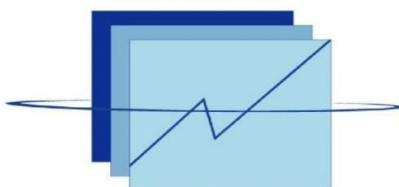




REPUBLIQUE DU SENEGAL

UN PEUPLE- UN BUT –UNE FOIE

MINISTERE DE L'ECONOMIE, DES FINANCES ET



ANSD

AGENCE NATIONALE DE LA STATISTIQUE ET DE LA DEMOGRAPHIE

ECOLE NATIONALE DE LA STATISTIQUE ET DE
L'ANALYSE ECONOMIQUE Pierre Ndiaye

Projet final - Logiciel statistique R

Rédigé par:

MOSSE Isabelle Danielle

Elèves ingénieurs statisticiens économistes

Sous la supervision de :

M. HEMA Aboubacar

Research Analyst

Contents

1	partie 1	3
1.1	Préparation des données	3
1.1.1	Importation et mise en forme	3
1.1.2	Création de variables	4
1.2	Analyses descriptives	4
1.3	cartographie	9
1.4	autre analyse spaciae	11
2	partie 2	12
2.1	Nettoyage et gestion des données	12
2.2	Analyse et visualisation des données	14
3	partie 3	17

1 partie 1

1.1 Préparation des données

1.1.1 Importation et mise en forme

```
##Importation et mise en forme##
projet <- read_excel("Base_Partie 1.xlsx",
  sheet = "sheet1",
  range = NULL,
  col_names = TRUE,
  col_types = NULL,
  na = "")

# Sélection les variables
variables <- select(projet, matches("^[qfgst]"))

#Résumé des valeurs manquantes par variable
val_manq <- variables %>%
summarise_all(~ sum(is.na(.)))

#tableau qui resume les valeurs manquantes par variable
# Calcul du nombre total de valeurs manquantes par variable
table_manquantes <- data.frame(
  Variable = names(projet),
  `Nombredevaleursmanquante` = colSums(is.na(projet))
)
# tableau "gtsummary" avec les variables et le nombre de valeurs manquantes
table_manquantes_table <-
  gt(table_manquantes)
table_manquantes_table
```

Variable	Nombredevaleursmanquante
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

```
# Vérifier les valeurs manquantes pour la variable "key"
valeurs_manquantes_key <- subset(projet, is.na(key))
if (nrow(valeurs_manquantes_key) > 0) {
  pmes_concernees <- valeurs_manquantes_key$key
  cat("Les PME concernées avec des valeurs manquantes pour 'key' sont :", pmes_concernees)
} else {
  cat("Il n'y a pas de valeurs manquantes pour la variable 'key'.")
}
```

Il n'y a pas de valeurs manquantes pour la variable 'key'.

1.1.2 Création de variables

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

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

# Sélectionner les variables key et les variables correspondantes dans un nouveau data.frame
langues <- projet %>%
  select(key, starts_with("q24a"))

# Créer la variable parle
langues$parle <- rowSums(langues[, -1])

# Sélectionner les variables key et parle
langues <- select(langues, key, parle)
# Fusionner les data.frames projet et langues
projet_r <- merge(projet, langues, by = "key")
```

1.2 Analyses descriptives

```
#labelisation des variables
var_label(projet_r$q25) <- "Niveau d'instruction"
var_label(projet_r$q12) <- "Statut juridique"
var_label(projet_r$q81) <- "Locataire/propriétaire"
```

```

#précisons le français comme langue de sortie
theme_gtsummary_language("fr", decimal.mark = ",", big.mark = " ")

#tableau récapitulatif des statistiques descriptives
tab1 <-projet_r %>%
tbl_summary(
include = c(sexe, q12, q25, q81),
statistic = all_categorical() ~ "{p} % ({n}/{N})",
)%>%
bold_labels()
## repartition des pme suivant le statut juridique et le sexe
tab2 <- projet_r %>%
tbl_summary(
include = c(sexe, q25, q12, q81),
by = sexe,
statistic = all_categorical() ~ "{p}% ({n}/{N})",
percent = "row"
)%>%
add_overall()%>%
modify_header(
all_stat_cols() ~ "**{level}** ({n} obs.)"
) %>%
bold_labels()
#combinons les tables tab1 et tab2
gtsummary::tbl_merge(
list(tab1, tab2),
tab_spanner = c("statistiques univarié", "statistiques bivariées")
)## intitulé des groupes de tableau associés

```

Caractéristique	N = 250	Total (250 obs.)	Femme (191 obs.)	Homme (59 obs.)
sexe				
Femme	76 % (191/250)			
Homme	24 % (59/250)			
Statut juridique				
Association	2,4 % (6/250)	100% (6/6)	50% (3/6)	50% (3/6)
GIE	72 % (179/250)	100% (179/179)	83% (149/179)	17% (30/179)
Informel	15 % (38/250)	100% (38/38)	84% (32/38)	16% (6/38)
SA	2,8 % (7/250)	100% (7/7)	14% (1/7)	86% (6/7)
SARL	5,2 % (13/250)	100% (13/13)	15% (2/13)	85% (11/13)
SUARL	2,8 % (7/250)	100% (7/7)	57% (4/7)	43% (3/7)
Niveau d'instruction				
Aucun niveau	32 % (79/250)	100% (79/79)	89% (70/79)	11% (9/79)
Niveau primaire	22 % (56/250)	100% (56/56)	86% (48/56)	14% (8/56)
Niveau secondaire	30 % (74/250)	100% (74/74)	76% (56/74)	24% (18/74)
Niveau Supérieur	16 % (41/250)	100% (41/41)	41% (17/41)	59% (24/41)
Locataire/propriétaire				
Locataire	9,6 % (24/250)	100% (24/24)	67% (16/24)	33% (8/24)
Propriétaire	90 % (226/250)	100% (226/226)	77% (175/226)	23% (51/226)

