

# Traitement des données sociodémographiques – EHCVM Sénégal 2018

2025-05-21

## Table des matières

0.1	Importation et création des identifiants	1
0.2	Taille des ménages	1
0.3	Graphique : Distribution de la taille des ménages	2
0.4	Répartition par sexe	3
0.5	Structure par âge	4
0.6	Pyramide des âges	6
0.7	Statut matrimonial	8
0.8	Ethnies	9
0.9	Religion	10
0.10	Possession de téléphone	11

### 0.1 Importation et création des identifiants

```
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 = "_")
  )
n_menages <- n_distinct(data$id_menage)
n_individus <- nrow(data)
paste("Nombre total d'individus :", n_individus)

## [1] "Nombre total d'individus : 66119"
paste("Nombre total de ménages :", n_menages)

## [1] "Nombre total de ménages : 7156"
```

### 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

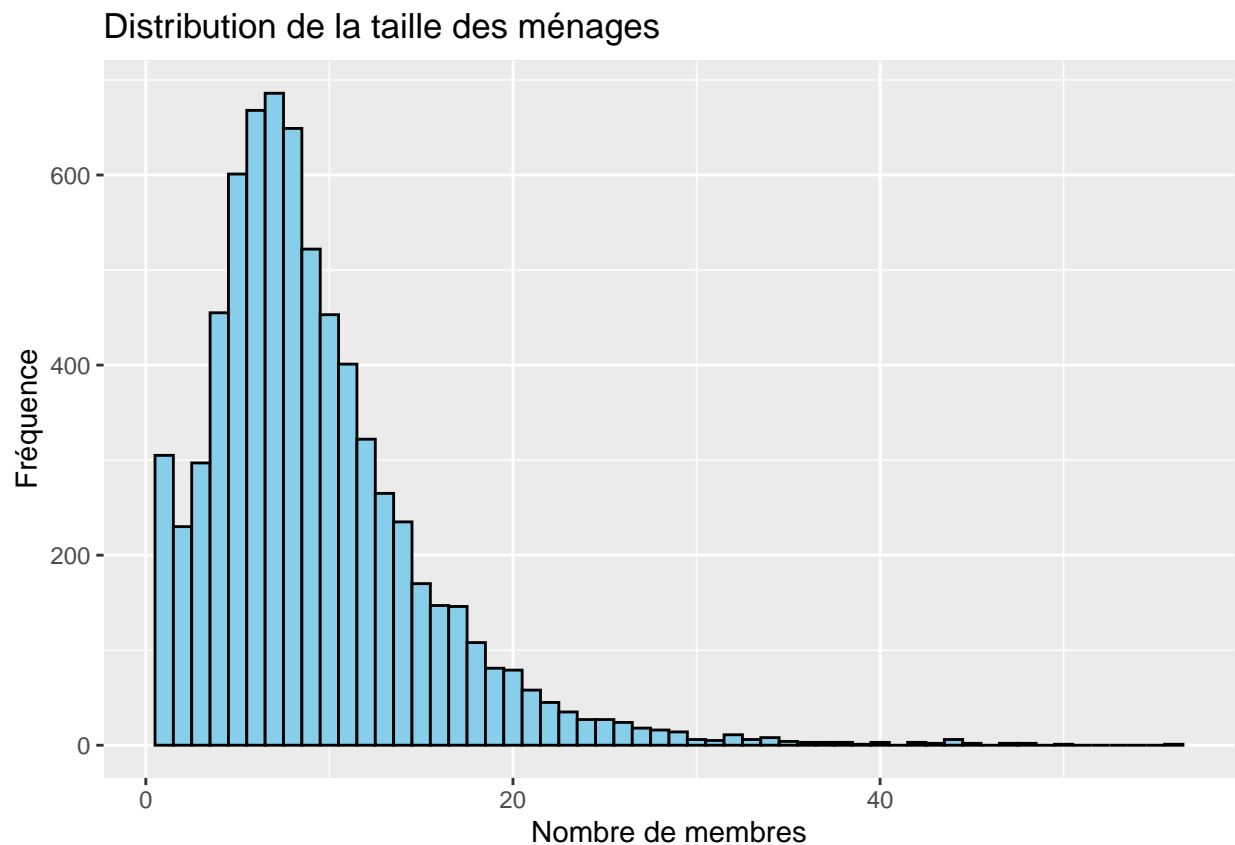
```

### 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")

