Traitement des données sociodémographiques – EHCVM Sénégal 2018

2025-05-21

Table des matières

0.1 Importation et préparation des données 1
0.2 Taille des ménages 1
0.3 Graphique : Distribution de la taille des ménages 2
0.4 Répartition par sexe 2
0.5 Structure par âge 3
0.6 Pyramide des âges 4
0.7 Ratio de masculinité 5
0.8 Statut matrimonial 6
0.9 Ethnies 7
0.10 Religion 8
0.11 Possession de téléphone 9

0.1 Importation et préparation des données

```
data <- read_dta("C:/Users/HP/Desktop/S4/COURS R/EXPOSE/s01_me_SEN2018.dta") %>%
  clean_names() %>%
  mutate(
    id_menage = paste(vague, grappe, menage, sep = "_"),
    id_individu = paste(id_menage, s01q00a, sep = "_")
  )
```

0.2 Taille des ménages

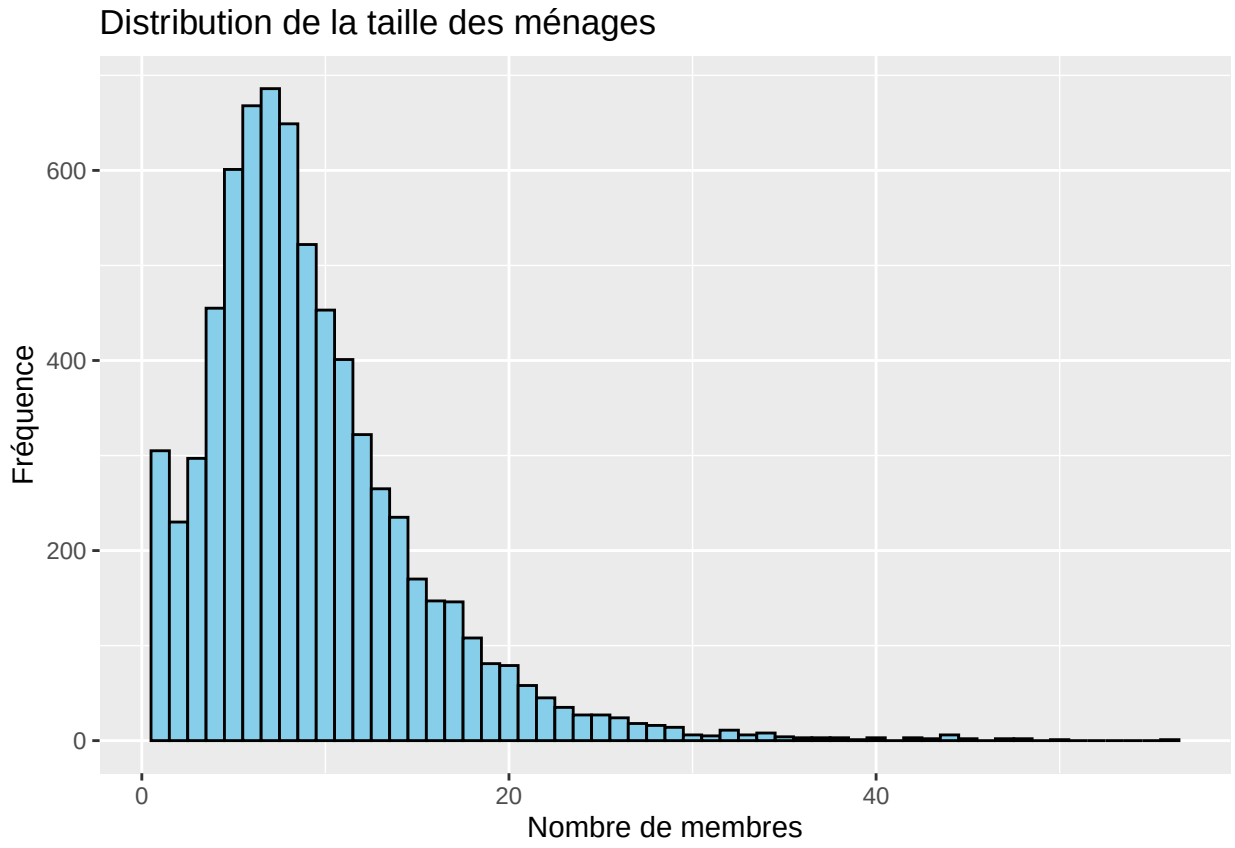
```
taille_menage <- data %>%
  group_by(id_menage) %>%
  summarise(taille = n()) %>%
  ungroup()

taille_menage %>%
  summarise(
    moyenne = mean(taille),
    mediane = median(taille),
    min = min(taille),
```

```
    max = max(taille)
  )
```

```
## # A tibble: 1 x 4
##   moyenne mediane   min   max
##     <dbl>   <dbl> <int> <int>
## 1    9.24       8     1    56
```

