

# Traitement des données - Accès au crédit et à l'épargne

Mouhamet Seck

2025-05-22

## 1. Introduction

Ce document présente le traitement des données issues de l'enquête harmonisée sur les conditions de vie des ménages (EHCVM) au Sénégal (édition 2018), dans le but de préparer une base exploitable pour l'analyse statistique de l'accès au crédit et à l'épargne.

## 2. Importation des données

```
senedata <- "C:/Users/HP/Desktop/ISEP2/Semestre2/Logiciel R/Cours/EHCVM_Données_SENEGAL"

Identification_data <- read_dta(paste0(senedata, "/s00_me_SEN2018.dta"))
socio_data <- read_dta(paste0(senedata, "/s01_me_SEN2018.dta"))
education_data <- read_dta(paste0(senedata, "/s02_me_SEN2018.dta"))
emploi_data <- read_dta(paste0(senedata, "/s04_me_SEN2018.dta"))
revenu_data <- read_dta(paste0(senedata, "/s05_me_SEN2018.dta"))
epargne_credit_data <- read_dta(paste0(senedata, "/s06_me_SEN2018.dta"))
logement_data <- read_dta(paste0(senedata, "/s11_me_SEN2018.dta"))
transfert_data <- read_dta(paste0(senedata, "/s13a_1_me_SEN2018.dta"))
```

## 3. Construction des identifiants et préparation des modules

Chaque module est enrichi avec des identifiants individuels et ménage. Des transformations et sélections de variables pertinentes sont réalisées afin de préparer la fusion des bases.

*Exemple :* - L'identifiant du ménage `id_menage` est obtenu par concaténation des variables `vague`, `grappe` et `menage`. - L'identifiant individuel `id_individu` ajoute la variable `s01q00a` (rang de l'individu dans le ménage).

Des indicateurs synthétiques comme `alphabetisation`, `a_credit`, `a_epargne`, ou `a_revenu_hors_emploi` sont également construits.

```
# Identifiants et sélection des colonnes pertinentes
Identification_data <- Identification_data %>%
  mutate(id_menage = paste(vague, grappe, menage, sep = "_")) %>%
  select(id_menage, Region = s00q01, Milieu_de_residence = s00q04)

socio_data <- socio_data %>%
  mutate(id_menage = paste(vague, grappe, menage, sep = "_"),
         id_individu = paste(vague, grappe, menage, s01q00a, sep = "_")) %>%
```

```

select(id_menage, id_individu, sexe = s01q01, annee_naissance = s01q03c,
       situation_matrimoniale = s01q07, age_premier_mariage = s01q10,
       religion = s01q14, nationalite = s01q15, ethnique = s01q16,
       telephone = s01q36, socio_pro_pere = s01q27)

education_data <- education_data %>%
  mutate(id_menage = paste(vague, grappe, menage, sep = "_"),
         id_individu = paste(vague, grappe, menage, s01q00a, sep = "_")) %>%
  select(id_menage, id_individu, etude_formelle = s02q03, niveau_etudes_max = s02q29,
         diplome_obtenu = s02q33) %>%
  mutate(alphabetisation = ifelse(etude_formelle == 1 & !is.na(niveau_etudes_max), 1, 0))

emploi_data <- emploi_data %>%
  mutate(id_menage = paste(vague, grappe, menage, sep = "_"),
         id_individu = paste(vague, grappe, menage, s01q00a, sep = "_")) %>%
  select(id_menage, id_individu, categorie_socio_pro = s04q39,
         branche_activite = s04q30c, salaire_total = s04q43)

revenu_data <- revenu_data %>%
  mutate(id_menage = paste(vague, grappe, menage, sep = "_"),
         id_individu = paste(vague, grappe, menage, s01q00a, sep = "_")) %>%
  mutate(
    a_recu_pension = if_else(s05q01 == 1 | s05q03 == 1 | s05q05 == 1 | s05q07 == 1, 1, 0),
    a_recu_loyer = if_else(s05q09 == 1, 1, 0),
    a_recu_revenu_financier = if_else(s05q11 == 1, 1, 0),
    a_recu_autre_revenu = if_else(s05q13 == 1, 1, 0)
  ) %>%
  mutate(revenu_hors_emploi_total = rowSums(across(c(s05q02, s05q04, s05q06, s05q08,
                                                    s05q10, s05q12, s05q14), ~replace_na(., 0))),
         a_revenu_hors_emploi = if_else(revenu_hors_emploi_total > 0, 1, 0))

logement_data <- logement_data %>%
  mutate(id_menage = paste(vague, grappe, menage, sep = "_")) %>%
  select(id_menage, statut_occupation = s11q04, titre_propriete = s11q09a,
         credit_immobilier = s11q16, eau_courante = s11q22,
         electricite = s11q34, internet = s11q46) %>%
  mutate(acces_infrastructures = ifelse(rowSums(across(c(eau_courante, electricite, internet))), na.rm =

epargne_credit_data <- epargne_credit_data %>%
  mutate(id_menage = paste(vague, grappe, menage, sep = "_"),
         id_individu = paste(vague, grappe, menage, s01q00a, sep = "_"),
         a_epargne = ifelse((s06q01__1 == 1 | s06q01__2 == 1 | s06q01__3 == 1 | s06q01__4 == 1 | s06q01__5 == 1), 1, 0),
         a_credit = ifelse((s06q03 == 1 & s06q05 == 1) | s06q09 == 1 | s06q10 > 0, 1, 0))

epargne_credit_data$a_epargne[is.na(epargne_credit_data$a_epargne)] <- 0
epargne_credit_data$a_credit[is.na(epargne_credit_data$a_credit)] <- 0

transfert_data <- transfert_data %>%
  mutate(id_menage = paste(vague, grappe, menage, sep = "_")) %>%
  select(id_menage, transfert_famille = s13aq01, transfert_non_membre = s13aq02,
         transfert_mobile_money = s13aq03, a_recu_transfert = s13aq04) %>%
  mutate(type_transfert = case_when(
    transfert_famille == 1 ~ "Famille",

