

TP_2

Thomas Mauran

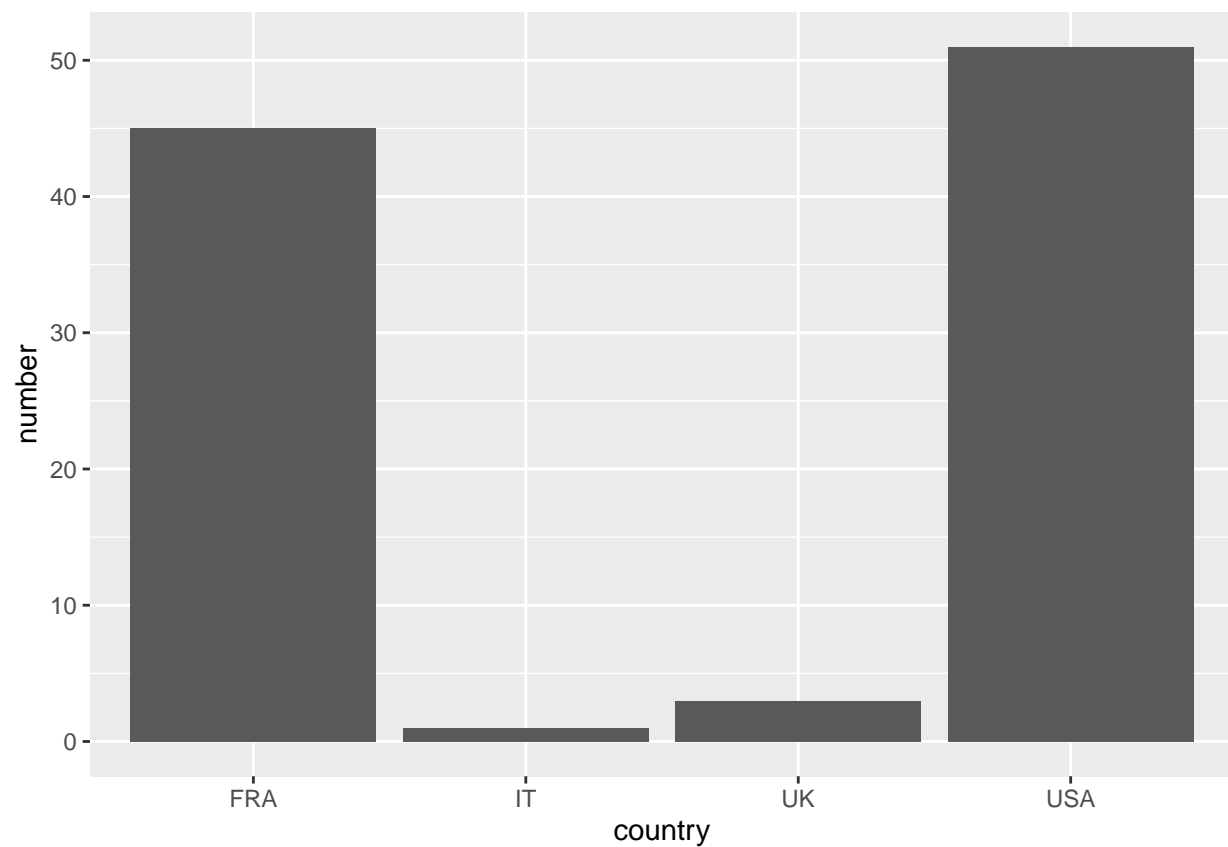
2023-03-30

TP 2

```
boxoffice <- data.frame (  
  country = c("USA", "FRA", "UK", "IT"),  
  number = c(51, 45, 03, 01)  
)
```

Diagramme en barre

```
library(ggplot2)  
ggplot(boxoffice,  
  aes(x = country,  
      y = number)) +  
  geom_col()
```



Trier les valeurs par ordre croissant + clean graph

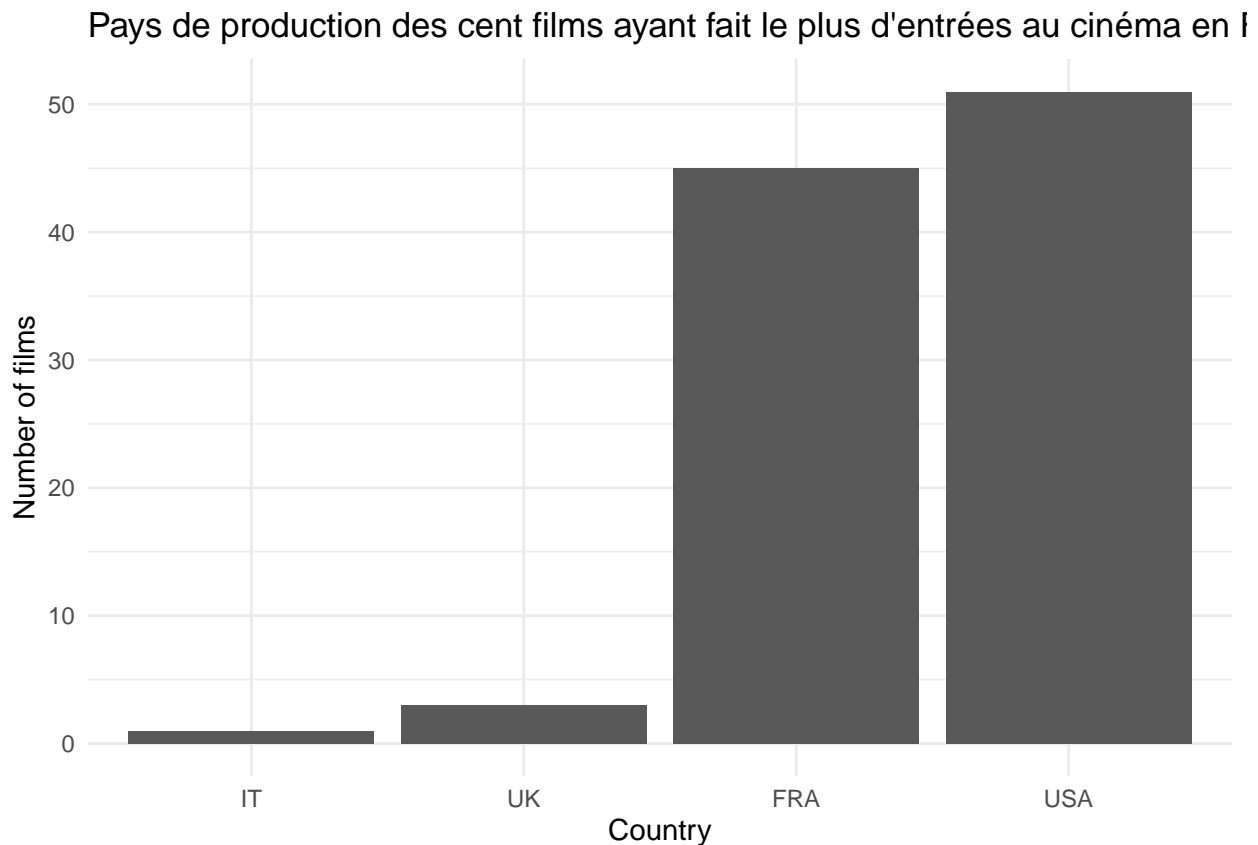
```
library(forcats)
```

```
# On transforme en facteur  
factor(boxoffice$country)
```

```
## [1] USA FRA UK IT  
## Levels: FRA IT UK USA
```

```
orderedList <- fct_reorder(boxoffice$country, boxoffice$number) # -boxoffice$number => décroissant
```

```
ggplot(boxoffice,  
      aes(x = orderedList,  
          y = number))+  
geom_col() +  
xlab("Country") +  
ylab("Number of films")+  
theme_minimal() +  
ggtitle("Pays de production des cent films ayant fait le plus d'entrées au cinéma en France depuis 1990")
```



