

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 != "",
```

```

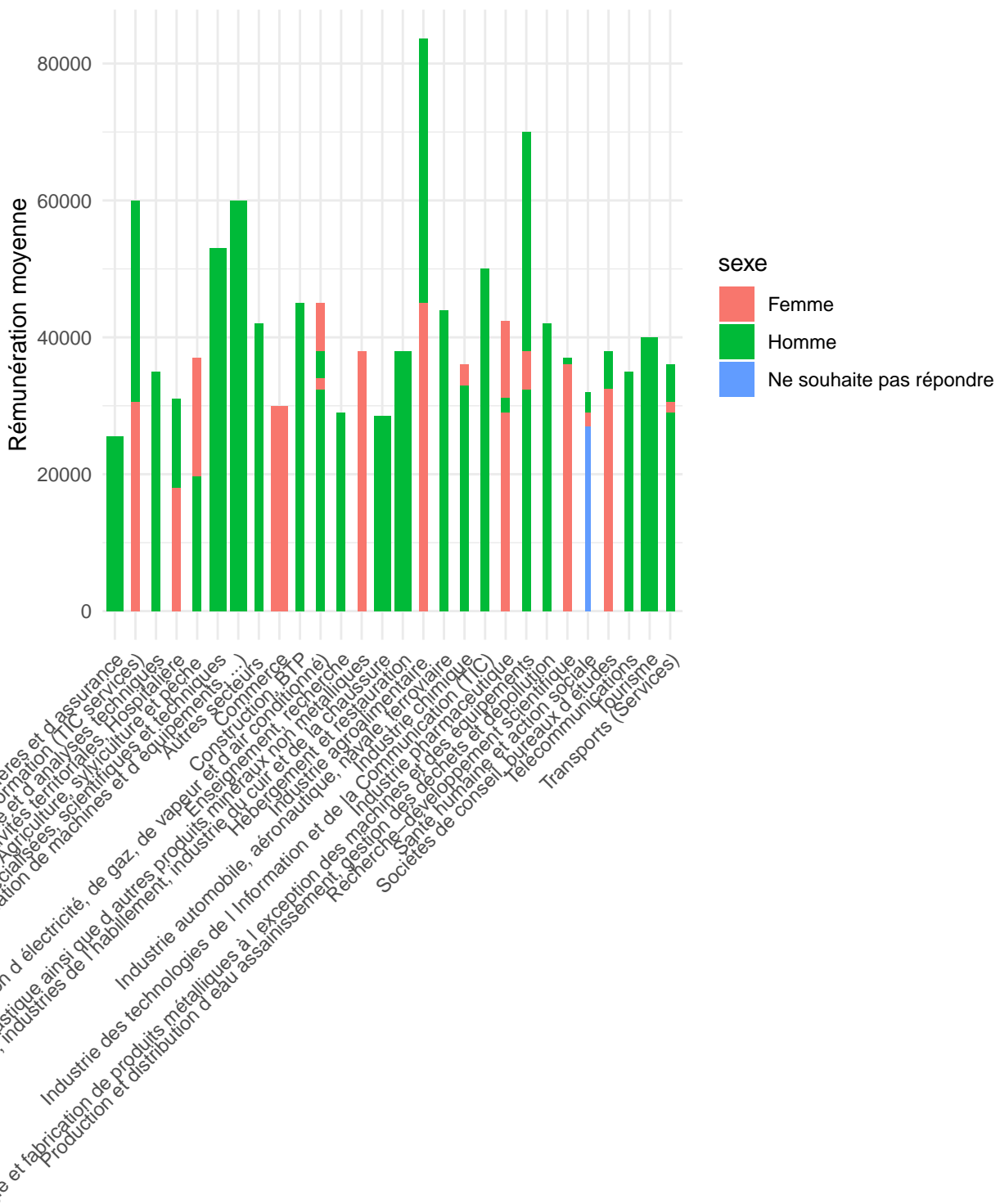
      !is.na(remuneration_annuelle_brute_avec_prime_premier_emploi) & remuneration_annuelle_brute_av
      !is.na(sexe) & sexe != "")

