TP 1 - Logiciel statistique R

Enseignant: Aboubacar HEMA

A rendre avant le Dimanche 7 Avril 2024

**Résumé**

L’objectif de ce projet est que vous appliquiez les outils que vous avez étudiés dans le cours du logiciel statistique R, dans le cas d’une étude de cas réelle. Le TP est à faire seul et à rendre au format **.html** ou **.pdf**. Les codes que vous utiliserez pour répondre aux questions seront à intégrer dans le corps de votre rapport. C’est pourquoi l’utilisation de **R Markdown** est à privilégier, mais vous pouvez également utiliser d’autres éditeurs de textes. . . En ce qui concerne l’organisation du travail à rendre, vous vous inspirerez de la façon dont est organisé le sujet du projet (vous n’êtes pas obliger de recopier l’énoncé). Le barème est indiqué en face de chaque question. Il tient compte du code proposé et dans le cas où nous vous demandons de faire des commentaires sur les résultats obtenus, nous tiendrons compte des commentaires que vous aurez fait dans la notation.

**IMPORTANT** : il est demandé de commenter vos lignes de code. Vous pourriez être pénalisés si vous ne commentez pas vos codes.

# Objectif de l’étude

Cette enquête vise à identifier et à caractériser des bioénergies durables pour les petites et moyennes entreprises (PME) agroalimentaires d’Afrique de l’Ouest.

# Préparation des données

**1.1 Description**

Le fichier **Base\_Projet.xlsx** contient 250 observations et 33 variables.

La première colonne **key** correspond à l’identifiant de la PME. Les variables sont les suivantes :

* *q1*: Région
* *q2*: Département
* *q23*: Sexe du dirigeant/responsable de la PME
* *q24*: Age du dirigeant/responsable de la PME
* *q24a\_1*=1 si le dirigeant/propriétaire de la PME parle le français
* *q24a\_2*=1 si le dirigeant/propriétaire de la PME parle le wolof
* *q24a\_3*=1 si le dirigeant/propriétaire de la PME parle le Diola
* *q24a\_4*=1 si le dirigeant/propriétaire de la PME parle le Serere
* *q24a\_5*=1 si le dirigeant/propriétaire de la PME parle le Peul
* *q24a\_6*=1 si le dirigeant/propriétaire de la PME parle le Mandingue
* *q24a\_7*=1 si le dirigeant/propriétaire de la PME parle le Balante
* *q24a\_9*=1 si le dirigeant/propriétaire de la PME parle le Bambara
* *q24a\_10*=1 si le dirigeant/propriétaire de la PME parle une autre langue
* *q25*: Niveau d’instruction du dirigeant/responsable de la PME
* *q26*: Nombre d’années d’experience professionnelle du dirigeant/responsable de la PME dans l’entreprise
* *q12*: Statut juridique (SARL, SA, SUARL, GIE, Association, Groupement,) • *q14b*: Autorisation de fabrication et de mise en vente (FRA)
* *q16*: L’entreprise est-elle désservie par une route bitumée ?
* *q17*: Etat de la route bitumée
* *q19*: l’état de la piste qui mène à l’entreprise
* *q20*: Avez-vous des associés dans l’entreprise
* *filiere\_1*=1 si la PME est dans la filière **arachide**
* *filiere\_2*=1 si la PME est dans la filière **anacarde**
* *filiere\_3*=1 si la PME est dans la filière **mangue** • *filiere\_4*=1 si la PME est dans la filière **riz**
* *q8*: Activité principale de l’entreprise :
* *q81*: propriétaire ou locataire
* *gps\_menlatitude*: coordonnées géographiques de la PME (latitude)
* *gps\_menlongitude*: coordonnées géographiques de la PME (longitude)
* *submissiondate*: la date de soumission des informations de la PME
* *start*: la date de début de l’enrégistrement des informations de la PME par l’enquêteur
* *today*: la date de l’enquête

## Importation et mise en forme (2.5 pts)

* Importer la base de données dans un objet de type **data.frame** nommé **projet**
* Donner le nombre de lignes (i.e. le nombre de PME) et colonnes (i.e. nombre de variables) de la base

**projet**

* 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).

## Création de variables (2.5 pts)

* Rénommer la variable **q1** en **region**
* Rénommer la variable **q2** en **departement**
* Rénommer la variable **q23** en **sexe**
* Créer la variable **sexe\_2** qui vaut **1** si **sexe égale à Femme** et **0** sinon.
* Créer un **data.frame** nommé **langues** qui prend les variables **key** et les variables correspondantes décrites plus haut.

**Indication**: Vous remarquerez que ces variables commencent par **q24a\_**.

* Créer une variable **parle** qui est égale au nombre de langue parlée par le dirigeant de la PME.
* Sélectionnez uniquement les variables **key** et **parle**, l’objet de retour sera **langues**.
* Merger les **data.frame projet** et **langues**:

# Analyses descriptives (10 pts)

Quelle est la répartion des PME suivant:

* le sexe?
* le niveau d’instruction?
* le statut juridique?
* le propriétaire/locataire?
* le statut juridique et le sexe?
* le niveau d’instruction et le sexe?
* Propriétaire/locataire suivant le sexe?

**Faite les statistiques descriptives de votre choix sur les autres variables**

**IMPORTANT**: Il faut créer deux fonctions *univarie()* pour des statistiques descriptives univariées et une autre fonction *bivarie()* pour plusieurs variables.

