

TP2-vq

Sarah THEOULLE

2024-04-09

On commence par charger la librairie et signifier que l'on se trouve "here" pour la suite du code.

```
library(ggplot2)
library(ggforce)
library(forcats)
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(here)
```

```
## here() starts at /home/sarah/Documents/D03/S6/DataViz/do3-dataviz
```

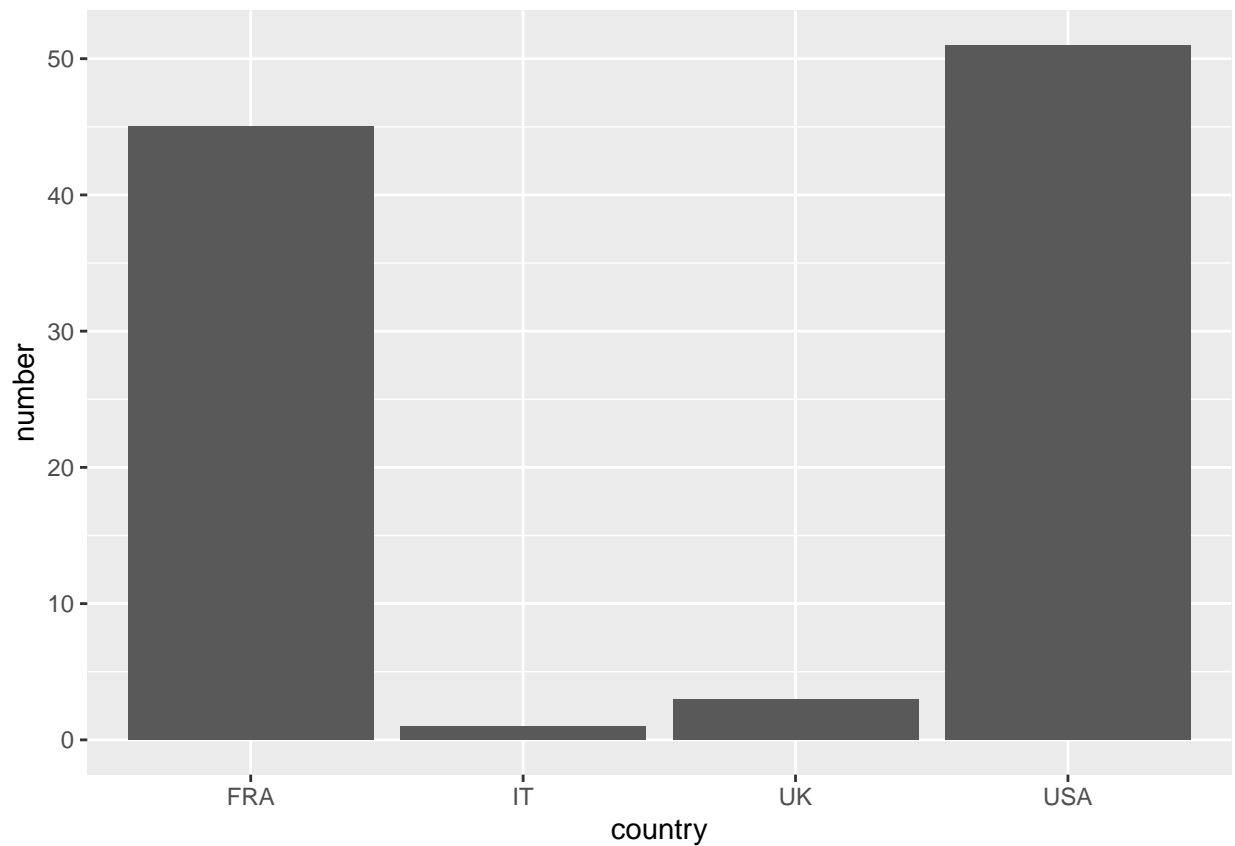
```
library(readxl)
library(tidyr)
```

```
boxoffice <- data.frame(
  country = c("USA", "FRA", "UK", "IT"), # pays de production
  number = c(51, 45, 03, 01)             # nombre de films
)
boxoffice
```

```
##   country number
## 1     USA     51
## 2     FRA     45
## 3      UK      3
## 4      IT      1
```

Diagramme en barres avec `geom_col()`

```
ggplot(boxoffice, aes(x=country, y=number)) + geom_col()
```

