

**AGENCE NATIONALE DE LA STATISTIQUE ET DE LA
DEMOGRAPHIE**



**ECOLE NATIONALE DE LA STATISTIQUE ET DE L'ANALYSE
ECONOMIQUE PIERRE NDIAYE**

TP 1 - Logiciel statistique R

Par

KPAKOU M'Mounéné ISEP3

Chargé du cours:

Mr. Aboubacar HEMA

Ingénieur des Travaux Statistiques

0.Libraries nécessaires

```
rm(list = ls())
```

```
library(tidyverse)
library(readxl)
library(sjPlot)#création de graphiques descriptifs
library(dplyr)
library(gtsummary)
library(kableExtra)
library(sf) # objet spatial
library(ggspatial) #representation spatial
```

1.Préparation des données

1.2 Importation et mise en forme

a.Repertoire de travail

```
setwd(getwd())
```

b.Importation

```
projet <- read_xlsx("Base_Projet.xlsx")
#View(projet)
```

```
dim(projet)
```

Nombre de lignes et colonnes

```
[1] 250 33
```

La base de données contient 250 Petites et Moyennes entreprises et 33 variables

```
sum(is.na(projet$key))
```

Verification de l'existence de NA pour la variable key

```
[1] 0
```

Il n'y a pas de valeurs manquantes pour la variable key c'est à dire qu'il n'y a pas d'entreprises non identifiables à travers la variable key

1.3 Création de variables

a.Renames des variables

```
projet <- projet %>% rename(!!!setNames(c("q1","q2","q23"),  
                                         c("region","departement","sexe")))
```

b.Création de la variable sexe_2

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

c.Création d'un dataframe langues

```
#selection des variables commençant par q24a_  
langues_variables = projet %>%  
  dplyr::select(gtsummary::starts_with("q24a_")) %>% names()  
#création du dataframe  
langues <- projet %>%  
  select(key, all_of(langues_variables))
```

d. creation de la variable

```
# nouvelle variable parle qui représente la somme des langues parlées
langues <- langues %>%
  mutate(parle = rowSums(select(., starts_with("q24a_"))))
```

e. selection des variables key et parle

```
# Sélection des variables key et parle dans le data.frame langues
langues <- select(langues, key, parle)

# Fusion des data.frames projet et langues
merge_data <- merge(projet, langues, by = "key")
```

2.Analyses descriptives

Fonction pour les statistiques descriptives univariées

```
# Fonction pour les statistiques descriptives univariées
univarie <- function(data, variable) {
  if (is.numeric(var)) {
    summary(data[[variable]])#statistiques
    hist(var, main = paste("Histogramme de", variable))## Histogramme
  } else if (is.factor(data[[variable]])) {
    data %>% tbl_summary(variable) %>% as_gt() # Statistiques descriptives
    data %>% plot_frq(variable,coord.flip =T,show.na = T)
  } else if (is.character(data[[variable]])) {
    data %>% tbl_summary(include=variable)%>%as_gt()# Statistiques
    data %>%plot_frq(variable,coord.flip =T,
      title=paste("Répartition des PME selon", variable),show.na = T)
  }
}
```

Fonction pour les statistiques descriptive bivariée

```

bivariee <- function(data, x, y) {
  # Déterminer les types des variables
  type_x <- class(data[[x]])
  type_y <- class(data[[y]])
  # Cas Quantitatif & Quantitatif
  if (type_x == "numeric" && type_y == "numeric") {
    summary(data[[x]] ~ data[[y]]) # Statistiques descriptives
    plot(data[[x]], data[[y]], main = paste(x, "vs", y)) # Graphique
    cor.test(data[[x]], data[[y]])
  }
  # Cas Quantitatif & Qualitatif
  if (type_x == "numeric" && type_y == "character" || type_y == "factor") {
    tapply(data[[x]], data[[y]], summary) # Statistiques descriptives
    boxplot(data[[x]] ~ data[[y]]) # Graphique
  }
  # Cas Qualitatif & Qualitatif
  if (type_x == "character" || type_x == "factor" && type_y == "character" || type_y == "factor") {
    tbl <- tbl_summary(data, include = c(x,y), by=x) # Statistiques
    tbl_pdf <- tbl %>% as_kable_extra()
  }
}

```

Nous allons renommer certains variables pour que leurs noms soit plus explicites

```

merge_data <-
  merge_data %>%
  rename(!!!setNames(c("q12", "q25", "q81", "filieres_1", "filieres_2",
    "filieres_3", "filieres_4"), c("Statut_juridique",
    "Niveau_instruction", "Proprietaire"

```

2.1 Statistiques descriptives univariées

La répartition des PME suivant:

a. sexe

```
univarie(data = merge_data, variable = "sexe")
```

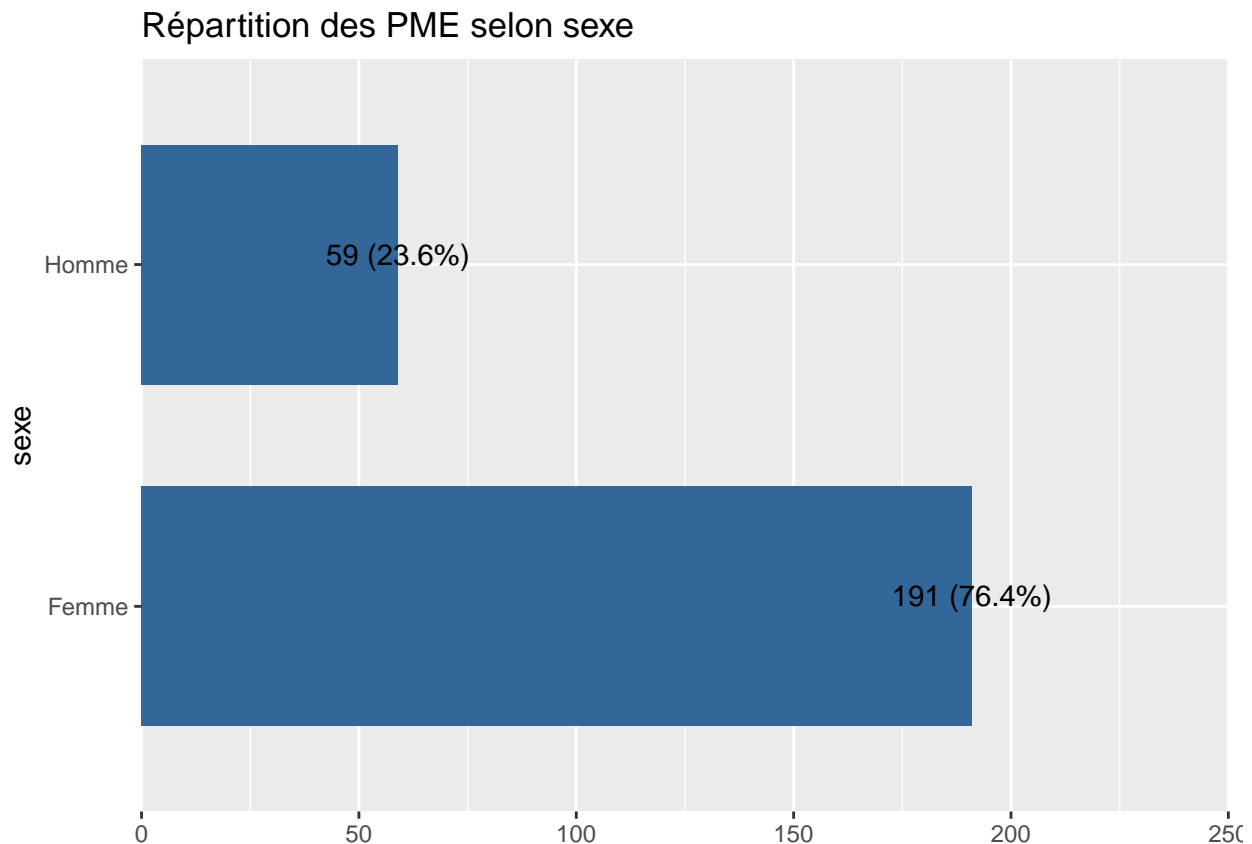


Table 1: répartition du sexe

Characteristic	N = 250 ¹
sexe	
Femme	191 (76%)
Homme	59 (24%)

