Prétraitement des données

Adjiwanou - Mounchingam

24 June 2024

Dresser la table

```
#install.packages("table1")

rm(list = ls())
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(summarytools)   # Analyse descriptive
```

```
##
## Attaching package: 'summarytools'
##
## The following object is masked from 'package:tibble':
##
##     view
```

```
library(gtsummary)      # Analyse descriptive
```

```
## #BlackLivesMatter
```

```
library(table1)         # Analyse descriptive
```

```
##
## Attaching package: 'table1'
##
## The following objects are masked from 'package:summarytools':
```

```
##
##     label, label<-
##
## The following objects are masked from 'package:base':
##
##     units, units<-
```

```
library(haven)          # Ouvrir les fichiers stata
```

Ouvrir une base de données

```
bf <- read_dta("../Donnee/BFKR81FL.DTA")
```

Liste des noms de fichiers de données pour chaque pays d'afrique de l'ouest

```
KR <- c("BFKR81FL","BJKR71FL","CIKR81FL","GMKR81FL","GNKR71FL","LBKR7AFL","MLKR7AFL","MRKR71FL","NGKR7B
```

Fonction pour traiter les fichiers enfants

- M18 poids qualitatif
- midx :

```
file_kr <- function(data) {
  data_kr <- read_dta(paste0("../Donnee/",data,".DTA"))
  data_kr_sub <- data_kr %>%
    select(caseid,v000,v001,v002,v005,v131,v157,v158,v159,v106,v743a,v743b,
           v743d,v744a,v744b,v744c,v744d,v744e,v717,v012,midx, m18, m19,b11,b4,m15, bord,v150,
           v151,v190,v136,v025,v130,v445,b5, v701,v024,v367,v426,v113,v115,
           v116,v445,m14,m70,v313)
  return(data_kr_sub)
}

# purrr

data_kr <- map(KR, file_kr)
```

```
dhs_child_west_africa <- bind_rows(data_kr)
```

```
## Warning: ‘..1$v131‘ and ‘..2$v131‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v131‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, and 8

## Warning: ‘..1$m15‘ and ‘..2$m15‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m15‘.
## x Values: 21, 22, 23, 26, 30, 31, and 36
```

```
## Warning: ‘..1$v130‘ and ‘..2$v130‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v130‘.
## x Values: 1, 2, 3, 4, and 5

## Warning: ‘..1$v024‘ and ‘..2$v024‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v024‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, 8, ..., 11, and 12

## Warning: ‘..1$v113‘ and ‘..2$v113‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v113‘.
## x Values: 21 and 72

## Warning: ‘..1$midx‘ and ‘..3$midx‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$midx‘.
## x Values: 0

## Warning: ‘..1$m15‘ and ‘..3$m15‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m15‘.
## x Values: 21, 22, 23, 24, 25, 26, 30, 31, ..., 36, and 40

## Warning: ‘..1$v150‘ and ‘..3$v150‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v150‘.
## x Values: 2, 3, 4, 5, 6, 7, and 8

## Warning: ‘..1$v130‘ and ‘..3$v130‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v130‘.
## x Values: 3, 4, 5, and 6

## Warning: ‘..1$v024‘ and ‘..3$v024‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v024‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, 8, ..., 12, and 13

## Warning: ‘..1$v113‘ and ‘..3$v113‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v113‘.
## x Values: 21 and 72

## Warning: ‘..1$v131‘ and ‘..4$v131‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v131‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, 8, 9, and 996

## Warning: ‘..1$v717‘ and ‘..4$v717‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v717‘.
## x Values: 96

## Warning: ‘..1$m15‘ and ‘..4$m15‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m15‘.
## x Values: 21, 22, 23, 26, 30, 31, 32, and 36

## Warning: ‘..1$v130‘ and ‘..4$v130‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v130‘.
## x Values: 1, 2, and 4
```

```
## Warning: ‘..1$v024‘ and ‘..4$v024‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v024‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, and 8

## Warning: ‘..1$v113‘ and ‘..4$v113‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v113‘.
## x Values: 21

## Warning: ‘..1$v131‘ and ‘..5$v131‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v131‘.
## x Values: 1, 2, 3, 4, 5, 6, 96, and 97

## Warning: ‘..1$v717‘ and ‘..5$v717‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v717‘.
## x Values: 10 and 96

## Warning: ‘..1$m15‘ and ‘..5$m15‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m15‘.
## x Values: 21, 22, 23, 24, 25, 26, 30, 31, ..., 33, and 36

## Warning: ‘..1$v150‘ and ‘..5$v150‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v150‘.
## x Values: 2, 3, 4, 5, 6, 7, 8, and 15

## Warning: ‘..1$v130‘ and ‘..5$v130‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v130‘.
## x Values: 2, 3, and 4

## Warning: ‘..1$v024‘ and ‘..5$v024‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v024‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, and 8

## Warning: ‘..1$v113‘ and ‘..5$v113‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v113‘.
## x Values: 21

## Warning: ‘..1$v131‘ and ‘..6$v131‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v131‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, 8, ..., 16, and 996

## Warning: ‘..1$v717‘ and ‘..6$v717‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v717‘.
## x Values: 0

## Warning: ‘..1$m18‘ and ‘..6$m18‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m18‘.
## x Values: 1, 2, 3, and 4

## Warning: ‘..1$m15‘ and ‘..6$m15‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m15‘.
## x Values: 21, 22, 23, 26, 30, 31, and 36
```