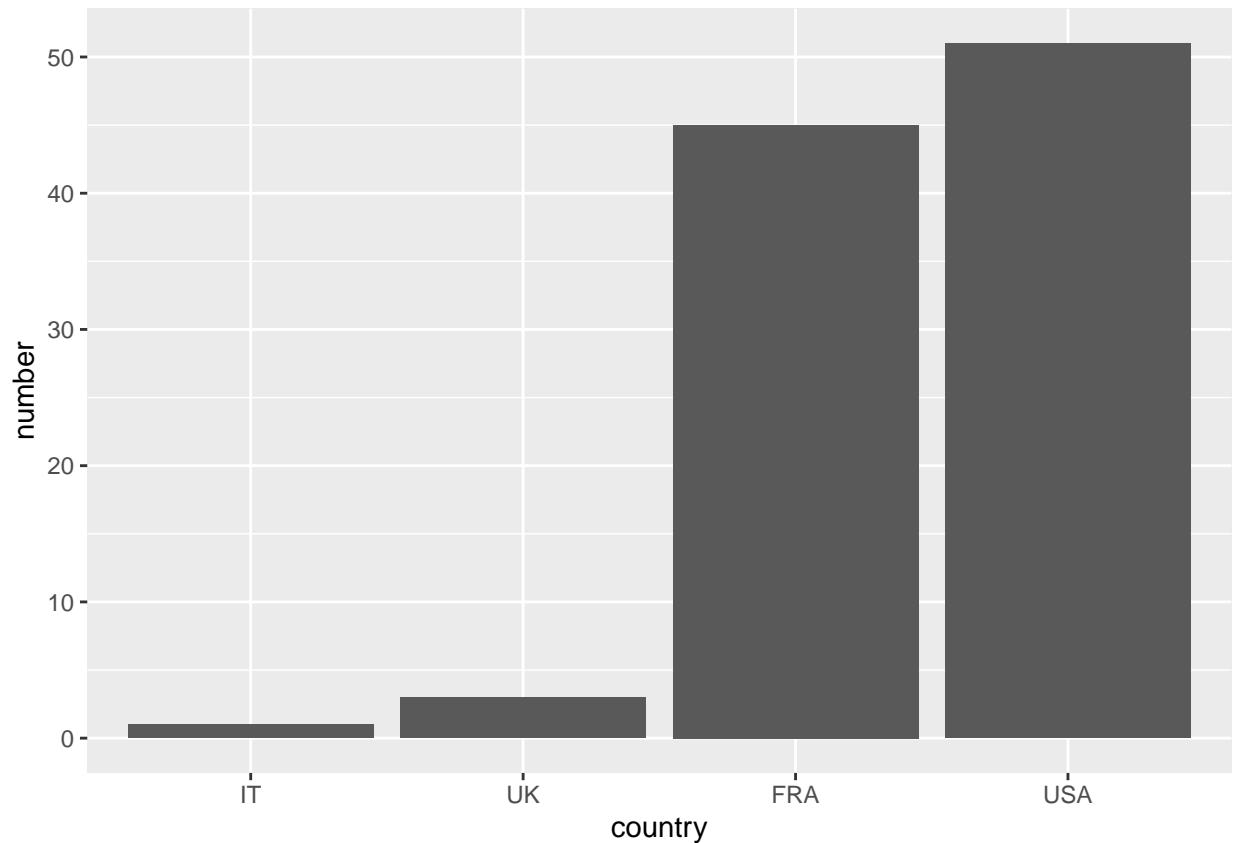


Si on inverse les rôle de x et de y on tourne le graphique de 90 degrés

Ordonnancement des facteurs

On remarque dans le graphique précédent que les facteurs sont ordonnés par ordre alphabétique. Cependant cela ne facilite pas la lecture. On va les représenter autrement à l'aide de la fonction `fct_reorder`.

```
boxoffice <- boxoffice %>%  
  mutate(country = fct_reorder(country, number))  
  
ggplot(boxoffice, aes(x = country, y = number)) +  
  geom_col()
```