0.3 Graphique : Distribution de la taille des ménages

```
ggplot(taille_menage, aes(x = taille)) +
  geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
  labs(title = "Distribution de la taille des ménages", x = "Nombre de membres", y = "Fréquence
```



0.4 Répartition par sexe

```
data %>%
  filter(!is.na(s01q01)) %>%
  mutate(sexe = case_when(
    s01q01 == 1 ~ "Homme",
    s01q01 == 2 ~ "Femme",
    TRUE ~ "Autre"
```

```
  )) %>%
  count(sexe) %>%
  mutate(pourcentage = round(n / sum(n) * 100, 1))
```

```
## # A tibble: 2 x 3
##   sexe      n pourcentage
##   <chr> <int>       <dbl>
## 1 Femme 35316        53.4
## 2 Homme 30802        46.6
```

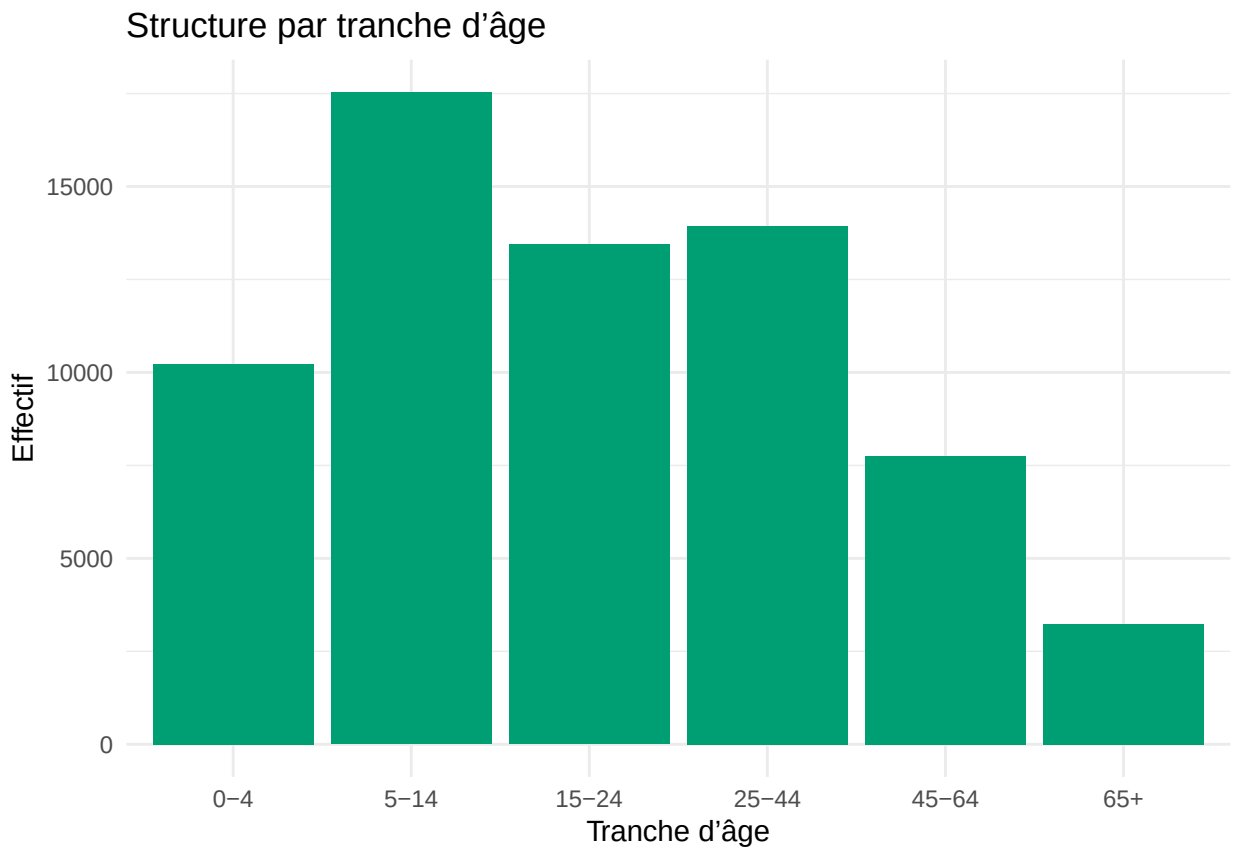
0.5 Structure par âge

```
data <- data %>%
  distinct(id_individu, .keep_all = TRUE) %>%
  mutate(
    age = if_else(vague == 1, 2021 - s01q03c, 2022 - s01q03c),
    tranche_age = case_when(
      age < 5 ~ "0-4",
      age < 15 ~ "5-14",
      age < 25 ~ "15-24",
      age < 45 ~ "25-44",
      age < 65 ~ "45-64",
      TRUE ~ "65+"
    ),
    tranche_age = factor(tranche_age, levels = c("0-4", "5-14", "15-24", "25-44", "45-64", "65+"))
  )

knitr::kable(
  data %>% count(tranche_age) %>% mutate(pourcentage = round(n / sum(n) * 100, 1)),
  caption = "Distribution des individus par tranche d'âge"
)
```

Table 1: Distribution des individus par tranche d'âge

tranche_age	n	pourcentage
0-4	10222	15.5
5-14	17532	26.5
15-24	13440	20.3
25-44	13936	21.1
45-64	7751	11.7
65+	3238	4.9

```
ggplot(data, aes(x = tranche_age)) +
  geom_bar(fill = "#009E73") +
  labs(title = "Structure par tranche d'âge", x = "Tranche d'âge", y = "Effectif") +
  theme_minimal()
```

