Partie 1

DIAKITE Moussa

library(readxl)  
library(dplyr)

##   
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':  
##   
## filter, lag

## Les objets suivants sont masqués depuis 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyr)  
library(gtsummary)  
library(sf)

## Linking to GEOS 3.9.3, GDAL 3.5.2, PROJ 8.2.1; sf\_use\_s2() is TRUE

library(gt)  
library(ggplot2)  
# chargement des packages nécessaires

##### corps du code

setwd("E:/Projet R")  
# et pour changer de répertoire  
  
#-----------------------------------------------  
# IMPORTATION ET MISE EN FORME  
#-----------------------------------------------  
  
  
projet<-read\_excel('Base\_Partie 1.xlsx')  
head(projet)

## # A tibble: 6 × 33  
## key q1 q2 q23 q24 q24a\_1 q24a\_2 q24a\_3 q24a\_4 q24a\_5 q24a\_6 q24a\_7  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 uuid… Diou… Bamb… Femme 65 0 1 0 1 0 0 0  
## 2 uuid… Thiès Mbour Femme 52 1 1 0 0 1 0 0  
## 3 uuid… Thiès Mbour Femme 65 1 1 0 0 0 0 0  
## 4 uuid… Thiès Mbour Femme 38 1 1 0 0 1 0 0  
## 5 uuid… Zigu… Bign… Homme 40 1 1 1 0 0 1 0  
## 6 uuid… Zigu… Ouss… Femme 43 1 1 1 0 0 0 0  
## # ℹ 21 more variables: q24a\_9 <dbl>, q24a\_10 <dbl>, q25 <chr>, q26 <dbl>,  
## # q12 <chr>, q14b <chr>, q16 <chr>, q17 <chr>, q19 <chr>, q20 <chr>,  
## # filiere\_1 <dbl>, filiere\_2 <dbl>, filiere\_3 <dbl>, filiere\_4 <dbl>,  
## # q8 <chr>, q81 <chr>, gps\_menlatitude <dbl>, gps\_menlongitude <dbl>,  
## # submissiondate <dttm>, start <dttm>, today <dttm>

# data.frame nommé projet  
# head : pour lecture du data.frame  
  
  
selection\_variable <- projet[, c("key", "q1", "q2", "q23", "q24",  
 "q24a\_1", "q24a\_2", "q24a\_3",   
 "q24a\_4", "q24a\_5", "q24a\_6",   
 "q24a\_7", "q24a\_9", "q24a\_10",  
 "q25", "q26", "q12", "q14b",  
 "q16", "q17", "q19", "q20",   
 "filiere\_1", "filiere\_2",   
 "filiere\_3", "filiere\_4",   
 "q8", "q81", "gps\_menlatitude",   
 "gps\_menlongitude",   
 "submissiondate", "start", "today")]  
# création de sous ensemble de data.frame  
  
nombre\_valeur\_manquante <- colSums(is.na(selection\_variable))  
# pour avoir le nombre de valeur manquante par variable  
  
tableau\_valeur\_manquante <- data.frame(variable=names(nombre\_valeur\_manquante),  
 nombre\_valeur\_manquante=nombre\_valeur\_manquante)  
  
# création de data.frame pour résumer les valeurs manquantes  
  
print(tableau\_valeur\_manquante)

## variable nombre\_valeur\_manquante  
## key key 0  
## q1 q1 0  
## q2 q2 0  
## q23 q23 0  
## q24 q24 0  
## q24a\_1 q24a\_1 0  
## q24a\_2 q24a\_2 0  
## q24a\_3 q24a\_3 0  
## q24a\_4 q24a\_4 0  
## q24a\_5 q24a\_5 0  
## q24a\_6 q24a\_6 0  
## q24a\_7 q24a\_7 0  
## q24a\_9 q24a\_9 0  
## q24a\_10 q24a\_10 0  
## q25 q25 0  
## q26 q26 0  
## q12 q12 0  
## q14b q14b 1  
## q16 q16 1  
## q17 q17 131  
## q19 q19 120  
## q20 q20 0  
## filiere\_1 filiere\_1 0  
## filiere\_2 filiere\_2 0  
## filiere\_3 filiere\_3 0  
## filiere\_4 filiere\_4 0  
## q8 q8 0  
## q81 q81 0  
## gps\_menlatitude gps\_menlatitude 0  
## gps\_menlongitude gps\_menlongitude 0  
## submissiondate submissiondate 0  
## start start 0  
## today today 0

