

Medailles Jeux Olympiques d'hiver de 1924 à 2010

Kathleen GALVIN & Marie BERGERET

30 mars 2017

```
library(readr)
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Présentation des données

Importation

```
donnees<- read_csv(file = "winter_olympics_medals.csv")

## Parsed with column specification:
## cols(
##   id = col_integer(),
##   year = col_integer(),
##   sport = col_character(),
##   medal = col_character(),
##   code = col_character(),
##   country = col_character(),
##   host = col_logical()
## )
head(donnees)

## # A tibble: 6 × 7
##   id year sport medal code country host
##   <int> <int> <chr> <chr> <chr> <chr> <lgl>
## 1 31666 1924 biathlon gold SUI Switzerland FALSE
## 2 31666 1924 biathlon silver FIN Finland FALSE
## 3 31666 1924 biathlon bronze FRA France TRUE
## 4 31688 1924 bobsleigh gold SUI Switzerland FALSE
## 5 31688 1924 bobsleigh silver GBR GreatBritain FALSE
## 6 31688 1924 bobsleigh bronze BEL Belgium FALSE
```

Dates

Nous possédons les données des Jeux Olympiques d'hivers qui ont eu lieu entre les deux années suivantes.

```
donnees%>%summarise(AnneeMin=min(year))
```

```
## # A tibble: 1 × 1
##   AnneeMin
##   <int>
## 1     1924
```

```
donnees%>%summarise(AnneeMax=max(year))
```

```
## # A tibble: 1 × 1
##   AnneeMax
##   <int>
## 1     2010
```

Les JO d'hiver ayant lieu tous les quatres ans, il n'est pas nécessaire d'afficher toutes les années.

Sports

Voici la liste des différents sports représentés de 1924 à 2010 :

```
donnees%>%select(sport)%>%unique()
```

```
## # A tibble: 15 × 1
##       sport
##   <chr>
## 1  biathlon
## 2  bobsleigh
## 3  crosscountry
## 4  figureskating
## 5  icehockey
## 6  nordiccombined
## 7  skijumping
## 8  speedskating
## 9  curling
## 10 skeleton
## 11 alpinesskiing
## 12 luge
## 13 freestyleskiing
## 14 shorttrack
## 15 snowboard
```

Pays

Lors de tous ces jeux, de nombreux pays ont participé. Ces données se situent dans un autre fichier csv que nous allons importer.

Participants

```
dpays<- read_csv(file = "pays.csv")
```

```
## Parsed with column specification:
## cols(
##   code = col_character(),
```

```
## country = col_character()
## )
```

```
head(dpays)
```

```
## # A tibble: 6 × 2
##   code      country
##   <chr>      <chr>
## 1  AFG  Afghanistan
## 2  RSA  South Africa
## 3  ALB    Albania
## 4  ALG    Algeria
## 5  GER    Germany
## 6  AND    Andorra
```

```
dpays%>%select(code)%>%unique()%>%summarise(N=n())
```

```
## # A tibble: 1 × 1
##       N
##   <int>
## 1   228
```

Ainsi, 228 pays ont participé à au moins un JO entre ces deux dates.

Gagnants

```
donnees%>%select(code)%>%unique()%>%summarise(N=n())
```

```
## # A tibble: 1 × 1
##       N
##   <int>
## 1    42
```

Ainsi, 42 pays ont gagné des médailles entre 1924 et 2010. Nous avons remarqué qu'en comptant par nom de pays ('country') plutôt que par le code ('code'), nous obtenions plus de résultats (46). Cela s'explique par le fait que certains pays ont changé de nom comme l'Allemagne : le code est resté 'GER' alors que le nom est passé de 'Federal Republic of Germany (1950-1990)' et 'German Democratic Republic (1955-1990)' à 'Germany'. Il nous ait paru plus cohérent de travailler avec le code des pays.

HYPOTHESE !!!!!!!!!!!!!!!

Quel pays gagne le plus/ le moins ?

Nous allons maintenant regarder la répartition du nombre de médailles gagnées par pays. ## Peu importe la couleur de la médaille On va donc compter le nombre de médaille par pays.

