

Analyse croisée

Margot Goudard

2023-2024

```
library(ggplot2)
library(dplyr)
```

```
##
## 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
```

```
library(plotly)
```

```
##
## Attachement du package : 'plotly'

## 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
```

```
library(stringi)
library(forcats)
```

```
# Diagramme 1 - Rémunération en fonction du secteur d'activité
```

```
data1 <- read.csv("merged_database.csv") %>%
  filter(!is.na(secteur_premiere_entreprise) & secteur_premiere_entreprise != "",
```

```

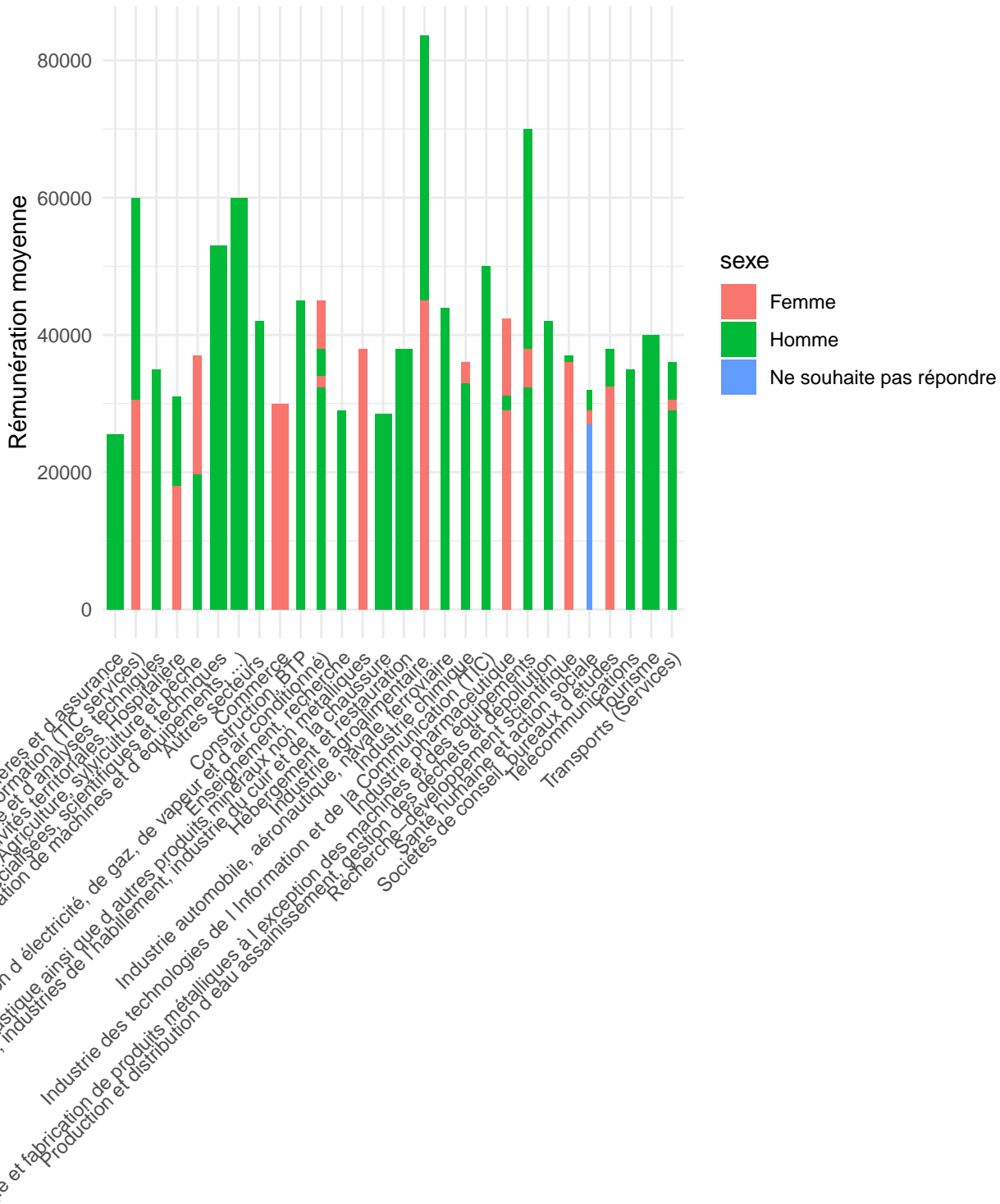
!is.na(remuneration_annuelle_brute_avec_prime_premier_emploi) & remuneration_annuelle_brute_av
!is.na(sexe) & sexe != "")

data1$secteur_premiere_entreprise <- gsub("", " ", data1$secteur_premiere_entreprise)

print(ggplot(data1, aes(x = secteur_premiere_entreprise, y = remuneration_annuelle_brute_avec_prime_prem
  geom_bar(stat = "identity", position = position_dodge(width = 0), width = 0.8) +
  labs(title = "Rémunération Premier Emploi en fonction du secteur d'activité",
        x = "Secteur",
        y = "Rémunération moyenne") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)))