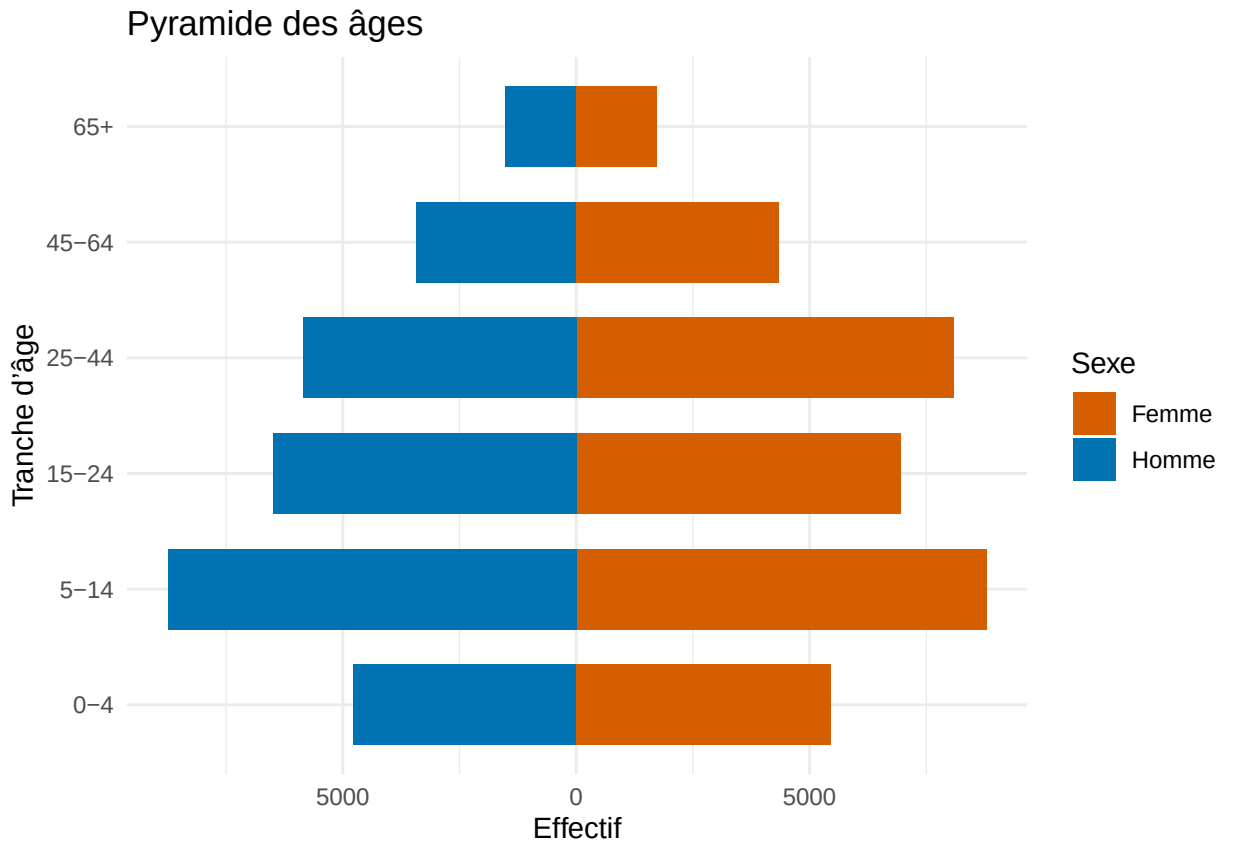


0.6 Pyramide des âges

```
data <- data %>%
  mutate(
    sexe_label = case_when(
      s01q01 == 1 ~ "Homme",
      s01q01 == 2 ~ "Femme",
      TRUE ~ NA_character_
    )
  )

pyramide_data <- data %>%
  filter(!is.na(sexe_label)) %>%
  count(tranche_age, sexe_label) %>%
  mutate(effectif = if_else(sexe_label == "Homme", -n, n))

ggplot(pyramide_data, aes(x = tranche_age, y = effectif, fill = sexe_label)) +
  geom_bar(stat = "identity", width = 0.7) +
  coord_flip() +
  scale_y_continuous(labels = abs) +
  scale_fill_manual(values = c("Homme" = "#0072B2", "Femme" = "#D55E00")) +
  labs(title = "Pyramide des âges", x = "Tranche d’âge", y = "Effectif", fill = "Sexe") +
```

tranche_age	Homme	Femme	sex_ratio
0-4	4768	5454	0.87
5-14	8747	8785	1.00
15-24	6496	6944	0.94
25-44	5853	8083	0.72
45-64	3423	4328	0.79
65+	1515	1722	0.88

```
  theme_minimal()
```

