Cours 06:Manipulation des données

DIOP_Ousseynou

2023-04-16

Manipulation des données

Importation

```
datapath <- "C:\\Users\\dell\\Desktop\\ENSAE\\ISEP2\\Semestre_2\\Programmation R\\EHCVM"
library(haven)
```

```
## Warning: le package 'haven' a été compilé avec la version R 4.2.3
```

```
cereal <- read_dta(file = paste0(datapath,"\\cereales.dta"))
```

Autre méthode d'importation des données avec la bibliothèque Foreign

- library(foreign)
- cereal1 <- read.dta(file = paste0(datapath,"\cereales.dta"), convert.dates = TRUE, convert.factors = TRUE)

Comprendre la structure des données

```
## [1] 11114    14
```

followed by the number of columns.

```
names(cereal) # donne le nom des colones
```

```
##  [1] "interview__key"         "interview__id"          "cereales__id"
##  [4] "s07Bq02_autre_cereales" "s07Bq03a_cereales"      "s07Bq03b_cereales"
##  [7] "s07Bq03c_cereales"      "s07Bq04_cereales"       "s07Bq05_cereales"
## [10] "s07Bq06_cereales"       "s07Bq07a_cereales"      "s07Bq07b_cereales"
## [13] "s07Bq07c_cereales"      "s07Bq08_cereales"
```

```
str(cereal) # donne la structuration des données,
```

```
## tibble [11,114 x 14] (S3: tbl_df/tbl/data.frame)
##  $ interview__key        : chr [1:11114] "55-75-97-43" "55-75-97-43" "55-75-97-43" "14-16-03-21" ...
##   ..- attr(*, "format.stata")= chr "%-9s"
##  $ interview__id         : chr [1:11114] "3827f11c978e42e6aa4b5b8ed9aa1348" "3827f11c978e42e6aa4b5b8
##   ..- attr(*, "format.stata")= chr "%-9s"
##  $ cereales__id          : dbl+lbl [1:11114]  1,  7, 20,  4, 22,  1,  7, 16, 21,  1, 20, 21,  1, ...
##    ..@ label       : chr "Id in cereales"
##    ..@ format.stata: chr "%10.0g"
##    ..@ labels      : Named num [1:27] 1 2 3 4 5 6 7 8 9 10 ...
##    .. ..- attr(*, "names")= chr [1:27] "Riz local brisé" "Riz local entier" "Riz importé brisé" "Riz
##  $ s07Bq02_autre_cereales: chr [1:11114] "" "" "" "" ...
##   ..- attr(*, "format.stata")= chr "%-9s"
##  $ s07Bq03a_cereales     : num [1:11114] 9 7 2 7.5 42 14 10 0.25 7 21 ...
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ s07Bq03b_cereales     : dbl+lbl [1:11114] 100, 100, 139, 100, 568, 100, 100, 100, 568, 100, 12...
##    ..@ label       : chr "7B.03b. Unité"
##    ..@ format.stata: chr "%10.0g"
##    ..@ labels      : Named num [1:25] 100 108 109 115 123 125 126 129 135 136 ...
##    .. ..- attr(*, "names")= chr [1:25] "Kg" "Boite de tomate" "Bol" "Calebasse" ...
##  $ s07Bq03c_cereales     : dbl+lbl [1:11114] 0, 0, 2, 0, 3, 0, 0, 0, 3, 0, 2, 3, 0, 2, 0, 3, 0, 0...
##    ..@ label       : chr "7B.03c. Taille"
##    ..@ format.stata: chr "%10.0g"
##    ..@ labels      : Named num [1:8] 0 1 2 3 4 5 6 7
##    .. ..- attr(*, "names")= chr [1:8] "Taille unique" "Petit" "Moyen" "Grand" ...
##  $ s07Bq04_cereales      : num [1:11114] 0 7 NA NA 0 0 10 0 0 0 ...
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ s07Bq05_cereales      : num [1:11114] 0 0 0 0 0 0 0 0 0 0 ...
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ s07Bq06_cereales      : dbl+lbl [1:11114] 2, 4, 2, 1, 1, 2, 4, 1, 1, 1, 2, 1, 4, 4, 2, 1, 1, 4...
##    ..@ label       : chr "7B.06. Quelle est la dernière fois que le %rostertitle% a été acheté dans l
##    ..@ format.stata: chr "%10.0g"
##    ..@ labels      : Named num [1:5] 1 2 3 4 5
##    .. ..- attr(*, "names")= chr [1:5] "Hier" "7 dernier jours" "30 derniers jours" "Plus de 30 jours"
##  $ s07Bq07a_cereales     : num [1:11114] 1 NA 2 2.5 6 2 NA 0.25 1 3 ...
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ s07Bq07b_cereales     : dbl+lbl [1:11114] 100,  NA, 139, 100, 568, 100,  NA, 100, 568, 100, 12...
##    ..@ label       : chr "7B.07b. Unité"
##    ..@ format.stata: chr "%10.0g"
##    ..@ labels      : Named num [1:25] 100 108 109 115 123 125 126 129 135 136 ...
##    .. ..- attr(*, "names")= chr [1:25] "Kg" "Boite de tomate" "Bol" "Calebasse" ...
##  $ s07Bq07c_cereales     : dbl+lbl [1:11114]  0, NA,  2,  0,  3,  0, NA,  0,  3,  0,  2,  3, NA, ...
##    ..@ label       : chr "7B.07c. Taille"
##    ..@ format.stata: chr "%10.0g"
##    ..@ labels      : Named num [1:7] 0 1 2 3 4 5 6
##    .. ..- attr(*, "names")= chr [1:7] "Taille unique" "Petit" "Moyen" "Grand" ...
##  $ s07Bq08_cereales      : num [1:11114] 350 NA 600 750 1050 325 NA 200 200 1200 ...
##   ..- attr(*, "format.stata")= chr "%10.0g"
```