```

#statistique descriptive de notre choix sur les autres variables
#analyse par filiere
var_label(projet_r$q24) <- "Age du dirigeant"
var_label(projet_r$filiere_1) <- "filiere arachide"
var_label(projet_r$filiere_2) <- "filiere anacarde"
var_label(projet_r$filiere_3) <- "filiere mangue"
var_label(projet_r$filiere_4) <- "filiere riz"
theme_gtsummary_language("fr", decimal.mark = ",", big.mark = " ")

projet_r %>%
tbl_summary(
  include = c(q24, region, departement, parle),
  type = q24 ~ "continuous2",
  by = sexe,
  statistic = list(
    all_continuous2() ~ c("{median} ({p25} - {p75})", "{mean} ({sd})", "{min} - {max}")
  )
)%>%
add_overall()%>%
bold_labels()

```

Caractéristique	Total, N = 250	Femme, N = 191	Homme, N = 59
Age du dirigeant			
Médiane (EI)	55 (45 - 62)	56 (45 - 63)	50 (40 - 59)
Moyenne (ET)	3 106 217 (49 112 065)	4 065 708 (56 187 733)	65 (124)
Étendue	18 - 776 530 031	18 - 776 530 031	26 - 999
region			
Dakar	1 (0,4%)	0 (0%)	1 (1,7%)
Diourbel	34 (14%)	28 (15%)	6 (10%)
Fatick	30 (12%)	23 (12%)	7 (12%)
Kaffrine	8 (3,2%)	8 (4,2%)	0 (0%)
Kaolack	21 (8,4%)	20 (10%)	1 (1,7%)
Kolda	9 (3,6%)	4 (2,1%)	5 (8,5%)
Saint-Louis	42 (17%)	22 (12%)	20 (34%)
Sédhiou	4 (1,6%)	1 (0,5%)	3 (5,1%)
Thiès	51 (20%)	48 (25%)	3 (5,1%)
Ziguinchor	50 (20%)	37 (19%)	13 (22%)
departement			
Bambey	20 (8,0%)	15 (7,9%)	5 (8,5%)
Bignona	13 (5,2%)	9 (4,7%)	4 (6,8%)
Birkelane	1 (0,4%)	1 (0,5%)	0 (0%)
Dagana	39 (16%)	20 (10%)	19 (32%)
Diourbel	12 (4,8%)	11 (5,8%)	1 (1,7%)
Fatick	15 (6,0%)	12 (6,3%)	3 (5,1%)
Foundiougne	15 (6,0%)	11 (5,8%)	4 (6,8%)
Goudomp	3 (1,2%)	0 (0%)	3 (5,1%)
Kaffrine	4 (1,6%)	4 (2,1%)	0 (0%)
Kaolack	16 (6,4%)	15 (7,9%)	1 (1,7%)
Kolda	7 (2,8%)	3 (1,6%)	4 (6,8%)
Koungheul	3 (1,2%)	3 (1,6%)	0 (0%)
Mbacké	2 (0,8%)	2 (1,0%)	0 (0%)
Mbour	22 (8,8%)	21 (11%)	1 (1,7%)
Médina Yoro Foul	1 (0,4%)	0 (0%)	1 (1,7%)
Nioro	5 (2,0%)	5 (2,6%)	0 (0%)

Caractéristique	Total, N = 250	Femme, N = 191	Homme, N = 59
Oussouye	2 (0,8%)	2 (1,0%)	0 (0%)
Podor	1 (0,4%)	0 (0%)	1 (1,7%)
Rufisque	1 (0,4%)	0 (0%)	1 (1,7%)
Saint-Louis	2 (0,8%)	2 (1,0%)	0 (0%)
Sédhiou	1 (0,4%)	1 (0,5%)	0 (0%)
Thiès	23 (9,2%)	23 (12%)	0 (0%)
Tivaouane	6 (2,4%)	4 (2,1%)	2 (3,4%)
Velingara	1 (0,4%)	1 (0,5%)	0 (0%)
Ziguinchor	35 (14%)	26 (14%)	9 (15%)
parle			
1	37 (15%)	32 (17%)	5 (8,5%)
2	104 (42%)	86 (45%)	18 (31%)
3	67 (27%)	44 (23%)	23 (39%)
4	34 (14%)	24 (13%)	10 (17%)
5	7 (2,8%)	5 (2,6%)	2 (3,4%)
6	1 (0,4%)	0 (0%)	1 (1,7%)