# on affiche le tableau des valeurs manquantes   
  
valeur\_manquante\_key<- is.na(projet$key)  
# verification pour l'existence d'une varible manquante pour key  
  
pme\_concerne<-projet[valeur\_manquante\_key,]  
# pour avoir les pme concernés  
  
print(pme\_concerne)

## # A tibble: 0 × 33  
## # ℹ 33 variables: key <chr>, q1 <chr>, q2 <chr>, q23 <chr>, q24 <dbl>,  
## # q24a\_1 <dbl>, q24a\_2 <dbl>, q24a\_3 <dbl>, q24a\_4 <dbl>, q24a\_5 <dbl>,  
## # q24a\_6 <dbl>, q24a\_7 <dbl>, q24a\_9 <dbl>, q24a\_10 <dbl>, q25 <chr>,  
## # q26 <dbl>, q12 <chr>, q14b <chr>, q16 <chr>, q17 <chr>, q19 <chr>,  
## # q20 <chr>, filiere\_1 <dbl>, filiere\_2 <dbl>, filiere\_3 <dbl>,  
## # filiere\_4 <dbl>, q8 <chr>, q81 <chr>, gps\_menlatitude <dbl>,  
## # gps\_menlongitude <dbl>, submissiondate <dttm>, start <dttm>, today <dttm>

# affichage des pme concernés  
  
  
#----------------------------------------------  
# CREATION DE VARIABLES  
#----------------------------------------------  
  
projet <- projet %>% rename(region=q1)  
# renom de q1 en region  
  
projet <- projet %>% rename(departement=q2)  
# renom de q2 en departement  
  
projet <- projet %>% rename(sexe=q23)  
#renom de q23 en sexe   
  
projet$sexe\_2 <- ifelse(projet$sexe=="femme",1,0)  
# creation de la variable sexé\_2 et application du choix binaire  
  
langues <- projet %>% select(key, starts\_with("q24a\_"))  
# creation du data.frame nommé langue  
  
head(langues)

## # A tibble: 6 × 10  
## key q24a\_1 q24a\_2 q24a\_3 q24a\_4 q24a\_5 q24a\_6 q24a\_7 q24a\_9 q24a\_10  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 uuid:68bff42b… 0 1 0 1 0 0 0 0 0  
## 2 uuid:d70b3c7e… 1 1 0 0 1 0 0 0 0  
## 3 uuid:0ac18b64… 1 1 0 0 0 0 0 0 0  
## 4 uuid:c52cf5e4… 1 1 0 0 1 0 0 0 0  
## 5 uuid:ac177870… 1 1 1 0 0 1 0 0 0  
## 6 uuid:578097cf… 1 1 1 0 0 0 0 0 0

#afficher les premières lignes de langues   
  
projet$parle<- projet %>% select(key, starts\_with("q24a\_"))  
# creation d'une variable langue et une autre alternative est :  
# proje$parle <- rowSums(projet[, starts\_with("q24a\_")])  
  
langues<- projet %>% select (key, parle)  
# creation de langue à partir de key et parle  
  
fusion\_projet <- merge(projet, langues, by="key")  
# on fussionne les data.frame projet et langues en   
# utilisant key comme la clé de fusion  
  
head(fusion\_projet)

## Warning in format.data.frame(if (omit) x[seq\_len(n0), , drop = FALSE] else x, :  
## tableau de données corrompu : les colonnes seront tronquées ou complétées avec  
## des NAs