¹n (%)

Sur un total de 250 responsables, 191 sont des femmes, représentant ainsi 76 % du total, tandis que 59 sont des hommes, soit 24 % du total. Cette répartition dénote une prépondérance

marquée des femmes parmi les responsables ou dirigeants de PME inclus dans notre base.

b. niveau d'instruction

```
univarie(data = merge_data, variable = "Niveau_instruction")
```

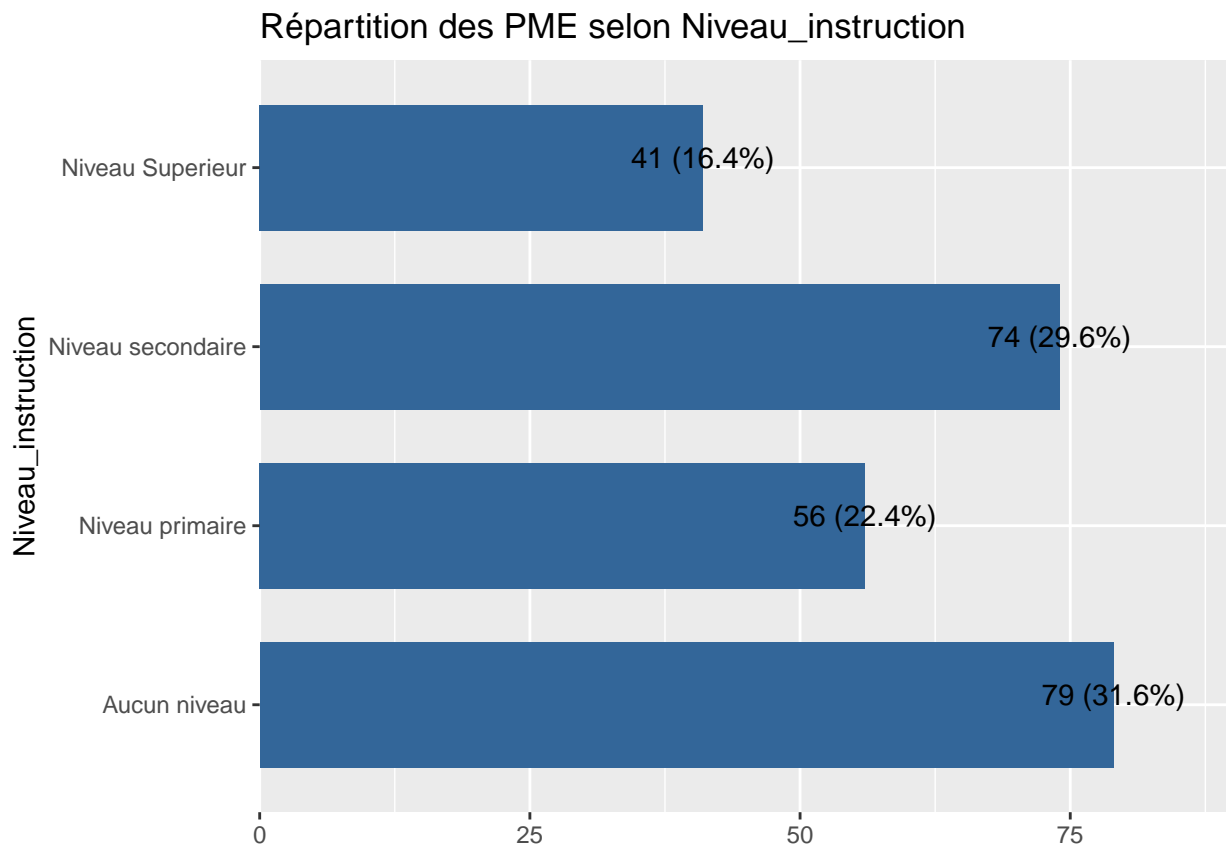


Table 2: répartition du Niveau d'instruction

Characteristic	N = 250 ¹
Niveau_instruction	
Aucun niveau	79 (32%)
Niveau primaire	56 (22%)
Niveau secondaire	74 (30%)
Niveau Supérieur	41 (16%)

¹n (%)

Le tableau fourni résume la répartition du niveau d'instruction des responsables ou dirigeants

des PME. Sur un total de 250 responsables, 79 d'entre eux (soit 32 %) n'ont aucun niveau d'instruction spécifié, 56 (22 %) ont un niveau primaire, 74 (30 %) ont un niveau secondaire, et 41 (16 %) ont un niveau supérieur. Cette répartition souligne une diversité de niveaux d'instruction parmi les responsables ou dirigeants de PME inclus dans l'échantillon.

c. statut juridique

```
univarie(data = merge_data, variable = "Statut_juridique")
```