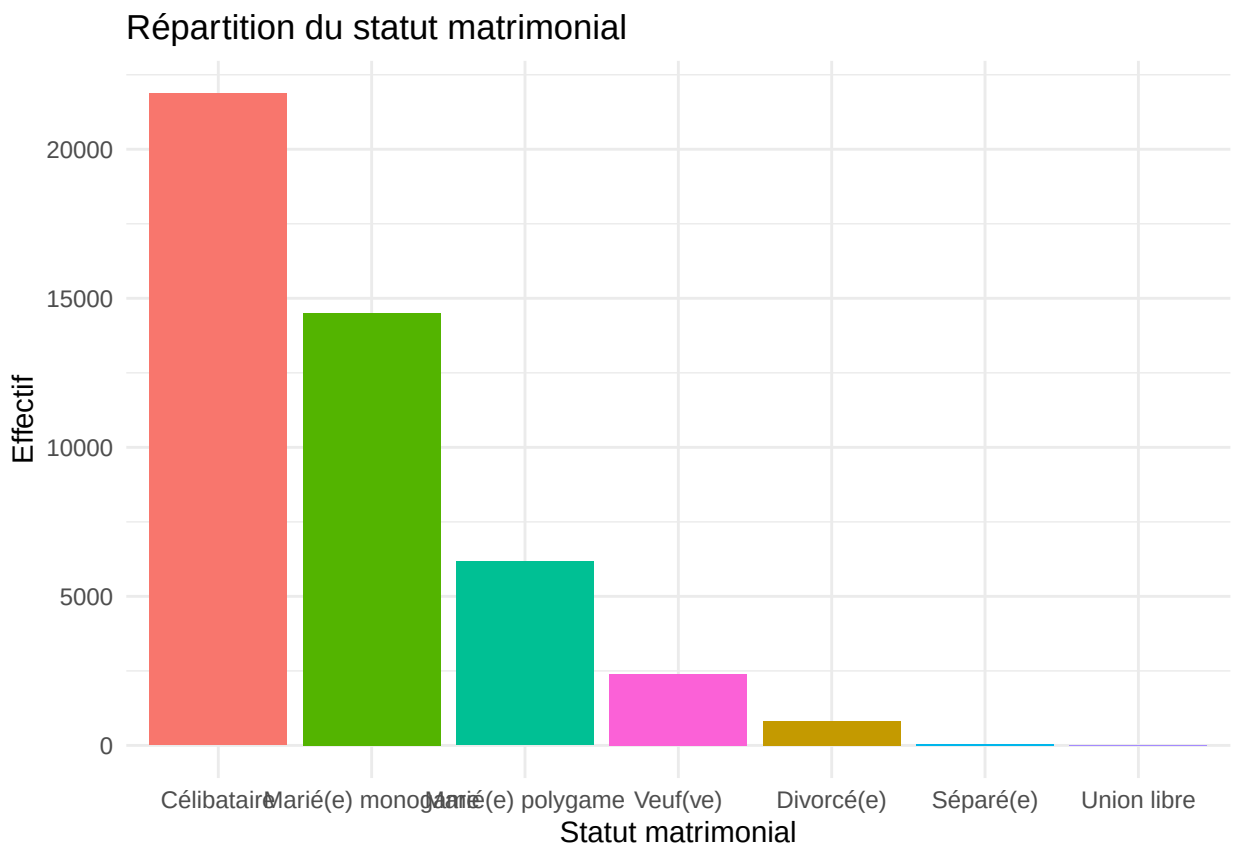


0.7 Ratio de masculinité

```
sexe_age <- data %>%
  filter(s01q01 %in% c(1, 2)) %>%
  count(tranche_age, sexe = s01q01) %>%
  pivot_wider(names_from = sexe, values_from = n, values_fill = 0) %>%
  rename(Homme = `1`, Femme = `2`) %>%
  mutate(sex_ratio = round(Homme / Femme, 2))

sexe_age %>% gt()
```