```



sexe	n	pourcentage
Femme	35316	53.4
Homme	30802	46.6

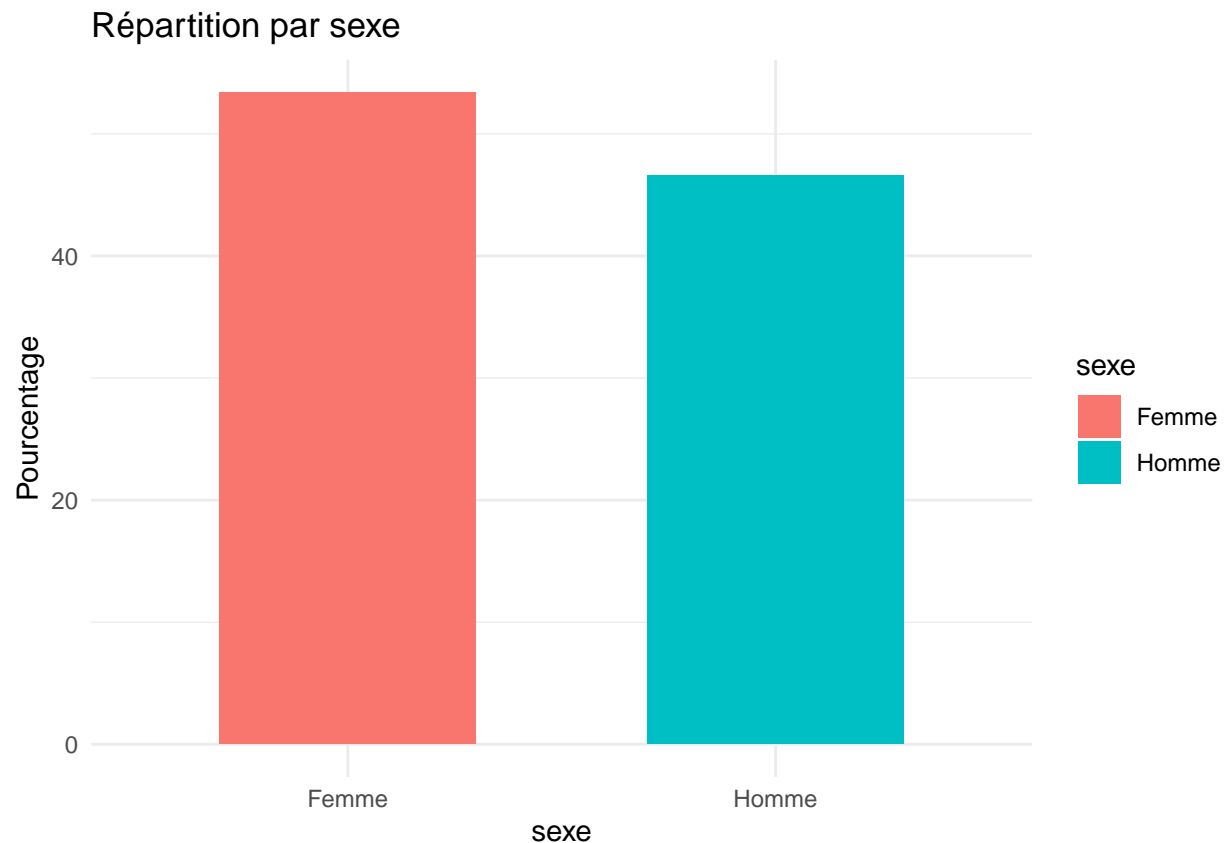
## 0.4 Répartition par sexe

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

sexe_tab <- data %>%
  count(sexe) %>%
  mutate(pourcentage = round(n / sum(n) * 100, 1))

gt(sexe_tab)

ggplot(sexe_tab, aes(x = sexe, y = pourcentage, fill = sexe)) +
  geom_bar(stat = "identity", width = 0.6) +
  labs(title = "Répartition par sexe", y = "Pourcentage") +
  theme_minimal()
```



## 0.5 Structure par âge

```
data <- data %>%
  mutate(
    s01q03c = na_if(s01q03c, 9999)
  ) %>%
  distinct(id_individu, .keep_all = TRUE) %>%
  mutate(
    age = if_else(vague == 1, 2018 - s01q03c, 2019 - 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",
      age >= 65 ~ "65+",
      TRUE ~ NA_character_ # pour les âges NA
    ),
    tranche_age = factor(tranche_age, levels = c("0-4", "5-14", "15-24", "25-44", "45-64", "65+"))
  )

