projet R ENSAE 2023

GUEYE AISSATA

r Sys.Date()

Contents

L'objectif de ce projet est que nous appliquions les outils que nous avons étudiés dans le cours du logiciel statistique R, dans le cas d'une étude de cas réelle. Sur une enquête vise à identifier et à caractériser des bioénergies durables pour les petites et moyennes entreprises (PME) agroalimentaires d'Afrique de l'Ouest.

Installation des packages nécessaire

```
##
## 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

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.4
## v ggplot2   3.4.2     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 3 > 1' dans la conversion automatique vers 'logical(1)'

##
## Attachement du package : 'kableExtra'
##
## L'objet suivant est masqué depuis 'package:dplyr':
##
##     group_rows
```

1- importation et mise en forme

1-1-1 Importer la base de données dans un objet de type data.frame nommé projet

#1-1-2 Selectionner les variales mentionnées dans la section description

Table 1: Tableau résumé des valeurs manquantes

	valeurs_manquantes	proportions
key	0	0.0000000
q1	0	0.0000000
q2	0	0.0000000
q23	0	0.0000000
q24	0	0.0000000
q24a_1	0	0.0000000
q24a_2	0	0.0000000
q24a_3	0	0.0000000
q24a_4	0	0.0000000
q24a_5	0	0.0000000
q24a_6	0	0.0000000
q24a_7	0	0.0000000
q24a_9	0	0.0000000
q24a_10	0	0.0000000
q25	0	0.0000000
q26	0	0.0000000
q12	0	0.0000000
q14b	1	0.3952569
q16	1	0.3952569
q17	131	51.7786561
q19	120	47.4308300
q20	0	0.0000000
filiere_1	0	0.0000000
filiere_2	0	0.0000000
filiere_3	0	0.0000000
filiere_4	0	0.0000000
q8	0	0.0000000
q81	0	0.0000000
gps_menlatitude	0	0.0000000
gps_menlongitude	0	0.0000000
submissiondate	0	0.0000000
start	0	0.0000000
today	0	0.0000000

1-1-3 Un tableau qui résume les valeurs manquantes par variable

1-1-4 Vérifier s'il y a des valeurs manquantes pour la variable key dans la base projet. Si oui, identifier la (ou les) PME concernée(s).

D'après le tableau précédent la variable key n'a aucune valeur manquante. On peut aussi le voir à partir du code suivant:

```
## integer(0)
```

1-2 Création de variables

#1-2-1 Rénommer les variable q1,q2 et q3 respectivement en region,en departement et en sexe

```
##       region departement  sexe
## 1   Diourbel      Bambey Femme
## 2      Thiès       Mbour Femme
## 3      Thiès       Mbour Femme
## 4      Thiès       Mbour Femme
## 5 Ziguinchor     Bignona Homme
## 6 Ziguinchor    Oussouye Femme
```

#1-2-2 Créer la variable sexe_2 qui vaut 1 si sexe égale à Femme et 0 sinon.

#1-2-3 Créer un data.frame nommé langues qui prend les variables key et les variables correspondantes décrites plus haut.

#1-2-3 Créer la variable parle qui est égale au nombre de langue parlée par le dirigeant de la PME.

#1-2-4 Sélection uniquement des variables key et parle, l'objet de retour sera langues.

#1-2-5 Merger les data.frame projet et langues:

2 analyse descriptive

Nous allons nous proposer de faire une analyse descriptive univariée et bivariée Création du tableau pour les variables univariées

2-1 analyse univarié

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

Characteristic	**N = 250**
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%)

2-1 analyse bivarié

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

Characteristic	**Femme**, N = 191	**Homme**, N = 59
q12		
Association	3 (50%)	3 (50%)
GIE	149 (83%)	30 (17%)
Informel	32 (84%)	6 (16%)
SA	1 (14%)	6 (86%)
SARL	2 (15%)	11 (85%)
SUARL	4 (57%)	3 (43%)
q25		
Aucun niveau	70 (89%)	9 (11%)
Niveau primaire	48 (86%)	8 (14%)
Niveau secondaire	56 (76%)	18 (24%)
Niveau Superieur	17 (41%)	24 (59%)
q81		
Locataire	16 (67%)	8 (33%)
Propriétaire	175 (77%)	51 (23%)