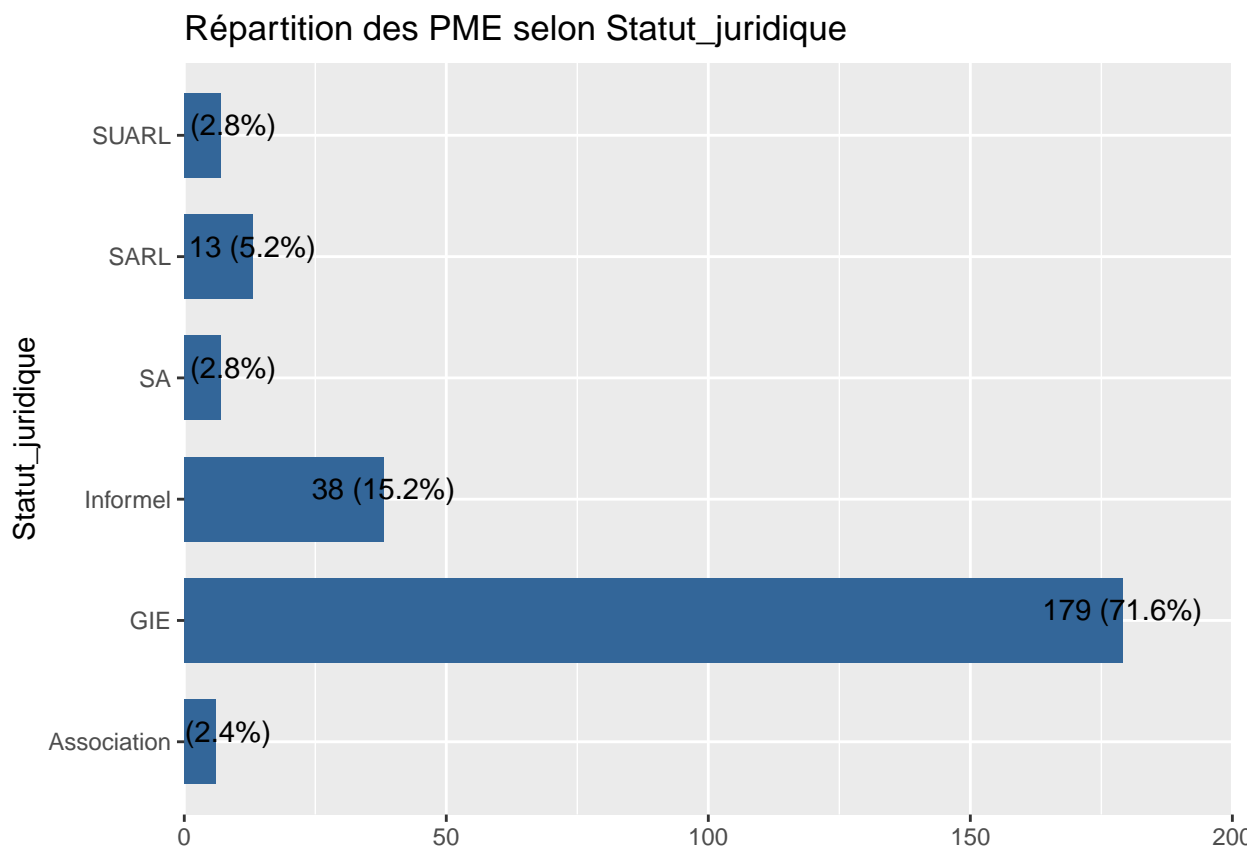


Table 3: répartition du Statut_juridique

Characteristic	N = 250 ¹
Statut_juridique	
Association	6 (2.4%)
GIE	179 (72%)
Informel	38 (15%)
SA	7 (2.8%)
SARL	13 (5.2%)
SUARL	7 (2.8%)

¹n (%)

Sur un total de 250 responsables, il apparaît que les Groupements d'Intérêt Économique (GIE) représentent la catégorie prédominante avec 179 responsables (72 %), suivis par les structures informelles comptant 38 responsables (15 %). Les autres statuts juridiques tels que les Sociétés Anonymes (SA), les Sociétés à Responsabilité Limitée (SARL), et les Sociétés Unipersonnelles à Responsabilité Limitée (SUARL) sont moins fréquents, représentant respectivement 2,8 %, 5,2 %, et 2,8 % des responsables. Seules six personnes (2,4 %) sont associées dans des associations.

```
### d.proprietaire/locataire
```

```
univarie(data = merge_data, variable = "Proprietaire")
```

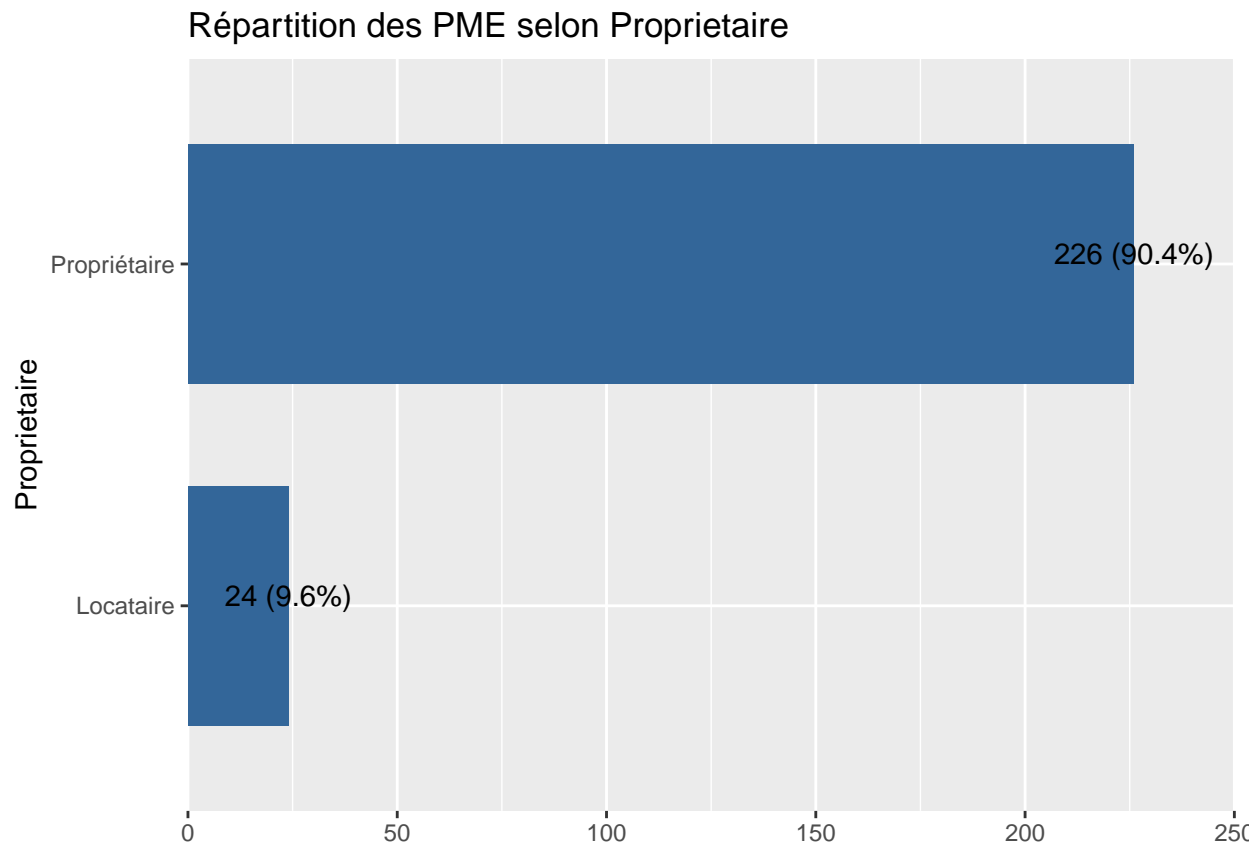


Table 4: répartition du Proprietaire/locataire

Characteristic	N = 250 ¹
Proprietaire	
Locataire	24 (9.6%)
Proprietaire	226 (90%)

¹n (%)

Le tableau fourni résume la répartition du statut de locataire ou de propriétaire parmi les responsables ou dirigeants des PME. Sur un total de 250 responsables, il apparaît que la grande majorité, soit 226 responsables (90 %), sont propriétaires, tandis que seulement 24 responsables (9,6 %) sont locataires. Cette répartition souligne une prévalence significative des propriétaires parmi les responsables de PME dans l'échantillon étudié.

2.2 Statistiques descriptives bivariées

a. Le statut juridique et le sexe

```
bivariee(merge_data, "sexe", "Statut_juridique")
```

Table 5: répartition selon le statut juridique et le sexe

Characteristic	Femme, N = 191 ¹	Homme, N = 59 ¹
Statut_juridique		
Association	3 (1.6%)	3 (5.1%)
GIE	149 (78%)	30 (51%)
Informel	32 (17%)	6 (10%)
SA	1 (0.5%)	6 (10%)
SARL	2 (1.0%)	11 (19%)
SUARL	4 (2.1%)	3 (5.1%)

¹n (%)

La répartition des PME selon le statut juridique et le sexe révèle des tendances significatives. Parmi les 191 PME dirigées par des femmes, la majorité opèrent sous forme de Groupements d'Intérêt Économique (GIE), représentant 78 % de cet échantillon, suivies par les structures informelles avec 17 %. Les autres statuts juridiques tels que les Associations, les SARL et les SUARL sont moins fréquents, chacun représentant moins de 3 % des PME dirigées par des femmes. En revanche, parmi les 59 PME dirigées par des hommes, la répartition est plus diversifiée : 51 % sont des GIE, 10 % sont des structures informelles ou des SA, tandis que les SARL et les SUARL représentent respectivement 19 % et 5 %.

b. Le niveau d'instruction et le sexe

```
bivariee(merge_data, "sexe", "Niveau_instruction")
```

