

Le projet R forces

Meziane Cherif

Olivier Cailloux

Contexte

Commission d'enquête de l'AN la situation, les missions et les moyens des forces de sécurité 22 mai au 28 juin 2019 agents de la police nationale et des polices municipales, aux militaires de la gendarmerie nationale et aux réservistes

“les éléments des réponses permettant d'identifier le répondant (coordonnées ou risque de ré-identification) ainsi que les réponses comprenant des imputations personnelles ou des insultes ont été supprimés.”

<https://data.assemblee-nationale.fr/autres/consultations-citoyennes/moyens-des-forces-de-securite>

Nous utilisons également les « Principaux indicateurs sur les revenus et la pauvreté aux niveaux national et local en 2019 » du dispositif Fichier localisé social et fiscal (Filosofi) publié par l'INSEE (présentation ici, téléchargement ici). Des données plus récentes existent mais nous utilisons les données reflétant la réalité au moment des réponses des forces de sécurité.

Mise en place

Chargeons quelques packages utiles.

```
library(conflicted)
conflicts_prefer(dplyr::filter)
library(tidyverse)
```

Téléchargeons les réponses des forces de sécurité, ou vérifions leur conformité si elles sont déjà présentes à l'aide du hash MD5 indiqué sur le site sus-mentionné.

```
answers_url <- paste0(
  "https://data.assemblee-nationale.fr/",
  "static/openData/repository/CONSULTATIONS_CITOYENNES/",
  "MOYENS_DES_FORCES_DE_SECURITE/Moyens-des-forces-de-securite.csv"
)
md5_expected <- "261b4244cc2e9ffcd54ff9a6bec0a0ac"
if (file.exists("Réponses original.csv")) {
  md5_observed <- tools::md5sum("Réponses original.csv")
} else {
  md5_observed <- 0L
}
if (md5_observed != md5_expected) {
  downloaded_return <- download.file(answers_url, "Réponses original.csv", mode = "wb")
  stopifnot(identical(downloaded_return, 0L))
}
md5_observed <- tools::md5sum("Réponses original.csv")
stopifnot(md5_observed == md5_expected)
```

Convertissons en UTF8.

```
input_original <- readLines("Réponses original.csv")
input_converted <- iconv(input_original, from = "WINDOWS-1252", to = "UTF8")
writeLines(input_converted, "Réponses.csv")
```

Téléchargeons de même les données sur les revenus et la pauvreté.

```
zip_file_name <- "base-cc-filosofi-2019_CSV.zip"
filosofi_url <- paste0("https://www.insee.fr/fr/statistiques/fichier/6036902/", zip_file_name)
if (!file.exists(zip_file_name)) {
  downloaded_return <- download.file(filosofi_url, zip_file_name, mode = "wb")
  stopifnot(identical(downloaded_return, 0L))
}
```

Ouvrons le zip et extrayons les fichiers cc_filosofi_2019_DEP.csv et meta_cc_filosofi_2019_DEP.csv.

```
to_extract <- c("cc_filosofi_2019_DEP.csv", "meta_cc_filosofi_2019_DEP.csv")
if (!all(file.exists(to_extract))) {
  unzip(zip_file_name, files = to_extract)
}
```

Tentons maintenant de télécharger le zip en mémoire et d'en extraire les fichiers.

```
# insee_response <- httr::GET(filosofi_url)
# stopifnot(insee_response$status_code == 200)
# zip_content <- httr::content(insee_response, as = "raw")
# zip_content <- unzip::unzip(zip_content, list = TRUE)
# to_extract <- c("cc_filosofi_2019_DEP.csv", "meta_cc_filosofi_2019_DEP.csv")
# stopifnot(all(to_extract %in% zip_content$Name))
# zip_content <- unzip::unzip(zip_content, files = to_extract)
```

Il nous faut maintenant écrire ces deux fichiers sur le disque.

```
# for (file_name in to_extract) {
#   file_content <- unzip::unzip(zip_content, files = file_name)
#   writeBin(file_content, file_name)
# }
```

Lecture des données

Réponses

Lisons les réponses des forces de sécurité.

```
answers <- read_delim("Réponses.csv",
  delim = ";", locale = locale(decimal_mark = ","),
  show_col_types = FALSE, name_repair = "minimal"
)
col_renaming <- read_csv("Colonnes.csv", show_col_types = FALSE)
stopifnot(all.equal(colnames(answers), col_renaming[["Nom original"]]))
colnames(answers) <- col_renaming[["Nouveau nom"]]
answers
```

```
## # A tibble: 13,735 x 73
##   rep d_start d_end you   you_other fct   fct_other belong belong_other dept
##   <dbl> <chr>   <chr> <chr> <chr>   <chr> <chr>   <chr> <chr>   <chr>
## 1     3 22/05/~ 22/0~ Fonc~ <NA>   Dans~ <NA>   Aux c~ <NA>   08 --
## 2     4 22/05/~ 22/0~ <NA> <NA>   <NA> <NA>   <NA> <NA>   <NA>
```

```
## 3      5 23/05/~ 23/0~ Fonc~ <NA>      Dans~ <NA>      Aux c~ <NA>      95 --
## 4      6 24/05/~ 24/0~ Fonc~ <NA>      Dans~ <NA>      La ge~ <NA>      11 --
## 5      7 24/05/~ 24/0~ Mili~ <NA>      Dans~ <NA>      La ge~ <NA>      74 --
## 6      8 27/05/~ 27/0~ <NA> <NA>      <NA> <NA>      <NA> <NA>      <NA>
## 7      9 28/05/~ 28/0~ Fonc~ <NA>      Dans~ <NA>      Autre DSPAP      75 --
## 8     10 28/05/~ 28/0~ <NA> <NA>      <NA> <NA>      <NA> <NA>      <NA>
## 9     11 29/05/~ 29/0~ Mili~ <NA>      Dans~ <NA>      La ge~ <NA>      17 --
## 10    12 29/05/~ 29/0~ Mili~ <NA>      Dans~ <NA>      La ge~ <NA>      58 --
## # i 13,725 more rows
## # i 63 more variables: `in` <chr>, task_1 <chr>, task_2 <chr>, impr <chr>,
## #   penal <chr>, penal_1 <chr>, penal_2 <chr>, penal_3 <chr>, agemaj <chr>,
## #   agemaj_other <chr>, hurt <chr>, hurt_then <chr>, prot <chr>,
## #   prot_adeq <chr>, prot_ext <chr>, train <chr>, train_ext <chr>,
## #   train_suff <chr>, train_days_2016 <dbl>, train_days_2017 <dbl>,
## #   train_days_2018 <dbl>, hab_1 <chr>, hab_2 <chr>, hab_3 <chr>, ...
```

Vérifions que les décimales sont lues correctement et que nous disposons du nombre de contributions annoncé sur le site ministériel.

```
stopifnot(answers |> filter(rep == 9) |> pull(train_days_2017) == 2.5)
stopifnot(nrow(answers) == 13735)
```