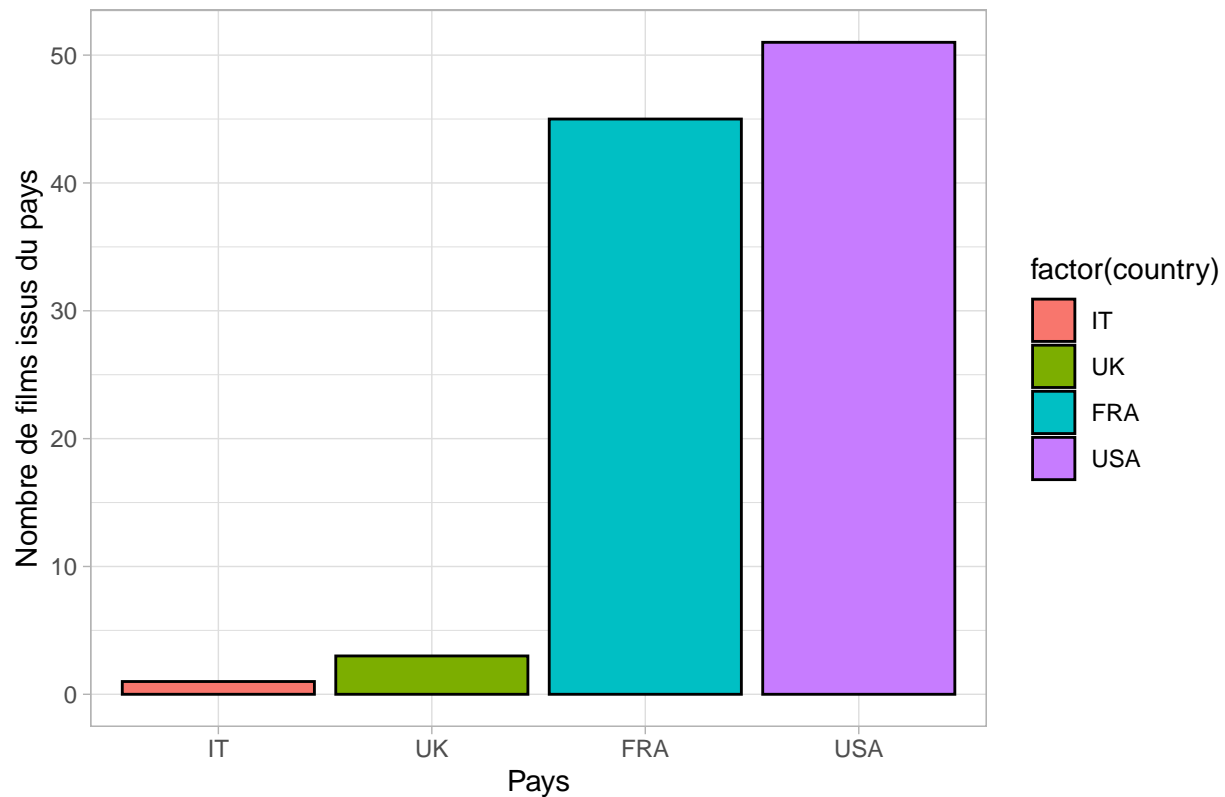


Nettoyage

On peut changer les couleurs du diagramme, ainsi que le thème du graphe et donner un titre aux axes pour donner du sens au graphique.

```
boxoffice <- boxoffice %>%  
  mutate(country = fct_reorder(country, number))  
  
ggplot(boxoffice, aes(x = country, y = number, fill = factor(country))) +  
  geom_col(color = "black") +  
  theme_light() +  
  labs(x = "Pays", y = "Nombre de films issus du pays") +  
  ggtitle("Pays de production des 100 films avec le plus d'entrées au cinéma français")
```

Pays de production des 100 films avec le plus d'entrées au cinéma français



Pingouins

```
library(palmerpenguins)
penguins_nomissing <- na.omit(penguins)
head(penguins_nomissing)
```

```
## # A tibble: 6 x 8
##   species island  bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
##   <fct>   <fct>         <dbl>         <dbl>         <int>         <int>
## 1 Adelie  Torgersen         39.1          18.7          181          3750
## 2 Adelie  Torgersen         39.5          17.4          186          3800
## 3 Adelie  Torgersen         40.3           18          195          3250
## 4 Adelie  Torgersen         36.7          19.3          193          3450
## 5 Adelie  Torgersen         39.3          20.6          190          3650
## 6 Adelie  Torgersen         38.9          17.8          181          3625
## # i 2 more variables: sex <fct>, year <int>
```

Jeu de données et objectif

À quelle question tente-t-on de répondre ?

- On veut représenter le nombre d'individus de chaque espèce.

Quels sont les individus ?

- Un individu est un pingouin

Quelle est la population ? l'échantillon ?

- La population est celle de tous les pingouins de l'Antarctique et l'échantillon un jeu de données avec les pingouins que l'on a pu attraper et examiner

Quelle est la variable ? quelles sont ses modalités ?

- La variable est un pingouin avec son espèce, l'île sur laquelle il a été observé, différentes mesures de sa taille ainsi que sa masse;

Comptage

```
penguins_nomissing$species
```

```
## [1] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [8] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [15] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [22] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [29] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [36] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [43] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [50] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [57] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [64] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [71] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [78] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [85] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [92] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [99] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [106] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [113] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [120] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [127] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [134] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [141] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [148] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [155] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [162] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [169] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [176] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [183] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [190] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [197] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [204] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [211] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [218] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
```



```
sum(penguins_nomissing$species == "Gentoo")
```

```
## [1] 119
```

```
sum(penguins_nomissing$species == "Chinstrap")
```

```
## [1] 68
```

On vient d'afficher le nom des espèces de tous les pingouins recensés, une table reprenant si ce sont des pingouins Adelie ainsi qu'un comptage du nombre d'individus de chaque espèce

On peut retrouver les mêmes résultats avec la fonction table

```
penguins_nomissing$species
```

```
## [1] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [8] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [15] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [22] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [29] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [36] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [43] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [50] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [57] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [64] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [71] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [78] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [85] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [92] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [99] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [106] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [113] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [120] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [127] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [134] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [141] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [148] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [155] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [162] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [169] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [176] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [183] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [190] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [197] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [204] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [211] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [218] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [225] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [232] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [239] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [246] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
## [253] Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo Gentoo
```

```
## [260] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Chinstrap
## [267] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
## [274] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
## [281] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
## [288] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
## [295] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
## [302] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
## [309] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
## [316] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
## [323] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
## [330] Chinstrap Chinstrap Chinstrap Chinstrap
## Levels: Adelie Chinstrap Gentoo
```

```
table_adelie <- table(penguins_nomissing$species == "Adelie")
table_adelie
```

```
##
## FALSE TRUE
##    187   146
```