```
donnees%>%group_by(code)%>%summarise(n=n())
```

```
## # A tibble: 42 × 2
##   code      n
##   <chr> <int>
## 1  AUS      9
## 2  AUT     201
## 3  BEL      5
## 4  BLR      9
```

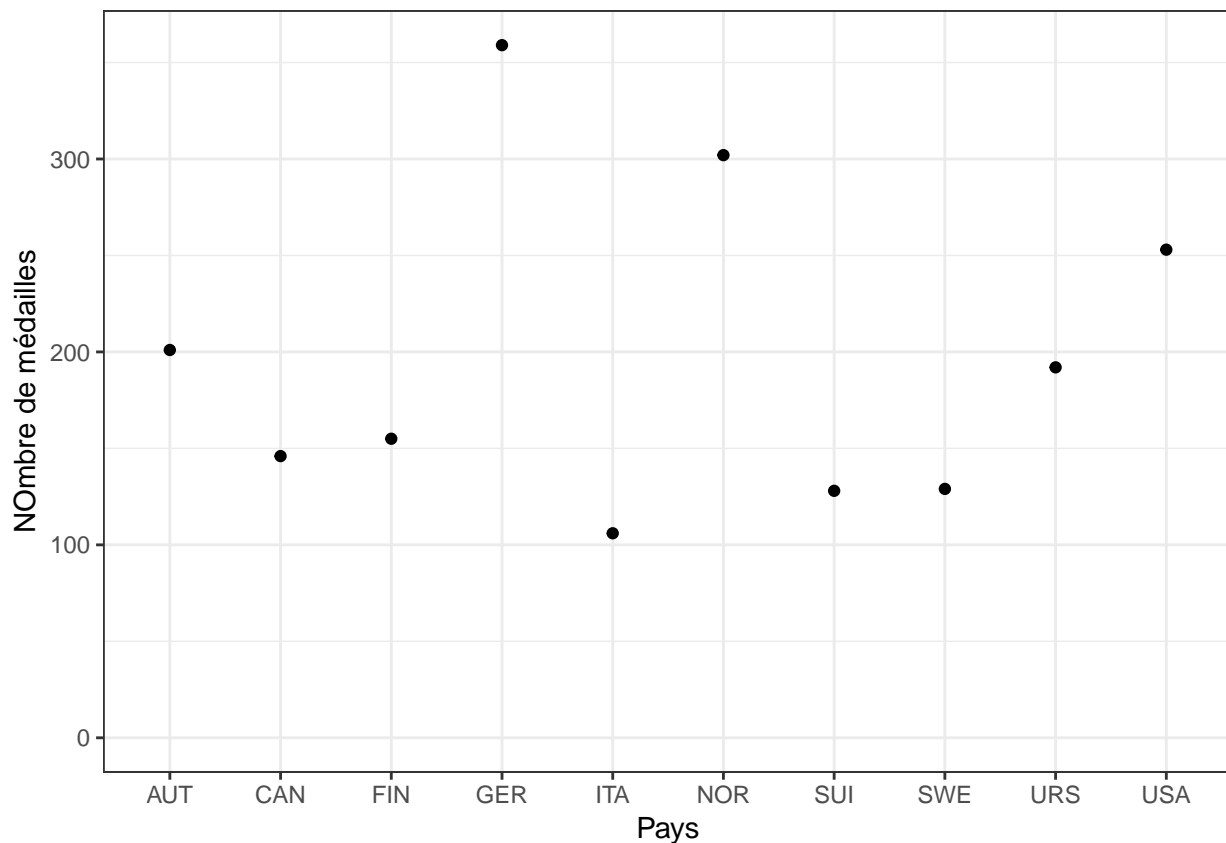
```
## 5    BUL     6
## 6    CAN   146
## 7    CHN    44
## 8    CRO    10
## 9    CZE    16
## 10   DEN     1
## # ... with 32 more rows
```

Le plus ?

Nous allons extraire les 10 pays qui ont gagné le plus de médailles sur ces 86 années.