Pingouins

```
library(palmerpenguins)
penguins_nomissing <- na.omit(penguins) # We want to have data without missing lines
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>
```

```
nrow(penguins_nomissing) # connaitre le nombre de ligne dans un tableau
```

```
## [1] 333
```

Experience statistique

- quelle question tente-t-on de répondre ?

On cherche à déterminer le nombre d'individus de chaque espèce de pingouins présente dans l'archipel

- quels sont les *individus* ?

Les pingouins

- quelle est la *population* ? l'*échantillon* ?

La population: L'ensemble des pingouins de l'archipel

L'échantillon: 333 pingouins de l'archipel

- quelle est la *variable* ? quelles sont ses *modalités* ?

Ici on considérera la variable comme étant l'espèce de pingouin.

Modalité: Chinstrap, gentoo adélie

Comptage

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

```
## [1] 146
```

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

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

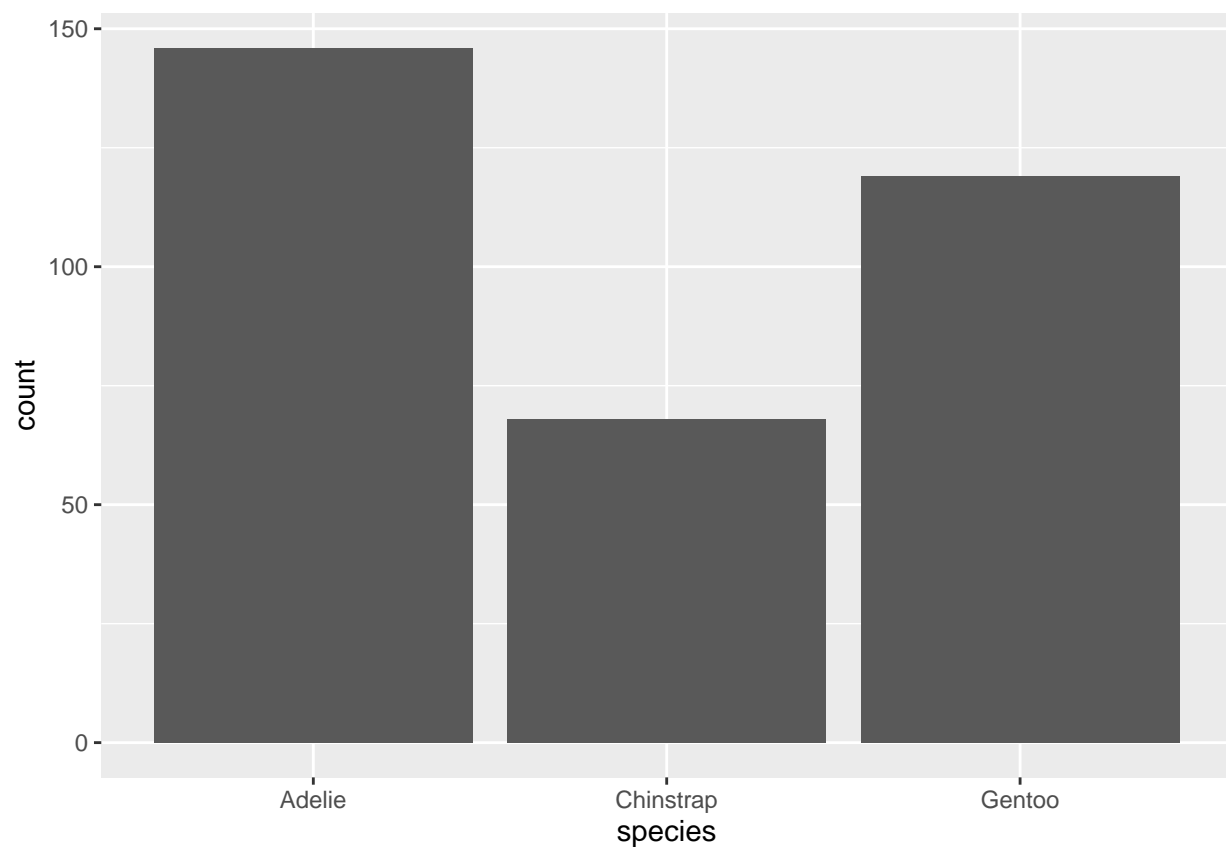
```
sum(penguins_nomissing$species == "Chinstrap")
```