Table 6: répartition selon le niveau d'instruction et le sexe

Characteristic	Femme, N = 191 ¹	Homme, N = 59 ¹
Niveau_instruction		
Aucun niveau	70 (37%)	9 (15%)
Niveau primaire	48 (25%)	8 (14%)
Niveau secondaire	56 (29%)	18 (31%)
Niveau Supérieur	17 (8.9%)	24 (41%)

¹n (%)

Parmi les 191 PME dirigées par des femmes, la majorité des responsables n'ont aucun niveau d'instruction spécifié, représentant 37 % de l'échantillon, suivis par ceux ayant un niveau secondaire (29 %). Les niveaux primaire et supérieur sont également représentés, avec respectivement 25 % et 8,9 % des responsables. En revanche, parmi les 59 PME dirigées par des hommes, une tendance différente se dessine : une plus grande proportion de responsables possèdent un niveau d'instruction supérieur (41 %), suivie par ceux ayant un niveau secondaire (31 %). Les proportions de responsables sans niveau d'instruction ou avec un niveau primaire sont plus faibles chez les hommes, représentant respectivement 15 % et 14 %. Ces différences dans la répartition selon le niveau d'instruction entre les sexes suggèrent des dynamiques distinctes en matière d'accès à l'éducation et d'entrepreneuriat,

c. Propriétaire/locataire suivant le sexe

```
bivariee(merge_data, "sexe", "Proprietaire")
```

Table 7: répartition Propriétaire/locataire suivant le sexe

Characteristic	Femme, N = 191 ¹	Homme, N = 59 ¹
Proprietaire		
Locataire	16 (8.4%)	8 (14%)
Proprietaire	175 (92%)	51 (86%)

¹n (%)

L'analyse de la répartition des PME selon le statut de locataire ou de propriétaire des responsables révèle des différences notables entre les sexes. Parmi les 191 PME dirigées par des femmes, la grande majorité des responsables sont propriétaires de leur local, représentant 92 % de l'échantillon, tandis que seulement 8,4 % sont des locataires. En revanche, parmi les 59 PME dirigées par des hommes, bien que la majorité soit également propriétaire (86 %), une proportion légèrement plus élevée de responsables (14 %) sont des locataires.

2.3 Autres analyses**a. Nombre d'entreprises par filière**

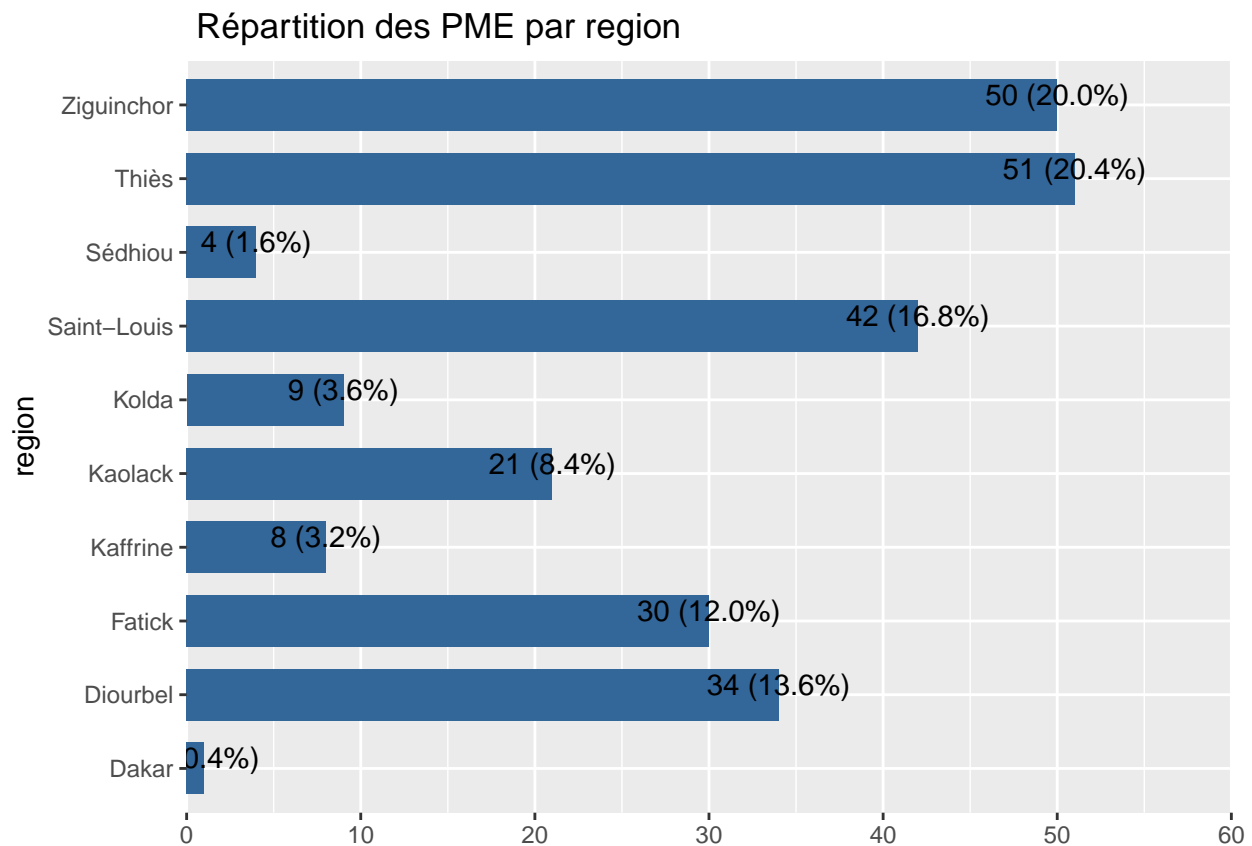
```
effectifs_filiere <- colSums(merge_data[, c("Arachide", "Anacarde", "Mangue", "Riz")])
effectifs_filiere
```

Arachide	Anacarde	Mangue	Riz
108	61	89	92

Nous avons 108 entreprises PME sont recensées dans la filière *arachide*, 61 dans la filière *anacarde*, 89 dans la filière *mangue*, et 92 dans la filière *Riz*.

b. Répartition des PME par region

```
merge_data %>% plot_frq(region,coord.flip =T,title=" Répartition des PME par region",sho
```



Nous observons que 20,4% des entreprises enquêtés sont dans la region de thiès suivi de la région de Ziguichor avec 20%.

