.704
## 3rd Qu.: 5.984      3rd Qu.:2.000      3rd Qu.:62.00      3rd Qu.:3.000
## Max.   :23.706      Max.   :2.000      Max.   :99.00      Max.   :7.000
##                               NA's      :2
##      hreligion      hnation      halfab      heduc
## Min.   :1.000      Min.   : 2.000      Min.   :0.0000      Min.   :1.000
## 1st Qu.:1.000      1st Qu.: 7.000      1st Qu.:0.0000      1st Qu.:1.000
## Median :1.000      Median : 7.000      Median :0.0000      Median :1.000
## Mean   :1.061      Mean   : 7.024      Mean   :0.4707      Mean   :2.188
## 3rd Qu.:1.000      3rd Qu.: 7.000      3rd Qu.:1.0000      3rd Qu.:3.000
## Max.   :5.000      Max.   :12.000      Max.   :1.0000      Max.   :9.000
##
##      hdiploma      hhandig      hactiv7j      hactiv12m
## Min.   : 0.0000      Min.   :0.00000      Min.   :1.000      Min.   :1.00
## 1st Qu.: 0.0000      1st Qu.:0.00000      1st Qu.:1.000      1st Qu.:1.00
## Median : 0.0000      Median :0.00000      Median :1.000      Median :1.00
## Mean   : 0.6385      Mean   :0.09083      Mean   :1.975      Mean   :1.49
## 3rd Qu.: 0.0000      3rd Qu.:0.00000      3rd Qu.:2.000      3rd Qu.:1.00
## Max.   :10.0000      Max.   :1.00000      Max.   :5.000      Max.   :3.00
##
##      hbranch      hsectins      hcsp      dali
## Min.   : 1.000      Min.   :1.000      Min.   : 1.000      Min.   : 113187
## 1st Qu.: 1.000      1st Qu.:3.000      1st Qu.: 5.000      1st Qu.:1143473
## Median : 6.000      Median :3.000      Median : 9.000      Median :1746986
## Mean   : 5.324      Mean   :2.894      Mean   : 7.428      Mean   :2063862
## 3rd Qu.: 9.000      3rd Qu.:3.000      3rd Qu.: 9.000      3rd Qu.:2575472

```

```
## Max. :11.000 Max. :6.000 Max. :10.000 Max. :31295272
## NA's :1722 NA's :1722 NA's :1722
## dnal dtot pcexp zzae
## Min. : 27585 Min. : 278116 Min. : 64161 Min. :296311
## 1st Qu.: 990977 1st Qu.: 2268859 1st Qu.: 300149 1st Qu.:305745
## Median : 1632298 Median : 3484846 Median : 442539 Median :326047
## Mean : 2268143 Mean : 4332005 Mean : 615630 Mean :332244
## 3rd Qu.: 2771981 3rd Qu.: 5391925 3rd Qu.: 697134 3rd Qu.:348125
## Max. :221582921 Max. :227607152 Max. :14286279 Max. :391340
##
## zref def_spa def_temp
## Min. :333441 Min. :0.8886 Min. :0.9916
## 1st Qu.:333441 1st Qu.:0.9169 1st Qu.:0.9916
## Median :333441 Median :0.9778 Median :0.9955
## Mean :333441 Mean :0.9964 Mean :1.0015
## 3rd Qu.:333441 3rd Qu.:1.0440 3rd Qu.:1.0089
## Max. :333441 Max. :1.1736 Max. :1.0147
##
```

summary (base2021)

```
## grappe menage country year
## Min. : 2.0 Min. : 1.000 Length:7120 Min. :2021
## 1st Qu.:152.0 1st Qu.: 4.000 Class :character 1st Qu.:2021
## Median :301.0 Median : 7.000 Mode :character Median :2021
## Mean :300.7 Mean : 7.187 Mean :2021
## 3rd Qu.:450.0 3rd Qu.:10.000 3rd Qu.:2021
## Max. :598.0 Max. :19.000 Max. :2021
##
## hhid vague month zae
## Min. : 201 Min. :1.000 Min. :2021-11-01 Min. : 1.000
## 1st Qu.:15208 1st Qu.:1.000 1st Qu.:2021-12-01 1st Qu.: 5.000
## Median :30103 Median :2.000 Median :2022-04-01 Median : 7.000
## Mean :30082 Mean :1.503 Mean :2022-02-16 Mean : 6.712
## 3rd Qu.:45004 3rd Qu.:2.000 3rd Qu.:2022-05-01 3rd Qu.: 9.000
## Max. :59812 Max. :2.000 Max. :2022-07-01 Max. :11.000
##
## region milieu hhweight hhsize
## Min. : 1.000 Min. :1.000 Min. : 17.73 Min. : 1.000
## 1st Qu.: 3.000 1st Qu.:1.000 1st Qu.: 131.07 1st Qu.: 5.000
## Median : 7.000 Median :1.000 Median : 221.81 Median : 8.000
## Mean : 6.799 Mean :1.449 Mean : 297.44 Mean : 8.747
## 3rd Qu.:10.000 3rd Qu.:2.000 3rd Qu.: 380.29 3rd Qu.:11.000
## Max. :14.000 Max. :2.000 Max. :3081.51 Max. :53.000
##
## eqadul eqadu2 hgender hage
## Min. : 0.660 Min. : 1.000 Min. :1.000 Min. : 16.00
## 1st Qu.: 3.940 1st Qu.: 3.008 1st Qu.:1.000 1st Qu.: 44.00
## Median : 5.790 Median : 4.180 Median :1.000 Median : 54.00
## Mean : 6.589 Mean : 4.625 Mean :1.284 Mean : 54.08
## 3rd Qu.: 8.350 3rd Qu.: 5.762 3rd Qu.:2.000 3rd Qu.: 64.00
## Max. :40.265 Max. :22.882 Max. :2.000 Max. :101.00
##
## hmstat hreligion hnation hethnie
## Min. :1.000 Min. :1.000 Min. : 4.00 Min. : 1.000
```