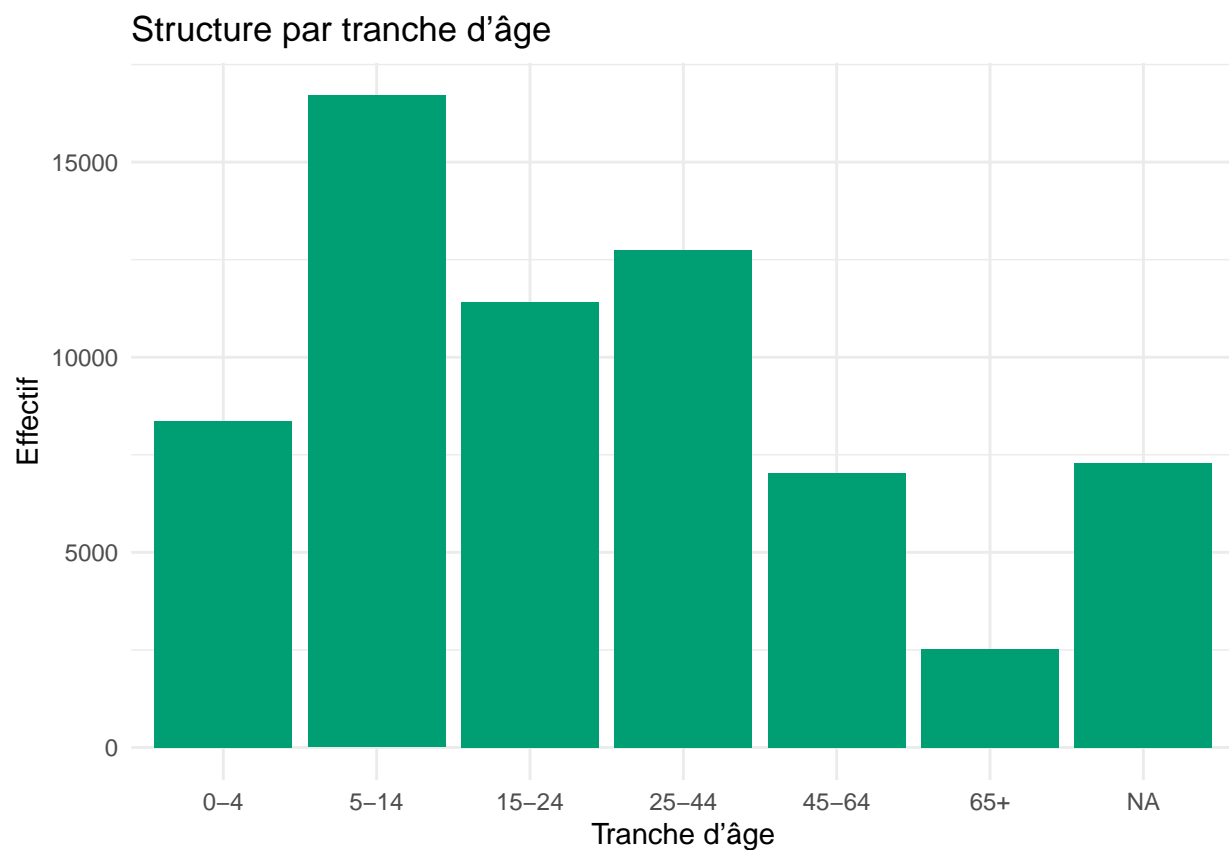
table_tranche <- data %>%
```

tranche_age	n	pourcentage
0-4	8375	12.7
5-14	16707	25.3
15-24	11424	17.3
25-44	12753	19.3
45-64	7037	10.6
65+	2534	3.8
NA	7288	11.0

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

gt(table_tranche)

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()
```



```
structure_menage <- data %>%
  group_by(id_menage) %>%
```

```

summarise(
  enfants = sum(age < 15, na.rm = TRUE),
  vieux = sum(age >= 65, na.rm = TRUE),
  actifs = sum(age >= 15 & age < 65, na.rm = TRUE)
) %>%
mutate(
  dep_ratio = round((enfants + vieux) / ifelse(actifs == 0, NA, actifs), 2)
)

summary(structure_menage$dep_ratio)

```

```

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.0000  0.4000  0.8300  0.9974  1.3300  8.0000   191