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

On a 146 Adelie, 119 Gentoo et 68 Chinstrap

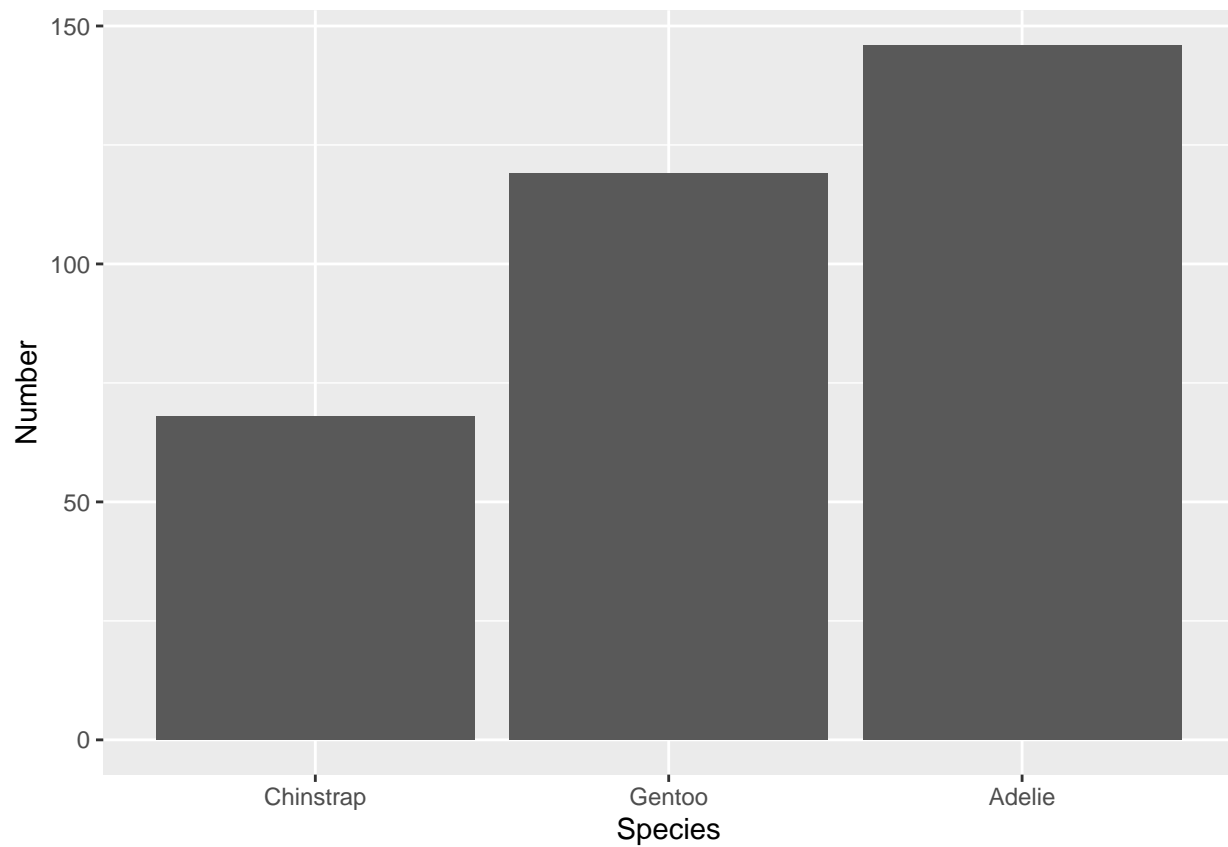
Diagramme en bar

```
ggplot(penguins_nomissing,  
       aes(x = species)) +  
  
geom_bar()
```



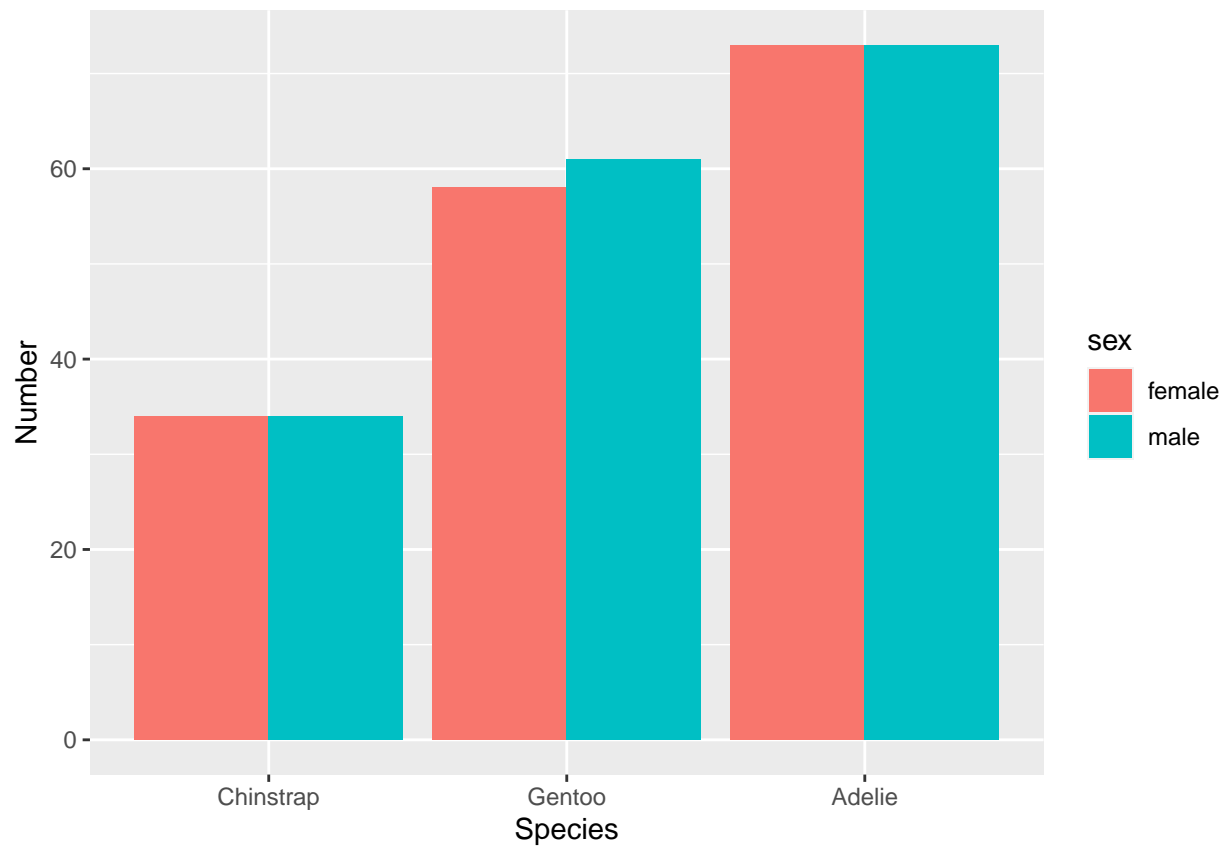
Nettoyage

```
ggplot(penguins_nomissing,  
       aes(x = fct_relevel(species, "Chinstrap", "Gentoo", "Adelie"))) +  
       xlab("Species") +  
       ylab("Number") +  
       geom_bar()
```