```

```
transfert_non_membre == 1 ~ "Non-membre",
transfert_mobile_money == 1 ~ "Mobile Money",
TRUE ~ "Aucun"))
```

## 4. Fusion des modules

```
logement_data_extended <- logement_data %>% left_join(socio_data[, c("id_menage", "id_individu")], by =
transfert_data_extended <- transfert_data %>% left_join(socio_data[, c("id_menage", "id_individu")], by =
Identification_data_extended <- Identification_data %>% left_join(socio_data[, c("id_menage", "id_individu")], by =
```

```
data_individu <- socio_data %>%
  left_join(education_data, by = c("id_menage", "id_individu")) %>%
  left_join(emploi_data, by = c("id_menage", "id_individu")) %>%
  left_join(epargne_credit_data, by = c("id_menage", "id_individu")) %>%
  left_join(revenu_data, by = c("id_menage", "id_individu")) %>%
  left_join(logement_data_extended, by = c("id_menage", "id_individu")) %>%
  left_join(transfert_data_extended, by = c("id_menage", "id_individu")) %>%
  left_join(Identification_data_extended, by = c("id_menage", "id_individu"))
```

*# Recalcul de l'âge*

```
data_individu$age <- 2018 - as.numeric(data_individu$annee_naissance)
```

## 5. Recodage des variables et regroupements de certaines modalités

Les variables catégorielles sont regroupées afin de simplifier l'interprétation et améliorer la robustesse des analyses :

- tranche\_age en 6 classes : 15-24, 25-34, ..., 65+
- niveau\_etude\_cat : Aucun / Primaire / Secondaire / Supérieur
- statut\_matrimonial : Célibataire / Marié(e) / Autre
- religion\_cat : Musulman / Chrétien / Autre ou sans religion

```
data_individu <- data_individu %>%
  filter(age >= 15) %>%
  mutate(tranche_age = cut(age, breaks = c(15, 24, 34, 44, 54, 64, Inf),
    labels = c("15-24", "25-34", "35-44", "45-54", "55-64", "65+")),
    Region = factor(Region, levels = 1:14,
    labels = c("Dakar", "Ziguinchor", "Diourbel", "Saint-Louis", "Tambacounda",
    "Kaolack", "Thiès", "Louga", "Fatick", "Kolda", "Matam", "Kaffrine",
    "Kédougou", "Sédhiou")),
    Milieu = factor(Milieu_de_residence, levels = c(1, 2), labels = c("Urbain", "Rural")),
    statut_matrimonial = case_when(
    situation_matrimoniale == 1 ~ "Célibataire",
    situation_matrimoniale %in% c(2, 3) ~ "Marié(e)",
    TRUE ~ "Autre"),
    telephone = factor(telephone, levels = c(1, 2), labels = c("Oui", "Non")),
    religion_cat = case_when(
    religion == 1 ~ "Musulman",
    religion == 2 ~ "Chrétien",
    religion %in% c(3, 4, 5) ~ "Autre/Sans",
```

```

    TRUE ~ NA_character_
  ),
  niveau_etude_cat = case_when(
    is.na(niveau_etudes_max) | niveau_etudes_max == 1 ~ "Aucun",
    niveau_etudes_max == 2 ~ "Primaire",
    niveau_etudes_max %in% 3:6 ~ "Secondaire",
    niveau_etudes_max %in% 7:8 ~ "Supérieur",
    TRUE ~ NA_character_
  ))

```

## 6. Sélection finale et gestion des valeurs manquantes

Nous conservons uniquement les variables utiles à la modélisation. Les observations contenant des valeurs manquantes sur les variables retenues sont supprimées.

```

vars_model <- c("a_credit", "a_epargne", "tranche_age", "sexe", "niveau_etude_cat",
               "statut_matrimonial", "telephone", "religion_cat", "Milieu", "Region",
               "a_recu_transfert", "a_revenu_hors_emploi")

data_model <- data_individu %>%
  select(all_of(vars_model)) %>%
  drop_na() %>%
  mutate(across(c(tranche_age, niveau_etude_cat, statut_matrimonial, religion_cat, Milieu, Region), as.factor),
         sexe = factor(sexe, levels = c(1, 2), labels = c("Homme", "Femme")))