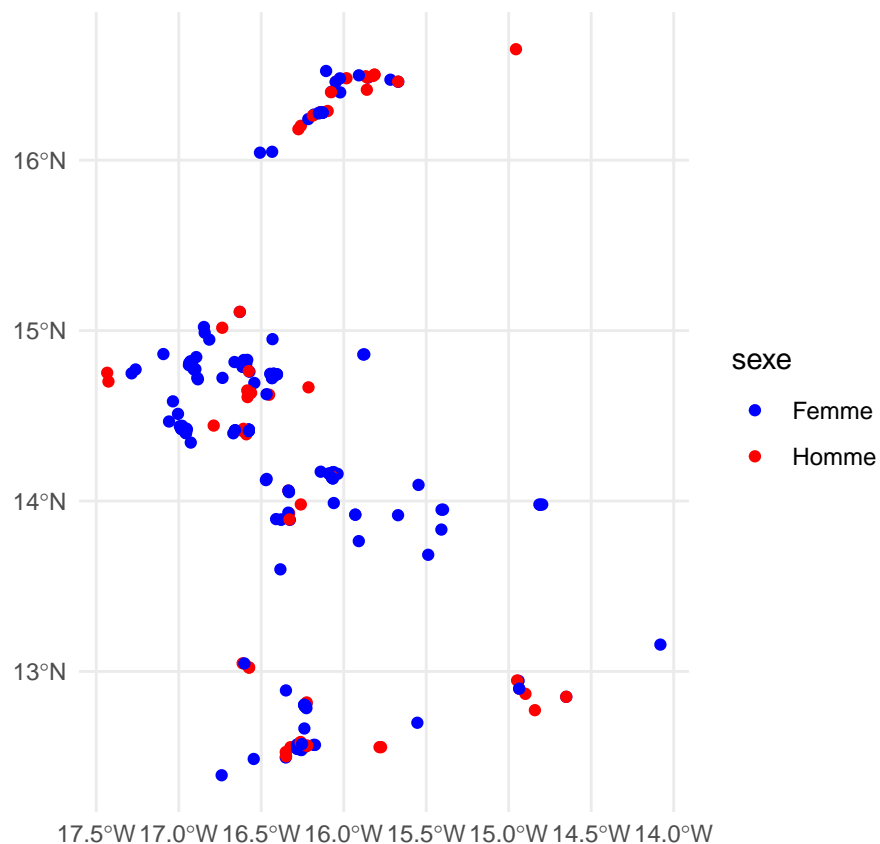
3.Cartographie

Transformons d'abord nos données en données spatiales

```
projet_map <- st_as_sf(merge_data, coords = c("gps_menlongitude", "gps_menlatitude"), crs = "EPSG:4326")
```

a.Code ayant permet de faire sortir la carte ci dessous

```
palette_couleurs <- c("blue", "red")
# Creation de la carte
map <- ggplot() +
  geom_sf(data = projet_map, aes(color = sexe)) + # "sexe" pour la couleur
  scale_color_manual(values = palette_couleurs) +
  theme_minimal()
# Affichage la carte
map
```

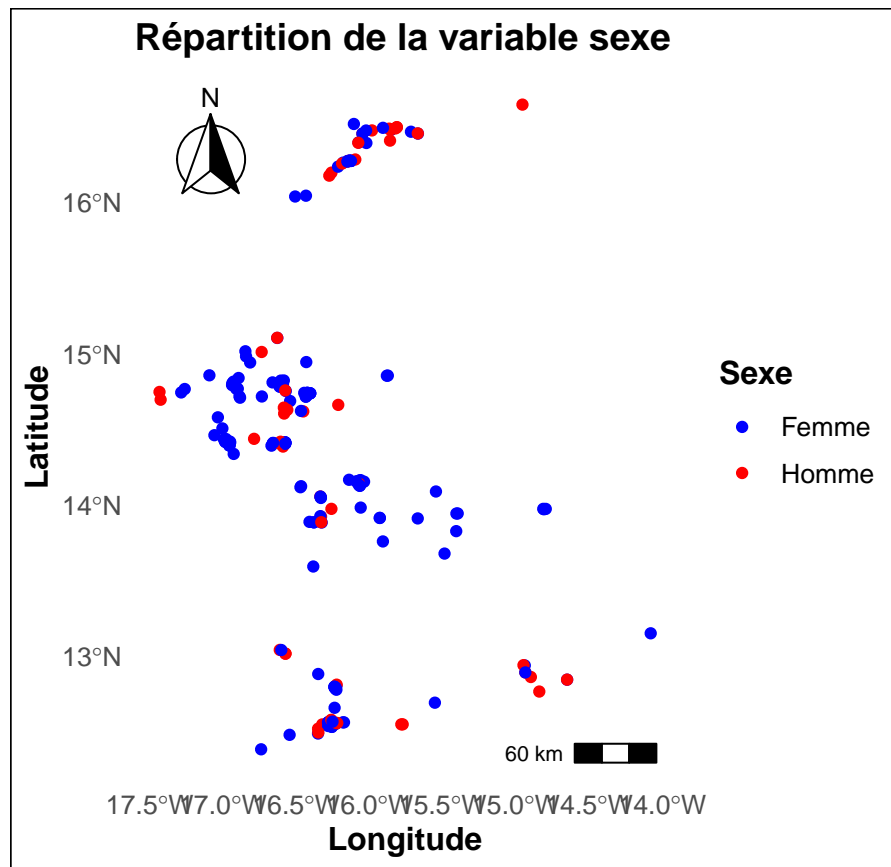


b. Carte avec echelle, titre des axes, legende

```

map <- ggplot() +
  geom_sf(data = projet_map, aes(color = sexe)) +
  scale_color_manual(values = palette_couleurs) +
  labs(title = "Répartition de la variable sexe", # Titre de la carte
        x = "Longitude", y = "Latitude", # Axes
        color = "Sexe") +
  theme_minimal() + # Thème minimal
  theme(legend.position = "right", # Position de la légende
        axis.text = element_text(size = 10), #Taille du texte des axes
        axis.title = element_text(size = 12, face = "bold"),# Style du texte des axes
        plot.title = element_text(size = 14, face = "bold"),# Style du texte du titre
        legend.text = element_text(size = 10), # Taille du texte de la légende
        legend.title = element_text(size = 12, face = "bold"),# la légende
        plot.caption = element_text(size = 10)) + # Légende de l'échelle
  annotation_scale(location = "br", width_hint = 0.2)+
  annotation_north_arrow(location = "tl", which_north = "true",
style = north_arrow_fancy_orienteering)+
  theme(
    plot.background = element_rect(color = "black", fill = NA),# Définir la couleur du c
    panel.grid.major = element_blank(), # Supprimer les lignes de grille principales
    panel.grid.minor = element_blank() # Supprimer les lignes de grille secondaires
  )
# Afficher la carte
map

```

```
# Importation des données géographiques pour chaque niveau administratif du Sénégal
SEN_adm0 <- st_read("gadm41_SEN_shp/gadm41_SEN_0.shp")
```

Reading layer `gadm41_SEN_0' from data source

```
`E:\KPAKOU\KPAM ISEP 3\SEMESTRE 6\Projet sous R\TP 1\TP1\gadm41_SEN_shp\gadm41_SEN_0.s
using driver `ESRI Shapefile'
```

Simple feature collection with 1 feature and 2 fields

Geometry type: MULTIPOLYGON

Dimension: XY

Bounding box: xmin: -17.54319 ymin: 12.30786 xmax: -11.34247 ymax: 16.69207

Geodetic CRS: WGS 84

```
SEN_adm1 <- st_read("gadm41_SEN_shp/gadm41_SEN_1.shp")
```

Reading layer `gadm41_SEN_1' from data source

```
`E:\KPAKOU\KPAM ISEP 3\SEMESTRE 6\Projet sous R\TP 1\TP1\gadm41_SEN_shp\gadm41_SEN_1.s
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

```
SEN_adm2 <- st_read("gadm41_SEN_shp/gadm41_SEN_2.shp")
```

Reading layer `gadm41_SEN_2' from data source

```
`E:\KPAKOU\KPAM ISEP 3\SEMESTRE 6\Projet sous R\TP 1\TP1\gadm41_SEN_shp\gadm41_SEN_2.s  
using driver `ESRI Shapefile'
```

Simple feature collection with 45 features and 13 fields

Geometry type: MULTIPOLYGON

Dimension: XY

Bounding box: xmin: -17.54319 ymin: 12.30786 xmax: -11.34247 ymax: 16.69207

Geodetic CRS: WGS 84

```
SEN_adm3 <- st_read("gadm41_SEN_shp/gadm41_SEN_3.shp")
```

Reading layer `gadm41_SEN_3' from data source

