REPUBLIQUE DU SENEGAL

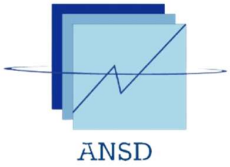
Un Peuple - Un But - Une Foi

****

Ministère de l'Economie, du Plan et de la Coopération

****

Agence Nationale de la Statistique et de la Démographie



Ecole Nationale de la Statistique et de l'Analyse Economique Pierre NDIAYE (ENSAE)

PROJET R

PROJET STATISTIQUE SUR R

Rédigé par :

Komi Amégbor Richard GOZAN

Elève Ingénieur

Statisticien économiste

Sous la supervision de :

M. Hema ABOUBACAR

Research analyst

Juillet 2023

Table des matières

1 Chargement des packages

```
library(readxl)
library(dplyr)
library(gtsummary)
library(gt)
library(sf)
library(leaflet)
library(ggplot2)
library(flextable)
library(ggspatial)
library(broom)
library(questionr)
library(lubridate)
library(GGally)
```

2 Partie 1

2.1 Préparation des données

2.1.1 Decription

2.1.2 Importation et mise en forme

- Importation de la base de données dans un objet de type data.frame nommé projet :

```
projet <- read_excel("Bases/Base_Partie 1.xlsx")
```

- Tableau résumant les valeurs manquantes par variable :

```
df = data.frame(variables = colnames(projet),Valeurs_manquantes = colSums(is.na(projet)))
df %>% gt()
```

variables	Valeurs_manquantes
key	0
q1	0
q2	0
q23	0
q24	0
q24a_1	0

q24a_2	0
q24a_3	0
q24a_4	0
q24a_5	0
q24a_6	0
q24a_7	0
q24a_9	0
q24a_10	0
q25	0
q26	0
q12	0
q14b	1
q16	1
q17	131
q19	120
q20	0
filiere_1	0
filiere_2	0
filiere_3	0
filiere_4	0
q8	0
q81	0
gps_menlatitude	0
gps_menlongitude	0
submissiondate	0
start	0
today	0

- Vérifions s'il y a des valeurs manquantes pour la variable key dans la base projet. Si oui, nous identifierons la (ou les) PME concernée(s).

2.1.3 Création de variables

- Renommage des variables

```
# q1 en region, q2 en déparement et q23 en sexe
projet <- projet %>%
  rename(region = q1, departement = q2, sexe = q23)
```

- Création de la variable sexe_2 vallant 1 si sexe égale à Femme et 0 sinon

```
projet <- projet %>%
  mutate(sexe_2 = ifelse(sexe == "Femme", 1, 0))
```

- Création du data.frame “langues”

```
langues <- projet %>%
  select(key, starts_with("q24a_"))
```

- Création de la variable “parle”

```
langues <- langues %>%
  mutate(parle = rowSums(.[2:ncol(.)]))
```

- Sélection des variables “key” et “parle”

```
langues <- langues %>%
  select(key, parle)
```

- Mergons les data.frame “projet” et “langues”

```
projet <- projet %>%
  left_join(langues, by = "key")
```

2.2 Statistiques descriptives

2.2.1 Satistiques descriptives demandées

Il nous est demandé la répartion des PME suivant:

- le sexe
- le niveau d’instruction
- le statut juridique
- le propriétaire/locataire
- le statut juridique et le sexe
- le niveau d’instruction et le sexe
- Propriétaire/locataire suivant le sexe

Nous résumons ces répartitions dans un seul tableau avec l’analyse univariée et l’anlyse buvariée

```r
# Répartition des PME suivant le sexe, le niveau d'instruction,
# le statut juridique et le propriétaire locataire
tbl1 <- projet %>% tbl_summary(
  include = c("sexe", "q25", "q12", "q81"),
  label=list(q25~ "Niveau d'instruction",
             q12~ "Statut juridique",
             q81~ "Propriétaire/locataire"))

# Répartition des PME suivant le statut juridique et le sexe,
# le niveau d'instruction et le sexe,   Propriétaire/locataire
# et le sexe
tbl2 <- projet %>% tbl_summary(
  include = c("q25", "q12", "q81"),
  by ="sexe", label=list(q12~ "Statut juridique",
                         q25~ "Niveau d'instruction",
                         q81~ "Propriétaire/locataire")) %>%
  modify_header(label ~ "**Variables**") %>%
  modify_spanning_header(all_stat_cols() ~ "**Sexe**") %>%
  bold_labels()

## Empillement l'une sur l'autre
tbl_merge(
    tbls = list(tbl1, tbl2),
    tab_spanner = c("**Analyse univariée**", "**Analye bivariée**")) %>%
  italicize_levels() %>%
  as_flex_table() %>%
  fontsize(size=10) %>%
  width(width = 1.3)
```

	Analyse univariée	Analye bivariée	
Characteristic	**N = 250**[1]	**Femme**, N = 191[1]	**Homme**, N = 59[1]
sexe			
Femme	191 (76%)		
[1]n (%)			

	Analyse univariée	Analye bivariée	
Characteristic	**N = 250**[1]	**Femme**, N = 191[1]	**Homme**, N = 59[1]
Homme	59 (24%)		
Niveau d'instruction			
Aucun niveau	79 (32%)	70 (37%)	9 (15%)
Niveau primaire	56 (22%)	48 (25%)	8 (14%)
Niveau secondaire	74 (30%)	56 (29%)	18 (31%)
Niveau Superieur	41 (16%)	17 (8.9%)	24 (41%)
Statut juridique			
Association	6 (2.4%)	3 (1.6%)	3 (5.1%)
GIE	179 (72%)	149 (78%)	30 (51%)
Informel	38 (15%)	32 (17%)	6 (10%)
SA	7 (2.8%)	1 (0.5%)	6 (10%)
SARL	13 (5.2%)	2 (1.0%)	11 (19%)
SUARL	7 (2.8%)	4 (2.1%)	3 (5.1%)
Propriétaire/locataire			
Locataire	24 (9.6%)	16 (8.4%)	8 (14%)
Propriétaire	226 (90%)	175 (92%)	51 (86%)

[1] n (%)

2.2.2 Statistiques descriptives de notre choix sur les autres variables:

2.2.2.1 Analyse univariée:

- **Analyse des filières**

```r
projet2 <- projet %>%
  rename(arachide = filiere_1, anacarde = filiere_2, mangue = filiere_3, riz =filiere_4)
projet2 %>% tbl_summary(
      include = c("arachide", "anacarde", "mangue", "riz"),
      label=list(arachide~ "Arachide",
             anacarde~ "anacarde",
             mangue~ "mangue",
             riz~ "riz")) %>%
      bold_labels()
```

Characteristic	N = 250
Arachide	108 (43%)
anacarde	61 (24%)
mangue	89 (36%)
riz	92 (37%)

- **Statistiques descriptives pour les variables de date**

```r
# Nous copie notre objet projet dans l'objet data
data = projet
# Convertir les colonnes en format de date
data$submissiondate <- as_date(data$submissiondate)
data$start <- as_date(data$start)
data$today <- as_date(data$today)

# Résumé statistique des colonnes de date
summary(data$submissiondate)
```

```
##         Min.      1st Qu.       Median         Mean      3rd Qu.         Max.
## "2021-05-17" "2021-06-08" "2021-06-11" "2021-06-09" "2021-06-15" "2021-06-21"
```

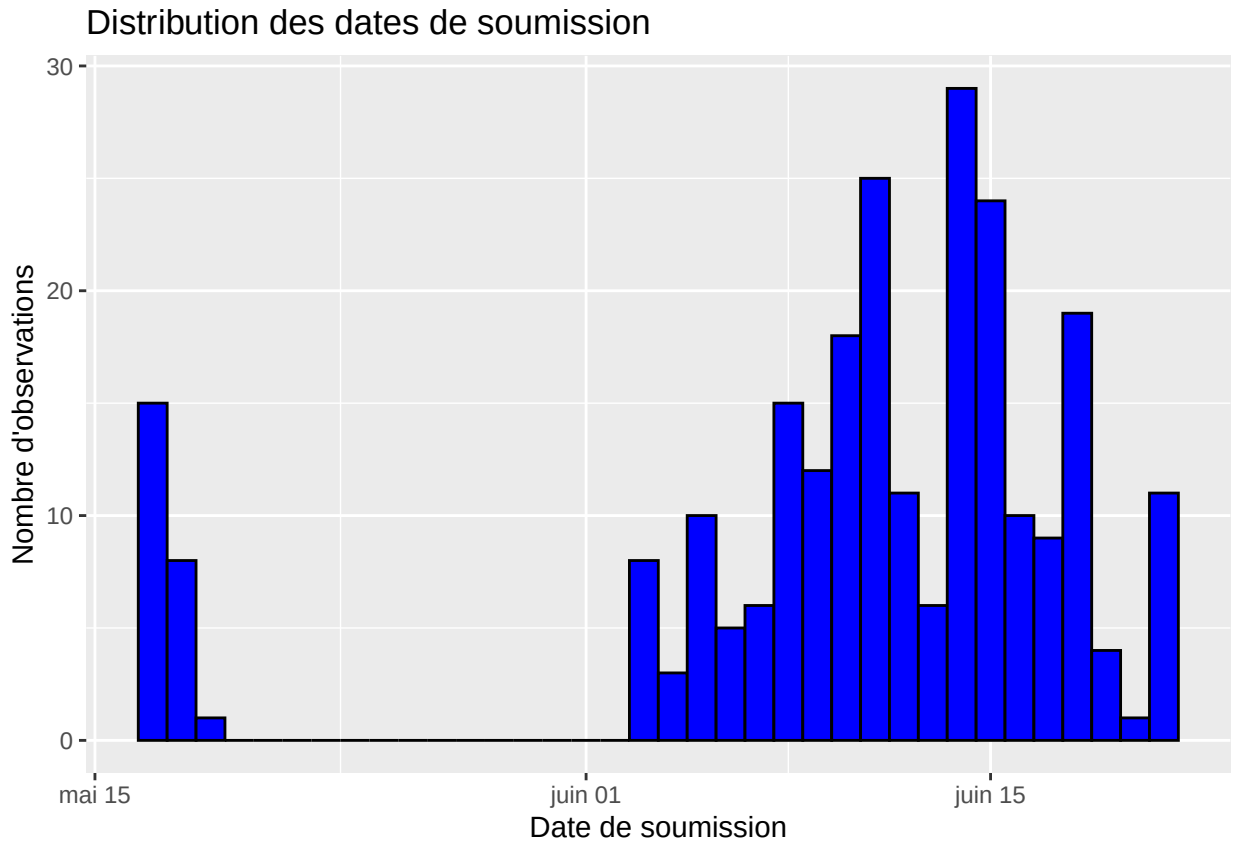
```r
summary(data$start)
```

```
##         Min.      1st Qu.       Median         Mean      3rd Qu.         Max.
## "2021-05-06" "2021-06-03" "2021-06-07" "2021-06-03" "2021-06-10" "2021-06-20"
```