#2-3 Analyse de la repartition des PM par filière suivant le sexe, statut juridique, propritaire

3.Cardiographie

Importation des packages nécessaire afin de réalise notre travail sur la cartographie

```
## Linking to GEOS 3.9.3, GDAL 3.5.2, PROJ 8.2.1; sf_use_s2() is TRUE
```

```
## The legacy packages maptools, rgdal, and rgeos, underpinning the sp package,
## which was just loaded, will retire in October 2023.
## Please refer to R-spatial evolution reports for details, especially
```

```
## https://r-spatial.org/r/2023/05/15/evolution4.html.
## It may be desirable to make the sf package available;
## package maintainers should consider adding sf to Suggests:.
## The sp package is now running under evolution status 2
##      (status 2 uses the sf package in place of rgdal)

## This project is in maintenance mode.
## Core functionalities of `cartography` can be found in `mapsf`.
## https://riatelab.github.io/mapsf/

## Please note that rgdal will be retired during October 2023,
## plan transition to sf/stars/terra functions using GDAL and PROJ
## at your earliest convenience.
## See https://r-spatial.org/r/2023/05/15/evolution4.html and https://github.com/r-spatial/evolution
## rgdal: version: 1.6-7, (SVN revision 1203)
## Geospatial Data Abstraction Library extensions to R successfully loaded
## Loaded GDAL runtime: GDAL 3.5.2, released 2022/09/02
## Path to GDAL shared files: C:/Users/Mbare/AppData/Local/R/win-library/4.2/rgdal/gdal
## GDAL binary built with GEOS: TRUE
## Loaded PROJ runtime: Rel. 8.2.1, January 1st, 2022, [PJ_VERSION: 821]
## Path to PROJ shared files: C:/Users/Mbare/AppData/Local/R/win-library/4.2/rgdal/proj
## PROJ CDN enabled: FALSE
## Linking to sp version:2.0-0
## To mute warnings of possible GDAL/OSR exportToProj4() degradation,
## use options("rgdal_show_exportToProj4_warnings"="none") before loading sp or rgdal.

## Please note that 'maptools' will be retired during October 2023,
## plan transition at your earliest convenience (see
## https://r-spatial.org/r/2023/05/15/evolution4.html and earlier blogs
## for guidance);some functionality will be moved to 'sp'.
##  Checking rgeos availability: TRUE

## Data: (c) OpenStreetMap contributors, ODbL 1.0 - http://www.openstreetmap.org/copyright

## Routing: OSRM - http://project-osrm.org/

## Le chargement a nécessité le package : spData

## To access larger datasets in this package, install the spDataLarge
## package with: `install.packages('spDataLarge',
## repos='https://nowosad.github.io/drat/', type='source')`

##
## Attachement du package : 'raster'

## L'objet suivant est masqué depuis 'package:gtsummary':
##
##     select

## L'objet suivant est masqué depuis 'package:dplyr':
##
##     select

## rgeos version: 0.6-4, (SVN revision 699)
##  GEOS runtime version: 3.9.3-CAPI-1.14.3
##  Please note that rgeos will be retired during October 2023,
## plan transition to sf or terra functions using GEOS at your earliest convenience.
## See https://r-spatial.org/r/2023/05/15/evolution4.html for details.
##  GEOS using OverlayNG
```

```
##  Linking to sp version: 2.0-0
##  Polygon checking: TRUE
##
## Attachement du package : 'rgeos'
## L'objet suivant est masqué depuis 'package:dplyr':
##
##     symdiff
##
## Attachement du package : 'igraph'
## L'objet suivant est masqué depuis 'package:rgeos':
##
##     union
## L'objet suivant est masqué depuis 'package:raster':
##
##     union
## Les objets suivants sont masqués depuis 'package:lubridate':
##
##     %--%, union
## Les objets suivants sont masqués depuis 'package:purrr':
##
##     compose, simplify
## L'objet suivant est masqué depuis 'package:tidyr':
##
##     crossing
## L'objet suivant est masqué depuis 'package:tibble':
##
##     as_data_frame
## Les objets suivants sont masqués depuis 'package:dplyr':
##
##     as_data_frame, groups, union
## Les objets suivants sont masqués depuis 'package:stats':
##
##     decompose, spectrum
## L'objet suivant est masqué depuis 'package:base':
##
##     union
##
## Attachement du package : 'plotly'
## L'objet suivant est masqué depuis 'package:igraph':
##
##     groups
## L'objet suivant est masqué depuis 'package:raster':
##
##     select
## L'objet suivant est masqué depuis 'package:ggplot2':
##
```