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

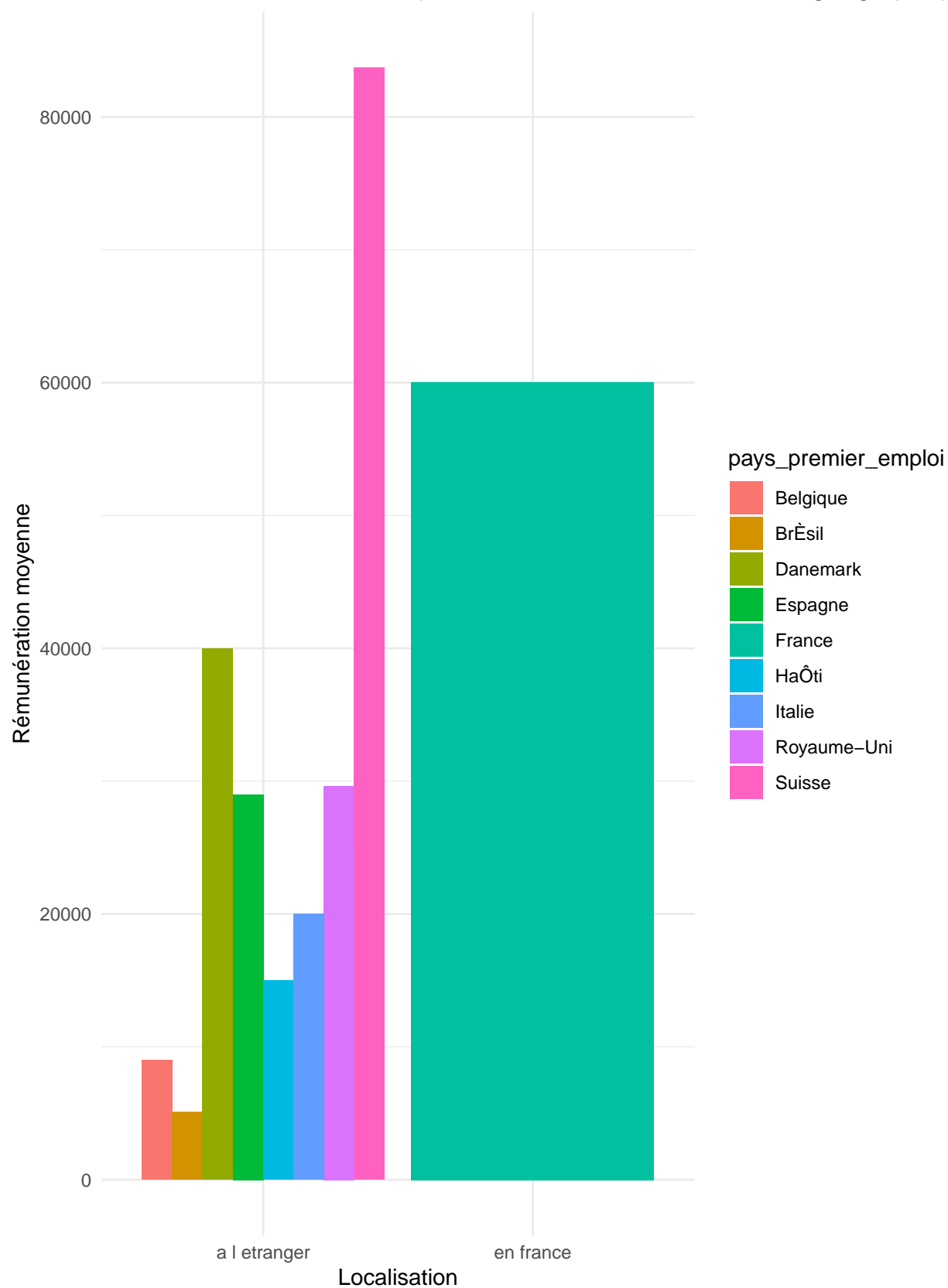


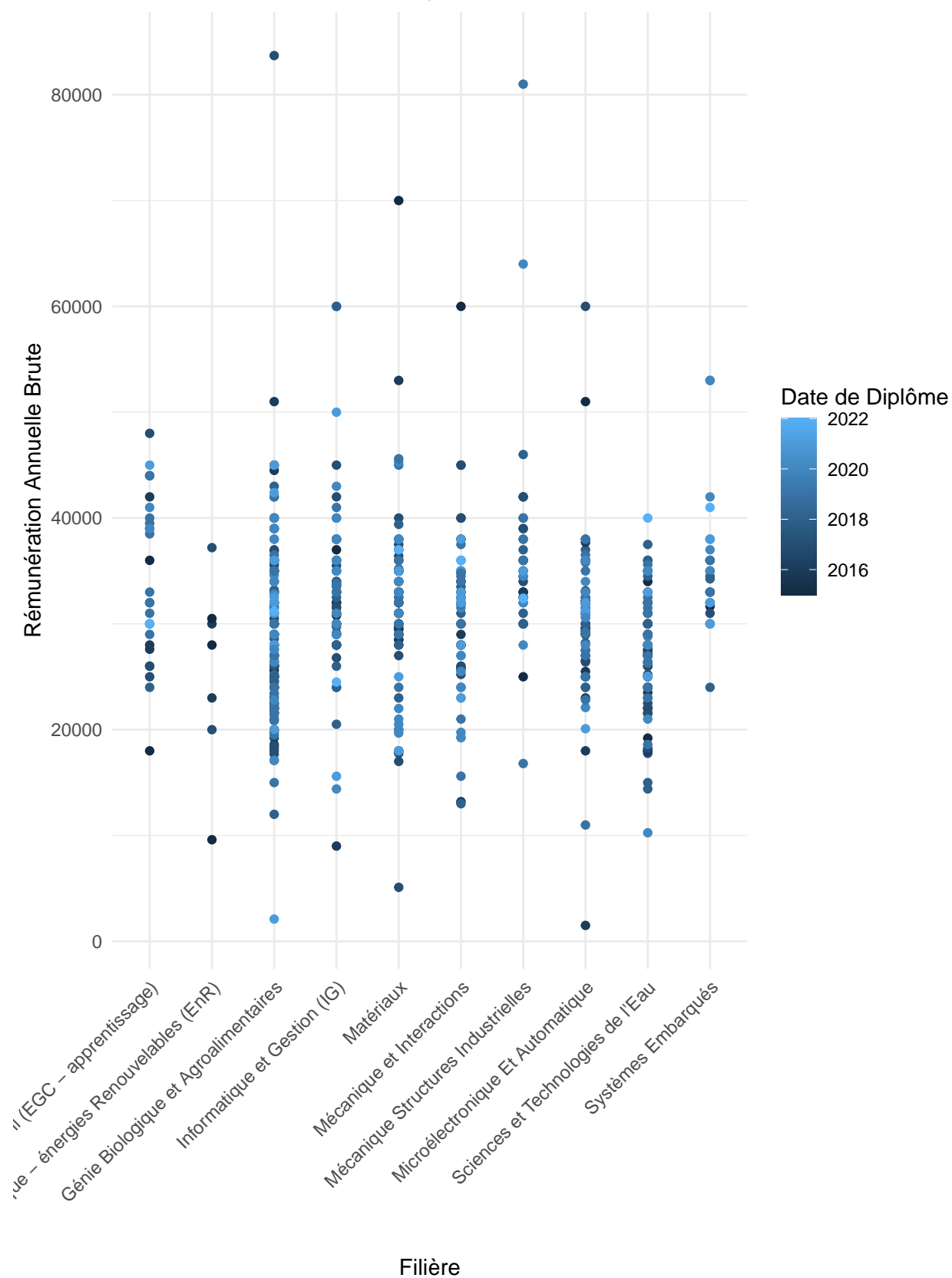
Diagramme 3 - Rémunération Premier Emploi en fonction de la Filière et de la Date de Diplôme

```
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



