Rapport

Grimaux Nicolas Sensey Valentin

27/10/2019

Table of Contents

# Présentation des données



coucoulol

Le travail demandé consiste à analyser un jeu de donnée concernant les députés français. Les sources des données sont les suivantes :

[Projet arcadie](https://projetarcadie.com/tableaux-thematiques)

[Données de l’Assemblée Nationale](http://data.assemblee-nationale.fr/)

Les variables et les méthodes à utiliser pour effectuer les analyses sont libres. Le jeu de données fourni est un fichier de format *.xlsx* comprenant 4 tables :

* **deputestable1** : comprend 575 observations et 11 variables notamment les Nom, Prénom, Région, Profession et Groupe politique (complet) et Groupe politique(abrégé)
* **deputestable2** : comprend 576 observations et 12 variables. Mis à part le Nom et le prénom cette table comprend les dates de naissance des députés, le statut des députés, les catégories de fonctionnaires…
* **deputestable3** : comprend 576 obersvations et 4 variables. Cette table récapitule l’historique parlementaire de chaque député.
* **deputestable4** : comprend 573 observations et 45 variables. Cette table récapitule tout ce qu’il y a à savoir concernant les duels politiques des dernières éléctions (abstention, candidat en face du vainqueur, etc)

# Travail préparatoire

## Prérequis

Voici les librairies requises pour notre analyse. Certains librairies nécessitent d’avoir Java correctement installé.

install.packages("xlsx")  
install.packages("readxl")  
install.packages("questionr")  
install.packages("ggplot2")  
install.packages("scales")

library(scales)  
library(xlsx)  
library(questionr)  
library(readxl)  
library(ggplot2)

## Importation des données et fusion des tables

Nous allons maintenant importer les données dans RStudio. Pour cela nous allons d’abord séparer notre fichier *.xlsx* comprenant 4 tables en 4 fichiers *.xlsx* comprenant chacun une table. Une fois ceci fait, nous allons fusionner nos 4 tables.

A l’issue de la fusion la table complète ne comprenait plus que 571 individus, nous avons donc cherché à savoir pourquoi ces derniers n’ont pas “matchés”. Il s’agissait en fait de problèmes d’accent, de majuscules ou de deuxième prénom. Voici les individus problématiques :

* Amal-Amélia Lakrafi : Dans les Table 2 et 3 elle se nomme “Amal-Amélia Lakrafi” et dans la Table 1 “Amélia Lakrafi”.
* Constance le Grip : Dans une des tables le “le” était écrit “Le”.
* Thierry Benoît : Présence de l’accent circonflexe dans certaines tables et pas dans les autres.
* Patricia Miralles : Le “e” est accentué ou non suivant la table.
* Pierre Venteau TABLE1: N’existe pas dans la table 1.

Nous allons corriger les problèmes d’accent et de second prénom avant d’importer les données. Ce faisant, on corrige bien les problemes de merge. Une seule personne disparait, cependant on veut la garder donc nous rajouterons l’argument “all.x=TRUE” lors de la fusion.

Il y a également des problèmes avec la table 4. Elle comprend plusieurs colonnes avec le même nom et la première ligne de la table ne comprend pas les intitulés de colonnes. Nous allons donc la nettoyer au préalable directement sur le fichier *.xlsx*.

## New names:  
## \* `N°Panneau` -> `N°Panneau...19`  
## \* Sexe -> Sexe...20  
## \* Nom -> Nom...21  
## \* Prénom -> Prénom...22  
## \* Nuance -> Nuance...23  
## \* … and 22 more problems

Il convient également de créer une nouvelle colonne Titre dans **deputestable1** pour pouvoir la fusionner avec les autres. Enfin, une fois que nous avons la même variable pour toutes les tables, nous la passons en majuscules afin d’éviter les erreurs mentionnées précédemment. Nous obtenons donc une table unique que nous appelerons **deputes**.

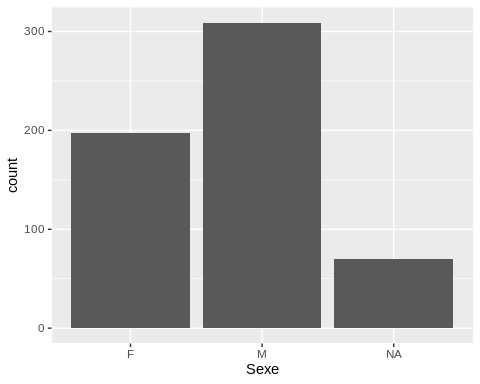
## [1] "On enlève les prenoms et noms en trop."

#Analyse

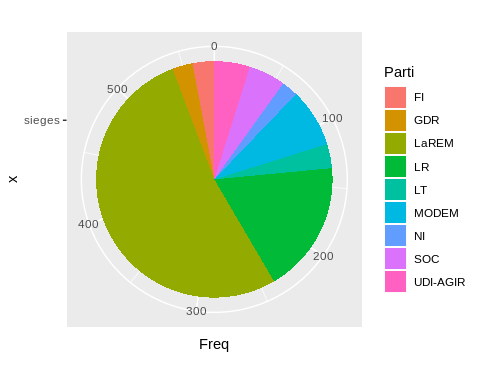
*Note : A partir de maintenant, le code ne figurera plus dans le corps du document mais sera mis en annexe*

## Répartition du genre au sein de l’AN

##   
## F M   
## 197 309



## Répartition des partis politques au sein de l’assemblée



## [1] "Je voudrai faire un graphique sur la frequence de premier mandat par partie politique"

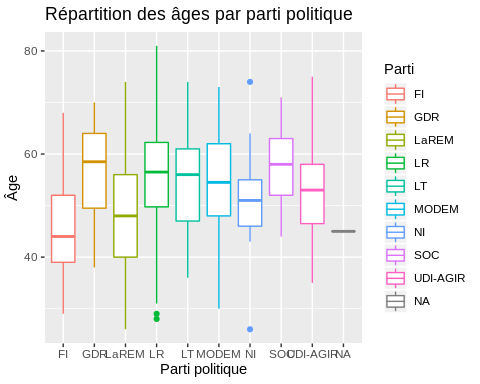
## Focus sur l’âge

To do : rappeler qu’on crée la colonne âge et dire que le code est en annexe

### Répartion générale des âges au sein de l’assemblée



### Répartition des âges par parti politique



### Âge moyen des députés par parti politique