```

## 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:13.00 1st Qu.: 1.000
## Median :2.000 Median :1.000 Median :13.00 Median : 3.000
## Mean :2.805 Mean :1.058 Mean :12.95 Mean : 3.096
## 3rd Qu.:3.000 3rd Qu.:1.000 3rd Qu.:13.00 3rd Qu.: 3.000
## Max. :7.000 Max. :5.000 Max. :18.00 Max. :13.000
## NA's :82
##
## halfa halfa2 heduc hdiploma
## Min. :0.0000 Min. :0.0000 Min. :1.000 Min. : 0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.: 0.0000
## Median :1.0000 Median :0.0000 Median :1.000 Median : 0.0000
## Mean :0.5117 Mean :0.4947 Mean :2.112 Mean : 0.5622
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:3.000 3rd Qu.: 0.0000
## Max. :1.0000 Max. :1.0000 Max. :9.000 Max. :10.0000
##
## hhandig hactiv7j hactiv12m hbranch
## Min. :0.00000 Min. :1.000 Min. :1.000 Min. : 1.00
## 1st Qu.:0.00000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.: 1.00
## Median :0.00000 Median :1.000 Median :1.000 Median : 5.00
## Mean :0.08919 Mean :2.067 Mean :1.382 Mean : 6.36
## 3rd Qu.:0.00000 3rd Qu.:5.000 3rd Qu.:1.000 3rd Qu.: 8.00
## Max. :1.00000 Max. :5.000 Max. :3.000 Max. :930.00
## NA's :1838
##
## hsectins hcsp dali dnal
## Min. :1.000 Min. : 1.000 Min. : 64205 Min. : 129749
## 1st Qu.:3.000 1st Qu.: 9.000 1st Qu.: 1325492 1st Qu.: 972937
## Median :3.000 Median : 9.000 Median : 1956762 Median : 1595027
## Mean :2.937 Mean : 7.717 Mean : 2276262 Mean : 2029004
## 3rd Qu.:3.000 3rd Qu.: 9.000 3rd Qu.: 2840293 3rd Qu.: 2556010
## Max. :6.000 Max. :10.000 Max. :15144512 Max. :19917053
## NA's :1359 NA's :1326
##
## dtot pcexp zzae zref
## Min. : 235210 Min. : 57610 Min. :331925 Min. :369666
## 1st Qu.: 2432066 1st Qu.: 328408 1st Qu.:335734 1st Qu.:369666
## Median : 3614746 Median : 472561 Median :373110 Median :369666
## Mean : 4305266 Mean : 621198 Mean :371487 Mean :369666
## 3rd Qu.: 5359042 3rd Qu.: 724571 3rd Qu.:385554 3rd Qu.:369666
## Max. :29248050 Max. :9990532 Max. :424179 Max. :369666
##
## def_spa def_temp def_temp_prix2021m11 def_temp_cpi
## Min. :0.8979 Min. :0.9455 Min. :1.000 Min. :0.9743
## 1st Qu.:0.9082 1st Qu.:0.9833 1st Qu.:1.001 1st Qu.:0.9752
## Median :1.0093 Median :0.9949 Median :1.020 Median :0.9935
## Mean :1.0049 Mean :0.9981 Mean :1.020 Mean :0.9935
## 3rd Qu.:1.0430 3rd Qu.:1.0200 3rd Qu.:1.027 3rd Qu.:1.0009
## Max. :1.1475 Max. :1.0590 Max. :1.087 Max. :1.0587
##
## def_temp_adj zali0 dtet monthly_cpi
## Min. :0.9606 Min. :196233 Min. : 50712 Min. :117.2
## 1st Qu.:0.9991 1st Qu.:196233 1st Qu.: 317994 1st Qu.:117.4
## Median :1.0109 Median :196233 Median : 469028 Median :119.6
## Mean :1.0140 Mean :196233 Mean : 641538 Mean :119.6
## 3rd Qu.:1.0363 3rd Qu.:196233 3rd Qu.: 746041 3rd Qu.:120.4
## Max. :1.0760 Max. :196233 Max. :10366096 Max. :127.4
##

```

```
##      cpi2017      icp2017      dollars
## Min.      :1.097   Min.      :238.6   Min.      : 0.5376
## 1st Qu.:1.097   1st Qu.:238.6   1st Qu.: 3.2473
## Median :1.097   Median :238.6   Median : 4.7915
## Mean    :1.097   Mean    :238.6   Mean    : 6.5526
## 3rd Qu.:1.097   3rd Qu.:238.6   3rd Qu.: 7.6134
## Max.    :1.097   Max.    :238.6   Max.    :102.5809
##
```

Identification des variables communes et différentes

Dans cette section, nous identifions les variables présentes dans les deux bases et celles qui diffèrent. Pour les variables communes, nous vérifierons que leur codification est identique. Pour les variables présentant des différences, nous ajouterons le suffixe `_2018` ou `_2021` afin de bien distinguer l'origine de chaque modalité.

