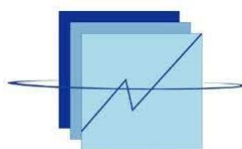


République du Sénégal



Un peuple - Un but - Une foi

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



ANSD

Agence Nationale de la Statistique et de la Démographie (ANSD)



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

PROJET STATISTIQUE SOUS R ET PYTHON

Objet :

PROJET FINAL – LOGICIEL STATISTIQUE R

Rédigé par :

Durel Valdes NZIALI TCHAMOU

Elève Ingénieur Statisticien Economiste, 1^{ère} Année

Sous l'encadrement de :

M. Aboubacar HEMA

Analyst research

juillet 2023

PROJET FINAL – LOGICIEL STATISTIQUE R

Rapport rédigé par :

Durel Valdes NZIALI TCHAMOU

Élève Ingénieur Statisticien Économiste (ISE) en 1^{ère} année à l'ENSAE

Sous l'encadrement de :

M. Aboubacar HEMA

Analyst research

Table des matières

Introduction	4
1 Partie 1	5
1.1 Préparation des données	5
1.1.1 Description	5
1.1.2 Importation et mise en forme	5
1.1.3 Création de variables	10
1.2 Analyses descriptives.	11
1.2.1 Statistiques Descriptives Univariées.	11
1.2.2 Statistiques descriptives bivariées.	16
1.2.3 Statistiques de choix.	20
1.3 Un peu de cartographie.	31
1.3.1 Transformation le data.frame en données géographiques dont l'objet sera nommé projetmap.	31
1.3.2 Représentation spatiale des PME suivant le sexe.	32
1.3.3 Faites une représentation spatiale des PME suivant le niveau d'instruction.	34
1.3.4 Analyse spatiale de choix.	35
2 Partie 2.	41
2.1 Nettoyage et gestion des données.	42
2.1.1 Renommation la variable country destination en destination, et définition les valeurs négatives comme manquantes	42
2.1.2 Créer une nouvelle variable contenant des tranches d'âge de 5 ans en utilisant la variable age.	42
2.1.3 Créer une nouvelle variable contenant le nombre d'entretiens réalisés par chaque agent recenseur.	43
2.1.4 Calculer la durée de l'entretien et indiquer la durée moyenne de l'entretien par enquêteur.	44
2.2 Analyse et visualisation des données.	46
2.2.1 Créez un tableau récapitulatif contenant l'âge moyen et le nombre moyen d'enfants par district.	46
2.3 Les tests	46
2.3.1 Annova	47
2.3.2 Test de student	47
2.3.3 Créer un nuage de points de l'âge en fonction du nombre d'enfants	47
2.3.4 Effet appartenance et intention	48
2.3.5 Presentation des modèles	49
3 Partie 3	50
3.1 Onglet "ACLED Western Africa" (Application 1)	50
3.2 Onglet "ACLED Western Africa" (Application 2)	50
3.3 Onglet "ACLED Western Africa" (Application 3)	51
Conclusion	51

Introduction

Le présent projet est une étude de cas visant à appliquer les connaissances acquises dans le cours sur le logiciel statistique R. Le projet est individuel et a pour objectif d’analyser des données réelles dans le contexte d’une enquête sur les bioénergies durables pour les petites et moyennes entreprises (PME) agroalimentaires en Afrique de l’Ouest.

La première partie du projet se concentre sur la préparation des données. Nous disposerons d’un fichier Excel intitulé “Base_Partie1.xlsx”, contenant 250 observations et 33 variables. Ces variables comprennent des informations telles que la région, le département, le sexe et l’âge du dirigeant de la PME, les langues parlées, le niveau d’instruction, le statut juridique, l’état des infrastructures et bien d’autres. Nous commencerons par importer les données dans R et sélectionner les variables pertinentes pour notre analyse. Ensuite, nous vérifierons s’il y a des valeurs manquantes et identifierons les PME concernées le cas échéant. Nous créerons également de nouvelles variables, comme le nombre de langues parlées par le dirigeant de la PME. Dans la deuxième partie, nous procéderons à des analyses descriptives approfondies pour explorer la répartition des PME en fonction du sexe, du niveau d’instruction, du statut juridique, du propriétaire/locataire, et d’autres facteurs. Nous effectuerons des statistiques descriptives pour fournir des informations clés sur les différentes variables étudiées. Ensuite, nous aborderons la cartographie des PME en utilisant les coordonnées géographiques fournies dans les données. Nous représenterons spatialement les PME en fonction du sexe et du niveau d’instruction pour visualiser leur répartition dans la région d’étude. La troisième partie sera consacrée à l’analyse des données de la deuxième base, “Base_Partie2.xlsx”. Nous effectuerons un nettoyage et une gestion appropriée des données, notamment en renommant les variables, en traitant les valeurs manquantes et en fusionnant les informations pertinentes. Ensuite, nous réaliserons des analyses approfondies et des visualisations pour mieux comprendre les caractéristiques des répondants à l’enquête. Enfin, nous conclurons ce travail en résumant les principales conclusions et en mettant en évidence les résultats les plus pertinents issus de nos analyses statistiques et cartographiques. Nous accorderons une attention particulière à la présentation visuelle des résultats, en utilisant des graphiques et des tableaux de manière à rendre les informations claires et accessibles.

Ce projet nous permettra ainsi d’appliquer de manière pratique les compétences en R acquises au cours de notre formation. Il mettra en évidence notre capacité à traiter des données réelles, à réaliser des analyses statistiques approfondies et à présenter nos résultats de manière claire et convaincante. Le tout dans le contexte stimulant de la recherche de solutions pour promouvoir les bioénergies durables au sein des PME agroalimentaires en Afrique de l’Ouest.

0.0.0.1 Importation des libraries nécessaires

```
library(readxl)
library(dplyr)
library(questionr)
library(gtsummary)
library(gt)
library(webshot)
library(ggplot2)
library(sf)
library(ggpubr)
library(ggspatial)
library(datawizard)
library(kableExtra)
library(janitor)
library(flextable)
library(GGally)
library(tidyverse)
library(tidyr)
library(gridExtra)
library(viridis)
library(tmap)
library(raster)
library(scales)
#install.packages("devtools")
#devtools::install_github("mikabr/ggpirate")
library(ggpirate)
```

1 Partie 1

1.1 Préparation des données

1.1.1 Description

Une description exhaustive des variables a été faite dans le document soumis à notre analyse.

1.1.2 Importation et mise en forme

1.1.2.1 Importer la base de données dans un objet de type *data.frame* nommé *projet*

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

1.1.2.2 Observation préliminaire nécessaire qui porte sur l'analyse des doublons

```
## vérifications des doublons
projet %>%
  duplicated() %>%
  table() %>%
  kbl()
```

.	Freq
FALSE	250

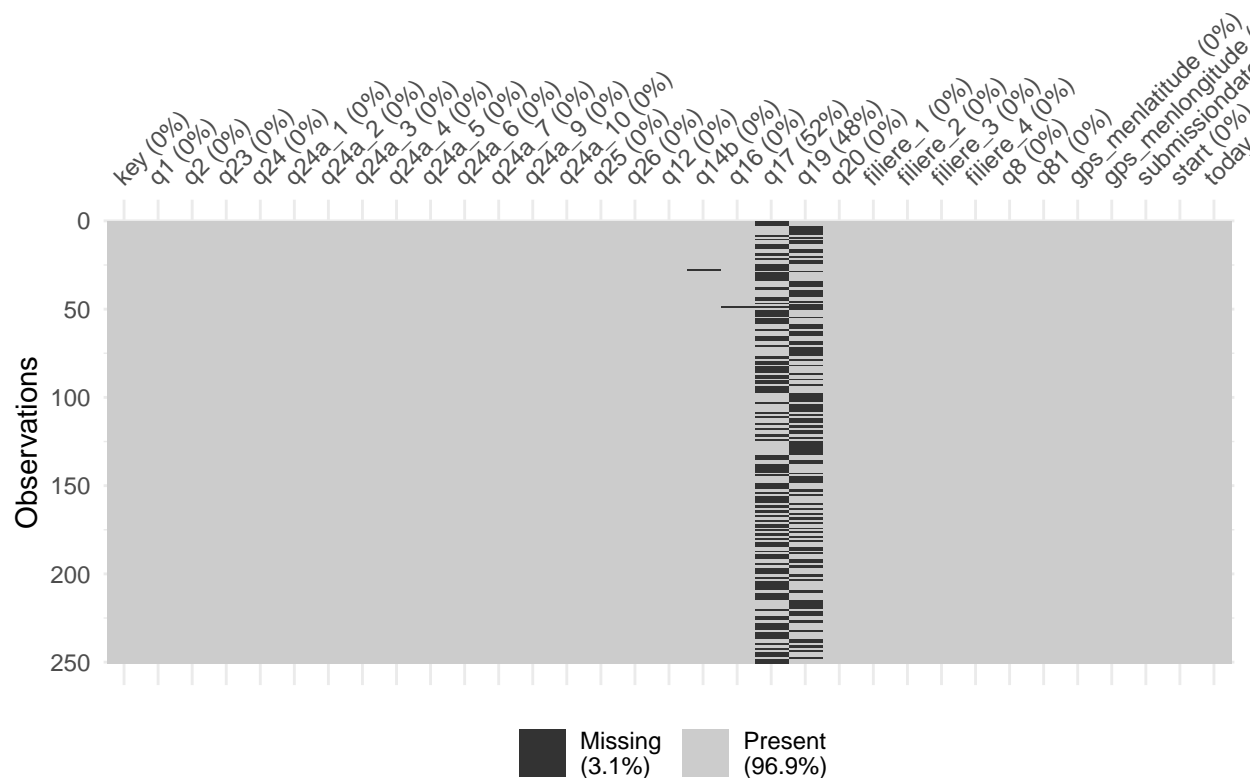
il n'y a pas de doublons dans la base.

1.1.2.3 Selection des variables mentionnées dans la section description.

Il est important de noter que notre base ne contient que des variables décrites dans la section de description.

1.1.2.4 Visualisation graphique d'un résumé des valeurs manquantes par variable

```
#graphique des valeurs manquantes
visdat::vis_miss(projet)
```



une première analyse serait de se dire que les Variables labélisés *Autorisation de fabrication et de mise en vente* et *L'entreprise est-elle desservie par une route bitumée ?* ont chacune 1 valeur manquante. et que celles labélisés* *Etat de la route bitumée* et *l'état de la piste qui mène à l'entreprise* ont chacune deux près de 52% et 48% des valeurs manquantes respectivement. Cependant,* au regard de leur description de la base (q17=*Etat de la route bitumée* et q19=*l'état de la piste qui mène à l'entreprise*) présentent *des hors champs et non des valeurs manquantes*.

1.1.2.5 Présentation sous forme de tableau un résumé des valeurs manquantes par variable

```
projet %>%
  is.na() %>% # Detection des valeurs manquantes

  colSums() %>% # sommes des valeurs par colonne
  sort(decreasing = TRUE) %>% # ranger par ordre décroissant

  as.data.frame() %>%
  setNames("valeurs_manquantes") %>% # donner un nom à la colonne du dataframe

  mutate(pourcentage = round(valeurs_manquantes / sum(valeurs_manquantes, na.rm = TRUE) * 100, 1)
  # Garder seulement les 10 premières lignes de la table
  head(10) %>%
```

```
# Créer la table avec kable
kbl() %>%
kable_paper(full_width = TRUE) %>%
row_spec(1:3, background = "red") %>%
row_spec(3:4, background = "blue")
```

	valeurs_manquantes	pourcentage
q17	131	51.78
q19	120	47.43
q14b	1	0.40
q16	1	0.40
key	0	0.00
q1	0	0.00
q2	0	0.00
q23	0	0.00
q24	0	0.00
q24a_1	0	0.00

Ce tableau vient confirmer les analyses faites plus hauts sur le concept de valeur manquante et de hors champs.

Note : le faite de colorier les 4 premieres lignes est la pour faciliter la présentation des données puisque avec l'observation graphique précédente on avait déjà une vision exhaustive de ce qui etait cherché. Remarqu'on que le table tableaux est par ordre croissant.

1.1.2.6 Ceci est le code dans le cas ou on veut afficher tout le tableau

```
# ici c'est dans le cas ou on veut afficher toutes les variables
projet %>%
  is.na() %>%           # Detection des valeurs manquantes
  colSums() %>%         # sommes des valeurs par colonne
  sort(decreasing = TRUE) %>% # ranger par ordre décroissant
  as.data.frame() %>%
  setNames("valeurs_manquantes") %>% # donner un nom a la colonne du dataframe
  mutate(pourcentage = round(valeurs_manquantes / sum(valeurs_manquantes, na.rm = TRUE) *
    kbl() %>%
    kable_paper(full_width = TRUE)%>%
    row_spec(1:3, background = "red") %>%
    row_spec(3:4, background = "blue")
```


	valeurs_manquantes	pourcentage
q17	131	51.78
q19	120	47.43
q14b	1	0.40
q16	1	0.40
key	0	0.00
q1	0	0.00
q2	0	0.00
q23	0	0.00
q24	0	0.00
q24a_1	0	0.00
q24a_2	0	0.00
q24a_3	0	0.00
q24a_4	0	0.00
q24a_5	0	0.00
q24a_6	0	0.00
q24a_7	0	0.00
q24a_9	0	0.00
q24a_10	0	0.00
q25	0	0.00
q26	0	0.00
q12	0	0.00
q20	0	0.00
filiere_1	0	0.00
filiere_2	0	0.00
filiere_3	0	0.00
filiere_4	0	0.00
q8	0	0.00
q81	0	0.00
gps_menlatitude	0	0.00
gps_menlongitude	0	0.00
submissiondate	0	0.00
start	0	0.00
today	0	0.00

1.1.2.7 Vérification s'il y a des valeurs manquantes pour la variable key dans la base projet.
Si c'est effectivement le cas , l'on identifira la (ou les) PME concernée(s).

```
projet %>%
  filter(is.na(key)) %>%
  dplyr::select(1:10) %>%
  gt()
```

key	q1	q2	q23	q24	q24a_1	q24a_2	q24a_3	q24a_4	q24a_5
-----	----	----	-----	-----	--------	--------	--------	--------	--------

Ce tableau vide ne fait que confirmer le resultat obtenu précédemment, en effet la variable key n'a pas de

valeurs manquantes

1.1.3 Création de variables

1.1.3.1 Renommer les variables q1 en region , q2 en departement et q23 en sexe.

```
# Renommer les variables q1, q2 et q23
projet <- projet %>%
  rename(region = q1, departement = q2, sexe = q23)
```

1.1.3.2 Créer la variable sexe_2 qui vaut 1 si sexe égale à Femme et 0 sinon.

```
# Créer la variable sexe_2
projet <- projet %>%
  mutate(sexe_2=ifelse(projet$sexe == "Femme", 1, 0) )
```

1.1.3.3 Créer un dataframe nommé langues qui prend les variables key et les variables correspondantes décrites plus haut. Indication: Vous remarquerez que ces variables commencent par q24a__.

```
# Créer le dataframe langues
langues <- projet %>%
  dplyr::select(key, starts_with("q24a_"))
```

