Analyse données INP

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

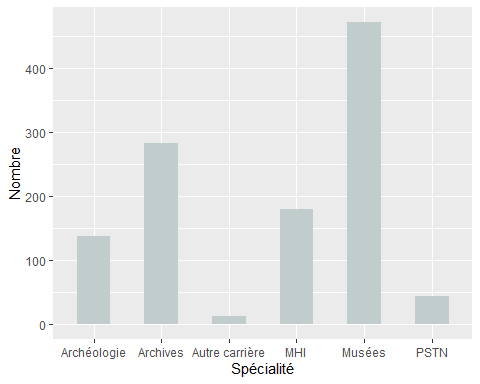
Le travail est réalisé sur la base d’un scrapping des données de [l’annuaire de l’INP](https://ent.inp.fr/mod/data/view.php?id=3227).

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 leur 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)

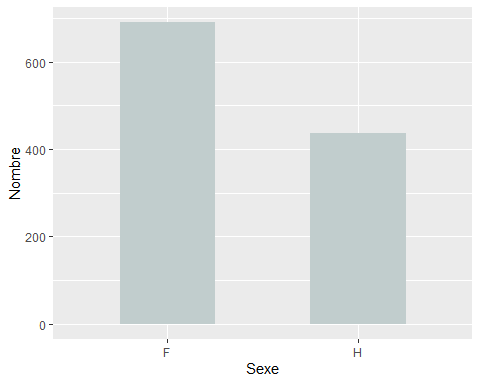
# Nombre de personne par spécialités

data <- read\_excel(path = "annuaireINP\_complet.xls")  
data %>%  
 count(Spécialité, name= "Nombre") %>%  
 ggplot( aes(x=Spécialité, y=Nombre)) +  
 geom\_bar(stat = "identity", width = 0.5, fill = "azure3")



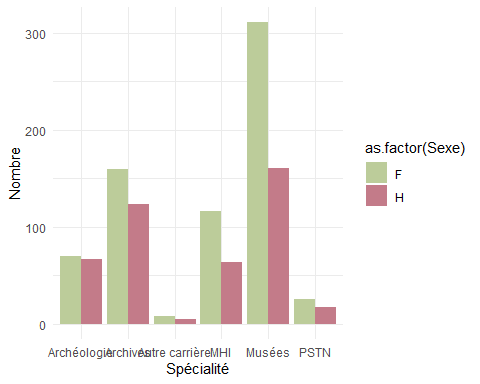
# Nombre de femmes et d’hommes toutes spécialités confondues

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, fill = "azure3")



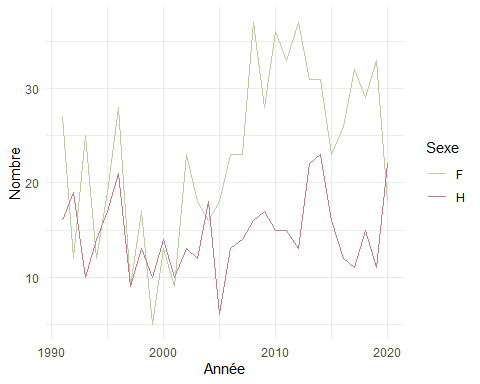
# Répartition femmes/hommes par spécialité

ggplot(data,aes(as.factor(Spécialité), fill=as.factor(Sexe)))+  
 geom\_bar(position = "dodge") +  
 labs(  
 x = "Spécialité",  
 y = "Nombre",  
 colour = "Sexe"  
 ) +  
 scale\_fill\_manual(values=c('#BCCC9A','#C37B89')) +  
 theme\_minimal()



# Evolution de la répartition femmes/hommes dans le temps

promo <- read\_excel(path = "annuaireINP\_complet.xls")  
promo %>%  
 group\_by(Année) %>%   
 count(Sexe, name = "Nombre") %>%   
 ggplot(aes(x = Année, y = Nombre, color = Sexe)) +  
 geom\_line() +  
 scale\_color\_manual(values=c('#BCCC9A','#C37B89')) +  
 theme\_minimal()



# 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 dans le temps et par spécialité

data <- read\_excel(path = "annuaireINP\_complet.xls")  
facettes <- data %>%  
 group\_by(Spécialité, Année) %>%   
 count(Sexe)  
  
facettes

## # A tibble: 267 x 4  
## # Groups: Spécialité, Année [147]  
## Spécialité Année Sexe n  
## <chr> <dbl> <chr> <int>  
## 1 Archéologie 1991 H 2  
## 2 Archéologie 1992 H 2  
## 3 Archéologie 1993 F 1  
## 4 Archéologie 1993 H 2  
## 5 Archéologie 1994 H 1  
## 6 Archéologie 1995 H 3  
## 7 Archéologie 1996 F 2  
## 8 Archéologie 1996 H 5  
## 9 Archéologie 1997 F 2  
## 10 Archéologie 1997 H 1  
## # ... with 257 more rows

ggplot(data = facettes, mapping = aes(x = Année, y = n, color = Sexe)) +  
 geom\_line() +  
 facet\_wrap(facets = vars(Spécialité)) +  
 scale\_color\_manual(values=c('#BCCC9A','#C37B89')) +  
 labs(  
 x = "Année",  
 y = "Nombre",  
 colour = "Sexe"  
 ) +  
 theme\_light()