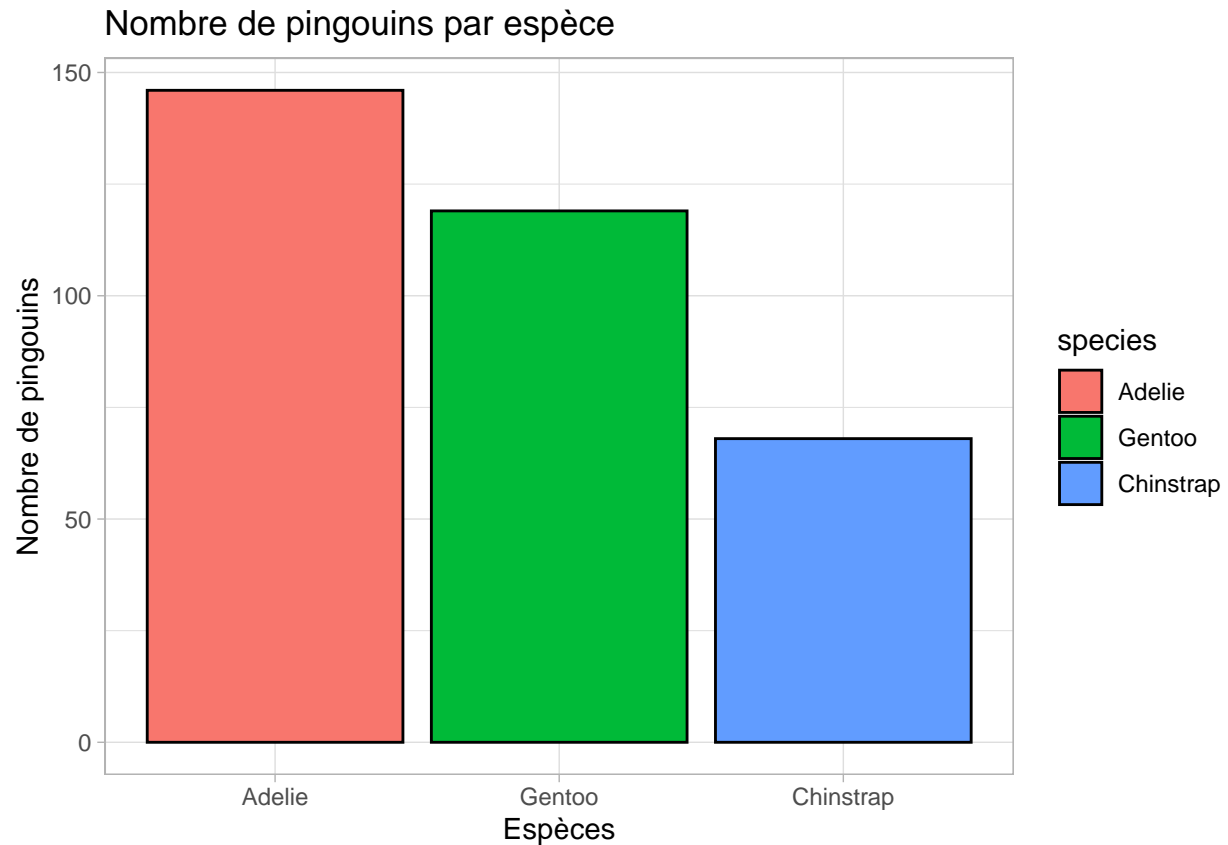
```
table_especes <- table(penguins_nomissing$species)
table_especes
```

```
##
##    Adelie Chinstrap    Gentoo
##      146         68      119
```

Graphique artisanal et diagramme en barres avec geom_bar()

```
penguins_nomissing <- penguins_nomissing %>%
  mutate(species = fct_reorder(species, species, function(x) -length(x)))

ggplot(penguins_nomissing, aes(x = species, fill = species)) +
  geom_bar(color = "black") +
  theme_light() +
  labs(x = "Espèces", y = "Nombre de pingouins") +
  ggtitle("Nombre de pingouins par espèce")
```