1.1.3.4 Créer une variable parle qui est égale au nombre de langue parlée par le dirigeant de la PME.

```
# Créer la variable "parle" en calculant la somme des réponses
langues <- langues %>%
  mutate(parle = rowSums(dplyr::select(., starts_with("q24a_"))))
```

1.1.3.5 Sélectionnez uniquement les variables key et parle, l'objet de retour sera langues.

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

1.1.3.6 Merger les data.frame projet et langues

```
projet<-merge(projet, langues,key_vars=key)
```

Note: Cette partie sur la langue il etait aussi possible de le faire en *une seule ligne de code*. le code reduit est donc

```
projet<-projet %>%  
  mutate(parle2 = rowSums(dplyr::select(., starts_with("q24a_"))))
```

1.2 Analyses descriptives.

1.2.1 Statistiques Descriptives Univariées.

1.2.1.1 Tableaux statistiques.

Quelle est la répartition des PME suivant:

- le sexe?
- le niveau d'instruction?
- le statut juridique?
- le propriétaire/locataire?

```
# Répartition des PME suivant ces variables  
tab2 = projet %>%  
  dplyr::select(sexe, q25,q12,q81) %>%  
  rename("Niveau d'instruction du dirigeant/responsable de la PME" = q25, "Statut juridique" = q12, "Statut juridique" = q81)  
  tbl_summary()%>%  
  modify_spanning_header(everything() ~ "**Répartition des PME selon les variables**")  
  
tab2%>% bold_labels() %>% #Mise en forme  
  italicize_levels() %>%  
  modify_header(update = list( label ~ "**VARIABLE**", all_stat_cols(stat_0 = FALSE) ~ "" ),  
    as_flex_table() %>%  
    fontsize(size = 8) %>%  
    width(width = 1.9)
```

Répartition des PME selon les variables	
VARIABLE	N = 250 ¹
sexe	
<i>Femme</i>	191 (76%)
<i>Homme</i>	59 (24%)
Niveau d'instruction du dirigeant/responsable de la PME	
<i>Aucun niveau</i>	79 (32%)
<i>Niveau primaire</i>	56 (22%)
<i>Niveau secondaire</i>	74 (30%)
<i>Niveau Supérieur</i>	41 (16%)
Statut juridique*	
<i>Association</i>	6 (2.4%)
<i>GIE</i>	179 (72%)
<i>Informel</i>	38 (15%)
<i>SA</i>	7 (2.8%)
<i>SARL</i>	13 (5.2%)
<i>SUARL</i>	7 (2.8%)
*propriétaire ou locataire	
<i>Locataire</i>	24 (9.6%)
<i>Propriétaire</i>	226 (90%)
¹ n (%)	

Commentaire : La plupart des PME sont dans des GIE ou exercent dans le secteur informel. Les dirigeants des PME sont en majorité des propriétaires, environ 90%. On constate aussi que très peu d'entre eux ont un niveau supérieur. Une analyse par la filière nous donnera certainement plus de précision.

1.2.1.2 Apport personnel

Avant d'aller plus loin il est important de faire d'observer la répartition des PME par filière :

```
# répartition des PME par filière
tab3= projet %>%
  dplyr::select(starts_with("filiere_")) %>%
  rename("arachide"=filiere_1,"Anacarde"=filiere_2,"Mangue"=filiere_3,"Riz"=filiere_4) %>%
  tbl_summary( missing = "no")
```

```

tab3%>% bold_labels() %>% #Mise en forme
  italicize_levels() %>%
  modify_header(update = list( label ~ "**VARIABLE**", all_stat_cols(stat_0 = FALSE) ~
    as_flex_table() %>%
      fontsize(size = 8) %>%
        width(width = 1)

```

VARIABLE	N = 250 ¹
arachide	108 (43%)
Anacarde	61 (24%)
Mangue	89 (36%)
Riz	92 (37%)
¹ _n (%)	

La filière 1 se démarque avec la plus grande représentation, comptant 108, soit 43% du total. Ensuite, la filière 4 suit de près avec 92, représentant 37%. La filière 3 occupe la troisième position avec 89 et 36% du total, tandis que la filière 2 se trouve en quatrième position avec 61 et 24% du total. Ces pourcentages fournissent une vue d'ensemble des préférences ou répartitions dans les différentes filières, mais une analyse plus approfondie semble nécessaire.

Maintenant que nous savons la répartition des PME par filière, on se demande ce qu'il en est de la répartition des PME par nombre de filière dans laquelle agit les PME.

```

# répartition des PME par nombre de filière dans laquelle agit les PME

```

```

tab4 = projet %>%
  mutate(filiere = as.factor(rowSums(dplyr::select(., starts_with("filiere_"))))) %>%
  dplyr::select(filiere) %>%
  rename("Nombre de filiere"=filiere) %>%
  tbl_summary( missing = "no")

```

```

tab4%>% bold_labels() %>% #Mise en forme
  italicize_levels() %>%
  modify_header(update = list( label ~ "**VARIABLE**", all_stat_cols(stat_0 = FALSE) ~ "**{level}** (
  as_flex_table() %>%
    fontsize(size = 8) %>%
      width(width = 1.2)

```

VARIABLE	N = 250 ¹
Nombre de filiere	
1	171 (68%)
2	59 (24%)
3	19 (7.6%)
4	1 (0.4%)
¹ _n (%)	

il ressort d'une analyse rapide que certains entreprises sont dans plusieurs filière à la fois. En effet **79** sont celles qui sont dans au moins une filière. L'analyse approfondie nous révèle que **59** d'entre elles sont dans deux ilières et une seule est dans quatre filières.

1.2.1.2.1 Création d'une variable avec le nom des filières.

Cette variable est crée pour faire des séparations stricte sur les filières, ceci dans l'optique de distinguer les PME qui sont dans plusieurs activités.

1.2.1.2.2 Représentation graphique.

-Répartition des PME selon le sexe, le niveau scolaire, le statut juridique et le type du dirigeant

Une représentation graphique semble intéressante pour observer ses variables.

```
# Sélection des PME selon le sexe, le niveau scolaire,
# le statut juridique et le type du dirigeant
pme_repartition <- projet %>%
  dplyr::select(sexe, q25, q12, q81)

# Créer les graphiques en secteurs pleins avec les pourcentages
plot1 <- ggplot(pme_repartition %>%
  count(sexe) %>%
  mutate(percentage = n / sum(n) * 100),
  aes(x = "", y = percentage, fill = sexe)) +
  geom_bar(stat = "identity", position = "fill") +
  coord_polar("y") +
  scale_fill_manual(values = c("#E45756", "#4C78A8")) +
  labs(x = NULL, y = NULL, fill = "SEXE") +
  geom_text(aes(label = paste0(round(percentage, 1), "%")),
    position = position_fill(vjust = 0.5), color = "white", size = 3) +
  theme_classic() +
  theme(plot.title = element_text(size = 8, hjust = 0.5))
```

```

plot2 <- ggplot(pme_repartition %>%
  count(q25) %>%
  mutate(percentage = n / sum(n) * 100),
  aes(x = "", y = percentage, fill = q25)) +
  geom_bar(stat = "identity", position = "fill") +
  coord_polar("y") +
  scale_fill_manual(values = c("#E45756", "#4C78A8", "darkgreen", "blue")) +
  labs(x = NULL, y = NULL, fill = "NIVEAU ACADEMIQUE") +
  geom_text(aes(label = paste0(round(percentage, 1), "%")),
    position = position_fill(vjust = 0.5), color = "white", size = 3) +
  theme_classic() +
  theme(plot.title = element_text(size = 12, hjust = 0.5))

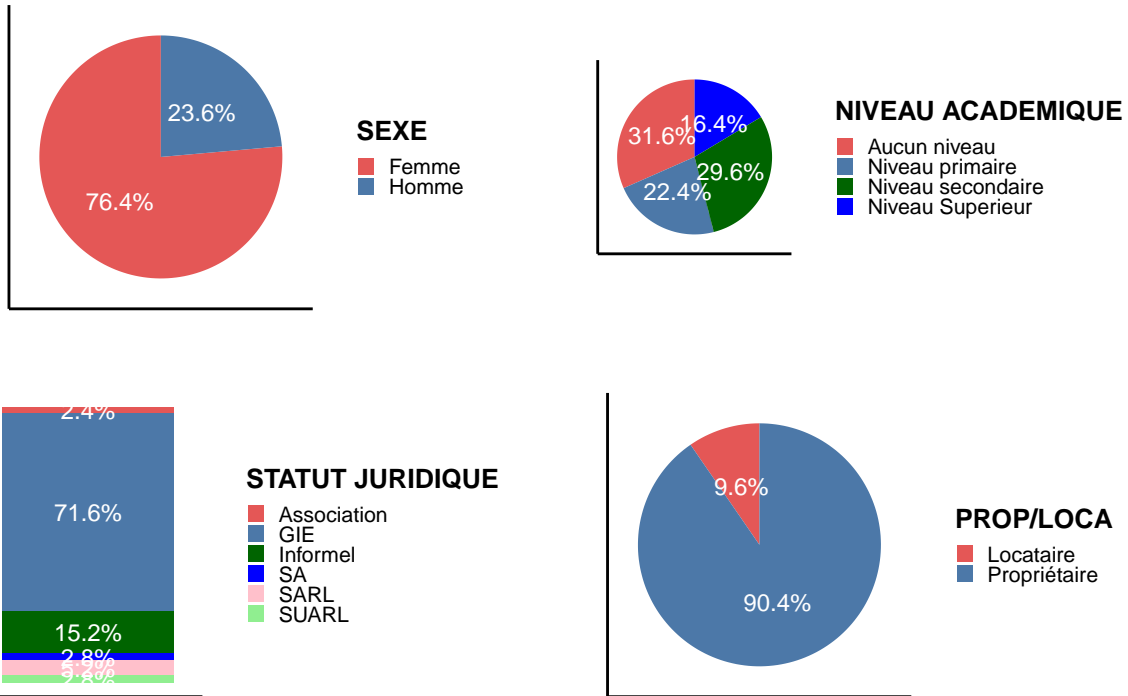
plot3 <- ggplot(pme_repartition %>%
  count(q12) %>%
  mutate(percentage = n / sum(n) * 100),
  aes(x = "", y = n, fill = q12)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = c("#E45756", "#4C78A8",
    "darkgreen", "blue", "pink", "lightgreen")) +
  labs(x = NULL, y = NULL, fill = "STATUT JURIDIQUE") +
  geom_text(aes(label = paste0(round(percentage, 1), "%")),
    position = position_stack(vjust = 0.5), color = "white", size = 3) +
  theme_classic() +
  theme(plot.title = element_text(size = 8, hjust = 0.5))

plot4 <- ggplot(pme_repartition %>%
  count(q81) %>%
  mutate(percentage = n / sum(n) * 100),
  aes(x = "", y = percentage, fill = q81)) +
  geom_bar(stat = "identity", position = "fill") +
  coord_polar("y") +
  scale_fill_manual(values = c("#E45756", "#4C78A8")) +
  labs(x = NULL, y = NULL, fill = "PROP/LOCA") +
  geom_text(aes(label = paste0(round(percentage, 1), "%")),
    position = position_fill(vjust = 0.5), color = "white", size = 3) +
  theme_classic() +
  theme(plot.title = element_text(size = 12, hjust = 0.5))
# Personnaliser le thème global
custom_theme <- theme(
  plot.margin = unit(c(1, 1, 1, 1), "lines"),
  axis.text = element_blank(),
  axis.ticks = element_blank(),
  legend.title = element_text(size = 10, face = "bold"),
  legend.text = element_text(size = 8),
  legend.key.size = unit(0.5, "lines"),
  legend.key = element_rect(color = "transparent"),
  panel.spacing = unit(3, "lines")
)
# Afficher les 4 graphiques côte à côte avec le thème
#personnalisé et des barres de séparation
grid.arrange(plot1 + custom_theme, plot2 + custom_theme,

```

```
plot3 + custom_theme, plot4 + custom_theme,
ncol = 2, nrow = 2, widths = c(1, 1), heights = c(1, 1),
top = "REPARTITION DES PME SELON
LES CARACTERISTIQUES DES DIRIGEANTS")
```

REPARTITION DES PME SELON LES CARACTERISTIQUES DES DIRIGEANTS



1.2.2 Statistiques descriptives bivariées.

1.2.2.1 Tableaux statistiques.

- le statut juridique et le sexe

- Niveau d'instruction du dirigeant/responsable de la PME et sexe
- propriétaire ou locataire et sexe

*# Répartition selon le statut juridique et le sexe, Niveau
d'instruction du dirigeant/responsable de la PME et sexe,
propriétaire ou locataire et sexe : profil colonne*


```

tab=projet %>%
  dplyr::select(q81,q25,q12, sexe) %>%
  rename("Niveau d'instruction du dirigeant/responsable de la PME" = q25,
    "Statut juridique*" = q12, "propriétaire ou locataire" = q81) %>%
  tbl_summary(by = sexe, missing = "no") %>%
  modify_spanning_header(everything() ~ "**Répartition
    des PME selonleur statut juridique
    et le sexe de son dirigeant**")
tab%>% bold_labels() %>%      #Mise en forme
  italicize_levels() %>%
  modify_header(update = list( label ~ "**VARIABLE**",
    all_stat_cols(stat_0 = FALSE) ~ "**{level}**
    (n={n}, {style_percent(p)}%)" )) %>%

  as_flex_table() %>%
  fontsize(size = 8) %>%
  width(width = 1.7)

```

Répartition des PME selonleur statut juridique et le sexe de son dirigeant		
VARIABLE	Femme (n=191, 76%) ¹	Homme (n=59, 24%) ¹
propriétaire ou locataire		
<i>Locataire</i>	16 (8.4%)	8 (14%)
<i>Propriétaire</i>	175 (92%)	51 (86%)
Niveau d'instruction du dirigeant/responsable de la PME		
<i>Aucun niveau</i>	70 (37%)	9 (15%)
<i>Niveau primaire</i>	48 (25%)	8 (14%)
<i>Niveau secondaire</i>	56 (29%)	18 (31%)
<i>Niveau Supérieur</i>	17 (8.9%)	24 (41%)
Statut juridique*		
<i>Association</i>	3 (1.6%)	3 (5.1%)
<i>GIE</i>	149 (78%)	30 (51%)
<i>Informel</i>	32 (17%)	6 (10%)
<i>SA</i>	1 (0.5%)	6 (10%)
<i>SARL</i>	2 (1.0%)	11 (19%)
<i>SUARL</i>	4 (2.1%)	3 (5.1%)
¹ n (%)		

Commentaire:

Parmi les propriétaires, 92% sont des femmes, tandis que parmi les locataires, la répartition est plus équilibrée avec 14% de femmes et 14% d'hommes.

En ce qui concerne le niveau d'éducation, les hommes ont tendance à avoir un niveau d'éducation supérieur (41%) par rapport aux femmes (8,9%).