0.8 Statut matrimonial

```
data %>%
  filter(!is.na(s01q07)) %>%
  count(statut = s01q07) %>%
  mutate(
    libelle = case_when(
      statut == 1 ~ "Célibataire",
      statut == 2 ~ "Marié(e) monogame",
      statut == 3 ~ "Marié(e) polygame",
      statut == 4 ~ "Union libre",
      statut == 5 ~ "Veuf(ve)",
      statut == 6 ~ "Divorcé(e)",
      statut == 7 ~ "Séparé(e)",
      statut == 11 ~ ".A"
    ),
    pourcentage = round(n / sum(n) * 100, 1)
  ) %>%
  ggplot(aes(x = reorder(libelle, -n), y = n, fill = libelle)) +
  geom_bar(stat = "identity") +
  labs(title = "Répartition du statut matrimonial", x = "Statut matrimonial", y = "Effectif") +
  theme_minimal() +
  theme(legend.position = "none")
```

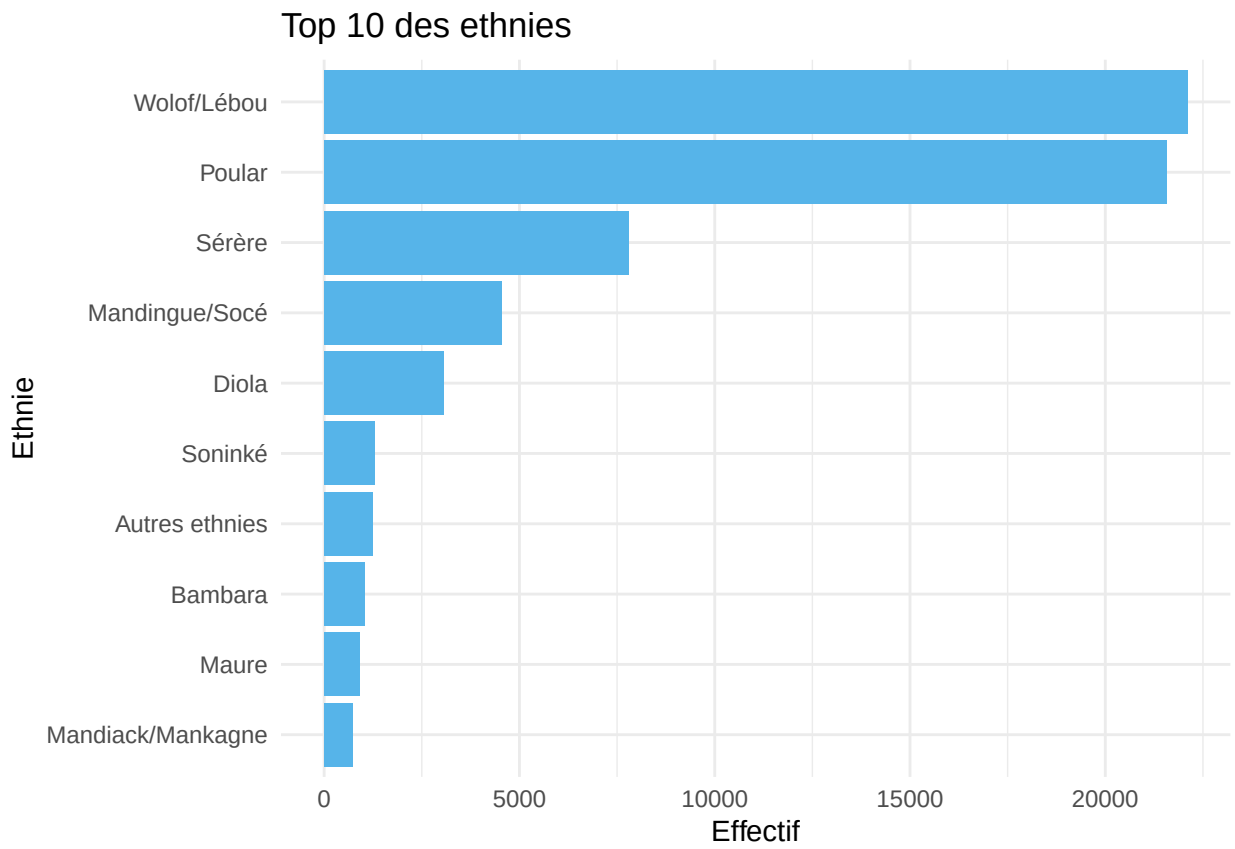


0.9 Ethnies

```
data <- data %>%
  mutate(ethnie_lib = case_when(
    s01q16 == 1 ~ "Wolof/Lébou",
    s01q16 == 2 ~ "Sérère",
    s01q16 == 3 ~ "Poular",
    s01q16 == 4 ~ "Soninké",
    s01q16 == 5 ~ "Diola",
    s01q16 == 6 ~ "Mandingue/Socé",
    s01q16 == 7 ~ "Balante",
    s01q16 == 8 ~ "Bambara",
    s01q16 == 9 ~ "Malinké",