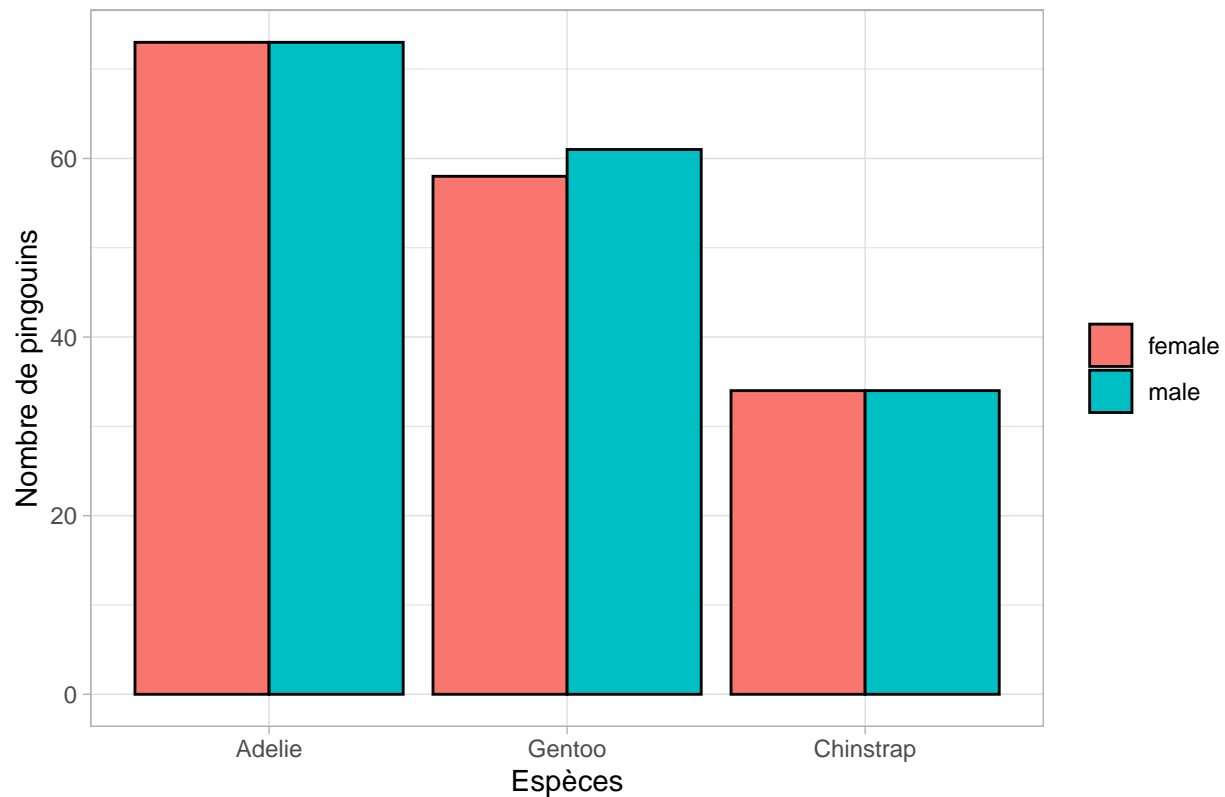



Ajout d'information

```
# Réorganisation des espèces par ordre décroissant de nombre de pingouins
penguins_nomissing <- penguins_nomissing %>%
  mutate(species = fct_reorder(species, species, function(x) -length(x)))

# Tracé du graphique
ggplot(penguins_nomissing, aes(x = species, fill = sex)) +
  geom_bar(color = "black", position = "dodge") +
  theme_light() +
  labs(x = "Espèces", y = "Nombre de pingouins",
       title = "Répartition des pingouins par espèce et par sexe") +
  theme(legend.title = element_blank()) # Pour supprimer le titre de la légende
```

Répartition des pingouins par espèce et par sexe



Ce graphique présente la répartition des pingouins par espèce et par sexe, avec le nombre de pingouins par espèce divisé par sexe.

Parlements

Données

On commence par charger les données du fichier

```
women <- read_excel("data/women_in_parliament-historical_database-1945_to_2018.xlsx",
  col_types = c("text", "text", "text",
    "numeric", "text", "text", "numeric",
    "numeric", "numeric", "text"))
head(women)
```

```
## # A tibble: 6 x 10
##   Country      Region 'Election / Renewal' Year Month 'Chamber Type'
##   <chr>        <chr>   <chr>                <dbl> <chr> <chr>
## 1 Afghanistan ASIA    Yes                    1965 July Lower
## 2 Afghanistan ASIA    Yes                    1965 July Upper
## 3 Afghanistan ASIA    Yes                    1969 <NA> Lower
## 4 Afghanistan ASIA    Yes                    1969 <NA> Upper
## 5 Afghanistan ASIA    Yes                    1988 <NA> Lower
## 6 Afghanistan ASIA    Yes                    1988 <NA> Upper
```

```
## # i 4 more variables: 'Chamber Total Seats' <dbl>, 'Total women' <dbl>,
## #   '% Of Women in Chamber' <dbl>, NOTES <chr>
```

Les données représentent des informations sur la représentation des femmes au parlement dans différents pays et régions du monde, avec des détails sur les élections ou le renouvellement, l'année et le mois de l'événement, le type de chambre (par exemple, la chambre basse), le nombre total de sièges dans la chambre, le nombre total de femmes siégeant, le pourcentage de femmes dans la chambre, et des notes supplémentaires.

- Taille : Il y a 3816 entrées dans le tableau
- Variables : Le jeu de données comprend les variables suivantes :
 1. Country (Pays)
 2. Region (Région)
 3. Election / Renewal (Élection / Renouvellement)
 4. Year (Année)
 5. Month (Mois)
 6. Chamber Type (Type de chambre)
 7. Chamber Total Seats (Nombre total de sièges dans la chambre)
 8. Total women (Nombre total de femmes)
 9. % Of Women in Chamber (Pourcentage de femmes dans la chambre)
 10. NOTES (Notes)

```
frwomen <- subset(women, Country == "France")
head(frwomen)
```

```
## # A tibble: 6 x 10
##   Country Region 'Election / Renewal' Year Month 'Chamber Type'
##   <chr>   <chr>   <chr>           <dbl> <chr> <chr>
## 1 France EUR     Yes             1945 <NA> Lower
## 2 France EUR     Yes             1946 <NA> Lower
## 3 France EUR     Yes             1946 <NA> Lower
## 4 France EUR     Yes             1946 <NA> Upper
## 5 France EUR     Yes             1948 <NA> Upper
## 6 France EUR     Yes             1951 <NA> Lower
## # i 4 more variables: 'Chamber Total Seats' <dbl>, 'Total women' <dbl>,
## #   '% Of Women in Chamber' <dbl>, NOTES <chr>
```

