R Notebook

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.5 v dplyr 1.0.7  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.0.2 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(ggplot2)  
library(readxl)

# Notes préliminaires

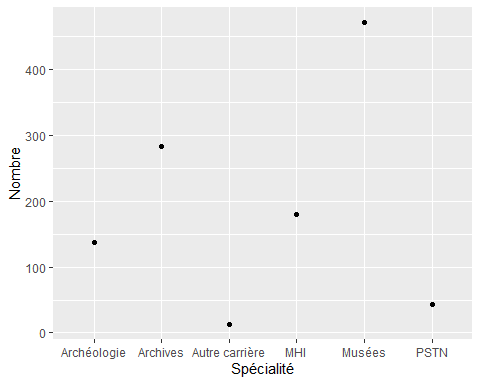
Le travail est réalisé sur la base d’un scrapping des données de l’annuaire de l’INP.

Le fichier a été nettoyé :

* Les élèves internationaux n’ont pas été conservés
* Suppression de doublons (femmes apparaissant sous leur nom de jeune fille puis sous le nom d’épouse)
* Le sexe a été ajouté manuellement (vérification par recherche soit sur l’annuaire, soit sur un moteur de recherche en cas de doute)
* Les spécialités ont été harmonisées
* Diverses coquilles et erreurs ont été corrigées (nom des promo)

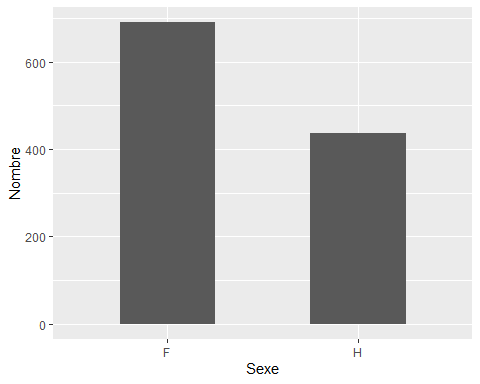
# Nombre de personne par spécialités

data <- read\_excel(path = "annuaireINP\_complet.xls")  
data %>%  
 count(Spécialité, name= "Nombre") %>%  
 ggplot( aes(x=Spécialité, y=Nombre)) +  
 geom\_point()



# Nombre de femmes et d’hommes quelque soit la spécialité

data <- read\_excel(path = "annuaireINP\_complet.xls")  
data %>%   
 count(Sexe, name= "Nombre") %>%  
 ggplot(aes(x=Sexe, y=Nombre)) +  
 geom\_bar(stat = "identity", width = 0.5)

 # Répartition par spécialité

## Archéologie

data <- read\_excel(path = "annuaireINP\_complet.xls")  
data %>%  
 filter(Spécialité == "Archéologie") %>%   
 count(Sexe, name = "Nombre")

## # A tibble: 2 x 2  
## Sexe Nombre  
## <chr> <int>  
## 1 F 70  
## 2 H 67

## Archives

data <- read\_excel(path = "annuaireINP\_complet.xls")  
data %>%  
 filter(Spécialité == "Archives") %>%   
 count(Sexe, name = "Nombre")

## # A tibble: 2 x 2  
## Sexe Nombre  
## <chr> <int>  
## 1 F 160  
## 2 H 123

## Autre carrière

data <- read\_excel(path = "annuaireINP\_complet.xls")  
data %>%  
 filter(Spécialité == "Autre carrière") %>%   
 count(Sexe, name = "Nombre")

## # A tibble: 2 x 2  
## Sexe Nombre  
## <chr> <int>  
## 1 F 8  
## 2 H 5

## MHI

data <- read\_excel(path = "annuaireINP\_complet.xls")  
data %>%  
 filter(Spécialité == "MHI") %>%   
 count(Sexe, name = "Nombre")

## # A tibble: 2 x 2  
## Sexe Nombre  
## <chr> <int>  
## 1 F 116  
## 2 H 64

## Musées

data <- read\_excel(path = "annuaireINP\_complet.xls")  
data %>%  
 filter(Spécialité == "Musées") %>%   
 count(Sexe, name = "Nombre")

## # A tibble: 2 x 2  
## Sexe Nombre  
## <chr> <int>  
## 1 F 311  
## 2 H 161

## PSTN

data <- read\_excel(path = "annuaireINP\_complet.xls")  
data %>%  
 filter(Spécialité == "PSTN") %>%   
 count(Sexe, name = "Nombre")

## # A tibble: 2 x 2  
## Sexe Nombre  
## <chr> <int>  
## 1 F 26  
## 2 H 17

# Evolution par promotion

promo <- read\_excel(path = "annuaireINP\_complet.xls")  
promo %>%  
 group\_by(Année) %>%   
 count(Sexe, name = "Nombre")

## # A tibble: 60 x 3  
## # Groups: Année [30]  
## Année Sexe Nombre  
## <dbl> <chr> <int>  
## 1 1991 F 27  
## 2 1991 H 16  
## 3 1992 F 12  
## 4 1992 H 19  
## 5 1993 F 25  
## 6 1993 H 10  
## 7 1994 F 12  
## 8 1994 H 14  
## 9 1995 F 19  
## 10 1995 H 17  
## # ... with 50 more rows