Par contre, les femmes ont une plus grande présence dans les niveaux d'éducation primaire (14%) et secondaire (29%) par rapport aux hommes (14% et 31% respectivement).

on constate que les entreprises dirigées par des hommes ont une plus grande présence dans les formes juridiques de type GIE (Groupement d'Intérêt Économique), SA (Société Anonyme), SARL (Société à Responsabilité Limitée) et SUARL (Société Unipersonnelle à Responsabilité Limitée).

D'autre part, les femmes dirigeantes dominent dans le statut juridique informel, qui représente 17% de leur part, tandis que les hommes n'en représentent que 10%.

1.2.2.2 Faire en sortir les statistiques univariées et Bivariées en un seul tableau

```
tbl_merge(list(tab2,tab),
          # definition des panels
          tab_spanner = c("ANALYSE UNIVARIE",
                          "ANALYSE BIVARIE"))%>%

# Mise en forme
# Mise en Gras
bold_labels() %>%

# Mise en italique
italicize_levels() %>%

# Mise en forme des entêtes

modify_header(update = list( label ~ "**VARIABLE**",
                              all_stat_cols(stat_0 = FALSE) ~ "**{level}**"
                              (n={n},{style_percent(p)}%)"))
as_flex_table() %>%
# Mise en forme des dimensions
fontsize(size = 8) %>%
width(width = 1.65)
```

Répartition des PME selon les variables

	ANALYSE UNIVARIE	ANALYSE BIVARIE	
VARIABLE	N = 250 ¹	Femme (n=191,76%) ¹	Homme (n=59,24%) ¹
sexe			
Femme	191 (76%)		
Homme	59 (24%)		

¹n (%)

Répartition des PME selon les variables		ANALYSE UNIVARIE		ANALYSE BIVARIE	
VARIABLE	N = 250 ¹	Femme (n=191,76%) ¹	Homme (n=59,24%) ¹		
Niveau d'instruction du dirigeant/responsable de la PME					
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 Supérieur	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 ou locataire					
Locataire	24 (9.6%)				
Propriétaire	226 (90%)				
propriétaire ou locataire					
Locataire		16 (8.4%)	8 (14%)		
Propriétaire		175 (92%)	51 (86%)		
¹ _n (%)					

¹n (%)

1.2.2.3 Représentation graphique

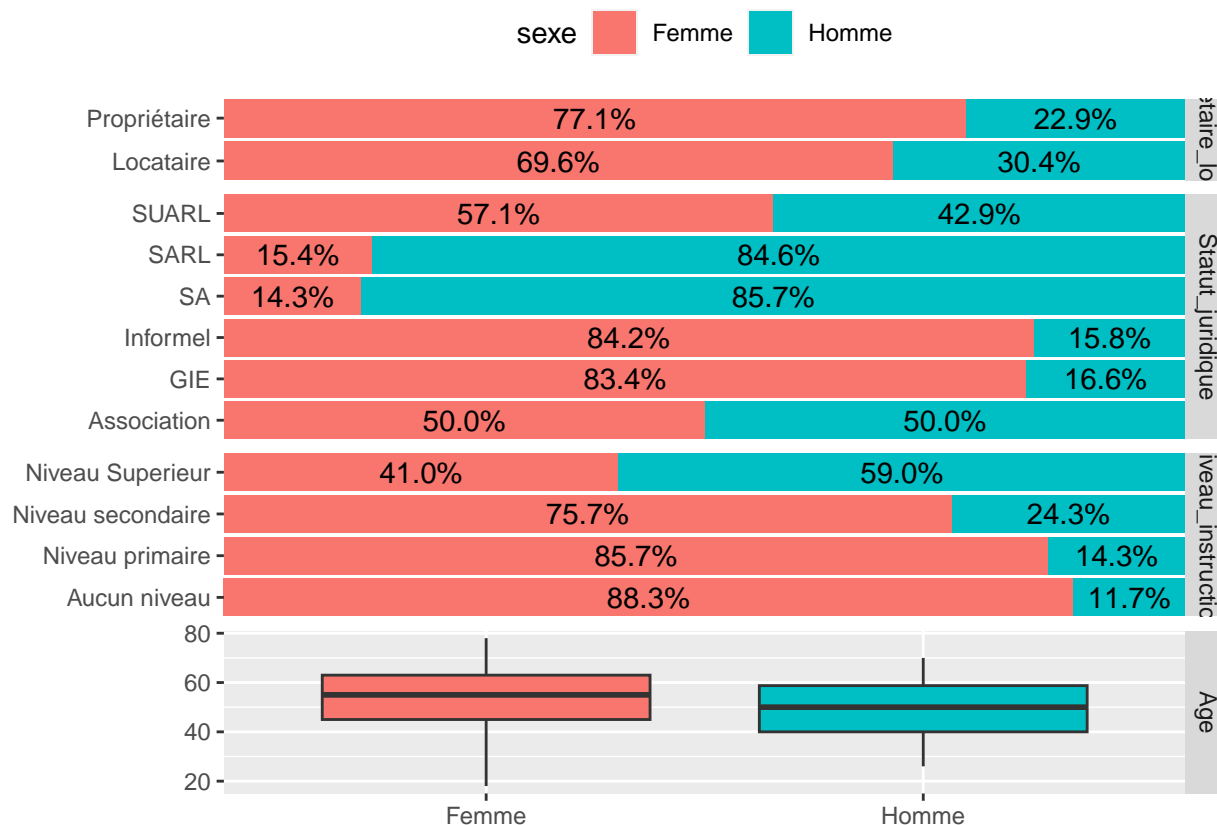
Pour une meilleur visibilité il est possible de faire un graphique qui est assez parlant et révélateur des mêmes analyses que les tableaux.

```

projet$Age = projet$q24 # Creation du graphique
Graph_bi = projet %>%
  filter( Age < 80 ) %>% rename(
    Proprietaire_locataire = q81,
    Statut_juridique = q12,
    Niveau_instruction = q25)
print(ggbivariate(data = Graph_bi, # Mise en forme
  outcome = "sexe",
  explanatory = c("Proprietaire_locataire",

```

```
"Statut_juridique",
"Niveau_instruction",
"Age")))
```



ce graphique est une illustration plus synthétique des analyses bivariées effectués

1.2.3 Statistiques de choix.

1.2.3.1 Activité principale de l'entreprise par filière.

1.2.3.1.1 Activite principale de l'entreprise par rapport au filière

```
#
tab5=projet %>%
  dplyr::select(q8, starts_with("filiere_")) %>%
  rename("Arachide" = filiere_1,
         "Anacarde"=filiere_2, "Mangue"= filiere_3,"Riz"= filiere_4) %>%
  tbl_summary(by = q8, missing = "no")%>%
  modify_spanning_header(everything() ~ "**Répartition des PME selon
la filière et l'Activité
principale de l'entreprise**")
```

```

tab5%>% bold_labels() %>%
  italicize_levels() %>%
  modify_header(update = list( label ~ "**VARIABLE**",
                                all_stat_cols(stat_0 = FALSE) ~ "**{level}**"
                                (n={n},{style_percent(p)}%)")
  )) %>% as_flex_table() %>%
  fontsize(size = 8) %>%
  width(width = 1.2)

```

****Répartition des PME selon
la filière et l'Activité
principale de l'entreprise****

VARIABLE	Aucun (n=5,2.0%) ¹	Autre a preciser (n=4,1.6%) ¹	Tansformation d'autres céréales (n=57,23%) ¹	Transformation d'autres fruits et legumes (n=14,5.6%) ¹	Transformation d'autres produits oléagineux (n=1,0.4%) ¹	Tran de (n
Arachide	5 (100%)	1 (25%)	38 (67%)	7 (50%)	1 (100%)	
Anacarde	0 (0%)	0 (0%)	1 (1.8%)	2 (14%)	0 (0%)	
Mangue	0 (0%)	0 (0%)	28 (49%)	6 (43%)	0 (0%)	
Riz	0 (0%)	3 (75%)	21 (37%)	9 (64%)	0 (0%)	

¹n (%)

L'analyse des croisements entre filières et activités principales des PME met en évidence des informations intéressantes. Parmi les différentes filières, la transformation de l'arachide et du riz se distingue comme les activités principales les plus fréquentes, représentant respectivement 19% et 18% des PME étudiées. Dans la filière de l'arachide, toutes les PME sont engagées dans cette activité, et une grande majorité (67%) sont exclusivement dédiées à cette transformation. Dans la filière du riz, une proportion significative (37%) se concentre uniquement sur cette transformation. La transformation d'autres céréales et de fruits et légumes est également représentée, avec des pourcentages de 23% et 5,6% respectivement. La transformation de la mangue est principalement présente dans la filière de l'anacarde (49%) et du riz (43%).

1.2.3.1.2 Activité principale suivant les spécificités des filières

```

#plus spécifiquement
tab6=projet %>%
  dplyr::select(q8, nom_filiere) %>%
  rename("Activité principale"=q8) %>%
  tbl_summary(by = nom_filiere, missing = "no") %>%
  modify_spanning_header(everything() ~
    "**Répartition des PME selon  
la filière et l'Activité principale  
de l'entreprise**")

```

```

tab6%>% bold_labels() %>%
  italicize_levels() %>%
  modify_header(update = list( label ~ "**VARIABLE**",
                                all_stat_cols(stat_0 = FALSE) ~
                                "**{level}** (n={n}, {style_percent(p)}%)"
  )) %>% as_flex_table() %>%
  fontsize(size = 8) %>%
  width(width = 1.1)

```

****Répartition des PME selon
la filière et l'Activité principale
de l'entreprise****

VARIABLE	anacarde (n=25, 10%) ¹	arachide (n=61, 24%) ¹	**arachide & anacarde & mangue & riz** (n=1, 0.4%) ¹	**arachide & anacarde & mangue** (n=1, 0.4%) ¹	arachide & anacarde (n=6, 2.4%) ¹	arachide & anacarde & (n=2, 0.8%)
Activité principale						
<i>Aucun</i>	0 (0%)	5 (8.2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<i>Autre a preciser</i>	0 (0%)	1 (1.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<i>Transformation d'autres céréales</i>	1 (4.0%)	15 (25%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<i>Transformation d'autres fruits et legumes</i>	1 (4.0%)	2 (3.3%)	0 (0%)	0 (0%)	1 (17%)	0 (0%)
<i>Transformation d'autres produits oléagineux</i>	0 (0%)	1 (1.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<i>Transformation de l'arachide</i>	0 (0%)	37 (61%)	0 (0%)	1 (100%)	3 (50%)	0 (0%)
<i>Transformation de la mangue</i>	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (50%)
<i>Transformation de la noix de cajoux</i>	23 (92%)	0 (0%)	0 (0%)	0 (0%)	2 (33%)	1 (50%)
<i>Transformation de la pomme de cajoux</i>	0 (0%)	0 (0%)	1 (100%)	0 (0%)	0 (0%)	0 (0%)
<i>Transformation du riz</i>	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

¹n (%)

En ce qui concerne la combinaison de filières, l'association de l'arachide et de la noix de cajou est également notable, représentant 9,2% des PME. De plus, l'association de l'arachide, de la noix de cajou et de la mangue est présente chez 1,2% des PME. La transformation du riz se distingue également, représentant 13% de toutes les PME. Parmi celles-ci, 79% sont exclusivement dédiées à cette activité. La combinaison de riz et d'anacarde est présente chez 9,2% des PME. Ces combinaisons de filières et d'activités principales mettent en évidence les

spécificités du secteur des PME dans le domaine de la transformation des produits agricoles, en particulier l'arachide, la noix de cajou, la mangue et le riz.

1.2.3.2 Analyse des langues parlées par les dirigeants de PME.

1.2.3.2.1 Répartition des PME selon le nombre filiere suivant les langues parlées par dirigeant et Identification des langues les plus couramment utilisées

```
tab7=projet %>%
  dplyr::select(filiere, starts_with("q24a_")) %>%
  rename("français"=q24a_1, "wolof"=q24a_2, "Diola"=q24a_3, "Serere"=q24a_4,
         "Peul"=q24a_5, "Mandingue"=q24a_6, "Balante"=q24a_7, "Bambara"=q24a_9,
         "autre langue"=q24a_10) %>%
  tbl_summary(by = filiere, missing = "no") %>%
  modify_spanning_header(everything() ~ "**Répartition
                           des PME selon le nombre filiere
                           suivant les langues parlées par dirigeant**")

tab7 %>% bold_labels() %>%
  italicize_levels() %>%
  modify_header(update = list( label ~ "**MODALITE**",
                              all_stat_cols(stat_0 = FALSE) ~
                                "**{level}** (n={n}, {style_percent(p)})"
                              )) %>% as_flex_table() %>%
  fontsize(size = 8) %>%
  width(width = 1.2)
```

Répartition des PME selon le nombre filiere suivant les langues parlées par dirigeant				
MODALITE	1 (n=171, 68%)¹	2 (n=59, 24%)¹	3 (n=19, 7.6%)¹	4 (n=1, 0.4%)¹
français	95 (56%)	50 (85%)	13 (68%)	1 (100%)
wolof	166 (97%)	59 (100%)	19 (100%)	1 (100%)
Diola	16 (9.4%)	18 (31%)	4 (21%)	1 (100%)
Serere	46 (27%)	11 (19%)	4 (21%)	0 (0%)
Peul	27 (16%)	12 (20%)	3 (16%)	1 (100%)
Mandingue	17 (9.9%)	11 (19%)	8 (42%)	0 (0%)
Balante	1 (0.6%)	0 (0%)	1 (5.3%)	0 (0%)
Bambara	2 (1.2%)	1 (1.7%)	0 (0%)	0 (0%)
autre langue	20 (12%)	10 (17%)	4 (21%)	1 (100%)

¹n (%)

Commentaire : Les dirigeants parlant français montrent une diversification modérée avec 56% impliqués dans une seule filière, tandis que les dirigeants parlant wolof se concentrent davantage sur une activité principale (97% dans une seule filière). Les dirigeants parlant diola et mandingue ont une diversification plus marquée, avec respectivement 31% et 42% impliqués dans trois filières. De plus, la langue parlée semble influencer la diversification globale des PME, avec certains dirigeants d'autres langues impliqués dans les quatre filières étudiées.