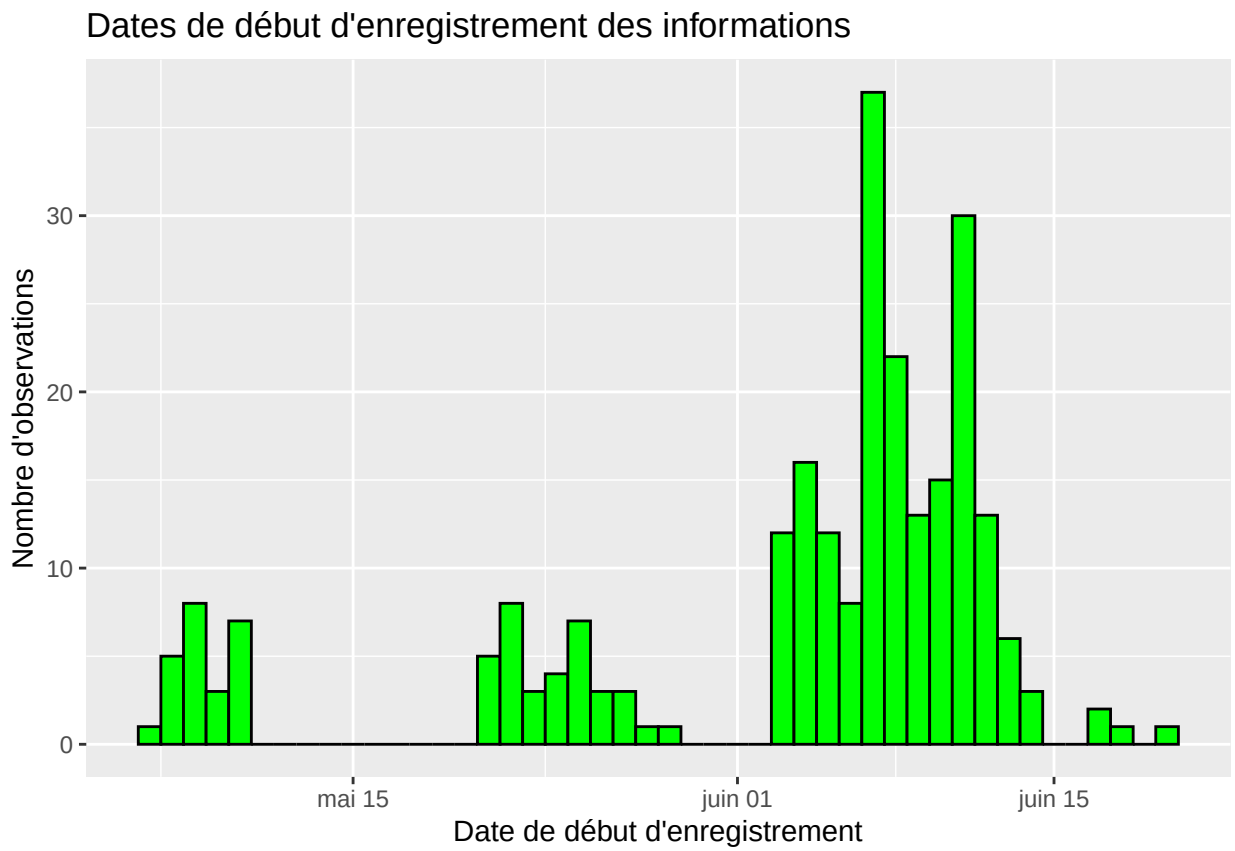
```r
summary(data$today)
```

```
##         Min.      1st Qu.       Median         Mean      3rd Qu.         Max.
## "2021-05-06" "2021-06-03" "2021-06-07" "2021-06-03" "2021-06-10" "2021-06-20"
```

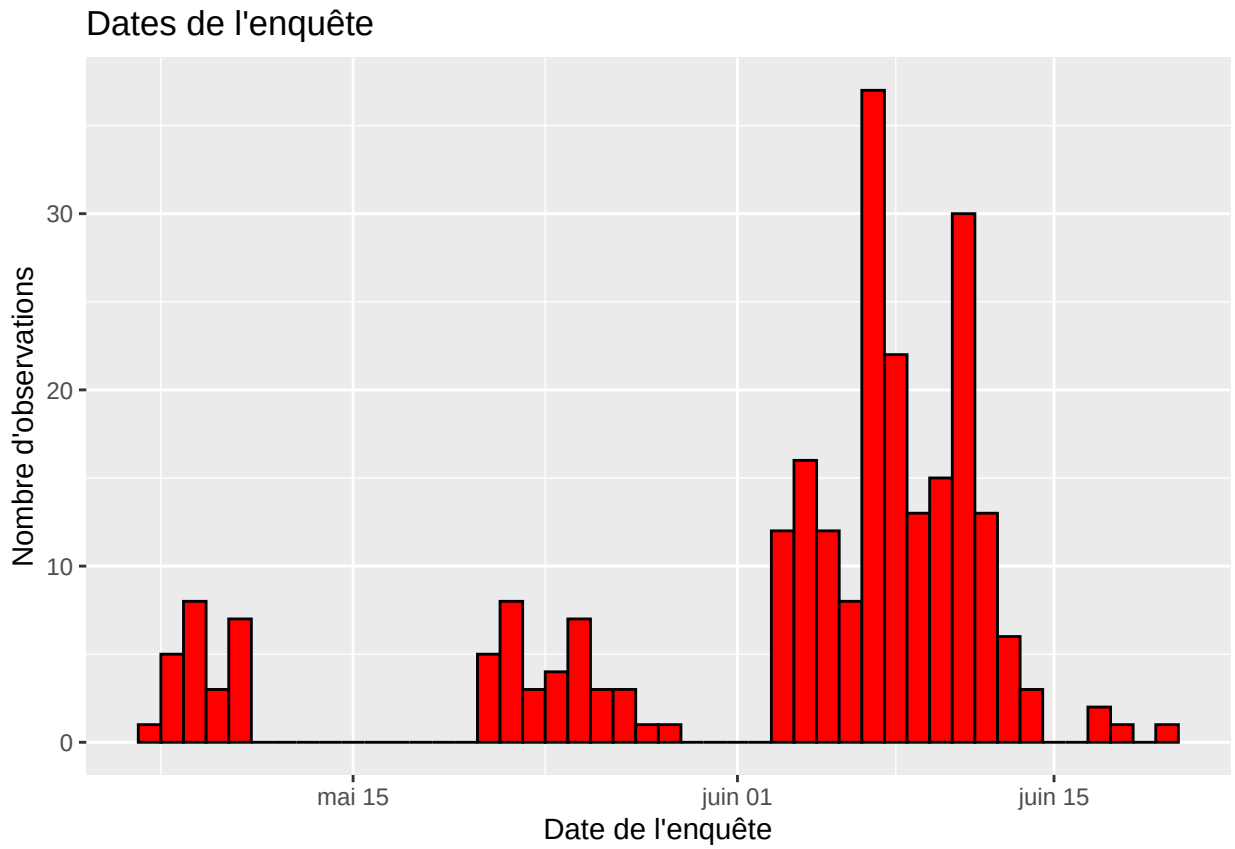
```r
# Graphique de la distribution des dates de soumission
ggplot(data, aes(x = submissiondate)) +
  geom_histogram(binwidth = 1, fill = "blue", color = "black") +
  labs(title = "Distribution des dates de soumission",
       x = "Date de soumission",
       y = "Nombre d'observations")
```




```r
# Graphique des dates de début d'enregistrement des informations
ggplot(data, aes(x = start)) +
  geom_histogram(binwidth = 1, fill = "green", color = "black") +
  labs(title = "Dates de début d'enregistrement des informations",
       x = "Date de début d'enregistrement",
       y = "Nombre d'observations")
```




```r
# Graphique des dates de l'enquête
ggplot(data, aes(x = today)) +
  geom_histogram(binwidth = 1, fill = "red", color = "black") +
  labs(title = "Dates de l'enquête",
       x = "Date de l'enquête",
       y = "Nombre d'observations")
```




```
# Tableau récapitulatif des dates
tbl_summary(data,
            missing = "no",
            include = c(submissiondate, start, today),
            label=list(submissiondate~ "Date de soumission",
                       start~ "Date de début de l'enrégistrement",
                      today~ "Date de l'enquête")) %>%
  modify_header(label ~ "**Date**") %>%
  italicize_labels() %>%
  as_flex_table()
```

Date	**N = 250**[1]
Date de soumission	2021-05-17 to 2021-06-21
Date de début de l'enrégistrement	2021-05-06 to 2021-06-20
Date de l'enquête	2021-05-06 to 2021-06-20

[1]Range

Date	**N = 250**[1]

[1]Range

Analyse :

• La date de soumission varie du 17 mai 2021 au 21 juin 2021 : Cela indique que les informations de l'enquête ont été soumises sur une période d'environ un mois.