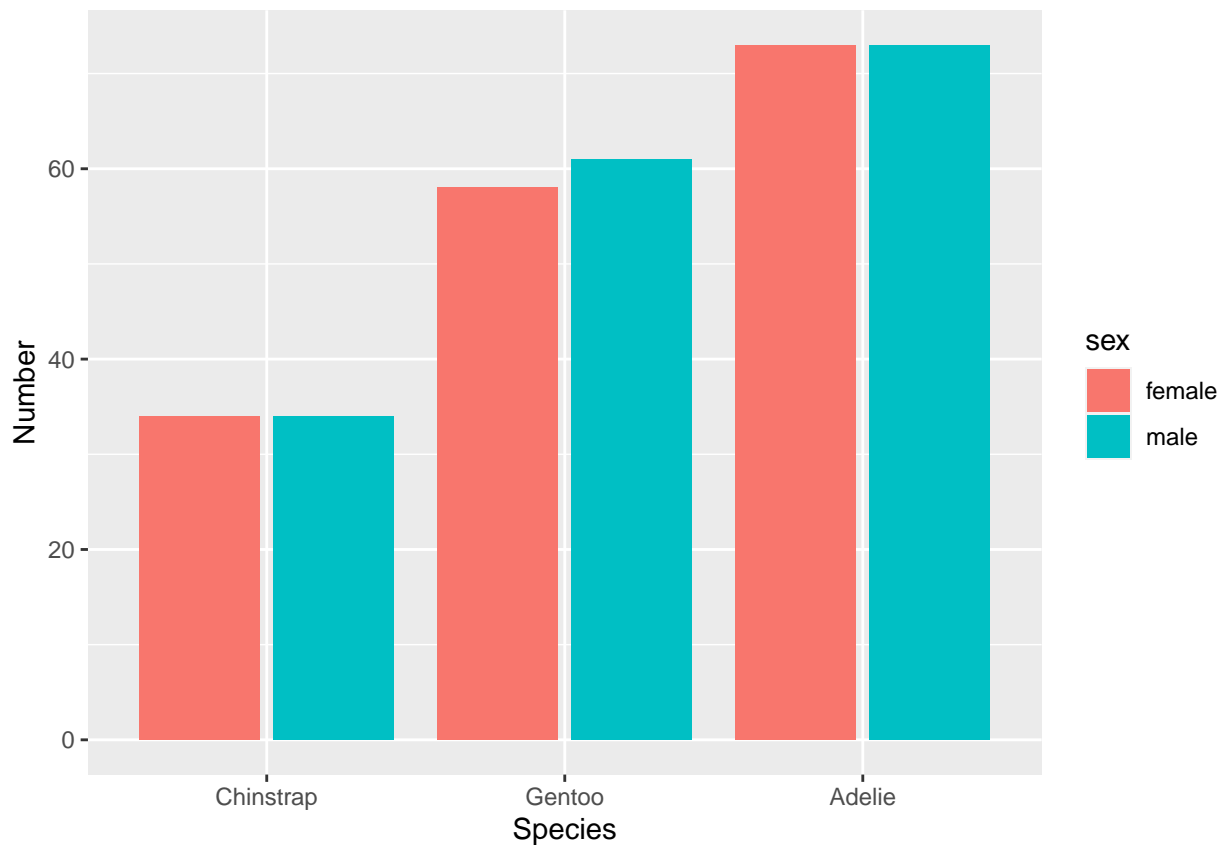
Ajout d'information

```
ggplot(penguins_nomissing,  
       aes(x = fct_relevel(species, "Chinstrap", "Gentoo", "Adelie"),  
           fill = sex)) +  
  xlab("Species") +  
  ylab("Number")+  
  geom_bar(position = "dodge")
```



Variation

```
ggplot(penguins_nomissing,  
  aes(x = fct_relevel(species, "Chinstrap", "Gentoo", "Adelie"),  
    fill = sex)) +  
  xlab("Species") +  
  ylab("Number")+  
  geom_bar(position = "dodge2")
```



Parlements

```
library(readxl)
library(here)
```

```
## here() starts at /home/thomas/Desktop/D0/R/do3-dataviz/RenduTP
```

```
data <- read_xlsx(here("data", "women_in_parliament-historical_database-1945_to_2018.xlsx"),
  col_types = c("text", "text", "text",
    "numeric", "text", "text", "numeric",
    "numeric", "numeric", "text"))
data
```

```
## # A tibble: 3,815 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
## 7 Afghanistan ASIA    Yes                  2005 <NA> Lower
## 8 Afghanistan ASIA    Yes                  2005 <NA> Upper
```

```
## 9 Afghanistan ASIA Yes 2010 <NA> Lower
## 10 Afghanistan ASIA Yes 2010 <NA> Upper
## # i 3,805 more rows
## # i 4 more variables: 'Chamber Total Seats' <dbl>, 'Total women' <dbl>,
## # '% Of Women in Chamber' <dbl>, NOTES <chr>
```

Les données pour chaque ligne:

Le pays, la région dans le pays, si c'est une Election ou une réélection, l'année, le mois, le type de chambre, le nombre de sièges dans la chambre, le nombre de femmes, le % de femme t des notes qui peuvent donner des infos sur le type d'assemblée

Assemblée constituante française

4eme république: 27 octobre 1946

```
subsetData <- subset(data, Year==1946 & Country=="France" & NOTES=="Constitutional Assembly")
subsetData
```

```
## # 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>
```

```
percentOfWomen <- subsetData[c("% Of Women in Chamber")] # le pourcentage de femmes
men <- 1 - percentOfWomen
men # pourcentage d'homme
```

```
## % Of Women in Chamber
## 1 0.9488055
```

Diagramme en barre

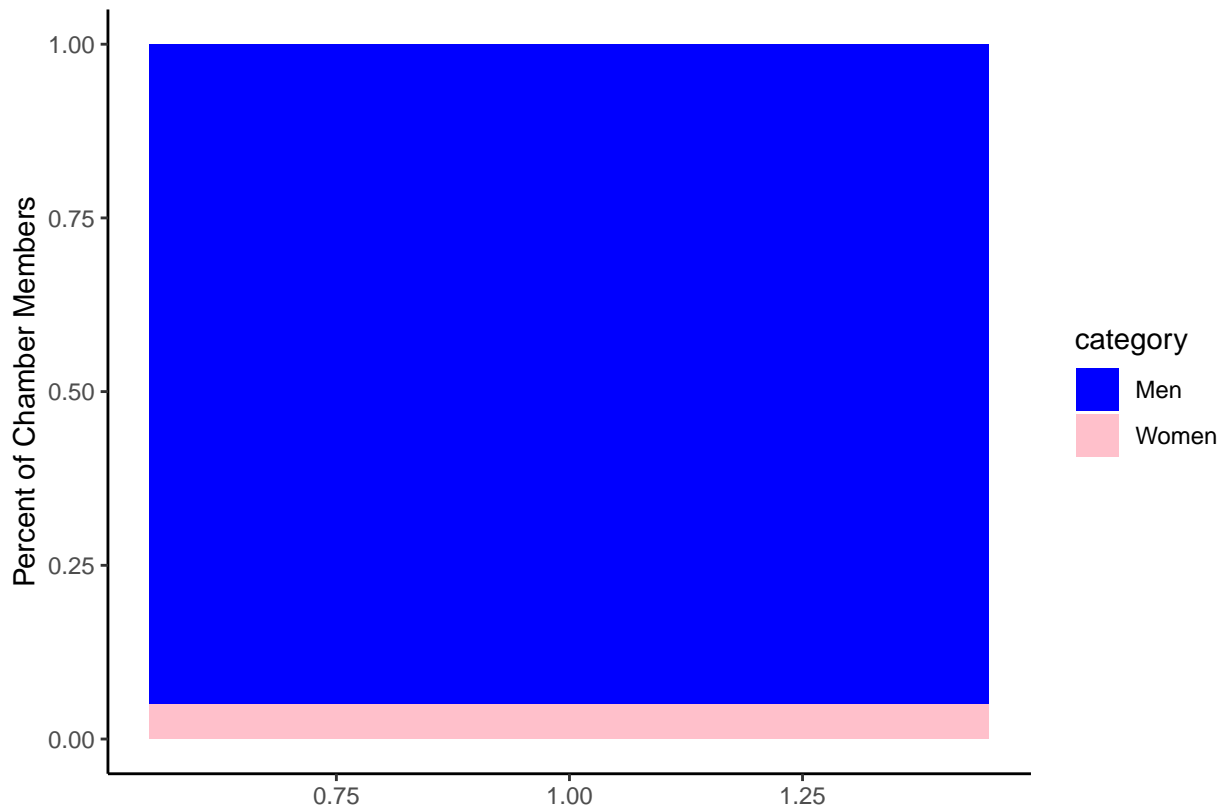
```
library(ggplot2)
subsetData$homme <- 1 - subsetData$"% Of Women in Chamber"

smallTable <- data.frame(category= c("Women", "Men"),
                          percent = c(subsetData$"% Of Women in Chamber", subsetData$homme))
smallTable

##   category    percent
## 1   Women 0.05119454
## 2    Men 0.94880546
```



```
ggplot(smallTable,
      aes(x = 1,
          y = percent,
          fill=category), position = "fill") +
  geom_col(position = "stack") +
  scale_fill_manual(values = c("blue", "pink")) +
  labs(x = "", y = "Percent of Chamber Members") +
  theme_classic()
```