```
# Récupération des noms de colonnes
```

```
vars_2018 <- names(base2018)
```

```
vars_2021 <- names(base2021)
```

```
# Variables communes aux deux bases
```

```
vars_communes <- intersect(vars_2018, vars_2021)
```

```
cat("Variables communes :", vars_communes, "\n")
```

```
## Variables communes : country year hhid grappe menage vague zae region milieu hhweig
```

```
# Variables spécifiques à chaque base
```

```
vars_uniques_2018 <- setdiff(vars_2018, vars_2021)
```

```
vars_uniques_2021 <- setdiff(vars_2021, vars_2018)
```

```
cat("Variables spécifiques à 2018 :", vars_uniques_2018, "\n")
```

```
## Variables spécifiques à 2018 : halfab
```

```
cat("Variables spécifiques à 2021 :", vars_uniques_2021, "\n")
```

```
## Variables spécifiques à 2021 : month hethnie halfa halfa2 def_temp_prix2021m11 def_
```

- **Les variables communes :**

Une intersection des noms de colonnes révèle 34 variables partagées entre les deux bases (par exemple, *country*, *year*, *hhid*, etc.). Pour ces variables, il sera possible de vérifier directement que leur codification est identique.

- **Les variables spécifiques à chaque base :**

La base 2018 présente une variable unique (*halphab*), tandis que la base 2021 (affichée ici comme 2021) contient 12 variables spécifiques (parmi lesquelles *month*, *hethnie*, *halfa*, *halfa2*, etc.).

Il est important de noter que la variable *halphab* de 2018 a pour correspondant *halpha2* dans la base 2021. Cela indique qu'une harmonisation manuelle est nécessaire pour aligner ces deux variables équivalentes avant de procéder à la fusion des données.

Recodage et harmonisation

Nous passons maintenant à l'analyse des variables communes pour vérifier leur codification. Pour les variables dont les modalités diffèrent entre 2018 et 2021, nous appliquons un recodage et/ou ajoutons un suffixe afin de conserver la provenance des données.

Dans un premier temps, nous allons faire correspondre la variable `halfa2` à la variable `halphab` en le renommant :

```
library(dplyr)

base2021 <- base2021 %>%
  rename(halfab = halfa2)
```

Nous allons ensuite vérifier si les variables communes sont codifiées de la même manière :

```
# Charger les bibliothèques nécessaires
library(dplyr)
library(labelled)
library(knitr)
library(kableExtra)

# Liste des variables à convertir en facteur
vars <- c("zae", "region", "milieu", "hgender", "hmstat", "hreligion", "hnation",
          "halfab", "heduc", "hdiploma", "hhandig", "hactiv7j", "hactiv12m",
          "hbranch", "hsectins", "hcsp")

# Fonction pour extraire les codes et libellés d'une variable
extract_labels <- function(var) {
  if (!is.null(attr(base2018[[var]], "labels"))) { # Vérifie si la variable a des
    N <- names(attr(base2018[[var]], "labels")) # Libellés
    L <- unname(attr(base2018[[var]], "labels")) # Codes
    df <- data.frame(Variable = var, Code = L, Libellé = N)
    return(df)
  } else {
    return(NULL) # Retourne NULL si pas de labels
  }
}

# Appliquer la fonction à toutes les variables
labels_list <- lapply(vars, extract_labels)

# Supprimer les NULL et fusionner les résultats en un seul tableau
labels_table <- bind_rows(labels_list)

# Afficher sous forme de tableau
kable(labels_table, caption = "Codes et libellés des variables")
```

Table 1: Codes et libellés des variables

Variable	Code	Libellé
region	1	DAKAR
region	2	ZIGUINCHOR

Variable	Code	Libellé
region	3	DIOURBEL
region	4	SAINT-LOUIS
region	5	TAMBACOUNDA
region	6	KAOLACK
region	7	THIES
region	8	LOUGA
region	9	FATICK
region	10	KOLDA
region	11	MATAM
region	12	KAFFRINE
region	13	KEDOUGOU
region	14	SEDHIOU
milieu	1	Urbain
milieu	2	Rural
hgender	1	Masculin
hgender	2	Féminin
hmstat	1	Célibataire
hmstat	2	Marié(e) monogame
hmstat	3	Marié(e) polygame
hmstat	4	Union libre
hmstat	5	Veuf(ve)
hmstat	6	Divorcé(e)
hmstat	7	Séparé(e)
hreligion	1	Musulman
hreligion	2	Chrétien
hreligion	3	Animiste
hreligion	4	Autre Religion
hreligion	5	Sans Religion
hnation	1	Benin
hnation	2	Burkina Faso
hnation	3	Côte d'Ivoire
hnation	4	Guinée Bissau
hnation	5	Mali
hnation	6	Niger
hnation	7	Sénégal
hnation	8	Togo
hnation	9	Nigéria
hnation	10	Autre CEDEAO
hnation	11	Autre Afrique
hnation	12	Autre pays hors Afrique
halfab	0	Non
halfab	1	Oui
heduc	1	Aucun
heduc	2	Maternelle
heduc	3	Primaire
heduc	4	Second. gl 1
heduc	5	Second. tech. 1
heduc	6	Second. gl 2
heduc	7	Second. tech. 2
heduc	8	Postsecondaire
heduc	9	Supérieur
hdiploma	0	Aucun

Variable	Code	Libellé
hdiploma	1	CEP/CFEE
hdiploma	2	BEPC/BFEM
hdiploma	3	CAP
hdiploma	4	BT
hdiploma	5	BAC
hdiploma	6	DEUG, DUT, BTS
hdiploma	7	Licence
hdiploma	8	Maitrise
hdiploma	9	Master/DEA/DESS
hdiploma	10	Doctorat/Phd
hhandig	0	Non
hhandig	1	Oui
hactiv7j	1	Occupe
hactiv7j	2	Chomeur
hactiv7j	3	TF cherchant emploi
hactiv7j	4	TF cherchant pas
hactiv7j	5	Inactif
hactiv7j	6	Moins de 5 ans
hactiv12m	1	Occupe
hactiv12m	2	Trav. fam.
hactiv12m	3	Non occupe
hactiv12m	4	Moins de 5 ans
hbranch	1	Agriculture
hbranch	2	Elevage/peche
hbranch	3	Indust. extr.
hbranch	4	Autr. indust.
hbranch	5	BTP
hbranch	6	Commerce
hbranch	7	Restaurant/Hotel
hbranch	8	Trans./Comm.
hbranch	9	Education/Sante
hbranch	10	Services perso.
hbranch	11	Aut. services
hsectins	1	Etat/Collectivités locales
hsectins	2	Entreprise publique/ parapublique
hsectins	3	Entreprise Privée
hsectins	4	Entreprise associative
hsectins	5	Ménage comme employeur de personnel domestique
hsectins	6	Organisme international /Ambassade
hcsp	1	Cadre supérieur
hcsp	2	Cadre moyen/agent de maîtrise
hcsp	3	Ouvrier ou employé qualifié
hcsp	4	Ouvrier ou employé non qualifié
hcsp	5	Manœuvre, aide ménagère
hcsp	6	Stagiaire ou Apprenti rémunéré
hcsp	7	Stagiaire ou Apprenti non rémunéré
hcsp	8	Travailleur familial contribuant à une entreprise familiale
hcsp	9	Travailleur pour compte propre
hcsp	10	Patron