## key region departement sexe q24  
## 1 uuid:004b9117-d180-4031-a6af-6b4efabb5f53 Diourbel Bambey Femme 62  
## 2 uuid:007d8eb4-45eb-44f4-aeac-722adc60aec8 Thiès Tivaouane Femme 60  
## 3 uuid:030ada55-8dd2-4f57-b1b7-aaccd707c118 Saint-Louis Dagana Homme 58  
## 4 uuid:04b4cd8d-0297-4dc0-9715-9c07120bab23 Diourbel Mbacké Femme 60  
## 5 uuid:080548bf-e68a-49b8-9f04-f920b44244aa Saint-Louis Dagana Homme 63  
## 6 uuid:087f6e40-9b2a-4252-a1b3-b4b8ec7b3dff Saint-Louis Dagana Femme 61  
## q24a\_1 q24a\_2 q24a\_3 q24a\_4 q24a\_5 q24a\_6 q24a\_7 q24a\_9 q24a\_10  
## 1 0 1 0 0 0 0 0 0 0  
## 2 0 1 0 1 0 0 0 0 0  
## 3 1 1 0 0 0 0 0 0 0  
## 4 1 1 0 0 0 0 0 0 0  
## 5 0 1 0 0 1 0 0 0 0  
## 6 1 1 0 0 0 0 0 0 0  
## q25 q26 q12 q14b q16 q17 q19 q20 filiere\_1  
## 1 Aucun niveau 20 GIE Non Non <NA> Mauvais état Oui 1  
## 2 Niveau secondaire 10 GIE Non Oui Bon état <NA> Non 1  
## 3 Niveau secondaire 30 GIE Non Non <NA> Mauvais état Oui 0  
## 4 Niveau primaire 25 GIE Oui Oui Bon état <NA> Oui 1  
## 5 Aucun niveau 21 SARL Non Oui Bon état <NA> Non 0  
## 6 Niveau primaire 25 GIE Non Non <NA> Mauvais état Oui 0  
## filiere\_2 filiere\_3 filiere\_4 q8 q81  
## 1 0 0 0 Aucun Propriétaire  
## 2 0 0 1 Tansformation d'autres céréales Locataire  
## 3 0 1 0 Transformation du riz Propriétaire  
## 4 0 0 0 Tansformation d'autres céréales Locataire  
## 5 0 1 0 Transformation du riz Propriétaire  
## 6 0 1 0 Transformation du riz Propriétaire  
## gps\_menlatitude gps\_menlongitude submissiondate start  
## 1 14.82743 -16.60590 2021-06-05 15:33:51 2021-06-04 15:14:14  
## 2 15.10929 -16.62974 2021-06-15 01:10:46 2021-06-08 14:40:28  
## 3 16.45945 -16.04850 2021-06-21 01:28:51 2021-06-07 18:24:19  
## 4 14.85961 -15.88164 2021-06-07 13:51:55 2021-06-07 09:50:58  
## 5 16.27839 -16.14392 2021-06-18 10:20:16 2021-05-24 15:33:59  
## 6 16.45927 -16.04855 2021-06-21 01:31:17 2021-06-07 18:50:42  
## today sexe\_2  
## 1 2021-06-04 0  
## 2 2021-06-08 0  
## 3 2021-06-07 0  
## 4 2021-06-07 0  
## 5 2021-05-24 0  
## 6 2021-06-07 0  
## parle.x  
## 1 # A tibble: 6 × 10  
## 2 key q24a\_1 q24a\_2 q24a\_3 q24a\_4 q24a\_5 q24a\_6 q24a\_7 q24a\_9 q24a\_10  
## 3 <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 4 1 uuid:004b9117… 0 1 0 0 0 0 0 0 0  
## 5 2 uuid:007d8eb4… 0 1 0 1 0 0 0 0 0  
## 6 3 uuid:030ada55… 1 1 0 0 0 0 0 0 0  
## parle.y  
## 1 # A tibble: 6 × 10  
## 2 key q24a\_1 q24a\_2 q24a\_3 q24a\_4 q24a\_5 q24a\_6 q24a\_7 q24a\_9 q24a\_10  
## 3 <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 4 1 uuid:004b9117… 0 1 0 0 0 0 0 0 0  
## 5 2 uuid:007d8eb4… 0 1 0 1 0 0 0 0 0  
## 6 3 uuid:030ada55… 1 1 0 0 0 0 0 0 0

