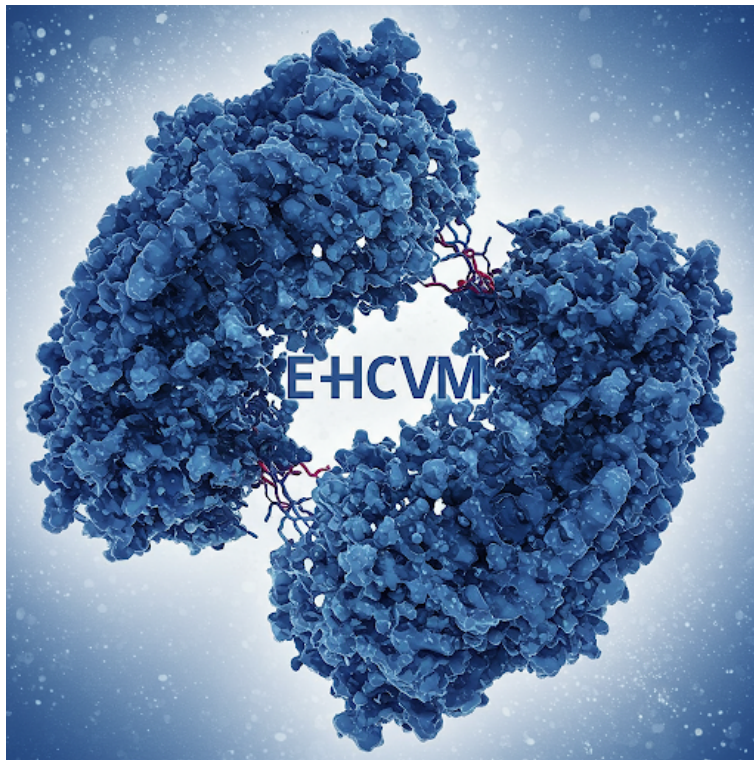


EHCVM 2018 et 2021

Aissatou Sega DIALLO

2025-03-20

Fusion des bases Welfare EHCVM



Sous la supervision de :
M. Hema

Introduction

Ce travail pratique (TP) a pour objectif de fusionner les bases de données welfare des Enquêtes sur les Conditions de Vie des Ménages (EHCVM) de 2018 et 2021 du Sénégal, afin de créer une base de données consolidée qui permettra de réaliser des analyses comparatives. Nous procéderons à l'exploration et à la préparation des données pour des analyses statistiques futures, en traitant des problématiques telles que les valeurs manquantes, les doublons et les valeurs aberrantes. À la fin de ce TP, la base de données fusionnée sera nettoyée et prête à être exploitée pour des études ultérieures sur les conditions de vie des ménages sénégalais.

I.Installation et chargement des packages

Cette section permet d'assurer que les bibliothèques requises sont correctement installées et chargées pour l'exécution des analyses.

```
libraries <- c("readr", "dplyr","labelled","tidyr", "gtsummary","haven", "utils", "ggplot2", "plotly")

for (x in libraries) {
  if (!requireNamespace(x, quietly = TRUE)) {
    install.packages(x)
  }
  library(x, character.only = TRUE)
}
```

II.Importation des bases de données welfare

Cette section consiste à importer les bases de données welfare des Enquêtes sur les Conditions de Vie des Ménages (EHCVM) afin de les préparer pour l'analyse.

```
getwd()
```

```
## [1] "C:/Formation ISE/ISE 1/Logiciel R/Projet-statistique-sous-R/TP 9/TP9_DIALLO_AissatouSega/Script"
```

```
welf2018 <- haven::read_dta("../Données/ehcvm_welfare_sen2018.dta") #importation base welfare 2018
welf2021 <- haven::read_dta("../Données/ehcvm_welfare_sen2021.dta") #importation base welfare 2021
```

III.Exploration des données :