```

kable(labels_table, caption = "Codes et libellés des variables") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"),
    full_width = FALSE, position = "center") %>%
  column_spec(1, bold = TRUE, color = "blue") %>%
  column_spec(2, bold = TRUE) %>%
  column_spec(3, width = "8cm")

```

Table 2: Codes et libellés des variables

Variable	Code	Libellé
region	1	DAKAR
region	2	ZIGUINCHOR
region	3	DIOURBEL
region	4	SAINT-LOUIS
region	5	TAMBACOUNDA
region	6	KAOLACK
region	7	THIES
region	8	LOUGA
region	9	FATICK
region	10	KOLDA
region	11	MATAM
region	12	KAFFRINE
region	13	KEDOUGOU
region	14	SEDHIOU
milieu	1	Urbain
milieu	2	Rural
hgender	1	Masculin
hgender	2	Féminin
hmstat	1	Célibataire
hmstat	2	Marié(e) monogame
hmstat	3	Marié(e) polygame
hmstat	4	Union libre
hmstat	5	Veuf(ve)
hmstat	6	Divorcé(e)
hmstat	7	Séparé(e)
hreligion	1	Musulman
hreligion	2	Chrétien
hreligion	3	Animiste
hreligion	4	Autre Religion
hreligion	5	Sans Religion
hnation	1	Benin
hnation	2	Burkina Faso
hnation	3	Côte d'Ivoire
hnation	4	Guinée Bissau
hnation	5	Mali
hnation	6	Niger
hnation	7	Sénégal
hnation	8	Togo
hnation	9	Nigéria
hnation	10	Autre CEDEAO

hnation	11	Autre Afrique
hnation	12	Autre pays hors Afrique
halfab	0	Non
halfab	1	Oui
heduc	1	Aucun
heduc	2	Maternelle
heduc	3	Primaire
heduc	4	Second. gl 1
heduc	5	Second. tech. 1
heduc	6	Second. gl 2
heduc	7	Second. tech. 2
heduc	8	Postsecondaire
heduc	9	Superieur
hdiploma	0	Aucun
hdiploma	1	CEP/CFEE
hdiploma	2	BEPC/BFEM
hdiploma	3	CAP
hdiploma	4	BT
hdiploma	5	BAC
hdiploma	6	DEUG, DUT, BTS
hdiploma	7	Licence
hdiploma	8	Maitrise
hdiploma	9	Master/DEA/DESS
hdiploma	10	Doctorat/Phd
hhandig	0	Non
hhandig	1	Oui
hactiv7j	1	Occupe
hactiv7j	2	Chomeur
hactiv7j	3	TF cherchant emploi
hactiv7j	4	TF cherchant pas
hactiv7j	5	Inactif
hactiv7j	6	Moins de 5 ans
hactiv12m	1	Occupe
hactiv12m	2	Trav. fam.
hactiv12m	3	Non occupe
hactiv12m	4	Moins de 5 ans
hbranch	1	Agriculture
hbranch	2	Elevage/peche
hbranch	3	Indust. extr.
hbranch	4	Autr. indust.
hbranch	5	BTP
hbranch	6	Commerce
hbranch	7	Restaurant/Hotel
hbranch	8	Trans./Comm.
hbranch	9	Education/Sante
hbranch	10	Services perso.
hbranch	11	Aut. services
hsectins	1	Etat/Collectivités locales
hsectins	2	Entreprise publique/ parapublique
hsectins	3	Entreprise Privée

hsectins	4	Entreprise associative
hsectins	5	Ménage comme employeur de personnel domestique
hsectins	6	Organisme international /Ambassade
hcsp	1	Cadre supérieur
hcsp	2	Cadre moyen/agent de maîtrise
hcsp	3	Ouvrier ou employé qualifié
hcsp	4	Ouvrier ou employé non qualifié
hcsp	5	Manœuvre, aide ménagère
hcsp	6	Stagiaire ou Apprenti rémunéré
hcsp	7	Stagiaire ou Apprenti non rémunéré
hcsp	8	Travailleur familial contribuant à une entreprise familiale
hcsp	9	Travailleur pour compte propre
hcsp	10	Patron