```
## Warning: ‘..1$v130‘ and ‘..6$v130‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v130‘.
## x Values: 1, 2, 3, and 4

## Warning: ‘..1$v024‘ and ‘..6$v024‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v024‘.
## x Values: 1, 2, 3, 4, and 5

## Warning: ‘..1$v113‘ and ‘..6$v113‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v113‘.
## x Values: 21

## Warning: ‘..1$v131‘ and ‘..7$v131‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v131‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, 8, ..., 16, and 23

## Warning: ‘..1$v717‘ and ‘..7$v717‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v717‘.
## x Values: 96

## Warning: ‘..1$m15‘ and ‘..7$m15‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m15‘.
## x Values: 21, 22, 23, 24, 25, 26, 30, 31, ..., 34, and 36

## Warning: ‘..1$v130‘ and ‘..7$v130‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v130‘.
## x Values: 4, 5, 6, and 8

## Warning: ‘..1$v024‘ and ‘..7$v024‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v024‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, 8, and 9

## Warning: ‘..1$v113‘ and ‘..7$v113‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v113‘.
## x Values: 21 and 72

## Warning: ‘..1$v717‘ and ‘..8$v717‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v717‘.
## x Values: 96

## Warning: ‘..1$m15‘ and ‘..8$m15‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m15‘.
## x Values: 21, 22, 23, 26, 30, 31, and 36

## Warning: ‘..1$v024‘ and ‘..8$v024‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v024‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, 8, ..., 13, and 14

## Warning: ‘..1$v113‘ and ‘..8$v113‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v113‘.
## x Values: 21
```

```
## Warning: ‘..1$v131‘ and ‘..9$v131‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v131‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, 8, ..., 10, and 96

## Warning: ‘..1$v717‘ and ‘..9$v717‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v717‘.
## x Values: 10

## Warning: ‘..1$m15‘ and ‘..9$m15‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m15‘.
## x Values: 21, 22, 23, 26, 30, 31, and 36

## Warning: ‘..1$v150‘ and ‘..9$v150‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v150‘.
## x Values: 13, 14, and 15

## Warning: ‘..1$v130‘ and ‘..9$v130‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v130‘.
## x Values: 1, 2, 3, and 4

## Warning: ‘..1$v024‘ and ‘..9$v024‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v024‘.
## x Values: 1, 2, 3, 4, 5, and 6

## Warning: ‘..1$v113‘ and ‘..9$v113‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v113‘.
## x Values: 21 and 92

## Warning: ‘..1$v157‘ and ‘..10$v157‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v157‘.
## x Values: 0, 1, 2, and 3

## Warning: ‘..1$v158‘ and ‘..10$v158‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v158‘.
## x Values: 0, 1, 2, and 3

## Warning: ‘..1$v159‘ and ‘..10$v159‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v159‘.
## x Values: 0, 1, 2, and 3

## Warning: ‘..1$v106‘ and ‘..10$v106‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v106‘.
## x Values: 0, 1, 2, and 3

## Warning: ‘..1$v743a‘ and ‘..10$v743a‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v743a‘.
## x Values: 1, 2, 3, 4, 5, and 6

## Warning: ‘..1$v743b‘ and ‘..10$v743b‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v743b‘.
## x Values: 1, 2, 3, 4, 5, and 6
```

```
## Warning: ‘..1$v743d‘ and ‘..10$v743d‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v743d‘.
## x Values: 1, 2, 3, 4, 5, and 6

## Warning: ‘..1$v744a‘ and ‘..10$v744a‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v744a‘.
## x Values: 0, 1, and 8

## Warning: ‘..1$v744b‘ and ‘..10$v744b‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v744b‘.
## x Values: 0, 1, and 8

## Warning: ‘..1$v744c‘ and ‘..10$v744c‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v744c‘.
## x Values: 0, 1, and 8

## Warning: ‘..1$v744d‘ and ‘..10$v744d‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v744d‘.
## x Values: 0, 1, and 8

## Warning: ‘..1$v744e‘ and ‘..10$v744e‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v744e‘.
## x Values: 0, 1, and 8

## Warning: ‘..1$v717‘ and ‘..10$v717‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v717‘.
## x Values: 0, 1, 2, 3, 4, 5, 6, 7, ..., 9, and 98

## Warning: ‘..1$m18‘ and ‘..10$m18‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m18‘.
## x Values: 1, 2, 3, 4, 5, and 8

## Warning: ‘..1$m19‘ and ‘..10$m19‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m19‘.
## x Values: 9996 and 9998

## Warning: ‘..1$b4‘ and ‘..10$b4‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$b4‘.
## x Values: 1 and 2

## Warning: ‘..1$m15‘ and ‘..10$m15‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m15‘.
## x Values: 10, 11, 12, 20, 21, 22, 23, 24, ..., 36, and 96

## Warning: ‘..1$v150‘ and ‘..10$v150‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v150‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, 8, ..., 12, and 98

## Warning: ‘..1$v151‘ and ‘..10$v151‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v151‘.
## x Values: 1 and 2
```