# lecture des premières lignes  
  
  
#------------------------------------------------  
# ANALYSE DESCRIPTIVE  
#------------------------------------------------  
  
repartition\_sexe <- table(projet$sexe)  
# compte le nombre de chaque type de sexe  
  
repartition\_niveau\_instruction <- table(projet$q25)  
# compte l'occurence de chaque niveau d'instruction  
  
repartition\_statut\_juridique <- table(projet$q12)  
# l'occurence de chaque statut juridique  
  
repartition\_proprietaire\_locataire <- table(projet$q81)  
# denombre l'occurence de chaque statut proprietaire\_locataire  
  
repartition\_statut\_sexe <- table(projet$q12, projet$sexe)  
# tableau croisé en statut juridique et sexe  
  
repartition\_niveau\_instruction\_sexe <- table(projet$q25, projet$sexe)  
# tableau croisé niveu d'instruction et sexe  
  
repartition\_proprietaire\_sexe <- table(projet$q81, projet$sexe)  
# tableau croisé entre statut proprietaire\_locataire et sexe  
  
tableau\_synthese\_1 <- projet %>% select(sexe, q25, q12, q81) %>% tbl\_summary() %>% as\_gt()  
tableau\_synthese\_1

| **Characteristic** | **N = 250***1* |
| --- | --- |
| sexe |  |
| Femme | 191 (76%) |
| Homme | 59 (24%) |
| q25 |  |
| Aucun niveau | 79 (32%) |
| Niveau primaire | 56 (22%) |
| Niveau secondaire | 74 (30%) |
| Niveau Superieur | 41 (16%) |
| q12 |  |
| Association | 6 (2.4%) |
| GIE | 179 (72%) |
| Informel | 38 (15%) |
| SA | 7 (2.8%) |
| SARL | 13 (5.2%) |
| SUARL | 7 (2.8%) |
| q81 |  |
| Locataire | 24 (9.6%) |
| Propriétaire | 226 (90%) |
| *1*n (%) | |

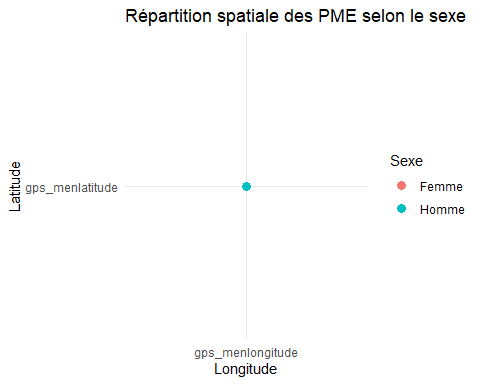
# tableau synthèse pour les premières répartitions  
  
  
variables\_pour\_resume <- c("q26", "q16", "q17", "q19", "q20")  
  
# Rassembler les variables de filière en une seule colonne  
fusion\_projet\_long <- fusion\_projet %>%  
 pivot\_longer(cols = starts\_with("filiere\_"), names\_to = "filiere", values\_to = "filiere\_value")  
  
# Créer un tableau de résumé pour les variables sélectionnées, groupées par "filiere"  
tableau\_resume <- fusion\_projet\_long %>%  
 select(filiere, all\_of(variables\_pour\_resume)) %>%  
 tbl\_summary(by = filiere, missing = "no") %>%  
 add\_stat\_label() %>%  
 modify\_caption("Statistiques descriptives pour les variables sélectionnées par filière") %>%  
 as\_gt()  
  
# Afficher le tableau de résumé  
tableau\_resume

| **Characteristic** | **filiere\_1**, N = 250 | **filiere\_2**, N = 250 | **filiere\_3**, N = 250 | **filiere\_4**, N = 250 |
| --- | --- | --- | --- | --- |
| q26, Median (IQR) | 12 (6, 20) | 12 (6, 20) | 12 (6, 20) | 12 (6, 20) |
| q16, n (%) |  |  |  |  |
| Non | 130 (52%) | 130 (52%) | 130 (52%) | 130 (52%) |
| Oui | 119 (48%) | 119 (48%) | 119 (48%) | 119 (48%) |
| q17, n (%) |  |  |  |  |
| Bon état | 92 (77%) | 92 (77%) | 92 (77%) | 92 (77%) |
| Etat moyen | 24 (20%) | 24 (20%) | 24 (20%) | 24 (20%) |
| Mauvais état | 3 (2.5%) | 3 (2.5%) | 3 (2.5%) | 3 (2.5%) |
| q19, n (%) |  |  |  |  |
| Bon état | 2 (1.5%) | 2 (1.5%) | 2 (1.5%) | 2 (1.5%) |
| Etat moyen | 73 (56%) | 73 (56%) | 73 (56%) | 73 (56%) |
| Mauvais état | 55 (42%) | 55 (42%) | 55 (42%) | 55 (42%) |
| q20, n (%) |  |  |  |  |
| Non | 121 (48%) | 121 (48%) | 121 (48%) | 121 (48%) |
| Oui | 129 (52%) | 129 (52%) | 129 (52%) | 129 (52%) |