1.2.3.2.2 Identification du nombre de langues parlé par les dirigeants en fonction de leur nombre de filière.

```
tab8=projet %>%
  dplyr::select(filiere, parle) %>%
  rename("Nombre de langue Parlé"= parle) %>%
  tbl_summary(by = filiere, missing = "no") %>%
  modify_spanning_header(everything() ~
    "**Répartition des PME selon le nombre
    filiere suivant le nombre de langues
    parlées par dirigeant**")

tab8 %>% bold_labels() %>%
  italicize_levels() %>%
  modify_header(update = list( label ~ "**MODALITE**",
    all_stat_cols(stat_0 = FALSE) ~
    "**{level}** (n={n}, {style_percent(p)})"))
  )) %>% as_flex_table() %>%
  fontsize(size = 8) %>%
  width(width = 1)
```

Répartition des PME selon le nombre filiere suivant le nombre de langues parlées par dirigeant				
MODALITE	1 (n=171, 68%) ¹	2 (n=59, 24%) ¹	3 (n=19, 7.6%) ¹	4 (n=1, 0.4%) ¹
Nombre de langue Parlé				
1	31 (18%)	4 (6.8%)	2 (11%)	0 (0%)
2	81 (47%)	18 (31%)	5 (26%)	0 (0%)
3	41 (24%)	20 (34%)	6 (32%)	0 (0%)
4	16 (9.4%)	14 (24%)	4 (21%)	0 (0%)
5	2 (1.2%)	2 (3.4%)	2 (11%)	1 (100%)
6	0 (0%)	1 (1.7%)	0 (0%)	0 (0%)
¹ n (%)				

La majorité des PME (68%) sont dirigées par des dirigeants parlant une seule langue, tandis que 24% et 7.6% des PME sont respectivement dirigées par des dirigeants parlant deux et trois langues différentes. Les PME ayant plus de trois filières sont rares, représentant moins de 1% de l'échantillon. En ce qui concerne le nombre de langues parlées par dirigeant, la majorité des dirigeants (47%) parlent deux langues, suivis de 24% parlant trois langues et 9.4% parlant quatre langues. Il est intéressant de noter que la plupart des dirigeants parlant une seule langue dirigent une seule filière (18%), tandis que parmi les dirigeants parlant deux langues, près de la moitié dirige deux filières (47%).

1.2.3.2.3 Un regard sur le wolof

```
f1 = projet %>%
filter(filiere_1 == 1) %>%
dplyr::mutate(Arachide = ifelse(filiere_1 == 1, "Arachide", "NA")) %>%
gtsummary::tbl_summary(
include = c("q24a_2", "Arachide"),
by = Arachide,
statistic = list( all_categorical() ~ "{p}% ({n}/{N})"),
label = q24a_2 ~ "WOLOF"
)

f2 = projet %>%
filter(filiere_2 == 1) %>%
dplyr::mutate(Anacarde = ifelse(filiere_2 == 1, "Anacarde", "NA")) %>%
gtsummary::tbl_summary(
include = c("q24a_2", "Anacarde"),
by = Anacarde,
statistic = list( all_categorical() ~ "{p}% ({n}/{N})"),
label = q24a_2 ~ "WOLOF"
)

f3 = projet %>%
filter(filiere_3 == 1) %>%
dplyr::mutate(Mangue = ifelse(filiere_3 == 1, "Mangue", "NA")) %>%
gtsummary::tbl_summary(
include = c("q24a_2", "Mangue"),
by = Mangue,
statistic = list( all_categorical() ~ "{p}% ({n}/{N})"),
label = q24a_2 ~ "WOLOF"
)

f4 = projet %>%filter(filiere_4 == 1) %>%
dplyr::mutate(Riz = ifelse(filiere_4 == 1, "Riz", "NA")) %>%
gtsummary::tbl_summary(
include = c("q24a_2", "Riz"),
by = Riz,
statistic = list( all_categorical() ~ "{p}% ({n}/{N})"),
label = q24a_2 ~ "WOLOF"
)

tbl_merge(list(f1,f2,f3,f4))%>%
bold_labels() %>%
italicize_levels() %>%
```

```

modify_header(update = list( label ~ "**FILIÈRES**",
                             all_stat_cols(stat_0 = FALSE) ~ "{level}" (n={n},
                             {style_percent(p)}%,size = 8)"
                             )) %>%
  modify_spanning_header(all_stat_cols() ~
  "**POURCENTAGE DE DIRIGEANT
  PARLANT WOLOF PAR FILIERE**") %>%
  as_flex_table() %>%
  fontsize(size = 8) %>%
  width(width = 1)

```

	POURCENTAGE DE DIRIGEANT PARLANT WOLOF PAR FILIERE			
FILIÈRES	Arachide (n=108, 100%,size = 8) ¹	Anacarde (n=61, 100%,size = 8) ¹	Mangue (n=89, 100%,size = 8) ¹	Riz (n=92, 100%,size = 8) ¹
WOLOF	100% (108/108)	95% (58/61)	100% (89/89)	98% (90/92)

¹% (n/N)

On observe que tous les dirigeants des entreprises opérant dans la filière Arachide (n=108) parlent wolof, soit un taux de 100%. Dans la filière Anacarde, la proportion de dirigeants parlant wolof est de 95% (58 sur 61 dirigeants). De même, tous les dirigeants de la filière Mangue (n=89) parlent wolof, ce qui représente 100% de la totalité. Pour la filière Riz, on constate que 98% des dirigeants (90 sur 92) parlent wolof. Ces résultats indiquent que le wolof est largement utilisé par les dirigeants d'entreprises dans toutes les filières étudiées, avec des taux de fréquence élevés, soulignant ainsi son importance dans le contexte économique des PME opérant dans ces secteurs spécifiques.

1.2.3.3 Représentations graphiques.

Répartition du sexe du dirigeant/responsable de la PME dans chaque filière pour évaluer les disparités éventuelles.

```

# ggplot code for filiere = 1
plot_filiere_1 <- ggplot(projet %>% filter(filiere == 1 |
                                           filiere == 2|filiere == 3 |
                                           filiere == 4),
  aes(x = nom_filiere, fill = sexe)) +
  geom_bar(position = "fill") +
  labs(x = "TYPE DE FILIERE", y = "POURCENTAGE",
  fill = "SEXE DU DIRIGEANT") +
  scale_fill_manual(values = c("pink1","dodgerblue1" ),
  labels = c( "Femme","Homme")) +
  theme_minimal()+
  theme(legend.text = element_text(face = "italic", size = 8))+
  theme(axis.text.x = element_text(angle = 90,
  hjust = 1))+ # Ajout du thème pour

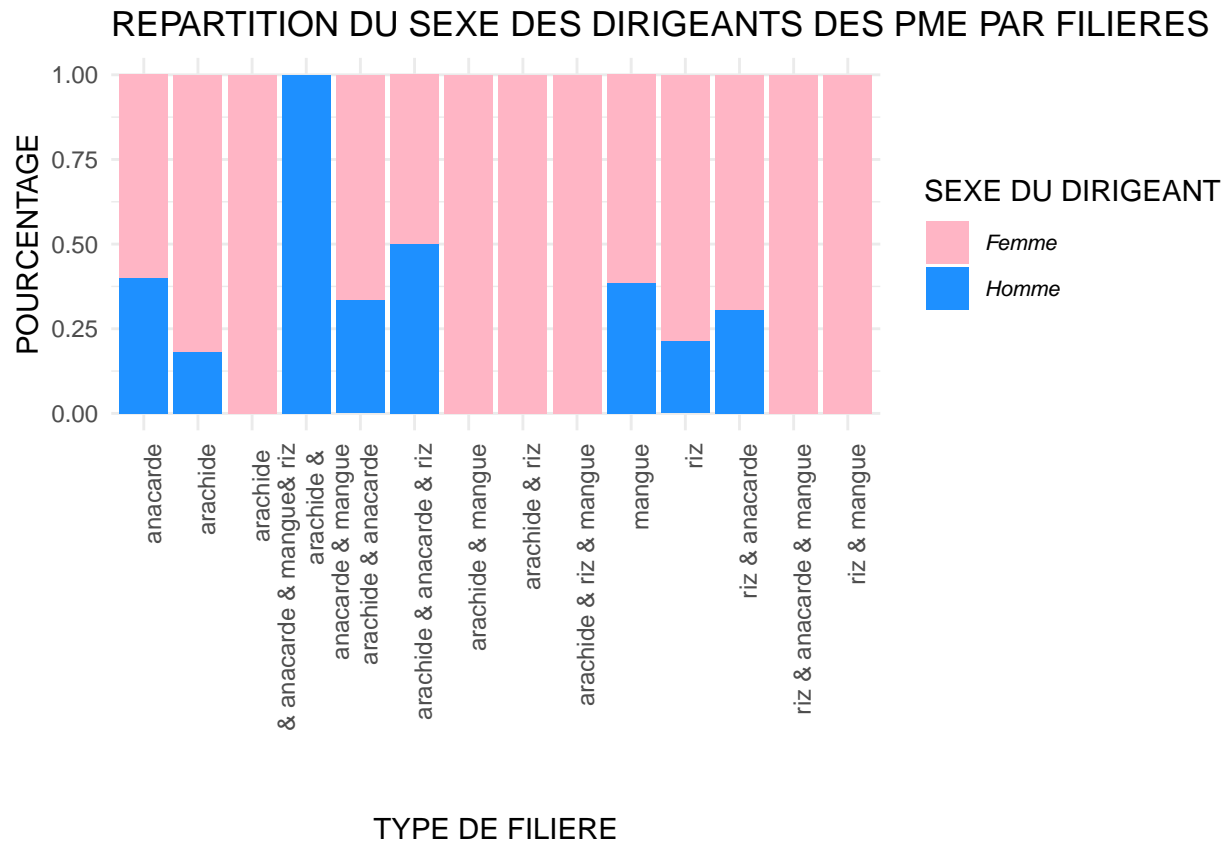
```

```

# pivoter les étiquettes de l'axe x
labs(title = "REPARTITION DU SEXE DES DIRIGEANTS DES PME PAR FILIERES")

plot_filiere_1

```



1.2.3.3.1 Un regard sur la période de l'enquête

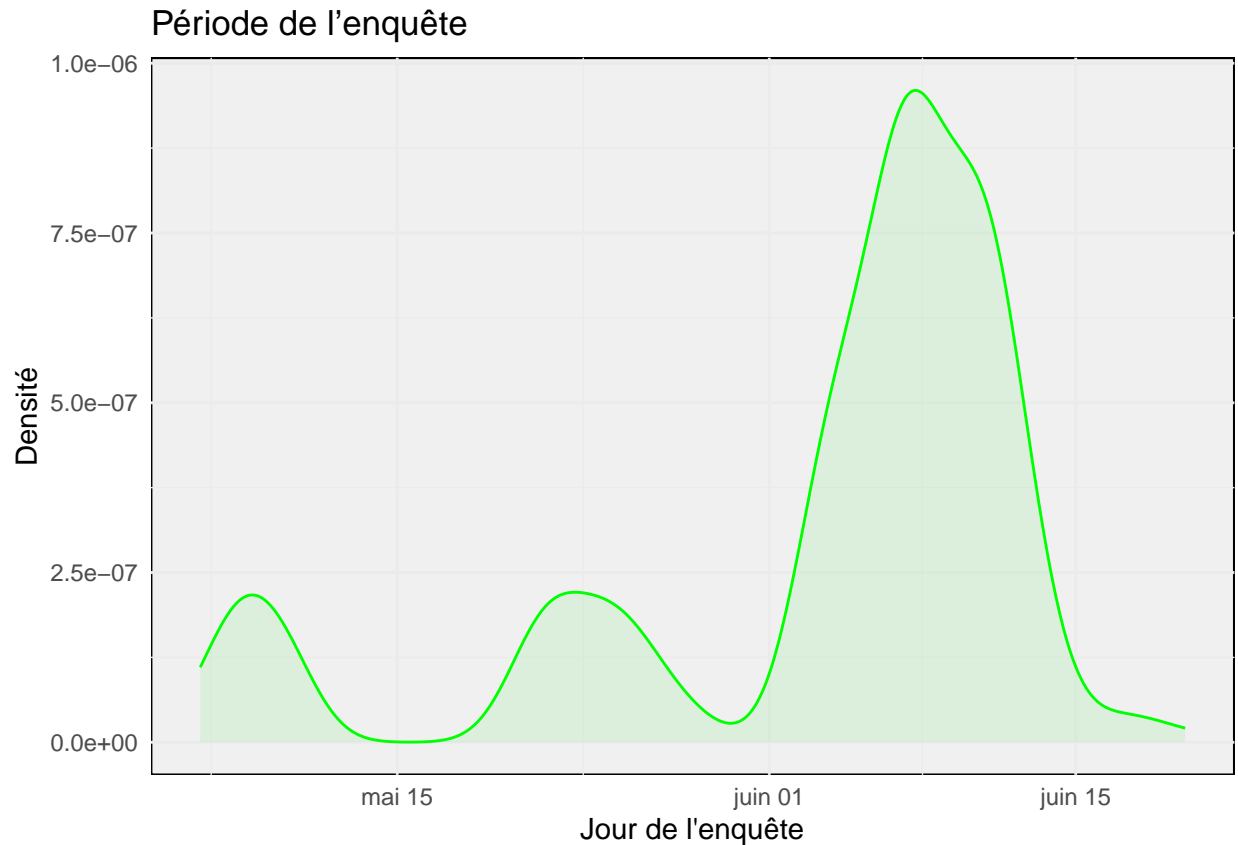
```

# Couleur personnalisée pour le fond du panneau principal
background_color <- "#f0f0f0"

# Créer le graphique de densité avec la couleur de fond personnalisée
ggplot(projet) +
  aes(x = today) +
  geom_density(fill = "lightgreen",
               color = "green", alpha = 0.2) + # Remplissage en
                                                #bleu avec bordures blanches

  ggtitle("Période de l'enquête") +
  xlab("Jour de l'enquête") +
  ylab("Densité") +
  theme_minimal() +
  theme(panel.background = element_rect(
    fill = background_color)) # Définir la couleur du

```



fond du panneau principal

Ceci montre que la période d'enquête la plus intense est du 01 juin au 15 juin.

1.2.3.3.2 Analyse de l'âge du dirigeant/responsable de la PME (variable q24) dans chaque filière pour déterminer les tendances générationnelles.