```

#Analyse par filiere
#filiere suivant le sexe du dirigeant
tab3 <- projet_r %>%
tbl_summary(
include = c(sexe, filiere_1, filiere_2, filiere_3, filiere_4),
by = sexe,
statistic = ~ "{p}% ({n})",
percent = "row"
)%>%
add_overall()%>%
modify_header(
all_stat_cols() ~ "***{level}** ({n} obs.)"
) %>%
bold_labels()

#filiere suivant la langue parlée
tab4 <- projet_r %>%
tbl_summary(
include = c(parle, filiere_1, filiere_2, filiere_3, filiere_4),
by = parle,
statistic = ~ "{p}% ({n})",
percent = "row"
)%>%
add_overall()%>%
modify_header(
all_stat_cols() ~ "***{level}** ({n} obs.)"
) %>%
bold_labels()

#filiere suivant la region
tab5 <- projet_r %>%
tbl_summary(
include = c(region, filiere_1, filiere_2, filiere_3, filiere_4),
by = region,
statistic = ~ "{p}% ({n})",

```

```

percent = "row"
)%>%
add_overall()%>%
modify_header(
all_stat_cols() ~ "**{level}** ({n} obs.)"
) %>%
bold_labels()
#combinons les tables tab1 et tab2
gtsummary::tbl_stack(
list(tab3, tab4, tab5),
group_header = c("filierie suivant le sexe du dirigeant", "filierie suivant la langue parlée", "#filierie s
)

```

Group	Caractéristique	Total	Femme	Homme	3	4	5	6	Saint-	Sédhiou	Thiès	Ziguinchor
		(250 obs.)	(191 obs.)	(59 obs.)	(67 obs.)	(34 obs.)	(7 obs.)	(1 obs.)	Louis (42 obs.)	(4 obs.)	(51 obs.)	(50 obs.)
filierie suivant le sexe du dirigeant	filierie arachide	100% (108)	86% (93)	14% (15)								
	filierie anacarde	100% (61)	66% (40)	34% (21)								
	filierie mangue	100% (89)	76% (68)	24% (21)								
	filierie riz	100% (92)	84% (77)	16% (15)								
	filierie arachide	100% (108)	18% (19)	45% (49)	26% (28)	9,3% (10)	1,9% (2)	0% (0)				
filierie suivant la langue parlée	filierie anacarde	100% (61)	1,6% (1)	34% (21)	28% (17)	25% (15)	9,8% (6)	1,6% (1)				
	filierie mangue	100% (89)	20% (18)	45% (40)	27% (24)	4,5% (4)	3,4% (3)	0% (0)				
	filierie riz	100% (92)	7,6% (7)	24% (22)	33% (30)	29% (27)	5,4% (5)	1,1% (1)				
	filierie arachide	100% (108)	0% (0)	31% (33)	11% (12)	7,4% (8)	19% (20)	0,9% (1)	0,9% (1)	0% (0)	25% (27)	5,6% (6)
	filierie anacarde	100% (61)	1,6% (1)	0% (0)	34% (21)	0% (0)	0% (0)	8,2% (5)	0% (0)	4,9% (3)	0% (0)	51% (31)
#filierie suivant la region	filierie mangue	100% (89)	0% (0)	1,1% (1)	3,4% (3)	5,6% (5)	7,9% (7)	0% (0)	47% (42)	0% (0)	28% (25)	6,7% (6)
	filierie riz	100% (92)	1,1% (1)	0% (0)	4,3% (4)	1,1% (1)	4,3% (4)	4,3% (4)	0% (0)	3,3% (3)	35% (32)	47% (43)

```

library("lubridate")
library("magrittr")

```



```

# Convertir les colonnes "date_depart" et "date_arrivee" en objets de type "POSIXct" (si nécessaire)
projet_r$start <- ymd_hms(projet_r$start)
projet_r$submissiondate <- ymd_hms(projet_r$submissiondate)
# Calculer l'intervalle de temps entre la date de départ et la date d'arrivée
projet_r$intervalle_temps <- interval(projet_r$start, projet_r$submissiondate)

# intervalle de temps entre la date de recueil et de soumissions des informations des informations de l
projet_r$duree <- as.duration(projet_r$intervalle_temps)
# creation d'un tatbleau qui donne la durée par region
projet_r %>%
  tbl_summary(
    include = c(region, duree),
    type = duree ~ "continuous2",
    by = region,
    statistic = list(
      all_continuous2() ~ c("{mean} ({sd})", "{min}", "{max}")
    )
  ) %>%
  add_overall()

## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.

```

	Total, N = 350	Dakar, N = 1	Diourbel, N = 34	Fatick, N = 30	Kaffrine, N = 8	Kaolack, N = 21	Kolda, N = 9	Saint- Louis, N = 42	Sédhiou, N = 4	Thiès, N = 51	Ziguinchor, N = 50
duree											
Moyenne	567	944	126	401	281	624	630	1 378	713	290	560 226
(ET)	820	916	586	594	582	660	489	780	334	979	(346
	(595	(NA)	(209	(485	(291	(558	(224	(714	(42	(348	353)
	691)		416)	986)	576)	324)	551)	012)	635)	563)	
Minimum2	192	944	3 130	19	22 975	151	236	201 110	689	2 192	13 297
		916		160		688	875		714		
Maximum2	389	944	1 036	1 631	686	1 724	994	2 389	777	1 116	1 468
	712	916	607	063	076	260	305	712	224	615	834

1.3 cartographie

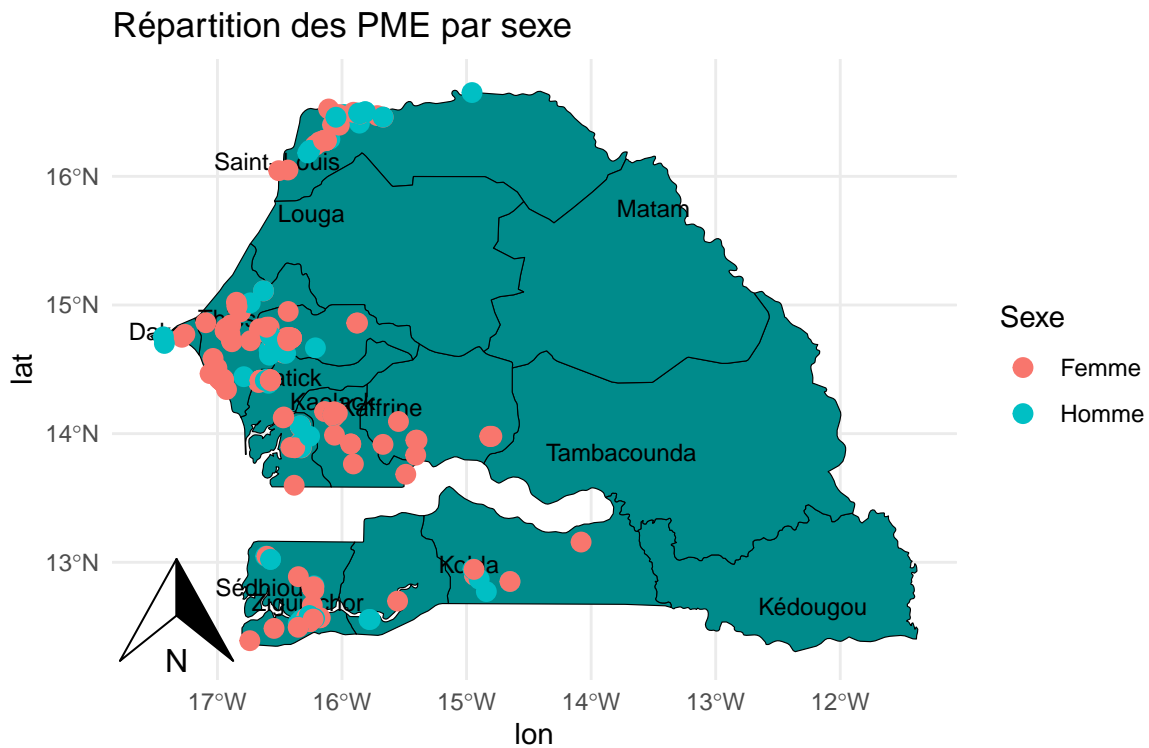
```

library(ggplot2)
library(sf)
library(rnaturalearth)
library(ggspatial)

# Créer un objet sf à partir du data.frame projet en utilisant les coordonnées géographiques
projet_map <- st_as_sf(projet, coords = c("gps_menlongitude", "gps_menlatitude"), crs = 4326)
# Récupérer les frontières des régions du Sénégal
senegal_regions <- rnaturalearth:: ne_states(country = "Senegal", returnclass = "sf")
# Création d'un data.frame contenant les coordonnées et les noms des régions du Sénégal
regions <- data.frame(
  region = c("Dakar", "Thiès", "Fatick", "Kaolack", "Kaffrine", "Kédougou", "Kolda", "Louga", "Matam", "S",
  lon = c(-17.455390, -16.920337, -16.412964, -16.073365, -15.687128, -12.220533, -14.981073, -16.246381,
  lat = c(14.693425, 14.798658, 14.339950, 14.151858, 14.101164, 12.559404, 12.887101, 15.614472, 15.6509
)