Ici je trouve cela plus pertinent de représenter les données dans un graph stacké pour bien mettre en exergue la proportion d'homme bien plus importante

Graph en secteur

```
library(ggplot2)
library(ggforce)

subsetData$homme <- 1 - subsetData$`% Of Women in Chamber`

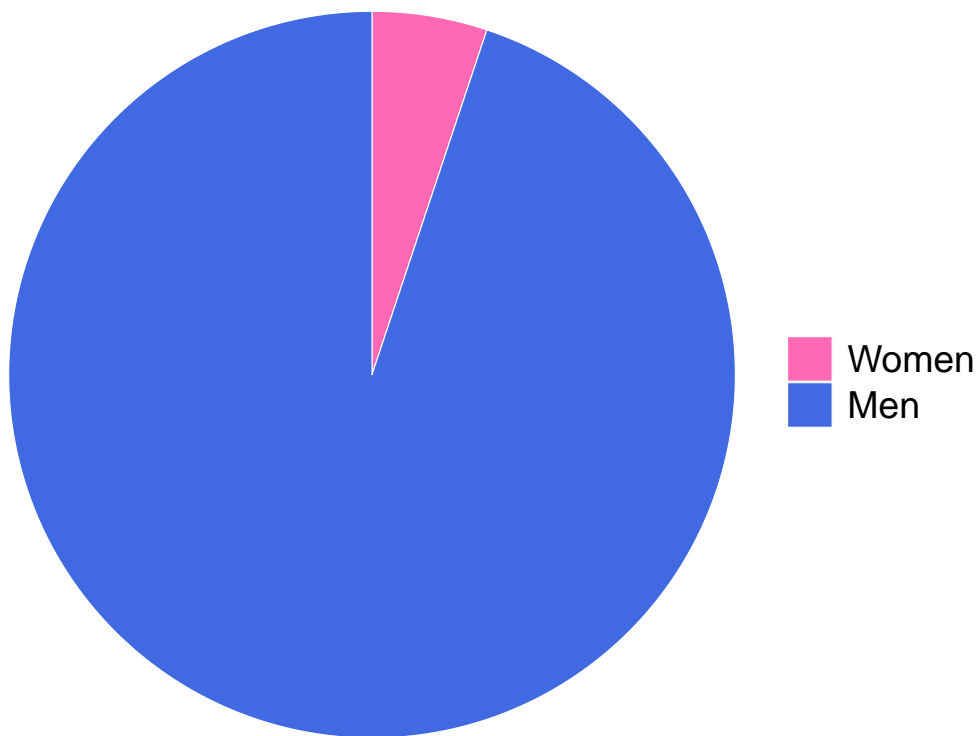
smallTable <- data.frame(category= c("Women", "Men"),
                        percent = c(subsetData$`% Of Women in Chamber`,subsetData$homme))

pie <- ggplot(smallTable,
      aes(x0 = 0, y0 = 0, r0 = 0, r = 1,
          amount = percent,
          fill = fct_relevel(category, "Women", "Men")) +
```

```
coord_fixed() +
ggtitle("Diagramme en secteurs") +
geom_arc_bar(stat = "pie", color = "white", size = 0.2) +
scale_fill_manual(values = c("#FF69B4", "#4169E1")) +
theme_void() +
theme(plot.title = element_text(size = 20, hjust = 0.5),
      legend.position = "right",
      legend.text = element_text(size = 14),
      legend.title = element_blank(),
      axis.title = element_blank())
```

pie

Diagramme en secteurs



Proportion de femme dans l'assemblée constituante de la cinquième république

Date: adoptée le 28 septembre 1958, promulguée le 4 octobre 1958

j'ai prit le type de chambre lower, je n'ai pas pu trouver la différence entre lower et upper mais lower avait plus de siège donc je l'ai sélectionné

```
subsetData5eme <- subset(data, Year==1958 & Country=="France" & `Chamber Type`=="Lower")
subsetData5eme
```

```
## # A tibble: 1 x 10
##   Country Region 'Election / Renewal' Year Month 'Chamber Type'
##   <chr>   <chr>   <chr>           <dbl> <chr> <chr>
## 1 France EUR     Yes             1958 <NA> Lower
## # i 4 more variables: 'Chamber Total Seats' <dbl>, 'Total women' <dbl>,
```

```
## #   '% Of Women in Chamber' <dbl>, NOTES <chr>

percentOfWomen <- subsetData[c("% Of Women in Chamber")] # le pourcentage de femmes

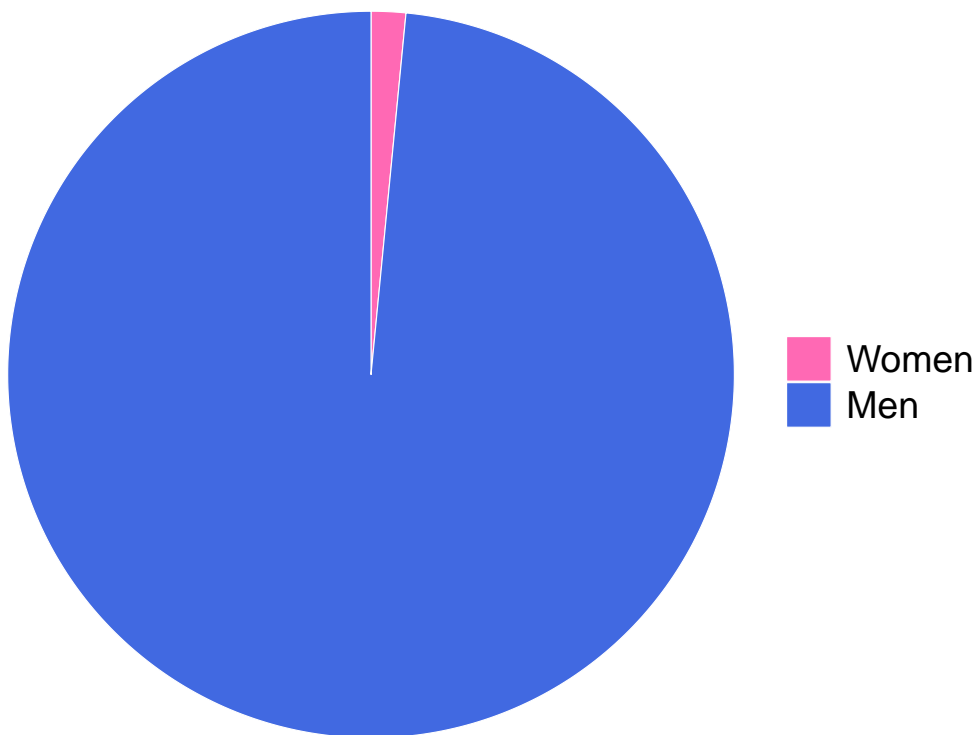
subsetData5eme$homme <- 1 - subsetData5eme$"% Of Women in Chamber"

smallTable <- data.frame(category= c("Women", "Men"),
                          percent = c(subsetData5eme$"% Of Women in Chamber",subsetData5eme$homme))

pie <- ggplot(smallTable,
              aes(x0 = 0, y0 = 0, r0 = 0, r = 1,
                  amount = percent,
                  fill = fct_relevel(category, "Women", "Men"))) +
  coord_fixed() +
  ggtitle("Diagramme en secteurs") +
  geom_arc_bar(stat = "pie", color = "white", size = 0.2) +
  scale_fill_manual(values = c("#FF69B4", "#4169E1")) +
  theme_void() +
  theme(plot.title = element_text(size = 20, hjust = 0.5),
        legend.position = "right",
        legend.text = element_text(size = 14),
        legend.title = element_blank(),
        axis.title = element_blank())

pie
```