Cette section permettra d'examiner la structure des données et d'obtenir un aperçu des variables. Nous analyserons les premières lignes de chaque fichier afin de confirmer leur bonne importation et leur préparation pour l'analyse.

```
# Cette étape permet d'examiner rapidement la structure des données
head(welf2018)
```

```
## # A tibble: 6 x 35
##   country year  hhid grappe menage vague   zae region  milieu hhweight hhsize
```

```
##   <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl+lbl> <dbl+lbl> <dbl> <dbl>
## 1 SEN     2018  1001    1    1    1    1 1 [daka~ 1 [Urb~ 1750.    2
## 2 SEN     2018  1002    1    2    1    1 1 [daka~ 1 [Urb~ 1750.    2
## 3 SEN     2018  1003    1    3    1    1 1 [daka~ 1 [Urb~ 1750.    1
## 4 SEN     2018  2001    2    1    2    1 1 [daka~ 1 [Urb~ 266.    10
## 5 SEN     2018  2002    2    2    2    1 1 [daka~ 1 [Urb~ 266.    6
## 6 SEN     2018  2003    2    3    2    1 1 [daka~ 1 [Urb~ 266.    4
## # i 24 more variables: eqadu1 <dbl>, eqadu2 <dbl>, hgender <dbl+lbl>,
## #   hage <dbl>, hmstat <dbl+lbl>, hreligion <dbl+lbl>, hnation <dbl+lbl>,
## #   halfab <dbl+lbl>, heduc <dbl+lbl>, hdiploma <dbl+lbl>, hhandig <dbl+lbl>,
## #   hactiv7j <dbl+lbl>, hactiv12m <dbl+lbl>, hbranch <dbl+lbl>,
## #   hsectins <dbl+lbl>, hcsp <dbl+lbl>, dali <dbl>, dnal <dbl>, dtot <dbl>,
## #   pcexp <dbl>, zzae <dbl>, zref <dbl>, def_spa <dbl>, def_temp <dbl>
```

```
head(welf2021)
```

```
## # A tibble: 6 x 47
##   grappe menage country year hhid vague month      zae      region milieu
##   <dbl>   <dbl> <chr>   <dbl> <dbl> <dbl> <date>    <dbl+lbl> <dbl+lbl> <dbl+lbl>
## 1     2     5 SEN     2021  205    2 2022-05-01 11 [Dakar] 1 [daka~ 1 [Urb~
## 2     2    15 SEN     2021  215    2 2022-05-01 11 [Dakar] 1 [daka~ 1 [Urb~
## 3     2     3 SEN     2021  203    2 2022-05-01 11 [Dakar] 1 [daka~ 1 [Urb~
## 4     2    13 SEN     2021  213    2 2022-05-01 11 [Dakar] 1 [daka~ 1 [Urb~
## 5     2     8 SEN     2021  208    2 2022-06-01 11 [Dakar] 1 [daka~ 1 [Urb~
## 6     2    16 SEN     2021  216    2 2022-06-01 11 [Dakar] 1 [daka~ 1 [Urb~
## # i 37 more variables: hhweight <dbl>, hhsize <dbl>, eqadu1 <dbl>,
## #   eqadu2 <dbl>, hgender <dbl+lbl>, hage <dbl>, hmstat <dbl+lbl>,
## #   hreligion <dbl+lbl>, hnation <dbl+lbl>, hethnie <dbl+lbl>, halfa <dbl+lbl>,
## #   halfa2 <dbl+lbl>, heduc <dbl+lbl>, hdiploma <dbl+lbl>, hhandig <dbl+lbl>,
## #   hactiv7j <dbl+lbl>, hactiv12m <dbl+lbl>, hbranch <dbl+lbl>,
## #   hsectins <dbl+lbl>, hcsp <dbl+lbl>, dali <dbl>, dnal <dbl>, dtot <dbl>,
## #   pcexp <dbl>, zzae <dbl>, zref <dbl>, def_spa <dbl>, def_temp <dbl>, ...
```

```
# Vérification des dimensions des deux bases de données
dim(welf2018)
```

```
## [1] 7156 35
```

```
dim(welf2021)
```

```
## [1] 7120 47
```

```
# Informations générales sur les variables/colonnes
str(welf2018)
```

```
## tibble [7,156 x 35] (S3: tbl_df/tbl/data.frame)
## $ country : chr [1:7156] "SEN" "SEN" "SEN" "SEN" ...
## ..- attr(*, "label")= chr "Pays"
## ..- attr(*, "format.stata")= chr "%3s"
## $ year : num [1:7156] 2018 2018 2018 2018 2018 ...
## ..- attr(*, "label")= chr "Annee enquete"
```

```

##   .- attr(*, "format.stata")= chr "%8.0g"
## $ hhid      : num [1:7156] 1001 1002 1003 2001 2002 ...
##   .- attr(*, "label")= chr "Idenfiant menage"
##   .- attr(*, "format.stata")= chr "%12.0g"
## $ grappe    : num [1:7156] 1 1 1 2 2 2 2 2 2 ...
##   .- attr(*, "label")= chr "Numero grappe"
##   .- attr(*, "format.stata")= chr "%8.0g"
## $ menage     : num [1:7156] 1 2 3 1 2 3 4 5 6 7 ...
##   .- attr(*, "label")= chr "Numero menage"
##   .- attr(*, "format.stata")= chr "%8.0g"
## $ vague      : num [1:7156] 1 1 1 2 2 2 2 2 2 ...
##   .- attr(*, "label")= chr "Vague"
##   .- attr(*, "format.stata")= chr "%8.0g"
## $ zae        : num [1:7156] 1 1 1 1 1 1 1 1 1 ...
##   .- attr(*, "label")= chr "Zone agroecologique"
##   .- attr(*, "format.stata")= chr "%8.0g"
## $ region     : dbl+lbl [1:7156] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
##   .@ label      : chr "Region residence"
##   .@ format.stata: chr "%8.0g"
##   .@ labels     : Named num [1:14] 1 2 3 4 5 6 7 8 9 10 ...
##   .- attr(*, "names")= chr [1:14] "dakar" "ziguinchor" "diourbel" "SAINT-LOUIS" ...
## $ milieu     : dbl+lbl [1:7156] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
##   .@ label      : chr "Milieu residence"
##   .@ format.stata: chr "%8.0g"
##   .@ labels     : Named num [1:2] 1 2
##   .- attr(*, "names")= chr [1:2] "Urbain" "Rural"
## $ hhweight   : num [1:7156] 1750 1750 1750 266 266 ...
##   .- attr(*, "label")= chr "Ponderation menage"
##   .- attr(*, "format.stata")= chr "%9.0g"
## $ hhsizes    : num [1:7156] 2 2 1 10 6 4 3 2 3 1 ...
##   .- attr(*, "label")= chr "Taille menage"
##   .- attr(*, "format.stata")= chr "%8.0g"
## $ eqadul     : num [1:7156] 1.32 1.45 1 7.66 4.28 ...
##   .- attr(*, "label")= chr "Nbr adultes-equiv. FAO"
##   .- attr(*, "format.stata")= chr "%12.0g"
## $ eqadu2     : num [1:7156] 1.61 1.61 1 5.54 3.25 ...
##   .- attr(*, "label")= chr "Nbr adultes-equiv. alt."
##   .- attr(*, "format.stata")= chr "%9.0g"
## $ hgender    : dbl+lbl [1:7156] 2, 1, 1, 1, 1, 2, 1, 2, 2, 2, 1, 1, 2, 1, 2, 1, 1, 1, 1, 1, ...
##   .@ label      : chr "Genre du CM"
##   .@ format.stata: chr "%8.0g"
##   .@ labels     : Named num [1:2] 1 2
##   .- attr(*, "names")= chr [1:2] "Masculin" "Féminin"
## $ hage       : num [1:7156] 59 59 27 85 40 68 31 54 67 42 ...
##   .- attr(*, "label")= chr "Age du CM"
##   .- attr(*, "format.stata")= chr "%8.0g"
## $ hmstat     : dbl+lbl [1:7156] 1, 2, 1, 2, 2, 5, 1, 2, 5, 1, 2, 2, 2, 1, 2, 2, 2, 2, 6, ...
##   .@ label      : chr "Situation famille du CM"
##   .@ format.stata: chr "%8.0g"
##   .@ labels     : Named num [1:7] 1 2 3 4 5 6 7
##   .- attr(*, "names")= chr [1:7] "Célibataire" "Marié(e) monogame" "Marié(e) polygame" "Union 1.
## $ hreligion  : dbl+lbl [1:7156] 1, 2, 2, 1, 1, 1, 2, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
##   .@ label      : chr "Religion du CM"
##   .@ format.stata: chr "%8.0g"