```
## Warning: ‘..1$v190‘ and ‘..10$v190‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v190‘.
## x Values: 1, 2, 3, 4, and 5

## Warning: ‘..1$v025‘ and ‘..10$v025‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v025‘.
## x Values: 1 and 2

## Warning: ‘..1$v130‘ and ‘..10$v130‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v130‘.
## x Values: 96

## Warning: ‘..1$v445‘ and ‘..10$v445‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v445‘.
## x Values: 9998

## Warning: ‘..1$b5‘ and ‘..10$b5‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$b5‘.
## x Values: 0 and 1

## Warning: ‘..1$v701‘ and ‘..10$v701‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v701‘.
## x Values: 0, 1, 2, 3, and 8

## Warning: ‘..1$v024‘ and ‘..10$v024‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v024‘.
## x Values: 1, 2, 3, 4, 5, 6, 7, and 8

## Warning: ‘..1$v367‘ and ‘..10$v367‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v367‘.
## x Values: 1, 2, and 3

## Warning: ‘..1$v426‘ and ‘..10$v426‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v426‘.
## x Values: 0, 100, 101, 199, 201, and 299

## Warning: ‘..1$v113‘ and ‘..10$v113‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v113‘.
## x Values: 10, 11, 12, 13, 20, 21, 30, 31, ..., 96, and 97

## Warning: ‘..1$v115‘ and ‘..10$v115‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v115‘.
## x Values: 996, 997, and 998

## Warning: ‘..1$v116‘ and ‘..10$v116‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v116‘.
## x Values: 10, 11, 12, 13, 14, 15, 20, 21, ..., 96, and 97

## Warning: ‘..1$m14‘ and ‘..10$m14‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m14‘.
## x Values: 0 and 98
```

```
## Warning: ‘..1$m70‘ and ‘..10$m70‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m70‘.
## x Values: 0, 1, and 8

## Warning: ‘..1$v313‘ and ‘..10$v313‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v313‘.
## x Values: 0, 1, 2, and 3

## Warning: ‘..1$v131‘ and ‘..11$v131‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v131‘.
## x Values: 11, 12, 13, 14, 15, 16, 17, 18, and 96

## Warning: ‘..1$v717‘ and ‘..11$v717‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v717‘.
## x Values: 96

## Warning: ‘..1$m15‘ and ‘..11$m15‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m15‘.
## x Values: 21, 22, 23, 26, 30, 31, and 36

## Warning: ‘..1$v150‘ and ‘..11$v150‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v150‘.
## x Values: 13, 14, and 15

## Warning: ‘..1$v130‘ and ‘..11$v130‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v130‘.
## x Values: 1, 2, 3, 4, and 5

## Warning: ‘..1$v024‘ and ‘..11$v024‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v024‘.
## x Values: 1, 2, 3, 4, and 5

## Warning: ‘..1$v113‘ and ‘..11$v113‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v113‘.
## x Values: 21, 40, and 72

## Warning: ‘..1$v131‘ and ‘..12$v131‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v131‘.
## x Values: 11, 12, 13, 14, 15, 95, and 96

## Warning: ‘..1$v717‘ and ‘..12$v717‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v717‘.
## x Values: 10

## Warning: ‘..1$m15‘ and ‘..12$m15‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$m15‘.
## x Values: 21, 22, 23, 24, 25, 26, 30, 31, 32, and 36

## Warning: ‘..1$v024‘ and ‘..12$v024‘ have conflicting value labels.
## i Labels for these values will be taken from ‘..1$v024‘.
## x Values: 1, 2, 3, 4, 5, and 6
```

```
## Warning: `..1$v113` and `..12$v113` have conflicting value labels.
## i Labels for these values will be taken from `..1$v113`.
## x Values: 13, 21, 40, and 72
```

```
save(dhs_child_west_africa, file = "dhs_child_west_africa.rda")

freq(dhs_child_west_africa$m18)
```

```
## Frequencies
## dhs_child_west_africa$m18
## Label: size of child at birth
## Type: Numeric
##
##                 Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- -------- --------- -------------- --------- --------------
##           1    14839    11.095         11.095    10.339         10.339
##           2    27031    20.210         31.305    18.834         29.173
##           3    67002    50.095         81.400    46.684         75.857
##           4    14148    10.578         91.978     9.858         85.715
##           5     7973     5.961         97.939     5.555         91.270
##           8     2722     2.035         99.974     1.897         93.167
##           9       35     0.026        100.000     0.024         93.191
##        <NA>     9772                               6.809        100.000
##       Total   143522   100.000        100.000   100.000        100.000
```