• La date de début d'enregistrement varie du 6 mai 2021 au 20 juin 2021 : Cela signifie que l'enregistrement des informations pour l'enquête a commencé à partir du 6 mai 2021 et s'est poursuivi jusqu'au 20 juin 2021.

• La date de l'enquête varie également du 6 mai 2021 au 20 juin 2021 : Cela indique que les enquêtes ont été menées sur la même période que l'enregistrement des informations.

2.2.2.2 Analyse bivariée:

- **Croisement de avec les filières**

```
# Nous renommons les variables filières aux noms des différents filières

# Nous créeons une fonction croiser_filiere qui prend en argument
# une filière et renvoie une analyse suivant la filière donnée
croiser_filiere <- function(projet, filiere) {
  tbl_summary_result <- projet %>%
    tbl_summary(
      include = c("sexe", "region", "q19"),
      by = {{ filiere }},
      label = list(sexe ~ "Sexe",
                   region ~ "Region",
                   q19 ~ "Etat de la piste qui mène à l'entreprise")
    ) %>%
    modify_header(label ~ "**Variable**") %>%
    bold_labels()

  return(tbl_summary_result)
}
```

```r
tbl_filiere_1 <- croiser_filiere(projet2, filiere = "arachide")
tbl_filiere_2 <- croiser_filiere(projet2, filiere = "anacarde")
tbl_filiere_3 <- croiser_filiere(projet2, filiere = "mangue")
tbl_filiere_4 <- croiser_filiere(projet2, filiere = "riz")

tbl_merge(
  list(tbl_filiere_1, tbl_filiere_2, tbl_filiere_3, tbl_filiere_4),
  tab_spanner = c("arachide", "anacarde", "mangue", "riz")) %>%
  italicize_levels() %>%
  as_flex_table() %>%
  fontsize(size=8) %>%
  width(width = 0.8 )
```

	arachide		anacarde		mangue		riz	
Variable	**0**, N = 142[1]	**1**, N = 108[1]	**0**, N = 189[1]	**1**, N = 61[1]	**0**, N = 161[1]	**1**, N = 89[1]	**0**, N = 158[1]	**1**, N = 92[1]
Sexe								
Femme	98 (69%)	93 (86%)	151 (80%)	40 (66%)	123 (76%)	68 (76%)	114 (72%)	77 (84%)
Homme	44 (31%)	15 (14%)	38 (20%)	21 (34%)	38 (24%)	21 (24%)	44 (28%)	15 (16%)
Region								
Dakar	1 (0.7%)	0 (0%)	0 (0%)	1 (1.6%)	1 (0.6%)	0 (0%)	0 (0%)	1 (1.1%)
Diourbel	1 (0.7%)	33 (31%)	34 (18%)	0 (0%)	33 (20%)	1 (1.1%)	34 (22%)	0 (0%)
Fatick	18 (13%)	12 (11%)	9 (4.8%)	21 (34%)	27 (17%)	3 (3.4%)	26 (16%)	4 (4.3%)
Kaffrine	0 (0%)	8 (7.4%)	8 (4.2%)	0 (0%)	3 (1.9%)	5 (5.6%)	7 (4.4%)	1 (1.1%)
Kaolack	1 (0.7%)	20 (19%)	21 (11%)	0 (0%)	14 (8.7%)	7 (7.9%)	17 (11%)	4 (4.3%)
Kolda	8 (5.6%)	1 (0.9%)	4 (2.1%)	5 (8.2%)	9 (5.6%)	0 (0%)	5 (3.2%)	4 (4.3%)
Saint-Louis	41 (29%)	1 (0.9%)	42 (22%)	0 (0%)	0 (0%)	42 (47%)	42 (27%)	0 (0%)
Sédhiou	4 (2.8%)	0 (0%)	1 (0.5%)	3 (4.9%)	4 (2.5%)	0 (0%)	1 (0.6%)	3 (3.3%)
Thiès	24 (17%)	27 (25%)	51 (27%)	0 (0%)	26 (16%)	25 (28%)	19 (12%)	32 (35%)

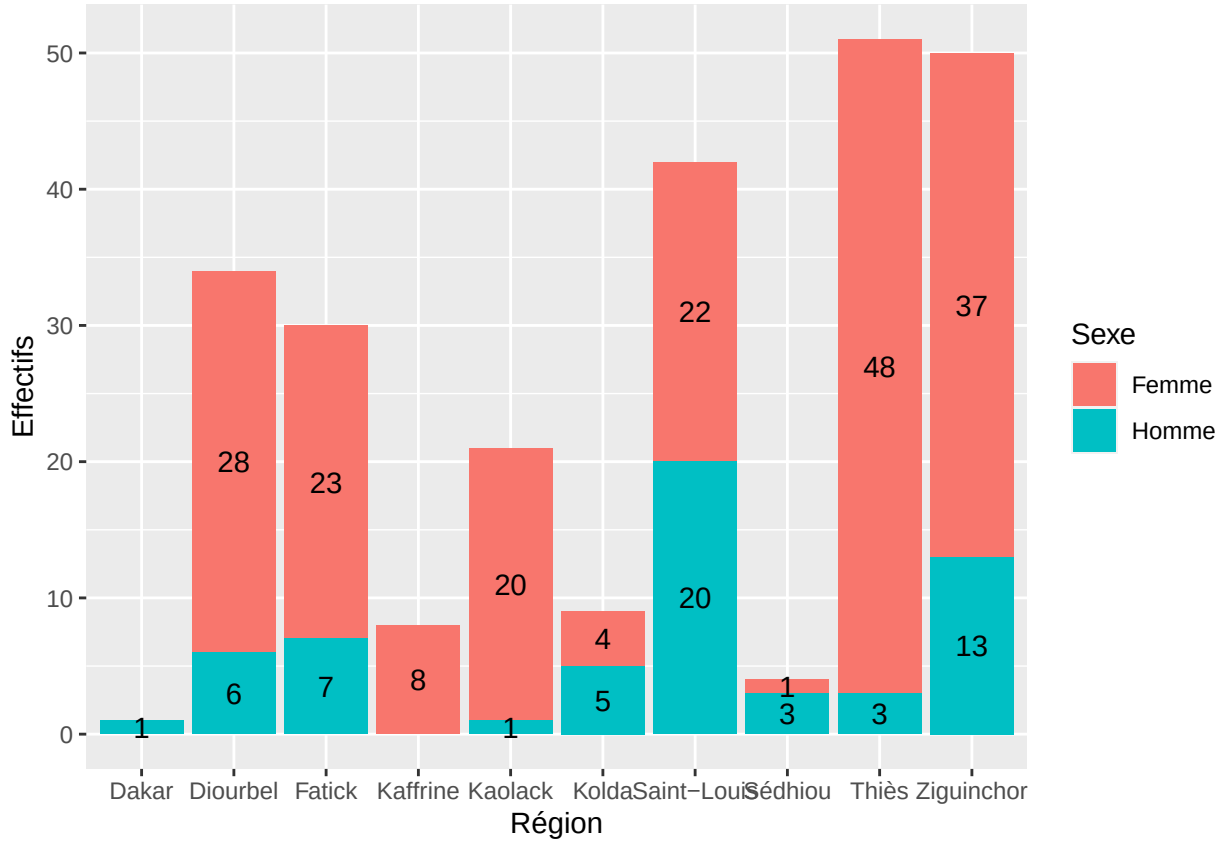
[1] n (%)

	arachide		anacarde		mangue		riz	
Variable	**0**, N = 142[1]	**1**, N = 108[1]	**0**, N = 189[1]	**1**, N = 61[1]	**0**, N = 161[1]	**1**, N = 89[1]	**0**, N = 158[1]	**1**, N = 92[1]
Ziguinchor	44 (31%)	6 (5.6%)	19 (10%)	31 (51%)	44 (27%)	6 (6.7%)	7 (4.4%)	43 (47%)
Etat de la piste qui mène à l'entreprise								
Bon état	2 (3.2%)	0 (0%)	1 (1.0%)	1 (2.9%)	2 (2.1%)	0 (0%)	0 (0%)	2 (4.5%)
Etat moyen	35 (56%)	38 (57%)	55 (57%)	18 (53%)	53 (55%)	20 (59%)	46 (53%)	27 (61%)
Mauvais état	26 (41%)	29 (43%)	40 (42%)	15 (44%)	41 (43%)	14 (41%)	40 (47%)	15 (34%)
Unknown	79	41	93	27	65	55	72	48

[1] n (%)