```

```

## ..@ labels      : Named num [1:5] 1 2 3 4 5
## .. ..- attr(*, "names")= chr [1:5] "Musulman" "Chrétien" "Animiste" "Autre Religion" ...
## $ hnation      : dbl+lbl [1:7156] 7, 12, 7, 7, 7, 7, 11, 7, 7, 7, 7, 7, 1...
## ..@ label       : chr "Nationalite du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels      : Named num [1:12] 1 2 3 4 5 6 7 8 9 10 ...
## .. ..- attr(*, "names")= chr [1:12] "Benin" "Burkina Faso" "Côte d'Ivoire" "Guinée Bissau" ...
## $ halfab       : dbl+lbl [1:7156] 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0,...
## ..@ label       : chr "Alphabetisation du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels      : Named num [1:2] 0 1
## .. ..- attr(*, "names")= chr [1:2] "Non" "Oui"
## $ heduc        : dbl+lbl [1:7156] 3, 9, 9, 3, 1, 1, 9, 1, 3, 4, 3, 1, 1, 1, 1, 9, 9, 1,...
## ..@ label       : chr "Education du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels      : Named num [1:9] 1 2 3 4 5 6 7 8 9
## .. ..- attr(*, "names")= chr [1:9] "Aucun" "Maternelle" "Primaire" "Second. gl 1" ...
## $ hdiploma     : dbl+lbl [1:7156] 1, 6, 10, 0, 0, 0, 10, 0, 0, 1, 0, 0, 0, ...
## ..@ label       : chr "Diplome du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels      : Named num [1:11] 0 1 2 3 4 5 6 7 8 9 ...
## .. ..- attr(*, "names")= chr [1:11] "Aucun" "CEP/CFEE" "BEPC/BFEM" "cap" ...
## $ hhandig      : dbl+lbl [1:7156] 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## ..@ label       : chr "Handicap majeur CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels      : Named num [1:2] 0 1
## .. ..- attr(*, "names")= chr [1:2] "Non" "Oui"
## $ hactiv7j     : dbl+lbl [1:7156] 1, 1, 1, 5, 1, 5, 1, 1, 5, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## ..@ label       : chr "Activite 7 jours du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels      : Named num [1:6] 1 2 3 4 5 6
## .. ..- attr(*, "names")= chr [1:6] "Occupe" "Chomeur" "TF cherchant emploi" "TF cherchant pas" ..
## $ hactiv12m    : dbl+lbl [1:7156] 1, 1, 1, 3, 1, 3, 1, 1, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## ..@ label       : chr "Activite 12 mois du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels      : Named num [1:4] 1 2 3 4
## .. ..- attr(*, "names")= chr [1:4] "Occupe" "Trav. fam." "Non occupe" "Moins de 5 ans"
## $ hbranch      : dbl+lbl [1:7156] 6, 6, 11, NA, 6, NA, 9, 11, NA, 6, 7, 8, 6, 1...
## ..@ label       : chr "Branche activite du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels      : Named num [1:11] 1 2 3 4 5 6 7 8 9 10 ...
## .. ..- attr(*, "names")= chr [1:11] "Agriculture" "Elevage/peche" "Indust. extr." "Autr. indust."
## $ hsectins     : dbl+lbl [1:7156] 3, 3, 1, NA, 3, NA, 1, 5, NA, 3, 3, 3, 3, ...
## ..@ label       : chr "Secteur instit. du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels      : Named num [1:6] 1 2 3 4 5 6
## .. ..- attr(*, "names")= chr [1:6] "Etat/Collectivités locales" "Entreprise publique/ parapublique"
## $ hcsp         : dbl+lbl [1:7156] 4, 1, 7, NA, 4, NA, 6, 5, NA, 9, 3, 4, 9, ...
## ..@ label       : chr "CSP du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels      : Named num [1:10] 1 2 3 4 5 6 7 8 9 10
## .. ..- attr(*, "names")= chr [1:10] "Cadre supérieur" "Cadre moyen/agent de maîtrise" "Ouvrier ou
## $ dali         : num [1:7156] 1978488 1886313 387989 6568518 2362250 ...
## ..- attr(*, "label")= chr "Conso annuelle alim. menage"

```

```
##      .. attr(*, "format.stata")= chr "%12.0g"
## $ dnal      : num [1:7156] 13139233 8894985 5055663 4889539 3930703 ...
##      .. attr(*, "label")= chr "Conso annuelle non alim. menage"
##      .. attr(*, "format.stata")= chr "%12.0g"
## $ dtot      : num [1:7156] 15117721 10781298 5443652 11458057 6292954 ...
##      .. attr(*, "label")= chr "Conso annuelle totale menage"
##      .. attr(*, "format.stata")= chr "%12.0g"
## $ pcexp      : num [1:7156] 6440507 4593088 4638249 976281 893649 ...
##      .. attr(*, "label")= chr "Indicateur de bien-être"
##      .. attr(*, "format.stata")= chr "%12.0g"
## $ zzae      : num [1:7156] 391340 391340 391340 391340 391340 ...
##      .. attr(*, "format.stata")= chr "%12.0g"
## $ zref      : num [1:7156] 333441 333441 333441 333441 333441 ...
##      .. attr(*, "label")= chr "Seuil pauvrete national"
##      .. attr(*, "format.stata")= chr "%12.0g"
## $ def_spa    : num [1:7156] 1.17 1.17 1.17 1.17 1.17 ...
##      .. attr(*, "label")= chr "Deflateur spatial"
##      .. attr(*, "format.stata")= chr "%9.0g"
## $ def_temp   : num [1:7156] 1.015 1.012 1.015 0.992 0.992 ...
##      .. attr(*, "label")= chr "Deflateur temporel"
##      .. attr(*, "format.stata")= chr "%9.0g"
```

```
## tibble [7,120 x 47] (S3: tbl_df/tbl/data.frame)
## $ grappe : num [1:7120] 2 2 2 2 2 2 2 2 2 2 ...
## .. attr(*, "label")= chr "grappe"
## .. attr(*, "format.stata")= chr "%8.0g"
## $ menage : num [1:7120] 5 15 3 13 8 16 7 4 1 12 ...
## .. attr(*, "label")= chr "Identifiant du ménage"
## .. attr(*, "format.stata")= chr "%8.0g"
## $ country : chr [1:7120] "SEN" "SEN" "SEN" "SEN" ...
## .. attr(*, "label")= chr "Pays"
## .. attr(*, "format.stata")= chr "%3s"
## $ year : num [1:7120] 2021 2021 2021 2021 2021 2021 ...
## .. attr(*, "label")= chr "Annee enquete"
## .. attr(*, "format.stata")= chr "%8.0g"
## $ hhid : num [1:7120] 205 215 203 213 208 216 207 204 201 212 ...
## .. attr(*, "label")= chr "Idenfiant menage"
## .. attr(*, "format.stata")= chr "%12.0g"
## $ vague : num [1:7120] 2 2 2 2 2 2 2 2 2 2 ...
## .. attr(*, "label")= chr "Vague"
## .. attr(*, "format.stata")= chr "%8.0g"
## $ month : Date[1:7120], format: "2022-05-01" "2022-05-01" ...
## $ zae : dbl+lbl [1:7120] 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 1...
## ..@ label : chr "Zone agroecologique"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:6] 1 3 5 7 9 11
## .. ..- attr(*, "names")= chr [1:6] "Kédougou" "Saint-Louis-Matam" "Thies-Diourbel-Louga" "Kaolack"
## $ region : dbl+lbl [1:7120] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## ..@ label : chr "Region residence"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:14] 1 2 3 4 5 6 7 8 9 10 ...
## .. ..- attr(*, "names")= chr [1:14] "dakar" "ziguinchor" "diourbel" "SAINT-LOUIS" ...
```

```

## $ milieu : dbl+lbl [1:7120] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## ..@ label : chr "Milieu residence"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:2] 1 2
## ..- attr(*, "names")= chr [1:2] "Urbain" "Rural"
## $ hhweight : num [1:7120] 386 386 386 386 290 ...
## ..- attr(*, "label")= chr "Ponderation menage"
## ..- attr(*, "format.stata")= chr "%12.0g"
## $ hhsz : num [1:7120] 3 2 3 5 1 5 1 3 9 6 ...
## ..- attr(*, "label")= chr "Taille menage"
## ..- attr(*, "format.stata")= chr "%8.0g"
## $ eqadu1 : num [1:7120] 2.28 1.66 2.82 4 0.79 ...
## ..- attr(*, "label")= chr "Nbr adultes-equiv. FAO"
## ..- attr(*, "format.stata")= chr "%12.0g"
## $ eqadu2 : num [1:7120] 2.2 1.61 1.87 3.17 1 ...
## ..- attr(*, "label")= chr "Nbr adultes-equiv. alt."
## ..- attr(*, "format.stata")= chr "%9.0g"
## $ hgender : dbl+lbl [1:7120] 2, 1, 2, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, 2, 1, 1, 1, 1, ...
## ..@ label : chr "Genre du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:2] 1 2
## ..- attr(*, "names")= chr [1:2] "Masculin" "Féminin"
## $ hage : num [1:7120] 33 42 47 56 54 45 46 34 88 51 ...
## ..- attr(*, "label")= chr "Age du CM"
## ..- attr(*, "format.stata")= chr "%8.0g"
## $ hmstat : dbl+lbl [1:7120] 3, 1, 3, 2, 6, 2, 1, 1, 2, 2, 3, 5, 1, 5, 3, 2, 2, 1, ...
## ..@ label : chr "Situation famille du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:7] 1 2 3 4 5 6 7
## ..- attr(*, "names")= chr [1:7] "Célibataire" "Marié(e) monogame" "Marié(e) polygame" "Union 1
## $ hreligion : dbl+lbl [1:7120] 1, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## ..@ label : chr "Religion du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:5] 1 2 3 4 5
## ..- attr(*, "names")= chr [1:5] "Musulman" "Chrétien" "Animiste" "Autre Religion" ...
## $ hnation : dbl+lbl [1:7120] 13, 13, 13, 13, 13, 13, 13, 17, 13, 13, 13, 13, 13, 1...
## ..@ label : chr "Nationalite du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:17] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:17] "Bénin" "Burkina Faso" "Cape-vert" "Cote d'Ivoire" ...
## $ hethnie : dbl+lbl [1:7120] 2, 3, 6, 1, 1, 1, 5, NA, 8, 2, 1, 1, 1, ...
## ..@ label : chr "Ethnie du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:13] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:13] "Wolof/Lébou" "Sérère" "Poullar" "Soninké" ...
## $ halfa : dbl+lbl [1:7120] 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, ...
## ..@ label : chr "Alpha. lire/scr. CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:2] 0 1
## ..- attr(*, "names")= chr [1:2] "Non" "Oui"
## $ halfa2 : dbl+lbl [1:7120] 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, ...
## ..@ label : chr "Alpha. lire/scr./comp. CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:2] 0 1