Noms des variables : - v367 : enfant était voulu (1: voulu à l'époque, 2: voulu pour plus tard, 3: ne plus vouloir du tout) - v426 : quand l'enfant a été mis au sein? - v113 : Source d'eau de boisson - v115 : temps jusqu'à la source d'eau de boisson - v116: type de toilette - v445 : indice de masse corporel - m14: nombre de visite prénatale durant la grossesse - m70: visites postnatales dans les 2 mois mois après l'accouchement (oui/non) - v313: Utilisation de contraceptive (à coder oui/non)

Création de nos différentes variables et leur labelisation

```
data <- dhs_child_west_africa

freq(data$b5)
```

```
## Frequencies
## data$b5
## Label: child is alive
## Type: Numeric
##
##                 Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- -------- --------- -------------- --------- --------------
##           0    10261      7.15           7.15      7.15           7.15
##           1   133261     92.85         100.00     92.85         100.00
##        <NA>        0                               0.00         100.00
##       Total   143522    100.00         100.00    100.00         100.00
```

```r
data <- data %>%
  mutate(dead = if_else(b5 == 0, 1, 0)) %>%
  mutate(dead = factor(dead,levels=c(0, 1),labels = c("alive","dead")),
         ponderation = v005/1000000)


freq(data$dead)
```

```
## Frequencies
## data$dead
## Type: Factor
##
##                Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- -------- --------- -------------- --------- --------------
##       alive   133261     92.85          92.85     92.85          92.85
##        dead    10261      7.15         100.00      7.15         100.00
##        <NA>        0                                0.00         100.00
##       Total   143522    100.00         100.00    100.00         100.00
```

- Poids qualitatif

```r
freq(data$m18)
```

```
## Frequencies
## data$m18
## Label: size of child at birth
## Type: Numeric
##
##                Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- -------- --------- -------------- --------- --------------
##           1    14839    11.095         11.095    10.339         10.339
##           2    27031    20.210         31.305    18.834         29.173
##           3    67002    50.095         81.400    46.684         75.857
##           4    14148    10.578         91.978     9.858         85.715
##           5     7973     5.961         97.939     5.555         91.270
##           8     2722     2.035         99.974     1.897         93.167
##           9       35     0.026        100.000     0.024         93.191
##        <NA>     9772                                6.809        100.000
##       Total   143522   100.000        100.000   100.000        100.000
```

```r
freq(data$midx)
```

```
## Frequencies
## data$midx
## Label: index to birth history
## Type: Numeric
##
##                Freq    % Valid   % Valid Cum.    % Total   % Total Cum.
## ----------- -------- ---------- -------------- ---------- --------------
##           1    92561    69.1879        69.1879    64.4926        64.4926
##           2    35999    26.9087        96.0966    25.0826        89.5751
```

```
##           3     4876     3.6447        99.7414     3.3974        92.9725
##           4      312     0.2332        99.9746     0.2174        93.1899
##           5       31     0.0232        99.9978     0.0216        93.2115
##           6        3     0.0022       100.0000     0.0021        93.2136
##        <NA>     9740                              6.7864       100.0000
##       Total   143522   100.0000       100.0000   100.0000       100.0000
```

```
data <- data %>%
  mutate(poids_quali = factor(case_when(
    m18 == 1 ~ "Very large",
    m18 == 2 ~ "Larger than average",
    m18 == 3 ~ "Average",
    m18 == 4 ~ "Smaller than average",
    m18 == 5 ~ "Very small",
    m18 == 8 | m18 == 9 ~ NA_character_,
    is.na(m18) ~ "à exclure"
  )))

ctable(data$poids_quali, factor(data$m18), "no")
```

```
## Cross-Tabulation
## poids_quali * factor(data$m18)
##
## --------------------- ------------------ ------- ------- ------- ------- ------ ------ ---- ------
##                          factor(data$m18)       1       2       3       4      5      8    9   <NA>
##            poids_quali
##              à exclure                          0       0       0       0      0      0    0   9772
##                 Average                         0       0   67002       0      0      0    0      0
##    Larger than average                          0   27031       0       0      0      0    0      0
##   Smaller than average                          0       0       0   14148      0      0    0      0
##             Very large                      14839       0       0       0      0      0    0      0
##             Very small                          0       0       0       0   7973      0    0      0
##                   <NA>                          0       0       0       0      0   2722   35      0
##                  Total                      14839   27031   67002   14148   7973   2722   35   9772
## --------------------- ------------------ ------- ------- ------- ------- ------ ------ ---- ------
```