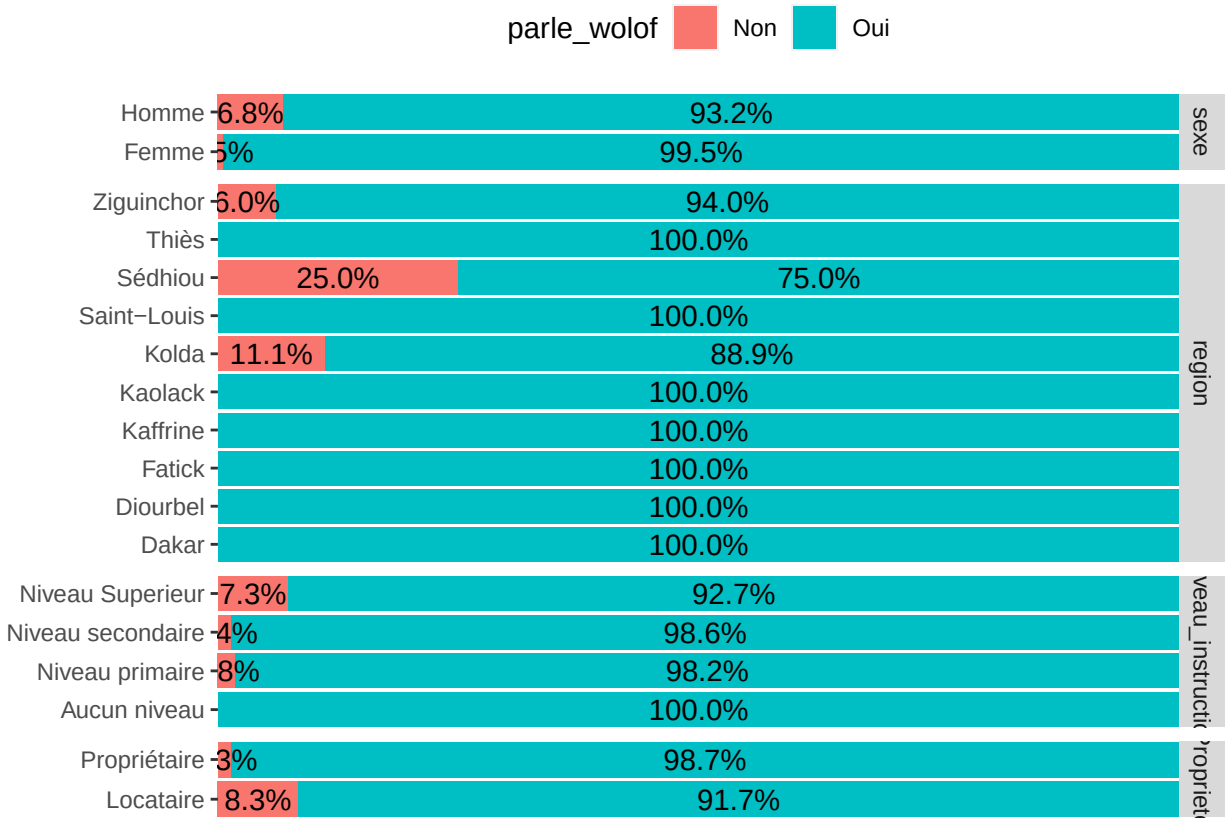
- **Distribution de la population suivant les Regions par sexe:**

```
ggplot(projet) +
  aes(x = region, fill = sexe) +
  geom_bar() +
  geom_text(aes(label = after_stat(count)), stat = "count", position = position_stack(.5)) +
  xlab("Région") +
  ylab("Effectifs") +
  labs(fill = "Sexe")
```

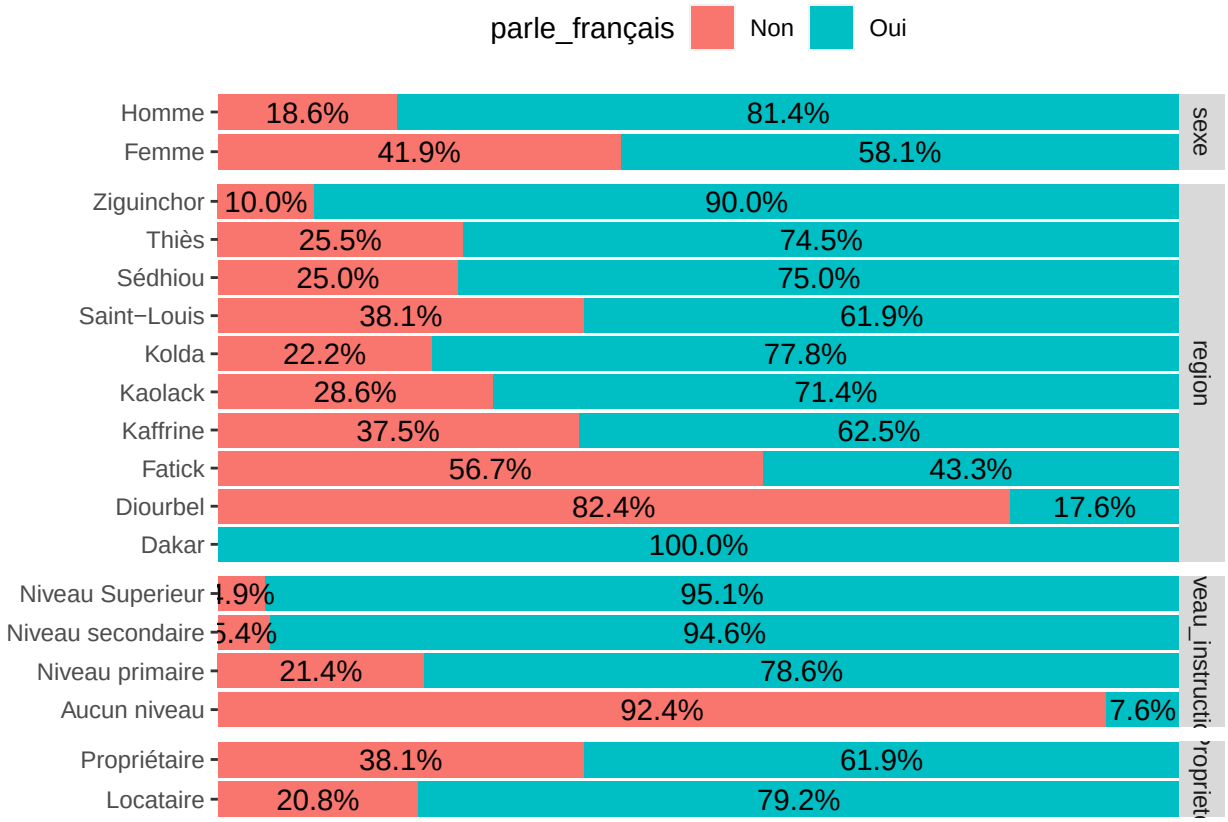


- **Distribution de la population suivant la langue parlée(Fraçais et wolof):**

```
data <- projet %>%
  rename(Niveau_instruction = q25, Propriete = q81)
data <- data %>%
  mutate(parle_français = ifelse(q24a_1 == 1, "Oui", "Non"),
         parle_wolof = ifelse(q24a_2 == 1, "Oui", "Non"))
ggbivariate(data = data , outcome = "parle_wolof", explanatory = c("sexe", "region", "Niveau_instructio
```




```
ggbivariate(data = data , outcome = "parle_français", explanatory = c("sexe", "region", "Niveau_instruc
```



2.3 Cartographie

- Transformation du data.frame en données géographiques dont l'objet est nommé projet_map :