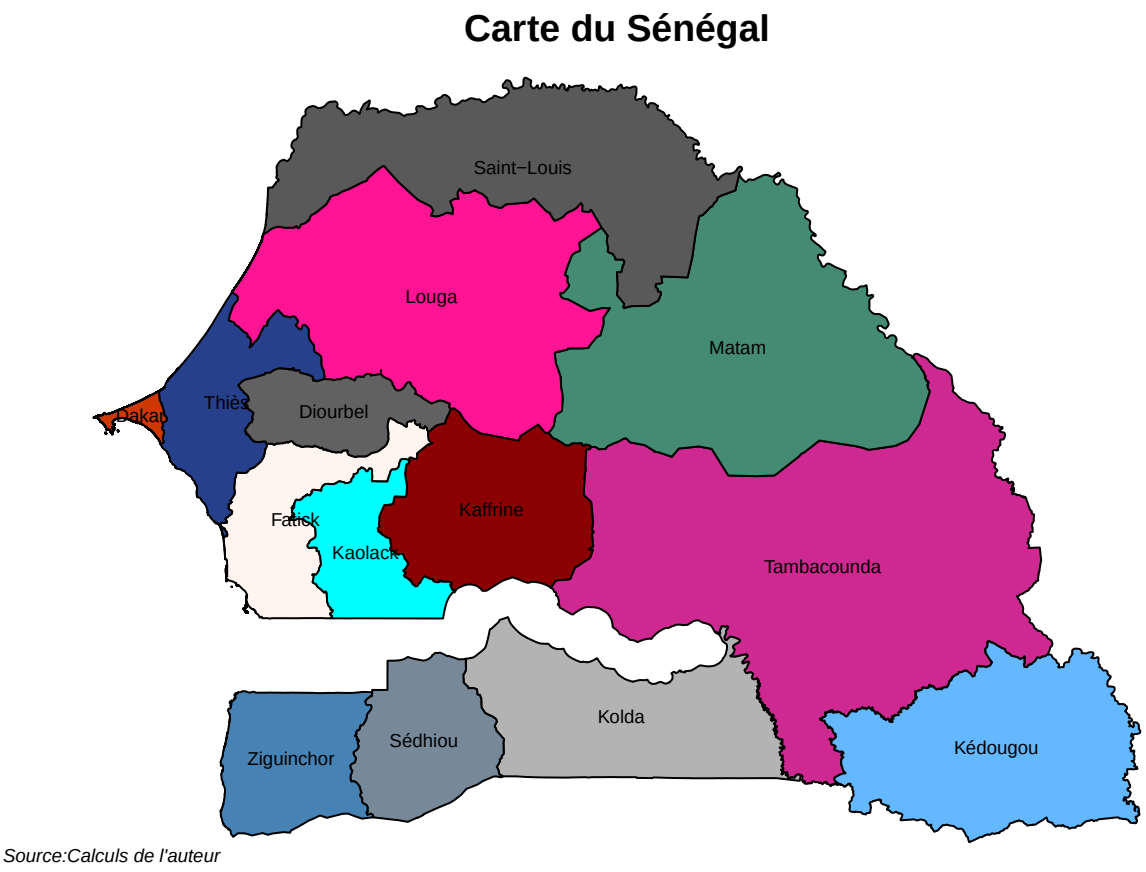
```
##     last_plot
```

```
## L'objet suivant est masqué depuis 'package:stats':
##
##     filter
```

```
## L'objet suivant est masqué depuis 'package:graphics':
##
##     layout
```

#3-1 Transformation du data.frame en données géographiques dont l'objet sera nommé projet_map.

```
## Warning in raster::getData("GADM", country = "Senegal", level = 1): getData will be removed in a futu
## . Please use the geodata package instead
```



Source:Calculs de l'auteur

Pour transformé le data.frame en données géographique on utilise la fonction st_as_sf du package "sf"

```
## [1] "sf"         "data.frame"
```