education de la femme (v106):

```
freq(data$v106)
```

```
## Frequencies
## data$v106
## Label: highest educational level
## Type: Numeric
##
##                Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- -------- --------- -------------- --------- --------------
##           0    84043    58.558         58.558    58.558         58.558
##           1    26867    18.720         77.277    18.720         77.277
##           2    28270    19.697         96.975    19.697         96.975
##           3     4321     3.011         99.985     3.011         99.985
##           9       21     0.015        100.000     0.015        100.000
##        <NA>        0                              0.000        100.000
##       Total   143522   100.000        100.000   100.000        100.000
```

```
data <- data %>% mutate(educ = factor(v106, levels = c(0,1,2,3), labels = c("Sans instruction","Primair
freq(data$educ)
```

```
## Frequencies
## data$educ
## Type: Factor
##
##                          Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ---------------------- -------- --------- -------------- --------- --------------
##       Sans instruction    84043    58.566         58.566    58.558         58.558
##               Primaire    26867    18.723         77.289    18.720         77.277
##             Secondaire    28270    19.700         96.989    19.697         96.975
##              Supérieur     4321     3.011        100.000     3.011         99.985
##                   <NA>       21                               0.015        100.000
##                  Total   143522   100.000        100.000   100.000        100.000
```

statut d'emploi (v717) [recoder comme suit : 0:Entrepreneures agricoles, 1:Travailleurs qualifiés ou non qualifiés

2:Sans emploi

```
data <- data %>% mutate(activite = case_when(v717 == 4 ~ 0,
                                             v717 == 1 | v717 == 2 | v717 == 3
                                             | v717 == 7 | v717 == 8 | v717 == 9 ~ 1,
                                             v717 == 0 ~ 2),
                        activite=factor(activite, levels=c(0:2),labels=c("Entrepreneures agricoles","Trav


freq(data$activite)
```

```
## Frequencies
## data$activite
## Type: Factor
##
##                                                Freq   % Valid   % Valid Cum.   % Total   % Tota
## ---------------------------------------------- -------- --------- -------------- --------- -------
##                          Entrepreneures agricoles    25485     20.86          20.86     17.76
##       Travailleuses qualifiées ou non qualifiées    54883     44.92          65.78     38.24
##                                     Sans emploi    41802     34.22         100.00     29.13
##                                            <NA>    21352                              14.88
##                                           Total   143522    100.00         100.00    100.00
```