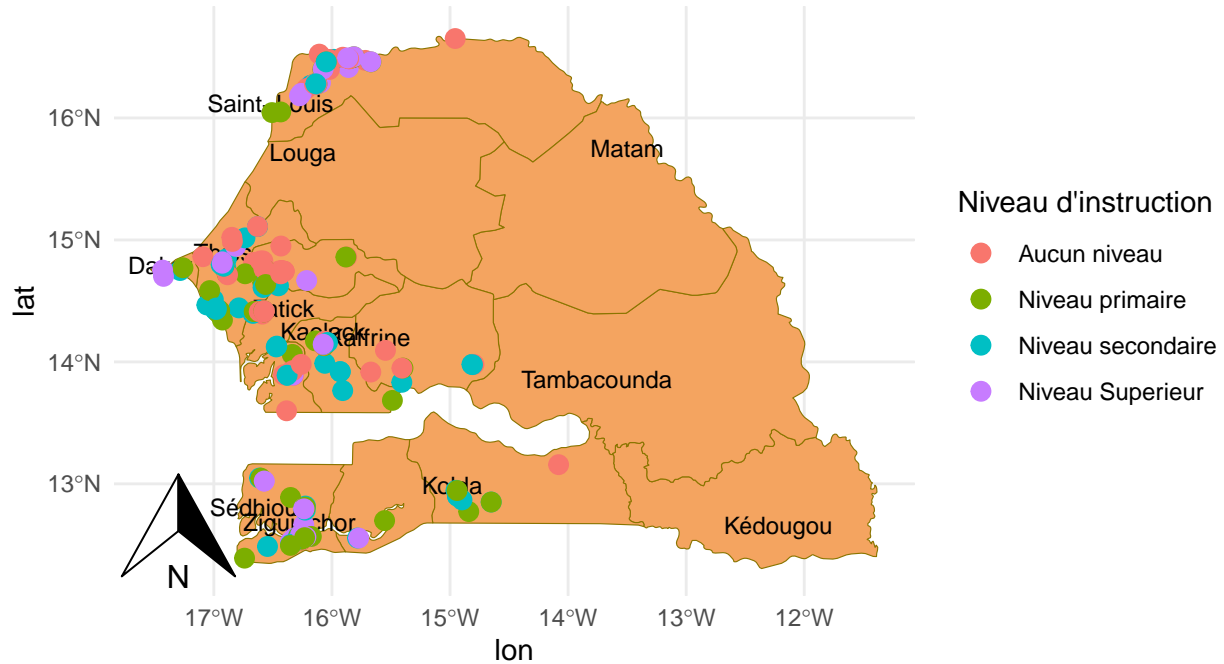
```

```
# Représentation spatiale des PME suivant le sexe avec ggplot2
ggplot() +
  geom_sf(data = senegal_regions, fill = "cyan4", color = "black") + # Fond de carte du Sénégal avec con
  geom_text(data = regions, aes(x = lon, y = lat, label = region), size = 3, nudge_y = 0.1) + # Ajout des
  annotation_north_arrow(location = "bl", which_north = "true", pad_x = unit(0.1, "cm"), pad_y = unit(0.1, "cm")) +
  geom_point(data = projet_map, aes(x = st_coordinates(geometry)[, 1], y = st_coordinates(geometry)[, 2]),
  labs(title = "Répartition des PME par sexe", color = "Sexe") +
  theme_minimal()
```



```
# Représentation spatiale des PME suivant le niveau d'instruction avec ggplot2
ggplot() +
  geom_sf(data = senegal_regions, fill = "sandybrown", color = "gold4") + # Fond de carte du Sénégal
  geom_text(data = regions, aes(x = lon, y = lat, label = region), size = 3, nudge_y = 0.1) + # Ajout des
  annotation_north_arrow(location = "bl", which_north = "true", pad_x = unit(0.1, "cm"), pad_y = unit(0.1, "cm")) +
  geom_point(data = projet_map, aes(x = st_coordinates(geometry)[, 1], y = st_coordinates(geometry)[, 2]),
  labs(title = "Répartition des PME par niveau d'instruction", color = "Niveau d'instruction") +
  theme_minimal()
```

Répartition des PME par niveau d'instruction



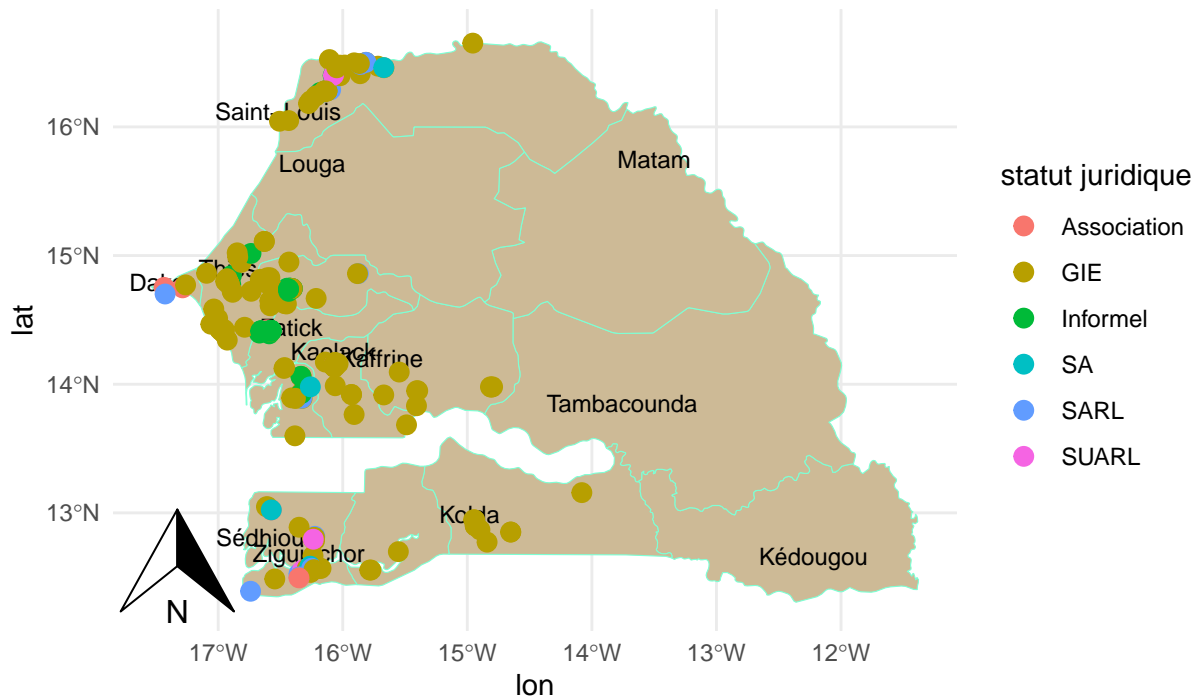
1.4 autre analyse spatiale

Nous allons faire une analyse spatiale des PME suivant le statut juridique

Représentation spatiale des PME suivant le niveau d'instruction avec ggplot2