On commencera d'abord par filtrer les données pour les âges inférieurs à 100 ans, puisque nous avons à faire à des dirigeants on suppose que ceux-ci sont en état de travail, l'âge de retraite étant d'un peu plus de 64 ans nous prenons donc une marge de +40 ceci nous permet d'éviter de fausser nos données par des erreurs de saisie comme par exemple saisir un numéro de téléphone à la place de l'âge

```
projet_filtered <- subset(projet, Age < 100)
# Défini les couleurs pour les niveaux du sexe
couleurs <- c("Femme" = "pink1", "Homme" = "dodgerblue1")
# Regrouper les âges en catégories
projet_filtered$Age_Group <- cut(projet_filtered$Age, breaks = c(0, 20, 30,
                                                                40, 50, 60,
                                                                70, 80, 100),
                                labels = c("0-20", "20-30", "30-40",
                                             "40-50", "50-60", "60-70",
```

```

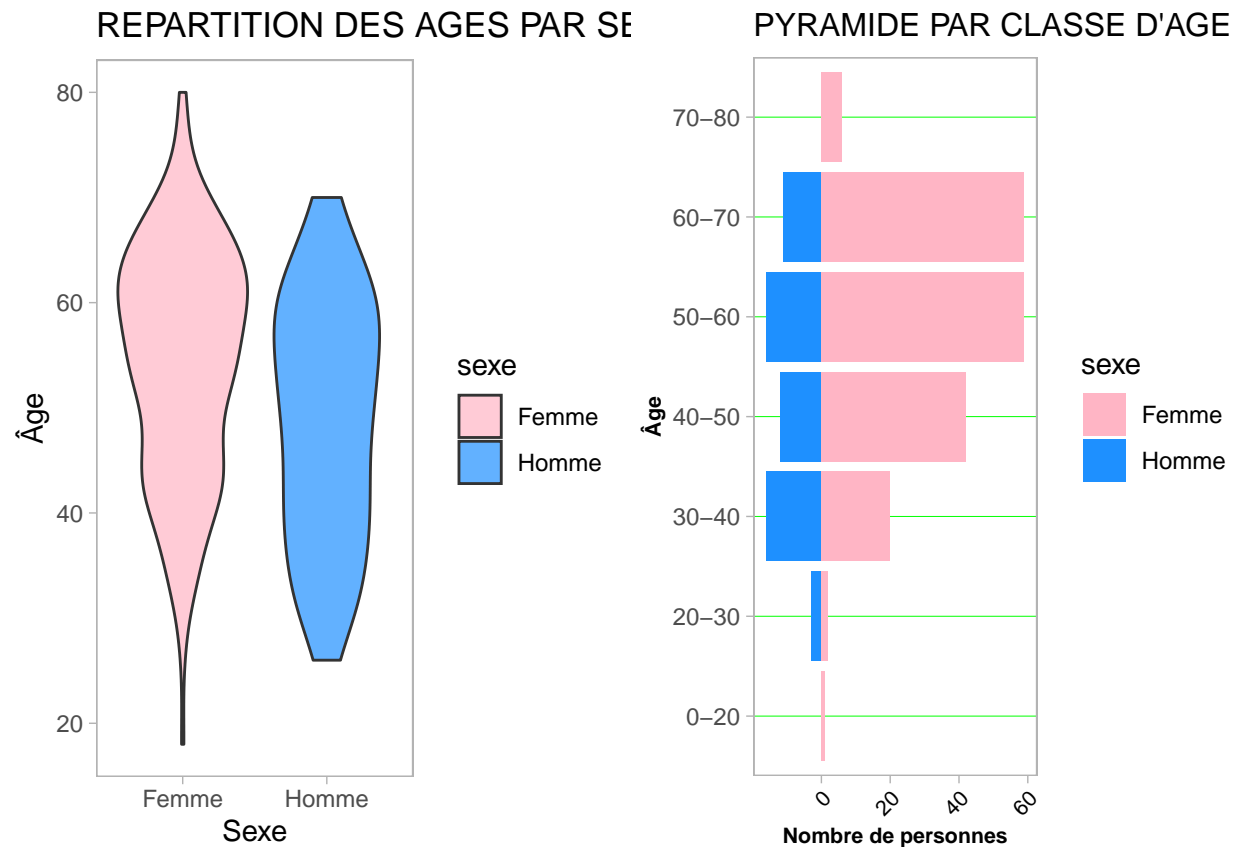
"70-80", "80+"))

# Créer la pyramide des âges avec ggplot2
pyramide <- ggplot(projet_filtered, aes(x = Age_Group, fill = sexe)) +
  geom_bar(data = subset(projet_filtered, sexe == "Homme"),
    aes(y = -..count..), position = "dodge") +
  geom_bar(data = subset(projet_filtered, sexe == "Femme"),
    aes(y = ..count..), position = "dodge") +
  scale_fill_manual(values = couleurs) +
  labs(x = "Âge", y = "Nombre de personnes",
    title = "PYRAMIDE PAR CLASSE D'AGE") +
  theme_minimal() +
  coord_flip()+
  theme_light()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1,
    vjust = 1, size = 8, color = "black"),
    axis.title = element_text(size=8, face = "bold"),
    plot.title = element_text(size = 12),
    panel.grid.major.x = element_blank(),
    panel.grid.minor.x = element_blank(),
    panel.grid.major.y = element_line(color = "green", size = 0.1),
    panel.background = element_rect(fill = "white"))

# Crée du graphique violon
Violon<-ggplot(projet_filtered, aes(x = sexe, y = Age, fill = sexe)) +
  geom_violin(alpha = 0.7) + # Opacité pour une meilleure lisibilité des violins
  scale_fill_manual(values = couleurs) + # Applique les couleurs définies
  xlab("Sexe") +
  ylab("Âge") +
  ggtitle("REPARTITION DES AGES PAR SEXE") +
  theme_minimal() + # Utiliser le thème minimal
  theme_light()+
  theme(
    plot.background = element_rect(fill = "white"), # Change la couleur de fond
    panel.grid.major = element_blank(), # Supprime les lignes de grille majeures
    panel.grid.minor = element_blank() # Supprime les lignes de grille mineures
  )

grid.arrange( Violin, pyramide, ncol = 2)

```



Une fois cette observation faite, il semble important de se demander quel est la répartition par Années d'expérience selon le sexe nous l'alignerons avec un boxplot sur les âges afin de mieux les analyser

```
Repartition1 = projet %>%
  rename(age = q24) %>%
  filter( age < 100 )
Repartition2 = projet %>%
  rename(Annees_experience = q26) %>%
  filter( Annees_experience < 50 )
plot1 <- ggplot(Repartition1, aes(x = sexe, y = age)) +
  geom_pirate(aes(colour = sexe)) +
  xlab("Sexe") +
  ylab("Âge") +
  ggtitle("Répartition par âge selon le sexe") +
  theme_light() +
  theme(
    plot.title = element_text(color = "black")
  )
plot2 <- ggplot(Repartition2, aes(x = sexe, y = Annees_experience)) +
  geom_pirate(aes(colour = sexe)) +
  xlab("Sexe") +
  ylab("Nombre d'années d'expérience") +
  ggtitle("Années d'expérience selon le sexe") +
  theme_light() +
```

```
theme(
  plot.title = element_text(color = "black")
  # Afficher les graphiques côte à côte sur une même ligne
  grid.arrange(plot1, plot2, ncol = 2)
```



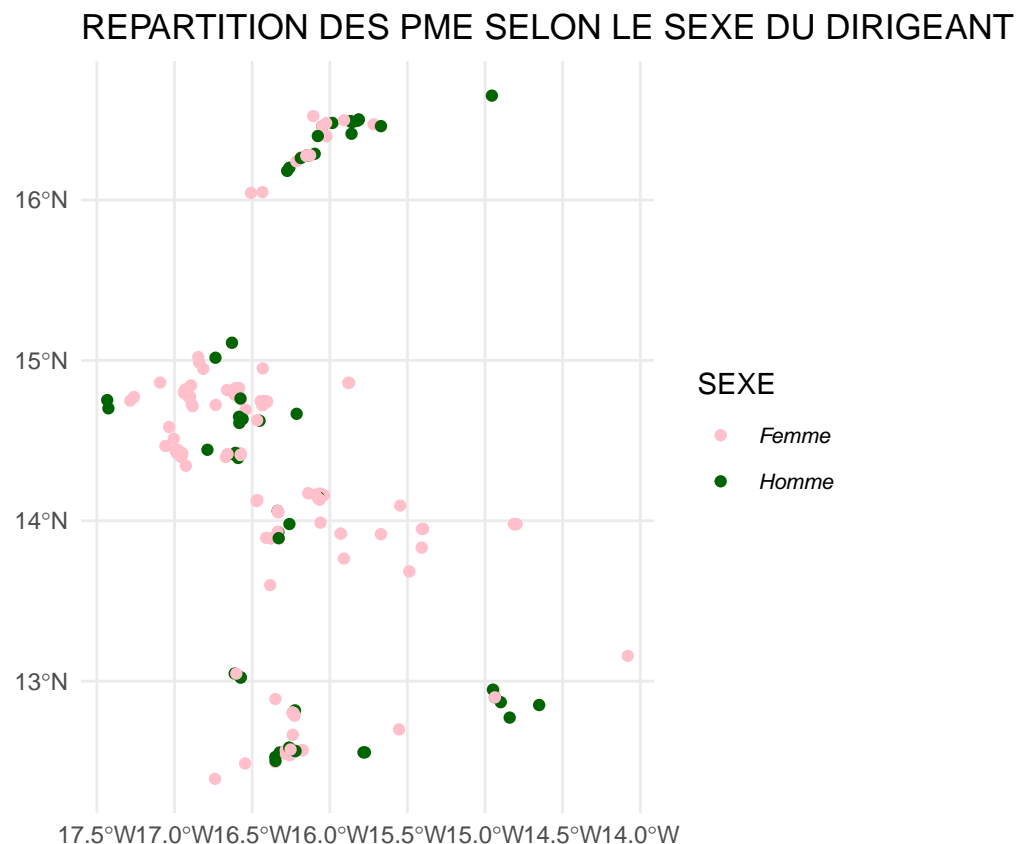
1.3 Un peu de cartographie.

1.3.1 Transformation le data.frame en données géographiques dont l'objet sera nommé projetmap.

```
# Création de l'objet projet_map avec des données géographiques
projet_map <- projet %>%
  st_as_sf(coords = c("gps_menlongitude", "gps_menlatitude"), crs = 4326)
#Ceci provient du referentiel mondiale
```

1.3.2 Représentation spatiale des PME suivant le sexe.

1.3.2.1 Création de la carte spatiale des PME selon le sexe



Pour une meilleure presentation ajoutons la carte du sénégal

1.3.2.2 Chargement des coordonnées géographiques de la région du pays

```
# Chargement des coordonnées géographiques de la région du pays  
occ_raw <- st_read(dsn = "data/sen.gpkg", stringsAsFactors = FALSE)
```

```
## Multiple layers are present in data source C:\Users\HP\Desktop\moi 22-23\sm2\R\Durel_Valdes_NZIALI_TCHAMOU\data\sen.gpkg  
## Use `st_layers` to list all layer names and their type in a data source.  
## Set the `layer` argument in `st_read` to read a particular layer.  
## Reading layer `sen_adm3` from data source  
##   C:\Users\HP\Desktop\moi 22-23\sm2\R\Durel_Valdes_NZIALI_TCHAMOU\data\sen.gpkg'  
##   using driver `GPKG`  
## Simple feature collection with 121 features and 8 fields
```



```
## Geometry type: MULTIPOLYGON
## Dimension: XY
## Bounding box: xmin: -17.53092 ymin: 12.30777 xmax: -11.34801 ymax: 16.69373
## Geodetic CRS: WGS 84

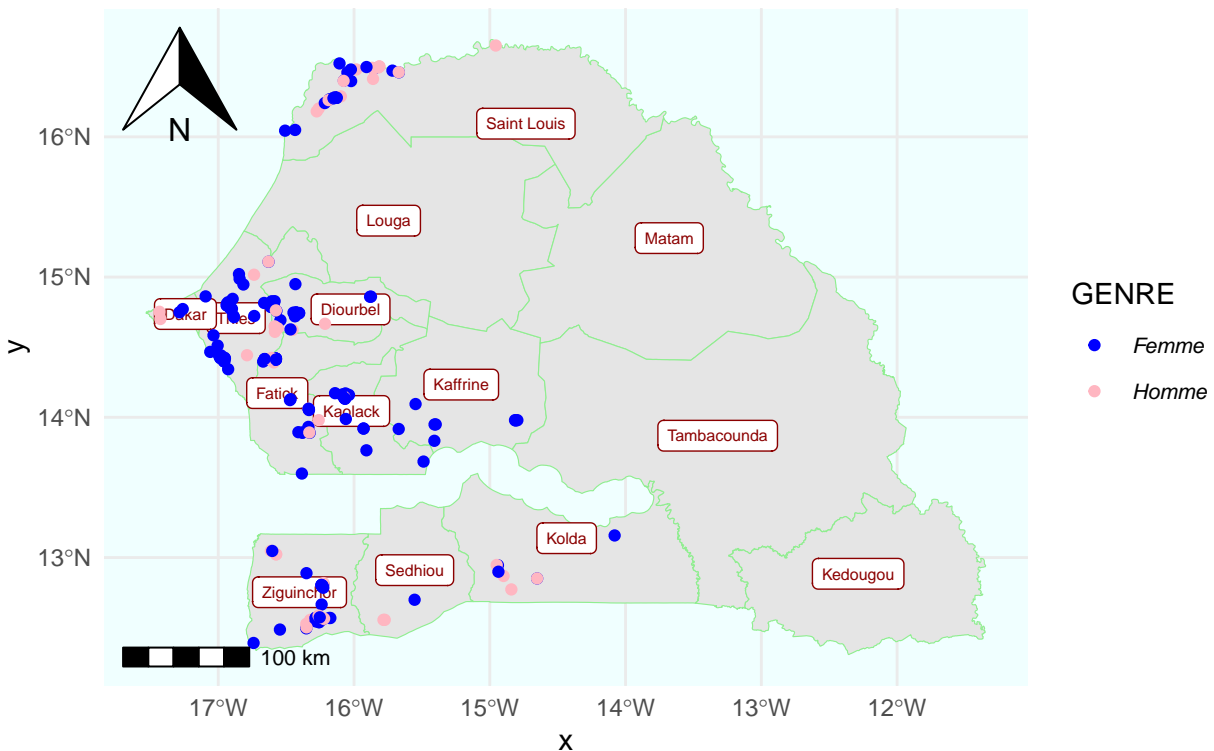
occ <- st_transform(x = occ_raw, crs = 4326)

OCCreg<-occ %>%
  dplyr::select(adm1_id,adm1_name, adm0_id,adm0_name,geom )%>%
  group_by(adm1_id,adm1_name, adm0_id,adm0_name) %>%
  summarise(sum())
```

1.3.2.3 Création de la carte spatiale des PME selon le sexe et la région avec la région en arrière-plan

```
# Création de la carte spatiale des PME selon le sexe et la région
ggplot() +
  geom_sf(data = OCCreg,
    color = "lightgreen") + # Carte de la région en arrière-plan
  geom_sf_label(data = OCCreg, aes(label = adm1_name),
    color = "darkred", size = 2)+
  coord_sf() +
  geom_sf(data = projet_map, aes(color = sexe)) +
  scale_color_manual(
    values = c("blue", "lightpink")) + # Couleurs des catégories de sexe
  labs(color = "GENRE") +
  theme_minimal()+
  theme(legend.text = element_text(face = "italic", size = 8))+
  theme(panel.background = element_rect(fill = "azure",color=NA)) +
  annotation_north_arrow(
    location = "tl", which_north = "true", style = north_arrow_orienteering)+
  annotation_scale(location = "bl", width_hint = 0.2)+
  labs(title = "REPARTITION DES PME SELON LE GENRE DU DIRIGEANT")
```