Assemblée constituante française

La Quatrième République française a été proclamée le 27 octobre 1946. Elle a succédé à la Troisième République après la Seconde Guerre mondiale et a duré jusqu'au 4 octobre 1958, date à laquelle la Cinquième République a été instaurée suite au référendum sur la Constitution de la Cinquième République française. La Quatrième République était caractérisée par un régime parlementaire faible et une instabilité politique, avec une série de gouvernements éphémères et une guerre d'indépendance en cours en Algérie.

```
assembly <- subset(frwomen, Year == 1946 & NOTES == "Constitutional Assembly")
assembly
```

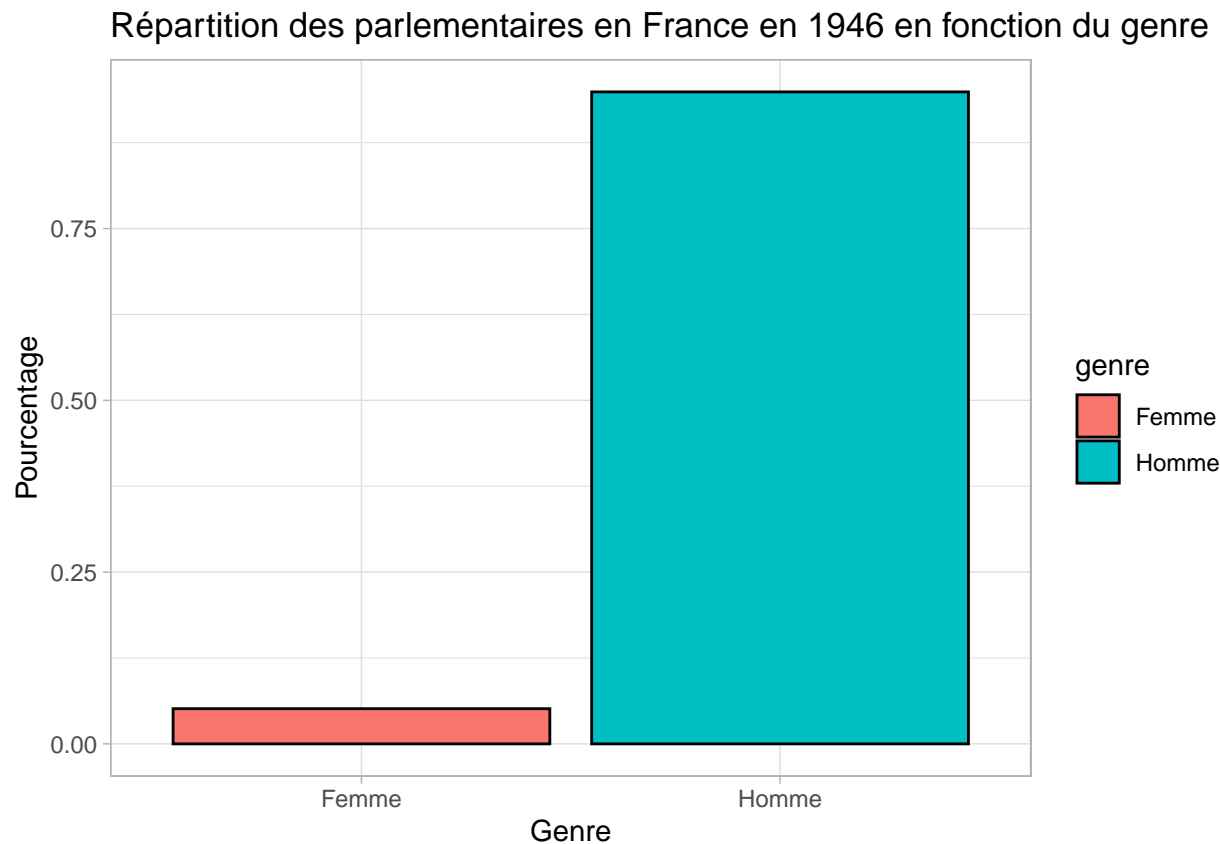
```
## # A tibble: 1 x 10
##   Country Region 'Election / Renewal' Year Month 'Chamber Type'
##   <chr>   <chr>   <chr>           <dbl> <chr> <chr>
```

```
## 1 France EUR Yes 1946 <NA> Lower
## # i 4 more variables: 'Chamber Total Seats' <dbl>, 'Total women' <dbl>,
## # '% Of Women in Chamber' <dbl>, NOTES <chr>
```

```
data_constituante <- data.frame(genre = c("Femme", "Homme"),
                                pourcent = c((assembly$"% Of Women in Chamber"), 1-assembly$"% Of Women in Chamber"))
data_constituante
```

```
##   genre   pourcent
## 1 Femme 0.05119454
## 2 Homme 0.94880546
```

```
ggplot(data_constituante, aes(x = genre, y = pourcent, fill = genre)) +
  geom_bar(stat = "identity", color = "black") +
  theme_light() +
  labs(x = "Genre", y = "Pourcentage",
       title = "Répartition des parlementaires en France en 1946 en fonction du genre")
```