```
ggplot() +
  geom_sf(data = senegal_regions, fill = "wheat3", color = "aquamarine") + # Fond de carte du Sénégal avec
  geom_text(data = regions, aes(x = lon, y = lat, label = region), size = 3, nudge_y = 0.1) + # Ajout des
  annotation_north_arrow(location = "bl", which_north = "true", pad_x = unit(0.1, "cm"), pad_y = unit(0.1, "cm")) +
  geom_point(data = projet_map, aes(x = st_coordinates(geometry)[, 1], y = st_coordinates(geometry)[, 2]),
    color = "statut juridique") +
  theme_minimal()
```

Répartition des PME suivant le statut juridique



2 partie 2

2.1 Nettoyage et gestion des données

```
library(broom)
# Charger les données depuis le fichier Excel
data <- read_excel("Base_Partie 2.xlsx",
  sheet = "data",
  range = NULL,
  col_names = TRUE,
  col_types = NULL,
  na = "")

# Nettoyage et gestion des données
# Remplacer les valeurs aberrantes (999) par NA dans la colonne endline_age
data <- data %>%
  mutate(age = ifelse(age == 999, NA, age))
# Calculer la moyenne des âges sans prendre en compte les valeurs manquantes (NA)
mean_age <- mean(data$age, na.rm = TRUE)
# Remplacer les valeurs manquantes (NA) par la moyenne des autres âges
data$age[is.na(data$age)] <- mean_age
# Arrondir à l'entier supérieur
data$age <- ceiling(data$age)
# Renommer la variable "country_destination" en "destination" et définir les valeurs négatives comme NA
data <- data %>%
  rename(destination = country_destination) %>%
  mutate(destination = ifelse(destination < 0, NA, destination))
```

```

# Remplacer les valeurs manquantes dans la variable "age" par NA
data <- data %>%
  mutate(age = ifelse(is.na(age), NA, age))
# Créer une nouvelle variable contenant des tranches d'âge de 5 ans en utilisant la variable "age"
data <- data %>%
  mutate(age_group = cut(age, breaks = seq(15, max(age) + 5, by = 5), include.lowest = TRUE),
         age_group = cut(age, breaks = seq(15, max(age) + 5, by = 5), include.lowest = TRUE))
# Afficher les premières lignes de la nouvelle variable age_group_label
head(data$age_group)

## [1] (30,35] (40,45] (25,30] (20,25] (25,30] (20,25]
## Levels: [15,20] (20,25] (25,30] (30,35] (35,40] (40,45]

# Créer une nouvelle variable contenant le nombre d'entretiens réalisés par chaque agent recenseur
data <- data %>%
  group_by(enumerator) %>%
  mutate(nombre_ent = n()) %>%
  ungroup()
# Créer une nouvelle variable qui affecte aléatoirement chaque répondant à un groupe de traitement (1)
set.seed(42) # fixons l'aléa
data <- data %>%
  mutate(treatment_group = sample(c(0, 1), nrow(.), replace = TRUE))

# Charger les données depuis le fichier Excel
district_data <- read_excel("Base_Partie 2.xlsx",
                           sheet = "district",
                           range = NULL,
                           col_names = TRUE,
                           col_types = NULL,
                           na = "")
# Fusionner la taille de la population de chaque district avec l'ensemble de données
data2 <- left_join(data, district_data, by = "district")
# afficher les premières lignes
head(data2)

## # A tibble: 6 x 14
##       id starttime      endtime      enumerator district  age  sex
##   <dbl> <dtm>      <dtm>      <dbl>      <dbl> <dbl> <dbl>
## 1     2 2019-01-14 14:56:37 2019-01-14 15:11:10         6         1    33    1
## 2     3 2019-01-14 16:12:22 2019-01-14 16:45:52         6         1    43    0
## 3     4 2019-01-14 17:15:47 2019-01-14 17:45:47         6         1    28    0
## 4     7 2019-01-14 13:04:51 2019-01-14 13:27:38         8         3    24    0
## 5     8 2019-01-14 13:38:00 2019-01-14 14:31:16         8         3    29    0
## 6    10 2019-01-14 15:52:17 2019-01-14 16:33:39         8         6    22    1
## # i 7 more variables: children_num <dbl>, intention <dbl>, destination <dbl>,
## #   age_group <fct>, nombre_ent <int>, treatment_group <dbl>, population <dbl>