```

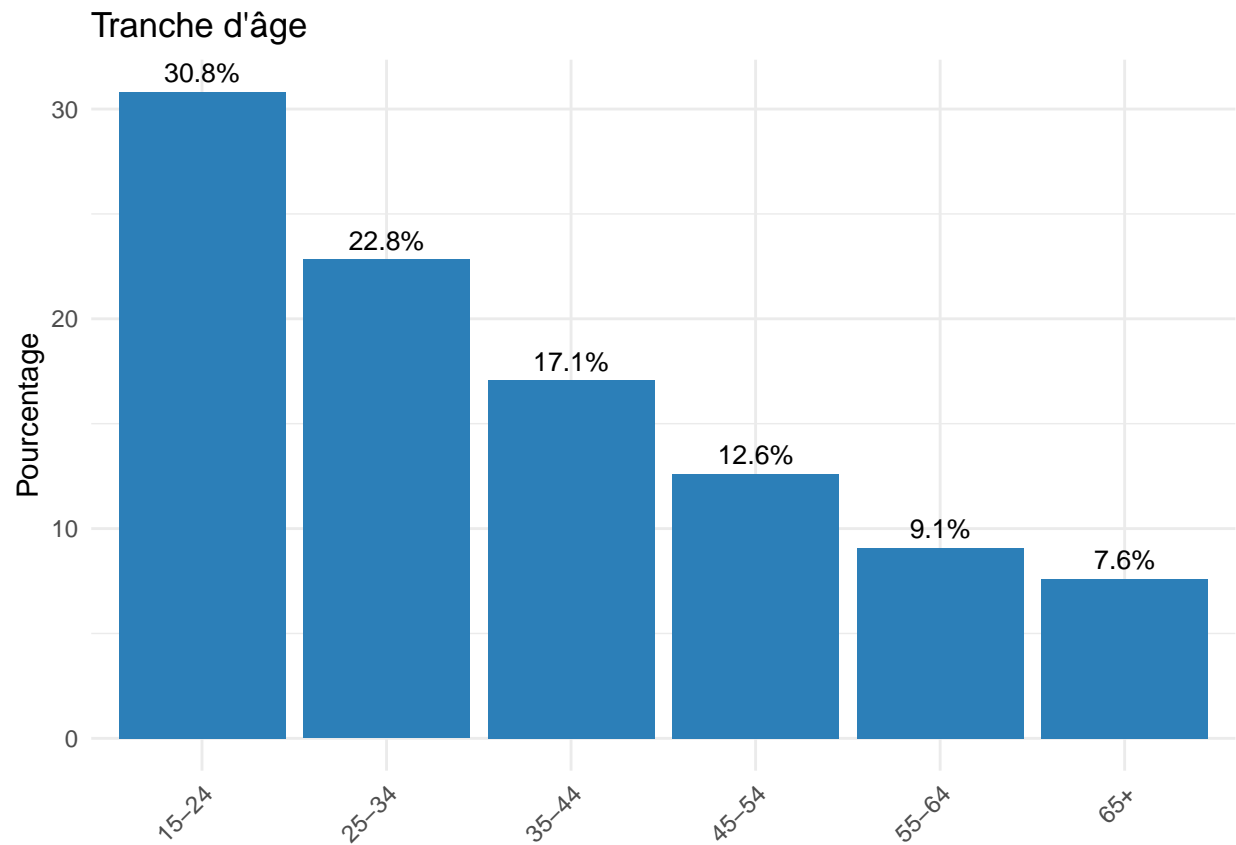
### 6.1 Visualisation des distributions (en pourcentage)

```

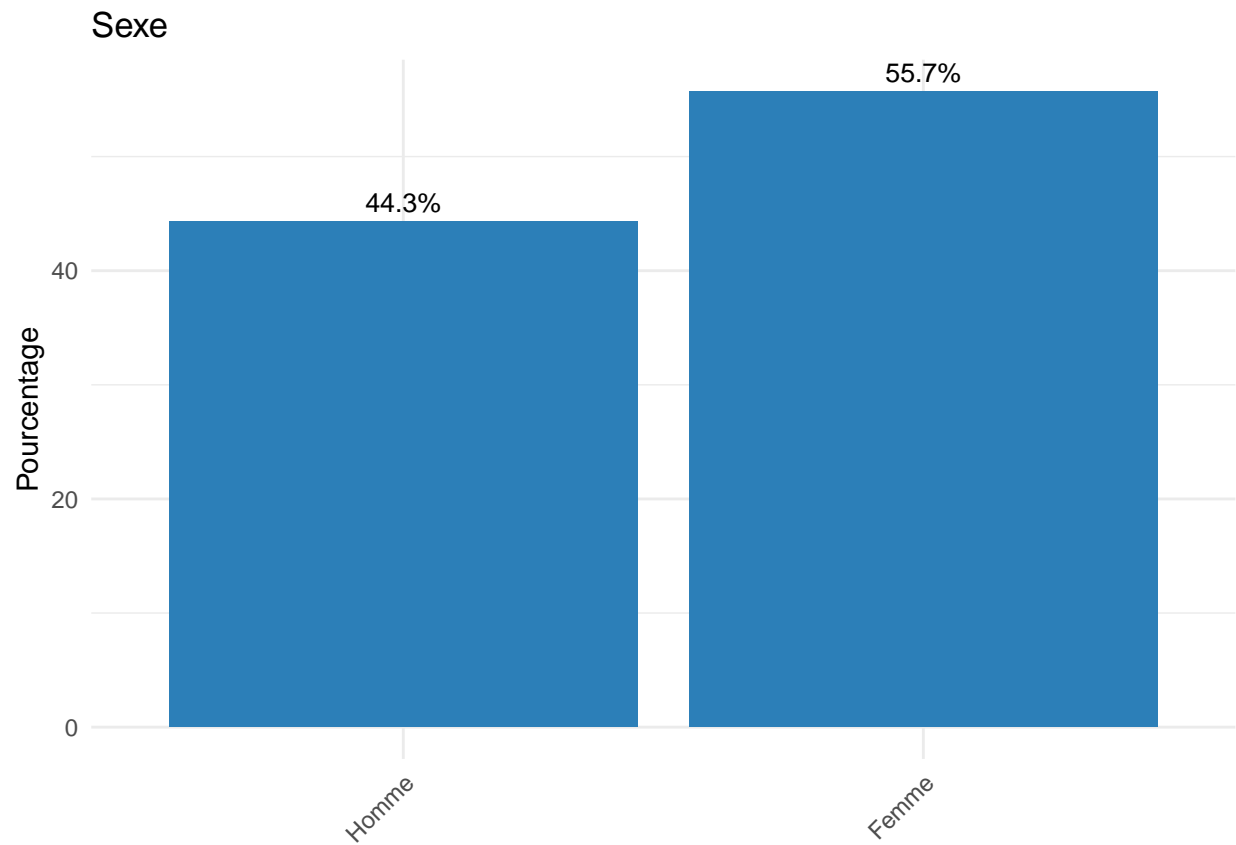
plot_bar <- function(var, data, title = NULL) {
  data %>%
    filter(!is.na(.data[[var]])) %>%
    count(var = .data[[var]]) %>%
    mutate(pct = 100 * n / sum(n)) %>%
    ggplot(aes(x = var, y = pct)) +
    geom_bar(stat = "identity", fill = "#2c7fb8") +
    geom_text(aes(label = paste0(round(pct, 1), "%")), vjust = -0.5, size = 3.5) +
    labs(title = title, x = NULL, y = "Pourcentage") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
}

# Affichage des graphiques
plot_bar("tranche_age", data_model, "Tranche d'âge")

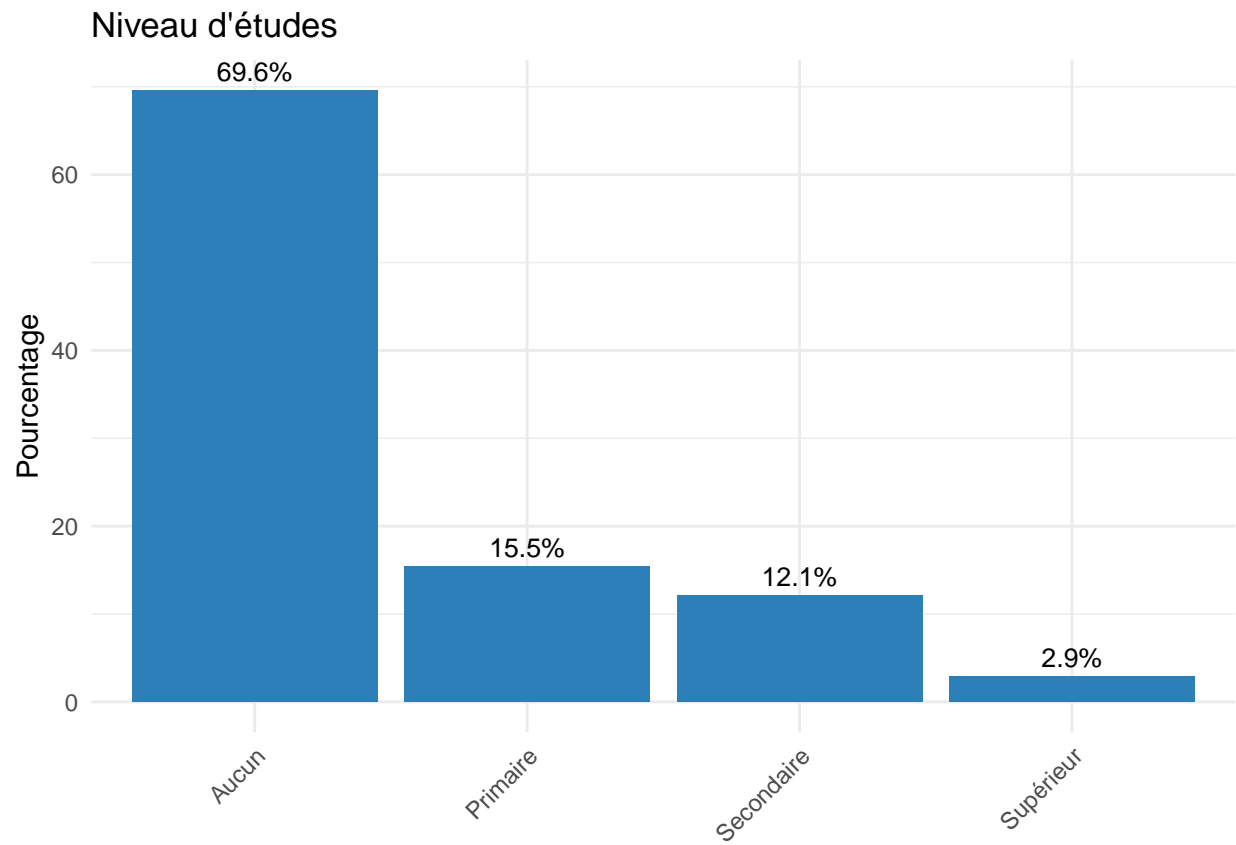
```