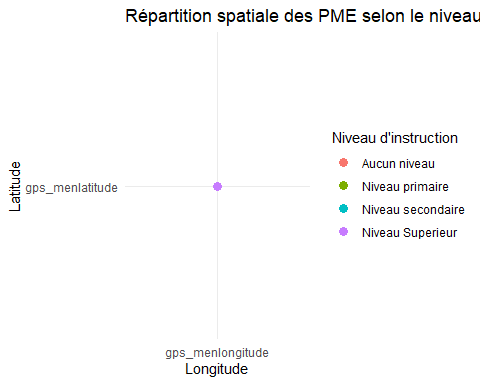
#-----------------------------------------------------------  
# UN PEU DE CARTOGRAPHIE  
#-----------------------------------------------------------  
  
library(sf)  
library(ggplot2)  
  
projet\_map <- st\_as\_sf(projet, coords = c("gps\_menlongitude", "gps\_menlatitude"), crs = 4326)  
# on crée un objet sf à partir du data.frame projet  
  
head(projet\_map)

## Simple feature collection with 6 features and 33 fields  
## Geometry type: POINT  
## Dimension: XY  
## Bounding box: xmin: -16.96077 ymin: 12.48649 xmax: -16.46786 ymax: 14.62691  
## Geodetic CRS: WGS 84  
## # A tibble: 6 × 34  
## key region departement sexe q24 q24a\_1 q24a\_2 q24a\_3 q24a\_4 q24a\_5 q24a\_6  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 uuid… Diour… Bambey Femme 65 0 1 0 1 0 0  
## 2 uuid… Thiès Mbour Femme 52 1 1 0 0 1 0  
## 3 uuid… Thiès Mbour Femme 65 1 1 0 0 0 0  
## 4 uuid… Thiès Mbour Femme 38 1 1 0 0 1 0  
## 5 uuid… Zigui… Bignona Homme 40 1 1 1 0 0 1  
## 6 uuid… Zigui… Oussouye Femme 43 1 1 1 0 0 0  
## # ℹ 23 more variables: q24a\_7 <dbl>, q24a\_9 <dbl>, q24a\_10 <dbl>, q25 <chr>,  
## # q26 <dbl>, q12 <chr>, q14b <chr>, q16 <chr>, q17 <chr>, q19 <chr>,  
## # q20 <chr>, filiere\_1 <dbl>, filiere\_2 <dbl>, filiere\_3 <dbl>,  
## # filiere\_4 <dbl>, q8 <chr>, q81 <chr>, submissiondate <dttm>, start <dttm>,  
## # today <dttm>, sexe\_2 <dbl>, parle <tibble[,10]>, geometry <POINT [°]>

ggplot(data = projet\_map) +  
 geom\_point(aes(x = "gps\_menlongitude", y = "gps\_menlatitude", color = sexe), size = 3) +  
 labs(title = "Répartition spatiale des PME selon le sexe",  
 x = "Longitude", y = "Latitude",  
 color = "Sexe") +  
 theme\_minimal()



#représentation spatiale des PME selon le sexe  
  
  
ggplot(data = projet\_map) +  
 geom\_point(aes(x = "gps\_menlongitude", y = "gps\_menlatitude", color = q25), size = 3) +  
 labs(title = "Répartition spatiale des PME selon le niveau d'instruction",  
 x = "Longitude", y = "Latitude",  
 color = "Niveau d'instruction") +  
 theme\_minimal()



# représentation spatiale des PME selon le niveau d'inscription