```

Rémunération Premier Emploi en fonction du secteur d'activité



Secteur

```

# Diagramme 2 - Rémunération en fonction de la localisation géographique
# Charger les données
data2 <- read.csv("merged_database.csv") %>%
  filter(!is.na(remuneration_annuelle_brute_avec_prime_premier_emploi) & remuneration_annuelle_brute_avec_prime_premier_emploi > 0)

# Standardiser les valeurs de localisation
data2$localisation_premier_emploi <- tolower(iconv(data2$localisation_premier_emploi, to = "UTF-8", sub = "?"))
data2$localisation_premier_emploi <- stringi::stri_trans_general(data2$localisation_premier_emploi, "Latin")
data2$localisation_premier_emploi <- gsub("'", " ", data2$localisation_premier_emploi)

# Remplacer les valeurs spécifiques dans la colonne 'pays_premier_emploi'
data2 <- data2 %>%
  mutate(pays_premier_emploi = ifelse(pays_premier_emploi == "Je travaillais depuis la Belgique pour un certain temps", "Belgique", pays_premier_emploi))
  mutate(pays_premier_emploi = ifelse(pays_premier_emploi %in% c("Danemark", "Danemark "), "Danemark", pays_premier_emploi))

# Ajouter "France" à la variable pays_premier_emploi lorsque localisation_premier_emploi est vide
data2 <- data2 %>%
  mutate(pays_premier_emploi = ifelse(localisation_premier_emploi == "en france", "France", pays_premier_emploi))

# Tracer le diagramme
print(ggplot(data2, aes(x = localisation_premier_emploi, y = remuneration_annuelle_brute_avec_prime_premier_emploi)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Rémunération Premier Emploi en fonction de la localisation géographique",
    x = "Localisation",
    y = "Rémunération moyenne") +
  theme_minimal())