```
# Charger les bibliothèques nécessaires
library(dplyr)
library(labelled)
library(knitr)
library(kableExtra)

# Liste des variables à convertir en facteur
vars <- c("zae", "region", "milieu", "hgender", "hmstat", "hreligion", "hnation",
          "halfab", "heduc", "hdiploma", "hhandig", "hactiv7j", "hactiv12m",
          "hbranch", "hsectins", "hcsp")

# Fonction pour extraire les codes et libellés d'une variable
extract_labels <- function(var) {
  if (!is.null(attr(base2021[[var]], "labels"))) { # Vérifie si la variable a des
    N <- names(attr(base2021[[var]], "labels")) # Libellés
    L <- unname(attr(base2021[[var]], "labels")) # Codes
    df <- data.frame(Variable = var, Code = L, Libellé = N)
    return(df)
  } else {
    return(NULL) # Retourne NULL si pas de labels
  }
}

# Appliquer la fonction à toutes les variables
labels_list <- lapply(vars, extract_labels)

# Supprimer les NULL et fusionner les résultats en un seul tableau
labels_table <- bind_rows(labels_list)

# Afficher sous forme de tableau
kable(labels_table, caption = "Codes et libellés des variables")
```

Table 3: Codes et libellés des variables

Variable	Code	Libellé
zae	1	Kédougou
zae	3	Saint-Louis-Matam
zae	5	Thies-Diourbel-Louga
zae	7	Kaolack-Fatick-Kaffrine
zae	9	Ziguinchor-Tamba-Kolda-Sédhiou
zae	11	Dakar
region	1	dakar
region	2	ziguinchor
region	3	diourbel
region	4	SAINT-LOUIS
region	5	tambacounda
region	6	kaolack
region	7	thies
region	8	louga
region	9	fatick
region	10	kolda
region	11	matam
region	12	kaffrine
region	13	kedougou
region	14	sedhiou
milieu	1	Urbain
milieu	2	Rural
hgender	1	Masculin
hgender	2	Féminin
hmstat	1	Célibataire
hmstat	2	Marié(e) monogame
hmstat	3	Marié(e) polygame
hmstat	4	Union libre
hmstat	5	Veuf(ve)
hmstat	6	Divorcé(e)
hmstat	7	Séparé(e)
hreligion	1	Musulman
hreligion	2	Chrétien
hreligion	3	Animiste
hreligion	4	Autre Religion
hreligion	5	Sans Religion
hnation	1	Bénin
hnation	2	Burkina Faso
hnation	3	Cape-vert
hnation	4	Cote d'ivoire
hnation	5	Gambie
hnation	6	Ghana
hnation	7	Guinee
hnation	8	Guinée Bissau
hnation	9	Liberia
hnation	10	Mali
hnation	11	Niger
hnation	12	Nigeria
hnation	13	Sénégal
hnation	14	Serra-Leonne
hnation	15	Togo

Variable	Code	Libellé
hnation	17	Autre Afrique
hnation	18	Autre pays hors Afrique
halfab	0	Non
halfab	1	Oui
heduc	1	Aucun
heduc	2	Maternelle
heduc	3	Primaire
heduc	4	Second. gl 1
heduc	5	Second. tech. 1
heduc	6	Second. gl 2
heduc	7	Second. tech. 2
heduc	8	Postsecondaire
heduc	9	Superieur
hdiploma	0	Aucun
hdiploma	1	cepe
hdiploma	2	bepc
hdiploma	3	cap
hdiploma	4	bt
hdiploma	5	bac
hdiploma	6	DEUG, DUT, BTS
hdiploma	7	Licence
hdiploma	8	Maitrise
hdiploma	9	Master/DEA/DESS
hdiploma	10	Doctorat/Phd
hhandig	0	Non
hhandig	1	Oui
hactiv7j	1	Occupe
hactiv7j	2	TF cherchant emploi
hactiv7j	3	TF cherchant pas
hactiv7j	4	Chomeur
hactiv7j	5	Inactif
hactiv7j	6	Moins de 5 ans
hactiv12m	1	Occupe
hactiv12m	2	Trav. fam.
hactiv12m	3	Non occupe
hactiv12m	4	Moins de 5 ans
hbranch	1	Agriculture
hbranch	2	Elevage/syl./peche
hbranch	3	Indust. extr.
hbranch	4	Autr. indust.
hbranch	5	btp
hbranch	6	Commerce
hbranch	7	Restaurant/Hotel
hbranch	8	Trans./Comm.
hbranch	9	Education/Sante
hbranch	10	Services perso.
hbranch	11	Aut. services
hsectins	1	Etat/Collectivités locales
hsectins	2	Entreprise publique/ parapublique
hsectins	3	Entreprise Privée
hsectins	4	Entreprise associative
hsectins	5	Ménage comme employeur de personnel domestique

Variable	Code	Libellé
hsectins	6	Organisme international /Ambassade
hcsp	1	Cadre supérieur
hcsp	2	Cadre moyen/agent de maîtrise
hcsp	3	Ouvrier ou employé qualifié
hcsp	4	Ouvrier ou employé non qualifié
hcsp	5	Manœuvre, aide ménagère
hcsp	6	Stagiaire ou Apprenti rémunéré
hcsp	7	Stagiaire ou Apprenti non rémunéré
hcsp	8	Travailleur Familial contribuant pour une entreprise familial
hcsp	9	Travailleur pour compte propre
hcsp	10	Patron