Diagramme en secteurs



Evolution de la proportion de femmes dans l'assemblée basse Colombienne de 1964 à 2014

```
subsetDataColombia <- subset(data, Country=="Colombia" & `Chamber Type` == "Lower")
subsetDataColombia
```

```
## # A tibble: 16 x 10
##   Country Region `Election / Renewal` Year Month `Chamber Type`
##   <chr>    <chr>    <chr>          <dbl> <chr> <chr>
## 1 Colombia AME      Yes          1966 <NA> Lower
## 2 Colombia AME      Yes          1968 <NA> Lower
## 3 Colombia AME      Yes          1970 <NA> Lower
## 4 Colombia AME      Yes          1974 <NA> Lower
## 5 Colombia AME      Yes          1978 <NA> Lower
## 6 Colombia AME      Yes          1982 <NA> Lower
## 7 Colombia AME      Yes          1986 <NA> Lower
## 8 Colombia AME      Yes          1990 <NA> Lower
## 9 Colombia AME      Yes          1991 <NA> Lower
## 10 Colombia AME      Yes          1994 <NA> Lower
## 11 Colombia AME      Yes          1998 <NA> Lower
## 12 Colombia AME      Yes          2002 <NA> Lower
## 13 Colombia AME      Yes          2006 <NA> Lower
## 14 Colombia AME      Yes          2010 <NA> Lower
## 15 Colombia AME      Yes          2014 <NA> Lower
## 16 Colombia AME      Yes          2018 MARCH Lower
## # i 4 more variables: `Chamber Total Seats` <dbl>, `Total women` <dbl>,
## #   `% Of Women in Chamber` <dbl>, NOTES <chr>
```

Données

```
subsetDataColombia$homme <- 1 - subsetDataColombia$`% Of Women in Chamber`

smallTableColombia <- data.frame(
  year = subsetDataColombia$Year,
  category = c(rep("Women", length(subsetDataColombia$`% Of Women in Chamber`)),
    rep("Men", length(subsetDataColombia$homme))),
  percent = c(subsetDataColombia$`% Of Women in Chamber`, subsetDataColombia$homme)
)
```

Diagramme barre empilées

```
ggplot(smallTableColombia,
  aes(x = year, y = percent, fill = category)) +
  geom_bar(position = "stack", stat = "identity") +
  scale_fill_manual(values = c("pink", "blue")) +
  labs(x = "", y = "Percent of Chamber Members") +
  theme_classic()
```

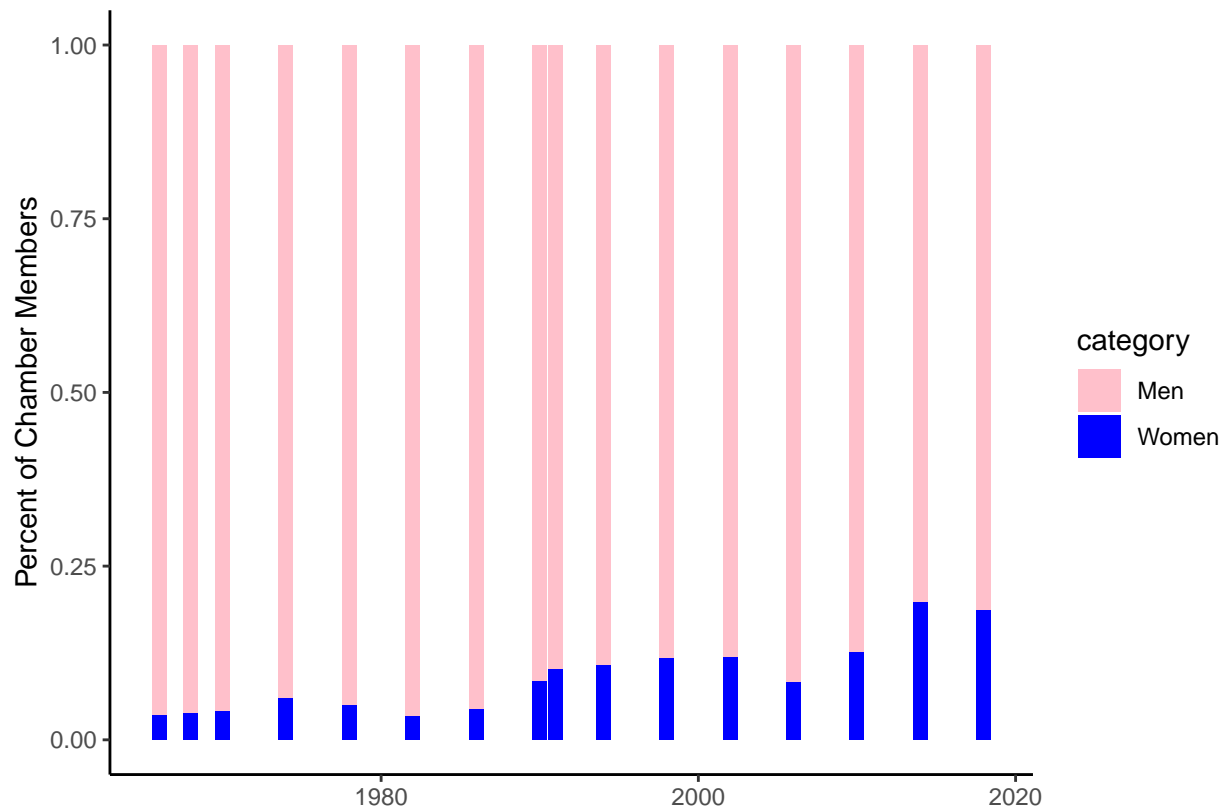
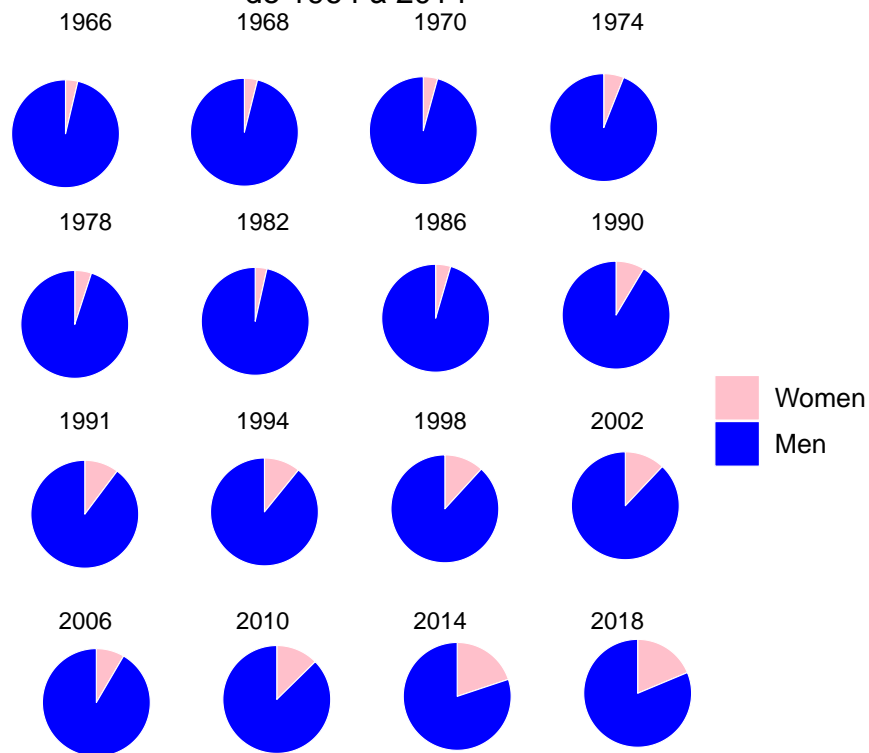