```
`E:\KPAKOU\KPAM ISEP 3\SEMESTRE 6\Projet sous R\TP 1\TP1\gadm41_SEN_shp\gadm41_SEN_3.s  
using driver `ESRI Shapefile'
```

Simple feature collection with 123 features and 16 fields

Geometry type: MULTIPOLYGON

Dimension: XY

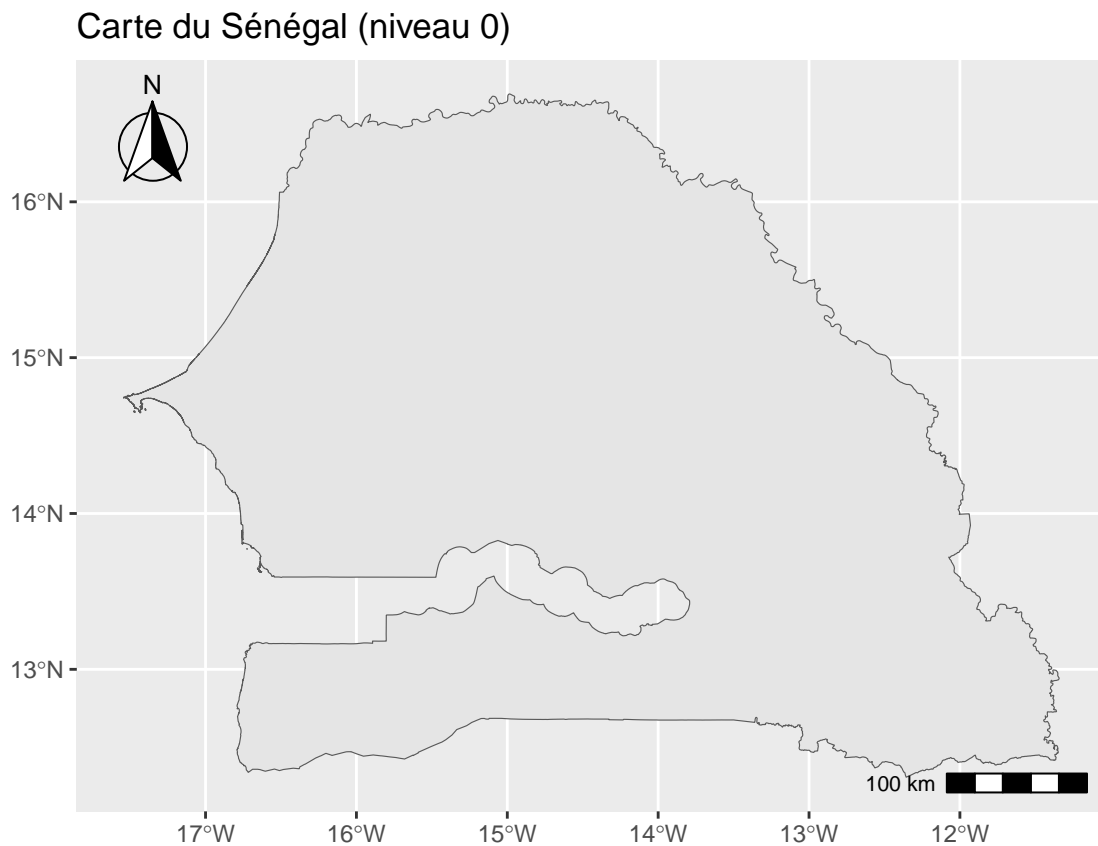
Bounding box: xmin: -17.54319 ymin: 12.30786 xmax: -11.34247 ymax: 16.69207

Geodetic CRS: WGS 84