```
kable(labels_table, caption = "Codes et libellés des variables") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"),
    full_width = FALSE, position = "center") %>%
  column_spec(1, bold = TRUE, color = "blue") %>%
  column_spec(2, bold = TRUE) %>%
  column_spec(3, width = "8cm")
```

Table 4: Codes et libellés des variables

Variable	Code	Libellé
zae	1	Kédougou
zae	3	Saint-Louis-Matam
zae	5	Thies-Diourbel-Louga
zae	7	Kaolack-Fatick-Kaffrine
zae	9	Ziguinchor-Tamba-Kolda-Sédhiou
zae	11	Dakar
region	1	dakar
region	2	ziguinchor
region	3	diourbel
region	4	SAINT-LOUIS
region	5	tambacounda
region	6	kaolack
region	7	thies
region	8	louga
region	9	fatick
region	10	kolda
region	11	matam
region	12	kaffrine
region	13	kedougou
region	14	sedhiou
milieu	1	Urbain
milieu	2	Rural
hgender	1	Masculin
hgender	2	Féminin
hmstat	1	Célibataire
hmstat	2	Marié(e) monogame
hmstat	3	Marié(e) polygame

hmstat	4	Union libre
hmstat	5	Veuf(ve)
hmstat	6	Divorcé(e)
hmstat	7	Séparé(e)
hreligion	1	Musulman
hreligion	2	Chrétien
hreligion	3	Animiste
hreligion	4	Autre Religion
hreligion	5	Sans Religion
hnation	1	Bénin
hnation	2	Burkina Faso
hnation	3	Cape-vert
hnation	4	Cote d'ivoire
hnation	5	Gambie
hnation	6	Ghana
hnation	7	Guinee
hnation	8	Guinée Bissau
hnation	9	Liberia
hnation	10	Mali
hnation	11	Niger
hnation	12	Nigeria
hnation	13	Sénégal
hnation	14	Serra-Leonne
hnation	15	Togo
hnation	17	Autre Afrique
hnation	18	Autre pays hors Afrique
halfab	0	Non
halfab	1	Oui
heduc	1	Aucun
heduc	2	Maternelle
heduc	3	Primaire
heduc	4	Second. gl 1
heduc	5	Second. tech. 1
heduc	6	Second. gl 2
heduc	7	Second. tech. 2
heduc	8	Postsecondaire
heduc	9	Superieur
hdiploma	0	Aucun
hdiploma	1	cepe
hdiploma	2	bepc
hdiploma	3	cap
hdiploma	4	bt
hdiploma	5	bac
hdiploma	6	DEUG, DUT, BTS
hdiploma	7	Licence
hdiploma	8	Maitrise
hdiploma	9	Master/DEA/DESS
hdiploma	10	Doctorat/Phd
hhandig	0	Non
hhandig	1	Oui

hactiv7j	1	Occupe
hactiv7j	2	TF cherchant emploi
hactiv7j	3	TF cherchant pas
hactiv7j	4	Chomeur
hactiv7j	5	Inactif
hactiv7j	6	Moins de 5 ans
hactiv12m	1	Occupe
hactiv12m	2	Trav. fam.
hactiv12m	3	Non occupe
hactiv12m	4	Moins de 5 ans
hbranch	1	Agriculture
hbranch	2	Elevage/syl./peche
hbranch	3	Indust. extr.
hbranch	4	Autr. indust.
hbranch	5	btp
hbranch	6	Commerce
hbranch	7	Restaurant/Hotel
hbranch	8	Trans./Comm.
hbranch	9	Education/Sante
hbranch	10	Services perso.
hbranch	11	Aut. services
hsectins	1	Etat/Collectivités locales
hsectins	2	Entreprise publique/ parapublique
hsectins	3	Entreprise Privée
hsectins	4	Entreprise associative
hsectins	5	Ménage comme employeur de personnel domestique
hsectins	6	Organisme international /Ambassade
hcsp	1	Cadre supérieur
hcsp	2	Cadre moyen/agent de maîtrise
hcsp	3	Ouvrier ou employé qualifié
hcsp	4	Ouvrier ou employé non qualifié
hcsp	5	Manœuvre, aide ménagère
hcsp	6	Stagiaire ou Apprenti rémunéré
hcsp	7	Stagiaire ou Apprenti non rémunéré
hcsp	8	Travailleur Familial contribuant pour une entreprise familiale
hcsp	9	Travailleur pour compte propre
hcsp	10	Patron

Passons maintenant à l'harmonisation des variables de la base de 2018 pour qu'elles correspondent à celles de 2021 :

```
library(dplyr)

# Harmonisation de la base 2018 pour qu'elle corresponde aux modalités de la base
base2018 <- base2018 %>%
  mutate(
    # 1. Région : harmonisation de la casse et des libellés
    region = recode(as.character(region),
      "1" = "dakar",
      "2" = "ziguinchor",
```