```

## 0.6 Pyramide des âges

```

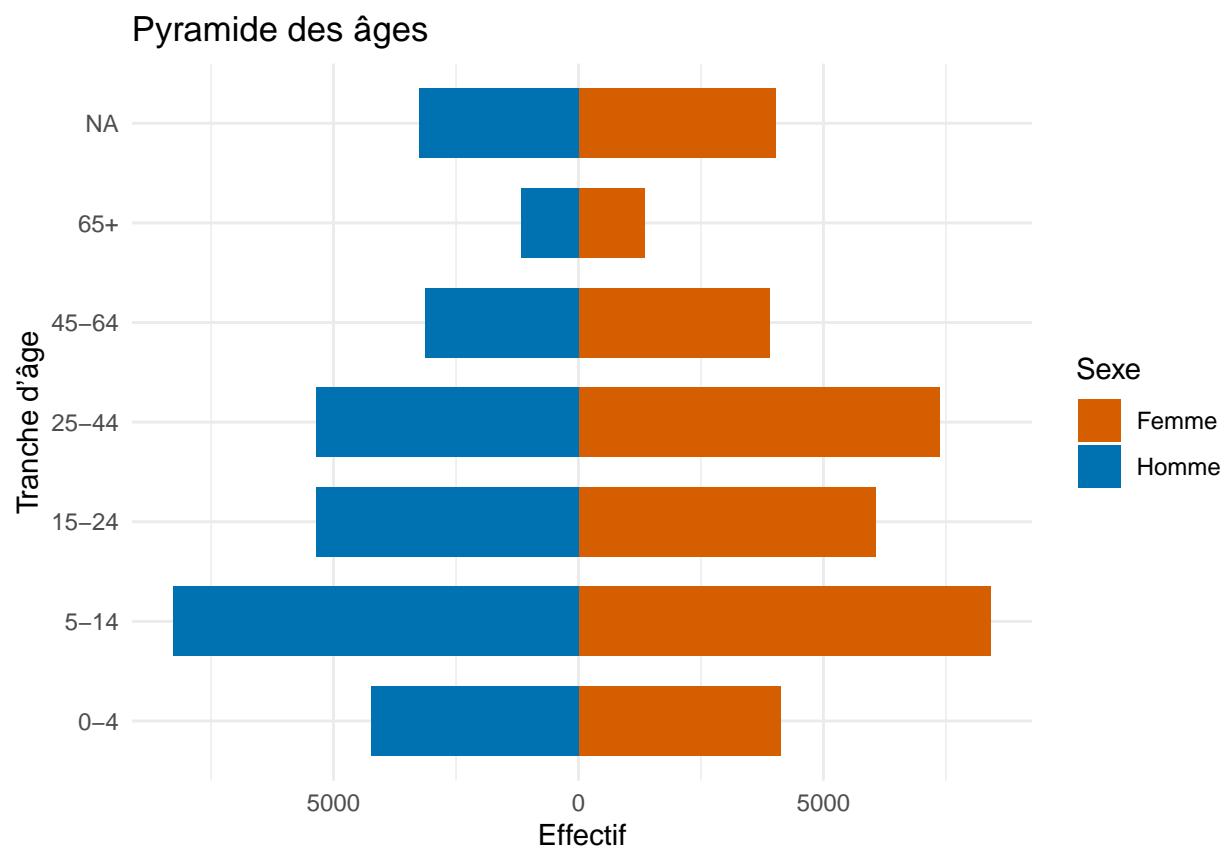
data <- data %>%
  mutate(
    sexe_label = case_when(
      s01q01 == 1 ~ "Homme",
      s01q01 == 2 ~ "Femme",
      TRUE ~ NA_character_
    )
  )

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") +
  theme_minimal()

```

tranche_age	Homme	Femme	sex_ratio
0-4	4232	4143	1.02
5-14	8280	8427	0.98
15-24	5356	6068	0.88
25-44	5368	7385	0.73
45-64	3138	3899	0.80
65+	1172	1362	0.86
NA	3256	4032	0.81