```

Rémunération Premier Emploi en fonction de la localisation géographique

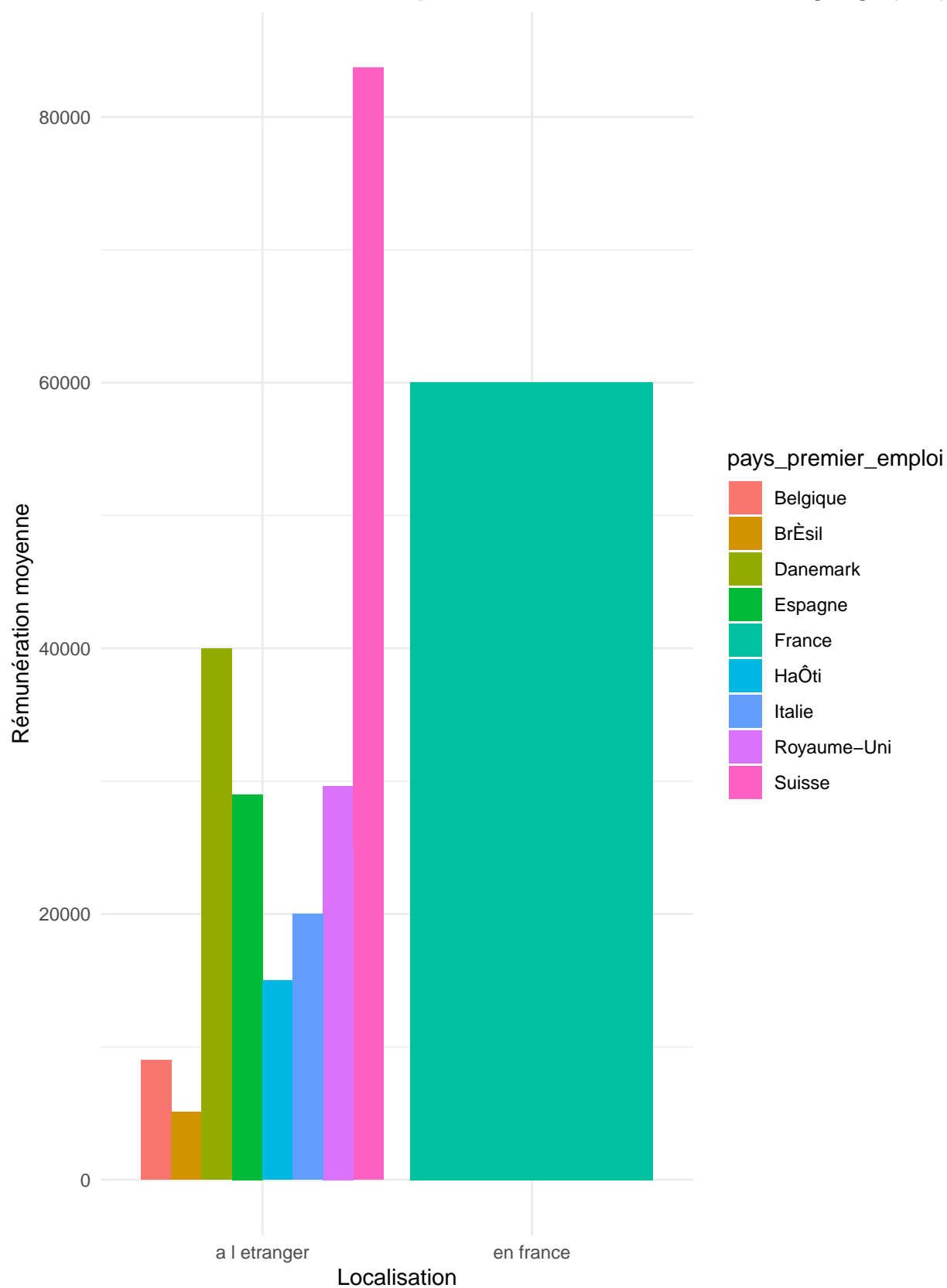


Diagramme 3 - Rémunération Premier Emploi en fonction de la Filière et de la Date de Diplôme

```
data3 <- read.csv("merged_database.csv") %>%
  filter(!is.na(filiere) & filiere != "",
         !is.na(remuneration_annuelle_brute_avec_prime_premier_emploi) & remuneration_annuelle_brute_avec_prime_premier_emploi != "",
         !is.na(date_diplome) & date_diplome != "")

data3$filiere <- fct_collapse(data3$filiere,
  "Eau et Génie Civil (EGC - apprentissage)" = c("Eau et Génie Civil (EGC - apprentissage)", "Eau et Génie Civil (EGC - apprentissage)"),
  "Génie Biologique et Agroalimentaires" = c("Génie Biologique et Agroalimentaires (GBA)", "Génie Biologique et Agroalimentaires (GBA)"),
  "Matériaux" = c("Matériaux (MAT)", "Matériaux (MAT)"),
  "Mécanique et Interactions" = c("Mécanique et Interactions (MI)", "Mécanique et Interactions (MI)"),
  "Mécanique Structures Industrielles" = c("Mécanique Structures Industrielles (MSI - apprentissage)", "Mécanique Structures Industrielles (MSI - apprentissage)"),
  "Microélectronique Et Automatique" = c("Microélectronique Et Automatique (MEA)", "Microélectronique Et Automatique (MEA)"),
  "Sciences et Technologies de l'Eau" = c("Sciences et Technologies de l'Eau (STE)", "Sciences et Technologies de l'Eau (STE)"),
  "Systèmes Embarqués" = c("Systèmes Embarqués (SE - apprentissage)", "Systèmes Embarqués (SE - apprentissage)"),
)

ggplot(data3, aes(x = filiere, y = remuneration_annuelle_brute_avec_prime_premier_emploi, color = date_diplome)) +
  geom_point() +
  labs(title = "Rémunération Premier Emploi en fonction de la Filière et de la Date de Diplôme",
       x = "Filière",
       y = "Rémunération Annuelle Brute",
       color = "Date de Diplôme") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Rémunération Premier Emploi en fonction de la Filière et de la Date de D