```
# Charger les bibliothèques nécessaires
library(car)
```

```
## Le chargement a nécessité le package : carData
```

```
##
## Attachement du package : 'car'
```

```
## L'objet suivant est masqué depuis 'package:dplyr':
##
##      recode
```

```
library(rstatix)
```

```
##
## Attachement du package : 'rstatix'
```

```
## L'objet suivant est masqué depuis 'package:stats':
##
##      filter
```

```
# Test de Tukey pour la rémunération du premier emploi avec la localisation et le pays
tukey_test_loc_pays <- TukeyHSD(aov(remuneration_annuelle_brute_avec_prime_premier_emploi ~ localisation_premier_emploi))
print("Test de Tukey pour la rémunération du premier emploi avec la localisation et le pays :")
```

```
## [1] "Test de Tukey pour la rémunération du premier emploi avec la localisation et le pays :"
```

```
print(tukey_test_loc_pays)
```

```
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = remuneration_annuelle_brute_avec_prime_premier_emploi ~ localisation_premier_emploi)
##
## $localisation_premier_emploi
##              diff              lwr              upr              p adj
## en france-a l etranger 467.0714 -3494.137 4428.28 0.8164551
##
## $pays_premier_emploi
##              diff              lwr              upr              p adj
## BrÈsil-Belgique    -3900.0000 -36182.374 28382.374 0.9999879
## Danemark-Belgique   29000.0000  1042.644 56957.356 0.0354794
## Espagne-Belgique    20000.0000 -12282.374 52282.374 0.5860689
## France-Belgique     20561.4286  -2318.929 43441.786 0.1167608
## Haïti-Belgique       6000.0000 -26282.374 38282.374 0.9996760
## Italie -Belgique    11000.0000 -21282.374 43282.374 0.9781832
## Royaume-Uni-Belgique 16612.0000  -8393.819 41617.819 0.4887534
## Suisse-Belgique     56850.0000 28892.644 84807.356 0.0000000
## Danemark-BrÈsil     32900.0000  4942.644 60857.356 0.0085881
## Espagne-BrÈsil      23900.0000  -8382.374 56182.374 0.3350226
```



```
## France-Br  sil      24461.4286    1581.071 47341.786 0.0261510
## Ha  ti-Br  sil      9900.0000   -22382.374 42182.374 0.9888783
## Italie -Br  sil     14900.0000   -17382.374 47182.374 0.8784826
## Royaume-Uni-Br  sil 20512.0000   -4493.819 45517.819 0.2054866
## Suisse-Br  sil      60750.0000    32792.644 88707.356 0.0000000
## Espagne-Danemark    -9000.0000   -36957.356 18957.356 0.9847693
## France-Danemark     -8438.5714   -24655.009  7777.866 0.7871087
## Ha  ti-Danemark     -23000.0000   -50957.356  4957.356 0.2022199
## Italie -Danemark     -18000.0000   -45957.356  9957.356 0.5330955
## Royaume-Uni-Danemark -12388.0000   -31486.510  6710.510 0.5226597
## Suisse-Danemark      27850.0000     5022.914 50677.086 0.0053123
## France-Espagne       561.4286   -22318.929 23441.786 1.0000000
## Ha  ti-Espagne      -14000.0000   -46282.374 18282.374 0.9118049
## Italie -Espagne      -9000.0000   -41282.374 23282.374 0.9941095
## Royaume-Uni-Espagne  -3388.0000   -28393.819 21617.819 0.9999705
## Suisse-Espagne       36850.0000     8892.644 64807.356 0.0016749
## Ha  ti-France       -14561.4286   -37441.786  8318.929 0.5492657
## Italie -France       -9561.4286   -32441.786 13318.929 0.9277468
## Royaume-Uni-France   -3949.4286   -14276.582  6377.725 0.9562079
## Suisse-France        36288.5714    20072.134 52505.009 0.0000000
## Italie -Ha  ti       5000.0000   -27282.374 37282.374 0.9999179
## Royaume-Uni-Ha  ti  10612.0000   -14393.819 35617.819 0.9214419
## Suisse-Ha  ti       50850.0000    22892.644 78807.356 0.0000014
## Royaume-Uni-Italie   5612.0000   -19393.819 30617.819 0.9987104
## Suisse-Italie        45850.0000    17892.644 73807.356 0.0000215
## Suisse-Royaume-Uni   40238.0000    21139.490 59336.510 0.0000000
```