    s01q16 == 10 ~ "Autres ethnies",
    s01q16 == 11 ~ "Naturalisé",
    s01q16 == 12 ~ "Mandiack/Mankagne",
    s01q16 == 13 ~ "Maure",
    s01q16 == 101 ~ ".A"
  ))

ethnie_table <- data %>%
  filter(!is.na(ethnie_lib)) %>%
  count(ethnie = ethnie_lib, sort = TRUE) %>%
  mutate(pourcentage = round(n / sum(n) * 100, 1)) %>%
  head(10)

print(ethnie_table)
```

```
## # A tibble: 10 x 3
##    ethnie                n pourcentage
##    <chr>             <int>       <dbl>
##  1 Wolof/Lébou       22098        33.8
##  2 Poular            21565        32.9
##  3 Sérère             7808        11.9
##  4 Mandingue/Socé     4555         7
##  5 Diola              3059         4.7
##  6 Soninké            1281         2
##  7 Autres ethnies     1253         1.9
##  8 Bambara            1041         1.6
##  9 Maure               902         1.4
## 10 Mandiack/Mankagne   725         1.1
```

```
ggplot(ethnie_table, aes(x = reorder(ethnie, n), y = n)) +
  geom_bar(stat = "identity", fill = "#56B4E9") +
  coord_flip() +
  labs(title = "Top 10 des ethnies", x = "Ethnie", y = "Effectif") +
  theme_minimal()
```