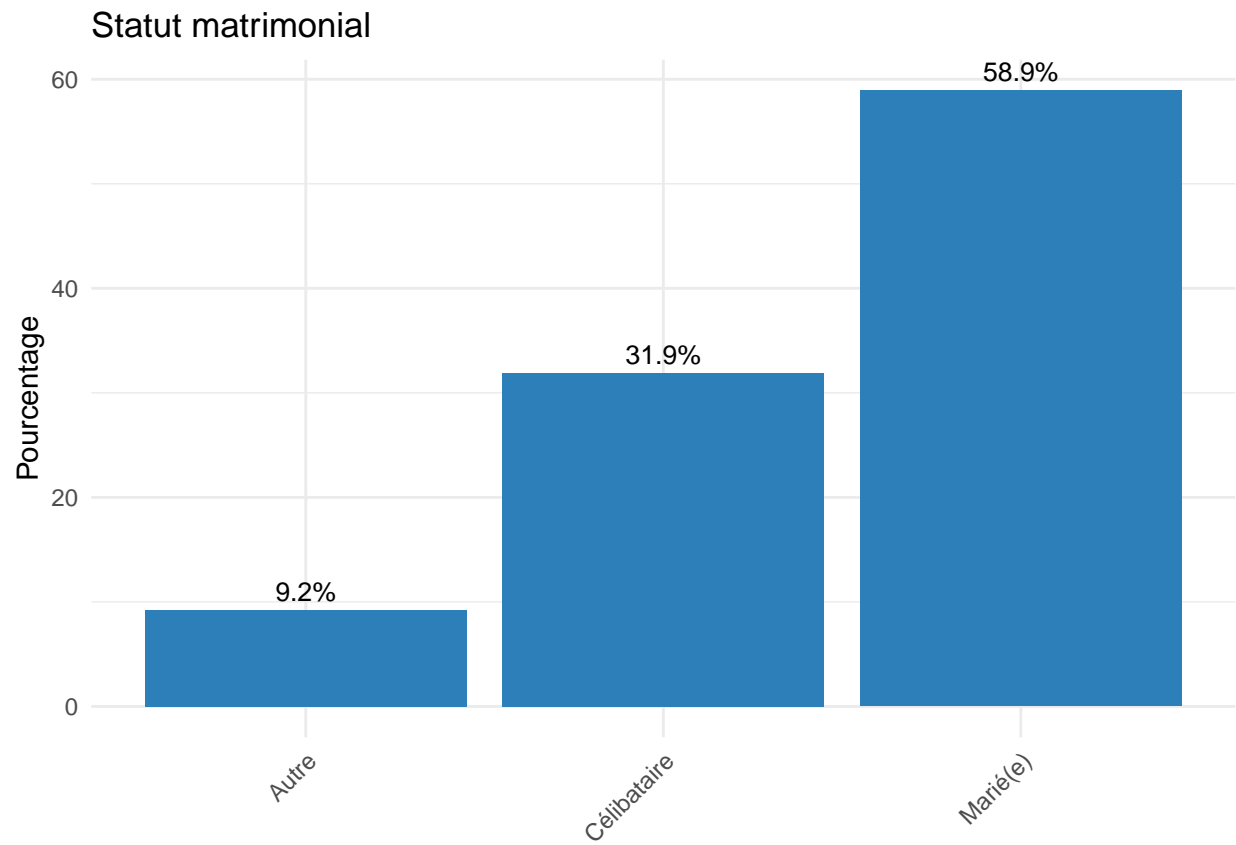
```
plot_bar("sexe", data_model, "Sexe")
```



```
plot_bar("niveau_etude_cat", data_model, "Niveau d'études")
```