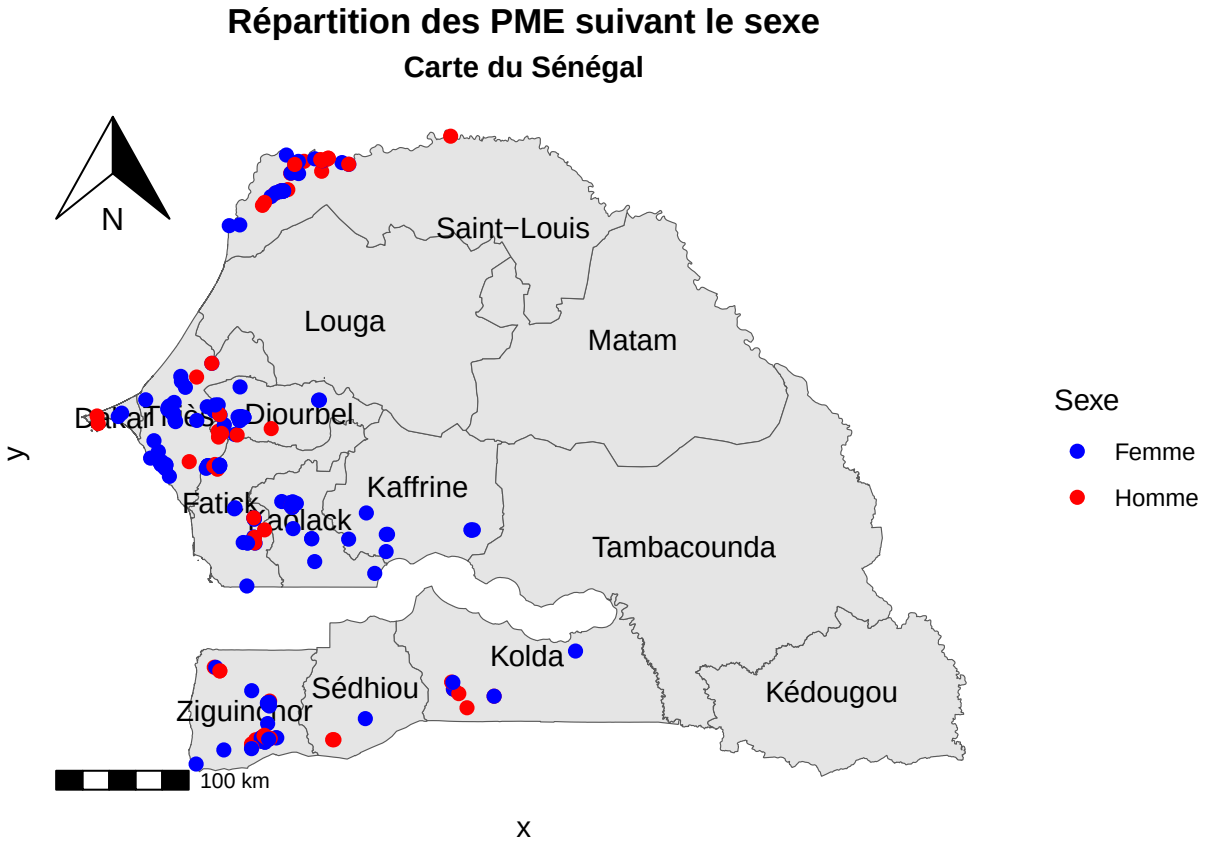
```
# Chargement des données de la carte du Sénégal
# Nous lisons le fichier contenant la carte du Sénégal et le stockons
# dans la variable "senegal"
senegal = st_read("Bases/gadm41_SEN_shp/gadm41_SEN_1.shp")
```

```
## Reading layer `gadm41_SEN_1' from data source
##   `D:\Mon phone\ISE\1Ã¨re annÃ©e\S2\Projet sur R\Projet_R_ENSAE_2023\Bases\gadm41_SEN_shp\gadm41_SEN_
##   using driver `ESRI Shapefile'
## Simple feature collection with 14 features and 11 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -17.54319 ymin: 12.30786 xmax: -11.34247 ymax: 16.69207
## Geodetic CRS:  WGS 84
```

```r
# La fonction st_as_sf convertit l'objet "projet" en objet
# sf (spatial feature) en spécifiant les coordonnées GPS à utiliser
projet_map = st_as_sf(projet, coords = c("gps_menlongitude", "gps_menlatitude"), crs=st_crs(senegal), ag

# Jointure des données du projet avec la carte du Sénégal
projet_map = st_join(projet_map, senegal)
```

- Représentation spatiale des PME suivant le sexe:

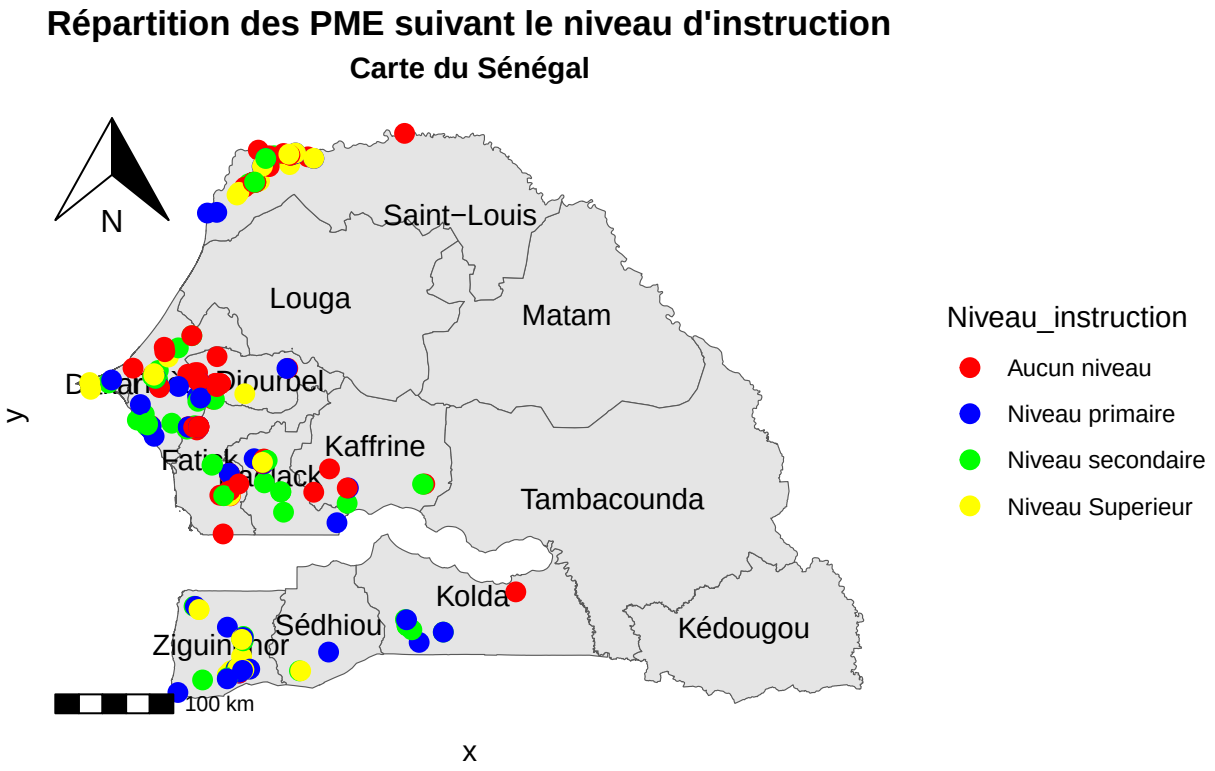
```r
# Création de la première carte avec la répartition des PME suivant le sexe
ggplot() +
  geom_sf(data=senegal)+
  geom_sf_text(data=senegal, aes(label=NAME_1))+
  geom_sf(data=projet_map, aes(color=sexe), size=2)+
  scale_color_manual(values=c("blue", "red"))+
  labs(title = "Répartition des PME suivant le sexe",
       subtitle = "Carte du Sénégal",
       color = "Sexe") +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5, face = "bold"))+
  coord_sf(datum = NA)+
  annotation_scale(location = "bl", text_col = "black")+
  annotation_north_arrow(location = "tl")
```



- Représentation spatiale des PME suivant le niveau d'instruction:

```
# Renommons la variable q25 en Niveau_instruction
projet_map <- projet_map %>% rename(Niveau_instruction = q25)


# Création de la deuxième carte avec la répartition des PME
# suivant le niveau d'instruction
ggplot() +
  geom_sf(data=senegal)+
  geom_sf_text(data=senegal, aes(label=NAME_1))+
  geom_sf(data = projet_map, aes(color = Niveau_instruction), size =3) +
  scale_color_manual(values=c("red", "blue", "green", "yellow")) +
  labs(title = "Répartition des PME suivant le niveau d'instruction",
       subtitle = "Carte du Sénégal") +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5, face = "bold"))+
```

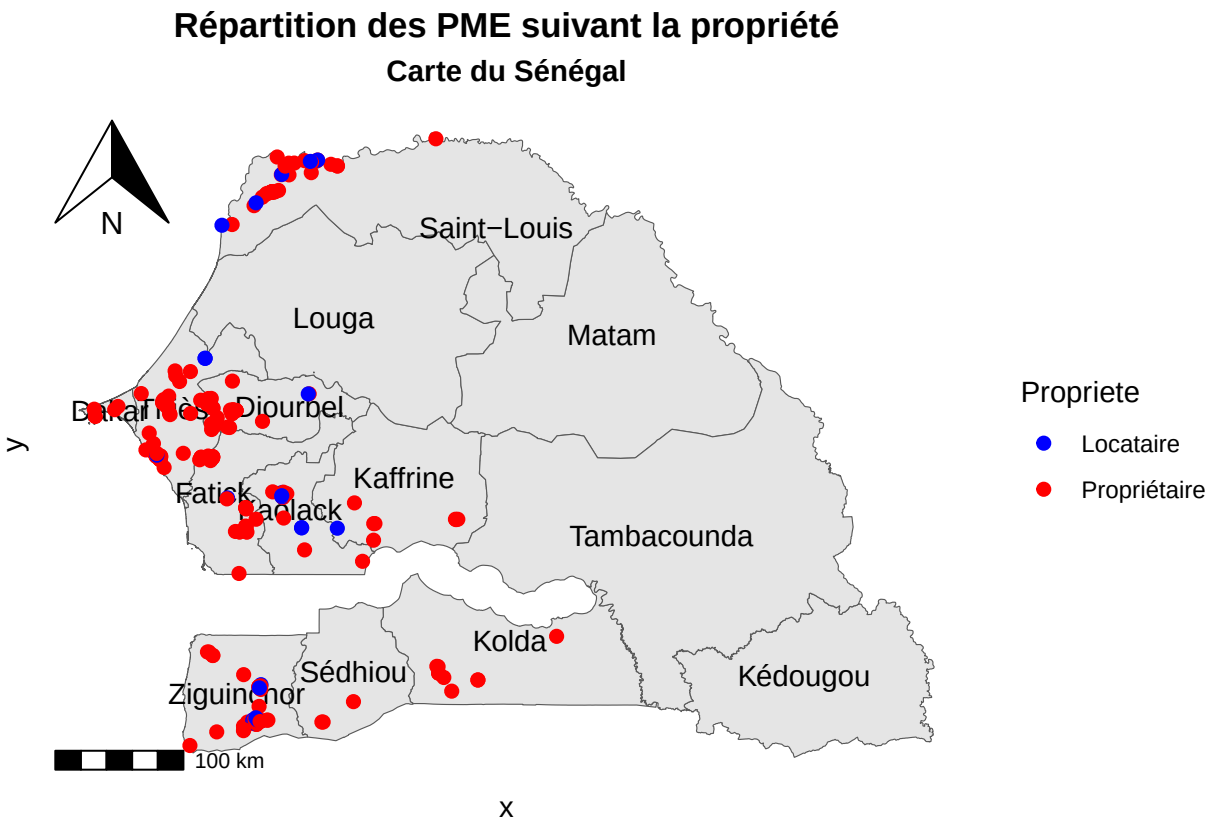
```
  coord_sf(datum = NA)+
  annotation_scale(location = "bl", text_col = "black")+
  annotation_north_arrow(location = "tl")
```