```
library(ggplot2)
donnees%>%group_by(code)%>%summarise(N=n())%>%arrange(-N)%>%slice(1:10)%>%ggplot(aes(x=code, y=N))+geom_point()

## Warning: Ignoring unknown parameters: width
```

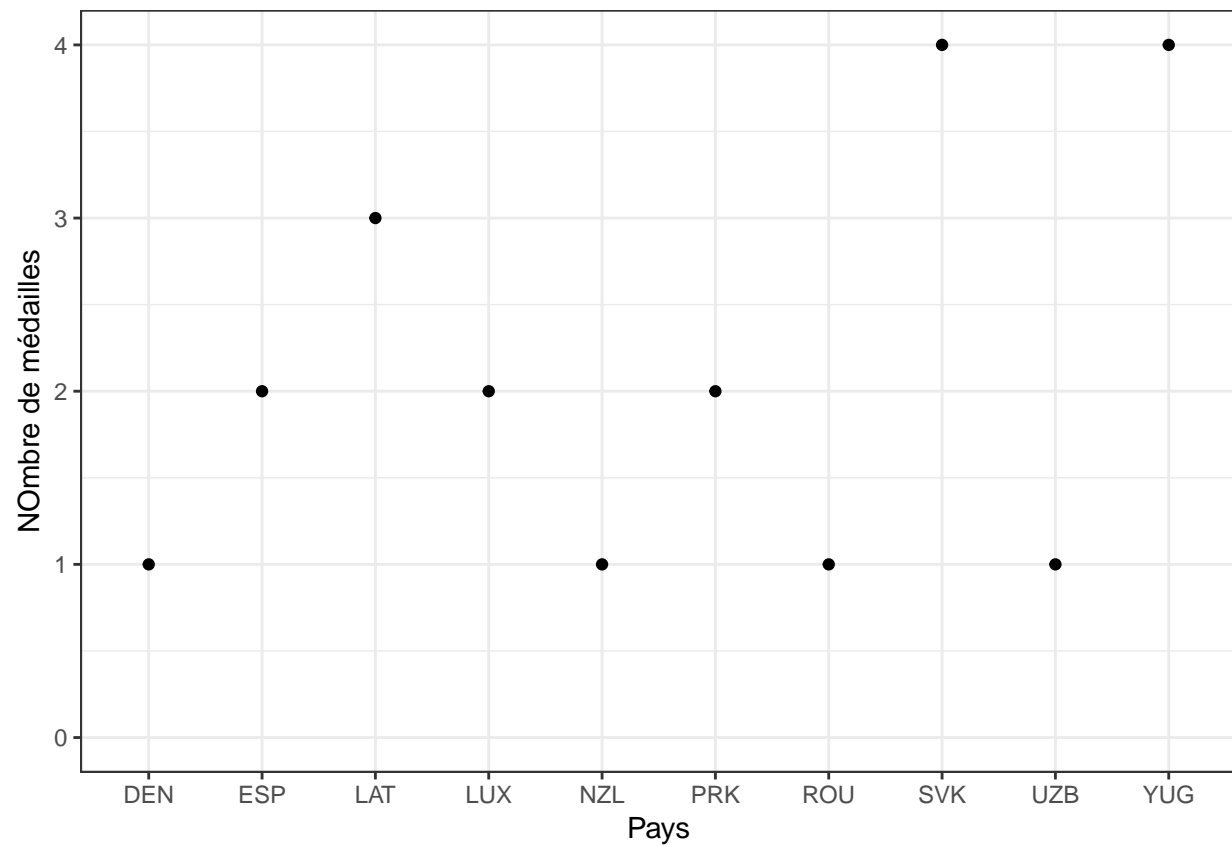


Le moins ?

Nous allons extraire les 10 pays qui ont gagné le moins de médailles sur ces 86 années.