**Recommandations** :

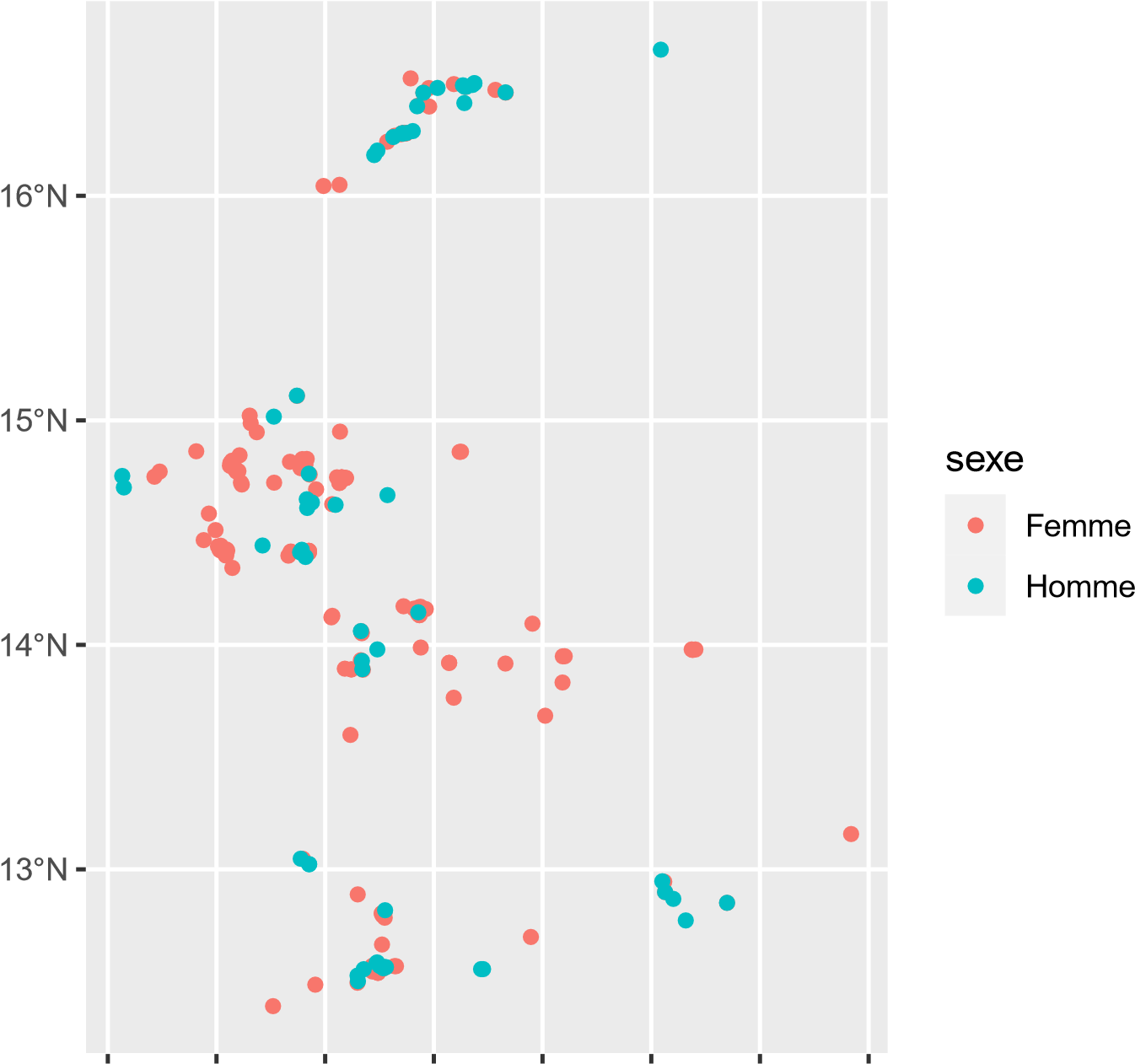
1. nous tiendrons compte de la façon dont vous présenterez les résultats en sortie : la qualité du graphique, l’affichage des résultats, etc.
2. quand vous créerez votre fonction, il faudra faire en sorte qu’elle soit le plus général possible. Par exemple, donner à l’utilisateur la possibilité de la réutiliser sur d’autres jeux de données, utiliser des options graphiques, etc.

**Indication**:

* + Priorisez une analyse par **filière**
  + Utilisez également des variables sur les dates

# Un peu de cartographie (5 pts)

* Transformer le **data.frame** en données géographiques dont l’objet sera nommé **projet\_map**.
* Trouver le code qui permet d’avoir la carte ci-dessous:



## 17.5°W17.0°W16.5°W16.0°W15.5°W15.0°W14.5°W14.0°W

* Donner des axes, titre, légende, orientation, échelle à la carte ci-dessus Nous allons utiliser maintenant la base **SEN\_adm**. Nous rappelons que:
* **SEN\_adm0** répresente le niveau 0 (la carte du Sénegal sans les découpages administratifs).
* **SEN\_adm1** répresente le niveau 1 (la carte du Sénegal suivant les régions).
* **SEN\_adm2** répresente le niveau 2 (la carte du Sénegal suivant les départements).
* **SEN\_adm3** répresente le niveau 3 (la carte du Sénegal suivant les communes).
* Importer et afficher les différentes cartes suivant les niveaux décrits ci-dessus.
* Faites une réprésentation spatiale des PME suivant le sexe
* Faites une réprésentation spatiale des PME suivant le niveau d’instruction
* Faites une analyse spatiale de votre choix