fourni un résumé utile et compact de sa structure interne.

alternative à str() avec The dplyr package offers a slightly different flavor of str().

```
library(dplyr)
```

```
## Warning: le package 'dplyr' a été compilé avec la version R 4.2.3
```

```
##
## Attachement du package : 'dplyr'
```

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

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

```
glimpse(cereal)
```

```
## Rows: 11,114
## Columns: 14
## $ interview__key         <chr> "55-75-97-43", "55-75-97-43", "55-75-97-43", "1~
## $ interview__id          <chr> "3827f11c978e42e6aa4b5b8ed9aa1348", "3827f11c97~
## $ cereales__id           <dbl+lbl>  1,  7, 20,  4, 22,  1,  7, 16, 21,  1, 20,~
## $ s07Bq02_autre_cereales <chr> "", "", "", "", "", "", "", "", "", "", "",~
## $ s07Bq03a_cereales      <dbl> 9.00, 7.00, 2.00, 7.50, 42.00, 14.00, 10.00, 0.~
## $ s07Bq03b_cereales      <dbl+lbl> 100, 100, 139, 100, 568, 100, 100, 100, 568~
## $ s07Bq03c_cereales      <dbl+lbl> 0, 0, 2, 0, 3, 0, 0, 0, 3, 0, 2, 3, 0, 2, 0~
## $ s07Bq04_cereales       <dbl> 0, 7, NA, NA, 0, 0, 10, 0, 0, 0, NA, 0, 0, 7, 0~
## $ s07Bq05_cereales       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ s07Bq06_cereales       <dbl+lbl> 2, 4, 2, 1, 1, 2, 4, 1, 1, 1, 2, 1, 4, 4, 2~
## $ s07Bq07a_cereales      <dbl> 1.00, NA, 2.00, 2.50, 6.00, 2.00, NA, 0.25, 1.0~
## $ s07Bq07b_cereales      <dbl+lbl> 100,  NA, 139, 100, 568, 100,  NA, 100, 568~
## $ s07Bq07c_cereales      <dbl+lbl>  0, NA,  2,  0,  3,  0, NA,  0,  3,  0,  2,~
## $ s07Bq08_cereales       <dbl> 350, NA, 600, 750, 1050, 325, NA, 200, 200, 120~
```