```

```

## ..- attr(*, "names")= chr [1:2] "Non" "Oui"
## $ heduc : dbl+lbl [1:7120] 1, 9, 4, 3, 1, 9, 4, 9, 3, 1, 1, 3, 4, 3, 1, 4, 9, 9,...
## ..@ label : chr "Education du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:9] 1 2 3 4 5 6 7 8 9
## ..- attr(*, "names")= chr [1:9] "Aucun" "Maternelle" "Primaire" "Second. gl 1" ...
## $ hdiploma : dbl+lbl [1:7120] 0, 7, 1, 1, 0, 6, 1, 10, 0, 0, 0, 0, 1, ...
## ..@ label : chr "Diplome du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:11] 0 1 2 3 4 5 6 7 8 9 ...
## ..- attr(*, "names")= chr [1:11] "Aucun" "cepe" "bepc" "cap" ...
## $ hhandig : dbl+lbl [1:7120] 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,...
## ..@ label : chr "Handicap majeur CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:2] 0 1
## ..- attr(*, "names")= chr [1:2] "Non" "Oui"
## $ hactiv7j : dbl+lbl [1:7120] 1, 1, 1, 1, 1, 1, 1, 5, 1, 1, 1, 5, 5, 1, 1, 1, 1, 1,...
## ..@ label : chr "Activite 7 jours du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:6] 1 2 3 4 5 6
## ....- attr(*, "names")= chr [1:6] "Occupe" "TF cherchant emploi" "TF cherchant pas" "Chomeur" ..
## $ hactiv12m : dbl+lbl [1:7120] 1, 1, 1, 1, 1, 1, 1, 3, 1, 1, 1, 3, 3, 1, 1, 1, 1, 1,...
## ..@ label : chr "Activite 12 mois du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:4] 1 2 3 4
## ..- attr(*, "names")= chr [1:4] "Occupe" "Trav. fam." "Non occupe" "Moins de 5 ans"
## $ hbranch : dbl+lbl [1:7120] 11, 9, 11, 11, 11, 9, 6, NA, 2, 4, 6, NA, NA, 1...
## ..@ label : chr "Branche activite du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:11] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:11] "Agriculture" "Eleavage/syl./peche" "Indust. extr." "Autr. ind..."
## $ hsectins : dbl+lbl [1:7120] 5, 3, 5, 3, 3, 2, 3, NA, 3, 3, 3, NA, NA, ...
## ..@ label : chr "Secteur instit. du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:6] 1 2 3 4 5 6
## ..- attr(*, "names")= chr [1:6] "Etat/Collectivités locales" "Entreprise publique/ parapublique..."
## $ hcsp : dbl+lbl [1:7120] 5, 3, 5, 4, 3, 3, 9, NA, 9, 9, 9, NA, NA, ...
## ..@ label : chr "CSP du CM"
## ..@ format.stata: chr "%8.0g"
## ..@ labels : Named num [1:10] 1 2 3 4 5 6 7 8 9 10
## ..- attr(*, "names")= chr [1:10] "Cadre supérieur" "Cadre moyen/agent de maîtrise" "Ouvrier ou..."
## $ dali : num [1:7120] 1357897 1330889 1510518 2759037 973303 ...
## ..- attr(*, "label")= chr "Conso annuelle alim. menage"
## ..- attr(*, "format.stata")= chr "%12.0g"
## $ dnal : num [1:7120] 930976 476051 1887723 2087600 1573509 ...
## ..- attr(*, "label")= chr "Conso annuelle non alim. menage"
## ..- attr(*, "format.stata")= chr "%12.0g"
## $ dtot : num [1:7120] 2288874 1806940 3398241 4846637 2546811 ...
## ..- attr(*, "label")= chr "Conso annuelle totale menage"
## ..- attr(*, "format.stata")= chr "%12.0g"
## $ pcexp : num [1:7120] 664906 787360 987171 844753 2219506 ...
## ..- attr(*, "label")= chr "Indicateur de bien-être"
## ..- attr(*, "format.stata")= chr "%12.0g"
## $ zzae : num [1:7120] 424179 424179 424179 424179 424179 ...