```

    "3" = "diourbel",
    "4" = "SAINT-LOUIS",
    "5" = "tambacounda",
    "6" = "kaolack",
    "7" = "thies",
    "8" = "louga",
    "9" = "fatick",
    "10" = "kolda",
    "11" = "matam",
    "12" = "kaffrine",
    "13" = "kedougou",
    "14" = "sedhiou"
),

# 2. Milieu (identique)
milieu = recode(as.character(milieu),
    "1" = "Urbain",
    "2" = "Rural"
),

# 3. Genre
hgender = recode(as.character(hgender),
    "1" = "Masculin",
    "2" = "Féminin"
),

# 4. Statut matrimonial
hmstat = recode(as.character(hmstat),
    "1" = "Célibataire",
    "2" = "Marié(e) monogame",
    "3" = "Marié(e) polygame",
    "4" = "Union libre",
    "5" = "Veuf(ve)",
    "6" = "Divorcé(e)",
    "7" = "Séparé(e)"
),

# 5. Religion
hreligion = recode(as.character(hreligion),
    "1" = "Musulman",
    "2" = "Chrétien",
    "3" = "Animiste",
    "4" = "Autre Religion",
    "5" = "Sans Religion"
),

# 6. Nationalité : harmonisation des libellés
hnation = recode(as.character(hnation),
    "1" = "Bénin",
    "2" = "Burkina Faso",
    "3" = "Cote d'Ivoire", # transformation de "Côte d'Ivoire"
    "4" = "Guinée Bissau",
    "5" = "Mali",

```

```

    "6" = "Niger",
    "7" = "Sénégal",
    "8" = "Togo",
    "9" = "Nigeria",          # transformation de "Nigéria"
    "10" = "Autre Afrique",   # fusion d' "Autre CEDEAO" et "Autre Afrique"
    "11" = "Autre Afrique",
    "12" = "Autre pays hors Afrique"
),

# 7. Niveau d'éducation
heduc = recode(as.character(heduc),
    "1" = "Aucun",
    "2" = "Maternelle",
    "3" = "Primaire",
    "4" = "Second. gl 1",
    "5" = "Second. tech. 1",
    "6" = "Second. gl 2",
    "7" = "Second. tech. 2",
    "8" = "Postsecondaire",
    "9" = "Superieur"
),

# 8. Diplôme : harmonisation (ex. transformation en minuscules)
hdiploma = recode(as.character(hdiploma),
    "0" = "Aucun",
    "1" = "cepe",
    "2" = "bepc",
    "3" = "cap",
    "4" = "bt",
    "5" = "bac",
    "6" = "DEUG, DUT, BTS",
    "7" = "Licence",
    "8" = "Maitrise",
    "9" = "Master/DEA/DESS",
    "10" = "Doctorat/Phd"
),

# 9. Handicape
hhandig = recode(as.character(hhandig),
    "0" = "Non",
    "1" = "Oui"
),

# 10. Activité sur 7 jours : réordonner et harmoniser les libellés
hactiv7j = recode(as.character(hactiv7j),
    "1" = "Occupe",
    "2" = "TF cherchant emploi",
    "3" = "TF cherchant pas",
    "4" = "Chomeur",
    "5" = "Inactif",
    "6" = "Moins de 5 ans"
),

```

```

# 11. Activité sur 12 mois
hactiv12m = recode(as.character(hactiv12m),
  "1" = "Occupe",
  "2" = "Trav. fam.",
  "3" = "Non occupe",
  "4" = "Moins de 5 ans"
),

# 12. Branche d'activité : harmonisation des libellés
hbranch = recode(as.character(hbranch),
  "1" = "Agriculture",
  "2" = "Elevage/syl./peche",
  "3" = "Indust. extr.",
  "4" = "Autr. indust.",
  "5" = "btp",
  "6" = "Commerce",
  "7" = "Restaurant/Hotel",
  "8" = "Trans./Comm.",
  "9" = "Education/Sante",
  "10" = "Services perso.",
  "11" = "Aut. services"
),

# 13. Secteur institutionnel
hsectins = recode(as.character(hsectins),
  "1" = "Etat/Collectivités locales",
  "2" = "Entreprise publique/ parapublique",
  "3" = "Entreprise Privée",
  "4" = "Entreprise associative",
  "5" = "Ménage comme employeur de personnel domestique",
  "6" = "Organisme international /Ambassade"
),

# 14. Catégorie socio-professionnelle
hcsp = recode(as.character(hcsp),
  "1" = "Cadre supérieur",
  "2" = "Cadre moyen/agent de maîtrise",
  "3" = "Ouvrier ou employé qualifié",
  "4" = "Ouvrier ou employé non qualifié",
  "5" = "Manœuvre, aide ménagère",
  "6" = "Stagiaire ou Apprenti rémunéré",
  "7" = "Stagiaire ou Apprenti non rémunéré",
  "8" = "Travailleur familial contribuant à une entreprise familiale",
  "9" = "Travailleur pour compte propre",
  "10" = "Patron"
),

# 15. zae : recodage selon les labels de 2021
zae = recode(as.character(zae),
  "1" = "Kédougou",
  "3" = "Saint-Louis-Matam",
  "5" = "Thies-Diourbel-Louga",
  "7" = "Kaolack-Fatick-Kaffrine",

```