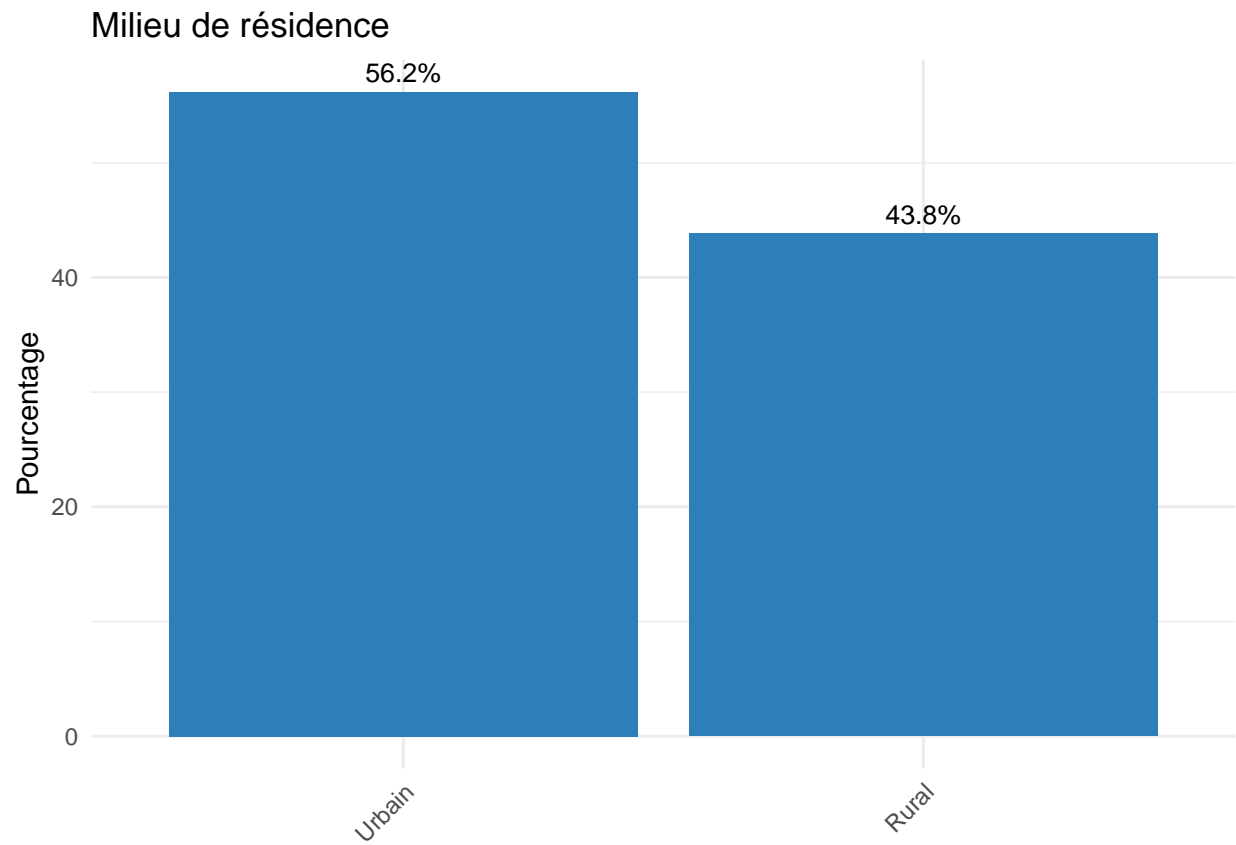


```
plot_bar("statut_matrimonial", data_model, "Statut matrimonial")
```