une autre façon de voir globalement la structuration

```
summary(cereal)  # plus exhaustive ;
```

```
##  interview__key     interview__id       cereales__id    s07Bq02_autre_cereales
##  Length:11114       Length:11114       Min.   :  1.00   Length:11114
##  Class :character   Class :character   1st Qu.:  3.00   Class :character
##  Mode  :character   Mode  :character   Median :  7.00   Mode  :character
##                                        Mean   : 12.13
```

```
##                                      3rd Qu.: 21.00
##                                      Max.   :169.00
##
##  s07Bq03a_cereales s07Bq03b_cereales s07Bq03c_cereales s07Bq04_cereales
##  Min.   :  0.05    Min.   :100.0     Min.   :0.0000    Min.   : 0.000
##  1st Qu.:  3.00    1st Qu.:100.0     1st Qu.:0.0000    1st Qu.: 0.000
##  Median :  8.00    Median :100.0     Median :0.0000    Median : 0.000
##  Mean   : 11.58    Mean   :226.7     Mean   :0.9865    Mean   : 1.014
##  3rd Qu.: 14.00    3rd Qu.:568.0     3rd Qu.:2.0000    3rd Qu.: 0.000
##  Max.   :168.00    Max.   :571.0     Max.   :7.0000    Max.   :91.000
##  NA's   :10        NA's   :10        NA's   :10        NA's   :2690
##  s07Bq05_cereales  s07Bq06_cereales s07Bq07a_cereales s07Bq07b_cereales
##  Min.   : 0.0000   Min.   :1.000    Min.   :  0.050   Min.   :100.0
##  1st Qu.: 0.0000   1st Qu.:1.000    1st Qu.:  1.000   1st Qu.:100.0
##  Median : 0.0000   Median :2.000    Median :  2.000   Median :138.0
##  Mean   : 0.3396   Mean   :2.052    Mean   :  4.153   Mean   :250.4
##  3rd Qu.: 0.0000   3rd Qu.:3.000    3rd Qu.:  3.000   3rd Qu.:568.0
##  Max.   :70.0000   Max.   :5.000    Max.   :600.000   Max.   :571.0
##  NA's   :10        NA's   :10       NA's   :1619      NA's   :1619
##  s07Bq07c_cereales s07Bq08_cereales
##  Min.   :0.000     Min.   :    5
##  1st Qu.:0.000     1st Qu.:  350
##  Median :0.000     Median :  600
##  Mean   :1.023     Mean   : 3652
##  3rd Qu.:2.000     3rd Qu.: 1400
##  Max.   :3.000     Max.   :75000
##  NA's   :1619      NA's   :1619
```

Voir les données

```
head(cereal, n=15) # affiche les 15 premières lignes;
```

```
## # A tibble: 15 x 14
##    interview__key interview__id                cereales__id s07Bq02_autre_cereales
##    <chr>          <chr>                        <dbl+lbl>    <chr>
##  1 55-75-97-43    3827f11c978e42e6aa4b5b8ed~  1 [Riz loc~ ""
##  2 55-75-97-43    3827f11c978e42e6aa4b5b8ed~  7 [Mil]     ""
##  3 55-75-97-43    3827f11c978e42e6aa4b5b8ed~ 20 [Pâtes a~ ""
##  4 14-16-03-21    65f04dfa96ce495788e6f3e97~  4 [Riz imp~ ""
##  5 14-16-03-21    65f04dfa96ce495788e6f3e97~ 22 [Pain tr~ ""
##  6 71-36-24-02    92d67f0681fb43f9a06ca89b6~  1 [Riz loc~ ""
##  7 71-36-24-02    92d67f0681fb43f9a06ca89b6~  7 [Mil]     ""
##  8 71-36-24-02    92d67f0681fb43f9a06ca89b6~ 16 [Farine ~ ""
##  9 71-36-24-02    92d67f0681fb43f9a06ca89b6~ 21 [Pain mo~ ""
## 10 87-87-56-15    61889f042e6e4c7aa8c5acab9~  1 [Riz loc~ ""
## 11 87-87-56-15    61889f042e6e4c7aa8c5acab9~ 20 [Pâtes a~ ""
## 12 87-87-56-15    61889f042e6e4c7aa8c5acab9~ 21 [Pain mo~ ""
## 13 44-02-17-24    b469453f8410449290ba3cd8c~  1 [Riz loc~ ""
## 14 44-02-17-24    b469453f8410449290ba3cd8c~  6 [Maïs en~ ""
## 15 44-02-17-24    b469453f8410449290ba3cd8c~  7 [Mil]     ""
## # i 10 more variables: s07Bq03a_cereales <dbl>, s07Bq03b_cereales <dbl+lbl>,
## #   s07Bq03c_cereales <dbl+lbl>, s07Bq04_cereales <dbl>,
```

```
## #   s07Bq05_cereales <dbl>, s07Bq06_cereales <dbl+lbl>,
## #   s07Bq07a_cereales <dbl>, s07Bq07b_cereales <dbl+lbl>,
## #   s07Bq07c_cereales <dbl+lbl>, s07Bq08_cereales <dbl>
```

```
tail(cereal, n=10) # affiche les 10 dernières lignes
```

```
## # A tibble: 10 x 14
##    interview__key interview__id              cereales__id s07Bq02_autre_cereales
##    <chr>          <chr>                      <dbl+lbl>    <chr>
##  1 97-81-02-07    95420c78a69f4cfc83cf43e14~  1 [Riz loc~ ""
##  2 83-62-00-16    3293c4c1692c41f7ae5f74182~  1 [Riz loc~ ""
##  3 83-62-00-16    3293c4c1692c41f7ae5f74182~ 21 [Pain mo~ ""
##  4 05-14-36-00    8295c0da027a4224a2100ea84~  3 [Riz imp~ ""
##  5 05-14-36-00    8295c0da027a4224a2100ea84~ 20 [Pâtes a~ ""
##  6 05-14-36-00    8295c0da027a4224a2100ea84~ 21 [Pain mo~ ""
##  7 19-00-93-55    277c4a049a1b4bcc8493625fe~  3 [Riz imp~ ""
##  8 19-00-93-55    277c4a049a1b4bcc8493625fe~ 15 [semoule~ ""
##  9 19-00-93-55    277c4a049a1b4bcc8493625fe~ 16 [Farine ~ ""
## 10 19-00-93-55    277c4a049a1b4bcc8493625fe~ 21 [Pain mo~ ""
## # i 10 more variables: s07Bq03a_cereales <dbl>, s07Bq03b_cereales <dbl+lbl>,
## #   s07Bq03c_cereales <dbl+lbl>, s07Bq04_cereales <dbl>,
## #   s07Bq05_cereales <dbl>, s07Bq06_cereales <dbl+lbl>,
## #   s07Bq07a_cereales <dbl>, s07Bq07b_cereales <dbl+lbl>,
## #   s07Bq07c_cereales <dbl+lbl>, s07Bq08_cereales <dbl>
```

```
View(cereal) # affiche la base (en quelques lignes)
```