```



```

##   ..- attr(*, "format.stata")= chr "%12.0g"
##   $ zref                : num [1:7120] 369666 369666 369666 369666 369666 ...
##   ..- attr(*, "label")= chr "Seuil pauvreté national"
##   ..- attr(*, "format.stata")= chr "%12.0g"
##   $ def_spa             : num [1:7120] 1.15 1.15 1.15 1.15 1.15 ...
##   ..- attr(*, "label")= chr "Déflateur spatial"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##   $ def_temp            : num [1:7120] 0.991 0.991 0.991 0.991 1.013 ...
##   ..- attr(*, "label")= chr "Déflateur temporel"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##   $ def_temp_prix2021m11: num [1:7120] 1.03 1.03 1.03 1.03 1.05 ...
##   ..- attr(*, "label")= chr "temporal deflator for international poverty, 1 = 2021m11 prices"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##   $ def_temp_cpi        : num [1:7120] 1 1 1 1 1.02 ...
##   ..- attr(*, "label")= chr "alternative temporal deflator based on official CPI, 2018/19 style"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##   $ def_temp_adj        : num [1:7120] 1.01 1.01 1.01 1.01 1.03 ...
##   ..- attr(*, "label")= chr "temporal deflator adjusted for difference between hh and market survey"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##   $ zali0               : num [1:7120] 196233 196233 196233 196233 196233 ...
##   ..- attr(*, "label")= chr "(sum) conso_pc_val_up"
##   ..- attr(*, "format.stata")= chr "%12.0g"
##   $ dtet               : num [1:7120] 769710 911466 1142771 977906 2515290 ...
##   ..- attr(*, "format.stata")= chr "%12.0g"
##   $ monthly_cpi         : num [1:7120] 120 120 120 120 123 ...
##   ..- attr(*, "label")= chr "Monthly CPI value"
##   ..- attr(*, "format.stata")= chr "%12.0g"
##   $ cpi2017             : num [1:7120] 1.1 1.1 1.1 1.1 1.1 ...
##   ..- attr(*, "format.stata")= chr "%9.0g"
##   $ icp2017            : num [1:7120] 239 239 239 239 239 ...
##   ..- attr(*, "format.stata")= chr "%12.0g"
##   $ dollars             : num [1:7120] 7.71 9.12 11.44 9.79 25.79 ...
##   ..- attr(*, "label")= chr "welfare in 2017 PPP USD per capita per day (not spatially deflated)"
##   ..- attr(*, "format.stata")= chr "%9.0g"

```

IV. Évaluation de la qualité des données :

Dans cette section, l'analyse de la qualité des données sera faite en vérifiant la présence de valeurs manquantes, de doublons, et en s'assurant que les types de données correspondent bien aux variables attendues. Cette étape nous aidera à identifier d'éventuelles anomalies.

1. Doublons

```

# Vérification des doublons
sum(duplicated(welf2018))

```

```
## [1] 0
```

```
sum(duplicated(welf2021))
```

```
## [1] 0
```

Aucun doublon n'a été trouvé dans les deux bases

2. Valeurs manquantes

```
# Détection des valeurs manquantes  
summary(welf2018)
```

```
##      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          hhsiz          eqadu1  
## 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\(welf2021\)](#)

```
##      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
##
##      eqadu1      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
##
```

```
# Calcul du pourcentage de valeurs manquantes pour chaque colonne
colSums(is.na(welf2018)) / nrow(welf2018) * 100
```

```
##      country      year      hhid      grappe      menage      vague
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      zae      region      milieu      hhweight      hhsizes      eqadu1
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      eqadu2      hgender      hage      hmstat      hreligion      hnation
## 0.00000000 0.00000000 0.00000000 0.02794857 0.00000000 0.00000000
##      halfab      heduc      hdiploma      hhandig      hactiv7j      hactiv12m
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      hbranch      hsectins      hcsp      dali      dnal      dtot
## 24.06372275 24.06372275 24.06372275 0.00000000 0.00000000 0.00000000
##      pcexp      zxae      zref      def_spa      def_temp
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
```

```
colSums(is.na(welf2021)) / nrow(welf2021) * 100
```

```
##      grappe      menage      country
## 0.000000 0.000000 0.000000
##      year      hhid      vague
## 0.000000 0.000000 0.000000
##      month      zae      region
## 0.000000 0.000000 0.000000
##      milieu      hhweight      hhsizes
## 0.000000 0.000000 0.000000
##      eqadu1      eqadu2      hgender
## 0.000000 0.000000 0.000000
```

```
##          hage          hmstat          hreligion
##      0.000000      0.000000      0.000000
##      hnation      hethnie          halfa
##      0.000000      1.151685      0.000000
##      halfa2          heduc          hdiploma
##      0.000000      0.000000      0.000000
##      hhandig      hactiv7j      hactiv12m
##      0.000000      0.000000      0.000000
##      hbranch      hsectins          hcsp
##      25.814607      19.087079      18.623596
##      dali          dnal          dtot
##      0.000000      0.000000      0.000000
##      pcexp          zzae          zref
##      0.000000      0.000000      0.000000
##      def_spa      def_temp      def_temp_prix2021m11
##      0.000000      0.000000      0.000000
##      def_temp_cpi      def_temp_adj          zali0
##      0.000000      0.000000      0.000000
##      dtet          monthly_cpi          cpi2017
##      0.000000      0.000000      0.000000
##      icp2017          dollars
##      0.000000      0.000000
```

Un calcul du pourcentage de valeurs manquantes sera fait pour chaque colonne des deux bases. Le mode

Pour la base de données welfare 2018, les variables hactiv7j, hsectins, dnal, pcexp, hcsp, hbranch, hactiv12m, dtot, zzae et dali présentent des pourcentages relativement faibles de valeurs manquantes, variant entre 0 % et 24 %. Il en va de même pour la base welfare 2021, avec les variables hcsp, hbranch, dali, hethnie et hsectins.

V. Comparaison des variables et modalités entre les bases welfare 2018 et 2021

Après la visualisation des données et vérification de la présence d'éventuelles anomalies, un procédé à une comparaison des variables et des modalités entre les bases welfare 2018 et 2021 sera engagé. Cette étape nous permettra d'identifier les divergences dans l'annotation des variables et des modalités entre les deux bases.