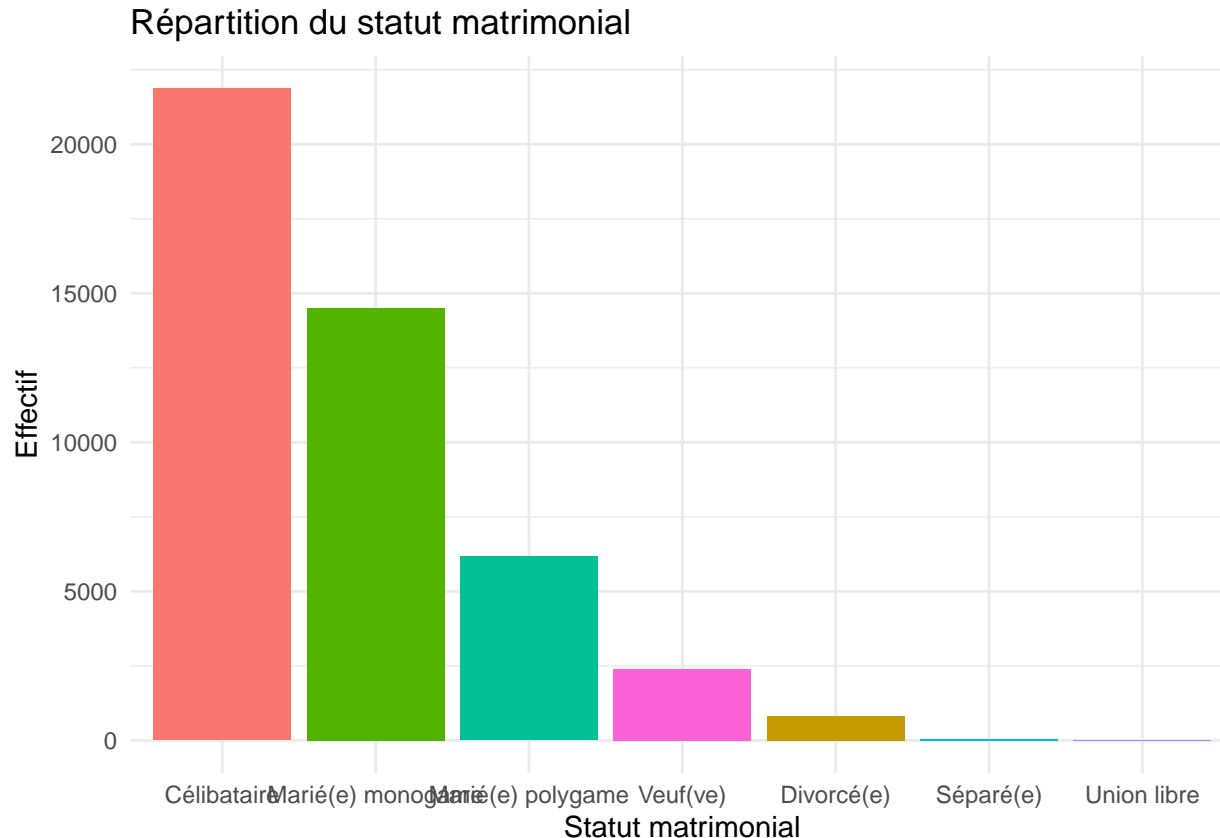
### ## Ratio de masculinité

```
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))

gt(sexe_age)
```

## 0.7 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")
```





ethnie	n	pourcentage
Wolof/Lébou	22098	33.8
Poular	21564	32.9
Sérère	7808	11.9
Mandingue/Socé	4555	7.0
Diola	3059	4.7
Soninké	1281	2.0
Autres ethnies	1253	1.9
Bambara	1041	1.6
Maure	902	1.4
Mandiack/Mankagne	725	1.1

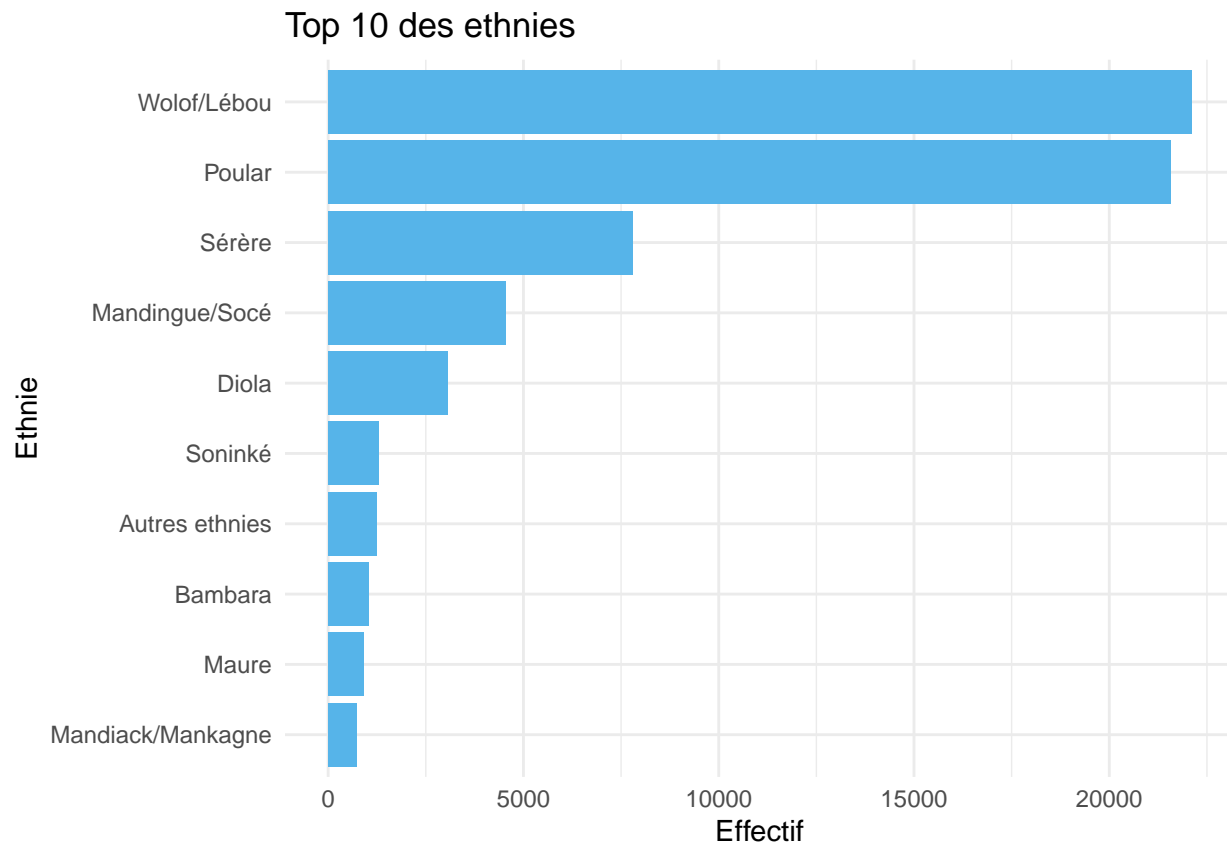
## 0.8 Ethnies