# Calculer la durée de l'entretien et indiquer la durée moyenne de l'entretien par enquêteur
data2 <- data2 %>%
  mutate(duree = endtime - starttime,
         duree_moy = mean(duree, na.rm = TRUE))
#affichage des premières lignes
head(data2)

## # A tibble: 6 x 16
##       id starttime      endtime      enumerator district  age  sex

```

```
##      <dbl> <dtm>                <dtm>                <dbl>      <dbl> <dbl> <dbl>
## 1      2 2019-01-14 14:56:37 2019-01-14 15:11:10          6          1    33    1
## 2      3 2019-01-14 16:12:22 2019-01-14 16:45:52          6          1    43    0
## 3      4 2019-01-14 17:15:47 2019-01-14 17:45:47          6          1    28    0
## 4      7 2019-01-14 13:04:51 2019-01-14 13:27:38          8          3    24    0
## 5      8 2019-01-14 13:38:00 2019-01-14 14:31:16          8          3    29    0
## 6     10 2019-01-14 15:52:17 2019-01-14 16:33:39          8          6    22    1
## # i 9 more variables: children_num <dbl>, intention <dbl>, destination <dbl>,
## #   age_group <fct>, nombre_ent <int>, treatment_group <dbl>, population <dbl>,
## #   duree <drtn>, duree_moy <drtn>

# Renommer toutes les variables en ajoutant le préfixe "endline_" à l'aide de supply et setNames

#creation un objet "prefix" qui prend la valeur "endline_"
prefix <- "endline_"
# Utilisation de la fonction rename_with() pour ajouter le préfixe
data2 <- data2 %>%
  rename_with(~ paste0(prefix, .), everything())
```

2.2 Analyse et visualisation des données

```
# Creation d'un tableau récapitulatif contenant l'âge moyen et le nombre moyen d'enfants par district

data2 %>%
  tbl_summary(
    include = c(endline_age, endline_children_num),
    by = endline_district,
    statistic = all_continuous() ~ "Moy. : {mean}"
  ) %>%
  add_overall()

## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

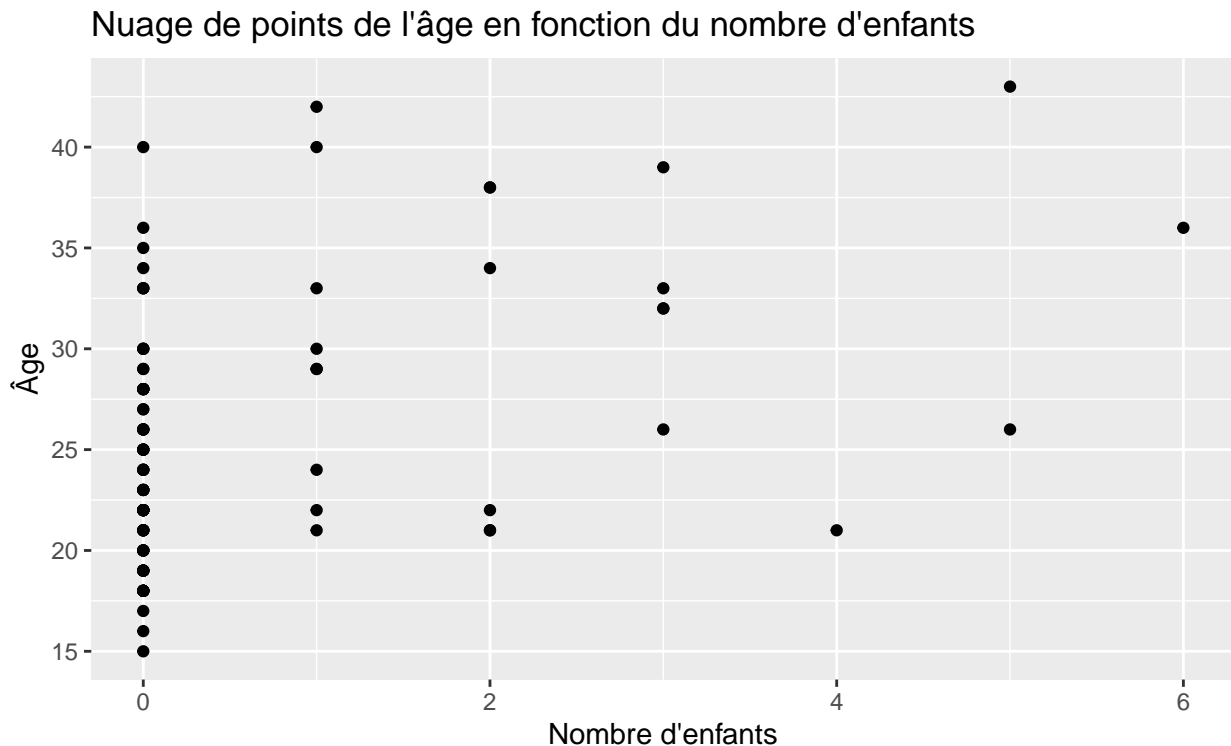
	Total, N	1, N =	2, N =	3, N =	4, N =	5, N =	6, N =	7, N =	8, N =
Caractéristique	= 97	8	27	8	5	6	26	6	11
endline_age	Moy. : 26	Moy. : 30	Moy. : 27	Moy. : 26	Moy. : 26	Moy. : 24	Moy. : 23	Moy. : 28	Moy. : 25
endline_children_num									
0	73 (75%)	4 (50%)	17 (63%)	8 (100%)	5 (100%)	4 (67%)	23 (88%)	5 (83%)	7 (64%)
1	9 (9,3%)	1 (13%)	3 (11%)	0 (0%)	0 (0%)	1 (17%)	3 (12%)	1 (17%)	0 (0%)
2	6 (6,2%)	1 (13%)	3 (11%)	0 (0%)	0 (0%)	1 (17%)	0 (0%)	0 (0%)	1 (9,1%)
3	5 (5,2%)	0 (0%)	3 (11%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (18%)
4	1 (1,0%)	1 (13%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
5	2 (2,1%)	1 (13%)	1 (3,7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
6	1 (1,0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (9,1%)

```
# Testez si la différence d'âge entre les sexes est statistiquement significative au niveau de 5 %
test <- lm(endline_age ~ endline_sex, data = data2)
test%>%gtsummary::tbl_regression(add_estimate_to_reference_rows = TRUE)
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Caractéristique	Beta	95% IC	p-valeur
endline_sex	-3,6	-7,6 – 0,38	0,076