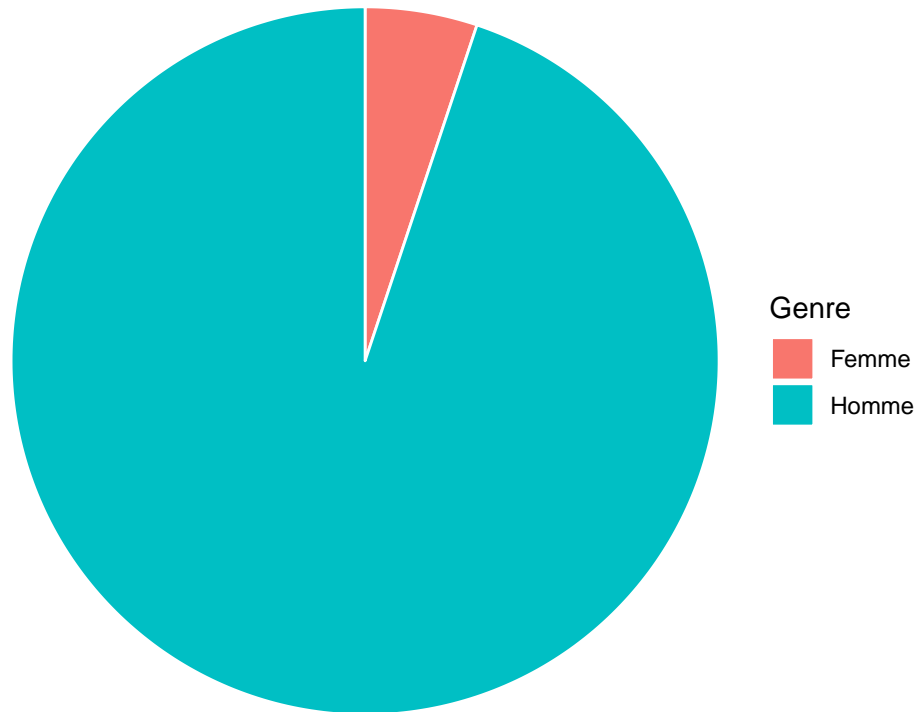


J'utilise des barres côtes à côtes afin de mieux voir la différence entre le pourcentage de femmes et d'hommes dans le l'assemblée constituante française de 1946.

```
ggplot(data_constituante, aes(x0 = 0, y0 = 0, r0 = 0, r = 1, amount=pourcent, fill=genre)) +
  coord_fixed() +
  ggtitle("Proportion de femmes/hommes dans l'assemblée constituante \nde la V république française") +
  geom_arc_bar(stat = "pie", color = "white") +
```

```
theme_void() +
labs(fill = "Genre")
```

Proportion de femmes/hommes dans l'assemblée constituante de la V république française



Evolution de la proportion de femmes dans l'assemblée d'un pays

Je choisis la ligne 2334 Netherlands EUR Yes 1987 Upper 75 20 26,7%

```
line <- subset(women, Country == "United States of America" & `Chamber Type` == 'Upper' )
line
```

```
## # A tibble: 37 x 10
##   Country          Region 'Election / Renewal' Year Month 'Chamber Type'
##   <chr>            <chr> <chr>          <dbl> <chr> <chr>
## 1 United States of Amer~ AME   Yes          1946 Nove~ Upper
## 2 United States of Amer~ AME   Yes          1948 Nove~ Upper
## 3 United States of Amer~ AME   Yes          1950 Nove~ Upper
## 4 United States of Amer~ AME   Yes          1952 Nove~ Upper
## 5 United States of Amer~ AME   Yes          1954 Nove~ Upper
## 6 United States of Amer~ AME   Yes          1956 Nove~ Upper
## 7 United States of Amer~ AME   Yes          1958 Nove~ Upper
## 8 United States of Amer~ AME   Yes          1960 Nove~ Upper
## 9 United States of Amer~ AME   Yes          1962 Nove~ Upper
## 10 United States of Amer~ AME   Yes          1964 Nove~ Upper
## # i 27 more rows
```

```
## # i 4 more variables: 'Chamber Total Seats' <dbl>, 'Total women' <dbl>,
## #   '% Of Women in Chamber' <dbl>, NOTES <chr>
```

J'ai choisi d'utiliser les données des Etats-Unis avec la chambre haute et je vais le transformer pour ggplot. J'ai choisi ce pays car il est l'un des seuls à ne pas avoir de données tronquées avec plusieurs élections dans une même année pour quelque raison que ce soit.

```
colonnes <- c("Year", "% Of Women in Chamber")
line_ordonated <- line[colonnes]
line_ordonated$`% Of Women in Chamber` <- line_ordonated$`% Of Women in Chamber` * 100
line_ordonated$`% Of Men in Chamber` <- 100 - line_ordonated$`% Of Women in Chamber`
line_ordonated
```

```
## # A tibble: 37 x 3
##   Year '% Of Women in Chamber' '% Of Men in Chamber'
##   <dbl>           <dbl>           <dbl>
## 1 1946             0             100
## 2 1948             1             99
## 3 1950             1             99
## 4 1952             1             99
## 5 1954             3             97
## 6 1956             1             99
## 7 1958             1             99
## 8 1960             2             98
## 9 1962             2             98
## 10 1964            2             98
## # i 27 more rows
```