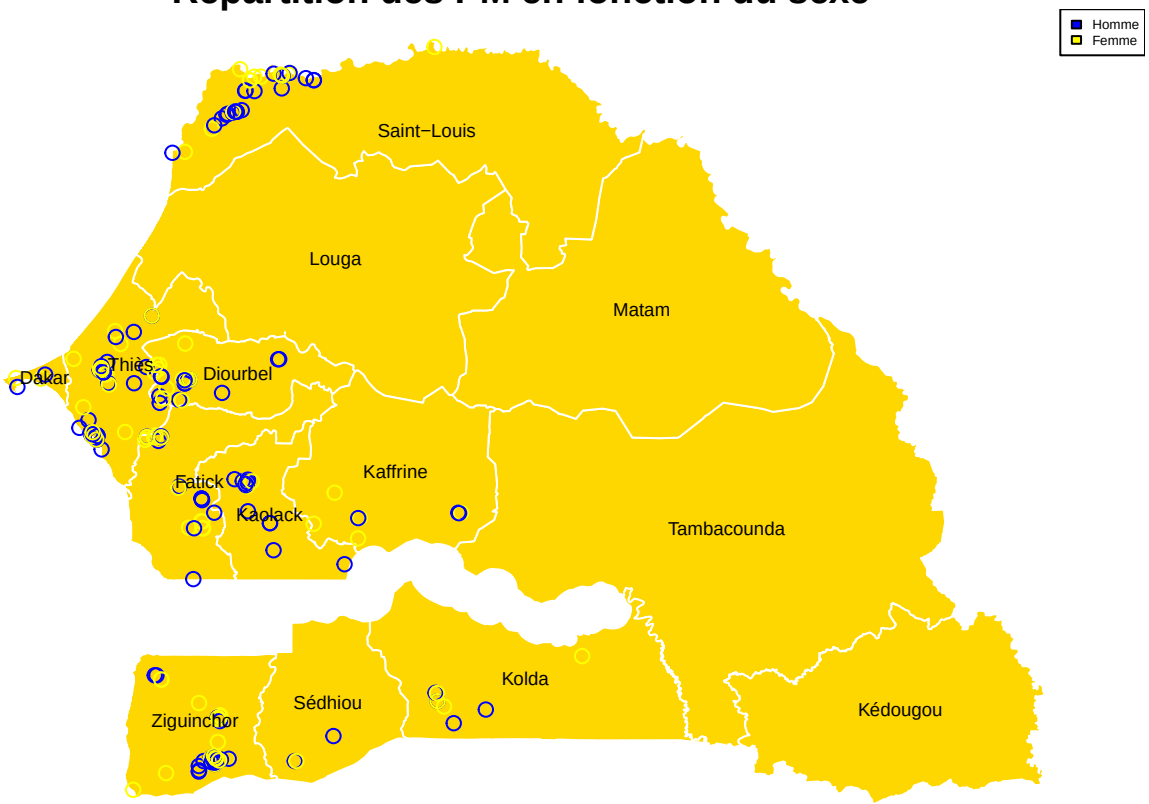
#3-2 Faites une réprésentation spatiale des PME suivant le sexe

```
## Warning in raster::getData("GADM", country = "Senegal", level = 1): getData will be removed in a futu
## . Please use the geodata package instead
```

```
## Warning in plot.sf(projet_map["sexe"], col = c("blue", "yellow"), add = TRUE):
## col is not of length 1 or nrow(x): colors will be recycled; use pal to specify
## a color palette
```

```
## Les objets suivants sont masqués depuis projet:
##
##     filiere_1, filiere_2, filiere_3, filiere_4, key, q12, q14b, q16,
```

```
##     q17, q19, q20, q24, q24a_1, q24a_10, q24a_2, q24a_3, q24a_4,
##     q24a_5, q24a_6, q24a_7, q24a_9, q25, q26, q8, q81, start,
##     submissiondate, today
```