```
# Créer un nuage de points de l'âge en fonction du nombre d'enfants (ggplot)
ggplot(data2, aes(x = endline_children_num, y = endline_age)) +
  geom_point() +
  labs(x = "Nombre d'enfants", y = "Âge") +
  ggtitle("Nuage de points de l'âge en fonction du nombre d'enfants")
```



```
# Effectuer l'analyse de régression linéaire
modele_regression <- lm(endline_intention ~ endline_treatment_group, data = data2)
# Obtenir les résultats de l'analyse
summary(modele_regression)
```

```
##
## Call:
## lm(formula = endline_intention ~ endline_treatment_group, data = data2)
##
## Residuals:
```

##	Min	1Q	Median	3Q	Max
----	-----	----	--------	----	-----

```
## -1.2500 -1.2500 -0.9623 1.0377 5.0377
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.2500     0.2598   8.660 1.2e-13 ***
## endline_treatment_group -0.2877     0.3515  -0.819  0.415
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.723 on 95 degrees of freedom
## Multiple R-squared:  0.007004, Adjusted R-squared:  -0.003448
## F-statistic: 0.6701 on 1 and 95 DF, p-value: 0.4151

# Obtenir les résultats de l'analyse dans un tableau
modele_regression%>%gtsummary::tbl_regression()## le coefficient de regression est négatif et on consta

## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Caractéristique	Beta	95% IC	p-valeur
endline_treatment_group	-0,29	-0,99 – 0,41	0,4

```
# Créez un tableau de régression avec 3 modèles
model_A <- lm(endline_intention ~ endline_treatment_group, data = data2)
model_B <- lm(endline_intention ~ endline_treatment_group + endline_age + endline_sex, data = data2)
model_C <- lm(endline_intention ~ endline_treatment_group + endline_age + endline_sex + endline_district)

# Stocker les résultats des modèles dans un tableau
tabA <- model_A%>%gtsummary::tbl_regression(add_estimate_to_reference_rows = TRUE)
tabB <- model_B%>%gtsummary::tbl_regression(add_estimate_to_reference_rows = TRUE)
tabC <- model_C%>%gtsummary::tbl_regression(add_estimate_to_reference_rows = TRUE)

#sortons le tableau de la regression
stacked_table <- gtsummary::tbl_stack(
  list(tabA, tabB, tabC),
  group_header = c("traitement sur les intentions", "traitement sur les intentions en tenant compte de l'âge et du sexe", "traitement sur les intentions en tenant compte de l'âge et du sexe en tenant compte du district")
)
stacked_table

## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Group	Caractéristique	Beta	95% IC	p-valeur
traitement sur les intentions	endline_treatment_group	-0,29	-0,99 – 0,41	0,4
traitement sur les intentions en tenant compte de l'âge et du sexe	endline_treatment_group	-0,29	-0,99 – 0,41	0,4
traitement sur les intentions en tenant compte de l'âge et du sexe en tenant compte du district	endline_treatment_group	-0,29	-0,99 – 0,41	0,4

3 partie 3

Voici le lien de l'application RShinny <https://isabellemosse.shinyapps.io/MIDapp/> Il est nécessaire de choisir un pays une fois l'application ouverte afin que la carte ne s'affiche.