Convertir en data frame

```
typeof(cereal)
```

```
## [1] "list"
```

```
class(cereal)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

```
cereal_df <- data.frame(cereal)
class(cereal_df)
```

```
## [1] "data.frame"
```

charger la table de conversion

```
tableconversion <- "C:\\Users\\dell\\Desktop\\ENSAE\\ISEP2\\Semestre_2\\Programmation R\\EHCVM\\ressour
library(readxl)
```

```
## Warning: le package 'readxl' a été compilé avec la version R 4.2.3
```

```
Sys.setenv(TZ='GMT') # set time zone
base_table <- read_excel(paste0(tableconversion,"\\Table_de_conversion_phase_2.xlsx"))
```

```
## New names:
## * `` -> `...8`
## * `` -> `...9`
```

```
str(base_table)
```

```
## tibble [1,367 x 9] (S3: tbl_df/tbl/data.frame)
##  $ produitID : num [1:1367] 1 1 1 1 1 1 1 1 2 2 ...
##  $ produitNom: chr [1:1367] "Riz local brisé" "Riz local brisé" "Riz local brisé" "Riz local brisé"
##  $ uniteID   : num [1:1367] 100 108 108 108 108 136 138 139 100 108 ...
##  $ uniteNom  : chr [1:1367] "Kg" "Boite de tomate/Pot" "Boite de tomate/Pot" "Boite de tomate/Pot" .
##  $ tailleID  : num [1:1367] 0 1 2 3 7 0 0 2 0 1 ...
##  $ tailleNom : chr [1:1367] "taille unique" "Petit" "Moyen" "Grand" ...
##  $ poids     : chr [1:1367] "1000" "521.5" "935" "1997.5" ...
##  $ ...8      : chr [1:1367] NA NA NA NA ...
##  $ ...9      : chr [1:1367] NA NA NA NA ...
```

```
base_table <- data.frame(base_table)
```

Renomer les variables

```
colnames(cereal_df) # affiche le nom des variables
```

```
##  [1] "interview__key"         "interview__id"          "cereales__id"
##  [4] "s07Bq02_autre_cereales" "s07Bq03a_cereales"      "s07Bq03b_cereales"
##  [7] "s07Bq03c_cereales"      "s07Bq04_cereales"       "s07Bq05_cereales"
## [10] "s07Bq06_cereales"       "s07Bq07a_cereales"      "s07Bq07b_cereales"
## [13] "s07Bq07c_cereales"      "s07Bq08_cereales"
```

#library(tidyverse)