- Analyse spatiale de notre choix

```
# Création de la troisième carte avec la répartition des PME suivant prop

# Renommons la variable q81 en Propriéte
projet_map <- projet_map %>% rename(Propriete = q81)
ggplot() +
  geom_sf(data=senegal)+
  geom_sf_text(data=senegal, aes(label=NAME_1))+
  geom_sf(data=projet_map, aes(color=Propriete), size=2)+
  scale_color_manual(values=c("blue", "red"))+
  labs(title = "Répartition des PME suivant la propriété",
       subtitle = "Carte du Sénégal") +
  theme_minimal() +
```

```
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5, face = "bold"))+
  coord_sf(datum = NA)+
  annotation_scale(location = "bl", text_col = "black")+
  annotation_north_arrow(location = "tl")
```



3 Partie 2

3.1 Nettoyage et gestion des données

- Chargement des données et effectuons le nettoyage initial

```
# Chargement les données
feuille1 <- read_excel("Bases/Base_Partie 2.xlsx", sheet = 1)


# Renommons la variable "country_destination" en "destination" et
# définissons les valeurs négatives comme manquantes
```

```
feuille1 <- feuille1 %>%
  rename(destination = country_destination) %>%
  mutate(destination = ifelse(destination < 0, NA, destination))
```

- Création d'une nouvelle variable contenant des tranches d'âge de 5 ans

```
# Imputons la valeur abérante par la médiane

# Calculer la médiane des valeurs non aberrantes (différentes de 999) dans la variable "age" et remplac
feuille1 <- feuille1 %>%
  mutate(age = replace(age, age == 999, median(age[age != 999], na.rm = TRUE)))



# Créons la variable tranche_age contenant des tranches d'âge de
# 5 ans
feuille1 <- feuille1 %>%
  mutate(tranche_age = cut(age, breaks = seq(0, max(age), by = 5), include.lowest = TRUE))
```

- Création d'une nouvelle variable contenant le nombre d'entretiens réalisés par chaque agent recenseur :

```
feuille1 <- feuille1 %>%
  group_by(enumerator) %>%
  mutate(nombre_entretiens = n()) %>%
  ungroup()
```

- Création d'une nouvelle variable qui affecte aléatoirement chaque répondant à un groupe de traitement (1) ou de contrôle (0)

```
feuille1 <- feuille1 %>%
  mutate(groupe = {set.seed(123); sample(c(0, 1), size = n(), replace = TRUE)})
```

- Fusion de la taille de la population de chaque district (feuille 2) avec l'ensemble de données (feuille 1) :

```
# Charger les données de la deuxième feuille
feuille2 <- read_excel("Bases/Base_Partie 2.xlsx", sheet = 2)

# Fusion
feuille1 <- feuille1 %>%
  left_join(feuille2, by = "district")
```

- Calcul la durée de l'entretien et indiquer la durée moyenne de l'entretien par enquêteur :

```r
feuille1 <- feuille1 %>%
  mutate(duree_entretien = endtime - starttime) %>%
  group_by(enumerator) %>%
  mutate(duree_moyenne_entretien = mean(duree_entretien)) %>%
  ungroup()
```

- Renommage de toutes les variables de l'ensemble de données en ajoutant le préfixe "endline_" à l'aide d'une boucle :

```r
copie = feuille1
feuille1 <- feuille1 %>%
  rename_all(~paste0("endline_", .))
```

3.2 Analyse et visualisation des données

- Création d'un tableau récapitulatif contenant l'âge moyen et le nombre moyen d'enfants par district :

```r
feuille1 %>%
  tbl_summary(
    by = endline_district,
    include = c("endline_age", "endline_children_num"),
    type = list(endline_children_num = "continuous"),
    statistic = all_continuous() ~ "{mean}",
    label=list( endline_age~ "Age moyen",
                endline_children_num ~ "Nombre moyen d'enfants")
    ) %>%
  add_overall() %>%
  modify_header(label ~ "") %>%
  modify_spanning_header(all_stat_cols() ~ "**District**") %>%
  bold_labels() %>%
  as_flex_table() %>%
  fontsize(size = 10) %>%
  width(width = 0.6)
```

	District								
	Overall, N = 97[1]	**1**, N = 8[1]	**2**, N = 27[1]	**3**, N = 8[1]	**4**, N = 5[1]	**5**, N = 6[1]	**6**, N = 26[1]	**7**, N = 6[1]	**8**, N = 11[1]
Age moyen	26	30	27	26	26	24	23	28	25
Nombre moyen d'enfants	0.58	1.50	0.85	0.00	0.00	0.50	0.12	0.17	1.27

[1]Mean

- Testons si la différence d'âge entre les sexes est statistiquement significative au niveau de 5 %:

```
### différence d'âge entre les sexe
diff_age <- feuille1 %>% tbl_summary(
include = c(endline_age),
by = endline_sex,
statistic = ~"{mean}",
label = list(endline_age ~ "Age du répondant"),
digits = ~ 1
)%>% bold_labels() %>% italicize_levels() %>%
add_difference() %>%
modify_header(
  list(
    all_stat_cols() ~ "{level} \n {n}, ({style_percent(p)}%)"
    )
) %>%
  as_flex_table()
diff_age
```

Characteristic	0 86, (89%)[1]	1 11, (11%)[1]	**Difference**[2]	**95% CI**[23]	**p-value**[2]
Age du répondant	26.0	22.2	3.8	0.23, 7.4	0.039

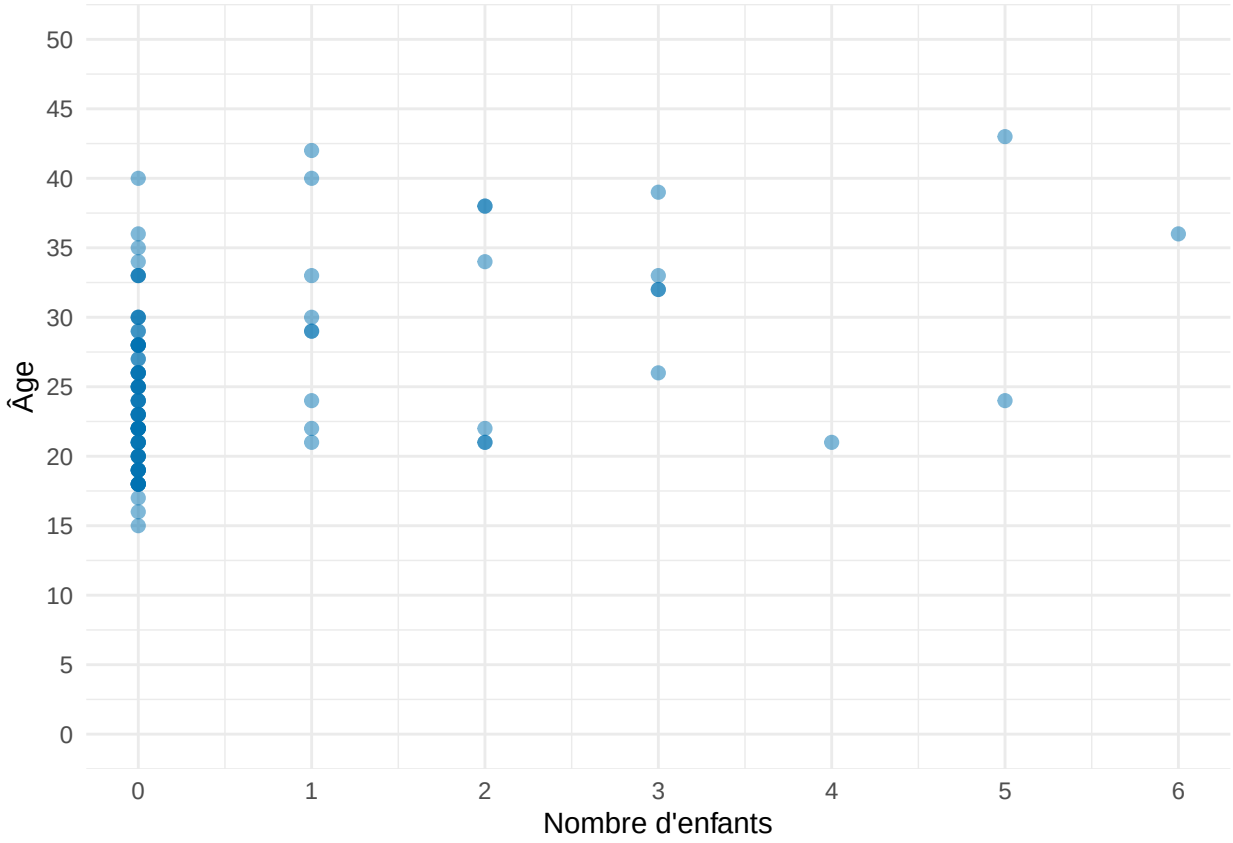
[1]Mean

[2]Welch Two Sample t-test

[3]CI = Confidence Interval

- Création d'un nuage de points de l'âge en fonction du nombre d'enfants:

```
ggplot(feuille1, aes(x = endline_children_num, y = endline_age))+
  geom_point(size = 2, alpha = 0.5, color = "#0072B2") +
  labs(x = "Nombre d'enfants", y = "Âge") +
  theme_minimal() +
  scale_x_continuous(limits = c(0, 6) ,breaks = seq(0, 6, 1)) +
  scale_y_continuous(limits = c(0, 50), breaks = seq(0, 50, 5))
```