```
#Liste des variables des deux bases
var2018=colnames(welf2018)
var2021=colnames(welf2021)

# Identification des variables communes aux deux bases
var_communes <- intersect(var2018, var2021)

# Variables spécifiques à la base 2018
var_sp2018 <- setdiff(var2018, var2021)

# Variables spécifiques à la base 2021
var_sp2021 <- setdiff(var2021, var2018)

var_sp2018
```

```
## [1] "halfab"
```

```
var_sp2021
```

```
## [1] "month"          "hethnie"          "halfa"
## [4] "halfa2"         "def_temp_prix2021m11" "def_temp_cpi"
## [7] "def_temp_adj"   "zali0"            "dtet"
## [10] "monthly_cpi"    "cpi2017"          "icp2017"
## [13] "dollars"
```

Cette catégorisation montre que les deux bases partagent 34 variables communes. La base welfare 2018 comporte une seule variable unique, halfab, tandis que 13 variables sont spécifiques à la base welfare 2021 : “month”, “hethnie”, “halfa”, “halfa2”, “def_temp_prix2021m11”, “def_temp_cpi”, “def_temp_adj”, “zali0”, “dtet”, “monthly_cpi”, “cpi2017”, “icp2017” et “dollars”.

Concernant la variable d’alphabétisation, il s’agit simplement d’une différence d’orthographe. Nous procéderons donc à renommer la variable de 2021 en halfab et l’inclure dans les variables communes.

```
## Renommage de la variable 'halfa' en 'halfab'
colnames(welf2021)[colnames(welf2021)=="halfa"] <- "halfab"

# Ajout de 'halfab' à la liste des variables communes
var_communes <- append(var_communes, "halfab")
```

Ainsi , a la suite , une vérification de la cohérence des labellisations.

1. Identification des incohérences de labellisation

```
discord_lab_var <- c()

for (var in var_communes) {

  if(labelled::is.labelled(welf2018[[var]])){ # Vérification si la variable en 2018 est labellisée

    lab_val2018 <- labelled::val_labels(welf2018[[var]]) # Récupération des labels de la variable en 2018

  }else{
    lab_val2018 <- NULL # Si la variable en 2018 n'est pas labellisée, on l'indique par NULL
  }

  if(labelled::is.labelled(welf2021[[var]])){ # Vérification si la variable en 2021 est labellisée

    lab_val2021 <- labelled::val_labels(welf2021[[var]]) # Récupération des labels de la variable en 2021

  }else{
    lab_val2021 <- NULL # Si la variable en 2021 n'est pas labellisée, on l'indique par NULL
  }

  if(!identical(lab_val2018, lab_val2021)){ # Vérification de la concordance des labels entre 2018 et 2021
```

```

    discord_lab_var <- append(discord_lab_var,var) # Si les labels diffèrent, on ajoute le nom de la v
    print(var)
  }
}

```

```

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

```

Il a été constaté que six variables présentent des incohérences dans leur labellisation entre les bases welfare 2018 et 2021 : “zae”, “hnation”, “hdiploma”, “hactiv7j”, “hbranch” et “hcsp”. Afin de corriger ces divergences, chaque variable sera traitée individuellement, recodée et ses modalités ajustées en fonction des incohérences détectées.

2. Gestion des incohérences

a) Traitement de la variable hnation

```

# Visualisation de la distribution de la variable hnation en 2018
labelled::val_labels(welf2018$hnation)

```

Visualisation de la distribution

##	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

```

# Visualisation de la distribution de la variable hnation en 2021
labelled::val_labels(welf2021$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	

Characteristic	N = 7,120 ^I
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

^In (%)

```
# Transformation de la variable hnation en facteur et affichage de la distribution en 2021
welf2021 %>%
  to_factor() %>%
  select(hnation) %>%
  tbl_summary(missing = "always", # Affichage des valeurs manquantes
              missing_text = "Valeurs manquantes")
```

Recherche de l’incohérence Certaines modalités, telles que “Autres CEDEAO” en 2018, ont été subdivisées en catégories distinctes, comme c’est le cas pour la Guinée, qui figure désormais comme une modalité spécifique.

Correction de l’incohérence Afin de corriger cette incohérence, la variable sera recodée dans la base welfare 2021, en combinant les modalités qui étaient regroupées sous “Autres CEDEAO” en 2018, y compris la Guinée et les autres pays concernés.

```
#Recodage dans la base de 2021
welf2021 <- welf2021 %>%
  mutate(hnation = dplyr::recode(hnation,
    `4` = 3, # Remplace la modalité 4 par 3
    `8` = 4,
    `10` = 5,
    `11` = 6,
    `13` = 7,
```

Characteristic	N = 7,120 ^I
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

^In (%)

```
`15` = 8,
`12` = 9,
`17` = 11,
`18` = 12,
`3` = 10, `5` = 10, `6` = 10, `7` = 10, `9` = 10, `14` = 10
))
```

```
#Uniformisation des labels
labelled::val_labels(welf2021$hnation) <- labelled::val_labels(welf2018$hnation)
```

```
welf2021 %>%
  to_factor() %>%
  select(hnation) %>%
  tbl_summary(missing = "always",
              missing_text = "Valeurs manquantes")
```

Contrôle des modifications

b) Traitement de la variable hdiploma

```
# Affichage des labels de la variable hdiploma en 2018
labelled::val_labels(welf2018$hdiploma)
```

Visualisation de la distribution

Characteristic	N = 7,156 ^I
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

^In (%)

```
##          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
```

```
# Affichage des labels de la variable hdiploma en 2021
labelled::val_labels(welf2021$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
```

```
# Résumé statistique de la variable hdiploma en 2018 avec affichage des valeurs manquantes
welf2018 %>%
  to_factor() %>%
  select(hdiploma) %>%
  tbl_summary(missing = "always",
              missing_text = "NA")
```

Recherche de l'incohérence Les différences observées entre les deux bases concernent uniquement des modifications d'étiquettes (labels) sans altération des valeurs sous-jacentes.

Correction de l'incohérence Il suffira donc d'affecter les labels de hdiploma dans welfare 2021 à ceux de welfare 2018 afin d'assurer une harmonisation des libellés entre les deux bases.

Characteristic	N = 7,156 ^I
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

^In (%)

```
#Harmonisation des labels
val_labels(welf2018$hdiploma) <- val_labels(welf2021$hdiploma)
```

```
welf2018 %>%
  to_factor() %>%
  select(hdiploma) %>%
  tbl_summary(missing = "always",
              missing_text = "NA")
```

Controle des modifications

c) Traitement de la variable hactiv7j

```
# Affichage des labels de la variable hactiv7j en 2018
labelled::val_labels(welf2018$hactiv7j)
```

Visualisation de la distribution

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

```
# Affichage des labels de la variable hactiv7j en 2021
labelled::val_labels(welf2021$hactiv7j)
```

Characteristic	N = 7,156 ^I
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

^In (%)

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

Résumé statistique de la variable hactiv7j dans la base 2018

```
welf2018 %>%
  to_factor() %>%
  select(hactiv7j) %>%
  tbl_summary(missing = "always", missing_text = "NA") # Génération du tableau avec affichage des valeurs
```

Recherche de l'incohérence Il y a une incohérence dans l'ordre des modalités entre les bases 2018 et 2021, bien que les labels soient identiques. Cela suggère une erreur dans le codage des valeurs. Une harmonisation des modalités est nécessaire pour assurer la comparabilité des données entre les deux années.