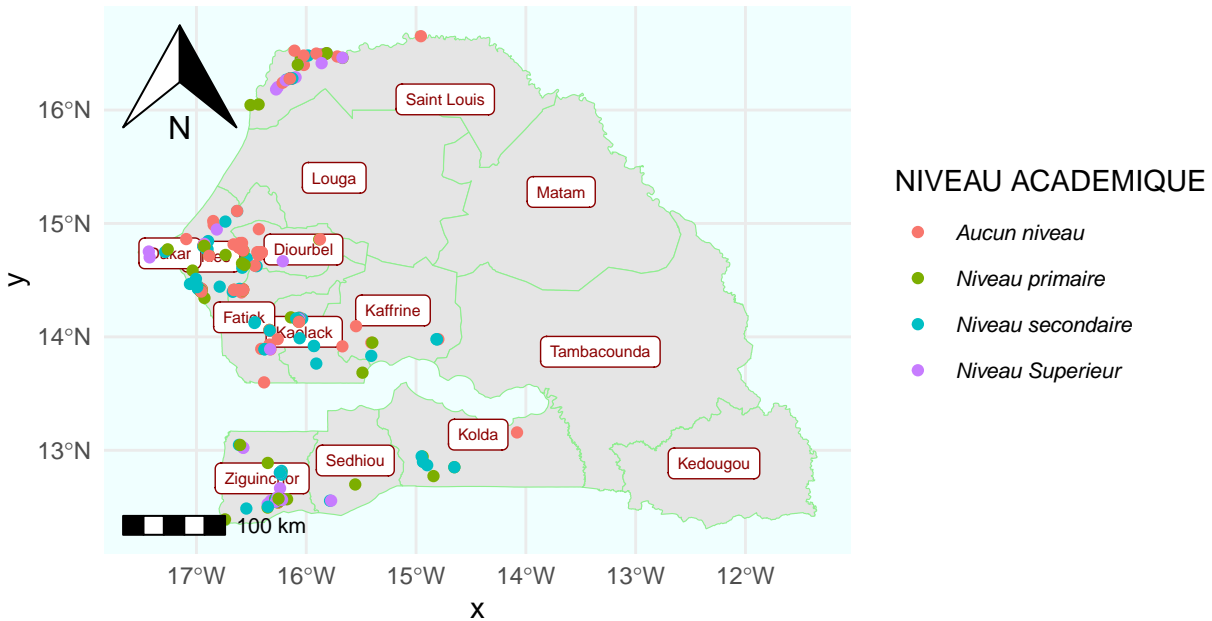
REPARTITION DES PME SELON LE GENRE DU DIRIGEANT



1.3.3 Faites une représentation spatiale des PME suivant le niveau d'instruction.

```
# Création de la carte spatiale des PME selon le sexe
ggplot() +
  geom_sf(data = OCCreg, color = "lightgreen") + # Carte de la région en arrière-plan
  geom_sf_label(data = OCCreg, aes(label = adm1_name), color = "darkred", size = 2) +
  coord_sf() +
  geom_sf(data = projet_map, aes(color = q25)) +
  scale_fill_manual(
    values = c("#E45756", "#4C78A8", "darkgreen", "blue")) + # Couleurs des
  labs(color = "NIVEAU ACADEMIQUE") + #catégories de sex
  theme_minimal() +
  theme(legend.text = element_text(face = "italic", size = 8)) +
  theme(panel.background = element_rect(fill = "azure", color = NA)) +
  annotation_north_arrow(
    location = "tl", which_north = "true", style = north_arrow_orienteering) +
  annotation_scale(location = "bl", width_hint = 0.2) +
  labs(title = "REPARTITION DES PME PAR LE NIVEAU D'INSTRUCTION DES DIRIGEANTS")
```

REPARTITION DES PME PAR LE NIVEAU D'INSTRUCTION DES DIRIGE

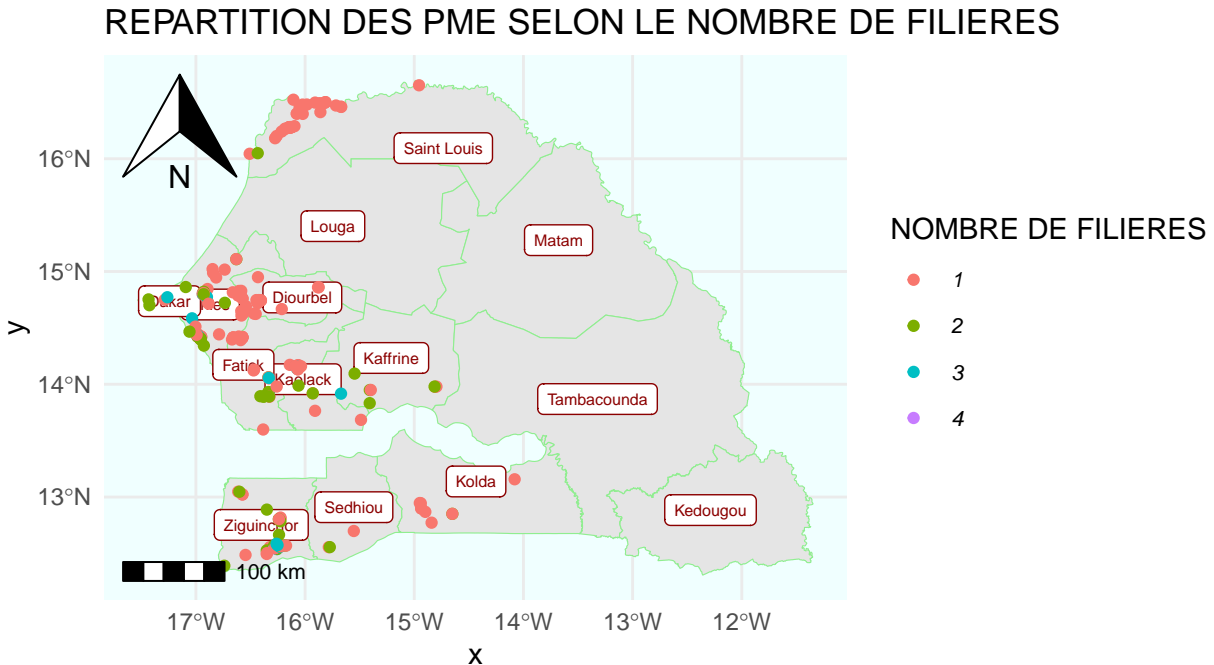


1.3.4 Analyse spatiale de choix.

1.3.4.1 Répartition géographique des PME selon le nombre de filières qu'elle possède

```
# répartition géographique des PME selon les filières
ggplot() +
  geom_sf(data = OCCreg, color = "lightgreen") +
  geom_sf_label(data = OCCreg,
    aes(label = adm1_name), color = "darkred", size = 2) +
  coord_sf() +
  geom_sf(data = projet_map, aes(color = filiere)) +
  scale_fill_manual(values = c("#E45756", "#4C78A8", "darkgreen", "blue")) +
  labs(color = "NOMBRE DE FILIERES") +
  theme_minimal() +
  theme(legend.text = element_text(face = "italic", size = 8)) +
  theme(panel.background = element_rect(fill = "azure", color = NA)) +
  annotation_north_arrow(
    location = "tl", which_north = "true", style = north_arrow_orienteering) +
```

```
annotation_scale(location = "bl", width_hint = 0.2)+
labs(title = "REPARTITION DES PME SELON LE NOMBRE DE FILIERES")
```



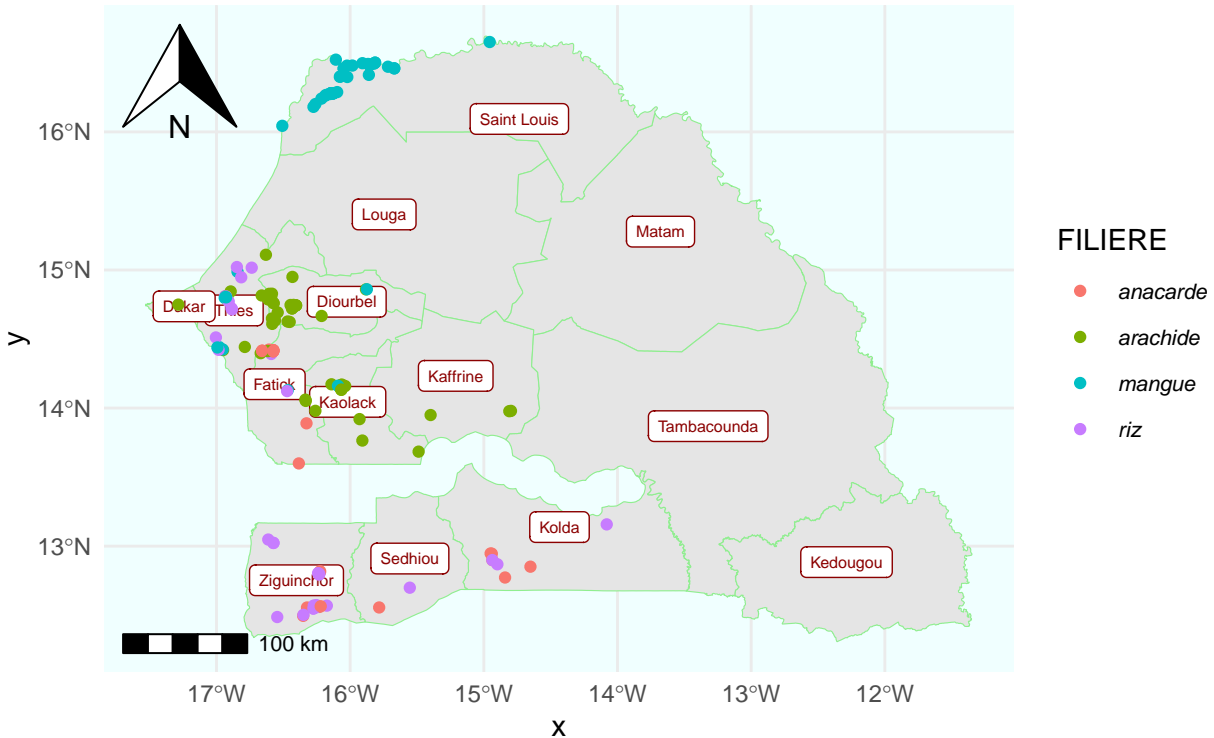
Commençons donc à filtrer notre carte

1.3.4.2 Répartition géographiquedes PME qui font dans une seule culture

```
# répartition géographiquedes PME qui font dans une seule culture
# filtre les PME qui font dans une seule filiere
projet_map1<-projet_map %>% filter(filiere==1)
# Carte de la région en arrière-plan
ggplot() +
  geom_sf(data = OCCreg, color = "lightgreen") +
  geom_sf_label(data = OCCreg,
    aes(label = adm1_name), color = "darkred", size = 2)+
  coord_sf() +
  geom_sf(data = projet_map1, aes(color = nom_filiere)) +
  scale_fill_manual(values = c("#E45756", "#4C78A8", "darkgreen", "blue")) +
  labs(color = "FILIERE") +
  theme_minimal()+
  theme(legend.text = element_text(face = "italic", size = 8))+
```

```
theme(panel.background = element_rect(fill = "azure",color=NA)) +
annotation_north_arrow(
  location = "tl", which_north = "true", style = north_arrow_orientteering)+
annotation_scale(location = "bl", width_hint = 0.2)+
labs(title = "REPARTITION GEOGRAPHIQUE DES PME SELON LES FILIERES")
```

REPARTITION GEOGRAPHIQUE DES PME SELON LES FILIERES



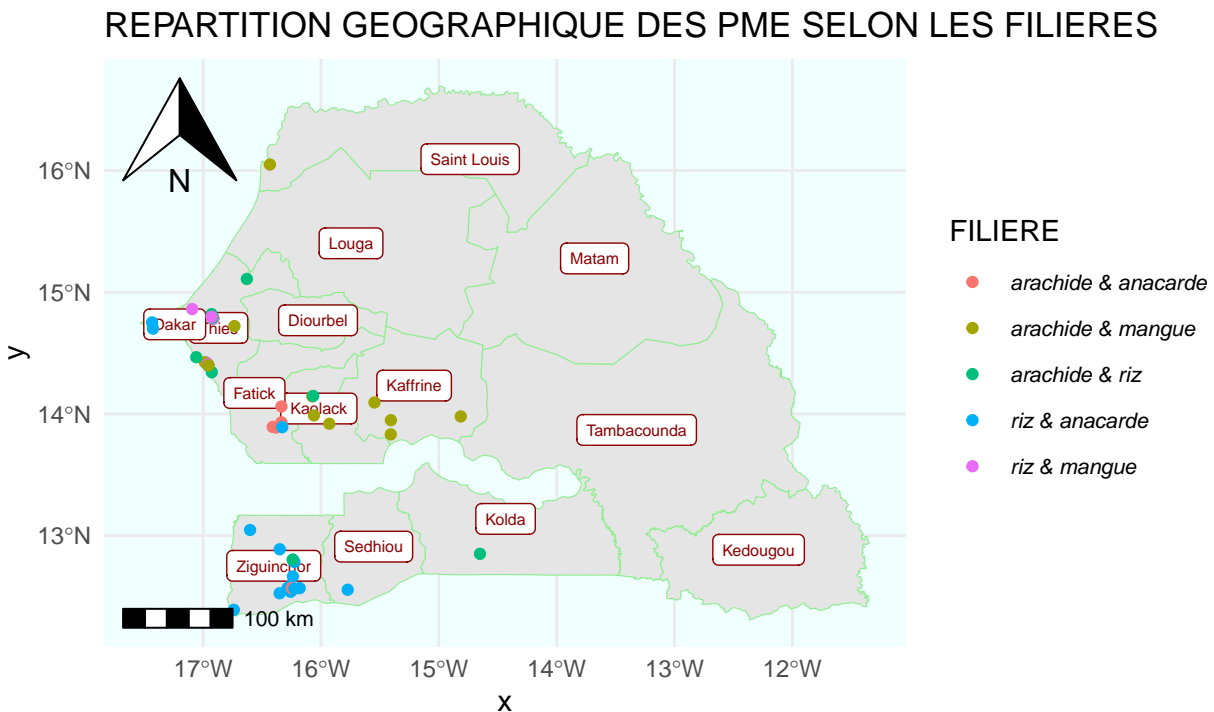
1.3.4.3 Répartition géographique des PME qui font dans deux culture

```
# répartition géographique des PME qui font dans deux cultures
# filtre les PME qui font dans deux filieres
projet_map1<-projet_map %>% filter(filieres==2)
ggplot() +
  geom_sf(data = OCCreg, color = "lightgreen") +
  geom_sf_label(data = OCCreg,
    aes(label = adm1_name), color = "darkred", size = 2)+
  coord_sf() +
  geom_sf(data = projet_map1, aes(color = nom_filieres)) +
  scale_fill_manual(values = c("#E45756", "#4C78A8", "darkgreen", "blue")) +
  labs(color = "FILIERE") +
  theme_minimal()+
  theme(legend.text = element_text(face = "italic", size = 8))+
  theme(panel.background = element_rect(fill = "azure",color=NA)) +
```

```

annotation_north_arrow(
  location = "tl", which_north = "true", style = north_arrow_orienteering)+
annotation_scale(location = "bl", width_hint = 0.2)+
labs(title = "REPARTITION GEOGRAPHIQUE DES PME SELON LES FILIERES")