```
library(tidyverse)
```

```
## Warning: le package 'tidyverse' a été compilé avec la version R 4.2.3

## Warning: le package 'ggplot2' a été compilé avec la version R 4.2.3

## Warning: le package 'tibble' a été compilé avec la version R 4.2.3

## Warning: le package 'tidyr' a été compilé avec la version R 4.2.3

## Warning: le package 'readr' a été compilé avec la version R 4.2.3
```

```
## Warning: le package 'purrr' a été compilé avec la version R 4.2.3

## Warning: le package 'stringr' a été compilé avec la version R 4.2.3

## Warning: le package 'forcats' a été compilé avec la version R 4.2.3

## Warning: le package 'lubridate' a été compilé avec la version R 4.2.3

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.4
## v ggplot2   3.4.1     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0
## -- 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
```

typeof(cereal_df$s07Bq03a_cereales) <- double

```
cereal_df <- rename(cereal_df, poids=s07Bq03a_cereales)
glimpse(cereal_df)
```

```
## Rows: 11,114
## Columns: 14
## $ interview__key         <chr> "55-75-97-43", "55-75-97-43", "55-75-97-43", "1~
## $ interview__id          <chr> "3827f11c978e42e6aa4b5b8ed9aa1348", "3827f11c97~
## $ cereales__id           <dbl+lbl>  1,  7, 20,  4, 22,  1,  7, 16, 21,  1, 20,~
## $ s07Bq02_autre_cereales <chr> "", "", "", "", "", "", "", "", "", "", "",~
## $ poids                  <dbl> 9.00, 7.00, 2.00, 7.50, 42.00, 14.00, 10.00, 0.~
## $ s07Bq03b_cereales      <dbl+lbl> 100, 100, 139, 100, 568, 100, 100, 100, 568~
## $ s07Bq03c_cereales      <dbl+lbl> 0, 0, 2, 0, 3, 0, 0, 0, 3, 0, 2, 3, 0, 2, 0~
## $ s07Bq04_cereales       <dbl> 0, 7, NA, NA, 0, 0, 10, 0, 0, 0, NA, 0, 0, 7, 0~
## $ s07Bq05_cereales       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ s07Bq06_cereales       <dbl+lbl> 2, 4, 2, 1, 1, 2, 4, 1, 1, 1, 2, 1, 4, 4, 2~
## $ s07Bq07a_cereales      <dbl> 1.00, NA, 2.00, 2.50, 6.00, 2.00, NA, 0.25, 1.0~
## $ s07Bq07b_cereales      <dbl+lbl> 100,  NA, 139, 100, 568, 100,  NA, 100, 568~
## $ s07Bq07c_cereales      <dbl+lbl>  0, NA,  2,  0,  3,  0, NA,  0,  3,  0,  2,~
## $ s07Bq08_cereales       <dbl> 350, NA, 600, 750, 1050, 325, NA, 200, 200, 120~
```

```
colnames(cereal_df)
```

```
##  [1] "interview__key"         "interview__id"          "cereales__id"
##  [4] "s07Bq02_autre_cereales" "poids"                  "s07Bq03b_cereales"
##  [7] "s07Bq03c_cereales"      "s07Bq04_cereales"       "s07Bq05_cereales"
## [10] "s07Bq06_cereales"       "s07Bq07a_cereales"      "s07Bq07b_cereales"
## [13] "s07Bq07c_cereales"      "s07Bq08_cereales"
```

limite ne prend pas un vecteur ???

Renomer avec select()

- library(dplyr)
- df_cereal <- select(cereal_df,autre_cereal=s07Bq02_autre_cereales)

renommer avec colnames

```
old_name <- colnames(cereal_df)[1:14]
old_name
```

```
##  [1] "interview__key"         "interview__id"          "cereales__id"
##  [4] "s07Bq02_autre_cereales" "poids"                  "s07Bq03b_cereales"
##  [7] "s07Bq03c_cereales"      "s07Bq04_cereales"       "s07Bq05_cereales"
## [10] "s07Bq06_cereales"       "s07Bq07a_cereales"      "s07Bq07b_cereales"
## [13] "s07Bq07c_cereales"      "s07Bq08_cereales"
```