Diagramme en secteur

```
pie <- ggplot(smallTableColombia,
  aes(x0 = year, y0 = year, r0 = 70, r = 0,
    amount = percent,
    fill = fct_relevel(category, "Women", "Men"))) +
  coord_fixed() +
  ggtitle("Evolution de la proportion de femmes dans l'assemblée basse Colombienne \n de 1964 à 2014") +
  geom_arc_bar(stat = "pie", color = "white", size = 0.2) +
  scale_fill_manual(values = c("pink", "blue")) +
  theme_void() +
  theme(plot.title = element_text(size = 12, hjust = 0.5),
    legend.position = "right",
    legend.text = element_text(size = 10),
    legend.title = element_blank(),
    axis.title = element_blank()) +
  facet_wrap(~year, ncol = 4)

pie
```

Évolution de la proportion de femmes dans l'assemblée basse Colombienne de 1964 à 2014



Autres questions

Question: Evolution de la proportion de femmes dans l'assemblée basse Italienne de 1948 à 2008

Data

```
subsetDataItalie <- subset(data, (Year == "1948" | Year == "1976" | Year == "2008") &
                             Country == "Italy" & `Chamber Type` == "Lower"
                             )
subsetDataItalie
```

```
## # A tibble: 3 x 10
##   Country Region `Election / Renewal` Year Month `Chamber Type`
##   <chr>   <chr>   <chr>           <dbl> <chr> <chr>
## 1 Italy   EUR       Yes           1948 <NA> Lower
## 2 Italy   EUR       Yes           1976 <NA> Lower
## 3 Italy   EUR       Yes           2008 <NA> Lower
## # i 4 more variables: `Chamber Total Seats` <dbl>, `Total women` <dbl>,
## #   `% Of Women in Chamber` <dbl>, NOTES <chr>
```

Premier graph

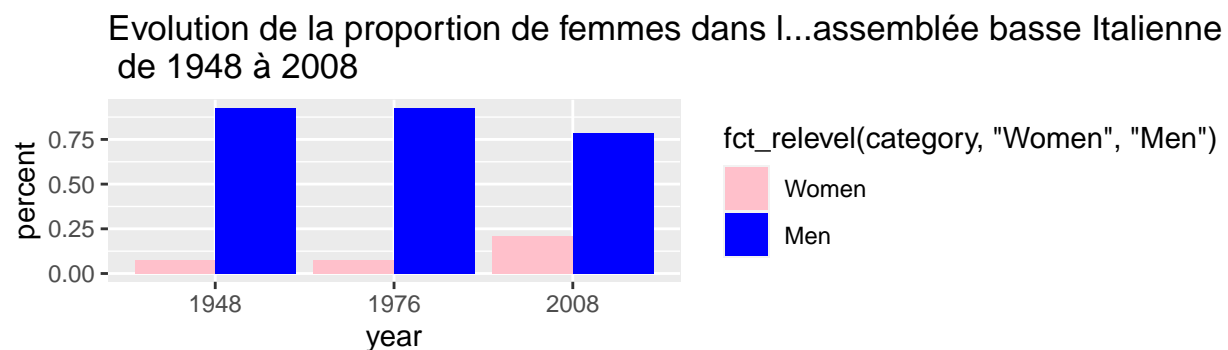
```
library(ggplot2)

subsetDataItalie$homme <- 1 - subsetDataItalie$`% Of Women in Chamber`
```

```
smallDataItalie <- data.frame(
  year = c("1948", "1976", "2008"),
  category = c(rep("Women", length(subsetDataItalie$`% Of Women in Chamber`)),
    rep("Men", length(subsetDataItalie$homme))),
  percent = c(subsetDataItalie$`% Of Women in Chamber`, subsetDataItalie$homme)
)
smallDataItalie
```

```
##   year category   percent
## 1 1948    Women 0.07665505
## 2 1976    Women 0.07460317
## 3 2008    Women 0.21144674
## 4 1948     Men 0.92334495
## 5 1976     Men 0.92539683
## 6 2008     Men 0.78855326
```

```
ggplot(smallDataItalie,
  aes(x = year,
    y = percent,
    fill = fct_relevel(category, "Women", "Men"))) +
  coord_fixed() +
  ggtitle("Evolution de la proportion de femmes dans l'assemblée basse Italienne \n de 1948 à 2008") +
  geom_col(position = "dodge") +
  scale_fill_manual(values = c("pink", "blue"))
```



```
subsetDataItaliePie <- subset(data,
  Country == "Italy" & `Chamber Type` == "Lower"
)
subsetDataItaliePie
```

```
## # A tibble: 19 x 10
##   Country Region 'Election / Renewal' Year Month 'Chamber Type'
##   <chr>    <chr>    <chr>      <dbl> <chr> <chr>
## 1 Italy   EUR      Yes      1948 <NA> Lower
## 2 Italy   EUR      Yes      1953 <NA> Lower
## 3 Italy   EUR      Yes      1958 <NA> Lower
## 4 Italy   EUR      Yes      1963 <NA> Lower
## 5 Italy   EUR      Yes      1968 <NA> Lower
## 6 Italy   EUR      Yes      1972 <NA> Lower
## 7 Italy   EUR      Yes      1976 <NA> Lower
```

```
## 8 Italy EUR Yes 1979 <NA> Lower
## 9 Italy EUR Yes 1983 <NA> Lower
## 10 Italy EUR Yes 1987 <NA> Lower
## 11 Italy EUR Yes 1992 <NA> Lower
## 12 Italy EUR Yes 1994 <NA> Lower
## 13 Italy EUR Yes 1996 <NA> Lower
## 14 Italy EUR Yes 2001 <NA> Lower
## 15 Italy EUR Yes 2006 <NA> Lower
## 16 Italy EUR Yes 2008 <NA> Lower
## 17 Italy EUR Yes 2013 March Lower
## 18 Italy EUR Yes 2013 May Lower
## 19 Italy EUR Yes 2018 March Lower
## # i 4 more variables: 'Chamber Total Seats' <dbl>, 'Total women' <dbl>,
## # '% Of Women in Chamber' <dbl>, NOTES <chr>
```

```
subsetDataItaliePie$homme <- 1 - subsetDataItaliePie$`% Of Women in Chamber`

smallTableItalie <- data.frame(
  year = subsetDataItaliePie$Year,
  category = c(rep("Women", length(subsetDataItaliePie$`% Of Women in Chamber`)),
    rep("Men", length(subsetDataItaliePie$homme))),
  percent = c(subsetDataItaliePie$`% Of Women in Chamber`, subsetDataItaliePie$homme)
)
smallTableColombia
```

```
##   year category   percent
## 1 1966   Women 0.03684211
## 2 1968   Women 0.03921569
## 3 1970   Women 0.04285714
## 4 1974   Women 0.06030151
## 5 1978   Women 0.05025126
## 6 1982   Women 0.03517588
## 7 1986   Women 0.04522613
## 8 1990   Women 0.08542714
## 9 1991   Women 0.10240964
## 10 1994   Women 0.10843373
## 11 1998   Women 0.11801242
## 12 2002   Women 0.12048193
## 13 2006   Women 0.08433735
## 14 2010   Women 0.12650602
## 15 2014   Women 0.19879518
## 16 2018   Women 0.18713450
## 17 1966    Men 0.96315789
## 18 1968    Men 0.96078431
## 19 1970    Men 0.95714286
## 20 1974    Men 0.93969849
## 21 1978    Men 0.94974874
## 22 1982    Men 0.96482412
## 23 1986    Men 0.95477387
## 24 1990    Men 0.91457286
## 25 1991    Men 0.89759036
## 26 1994    Men 0.89156627
## 27 1998    Men 0.88198758
## 28 2002    Men 0.87951807
```