```



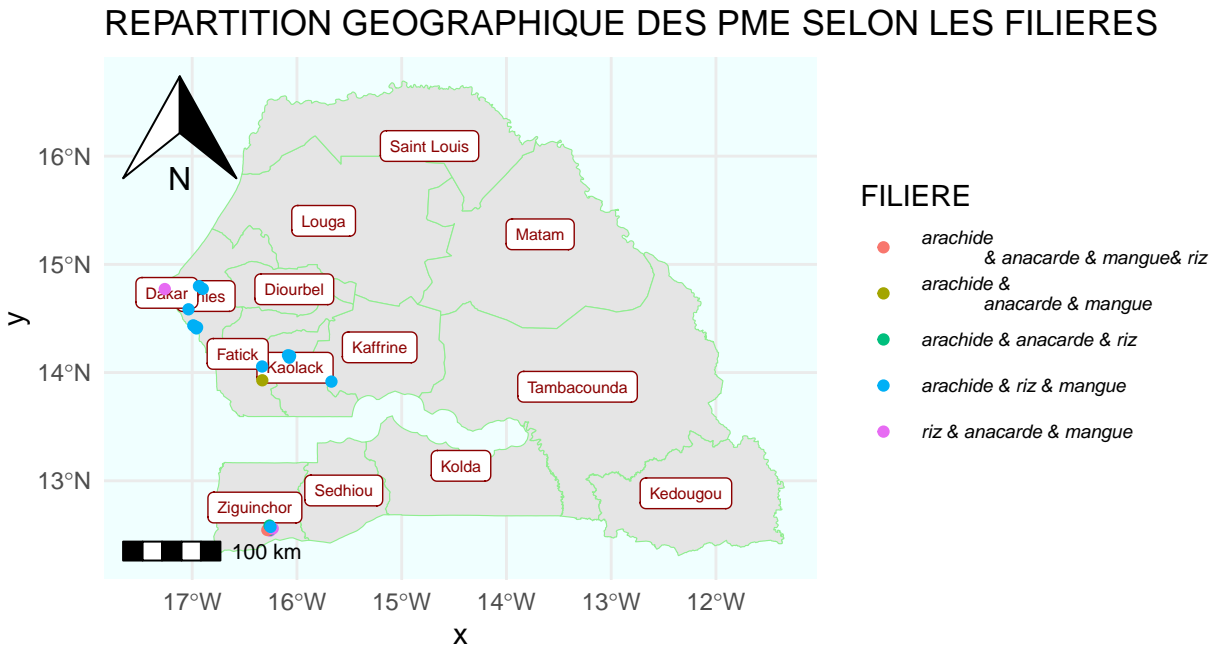
1.3.4.4 Répartition géographique des PME qui font dans trois cultures

```

# répartition géographique des PME qui font dans trois cultures
# filtre les PME qui font dans trois filieres
projet_map1<-projet_map %>% filter(filiere==3| filiere == 4)
# Carte de la région en arrière-plan
ggplot() +
  geom_sf(data = OCCreg, color = "lightgreen") +
  geom_sf_label(data = OCCreg, aes(label = adm1_name), color = "darkred", size = 2)+
  coord_sf() +
  geom_sf(data = projet_map1, aes(color = nom_filiere)) +
  scale_fill_manual(values = c("#E45756", "#4C78A8", "darkgreen", "blue")) +
  labs(color = "FILIERE") +
  theme_minimal()+theme(legend.text = element_text(face = "italic", size = 7))+
  theme(panel.background = element_rect(fill = "azure",color=NA)) +
  annotation_north_arrow(location = "tl", which_north = "true", style = north_arrow_orienteering)+
  annotation_scale(location = "bl", width_hint = 0.2)+

```

```
labs(title = "REPARTITION GEOGRAPHIQUE DES PME SELON LES FILIERES")
```



1.3.4.5 Évaluation des PME de chaque filière qui sont desservies par des routes bitumées et examen de l'état de ces routes.

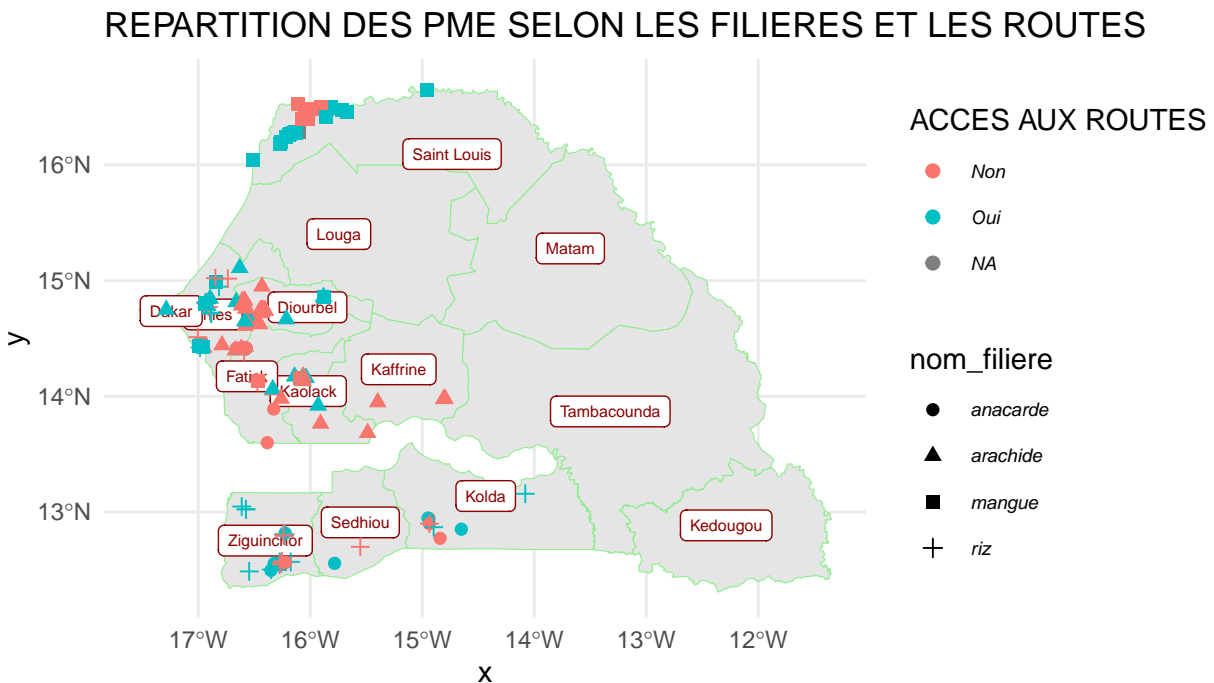
1.3.4.5.1 Répartition géographique des PME selon les filières uniques desservies par des routes bitumées

```
# filtre les PME qui font dans une seule filiere
projet_map1 <- projet_map %>% filter(filiere == 1)

# Création de la carte avec routes bitumées

ggplot() +
  geom_sf(data = OCCreg, color = "lightgreen") +
  geom_sf_label(data = OCCreg,
    aes(label = adm1_name), color = "darkred", size = 2) +
  geom_sf(data = projet_map1,
    aes(fill = nom_filiere, shape = nom_filiere, color = q16), size = 2) +
  labs(title = "REPARTITION DES PME SELON LES FILIERES ET LES ROUTES", size = 2,
```

```
color = "ACCES AUX ROUTES") +
theme_minimal()+
theme(legend.text = element_text(face = "italic", size = 7))# Mise en forme
```



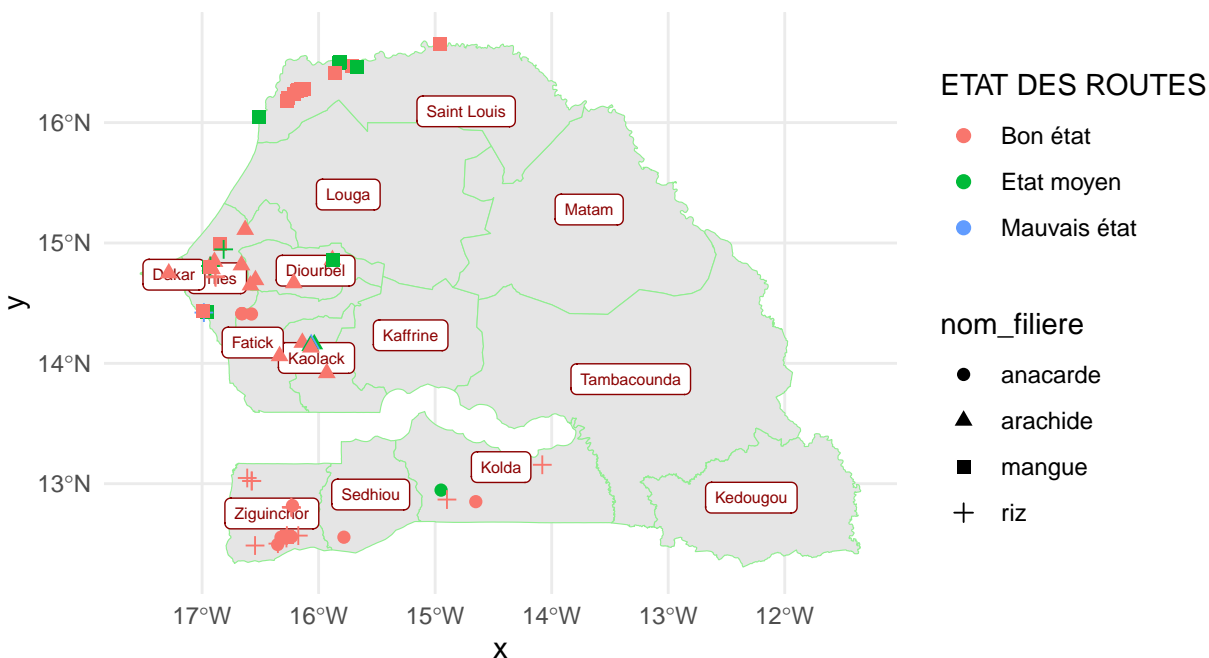
de la légende en italique et réduction de la taille

1.3.4.5.2 Création de la carte avec l'état des routes bitumées

Création de la carte avec l'état des routes bitumées

```
projet_map1<-projet_map1 %>% filter(q16=="Oui")
ggplot() +
  geom_sf(data = OCCreg, color = "lightgreen") +
  geom_sf_label(data = OCCreg, aes(label = adm1_name), color = "darkred", size = 2) +
  geom_sf(data = projet_map1, aes(fill = nom_filiere, shape= nom_filiere, color = q17) , size=2)+
  labs(title = "REPARTITION DES PME SELON LES FILIERES ET L'ETAT DES ROUTES",size = 2,
        color = "ETAT DES ROUTES") +
  theme_minimal()
```


REPARTITION DES PME SELON LES FILIERES ET L'ETAT DES ROUTE



2 Partie 2.

Le fichier excel Base_Partie 2.xlsx contient un ensemble de données artificielles créé dans le cadre de ce projet. La première feuille contient des micro-données au niveau individuel des répondants à l'enquête, la deuxième feuille contient des données pour les districts dans lesquels les répondants de la première feuille ont été interrogés. La troisième feuille contient des explications sur les variables incluses dans toutes les feuilles précédentes.

- **Chargement de données**

```
data <- read_excel("data/Base_Partie 2.xlsx",
  sheet = "data")

district <- read_excel("data/Base_Partie 2.xlsx",
  sheet = "district")
```

```
codebook <- read_excel("data/Base_Partie 2.xlsx",
  sheet = "codebook")
```

2.1 Nettoyage et gestion des données.

2.1.1 Renommer la variable country destination en destination, et définition les valeurs négatives comme manquantes

```
# Renommer les variables country_destination
data <- data %>%
  mutate(country_destination=ifelse(country_destination>=0,
                                     country_destination,NA)) %>%
  rename(destination = country_destination)
```

2.1.2 Créer une nouvelle variable contenant des tranches d'âge de 5 ans en utilisant la variable age.

```
data<- data %>%
  mutate(age=ifelse(age==999, NA, age),
         Age_group= cut(age,seq(14, by=5, length.out=7)))

tab11=data %>%
  dplyr::select(Age_group) %>%
  tbl_summary() %>%
  bold_labels()

tab11%>% bold_labels() %>%
  italicize_levels() %>%
  modify_header(update = list(
    label ~ "**VARIABLE**", all_stat_cols(
      stat_0 = FALSE) ~ "**{level}** (n={n}, {style_percent(p)})%"
  )) %>% as_flex_table() %>%
  fontsize(size = 8) %>%
  width(width = 1.1)
```

VARIABLE	N = 97 ¹
Age_group	
(14,19]	16 (17%)
(19,24]	34 (35%)
(24,29]	23 (24%)
(29,34]	13 (14%)
(34,39]	6 (6.3%)
(39,44]	4 (4.2%)
Unknown	1
¹ _n (%)	

2.1.3 Créer une nouvelle variable contenant le nombre d'entretiens réalisés par chaque agent recenseur.

```
data<-data %>%
  mutate(enumerator=as.factor(enumerator))
# Création du tableau
tab12=data %>%
  dplyr::select(enumerator) %>%
    rename("Numero de l'agent recenseur"=enumerator) %>%
    tbl_summary(sort = list(everything() ~ "frequency")) %>% bold_labels()

tab12_kable <- as_kable(tab12) %>%
  kable_paper(full_width = TRUE)%>%
  row_spec(0:1, background = "green", color = "black") %>%
  row_spec(2:3, background = "#90EE90", color = "black") %>%
  row_spec(4:12, background = "#FFFFE0", color = "black")

# Afficher la table avec les deux premières lignes colorées en vert
tab12_kable
```

Characteristic	**N = 97**
Numero de l'agent recenseur	
4	9 (9.3%)
20	9 (9.3%)
13	8 (8.2%)
7	7 (7.2%)
11	7 (7.2%)
5	6 (6.2%)
8	6 (6.2%)
9	6 (6.2%)
14	6 (6.2%)
17	6 (6.2%)
18	6 (6.2%)
1	5 (5.2%)
6	5 (5.2%)
10	5 (5.2%)
12	5 (5.2%)
15	1 (1.0%)

il apparait que se sont les resseurs 4 et 20 qui ont le plus grand nombre de recensement soit 9 recensements

2.1.3.1 Créer une nouvelle variable qui affecte aléatoirement chaque répondant à un groupe de traitement (1) ou de controle (0).

```
data<-data %>%
  mutate(traitement=sample(c(0,1), 97, replace = TRUE))
```

```
# jointure a gauche
d<- data_merge(data, district, join = "left", id = district)
```

2.1.3.2 Fusionner la taille de la population de chaque district (feuille 2) avec l'ensemble de données (feuille 1) afin que toutes les personnes interrogées aient une valeur correspondante représentant la taille de la population du district dans lequel elles vivent.