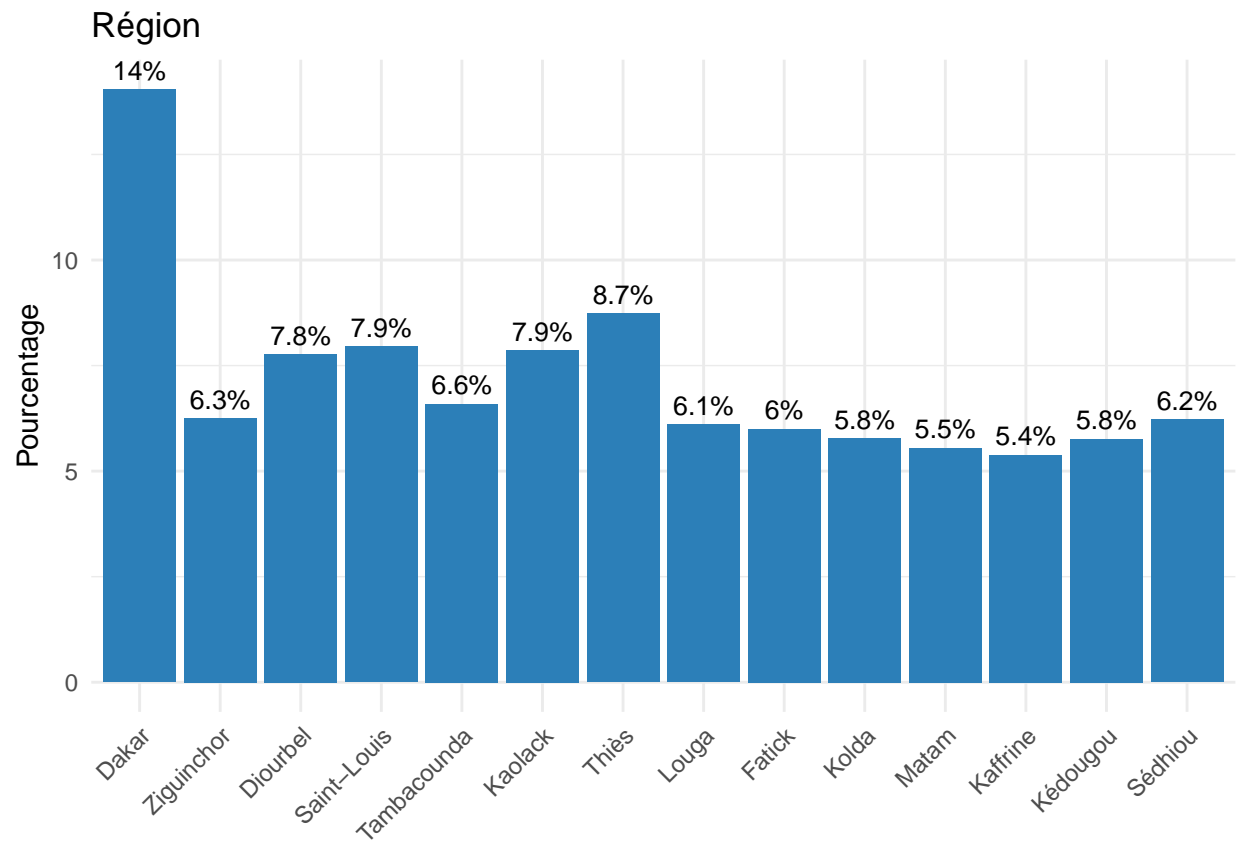


```
plot_bar("Milieu", data_model, "Milieu de résidence")
```

