Cours Manipulation de données

Mame Thierno Ndiaye

2023-04-15

Manipulation des données

Importation

```
datapath <- "C:/Users/user/Desktop/mtn/ENSAE/ISE/ISEP2/SEMESTRE 2/Informatique/R_2023/Bases"
library(haven)
cereal <- read_dta(file = paste0(datapath,"\\cereales.dta"))
```

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

Comprendre la structure des données

```
dim(cereal) # dim()` always displays the number of rows first,
```

followed by the number of columns.

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

fourni un résumé utile et compact de sa structure interne.

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

```
library(dplyr)
glimpse(cereal)
```

une autre faÃ§on de voir globalement la structuration

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

Voir les données

```
head(cereal, n=15) # affiche les 15 premiÃ¨res lignes;
tail(cereal, n=10) # affiche les 10 derniÃ¨res lignes
View(cereal) # affiche la base (en quelques lignes)
```

Convertir en data frame

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

charger la table de conversion

```
tableconversion <- "C:/Users/user/Desktop/mtn/ENSAE/ISE/ISEP2/SEMESTRE 2/Informatique/R_2023/Bases"
library(readxl)
Sys.setenv(TZ='GMT') # set time zone
base_table <- read_excel(paste0(tableconversion,"\\Table de conversion phase 2.xlsx"))
str(base_table)
base_table <- data.frame(base_table)
```

Renomer les variables

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

#library(tidyverse) ## typeof(cereal_df$s07Bq03a_cereales) <- double

```
cereal_df <- rename(cereal_df, poids=s07Bq03a_cereales)
glimpse(cereal_df)
colnames(cereal_df) # affiche le nom des variables
```

limite ne prend pas un vecteur ???

Renomer avec select()

```
df_cereal <- select(cereal_df,autre_cereal=s07Bq02_autre_cereales)
```

renommer avec colnames

```
old_name <- colnames(cereal_df)[1:14]
old_name
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)) # vÃ©fie si les longeurs sont idtique
```

check

```
colnames(cereal_df)
```

renomer l'ensemble

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

renomer une seule varaible

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

labelisation des modalités

avec la bibliotheque lessR

```
library(lessR)
cereal_df <- label(quanite_cons, "La quantitÃ© consommÃ©e des 7 derniers jours", data=cereal_df)
cereal_df <- label(cereales_id1, "Le produit consommÃ©", data=cereal_df)
cereal_df <- label(autre_cereales, "Le produit consommÃ©, autre Ã prÃ©ciser", data=cereal_df)
cereal_df <- label(unites_cons, "l'unitÃ© de la quantitÃ© consommÃ©e", data=cereal_df)
cereal_df <- label(taille_cons, "la taille de l'unitÃ© de la quantitÃ© consommÃ©e", data=cereal_df)
cereal_df <- label(provenance_auto, "La provenance de la consommation (autoconsommation)", data=cereal_
cereal_df <- label(provenance_other, "Autre provenance", data=cereal_df)
cereal_df <- label(freq_achat, "La frÃ©quence d'achat du produit", data=cereal_df)
cereal_df <- label(quatite_achat, "La quantitÃ© achetÃ©", data=cereal_df)
cereal_df <- label(unite_achat, "L'unitÃ© de la qqtitÃ© achetÃ©", data=cereal_df)
cereal_df <- label(taille_achat, "la taille de la de l'unitÃ© de la quantitÃ© achetÃ©", data=cereal_df)
cereal_df <- label(value_lastachat, "La valeur de la quantitÃ© achetÃ©", data=cereal_df)
```

verification

#“label(quanite_cons, data = cereal_df)

```
db(cereal_df)
```

Recoder les modalités

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

table(cereal_df$s07Bq03c_cereales)
#edit (cereal_df$s07Bq03c_cereales)
```

labels = c(“Taille unique” = 0, Petit = 1,

Moyen = 2, Grand = 3, Quart = 4, Demi = 5, Entier = 6, “Très Petite” = 7

```
#cereal_df$s07Bq03c_cereales[0]=10
```

Arrêt temporaire ;