```
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)

gt(ethnie_table)

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()
```



## 0.9 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"
  ))

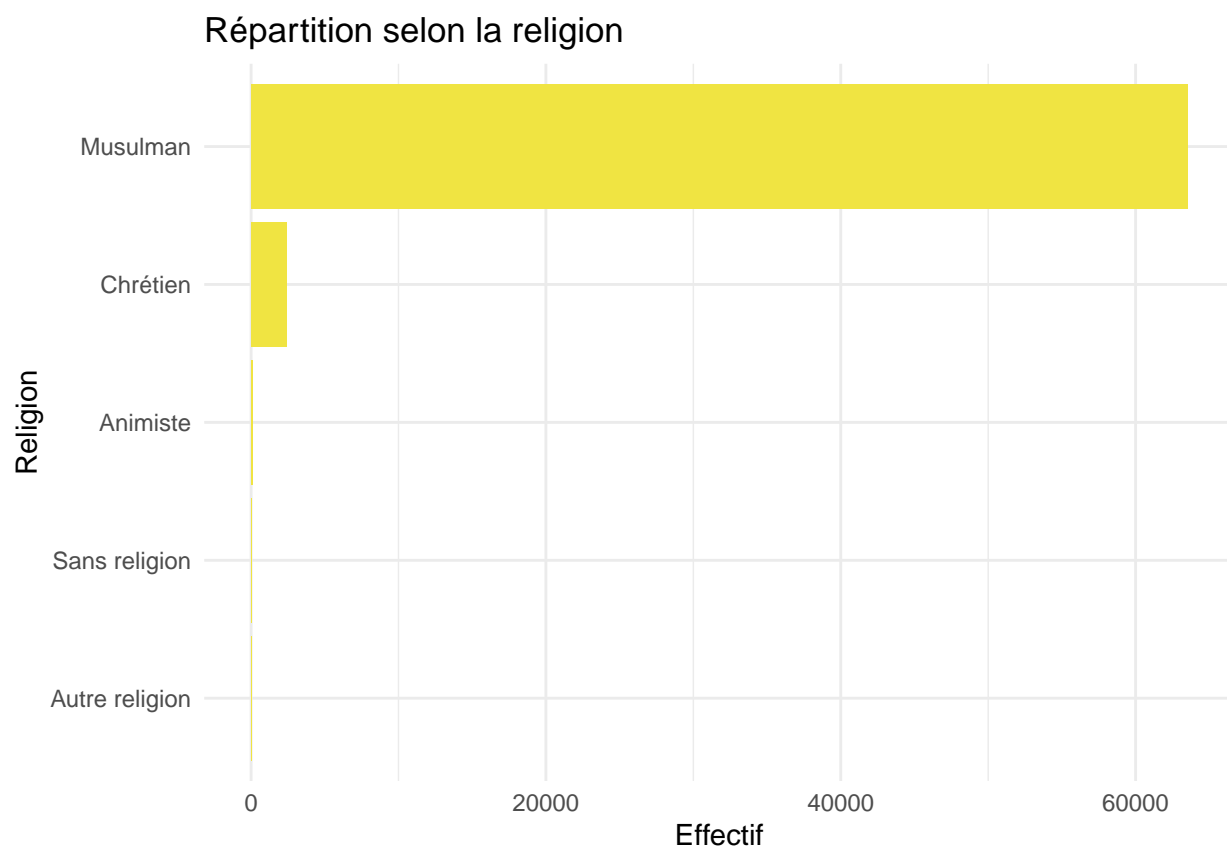
religion_table <- data %>%
  filter(!is.na(religion_lib)) %>%
  count(religion = religion_lib, sort = TRUE) %>%
  mutate(pourcentage = round(n / sum(n) * 100, 1))

gt(religion_table)

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") +
```

religion	n	pourcentage
Musulman	63524	96.1
Chrétien	2402	3.6
Animiste	74	0.1
Autre religion	54	0.1
Sans religion	54	0.1

```
theme_minimal()
```