2.1.4 Calculer la durée de l'entretien et indiquer la durée moyenne de l'entretien par enquêteur.

```
#calcul de la duree
d<-d %>%
  mutate(duree=endtime-starttime)
```

```
#calcul de la duree moyenne par enqueteur

d %>%
  group_by(enumerator) %>%
    rename("Numero de l'agent recenseur"=enumerator) %>%
    summarise(dureemoy = mean(duree), nbre_enquete = n()) %>%
    kbl() %>%
    kable_paper(full_width = TRUE) %>%
    row_spec(0:1, background = "green",color="black") %>%
    row_spec(1:16, background = "lightgray",color="black")
```

Numero de l'agent recenseur	dureemoy	nbre_enquete
1	68.14667 mins	5
4	36.48333 mins	9
5	33.55833 mins	6
6	25.84667 mins	5
7	37.16429 mins	7
8	40.13056 mins	6
9	114.76667 mins	6
10	55.27667 mins	5
11	33.48333 mins	7
12	48.16667 mins	5
13	31.59583 mins	8
14	25.56111 mins	6
15	28.65000 mins	1
17	29.28611 mins	6
18	36.85833 mins	6
20	28.76852 mins	9

2.1.4.1 Renommez toutes les variables de l'ensemble de données en ajoutant le préfixe “endline_” à l'aide d'une boucle.}}

```
d1<-d
#nom des colonnes
colon<-colnames(d)

# Nouveaux noms de variables avec le préfixe "endline_"
new_col <- paste0("endline_", colon)

# Renommage des variables à l'aide d'une boucle
for (i in seq_along(colon)) {
  colnames(d)[i] <- new_col[i]
}
```

2.1.4.2 Renommez toutes les variables de l'ensemble de données en ajoutant le préfixe “endline_” avec lapply

```
noms_nouveaux <- lapply(names(d1), function(col) paste0("endline_", col))
names(d1) <- unlist(noms_nouveaux)
d1 %>%
names()

## [1] "endline_id"          "endline_starttime"   "endline_endtime"
## [4] "endline_enumerator" "endline_district"    "endline_age"
## [7] "endline_sex"         "endline_children_num" "endline_intention"
## [10] "endline_destination" "endline_Age_group"   "endline_traitement"
## [13] "endline_population"  "endline_duree"
```

2.2 Analyse et visualisation des données.

2.2.1 Créez un tableau récapitulatif contenant l'âge moyen et le nombre moyen d'enfants par district.

```
data %>%
  group_by(district) %>%
    summarise(age_moyen = round(mean(age, na.rm=TRUE),2),
              enfants_moyen = round(mean(children_num, na.rm=TRUE),2)) %>%
  kbl() %>%
    kable_paper(full_width = TRUE) %>%
    row_spec(0:1, background = "green",color="black") %>%
    row_spec(1:8, background = "lightgray",color="black")
```

	district	age_moyen	enfants_moyen
	1	29.62	1.50
	2	26.62	0.85
	3	26.12	0.00
	4	26.00	0.00
	5	24.33	0.50
	6	23.15	0.12
	7	28.00	0.17
	8	24.64	1.27

2.3 Les tests

- Testez si la différence d'âge entre les sexes est statistiquement significative au niveau de 5 %.
Il est nécessaire de faire un test d'annova afin de savoir comment définir notre test de significativité

2.3.1 Annova

```
#test d'égalité des variance
mod <- anova(lm(age ~ sex, data))
bind_rows(
  broom::tidy(mod)
)%>%
  kbl() %>%
    kable_paper(full_width = TRUE) %>%
    row_spec(0:1, background = "green",color="black") %>%
    row_spec(1:2, background = "lightgray",color="black")
```

term	df	sumsq	meansq	statistic	p.value
sex	1	142.5012	142.50121	3.55781	0.062353
Residuals	94	3764.9884	40.05307	NA	NA

donc les varainces sont égales

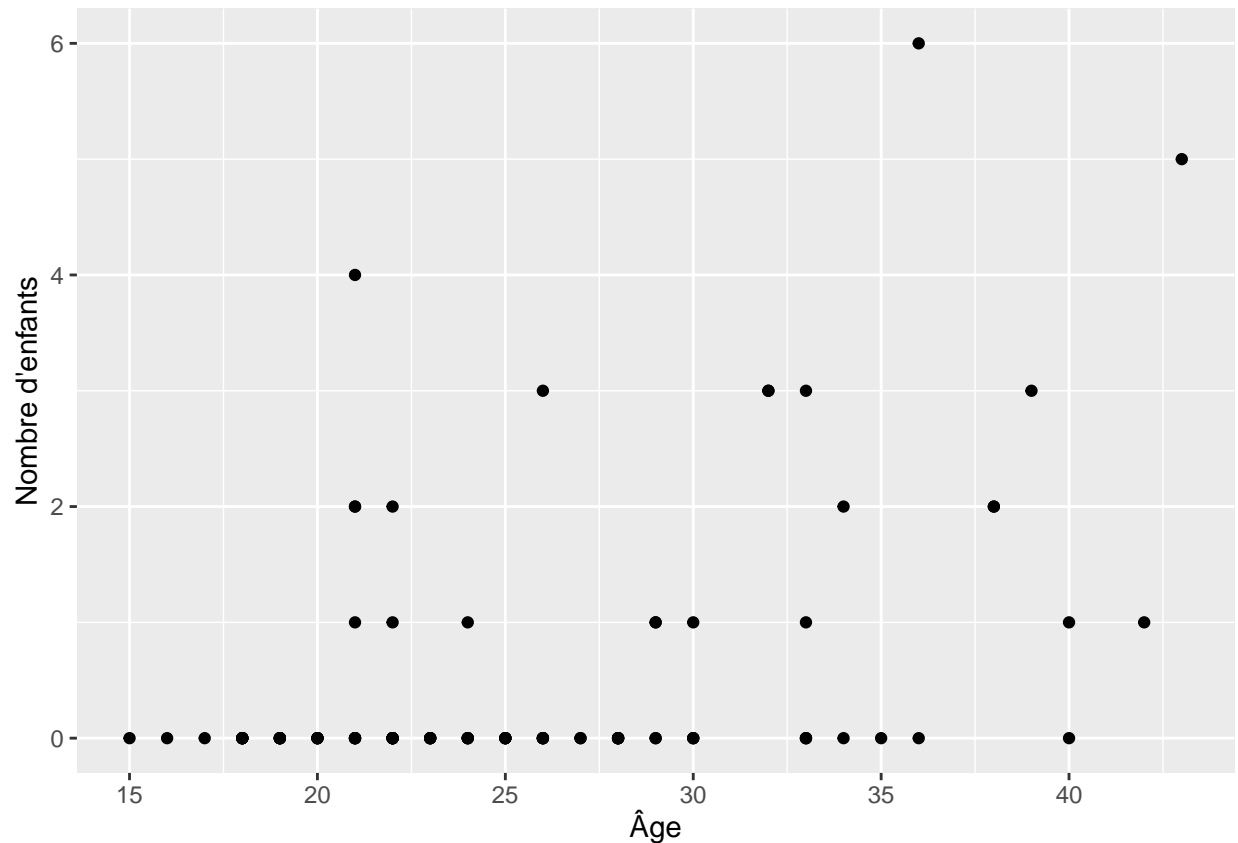
2.3.2 Test de student

```
#test de student
mod<-t.test(age ~ sex, data, var.equal = TRUE)
bind_rows(
  broom::tidy(mod)
)%>%
  kbl() %>%
    kable_paper(full_width = TRUE) %>%
    row_spec(0:1, background = "green",color="black") %>%
    row_spec(1:1, background = "lightgray",color="black")
```

estimate	estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	method	alternative
3.988372	25.98837	22	1.886216	0.062353	94	-	8.186728	Two	two.sided
						0.2099843		Sample	
								t-test	

2.3.3 Créer un nuage de points de l'âge en fonction du nombre d'enfants

```
ggplot(data, aes(x = age, y = children_num)) +
  geom_point() +
  labs(x = "Âge", y = "Nombre d'enfants")
```



2.3.4 Effet appartenance et intention

2.3.4.1 La variable “intention” indique si les migrants potentiels ont l’intention de migrer sur une échelle de 1 à 7. Estimez l’effet de l’appartenance au groupe de traitement sur l’intention de migrer.

```
modele_traitement<-lm(intention ~ traitement-1, data)
effet_traitement <- coef(modele_traitement)["traitement"]

mod <- summary(modele_traitement)
bind_rows(
  broom::tidy(mod)
)%>%
  kbl() %>%
  kable_paper(full_width = TRUE) %>%
  row_spec(0:1, background = "green",color="black") %>%
  row_spec(1:1, background = "lightgray",color="black")
```

term	estimate	std.error	statistic	p.value
traitement	2	0.2926911	6.833142	0

2.3.5 Présentation des modèles

Crée un tableau de régression avec 3 modèles. La variable de résultat est toujours “intention”. Modèle A : Modèle vide - Effet du traitement sur les intentions. Modèle B : Effet du traitement sur les intentions en tenant compte de l’âge et du sexe. Modèle C : Identique au modèle B mais en contrôlant le district. Les résultats des trois modèles doivent être affichés dans un seul tableau.

```
modele_a <- lm(intention ~ traitement-1, data)
modele_b <- lm(intention ~ traitement -1+ age + sex, data )
modele_c <- lm(intention ~ traitement -1+ age + sex + district, data)

tableau_regression <- bind_rows(
  broom::tidy(modele_a) ,
  broom::tidy(modele_b) ,
  broom::tidy(modele_c)
)%>%
mutate(
  estimate = round(estimate, 3),
  std.error = round(std.error, 3),
  statistic = round(statistic, 3),
  p.value = round(p.value, 3)
)

tableau_regression %>%
  kbl() %>%
  kable_paper(full_width = TRUE) %>%
  row_spec(0:1, background = "green",color="black") %>%
  row_spec(1:2, background = "lightgray",color="black") %>%
  row_spec(2:5, background = "lightgray",color="black") %>%
  row_spec(5:8, background = "lightgray",color="black") %>%
  group_rows(group_label = "Modèle A", start_row =1, end_row =2) %>%
  group_rows(group_label = "Modèle B", start_row =3, end_row =5) %>%
  group_rows(group_label = "Modèle C", start_row =6, end_row =8) %>%
  kable_styling() %>%
  column_spec(1, bold = TRUE) %>%
  column_spec(5, color = c(tableau_regression$p_value_color))
```

term	estimate	std.error	statistic	p.value
Modèle A				
traitement	2.000	0.293	6.833	0.000
traitement	0.182	0.348	0.522	0.603
Modèle B				
age	0.076	0.010	7.402	0.000
sex	-0.474	0.591	-0.802	0.424
traitement	0.013	0.345	0.038	0.970
Modèle C				
age	0.053	0.013	3.975	0.000
sex	-0.514	0.576	-0.892	0.375
district	0.166	0.067	2.485	0.015

3 Partie 3

Voici le lien permettant d'accéder à https://nziali-durel.shinyapps.io/Durel_Valdes_NZIALI_TCHAMOU/
 Cette application Shiny est une interface interactive permettant de visualiser et d'explorer des données d'événements en Afrique de l'Ouest à partir du fichier CSV "ACLED-Western_Africa.csv" et des données géographiques des pays au format .shp.

L'application est composée de trois onglets différents :

3.1 Onglet "ACLED Western Africa" (Application 1)

Panel "Carte du monde par événements" Dans ce panel, une carte du monde est affichée avec des marqueurs représentant différents événements en Afrique de l'Ouest. Sur la barre latérale gauche, vous pouvez filtrer les types d'événements à afficher en cochant ou en décochant les événements dans la liste. Après avoir fait une sélection, vous pouvez cliquer sur le bouton "Filtrer" pour mettre à jour la carte avec les marqueurs des événements choisis. Chaque marqueur affiche des informations telles que le pays, le type d'événement et l'année.

Panel "Filtrage des événements" Dans ce panel, vous pouvez filtrer les événements en fonction du pays, du type d'événement et de l'année. Utilisez les menus déroulants pour choisir un pays, un type d'événement et une plage d'années. En fonction des filtres appliqués, une carte filtrée et un boxplot par événement du pays sélectionné seront affichés. Le boxplot montre la distribution des années d'événements pour chaque type d'événement du pays sélectionné.

3.2 Onglet "ACLED Western Africa" (Application 2)

Dans cet onglet, vous pouvez choisir un ou plusieurs pays et un type d'événement à partir des menus déroulants. Une carte interactive est affichée, mettant en évidence les événements dans les pays sélectionnés avec des marqueurs rouges. Vous pouvez également réinitialiser la carte en utilisant le bouton "Réinitialiser la carte". Sous l'onglet "Statistiques", un tableau présente des informations statistiques sur le pays sélectionné, telles que le nombre total d'événements, l'année la plus ancienne, l'année la plus récente et les types d'événements les plus et les moins fréquents.

3.3 Onglet "ACLED Western Africa" (Application 3)

Cet onglet permet de choisir un pays et/ou un type d'événement à partir de menus déroulants. Une carte interactive est affichée, avec des marqueurs colorés pour chaque type d'événement, et un tableau statistique présente les pourcentages des types d'événements par rapport au pays sélectionné et par rapport à tous les pays. Vous pouvez également voir un tableau détaillant les événements filtrés en fonction de vos sélections.

En résumé, cette application Shiny offre une manière conviviale d'explorer et d'analyser les données d'événements en Afrique de l'Ouest, en fournissant des cartes interactives et des visualisations statistiques pour différentes sélections de pays et de types d'événements. L'utilisateur peut filtrer et explorer les données pour obtenir des informations pertinentes sur les événements dans la région.

Conclusion

Ce projet d'étude de cas avait pour objectif d'identifier et de caractériser des bioénergies durables pour les petites et moyennes entreprises (PME) agroalimentaires d'Afrique de l'Ouest. Pour ce faire, nous avons effectué une préparation des données en important un fichier Excel contenant 250 observations et 33 variables, puis en sélectionnant les variables pertinentes pour notre analyse. Nous avons également créé de nouvelles variables pour faciliter l'analyse.

Dans la partie 1, nous avons réalisé des analyses descriptives pour comprendre la répartition des PME en fonction de différentes variables telles que le sexe, le niveau d'instruction, le statut juridique, etc. Nous avons utilisé le package `gtsummary` pour présenter les résultats sous forme de tableaux concis et informatifs. La partie 2 portait sur la gestion et l'analyse des données d'une deuxième base de données contenant des micro-données individuelles et des données de district. Nous avons nettoyé les données en renommant les variables, créé de nouvelles variables pertinentes et fusionné des informations sur la taille de la population des districts. Ensuite, nous avons réalisé des analyses et visualisations pour explorer les relations entre différentes variables, comme l'âge et le nombre d'enfants, l'intention de migrer en fonction de l'appartenance à un groupe de traitement, etc. Les résultats ont été présentés à l'aide de tableaux et de graphiques. Enfin, dans la partie 3, nous avons développé une application Shiny pour visualiser les événements politiques et les manifestations en Afrique de l'Ouest par pays, type, année et localisation géographique.

Ce projet a permis d'appliquer les connaissances en statistiques et en manipulation de données à une étude de cas réelle, en utilisant le langage de programmation R et ses packages. Les analyses descriptives, les visualisations et l'application interactive ont fourni des informations pertinentes pour la caractérisation des bioénergies durables et pour comprendre les dynamiques des événements politiques en Afrique de l'Ouest.

En somme, ce projet a été une occasion d'approfondir nos compétences en R, de traiter des données réelles et de présenter les résultats de manière claire et concise. Il démontre également l'importance de l'analyse de données dans la prise de décisions éclairées et la résolution de problèmes dans des domaines variés.