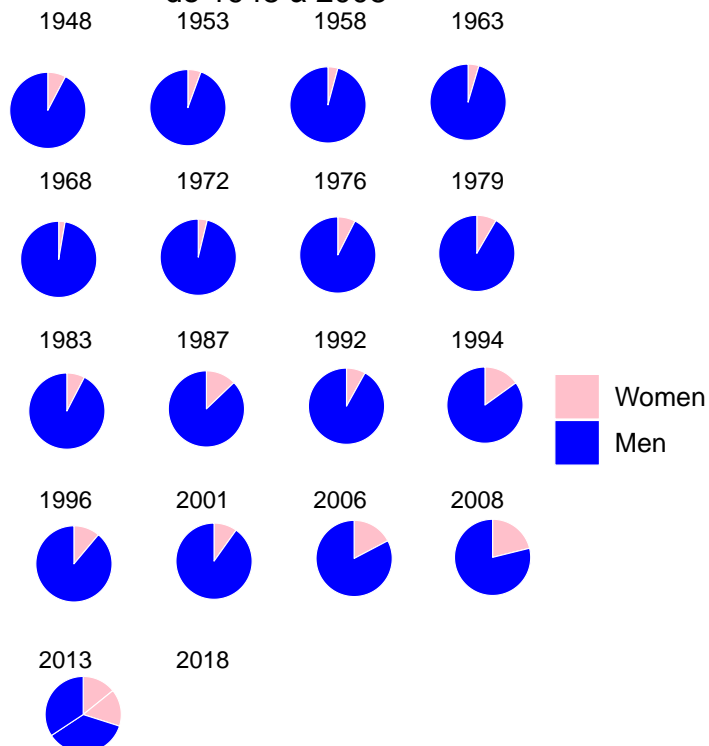


```
## 29 2006      Men 0.91566265
## 30 2010      Men 0.87349398
## 31 2014      Men 0.80120482
## 32 2018      Men 0.81286550
```

```
pie <- ggplot(smallTableItalie,
             aes(x0 = year, y0 = year, r0 = 70, r = 0,
                 amount = percent,
                 fill = fct_relevel(category, "Women", "Men"))) +
  coord_fixed() +
  ggtitle("Evolution de la proportion de femmes dans l'assemblée basse Italienne \n de 1948 à 2008") +
  geom_arc_bar(stat = "pie", color = "white", size = 0.2) +
  scale_fill_manual(values = c("pink", "blue")) +
  theme_void() +
  theme(plot.title = element_text(size = 12, hjust = 0.5),
        legend.position = "right",
        legend.text = element_text(size = 10),
        legend.title = element_blank(),
        axis.title = element_blank()) +
  facet_wrap(~year, ncol = 4)

pie
```

Evolution de la proportion de femmes dans l'assemblée basse Italienne de 1948 à 2008



Conclusion

Dans la première section, j'ai exploré les données en me concentrant sur un pays donné à une année donnée. Ensuite, dans la seconde section, j'ai examiné les données d'un pays donné au fil du temps.

J'ai décidé de poser ma propre question sur les données en utilisant les exemples donnés. J'ai choisi de me demander quelle était la proportion de femmes dans les chambres basses d'Italie de 1948 à 2008. Pour répondre à cette question, j'ai créé un graphique en utilisant les données correspondantes.

Tout d'abord, j'ai essayé de représenter les données sous forme de diagramme en barres. Cependant, le résultat n'était pas satisfaisant, car le graphique ne permettait pas une comparaison claire des proportions de femmes et d'hommes dans les différentes années. J'ai donc décidé de représenter les données sous forme de diagramme circulaire.

Le graphique montre l'évolution de la proportion de femmes dans l'assemblée basse italienne de 1948 à 2008. Les données sont présentées sous forme de pourcentages, avec la proportion de femmes en rose et celle des hommes en bleu. Chaque année est représentée dans un cercle séparé, permettant de comparer facilement les proportions entre les différentes années.

En somme, la représentation choisie permet une comparaison facile des proportions de femmes et d'hommes dans les différentes années, permettant ainsi de répondre à la question posée de manière claire et concise