## 0.10 Possession de téléphone

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

tel_table <- data %>%
  filter(!is.na(tel_possede)) %>%
```

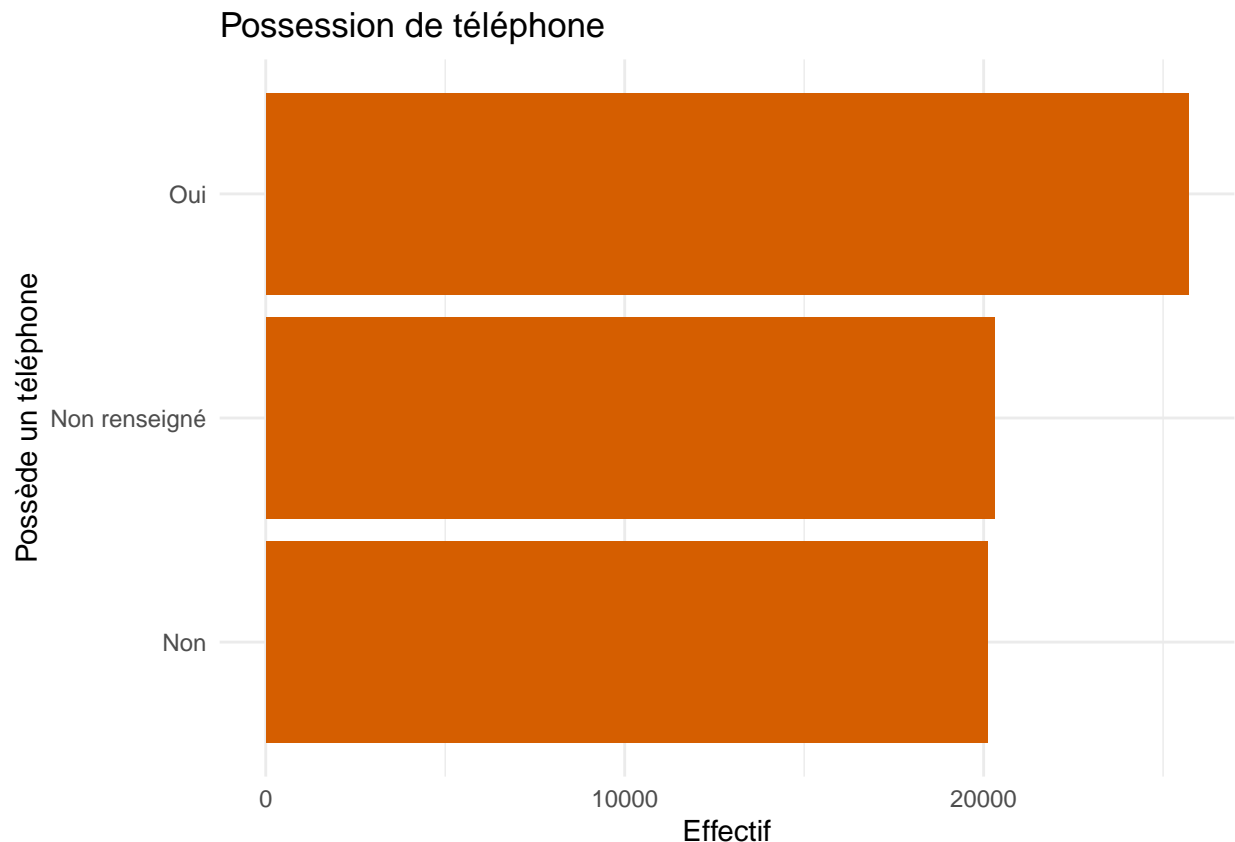
```
count(tel_possede, sort = TRUE) %>%
mutate(pourcentage = round(n / sum(n) * 100, 1))