Source:Calculs de l'auteur

#3-3 Faites une réprésentation spatiale des PME suivant le niveau d'instruction

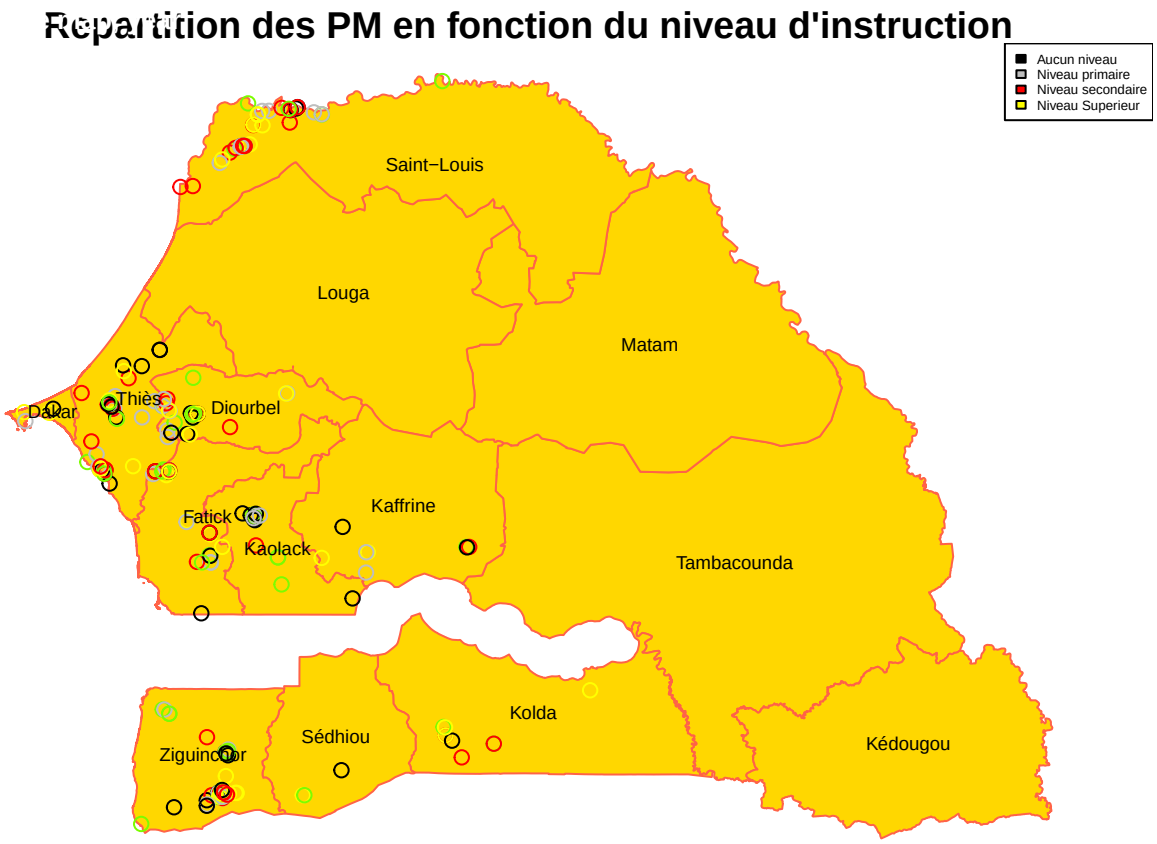
```
## [1] "GIE"         "Informel"    "SUARL"       "SARL"        "Association"
## [6] "SA"
```

```
## Warning in raster::getData("GADM", country = "Senegal", level = 1): getData will be removed in a futu
## . Please use the geodata package instead
```

```
## Warning in plot.sf(projet_map["q12"], col = c("black", "gray", "red", "yellow",
## : col is not of length 1 or nrow(x): colors will be recycled; use pal to
## specify a color palette
```

```
## Les objets suivants sont masqués depuis projet_map (pos = 3):
##
##     departement, filiere_1, filiere_2, filiere_3, filiere_4, geometry,
##     key, q12, q14b, q16, q17, q19, q20, q24, q24a_1, q24a_10, q24a_2,
##     q24a_3, q24a_4, q24a_5, q24a_6, q24a_7, q24a_9, q25, q26, q8, q81,
##     region, sexe, sexe_2, start, submissiondate, today
```

```
## Les objets suivants sont masqués depuis projet:
##
##     filiere_1, filiere_2, filiere_3, filiere_4, key, q12, q14b, q16,
##     q17, q19, q20, q24, q24a_1, q24a_10, q24a_2, q24a_3, q24a_4,
##     q24a_5, q24a_6, q24a_7, q24a_9, q25, q26, q8, q81, start,
##     submissiondate, today
```



Source:Calculs de l'auteur

#3-4 Faites une analyse spatiale de votre choix

Mon choix se fait sur la variable *statut juridique* .En répartissant les PME agroalimentaires par les statuts juridiques sur une carte, nous pouvons obtenir plusieurs informations intéressantes qui pourraient nous aider à mieux comprendre le paysage économique des PME dans la région, **Distribution géographique des statuts juridiques** : nous pourrez visualiser comment les différents statuts juridiques sont répartis géographiquement dans la région et cela nous permettra de voir s'il y a des concentrations particulières de certains types de statuts juridiques dans des régions du Sénégal.