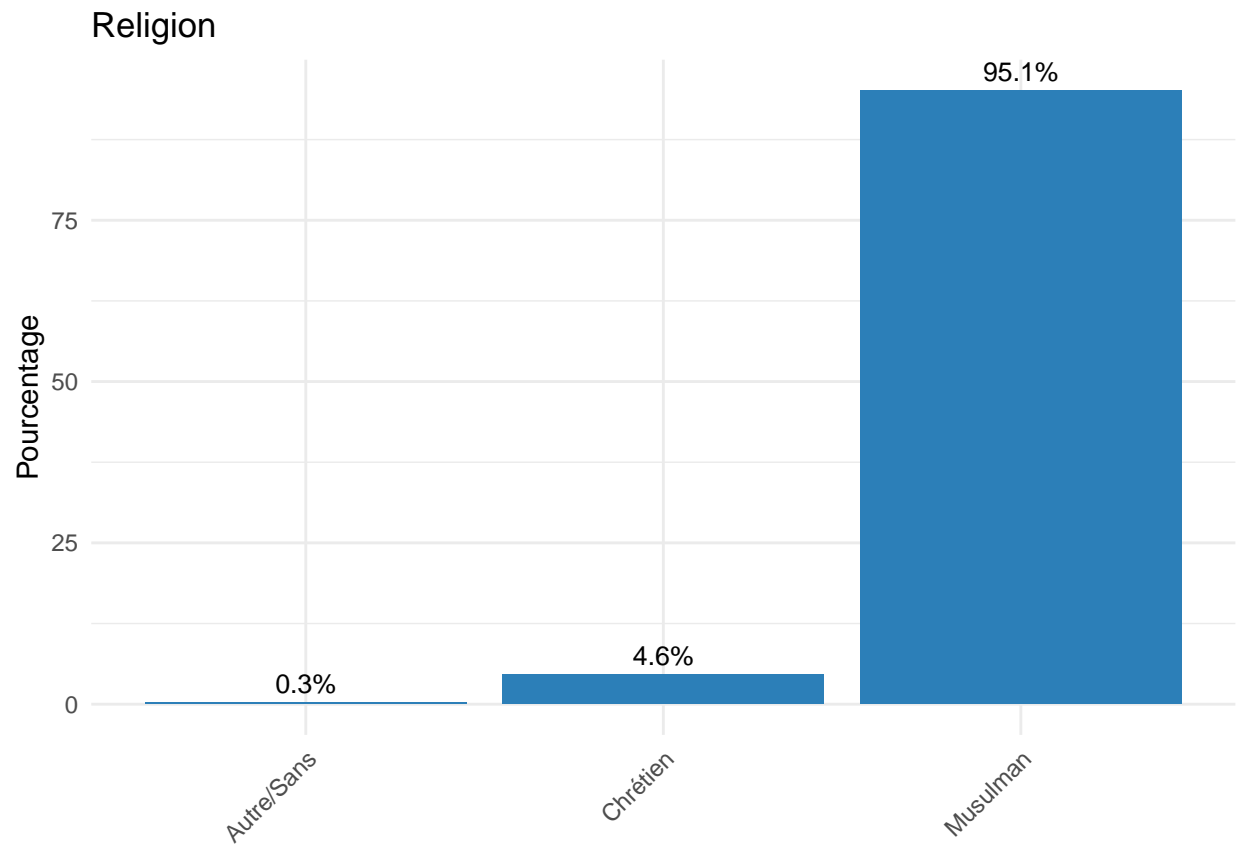




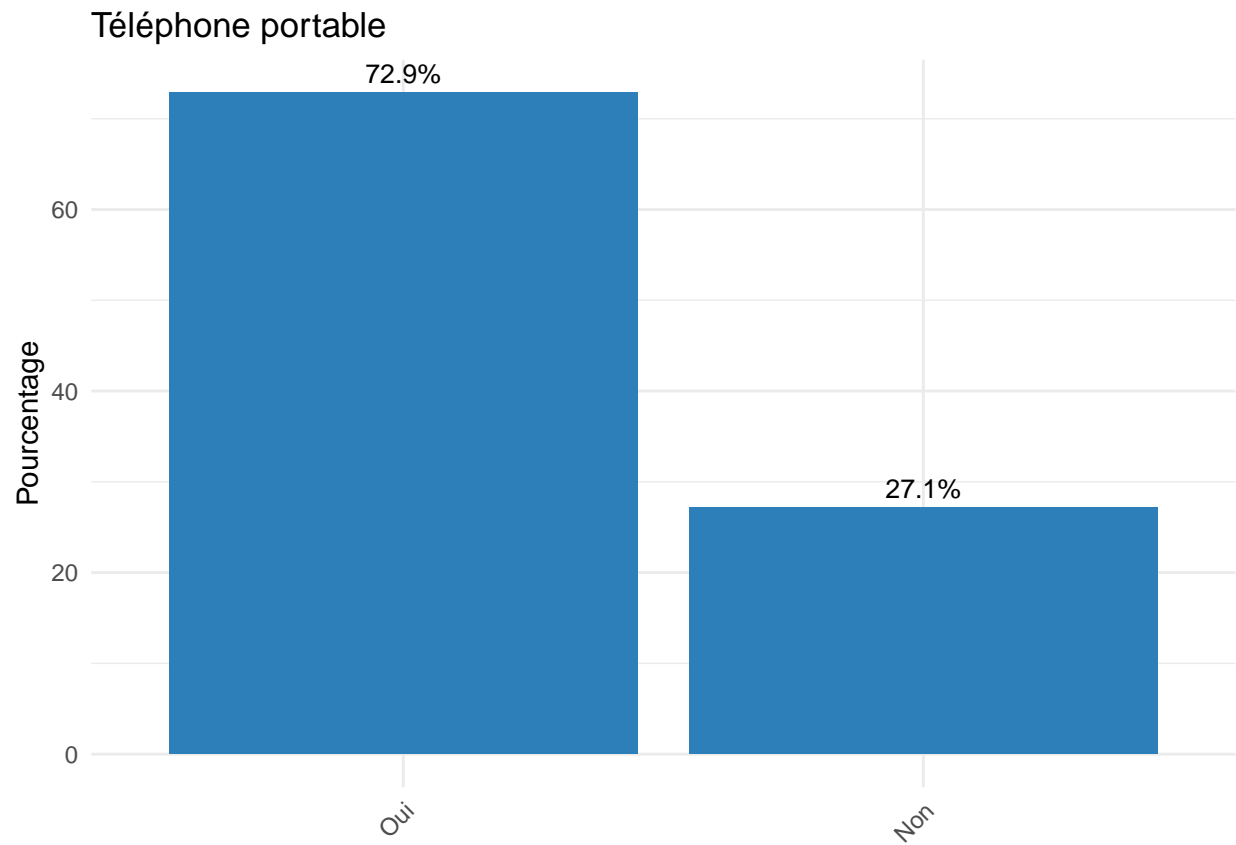
```
plot_bar("Region", data_model, "Région")
```



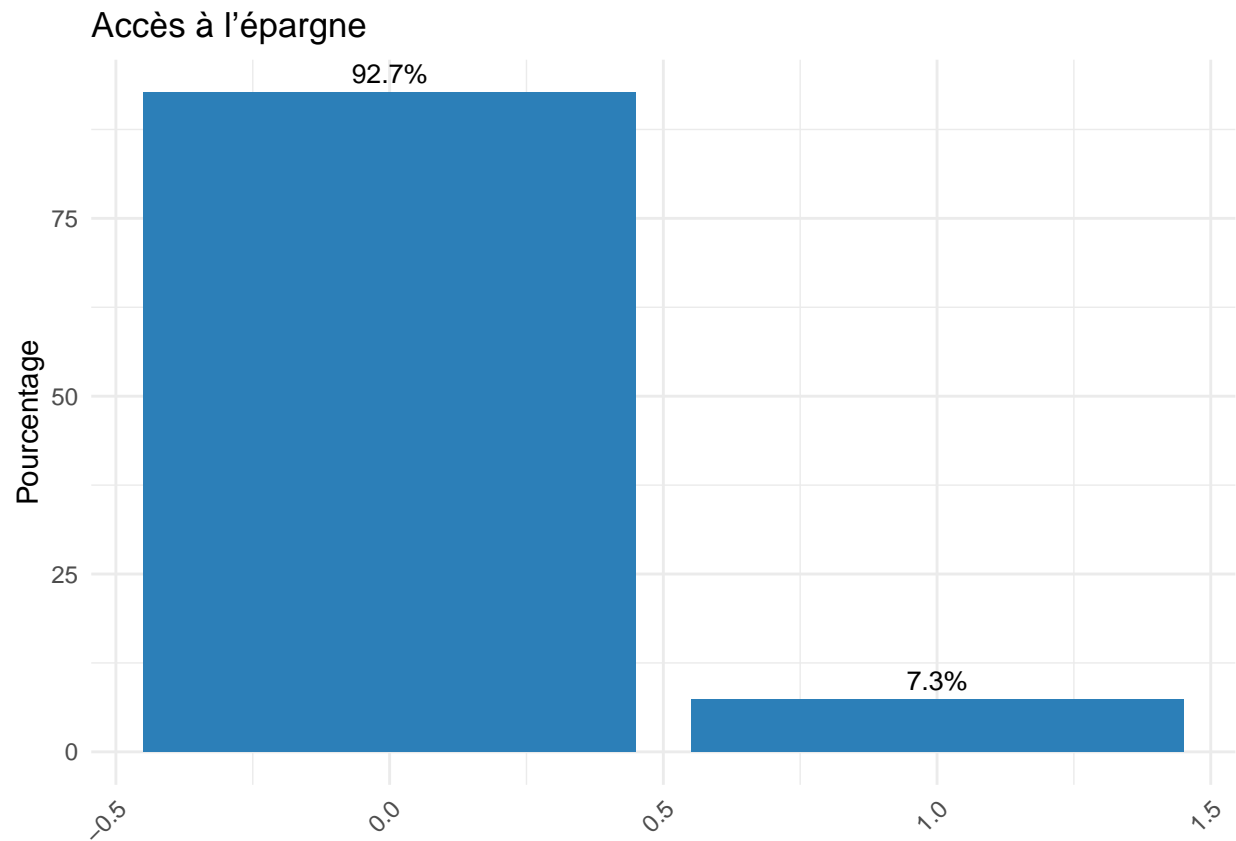
```
plot_bar("religion_cat", data_model, "Religion")
```