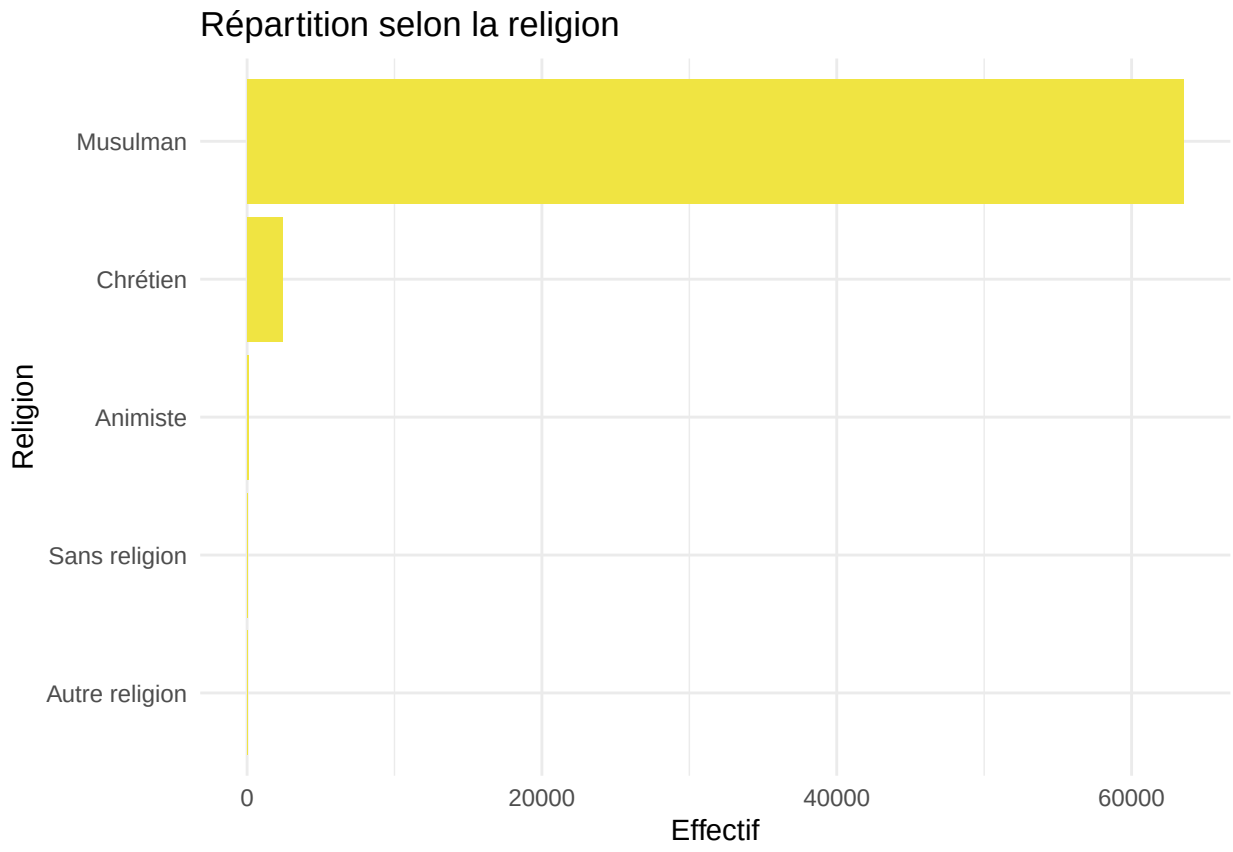
0.10 Religion

```
data <- data %>%
  mutate(religion_lib = case_when(
    s01q14 == 1 ~ "Musulman",
    s01q14 == 2 ~ "Chrétien",
    s01q14 == 3 ~ "Animiste",
    s01q14 == 4 ~ "Autre religion",
    s01q14 == 5 ~ "Sans religion"
  ))

religion_table <- data %>%
  filter(!is.na(religion_lib)) %>%
  count(religion = religion_lib, sort = TRUE) %>%
  mutate(pourcentage = round(n / sum(n) * 100, 1))

print(religion_table)
```

```
## # A tibble: 5 x 3
##   religion           n pourcentage
##   <chr>          <int>       <dbl>
## 1 Musulman       63525        96.1
```

```
## 2 Chrétien        2402         3.6
## 3 Animiste          74         0.1
## 4 Autre religion    54         0.1
## 5 Sans religion     54         0.1
```

```r
ggplot(religion_table, aes(x = reorder(religion, n), y = n)) +
  geom_bar(stat = "identity", fill = "#F0E442") +
  coord_flip() +
  labs(title = "Répartition selon la religion", x = "Religion", y = "Effectif") +
  theme_minimal()
```



0.11 Possession de téléphone

```r
data <- data %>%
  mutate(tel_possede = case_when(
    s01q36 == 1 ~ "Oui",
    s01q36 == 2 ~ "Non",
    TRUE ~ "Non renseigné"
  ))

tel_table <- data %>%
  filter(!is.na(tel_possede)) %>%
  count(tel_possede, sort = TRUE) %>%
```

```
  mutate(pourcentage = round(n / sum(n) * 100, 1))

knitr::kable(tel_table, caption = "Distribution de la possession de téléphone")
```

Table 2: Distribution de la possession de téléphone

tel_possede	n	pourcentage
Oui	25701	38.9
Non renseigné	20308	30.7
Non	20110	30.4

```
ggplot(tel_table, aes(x = reorder(tel_possede, n), y = n)) +
  geom_bar(stat = "identity", fill = "#D55E00") +
  coord_flip() +
  labs(title = "Possession de téléphone", x = "Possède un téléphone", y = "Effectif") +
  theme_minimal()
```