```
# ANOVA   deux facteurs pour la r  mun  ration du premier emploi avec le sexe et le secteur d'activit  
anova_sex_secteur <- aov(remuneration_annuelle_brute_avec_prime_premier_emploi ~ sexe * secteur_premier_emploi, data = data1)
print("ANOVA   deux facteurs pour la r  mun  ration du premier emploi avec le sexe et le secteur d'activit  ")
```

```
## [1] "ANOVA   deux facteurs pour la r  mun  ration du premier emploi avec le sexe et le secteur d'activit  "
```

```
print(anova_sex_secteur)
```

```
## Call:
## aov(formula = remuneration_annuelle_brute_avec_prime_premier_emploi ~
##     sexe * secteur_premiere_entreprise, data = data1)
##
## Terms:
##                sexe secteur_premiere_entreprise
## Sum of Squares    730413380                3262344663
## Deg. of Freedom         2                  27
##      sexe:secteur_premiere_entreprise  Residuals
## Sum of Squares                1114535568 19237684948
## Deg. of Freedom                 20          296
##
## Residual standard error: 8061.773
## 34 out of 84 effects not estimable
## Estimated effects may be unbalanced
```

```
# Régression linéaire pour la rémunération du premier emploi avec la filière et la date d'obtention du
lm_model <- lm(remuneration_annuelle_brute_avec_prime_premier_emploi ~ filiere + date_diplome, data = data3)
print("Régression linéaire pour la rémunération du premier emploi avec la filière et la date d'obtention du diplôme")
```

```
## [1] "Régression linéaire pour la rémunération du premier emploi avec la filière et la date d'obtention du diplôme"
```

```
print(summary(lm_model))
```

```
##
## Call:
## lm(formula = remuneration_annuelle_brute_avec_prime_premier_emploi ~
##     filiere + date_diplome, data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28719  -3895       37    3693   54799
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -406696.4    336735.0  -1.208  0.22756
## filiereEnergétique - énergies Renouvelables (EnR)    -8232.5     3255.5  -2.529  0.01167 *
## filiereGénie Biologique et Agroalimentaires    -5052.8     1598.5  -3.161  0.00164 **
## filiereInformatique et Gestion (IG)    -1283.2     1717.3  -0.747  0.45520
## filiereMatériaux    -2428.0     1687.4  -1.439  0.15065
## filiereMécanique et Interactions    -3420.8     1751.6  -1.953  0.05124 .
## filiereMécanique Structures Industrielles     2278.2     1946.7   1.170  0.24230
## filiereMicroélectronique Et Automatique    -3516.1     1709.3  -2.057  0.04007 *
## filiereSciences et Technologies de l'Eau    -6781.2     1685.3  -4.024  6.38e-05 ***
## filiereSystèmes Embarqués       1192.1     2114.1   0.564  0.57304
## date_diplome           218.5       166.9   1.309  0.19092
##
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7606 on 679 degrees of freedom
## Multiple R-squared:  0.1004, Adjusted R-squared:  0.08716
## F-statistic: 7.579 on 10 and 679 DF, p-value: 1.616e-11
```