```

    "9" = "Ziguinchor-Tamba-Kolda-Sédhiou",
    "11" = "Dakar"
  )
)

```

Dans notre cas, la base 2018 a été recodée de manière à adopter exactement les mêmes modalités que la base 2021. Ainsi, nous avons :

- Uniformisé la casse et les libellés : Par exemple, les régions ont été harmonisées en passant de libellés majuscules (comme « DAKAR ») à des formats identiques (ex. « dakar » ou « SAINT-LOUIS ») pour assurer une lecture cohérente.
- Aligné les catégories communes : Les variables telles que le milieu de résidence, le genre, le statut matrimonial, la religion, la nationalité, le niveau d'éducation, le diplôme, l'activité et la branche d'activité ont été recodées pour garantir que chaque modalité représente exactement le même concept dans les deux périodes.
- Standardisé les variables spécifiques : LA variable `zae`, bien qu'existante dans la base 2018, n'était pas encore codifiées. En l'affectant les mêmes labels que dans la base 2021, nous assurons une cohérence totale lors de la fusion des jeux de données.

Cette démarche permet ainsi de s'assurer que toute analyse comparative entre 2018 et 2021 repose sur des variables homogènes et comparables, éliminant ainsi les biais liés aux différences de codification.

Append des bases de données

Une fois l'harmonisation effectuée, nous pouvons procéder à l'append des deux bases afin de constituer une base finale homogène.

```

base2021 <- base2021 %>%
  rename(month_2021 = month,
         hethnie_2021 = hethnie,
         halfa_2021 = halfa,
         def_temp_prix2021m11_2021 = def_temp_prix2021m11,
         def_temp_cpi_2021 = def_temp_cpi,
         def_temp_adj_2021 = def_temp_adj,
         zali0_2021 = zali0,
         dtet_2021 = dtet,
         monthly_cpi_2021 = monthly_cpi,
         cpi2017_2021 = cpi2017,
         icp2017_2021 = icp2017,
         dollars_2021 = dollars
  )

```

```

base2018 <- labelled::to_factor(base2018)
base2021 <- labelled::to_factor(base2021)

```

```

# Définition de l'ordre commun des variables
common_vars <- c("country", "year", "hhid", "grappe", "menage", "vague",
               "zae", "region", "milieu", "hhweight", "hhsize",
               "eqadul", "eqadu2", "hgender", "hage", "hmstat",
               "hreligion", "hnation", "halfab", "heduc", "hdiploma",
               "hhandig", "hactiv7j", "hactiv12m", "hbranch",
               "hsectins", "hcsp", "dali", "dnal", "dtot", "pcexp",

```

```

      "zxae", "zref", "def_spa", "def_temp")

# Réorganisation des colonnes dans la base 2018
base2018 <- base2018 %>%
  select(all_of(common_vars))

# Réorganisation des colonnes dans la base 2021
base2021 <- base2021 %>%
  select(all_of(common_vars))

# Fusion des deux bases (bind_rows() complétera par NA les colonnes manquantes, le
data_final_intermediaire <- bind_rows(base2018, base2021)

# Création d'un dataframe avec les variables supplémentaires de base2021 (et les c
base2021 <- haven::read_dta("data/ehcvm_welfare_sen2021.dta")
base2021 <- base2021 %>%
  rename(month_2021 = month,
         hethnie_2021 = hethnie,
         halfa_2021 = halfa,
         def_temp_prix2021m11_2021 = def_temp_prix2021m11,
         def_temp_cpi_2021 = def_temp_cpi,
         def_temp_adj_2021 = def_temp_adj,
         zali0_2021 = zali0,
         dtet_2021 = dtet,
         monthly_cpi_2021 = monthly_cpi,
         cpi2017_2021 = cpi2017,
         icp2017_2021 = icp2017,
         dollars_2021 = dollars
  )
extra_vars <- base2021 %>%
  select(hhid, year, month_2021, hethnie_2021, halfa_2021,
         def_temp_prix2021m11_2021, def_temp_cpi_2021,
         def_temp_adj_2021, zali0_2021, dtet_2021,
         monthly_cpi_2021, cpi2017_2021, icp2017_2021, dollars_2021)

# Fusion de la base finale (issu du bind_rows des bases 2018 et 2021 harmonisées)
data_final <- data_final_intermediaire %>%
  left_join(extra_vars, by = c("hhid", "year"))

```

Voici les étapes clés :

- **Conversion des variables étiquetées en facteurs :**

Les commandes

```

base2018 <- labelled::to_factor(base2018)
base2021 <- labelled::to_factor(base2021)

```

transforment les variables étiquetées (provenant notamment de formats Stata ou SPSS) en facteurs R, assurant ainsi que les types des variables soient directement comparables entre les deux bases.

- **Réorganisation et fusion des variables communes :**

En définissant un vecteur `common_vars` qui liste les variables présentes dans les deux bases, on réorganise les colonnes pour qu'elles suivent le même ordre. La fusion initiale avec `bind_rows()` permet de combiner les observations des deux bases tout en remplissant les valeurs manquantes pour les variables absentes.

- **Ajout des variables supplémentaires de 2021 :**

Les variables spécifiques à la collecte 2021 sont extraites (après renommage pour éviter toute confusion) et jointes à la base fusionnée via une jointure sur les identifiants communs (`hhid` et `year`).

L'ensemble de ces opérations garantit une base de données finale homogène, où les variables sont alignées et comparables, facilitant ainsi les analyses longitudinales et comparatives entre 2018 et 2021.

Saving

Nous exportons enfin la base de données obtenue sous format stata.

```
# Sauvegarde de la base finale au format .dta
haven::write_dta(data_final, "output/data_final.dta")
```