attitude face à la violence (v744a,v744b,v744c,v744d,v744e) - justifié de battre si la femme sort sans averti son mari (V744a) : 0 : Non 1 : Oui 8 : missing - justifié de battre si la femme si elle néglige les enfants (v744b) : 0 : Non 1 : Oui 8 : missing - Justifié de battre si la femme se dispute avec son mari(v744c) : 0 : Non 1 : Oui 8 : missing - Justifié de battre si la femme refuse le sexe à son mari(v744d) : 0 : Non 1 : Oui 8 : missing - Justifié de battre si la femme brule la nourriture (v744e) : 0 : Non 1 : Oui 8 : missing

```
data$v744a[data$v744a == 8] <- NA
data$v744b[data$v744b == 8] <- NA
data$v744c[data$v744c == 8] <- NA
```

```r
data$v744d[data$v744d == 8] <- NA
data$v744e[data$v744e == 8] <- NA

data <- data %>% mutate(attitude_violence = v744a+v744b+v744c+v744d+v744e,
                        attitude_violence = case_when(attitude_violence==0~1,
                                                      attitude_violence %in% c(1:5)~2),
                        attitude_violence=factor(attitude_violence, levels = c(1,2),labels=c("Non favoral



data %>% filter(is.na(attitude_violence)) %>% select(v744a,v744b,v744c,v744d,v744e,attitude_violence)
```

```
## # A tibble: 2,774 x 6
##    v744a     v744b     v744c     v744d     v744e     attitude_violence
##    <dbl+lbl> <dbl+lbl> <dbl+lbl> <dbl+lbl> <dbl+lbl> <fct>
##  1 NA        NA         1 [yes]   1 [yes]   0 [no]   <NA>
##  2 NA        NA        NA        NA        NA        <NA>
##  3 NA        NA        NA        NA        NA        <NA>
##  4 NA        NA         1 [yes]   0 [no]    0 [no]   <NA>
##  5 NA        NA         1 [yes]   0 [no]    0 [no]   <NA>
##  6  0 [no]   NA         1 [yes]   1 [yes]   0 [no]   <NA>
##  7  1 [yes]   1 [yes]  NA         1 [yes]   1 [yes]  <NA>
##  8  0 [no]   NA         0 [no]    0 [no]    0 [no]   <NA>
##  9  1 [yes]   1 [yes]   1 [yes]  NA         1 [yes]  <NA>
## 10  1 [yes]   1 [yes]   1 [yes]  NA         1 [yes]  <NA>
## # i 2,764 more rows
```

```r
freq(data$attitude_violence)
```

```
## Frequencies
## data$attitude_violence
## Type: Factor
##
##                        Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## -------------------- -------- --------- -------------- --------- --------------
##       Non favorable    76909     54.64          54.64     53.59          53.59
##           Favorable    63839     45.36         100.00     44.48          98.07
##                <NA>     2774                               1.93         100.00
##               Total   143522    100.00         100.00    100.00         100.00
```

pouvoir decisionnel de la femme au sein du ménage (v743a,v743b,v743d) - Prise de décision concernant les soins de santé (v743a) : 0 : femme seule 1 : femme et mari/partenaire 2 : mari/partenaire seul 3: [5,6] Autres personnes NA: missing - Prise de décision concernant les achats du ménages (v743b) : 0 : femme seule 1 : femme et mari/partenaire 2 : mari/partenaire seul 3 : [5,6] Autres personnes NA: missing

- Prise de décision concernant la visite en famille (v743d) : 0 : femme seule 1 : femme et mari/partenaire 2 : mari/partenaire seul 3 : Autre personnes NA: missing

```r
freq(data$v743d)
```

```
## Frequencies
```

```
## data$v743d
## Label: person who usually decides on: visits to family or relatives
## Type: Numeric
##
##                  Freq    % Valid   % Valid Cum.    % Total   % Total Cum.
## ----------- -------- ---------- -------------- ---------- --------------
##           1    17867    13.4390        13.4390    12.4490        12.4490
##           2    51033    38.3854        51.8244    35.5576        48.0066
##           4    62941    47.3422        99.1666    43.8546        91.8612
##           5      829     0.6235        99.7901     0.5776        92.4388
##           6      276     0.2076        99.9977     0.1923        92.6311
##           9        3     0.0023       100.0000     0.0021        92.6332
##        <NA>    10573                              7.3668       100.0000
##       Total   143522   100.0000       100.0000   100.0000       100.0000
```

```
data <- data %>%
  mutate(decision_sante = case_when(
    v743a == 1 ~ 0,
    v743a == 2 ~ 1,
    v743a == 4 ~ 2,
    v743a == 5 | v743a == 6 ~ 3),
    decision_achat = case_when(
      v743b == 1 ~ 0,
      v743b == 2 ~ 1,
      v743b == 4 ~ 2, v743b == 5 | v743b == 6 ~ 3),
    decision_visite=case_when(
      v743d==1~0,
      v743d==2~1,
      v743d==4~2,
      v743d==5 | v743d==6 ~3),
    pouvoir_decision=decision_sante+decision_achat+decision_visite,
    pouvoir_decision=case_when(
      pouvoir_decision<2~1,
      pouvoir_decision==2 |pouvoir_decision==3~2,
      pouvoir_decision >=4~3),
    pouvoir_decision=factor(pouvoir_decision,levels=c(1:3),labels=c("Elévé","Moyen","Faible")))

freq(data$pouvoir_decision)
```

```
## Frequencies
## data$pouvoir_decision
## Type: Factor
##
##                   Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ------------ -------- --------- -------------- --------- --------------
##        Elévé     6516      4.90           4.90      4.54           4.54
##        Moyen    38374     28.87          33.77     26.74          31.28
##       Faible    88050     66.23         100.00     61.35          92.63
##         <NA>    10582                              7.37         100.00
##        Total   143522    100.00         100.00    100.00         100.00
```

- Religion de la mère (v130)
- Age de la mère (v012)