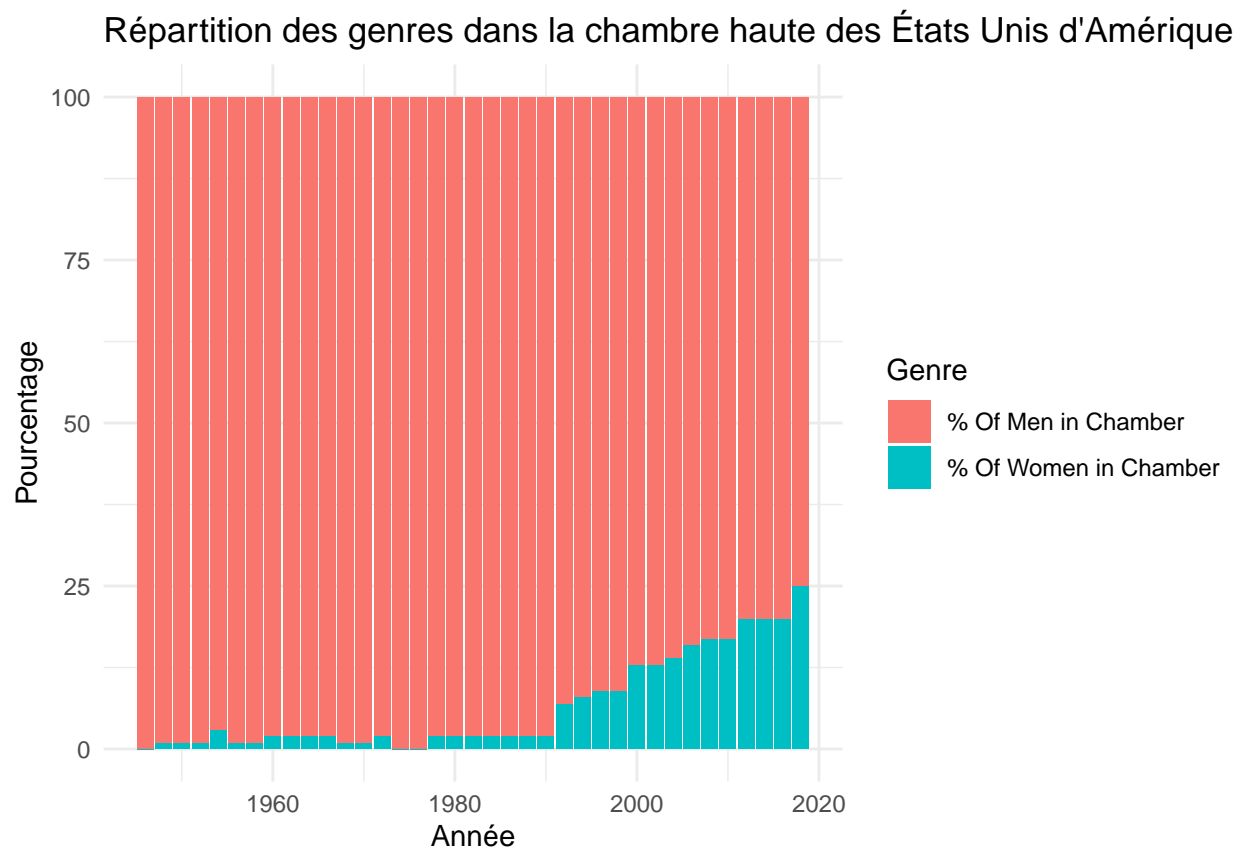
Maintenant que l'on a les années avec les répartitions de genre en pourcentage on peut utiliser un pivot pour finir de formater nos données

```
line_pivot <- pivot_longer(line_ordonated, cols = c("% Of Women in Chamber", "% Of Men in Chamber"))
line_pivot
```

```
## # A tibble: 74 x 3
##   Year name          value
##   <dbl> <chr>         <dbl>
## 1 1946 % Of Women in Chamber    0
## 2 1946 % Of Men in Chamber   100
## 3 1948 % Of Women in Chamber    1
## 4 1948 % Of Men in Chamber   99
## 5 1950 % Of Women in Chamber    1
## 6 1950 % Of Men in Chamber   99
## 7 1952 % Of Women in Chamber    1
## 8 1952 % Of Men in Chamber   99
## 9 1954 % Of Women in Chamber    3
## 10 1954 % Of Men in Chamber   97
## # i 64 more rows
```

Avec les données ainsi obtenues on peut tracer le diagramme associé avec des barres empilées.

```
# Tracer le graphique à barres empilées
ggplot(line_pivot, aes(x = Year, y = value, fill = name)) +
  geom_bar(stat = "identity") +
  labs(title = "Répartition des genres dans la chambre haute des États Unis d'Amérique",
        x = "Année",
        y = "Pourcentage",
        fill = "Genre") +
  theme_minimal()
```



Finalement on trace les diagrammes camembert correspondant. On a un diagramme par année avec des élections, et la répartition des genres dans la chambre haute.

```
# Tracer le graphique en secteurs
ggplot(line_pivot, aes(x = "", y = value, fill = name)) +
  geom_bar(stat = "identity") +
  coord_polar("y", start = 0) +
  facet_wrap(~ Year) +
  labs(title = "Répartition des genres dans la chambre haute des États-Unis d'Amérique",
        x = NULL,
        y = NULL,
        fill = "Genre") +
  theme_void()
```

Répartition des genres dans la chambre haute des États-Unis d'Amérique