```
# Affichage des différentes cartes suivant les niveaux administratifs
```

Carte du Sénégal (niveau 0)

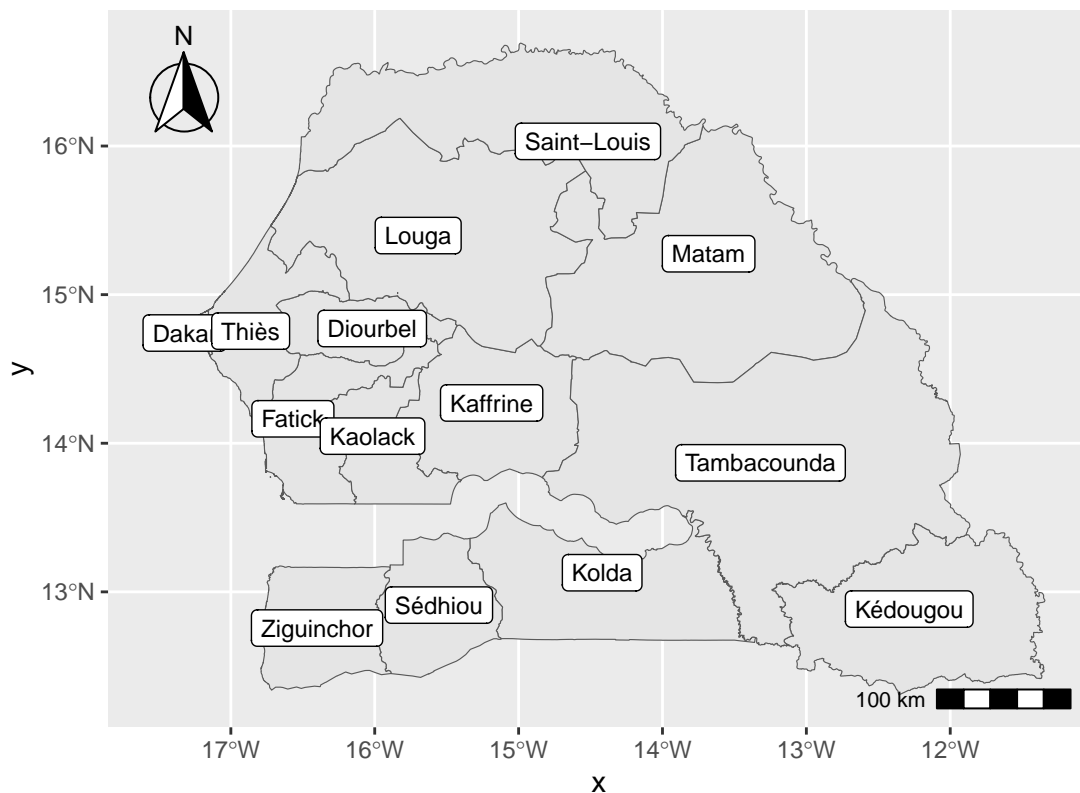
```
ggplot() +  
  geom_sf(data = SEN_adm0) +  
  labs(title = "Carte du Sénégal (niveau 0)") +  
  annotation_scale(location = "br", width_hint = 0.2) +  
  annotation_north_arrow(location = "tl", which_north = "true",  
    style = north_arrow_fancy_orienteering)
```



Carte du Sénégal (niveau 1 - Régions)

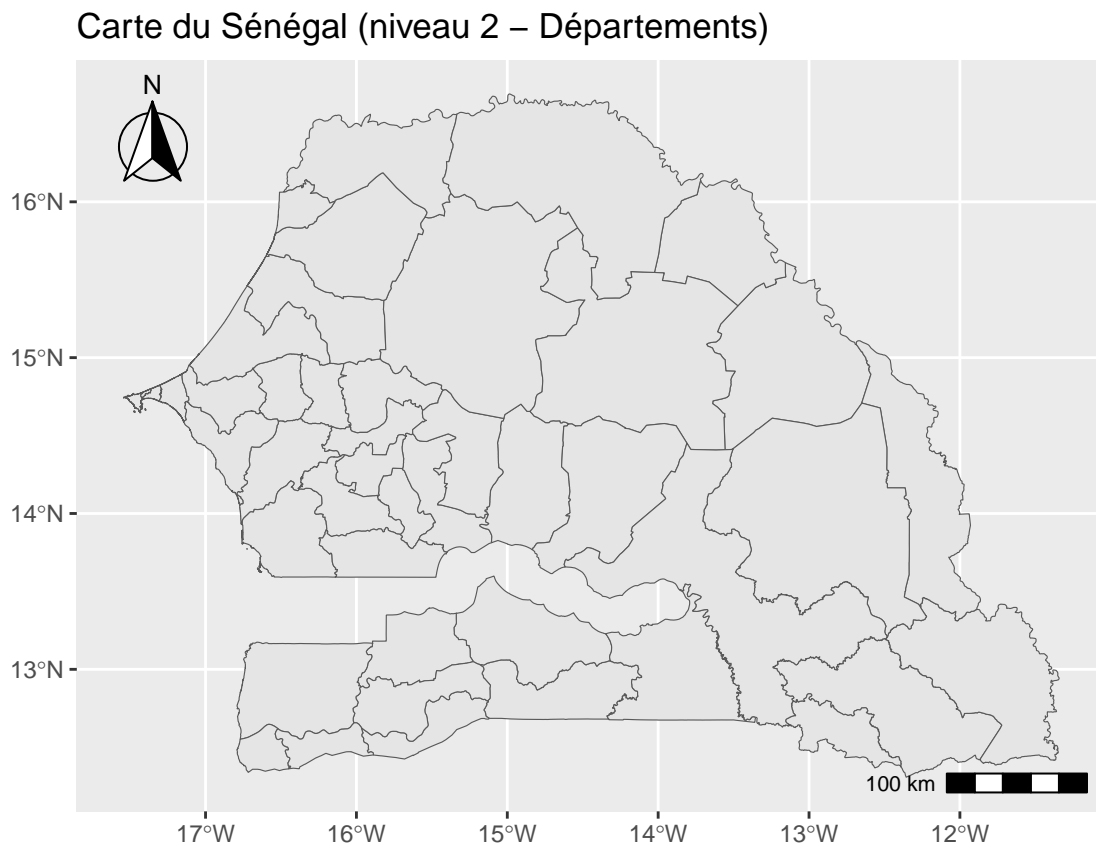
```
ggplot() +
  geom_sf(data = SEN_adm1) +
  labs(title = "Carte du Sénégal (niveau 1 - Régions)") +
  geom_sf_label(data = SEN_adm1, aes(label = NAME_1), color = "black", size = 3, check_o
  annotation_scale(location = "br", width_hint = 0.2) +
  annotation_north_arrow(location = "tl", which_north = "true",
  style = north_arrow_fancy_orienteering)
```

Carte du Sénégal (niveau 1 – Régions)



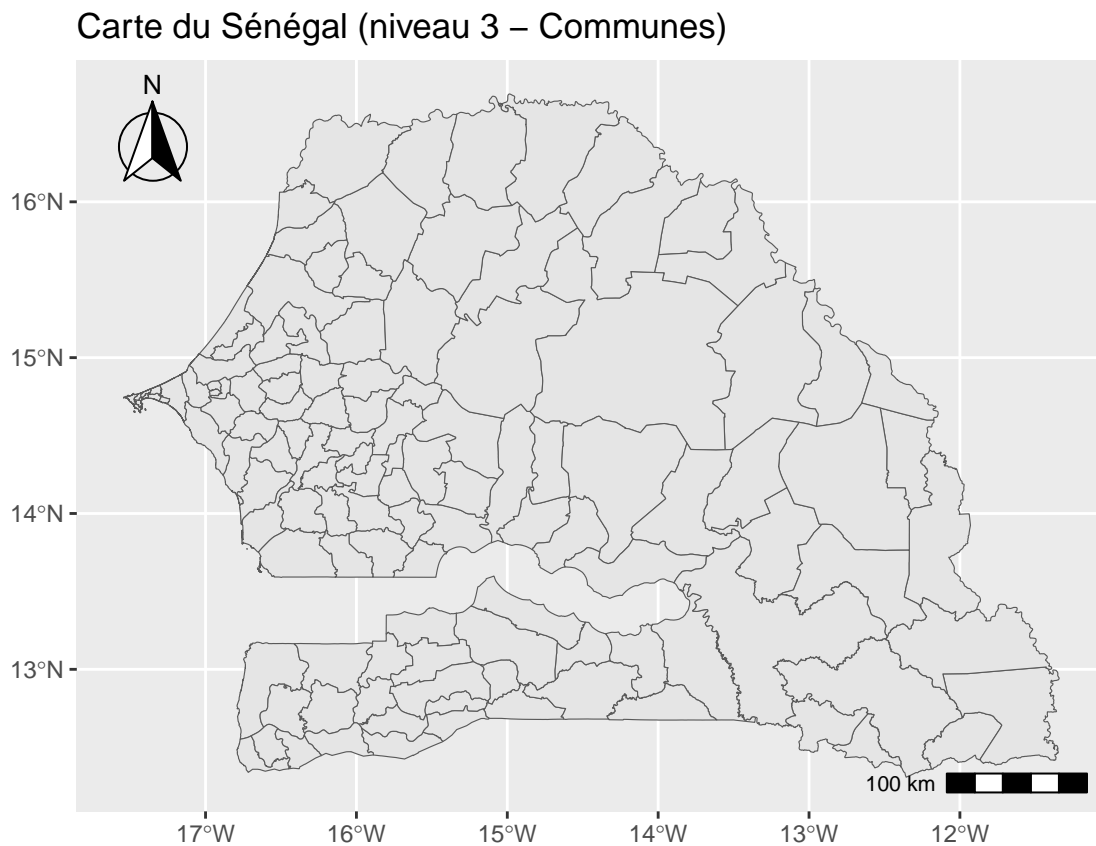
Carte du Sénégal (niveau 2 - Départements)

```
ggplot() +
  geom_sf(data = SEN_adm2) +
  labs(title = "Carte du Sénégal (niveau 2 - Départements)") +
  annotation_scale(location = "br", width_hint = 0.2) +
  annotation_north_arrow(location = "tl", which_north = "true",
    style = north_arrow_fancy_orienteering)
```



Carte du Sénégal (niveau 3 - Communes)