- Sexe de l'enfant (b4)
- Poids de l'enfant à la naissance (m19)
- Rang de l'enfant à la naissance (bord)
- Intervalle entre l'enfant et la naissance précédente (b11)
- Lieu d'accouchement (m15)
- Niveau d'instruction du conjoint (v701)

```
data <- data %>% mutate(religion=case_when(v130==1 | v130==2 | v130==3~1,
                                           v130==4~2,
                                           v130==5 | v130==7 | v130==96~3
                                             ),
                        religion=factor(religion, levels = c(1:3),labels = c("Chrétien","Musulman","An
                        age_mere=case_when(v012<20~1,
                                           v012 >=20 & v012<30~2,
                                           v012 >=30 & v012<40~3,
                                           v012 >=40~4),
                      age_mere=factor(age_mere,levels = c(1:4),labels = c("Moins de 20 ans","Entre 20
                      sex_enfant=factor(if_else(b4==1,1,2),levels = c(1,2), labels = c("Masculin","Fem

                      poids_nais = case_when(m19<2500~1,
                                             m19 >=2500 & m19<4000~2,
                                             m19 >=4000 & m19<7500~3,
                                             m19==9996 | m19==9998~ NA_real_),
                      poids_nais=factor(poids_nais,levels = c(1:3), labels = c("Faible","Normal","Elev
                      rang_naiss=case_when(bord==1~1,
                                             bord==2 | bord==3~2,
                                             bord >=4~3),

                      rang_naiss=factor(rang_naiss,levels = c(1:3),labels = c("Premier né","Rang 2 ou

                      interval_precedent=case_when(b11<24~1,
                                                   b11 >=24~2,

                                                   is.na(b11)~3),
                      interval_precedent=factor(interval_precedent,levels = c(1:3),labels = c("Moins d

                      lieu_accouch=factor(if_else(m15==10 | m15==11 | m15==12,1,2), levels = c(1,2),la

data <- data %>% mutate(ins_conj=factor(v701,levels=c(0:3),labels=c("Sans instruction","Primaire","Seco
```

- v367 : enfant était voulu (1: voulu à l'époque, 2: voulu pour plus tard, 3: ne plus vouloir du tout)
- v426 : quand l'enfant a été mis au sein? (rearranger comme suit :mis au sein dans les une heures suivant sa naissance ?)
- v113 : Source d'eau de boisson
- v116: type de toilette
- v445 : indice de masse corporel
- m14: nombre de visite prénatale durant la grossesse
- m70: visites postnatales dans les 2 mois mois après l'accouchement (oui/non)
- v313: Utilisation de contraceptive (à coder oui/non)

```
data <- data %>% mutate(naissance_voulu=factor(v367,levels=c(1:3),labels=c("avait voulu","voulu pour plu
                        allaiter_heure=case_when(v426==0 | v426==100~1,
                                                 v426==101 | v426==102 |v426==102 | v426==103 |v426==103
```

```
                        allaiter_heure=factor(allaiter_heure,levels =c(1,2),labels=c("Oui","Non")),
                        source_eau=if_else(v113==10 |v113==11 |v113==12 |v113==13 |v113==14 |v113==41 |v
                    source_eau=factor(source_eau,levels = c(1,2),labels=c("Amelioree","Non amelioree")),
                    type_toilet=if_else(v116==10 |v116==11| v116==12 |v116==13 |v116==14 |v116==15 |v116
                    type_toilet=factor(type_toilet,levels = c(1,2),labels=c("Amelioree","Non amelioree"))

                    contraception=if_else(v313==0,1,2),
                    contraception=factor(contraception,levels=c(1,2),labels=c("Non","Oui")))
```

- Degré d'exposition de la femme aux médias (à partir de v157,v158,v159)

```
freq(data$v159)
```

```
## Frequencies
## data$v159
## Label: frequency of watching television
## Type: Numeric
##
##                 Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- -------- --------- -------------- --------- --------------
##           0    85707    59.724         59.724    59.717         59.717
##           1    22851    15.924         75.648    15.922         75.639
##           2    33727    23.502         99.151    23.500         99.138
##           3     1178     0.821         99.971     0.821         99.959
##           9       41     0.029        100.000     0.029         99.987
##        <NA>       18                               0.013        100.000
##       Total   143522   100.000        100.000   100.000        100.000
```

```
data <- data %>% mutate(degmedia=v157+v158+v159,
                        degmedia=case_when(degmedia==0~0,
                                           degmedia==1 | degmedia==2 ~1,
                                           degmedia==3 | degmedia==4~2,
                                           degmedia==5 | degmedia==6~3),
                        degmedia=factor(degmedia, levels =c(0:3),labels =c("Nul","Faible","moyenne","Ele

freq(data$degmedia)
```

```
## Frequencies
## data$degmedia
## Type: Factor
##
##                  Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ------------- -------- --------- -------------- --------- --------------
##           Nul    51796     36.17          36.17     36.09          36.09
##        Faible    54895     38.33          74.50     38.25          74.34
##       moyenne    31370     21.90          96.40     21.86          96.20
##         Elevé     5154      3.60         100.00      3.59          99.79
##          <NA>      307                               0.21         100.00
##         Total   143522    100.00         100.00    100.00         100.00
```