```
library(ggplot2)
donnees%>%group_by(code)%>%summarise(N=n())%>%arrange(N)%>%slice(1:10)%>%ggplot(aes(x=code, y=N))+geom_point()

## Warning: Ignoring unknown parameters: width
```

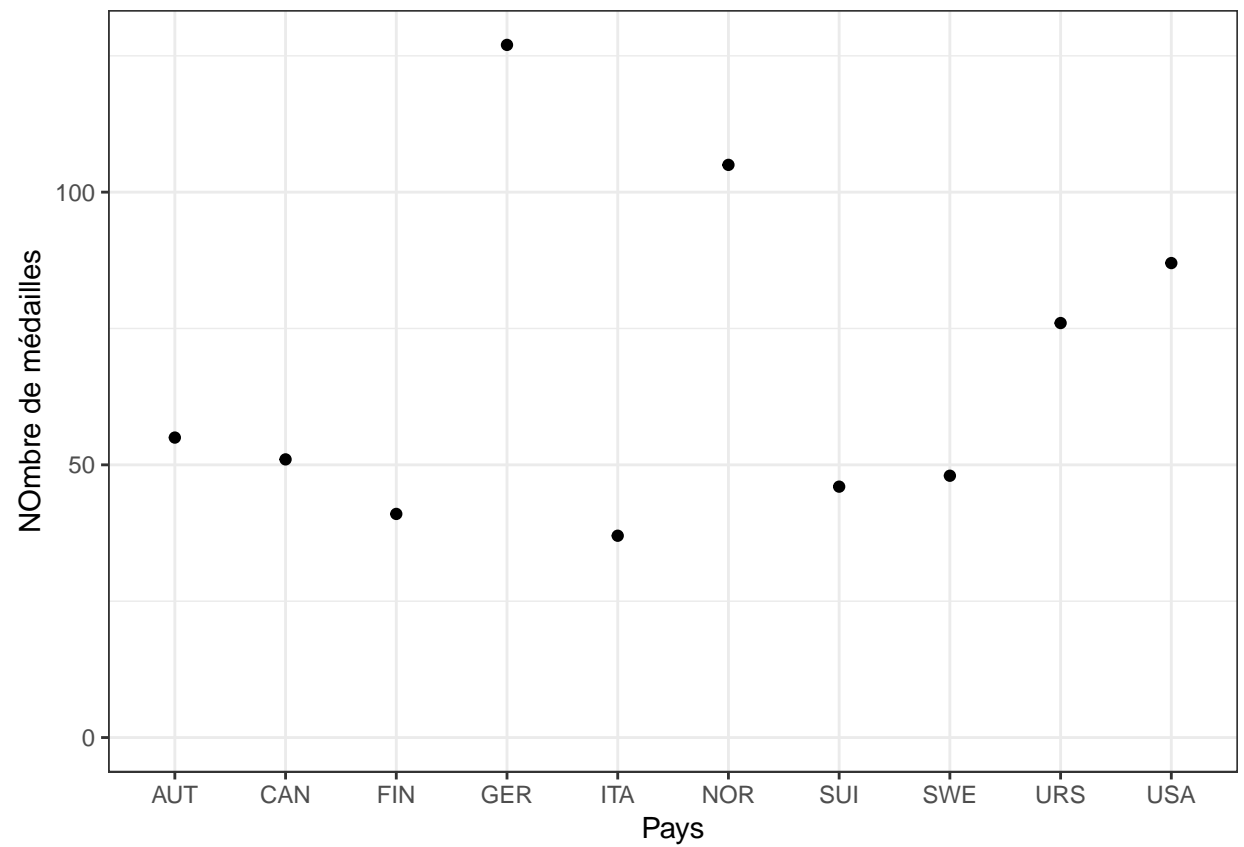


Selon la couleur de la médaille

Or

```
donnees%>%group_by(code)%>%filter(medal=='gold')%>%summarise(N=n())%>%arrange(-N)%>%slice(1:10)%>%ggplot
```

```
## Warning: Ignoring unknown parameters: width
```



Argent

A completer

Bronze

A completer