

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))
}

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)
}
```

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: works <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)
```

Revenus et pauvreté

Lisons maintenant les données économiques.

```
revenues_poverty <- read_delim("cc_filosofi_2019_DEP.csv",
  delim = ";", locale = locale(decimal_mark = ","),
  show_col_types = FALSE
)
revenues_poverty
```

```
## # A tibble: 101 x 28
##   CODGEO NBMENFISC19 NBPERSMENFISC19 MED19 PIMP19 TP6019 TP60AGE119 TP60AGE219
##   <chr>      <dbl>          <dbl> <dbl> <dbl> <dbl>      <dbl>      <dbl>
## 1 01          264074          629120 23490  59.6  10.7      15.4      12.4
## 2 02          223635          513278 19880  49.5  18.4       30      23.4
## 3 03          158967          326379 20570  49    15.4      24.9      19.8
## 4 04           74092          154195 20690  51.5  16.6      26.1      21.7
## 5 05           64688          134672 21020  54.2  13.9      21.6       17
## 6 06          527841         1109491 22300  60.4  15.8      22.2      18.1
## 7 07          143925          313991 21010  50.9  14.3      21.5      18.1
## 8 08          117854          260778 19840  47.2  18.6      30.6      22.8
## 9 09           70184          146066 20010  46.7  17.9      27.7      23.2
## 10 10          132913          290472 20580  52.6  16.3      27.8      21.5
## # i 91 more rows
## # i 20 more variables: TP60AGE319 <dbl>, TP60AGE419 <dbl>, TP60AGE519 <dbl>,
## #   TP60AGE619 <dbl>, TP60TOL119 <dbl>, TP60TOL219 <dbl>, PACT19 <dbl>,
## #   PTSA19 <dbl>, PCH019 <dbl>, PBEN19 <dbl>, PPEN19 <dbl>, PPAT19 <dbl>,
## #   PPSOC19 <dbl>, PPFAM19 <dbl>, PPMINI19 <dbl>, PPLOGT19 <dbl>,
## #   PIMPOT19 <dbl>, D119 <dbl>, D919 <dbl>, RD19 <dbl>
```

Vérifions que le revenu médian et le taux de pauvreté de l'Ain sont ceux indiqués sur le site.

```
ain <- revenues_poverty |> filter(CODGEO == "01")
stopifnot(ain |> pull(MED19) == 23490)
stopifnot(ain |> pull(TP6019) == 10.7)
```

Croisement des données

Extrayons le premier mot de la colonne `dept` pour obtenir le code de département (on vérifie avec une réponse donnée que la conversion a fonctionné). Notons que les départements corses ne s'encodent pas comme des nombres, donc ce code doit être de type chaîne de caractères.

```
stopifnot(answers |> filter(rep == 3) |> pull(dept) == "08 - ARDENNES")
answers <- mutate(answers, dept_nb = str_extract(dept, "[0-9AB]+"), .after = dept)
stopifnot(answers |> filter(rep == 3) |> pull(dept_nb) == "08")
```

Nous pouvons maintenant joindre les données économiques aux réponses des forces de sécurité.

```
data <- left_join(answers, revenues_poverty, by = c("dept_nb" = "CODGEO"))
data
```

```
## # A tibble: 13,735 x 101
```

```
##      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 91 more variables: dept_nb <chr>, works <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>, ...
```