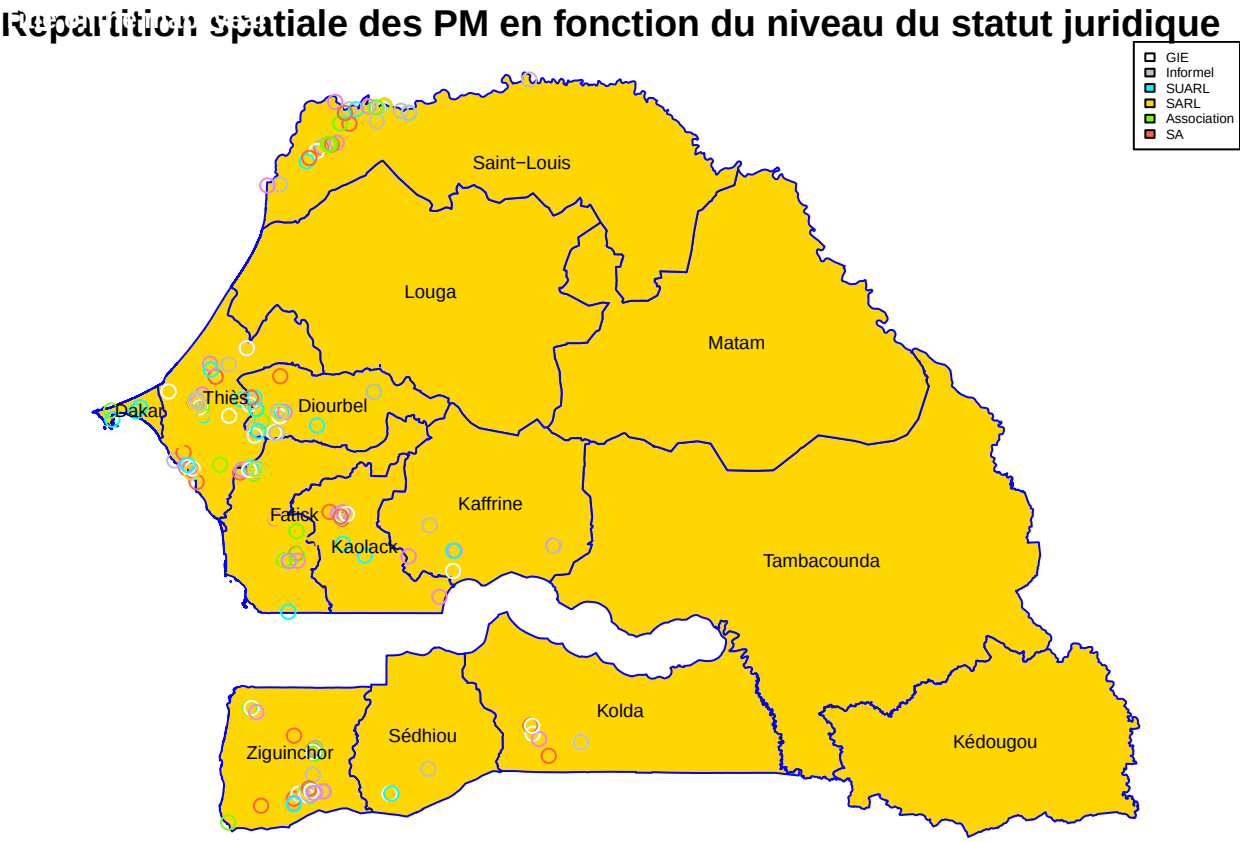
```
## Warning in raster::getData("GADM", country = "Senegal", level = 1): getData will be removed in a futu
## . Please use the geodata package instead
```

```
## Warning in plot.sf(projet_map["q12"], col = c("azure", "gray", "cyan", "gold",
## : col is not of length 1 or nrow(x): colors will be recycled; use pal to
## specify a color palette
```

```
## Les objets suivants sont masqués depuis projet_map (pos = 3):
##
##     departement, filiere_1, filiere_2, filiere_3, filiere_4, geometry,
##     key, q12, q14b, q16, q17, q19, q20, q24, q24a_1, q24a_10, q24a_2,
##     q24a_3, q24a_4, q24a_5, q24a_6, q24a_7, q24a_9, q25, q26, q8, q81,
##     region, sexe, sexe_2, start, submissiondate, today
```

```
## Les objets suivants sont masqués depuis projet_map (pos = 4):
##
##     departement, filiere_1, filiere_2, filiere_3, filiere_4, geometry,
##     key, q12, q14b, q16, q17, q19, q20, q24, q24a_1, q24a_10, q24a_2,
##     q24a_3, q24a_4, q24a_5, q24a_6, q24a_7, q24a_9, q25, q26, q8, q81,
##     region, sexe, sexe_2, start, submissiondate, today
```

```
## Les objets suivants sont masqués depuis projet:
##
##     filiere_1, filiere_2, filiere_3, filiere_4, key, q12, q14b, q16,
##     q17, q19, q20, q24, q24a_1, q24a_10, q24a_2, q24a_3, q24a_4,
##     q24a_5, q24a_6, q24a_7, q24a_9, q25, q26, q8, q81, start,
##     submissiondate, today
```