```
ggplot() +  
  geom_sf(data = SEN_adm3) +  
  labs(title = "Carte du Sénégal (niveau 3 - Communes)") +  
  annotation_scale(location = "br", width_hint = 0.2) +  
  annotation_north_arrow(location = "tl", which_north = "true",  
    style = north_arrow_fancy_orienteering)
```

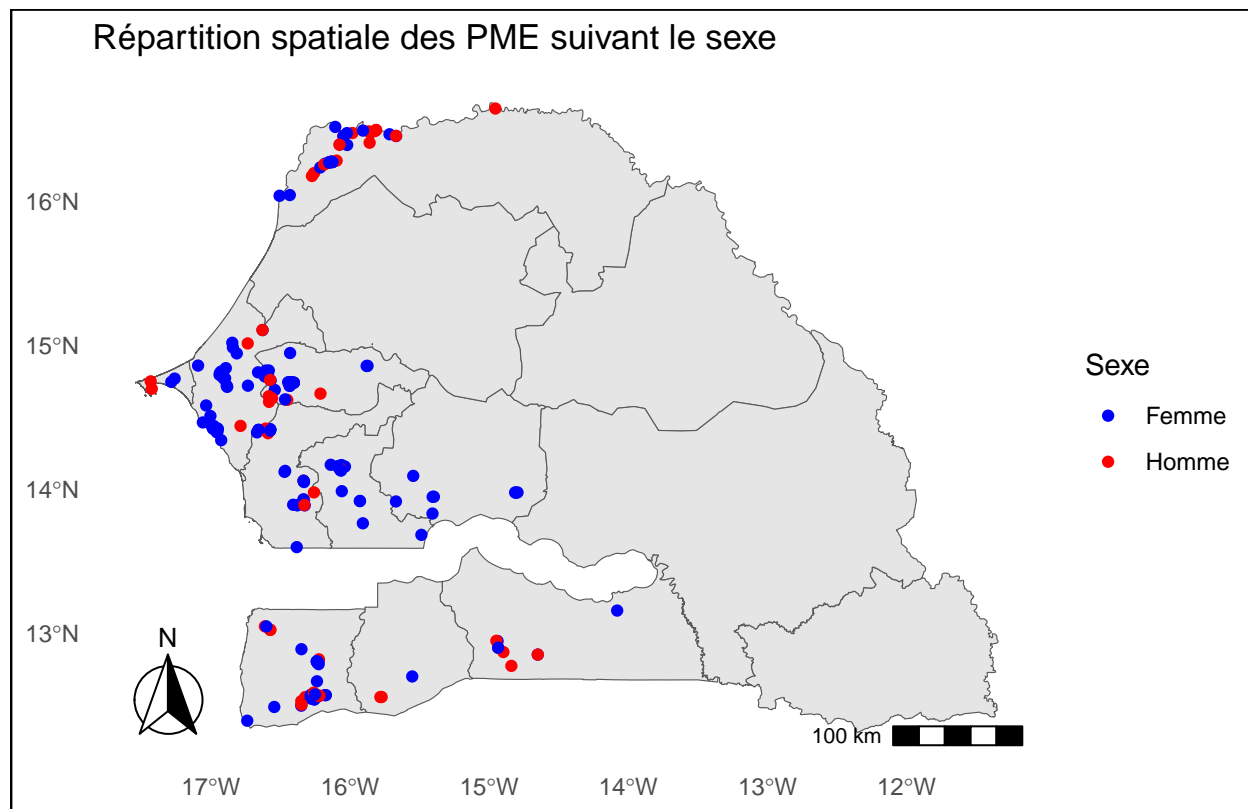


C.une représentation spatiale des PME suivant le sexe à l'échelle regionale

```

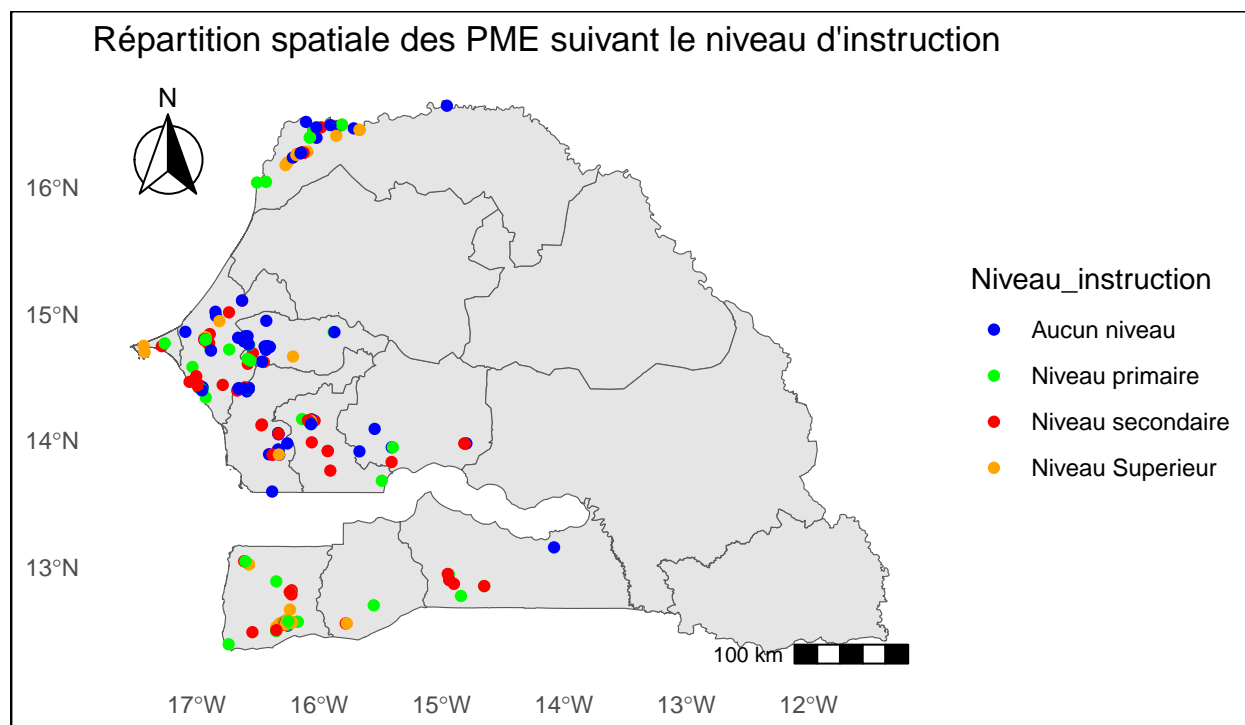
# Nous devons joindre les données de PME avec les données géographiques regionales
pme_joined <- st_join(projet_map, SEN_adm1)
# Affichage des PME sur la carte
ggplot() +
  geom_sf(data = SEN_adm1) + # Affichage des regions
  geom_sf(data = pme_joined, aes(color = sexe)) + # Affichage des PME avec une couleur di
  scale_color_manual(values = c("blue", "red")) + # Définissons les couleurs pour chaque
  labs(title = "Répartition spatiale des PME suivant le sexe", color = "Sexe") + # Titr
  theme_minimal() + #Thème minimal
  annotation_scale(location = "br", width_hint = 0.2) +
  annotation_north_arrow(location = "bl", which_north = "true",
style = north_arrow_fancy_orienteering) + # Echelle
  theme(plot.background = element_rect(color = "black", fill = NA), # Définir la couleur
        panel.grid.major = element_blank(), # Supprimer les lignes de grille principale
        panel.grid.minor = element_blank() # Supprimer les lignes de grille secondaires
  )

```



Réprésentation spatiale des PME suivant le niveau d'instruction à l'échelle regionale

```
# Nous devons joindre les données de PME avec les données géographiques regionales
pme_joined <- st_join(projet_map, SEN_adm1)
# Affichage des PME sur la carte
ggplot() +
  geom_sf(data = SEN_adm1) + # Affichage des regions
  geom_sf(data = pme_joined, aes(color = Niveau_instruction)) + # Affichage des PME avec
  scale_color_manual(values = c("blue", "green", "red", "orange")) + # Définissons les
  labs(title = "Répartition spatiale des PME suivant le niveau d'instruction", color = "Niveau_instruction") +
  theme_minimal() + # Thème minimal
  annotation_scale(location = "br", width_hint = 0.2) + # Echelle
  theme(plot.background = element_rect(color = "black", fill = NA), # Définir la couleur de fond
        panel.grid.major = element_blank(), # Supprimer les lignes de grille principale
        panel.grid.minor = element_blank() # Supprimer les lignes de grille secondaires
  ) +
  annotation_north_arrow(location = "tl", which_north = "true",
style = north_arrow_fancy_orienteering)
```



Analyse spatiale

On constate que une absence totale de PME dans certaines regions notamment: Louga , Matam, Kedougou, Tambacounda. Peut être que ces régions n'ont pas été enquêtés

Dans les region de Kaffrine et Koalack toute les PME sont dirigés par les responsables de sexe Feminim et dans la region de thiès la plupart sont dirigés par les femmes également.