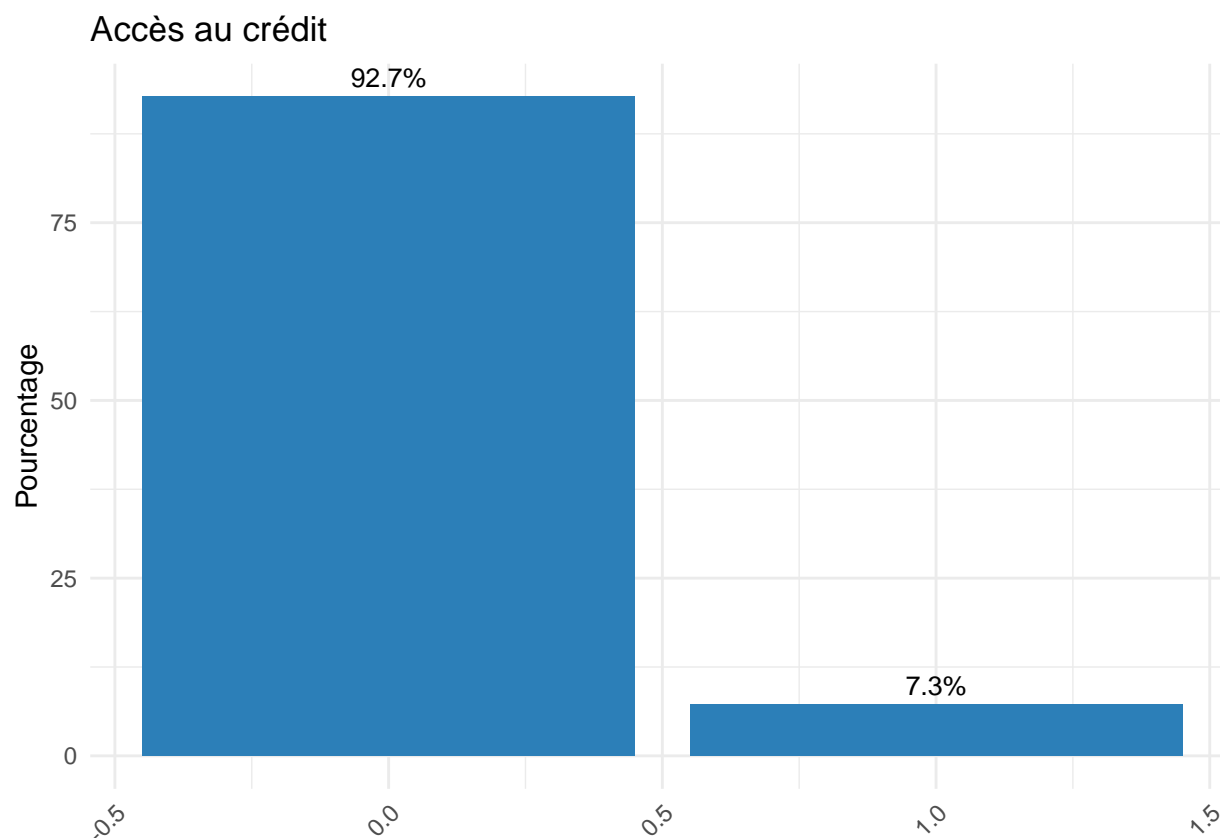
```
plot_bar("telephone", data_model, "Téléphone portable")
```



```
plot_bar("a_epargne", data_model, "Accès à l'épargne")
```



```
plot_bar("a_credit", data_model, "Accès au crédit")
```



## 7. Conclusion

Ce document présente les principales étapes de traitement des données issues de l'enquête EHCVM 2018. Nous avons nettoyé, recodé et structuré les variables nécessaires à l'analyse de l'accès au crédit et à l'épargne. Ces données prêtes à l'emploi serviront à la suite de l'étude, qui portera sur la modélisation statistique des facteurs d'inclusion financière. Le présent document constitue une base propre et structurée, prête pour l'analyse statistique (modélisation logistique). La suite de l'analyse est présentée dans un fichier distinct portant sur la modélisation des déterminants de l'inclusion financière.