Caractéristiques des ménages

Taille du menage (v136) Sexe du chef de ménage (v151) Niveau de vie du ménage (v190)

```
data <- data %>% mutate(taille_menage=case_when(v136<=3~1,
                                                v136 >=4 & v136<=6~2,
                                                v136 >=7~3),
                        taille_menage=factor(taille_menage,levels = c(1:3),labels = c("2-3","4-6","7 et p
                        sex_chef=factor(if_else(v151==1,1,2),levels = c(1,2), labels = c("Masculin","Femi
                        niveau_vie=case_when(v190==1 | v190==2~1,
                                             v190==3~2,
                                             v190==4 |v190==5~3),
                      niveau_vie=factor(niveau_vie,levels=c(1:3),labels = c("Pauvre","Moyen","Riche"))
                        )

freq(data$niveau_vie)
```

```
## Frequencies
## data$niveau_vie
## Type: Factor
##
##                  Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ------------ -------- --------- -------------- --------- --------------
##       Pauvre    66260     46.17          46.17     46.17          46.17
##        Moyen    29967     20.88          67.05     20.88          67.05
##        Riche    47295     32.95         100.00     32.95         100.00
##         <NA>        0                               0.00         100.00
##        Total   143522    100.00         100.00    100.00         100.00
```

Caractéristiques Communautaires

Milieu de résidence (v025)

```
data <- data %>% mutate(milieu_residence=factor(v025,levels = c(1,2),labels = c("Urbain","Rural")))

freq(data$milieu_residence)
```

```
## Frequencies
## data$milieu_residence
## Type: Factor
##
##                  Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ------------ -------- --------- -------------- --------- --------------
##       Urbain    47664     33.21          33.21     33.21          33.21
##        Rural    95858     66.79         100.00     66.79         100.00
##         <NA>        0                               0.00         100.00
##        Total   143522    100.00         100.00    100.00         100.00
```

```
#save(data, file = "01-pretatement/code/dhs_child_west_africa.rda")

#save(data, file = "../../02_Apprentissage_Supervise/01_Regression/dhs_child_west_africa.rda")
```

```r
## Tri à plat des variables

table1(~ dead+educ+activite+attitude_violence+pouvoir_decision+religion+age_mere+degmedia+ins_conj+sex_
```

```
## Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

	Overall
	(N=143522)
dead	
alive	133261 (92.9%)
dead.1	10261 (7.1%)
educ	
Sans instruction	84043 (58.6%)
Primaire	26867 (18.7%)
Secondaire	28270 (19.7%)
Supérieur	4321 (3.0%)
Missing	21 (0.0%)
activite	
Entrepreneures agricoles	25485 (17.8%)
Travailleuses qualifiées ou non qualifiées	54883 (38.2%)
Sans emploi	41802 (29.1%)
Missing	21352 (14.9%)
attitude_violence	
Non favorable	76909 (53.6%)
Favorable	63839 (44.5%)
Missing	2774 (1.9%)
pouvoir_decision	
Elévé	6516 (4.5%)
Moyen	38374 (26.7%)
Faible	88050 (61.3%)
Missing	10582 (7.4%)
religion	
Chrétien	99512 (69.3%)
Musulman	5780 (4.0%)
Animistes et autres	2585 (1.8%)
Missing	35645 (24.8%)
age_mere	
Moins de 20 ans	8024 (5.6%)
Entre 20 et 29 ans	68000 (47.4%)
Entre 30 ans et 39 ans	53494 (37.3%)
40 ans et plus	14004 (9.8%)
degmedia	
Nul	51796 (36.1%)
Faible	54895 (38.2%)
moyenne	31370 (21.9%)
Elevé	5154 (3.6%)
Missing	307 (0.2%)
ins_conj	
Sans instruction	72908 (50.8%)
Primaire	18152 (12.6%)
Secondaire	27567 (19.2%)
Supérieur	9125 (6.4%)
Missing	15770 (11.0%)
sex_enfant	
Masculin	72978 (50.8%)
Feminin	70544 (49.2%)
poids_nais	
Faible	6174 (4.3%)
Normal	44922 (31.3%)
Elevé	6482 (4.5%)
Missing	85944 (59.9%)

Manque deux autres varieables

- rang de naissance de l'enfant
- grossesse gémélaire
- Age à la grossesse