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

Table 1: Distribution de la possession de téléphone

tel_possede	n	pourcentage
Oui	25701	38.9
Non renseigné	20307	30.7
Non	20110	30.4

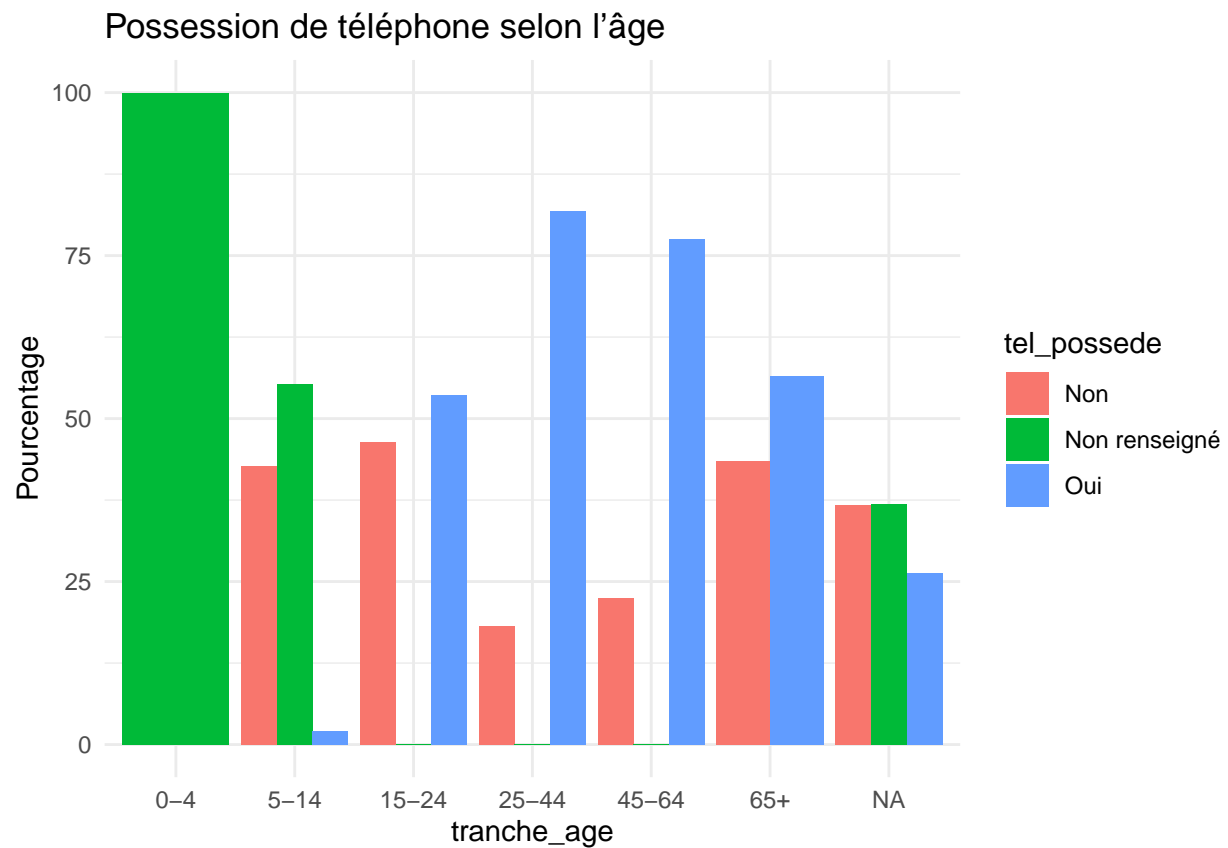
```
ggplot(tel_table, aes(x = reorder(tel_possede, n), y = n)) +
  geom_bar(stat = "identity", fill = "#D55E00") +
  coord_flip() +
  labs(title = "Possession de téléphone", x = "Possède un téléphone", y = "Effectif") +
  theme_minimal()
```



```
tel_par_age <- data %>%
  filter(!is.na(tel_possede)) %>%
  count(tranche_age, tel_possede) %>%
  group_by(tranche_age) %>%
```

```
mutate(pct = round(n / sum(n) * 100, 1))
gt(gt_par_age)
```

```
ggplot(gt_par_age, aes(x = tranche_age, y = pct, fill = tel_possede)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Possession de téléphone selon l'âge", y = "Pourcentage") +
  theme_minimal()
```



tel_possede	n	pct
0-4		
Non renseigné	8375	100.0
5-14		
Non	7134	42.7
Non renseigné	9232	55.3
Oui	341	2.0
15-24		
Non	5297	46.4
Non renseigné	2	0.0
Oui	6125	53.6
25-44		
Non	2316	18.2
Non renseigné	6	0.0
Oui	10431	81.8
45-64		
Non	1578	22.4
Non renseigné	1	0.0
Oui	5458	77.6
65+		
Non	1103	43.5
Oui	1431	56.5
NA		
Non	2682	36.8
Non renseigné	2691	36.9
Oui	1915	26.3