```
new_name <- c(old_name[1],old_name[2],old_name[3],"autre_cereales", "quanite_cons",
              "unites_cons","taille_cons","provenance_auto","provenance_other",
              "freq_achat","quatite_achat",
              "unite_achat","taille_achat","value_lastachat")
isTRUE(length(new_name)==length(old_name)) # véfie si les longeurs sont idtique
```

```
## [1] TRUE
```

check

```
colnames(cereal_df)
```

```
##  [1] "interview__key"         "interview__id"          "cereales__id"
##  [4] "s07Bq02_autre_cereales" "poids"                  "s07Bq03b_cereales"
##  [7] "s07Bq03c_cereales"      "s07Bq04_cereales"       "s07Bq05_cereales"
## [10] "s07Bq06_cereales"       "s07Bq07a_cereales"      "s07Bq07b_cereales"
## [13] "s07Bq07c_cereales"      "s07Bq08_cereales"
```

renomer l'ensemble

```
colnames(cereal_df) <- new_name
colnames(cereal_df)
```

```
##  [1] "interview__key"   "interview__id"    "cereales__id"     "autre_cereales"
##  [5] "quanite_cons"     "unites_cons"      "taille_cons"      "provenance_auto"
##  [9] "provenance_other" "freq_achat"       "quatite_achat"    "unite_achat"
## [13] "taille_achat"     "value_lastachat"
```

renomer l'ensemble

```
colnames(cereal_df) <- new_name
colnames(cereal_df)
```

```
##  [1] "interview__key"   "interview__id"    "cereales__id"     "autre_cereales"
##  [5] "quanite_cons"     "unites_cons"      "taille_cons"      "provenance_auto"
##  [9] "provenance_other" "freq_achat"       "quatite_achat"    "unite_achat"
## [13] "taille_achat"     "value_lastachat"
```

renomer une seule varaible

```
colnames(cereal_df)[3] <- "cereales_id1"
names(cereal_df)
```

```
##  [1] "interview__key"   "interview__id"    "cereales_id1"     "autre_cereales"
##  [5] "quanite_cons"     "unites_cons"      "taille_cons"      "provenance_auto"
##  [9] "provenance_other" "freq_achat"       "quatite_achat"    "unite_achat"
## [13] "taille_achat"     "value_lastachat"
```

labelisation des modalités

avec la bibliotheque lessR

```
library(lessR)
```

```
## Warning: le package 'lessR' a été compilé avec la version R 4.2.3
```

```
##
## lessR 4.2.8                         feedback: gerbing@pdx.edu
## ------------------------------------------------------------
## > d <- Read("")   Read text, Excel, SPSS, SAS, or R data file
##   d is default data frame, data= in analysis routines optional
##
## Learn about reading, writing, and manipulating data, graphics,
## testing means and proportions, regression, factor analysis,
## customization, and descriptive statistics from pivot tables.
##   Enter:  browseVignettes("lessR")
##
## View changes in this and recent versions of lessR.
##   Enter: news(package="lessR")
##
## Interactive data analysis.
##   Enter: interact()

##
## Attachement du package : 'lessR'
```

```
## Les objets suivants sont masqués depuis 'package:dplyr':
##
##     recode, rename
```

```
cereal_df <- label(quanite_cons, "La quantité consommée des 7 derniers jours", data=cereal_df)
```

```
##
## Variable Name: quanite_cons
## Variable Label: La quantité consommée des 7 derniers jours
```

```
cereal_df <- label(cereales_id1, "Le produit consommé", data=cereal_df)
```

```
##
## Variable Name: cereales_id1
## Variable Label: Le produit consommé
```

```
cereal_df <- label(autre_cereales, "Le produit consommé, autre à préciser", data=cereal_df)
```

```
##
## Variable Name: autre_cereales
## Variable Label: Le produit consommé, autre à préciser
```

```
cereal_df <- label(unites_cons, "l'unité de la quantité consommée", data=cereal_df)
```

```
##
## Variable Name: unites_cons
## Variable Label: l'unité de la quantité consommée
```

```
cereal_df <- label(taille_cons, "la taille de l'unité de la quantité consommée", data=cereal_df)
```

```
##
## Variable Name: taille_cons
## Variable Label: la taille de l'unité de la quantité consommée
```