Correction de l'incohérence Pour corriger cette incohérence, il est nécessaire de réorganiser les modalités afin qu'elles correspondent exactement entre les deux bases. Une fois l'ordre des modalités ajusté dans welfare 2021, il sera alors possible d'affecter ses labels à welfare 2018 pour assurer une harmonisation complète.

```
welf2018 <- welf2018 %>%
  mutate(hactiv7j = dplyr::recode(hactiv7j,
    `2` = 4,
    `3` = 2,
    `4` = 3))
```

Affectation des labels de la variable 'hactiv7j' de la base 2021 à la base 2018

```
val_labels(welf2018$hactiv7j) <- val_labels(welf2021$hactiv7j)
```

```
welf2018 %>%
  to_factor() %>%
  select(hactiv7j) %>%
  tbl_summary(missing = "always",
    missing_text = "NA")
```

Characteristic	N = 7,156 ^I
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

^In (%)

Controle des modifications

d) Traitement de la variable hbranch

```
# Visualisation des labels de la variable 'hbranch' dans la base 2018
labelled::val_labels(welf2018$hbranch)
```

Visualisation de la distribution

```
##      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
```

```
# Visualisation des labels de la variable 'hbranch' dans la base 2021
labelled::val_labels(welf2021$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
```

```
# Vérification de la distribution des valeurs manquantes pour la variable 'hbranch' dans la base 2018
welf2018 %>%
  to_factor() %>%
  select(hbranch) %>%
  tbl_summary(missing = "always",
              missing_text = "NA")
```

Characteristic	N = 7,156 ^I
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

^In (%)

Recherche de l'incohérence Il est observé qu'en 2021, une nouvelle modalité correspondant au secteur de la sylvopasture a été ajoutée et combinée avec les secteurs de l'élevage et de la pêche, ce qui diffère de la catégorisation de 2018.

Correction de l'incohérence Cette incohérence peut être corrigée par une simple affectation, similaire à celle effectuée pour les autres variables.

```
# Affectation des labels de la variable 'hactiv7j' de la base 2021 à la base 2018
val_labels(welf2018$hbranch) <- val_labels(welf2021$hbranch)
```

```
welf2018 %>%
  to_factor() %>%
  select(hbranch) %>%
  tbl_summary(missing = "always",
              missing_text = "NA")
```

Controle des modifications

v) Traitement de la variable hcsp

```
# Affichage des labels de la variable hcsp pour la base welfare 2018
labelled::val_labels(welf2018$hcsp)
```

Visualisation de la distribution

Characteristic	N = 7,156 ^I
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

^In (%)

```
##                               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
```

```
# Affichage des labels de la variable hcsp pour la base welfare 2021
labelled::val_labels(welf2021$hcsp)
```

```
##                               Cadre supérieur
##                               1
##                               Cadre moyen/agent de maîtrise
##                               2
##                               Ouvrier ou employé qualifié
##                               3
##                               Ouvrier ou employé non qualifié
##                               4
```


Characteristic	N = 7,120 ^I
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 rémunéré	34 (0.6%)
Stagiaire ou Apprenti non rémunéré	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

^I_n (%)

```
##                               Manœuvre, aide ménagère
##                               5
##                               Stagiaire ou Apprenti rémunéré
##                               6
##                               Stagiaire ou Apprenti non rémunéré
##                               7
## Travailleur Familial contribuant pour une entreprise familial
##                               8
##                               Travailleur pour compte propre
##                               9
##                               Patron
##                               10
```

Recherche de l'incohérence Comme la différence réside uniquement dans la labellisation de la neuvième modalité, il suffit de mettre à jour les labels de la base welfare 2021 pour les aligner sur ceux de 2018.

Correction de l'incohérence Ainsi, les labels de 2018 seront affectés à ceux de 2021.

```
#Affectation des labels de 2018 à 2021
val_labels(welf2021$hcsp) <- val_labels(welf2018$hcsp)
```

```
welf2021 %>%
  to_factor() %>%
  select(hcsp) %>%
  tbl_summary(missing = "always",
              missing_text = "NA")
```

Contrôle des modifications

Characteristic	N = 7,156 ^I
Zone agroécologique	
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

^In (%)

e) Traitement de la variable zae

```
# Vérification des labels de la variable zae dans la base welfare 2018
labelled::val_labels(welf2018$zae)
```

Visualisation de la distribution

```
## NULL
```

```
# Vérification des labels de la variable zae dans la base welfare 2021
labelled::val_labels(welf2021$zae)
```

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

```
# Résumé de la distribution de la variable zae dans la base welfare 2018, avec gestion des valeurs manquantes
welf2018 %>%
  to_factor() %>%
  select(zae) %>%
  tbl_summary(missing = "always",
              missing_text = "NA")
```

```
# Résumé de la distribution de la variable zae dans la base welfare 2021, avec gestion des valeurs manquantes
welf2021 %>%
  select(zae) %>%
  tbl_summary(missing = "always",
              missing_text = "NA")
```

```
## ! Column(s) "zae" are class "haven_labelled".
## i This is an intermediate data structure not meant for analysis.
```

Characteristic	N = 7,120 ^I
Zone agroécologique	
1	452 (6.3%)
3	911 (13%)
5	1,599 (22%)
7	1,413 (20%)
9	1,740 (24%)
11	1,005 (14%)
NA	0

^In (%)

```
## i Convert columns with 'haven::as_factor()', 'labelled::to_factor()',
## 'labelled::unlabelled()', and 'unclass()'. Failure to convert may have
## unintended consequences or result in error.
## <https://haven.tidyverse.org/articles/semantics.html>
## <https://larmarange.github.io/labelled/articles/intro_labelled.html#unlabelled>
```

Identification de l'incohérence Il a été observé que dans la base welfare 2018, la variable zae n'est pas labellisée, et les codes diffèrent entre les deux bases. En 2018, les codes sont 1, 2, 3, 4, 5 et 6, tandis qu'en 2021, ils sont 1, 3, 5, 7, 9 et 11. De plus, les codes correspondant à Kedougou et Dakar sont inversés dans la base 2018, bien qu'il faille prendre en compte les fréquences de chaque code dans les deux bases.

Correction de l'incohérence Il sera d'abord nécessaire d'aligner les codes de 2018 avec ceux de 2021. Une fois cette correspondance effectuée, il sera possible d'affecter les labels appropriés à la variable dans la base welfare 2018, en suivant le même processus que pour la base welfare 2021.