Source:Calculs de l'auteur

###PARTIE 2

Nettoyage et gestion des données

Renommer la variable "country_destination" en "destination" et définir les valeurs négatives comme manquantes.

```
##  [1]  3 11 13 14 21 27 29 30 39 53 56 58 67 71 74 78 83 85 87 89
```

```
## integer(0)
```

```
##
##  3  4  5  6  8  9 10 11 13
##  7  2  8  3 10 22 18  1  6
```

Partie 2

2-1 Nettoyage et gestion des données

2-1-1 Créer une nouvelle variable contenant des tranches d'âge de 5 ans en utilisant la variable *"age"*.

```
## [1] 15 20 25 30 35 40
```

```
##
## (15,20] (20,25] (25,30] (30,35] (35,40]
##      20      34      22      10      10
```

#2-1-2 Créer une nouvelle variable contenant le nombre d'entretiens réalisés par chaque agent recenseur.

```
## Les objets suivants sont masqués depuis donne_art (pos = 3):
##
##     age, children_num, destination, district, endtime, enumerator, id,
##     intention, sex, starttime
```

#2-1-3 Créer une nouvelle variable qui affecte aléatoirement chaque répondant à un groupe de traitement (1) ou de controle (0)

```
## Les objets suivants sont masqués depuis donne_art (pos = 3):
##
##     age, age_aberrante, children_num, destination, district, endtime,
##     enumerator, id, intention, sex, starttime
```

```
## Les objets suivants sont masqués depuis donne_art (pos = 4):
##
##     age, children_num, destination, district, endtime, enumerator, id,
##     intention, sex, starttime
```

#2-1-4 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.

#2-1-5 Calculer la durée de l'entretien et indiquer la durée moyenne de l'entretien par enquêteur.

```
##  [1] "district"       "id"             "starttime"      "endtime"
##  [5] "enumerator"     "age"            "sex"            "children_num"
##  [9] "intention"      "destination"    "age_aberrante"  "nbre_entretien"
## [13] "grp_traite"     "population"     "dure_entre"
```

enumerator	moyenne
6	0.4307778
14	0.4260185
11	0.5580556
20	0.4794753
18	0.6143056
13	0.5265972
4	0.6080556
1	1.1357778
12	0.8027778
8	0.6688426
15	0.4775000
9	1.9127778
10	0.9212778
5	0.5593056
17	0.4881019
7	0.6194048

#2-1-6 Renommez toutes les variables de l'ensemble de données en ajoutant le préfixe "endline_" à l'aide d'une boucle.

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.

2-2-2 Testez si la différence d'âge entre les sexes est statistiquement significative au niveau de 5 %.

2-2-3Créer un nuage de points de l'âge en fonction du nombre d'enfants

2-2-4La 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.

2-2-4 Créez 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.

Partie 3

Cette partie correspond au dossier *application-shiny* plus précisement à l'application *App.R*