```
cereal_df <- label(provenance_auto, "La provenance de la consommation (autoconsommation)", data=cereal_
```

```
##
## Variable Name: provenance_auto
## Variable Label: La provenance de la consommation (autoconsommation)
```

```
cereal_df <- label(provenance_other, "Autre provenance", data=cereal_df)
```

```
##
## Variable Name: provenance_other
## Variable Label: Autre provenance
```

```
cereal_df <- label(freq_achat, "La fréquence d'achat du produit", data=cereal_df)
```

```
##
## Variable Name: freq_achat
## Variable Label: La fréquence d'achat du produit
```

```
cereal_df <- label(quatite_achat, "La quantité acheté", data=cereal_df)
```

```
##
## Variable Name: quatite_achat
## Variable Label: La quantité acheté
```

```
cereal_df <- label(unite_achat, "L'unité de la qqtité acheté", data=cereal_df)
```

```
##
## Variable Name: unite_achat
## Variable Label: L'unité de la qqtité acheté
```

```
cereal_df <- label(taille_achat, "la taille de la de l'unité de la quantité acheté", data=cereal_df)
```

```
##
## Variable Name: taille_achat
## Variable Label: la taille de la de l'unité de la quantité acheté
```

```
cereal_df <- label(value_lastachat, "La valeur de la quantité acheté", data=cereal_df)
```

```
##
## Variable Name: value_lastachat
## Variable Label: La valeur de la quantité acheté
```

verification

##label(quanite_cons, data = cereal_df)

```
db(cereal_df)
```

```
## Data Types
## ----------------------------------------------------------------
## character: Non-numeric data values
## double: Numeric data values with decimal digits
## ----------------------------------------------------------------
##
##             Variable                   Missing  Unique
##                 Name     Type  Values  Values  Values   First and last values
## ------------------------------------------------------------------------------------------
##  1    interview__key character  11114       0    3479   55-75-97-43 ... 19-00-93-55
##  2     interview__id character  11114       0    3479   3827f11c978e42e6aa4b5b8ed9aa1348 ... 277c4a0
##  3      cereales_id1 haven_labelled  11114       0      27   1  7  20 ... 15  16  21
```

```
##  4    autre_cereales character  11114       0      32        ...
##  5      quanite_cons    double  11104      10     141   9  7  2 ... 2  1  35
##  6       unites_cons haven_labelled  11104      10      17   100  100  139 ... 100  139  568
##  7       taille_cons haven_labelled  11104      10       5   0  0  2 ... 0  2  3
##  8   provenance_auto    double   8424    2690      51   0  7  NA ... 0  0  0
##  9  provenance_other    double  11104      10      48   0  0  0 ... 0  0  0
## 10        freq_achat haven_labelled  11104      10       5   2  4  2 ... 1  2  1
## 11     quatite_achat    double   9495    1619      58   1  NA  2 ... 2  1  5
## 12       unite_achat haven_labelled   9495    1619      18   100  NA  139 ... 100  139  568
## 13      taille_achat haven_labelled   9495    1619       4   0  NA  2 ... 0  2  3
## 14   value_lastachat    double   9495    1619     224   350  NA  600 ... 700  200  1075
## ---------------------------------------------------------------------------------------
##
##
## Variable Names   Variable Labels
## ---------------------------------------------------------------
## quanite_cons       La quantité consommée des 7 derniers jours
## cereales_id1       Le produit consommé
## autre_cereales     Le produit consommé, autre à préciser
## unites_cons        l'unité de la quantité consommée
## taille_cons        la taille de l'unité de la quantité consommée
## provenance_auto    La provenance de la consommation (autoconsommation)
## provenance_other   Autre provenance
## freq_achat         La fréquence d'achat du produit
## quatite_achat      La quantité acheté
## unite_achat        L'unité de la qqtité acheté
## taille_achat       la taille de la de l'unité de la quantité acheté
## value_lastachat    La valueur de la quantité acheté
## ---------------------------------------------------------------
```

Recoder les modalités

```
typeof(cereal_df$cereales_id1) # double donc numeric
```

```
## [1] "double"
```

```
summary(cereal_df$cereales_id1) #
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    3.00    7.00   12.13   21.00  169.00
```

```
table(cereal_df$s07Bq03c_cereales)
```

```
## < table of extent 0 >
```

```
edit (cereal_df$s07Bq03c_cereales)
```

```
## NULL
```