- Estimation de l'effet de l'appartenance au groupe de traitement sur l'intention de migrer:

```
library(nnet)
regm <- multinom(intention ~ groupe ,data = copie)
```

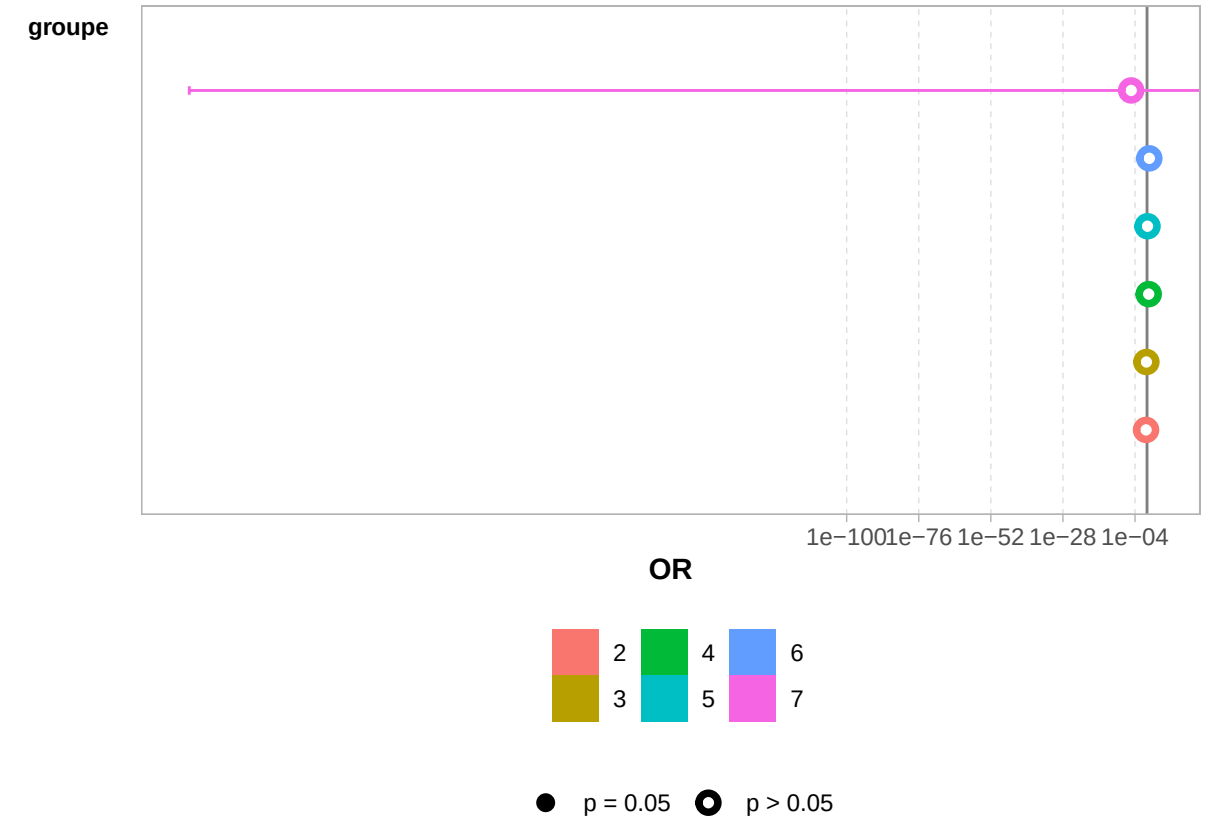
```
## # weights:  21 (12 variable)
## initial  value 188.753284
## iter  10 value 116.109117
## iter  20 value 115.901772
## final  value 115.901310
## converged
```

```
tbl <- tbl_regression(regm, exponentiate = TRUE)
tbl
```

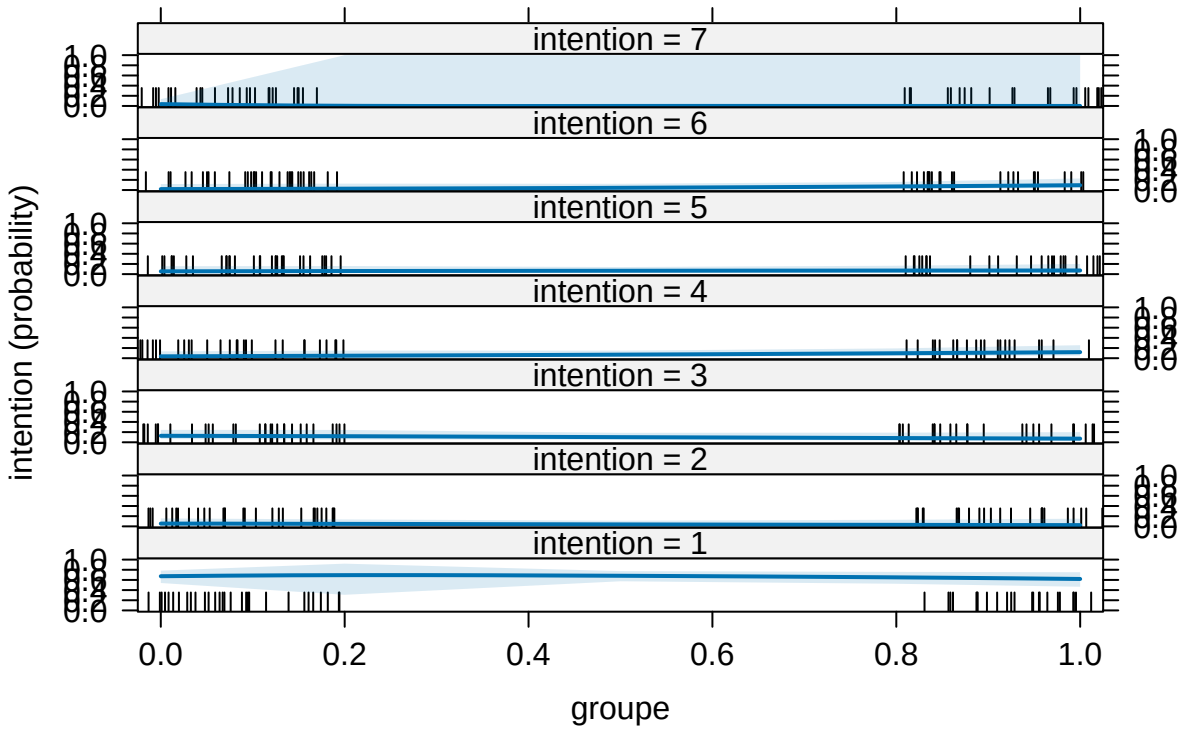
Outcome	Characteristic	OR	95% CI	p-value
2	groupe	0.47	0.05, 4.82	0.5
3	groupe	0.61	0.14, 2.58	0.5
4	groupe	3.56	0.64, 19.8	0.15

Outcome	Characteristic	OR	95% CI	p-value
5	groupe	1.42	0.27, 7.61	0.7
6	groupe	5.69	0.60, 53.9	0.13
7	groupe	0.00	0.00, Inf	>0.9

```
library(GGally)
ggcoef_multinom(
  regm,
  exponentiate = TRUE)
```




```
library(effects)
plot(allEffects(regm))
```



- Création d'un tableau de régression avec 3 modèles:

```
# Modèle A : Modèle vide - Effet du traitement sur les intentions
modele_A <- multinom(intention ~ groupe, data = copie)
```

```
## # weights:  21 (12 variable)
## initial  value 188.753284
## iter  10 value 116.109117
## iter  20 value 115.901772
## final  value 115.901310
## converged
```

```
tableau_A <- tbl_regression(modele_A)


# Modèle B : Effet du traitement sur les intentions en tenant compte de l'âge et du sexe
modele_B <- multinom(intention ~ groupe + age + sex, data = copie)
```