Rétablissement de l'ordre pour Kedougou et Dakar en 2018 Afin de corriger l'inversion des codes pour Kedougou et Dakar dans la base 2018, il convient d'échanger les codes correspondants dans cette base. Cette modification permettra d'aligner l'ordre des modalités sur celui de la base 2021, garantissant ainsi la cohérence des données.“{r}

```
#Correction de l'inversion de codes pour Kedougou et Dakar en 2018
#Ce code permet d'inverser les codes pour Kedougou et Dakar dans la base de données 2018 afin de rétablir
welf2018 <- welf2018 %>%
  mutate(zae = dplyr::recode(zae,
    '1' = 6, # Kédougou remplace Dakar
    '2' = 2,
    '3' = 3,
    '4' = 4,
    '5' = 5,
    '6' = 1 # Dakar devient Kedougou
  ))
```

Characteristic	N = 7,156 ^I
zae	
1	456 (6.4%)
2	912 (13%)
3	1,602 (22%)
4	1,414 (20%)
5	1,752 (24%)
6	1,020 (14%)
NA	0

^In (%)

```
#Vérification après correction des codes de Kedougou et Dakar en 2018
#Ce premier bloc permet de vérifier l'effet de la mutation dans la base de données 2018 après la correc
welf2018 %>%
  to_factor() %>%
  select(zae) %>%
  tbl_summary(missing = "always",
              missing_text = "NA")
```

```
#Harmonisation des codes géographiques dans la base 2021
#Le deuxième bloc effectue l'harmonisation des codes géographiques dans la base 2021 en réaffectant les
welf2021 <- welf2021 %>%
  mutate(zae = dplyr::recode(zae,
    `1` = 1,
    `3` = 2,
    `5` = 3,
    `7` = 4,
    `9` = 5,
    `11` = 6
  ))
```

Vérification après correction effectuée entre Kedougou et Dakar

```
#Vérification du recodage en 2021
welf2021 %>%
  select(zae) %>%
  tbl_summary(missing = "always",
              missing_text = "NA")
```

Controle des modifications

```
## ! Column(s) "zae" are class "haven_labelled".
## i This is an intermediate data structure not meant for analysis.
```

Characteristic	N = 7,120 ^I
Zone agroécologique	
1	452 (6.3%)
2	911 (13%)
3	1,599 (22%)
4	1,413 (20%)
5	1,740 (24%)
6	1,005 (14%)
NA	0

^In (%)

```
## i Convert columns with 'haven::as_factor()', 'labelled::to_factor()',
## 'labelled::unlabelled()', and 'unclass()'. Failure to convert may have
## unintended consequences or result in error.
## <https://haven.tidyverse.org/articles/semantics.html>
## <https://larmarange.github.io/labelled/articles/intro_labelled.html#unlabelled>
```

```
# Recodage des modalités de `zae` dans la base de données 2021
welf2021 <- welf2021 %>%
  mutate(zae = dplyr::recode(zae,
    `1` = "Kédougou",
    `2` = "Saint-Louis-Matam",
    `3` = "Thies-Diourbel-Louga",
    `4` = "Kaolack-Fatick-Kaffrine",
    `5` = "Ziguinchor-Tamba-Kolda-Sédhiou",
    `6` = "Dakar"
  ))
```

```
# Recodage des modalités de `zae` dans la base de données 2018
welf2018 <- welf2018 %>%
  mutate(zae = dplyr::recode(zae,
    `1` = "Kédougou",
    `2` = "Saint-Louis-Matam",
    `3` = "Thies-Diourbel-Louga",
    `4` = "Kaolack-Fatick-Kaffrine",
    `5` = "Ziguinchor-Tamba-Kolda-Sédhiou",
    `6` = "Dakar"
  ))
```

```
# Vérification des changements dans la base de données 2018
welf2018 %>%
  to_factor() %>%
  select(zae) %>%
  tbl_summary(missing = "always",
    missing_text = "NA")
```

Characteristic	N = 7,156 ^I
zae	
Dakar	1,020 (14%)
Kaolack-Fatick-Kaffrine	1,414 (20%)
Kédougou	456 (6.4%)
Saint-Louis-Matam	912 (13%)
Thies-Diourbel-Louga	1,602 (22%)
Ziguinchor-Tamba-Kolda-Sédhiou	1,752 (24%)
NA	0
^I n (%)	

Characteristic	N = 7,120 ^I
Zone agroécologique	
Dakar	1,005 (14%)
Kaolack-Fatick-Kaffrine	1,413 (20%)
Kédougou	452 (6.3%)
Saint-Louis-Matam	911 (13%)
Thies-Diourbel-Louga	1,599 (22%)
Ziguinchor-Tamba-Kolda-Sédhiou	1,740 (24%)
NA	0
^I n (%)	

```
# Vérification des changements dans la base de données 2021
welf2021 %>%
  to_factor() %>%
  select(zae) %>%
  tbl_summary(missing = "always",
              missing_text = "NA")
```

Vérification des modifications Le codage et la labellisation de la variable zae sont désormais cohérents entre les deux bases. Il reste à vérifier qu'aucune différence n'existe dans la labellisation ou le codage des autres variables communes aux deux bases.

3. Contrôle du traitement des données

```
discord_lab_var <- c()

for (var in var_communes) { # Parcourir les variables communes entre les deux bases

  # Vérifier si la variable est labellisée dans la base 2018
  if (labelled::is.labelled(welf2018[[var]])) {
```

```

    lab_val2018 <- labelled::val_labels(welf2018[[var]]) # Récupérer les labels de la variable en 2018
  } else {
    lab_val2018 <- NULL # Assigner NULL si la variable n'est pas labellisée en 2018
  }

  # Vérifier si la variable est labellisée dans la base 2021
  if (labelled::is.labelled(welf2021[[var]])) {
    lab_val2021 <- labelled::val_labels(welf2021[[var]]) # Récupérer les labels de la variable en 2021
  } else {
    label_val_2021 <- NULL # Assigner NULL si la variable n'est pas labellisée en 2021
  }

  # Comparer les labels des deux bases
  if (!identical(lab_val2018, lab_val2021)) {
    discord_lab_var <- append(discord_lab_var, var)
    print(var)
  }
}

```

```

## [1] "country"
## [1] "year"
## [1] "hhid"
## [1] "grappe"
## [1] "menage"
## [1] "vague"
## [1] "zae"
## [1] "hhweight"
## [1] "hhszise"
## [1] "eqadu1"
## [1] "eqadu2"
## [1] "hage"
## [1] "dali"
## [1] "dnal"
## [1] "dtot"
## [1] "pcexp"
## [1] "zzae"
## [1] "zref"
## [1] "def_spa"
## [1] "def_temp"

```

Fusion des bases welfare 2018 et 2021

Après avoir harmonisé le codage et la labellisation des variables communes entre les deux bases, la prochaine étape consiste à les fusionner. Cette fusion permettra de créer une base de données consolidée, comprenant les informations des deux années, afin d'effectuer des analyses comparatives et des traitements ultérieurs sur les données combinées.

```
welf_finale <- bind_rows(welf2018, welf2021)
```

Récupération de la nouvelle base de données

```
# Exporter la base de données finale en dta  
write_dta(welf_finale, "../Sorties/welf_finale.dta")
```