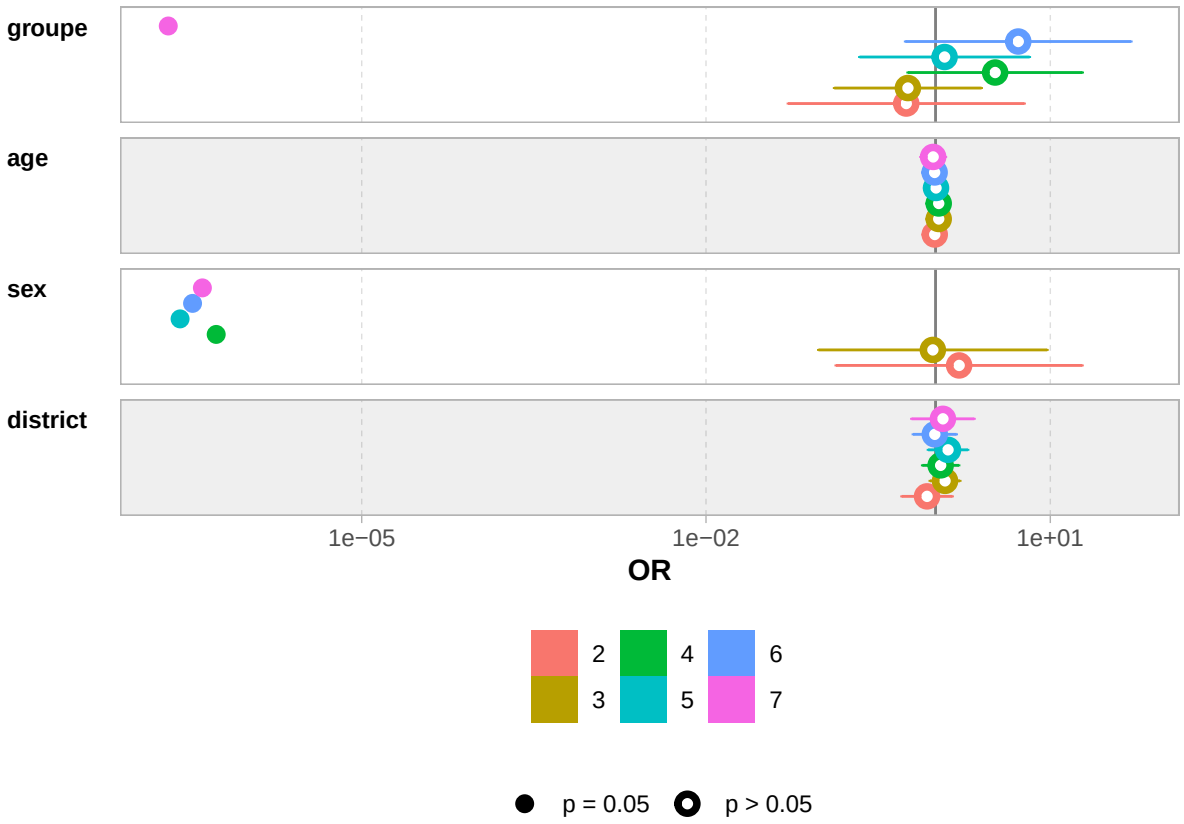
```
## # weights:  35 (24 variable)
## initial  value 188.753284
## iter  10 value 114.260695
```

```
## iter  20 value 112.317802
## iter  30 value 112.241966
## iter  40 value 112.240901
## final  value 112.240885
## converged
```

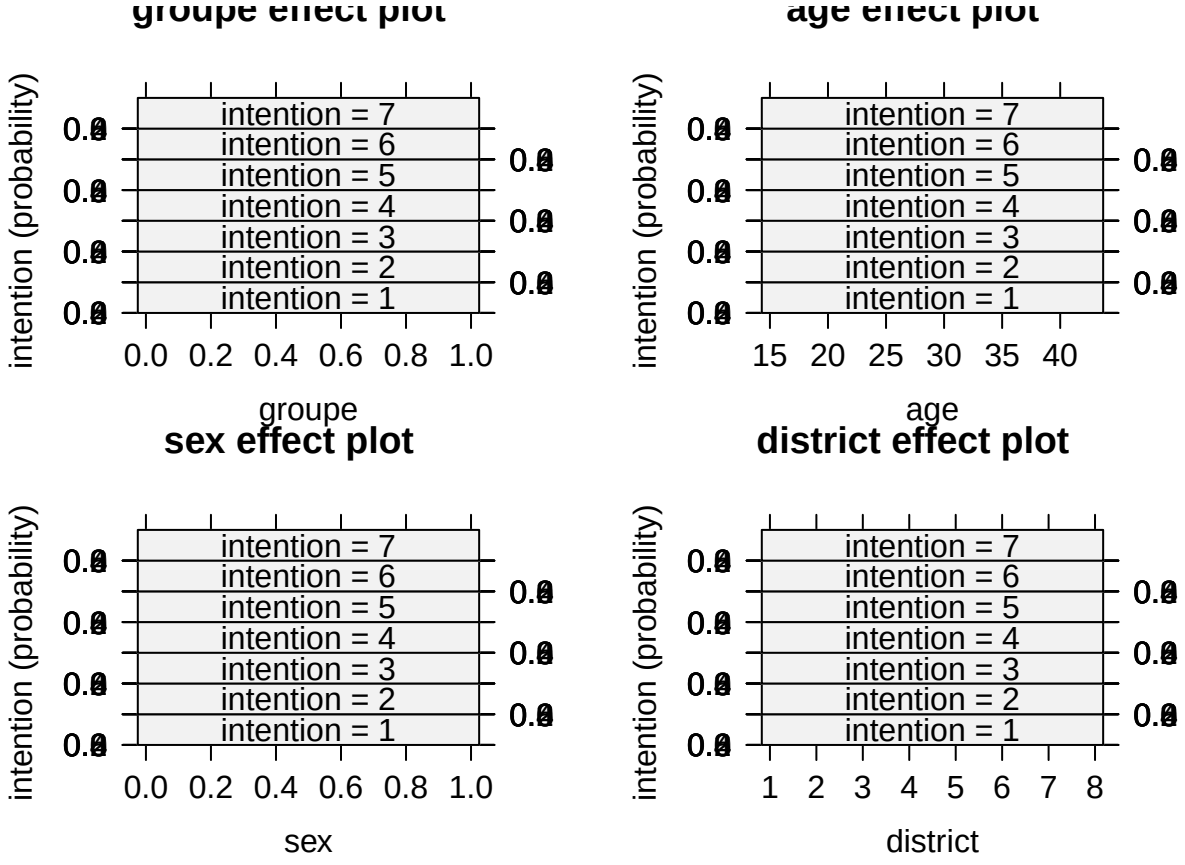
```
tableau_B <- tbl_regression(modele_B)


# Modèle C : Identique au modèle B mais en contrôlant le district
modele_C <- multinom(intention ~ groupe + age + sex + district, data = copie)
```

```
## # weights:  42 (30 variable)
## initial  value 188.753284
## iter  10 value 123.069300
## iter  20 value 110.582195
## iter  30 value 110.364853
## iter  40 value 110.317891
## iter  50 value 110.316861
## final  value 110.316851
## converged
```

```
ggcoef_multinom(
  modele_C,
  exponentiate = TRUE)
```




```
plot(allEffects(modele_C))
```




```
tableau_C <- tbl_regression(modele_C)


# Création du tableau récapitulatif des résultats des trois modèles
tableau_final <- tbl_stack(
  list(tableau_A, tableau_B, tableau_C),
  group_header = c("Modèle A","Modèle B","Modèle C"))%>%
  as_flex_table()


# Affichage du tableau final
tableau_final
```

Group	Characteristic	log(OR)[1]	95% CI[1]	p-value
Modèle A	groupe	-0.75	-3.1, 1.6	0.5
	groupe	-0.49	-1.9, 0.95	0.5
	groupe	1.3	-0.45, 3.0	0.15

[1]OR = Odds Ratio, CI = Confidence Interval

Group	Characteristic	log(OR)[1]	95% CI[1]	p-value
	groupe	0.35	-1.3, 2.0	0.7
	groupe	1.7	-0.51, 4.0	0.13
	groupe	-12	-734, 710	>0.9
Modèle B	groupe	-0.69	-3.0, 1.6	0.6
	age	0.00	-0.16, 0.17	>0.9
	sex	0.61	-1.8, 3.1	0.6
	groupe	-0.47	-1.9, 1.0	0.5
	age	0.05	-0.05, 0.15	0.4
	sex	-0.28	-2.5, 2.0	0.8
	groupe	1.2	-0.50, 3.0	0.2
	age	0.06	-0.07, 0.18	0.4
	sex	-14	-14, -14	<0.001
	groupe	0.27	-1.4, 2.0	0.8
	age	-0.01	-0.15, 0.13	0.9
	sex	-16	-16, -16	<0.001
	groupe	1.7	-0.60, 3.9	0.2
	age	-0.02	-0.19, 0.15	0.8
	sex	-14	-14, -14	<0.001
	groupe	-14	-14, -14	<0.001
	age	-0.06	-0.31, 0.19	0.6
	sex	-15	-15, -15	<0.001
Modèle C	groupe	-0.59	-3.0, 1.8	0.6

[1]OR = Odds Ratio, CI = Confidence Interval

Group	Characteristic	log(OR)[1]	95% CI[1]	p-value
	age	-0.02	-0.19, 0.16	0.9
	sex	0.47	-2.0, 2.9	0.7
	district	-0.17	-0.68, 0.34	0.5
	groupe	-0.55	-2.0, 0.92	0.5
	age	0.06	-0.04, 0.17	0.2
	sex	-0.05	-2.3, 2.2	>0.9
	district	0.19	-0.11, 0.49	0.2
	groupe	1.2	-0.56, 2.9	0.2
	age	0.06	-0.06, 0.19	0.3
	sex	-14	-14, -14	<0.001
	district	0.10	-0.27, 0.47	0.6
	groupe	0.18	-1.5, 1.9	0.8
	age	0.01	-0.13, 0.15	0.9
	sex	-15	-15, -15	<0.001
	district	0.25	-0.15, 0.65	0.2
	groupe	1.7	-0.61, 3.9	0.2
	age	-0.02	-0.19, 0.15	0.8
	sex	-15	-15, -15	<0.001
	district	-0.02	-0.45, 0.42	>0.9
	groupe	-15	-15, -15	<0.001
	age	-0.05	-0.30, 0.20	0.7
	sex	-15	-15, -15	<0.001

[1]OR = Odds Ratio, CI = Confidence Interval

Group	Characteristic	log(OR)[1]	95% CI[1]	p-value
	district	0.15	-0.48, 0.78	0.6

[1]OR = Odds Ratio, CI = Confidence Interval

4 Partie 3: