

# Médailles Jeux Olympiques d'hiver de 1924 à 2010

*Kathleen GALVIN & Marie BERGERET*

*30 mars 2017*

Dans le cadre du cours de Modèles Statistiques, nous allons analyser un jeu de données, selon les méthodes étudiées en TD. Nous avons donc cherché des données sur un sujet qui nous intéressait. Grâce au site [www.data.gouv.fr](http://www.data.gouv.fr), nous avons trouvé une base de données contenant des informations sur les nations médaillées au Jeux Olympiques (JO) d'hiver entre les années 1924 et 2010.

Dans un premier temps il a été nécessaire d'installer plusieurs packages :

```
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
library(ggplot2)
```

## 1 Présentation des données

Les données sources se situent dans le fichier 'winter\_olympics\_medals.csv', nous allons donc les importer.

### 1.1 Importation

Ci-dessous se trouve un extrait des données, à savoir les 7 premières lignes du fichier (Intitulé des colonnes et lignes de données).

```
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
```

Cette base de données est constituée de 2 variables quantitatives ('int') et de 5 variables qualitatives ('chr' et 'lgl') : - id : Identifiant (il est le même pour un même sport et une même année, nous ne possédons pas le nom des différentes épreuves)

- year : Année des JO

- sport : Nom du sport (et pas celui de l'épreuve)

- medal : Couleur de la médaille obtenue - gold, silver, bronze -

- code : Code de la nation

- country : Nom de la nation

- host : Booléen qui indique si le pays ayant gagné la médaille était le pays d'accueil ou non - TRUE, FALSE -

Une ligne correspond à la victoire d'une nation dans un sport lors d'une année de jeux.

## 1.2 Dates

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

```
min(donnees$year)
```

```
## [1] 1924
```

```
max(donnees$year)
```

```
## [1] 2010
```

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

## 1.3 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 alpinesthinking
## 12     luge
## 13 freestyleskiing
```

```
## 14      shorttrack
## 15      snowboard
```

Il y a donc 15 sport différents aux JO d'hiver (du moins jusqu'en 2010).

## 1.4 Pays

Lors de tous ces jeux, de nombreux pays ont participé.

### 1.4.1 Participants

Ces données se situent dans un autre fichier csv que nous allons importer : 'pays.csv'. Ci-dessous un extrait de ces données (Code et nom de nation).

```
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
```

Nous allons maintenant compter le nombre de nations ayant participé aux JO.

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

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

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

### 1.4.2 Gagnants

Nous allons désormais nous concentrer sur les pays ayant célébré au moins une victoire, c'est à dire ayant gagné une médaille.

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

```
## # A tibble: 1 × 1
##       N
##   <int>
## 1    45
```

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

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

Ainsi, 42 nations 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. Cela s'explique par le fait que certains pays ont changé de nom au cours du temps, comme l'Allemagne : le code est resté 'GER' alors que le nom est passé de 'Federal Republic of Germany (1950-1990)', 'German Democratic Republic (1955-1990)' et 'UnitedTeamofGermany(1956-1960)' à 'Germany'. Il nous ait paru plus cohérent de travailler avec le code des pays. Vous trouverez en annexe un tableau de correspondance entre les codes et les nom des nations.

Maintenant que nous avons présenté les données, nous allons les analyser selon une certaine ligne de réflexion. Dans un premier temps, nous allons essayer de trouver les pays qui ont gagné le plus (et réciproquement le moins) de médailles sur l'ensemble de ces jeux puis par sport. Ensuite, nous allons regarder si ce sont toujours les mêmes pays qui restent "au sommet" ou si, au cours des années, de nouveaux pays "compétitifs" ont émergé. Enfin, dans une dernière partie qui résulte surtout de notre curiosité, nous étudierons le cas de la France.

## 2 Quel pays gagne le plus/ le moins ?

Nous allons maintenant regarder la répartition du nombre de médailles gagnées par pays sur l'ensemble de JO.

### 2.1 Peu importe la couleur de la médaille

On va donc compter le nombre de médailles par pays (nous raisonnons toujours sur le code) sur les 86 ans de données que nous possédons.

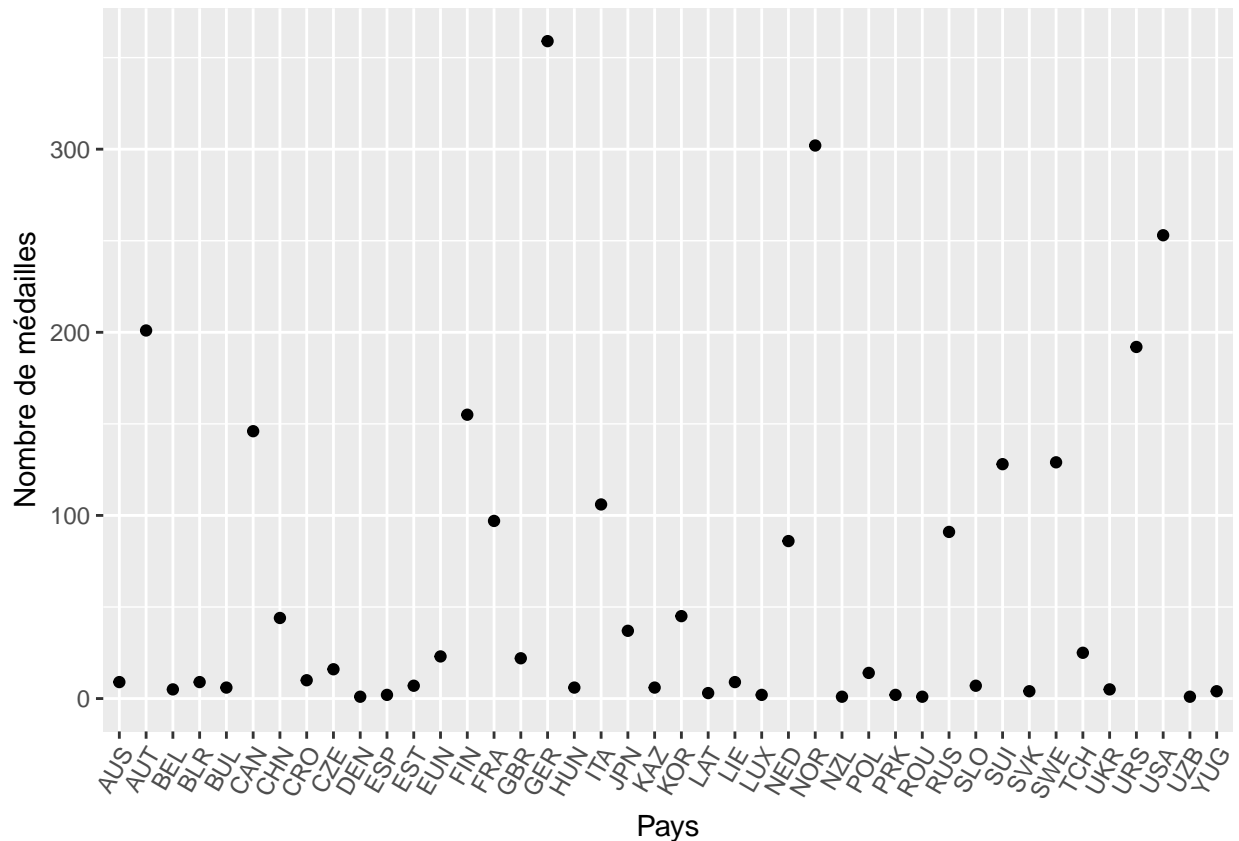
```
donnees%>%group_by(code)%>%summarise(N=n())
```

```
## # 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
```

Nous allons maintenant transformer ces données en graphique afin de mieux les visualiser.

```
donnees%>%group_by(code)%>%summarise(N=n())%>%arrange(-N)%>%ggplot(aes(x=code, y=N)) + geom_point(width=
```

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

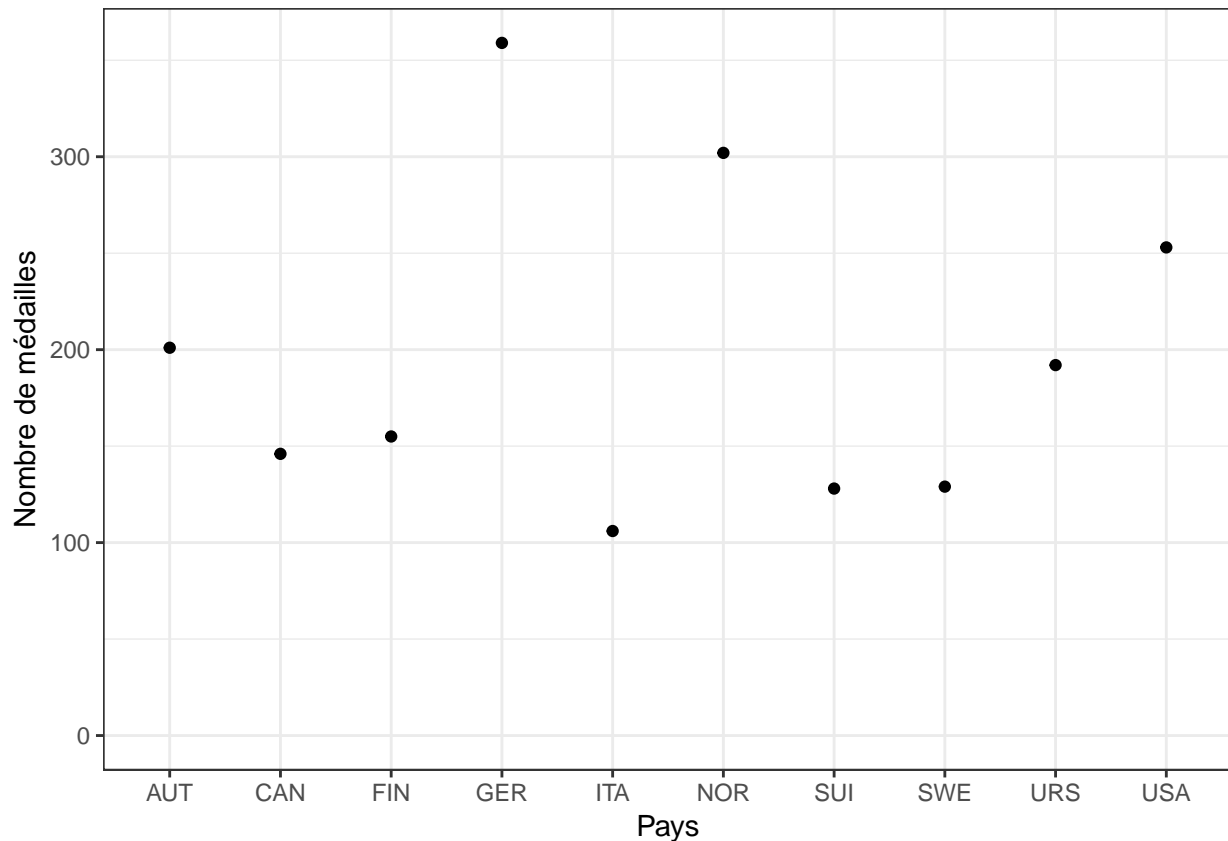


Le nombre de pays étant grand il n'est pas forcément aisé de lire ce graphique mais on voit quand même très bien les nations qui ont le plus gagné: Autriche, Allemagne, Norvège... De plus il est également possible de voir en un coup d'oeil les pays qui ont au contraire, gagné le moins de médailles : Roumanie, Danemark...

Pour une meilleur lisibilité, nous allons maintenant extraire les 10 pays qui ont gagné le plus de médailles sur ces années

```
MedailleGeneral<-donnees%>%group_by(code)%>%summarise(N=n())%>%arrange(-N)%>%slice(1:10)
MedailleGeneral%>%ggplot(aes(x=code, y=N)) + geom_point(width=0.3)+xlab("Pays")+ylim(0,NA)+ylab("Nombre de médailles")

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

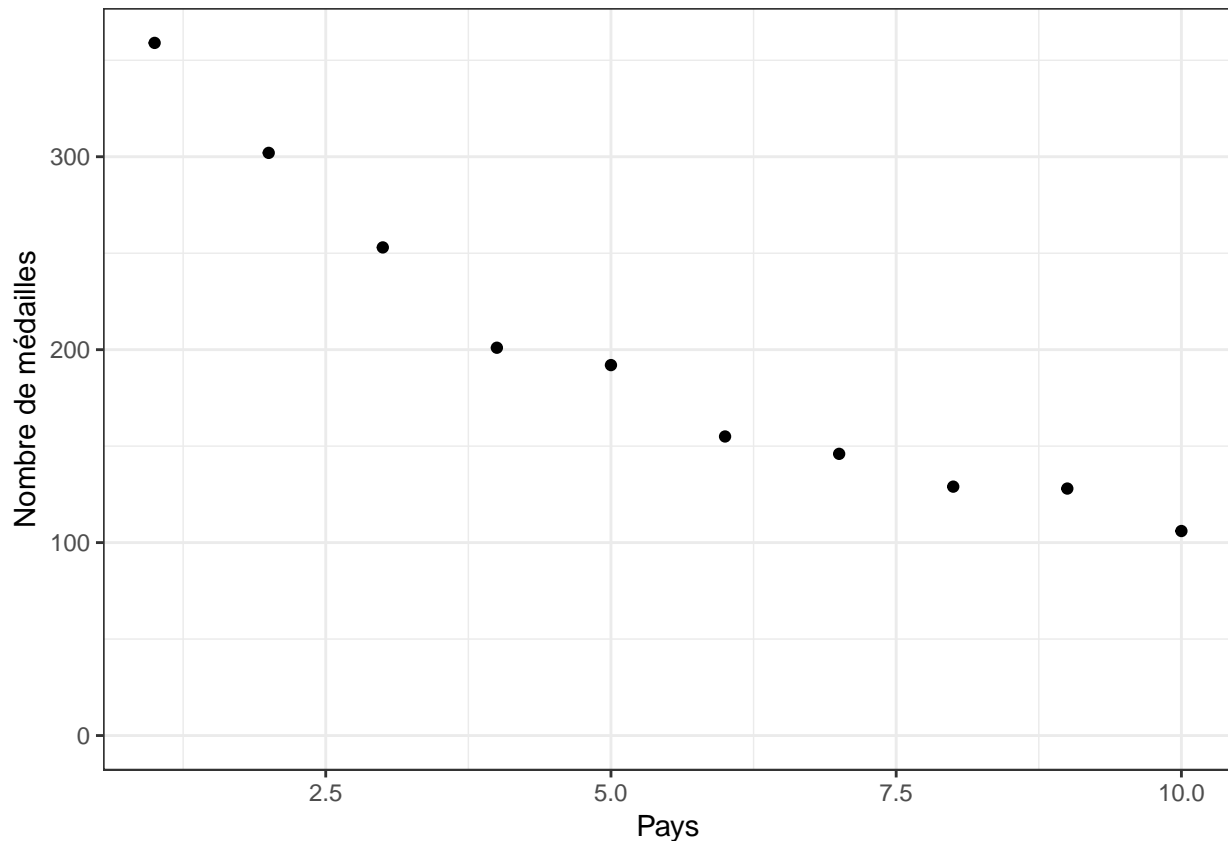


Nous pouvons donc voir les 10 “meilleurs” pays, enfin, plus rigoureusement les pays qui ont gagné le plus de médailles (toutes couleurs confondues) lors de l’ensemble des JO. Comme nous le pensions les pays tels que l’Allemagne, les pays scandinaves ou encore les Etats Unis y figurent. Cependant la France n’entre pas dans le top 10, mais nous verrons son cas dans la dernière partie du rapport.

Nous n’avons pas au premier coup d’oeil un classement par ordre croissant ou décroissant, nous allons donc essayer d’ordonner le graphique.

```
donnees%>%group_by(code)%>%summarise(N=n())%>%arrange(-N)%>%slice(1:10)%>%mutate(ordre=1:10)%>% ggplot()
```

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



Ainsi, nous visualisons beaucoup plus facilement le “classement” et le nombre de médailles de chaque nation : Allemagne, Norvège... Cependant nous n’avons pas réussi à afficher le code des nations plutôt que leur ordre et il faut donc se référer au tableau suivant, ce qui n’est pas très pratique.

```
donnees%>%group_by(code)%>%summarise(N=n())%>%arrange(-N)%>%slice(1:10)%>%mutate(ordre=1:10)%>%select(
```

```
## # A tibble: 10 × 2
##   code ordre
##   <chr> <int>
## 1   GER     1
## 2   NOR     2
## 3   USA     3
## 4   AUT     4
## 5   URS     5
## 6   FIN     6
## 7   CAN     7
## 8   SWE     8
## 9   SUI     9
## 10  ITA    10
```

### 2.1.1 Le moins ?

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

```
donnees%>%group_by(code)%>%summarise(N=n())%>%arrange(N)%>%slice(1:10)
```

```
## # A tibble: 10 × 2
##   code      N
##   <chr> <int>
```

```
##      <chr> <int>
## 1    DEN      1
## 2    NZL      1
## 3    ROU      1
## 4    UZB      1
## 5    ESP      2
## 6    LUX      2
## 7    PRK      2
## 8    LAT      3
## 9    SVK      4
## 10   YUG      4
```

On peut voir que tous ces pays ont gagné moins de 5 médailles au JO d'hiver.

Maintenant que nous avons vu les pays qui gagnent le plus et le moins de médailles toutes couleurs confondues, nous allons regarder pour chaque couleur de médaille afin de voir si ce sont toujours les mêmes pays et dans le même ordre.

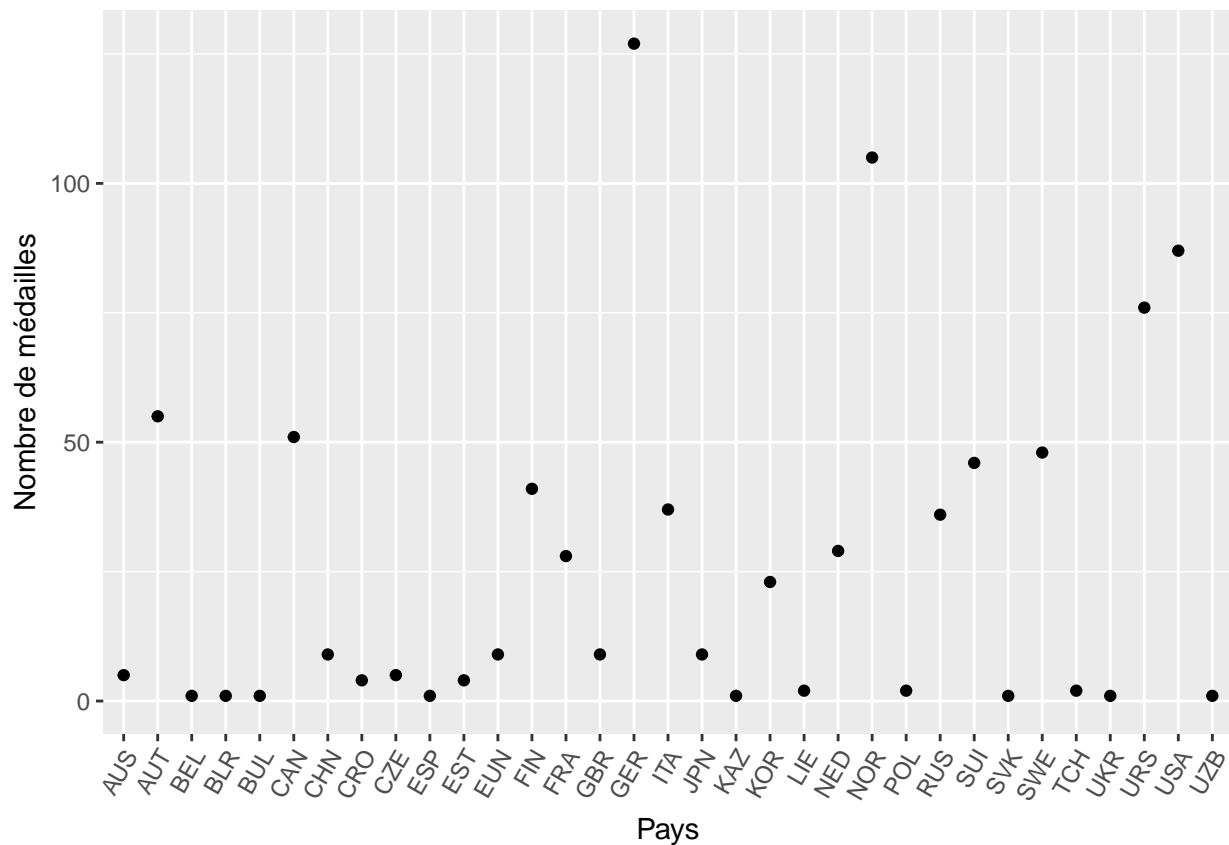
## 2.2 Selon la couleur de la médaille

### 2.2.1 Or

Nous allons faire comme précédemment mais en ne sélectionnant que les médailles d'or.

```
MedailleOr<-donnees%>%group_by(code)%>%filter(medal=='gold')%>%summarise(N=n())%>%arrange(-N)%>%slice(1:10)
donnees%>%group_by(code)%>%filter(medal=='gold')%>%summarise(N=n())%>%ggplot(aes(x=code, y=N)) + geom_point()
```

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





Comme il n'est pas très facile de confronter le plot des médailles toutes couleurs confondues et des médailles d'or, et afin de répondre à notre question, nous allons confronter les liste des 10 pays qui ont gagné le plus dans les deux cas.

```
tout<-MedailleGeneral%>%mutate(ordre=1:10)%>%select(code,ordre)%>%arrange(code)
```

```
or<-MedailleOr%>%mutate(ordre=1:10)%>%select(code,ordre)%>%arrange(code)
```

```
tout$code
```

```
## [1] "AUT" "CAN" "FIN" "GER" "ITA" "NOR" "SUI" "SWE" "URS" "USA"
```

```
tout$ordre
```

```
## [1] 4 7 6 1 10 2 9 8 5 3
```

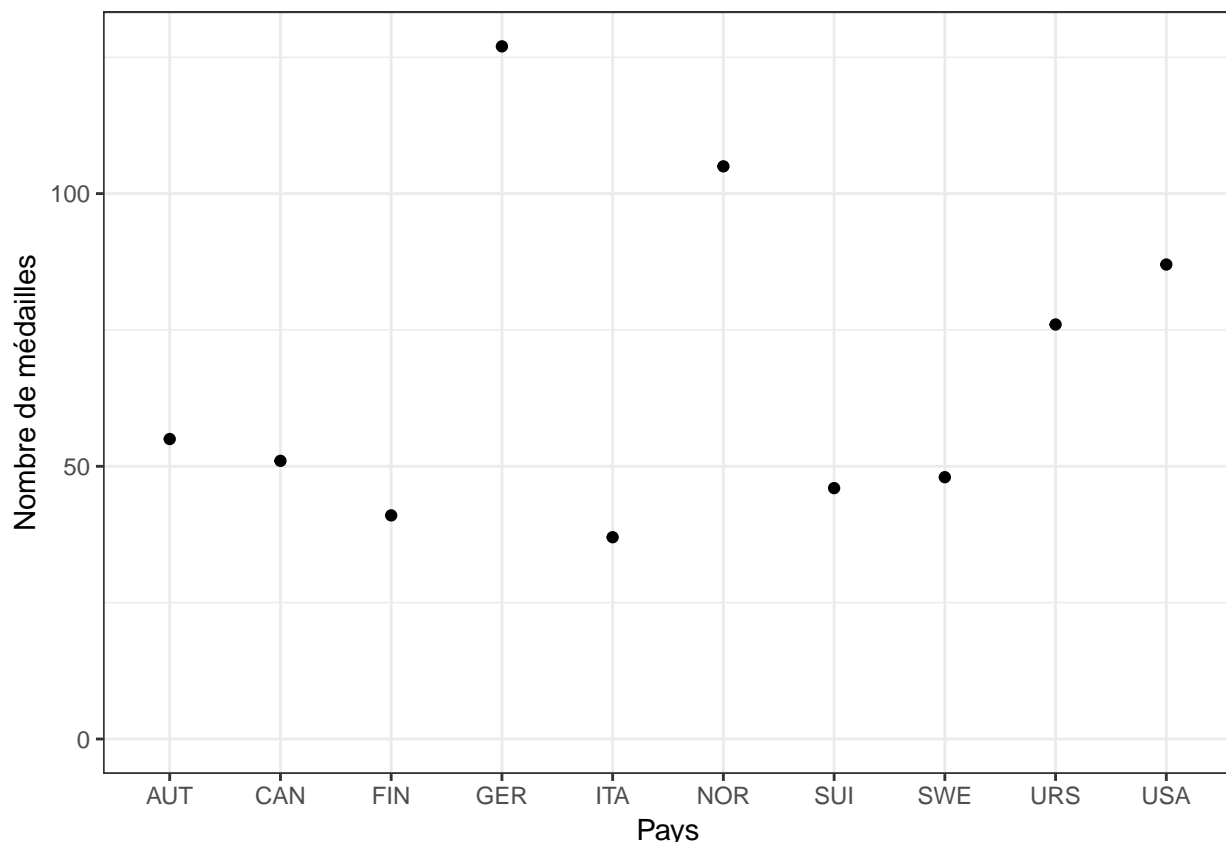
```
or$ordre
```

```
## [1] 5 6 9 1 10 2 8 7 4 3
```

Donc même si l'affichage n'est pas optimal, on peut déjà voir que c'est la même liste de pays qui gagnent le plus de médailles en général et les médailles d'or, seul l'ordre varie légèrement. En effet les trois premiers pays (Allemagne, Norvège et Finlande) restent classés dans le même ordre. Pour la plupart des nations nous observons une variation d'un rang sauf pour la Finlande qui est trois places derrière quand on regarde les médailles d'or.

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

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



L'allure de ce graphique est similaire au précédent concernant les médailles en général, les écarts entre les différents pays sont respectés, par exemple l'Allemagne et la Norvège se distinguent toujours autant.

### 2.2.2 Argent

Nous allons rapidement regarder si c'est la même chose pour les médailles d'argent

```
MedailleArgent<-donnees%>%group_by(code)%>%filter(medal=='silver')%>%summarise(N=n()) %>%arrange(-N)%>%
MedailleArgent
```

```
## # A tibble: 10 × 2
##   code      N
##   <chr> <int>
## 1   AUT    70
## 2   CAN    48
## 3   FIN    59
## 4   GER   129
## 5   ITA    32
## 6   NOR   107
## 7   SUI    36
## 8   SWE    34
## 9   URS    58
## 10  USA    90
```

Ce sont toujours les 10 mêmes pays.

```
argent<-MedailleOr%>%mutate(ordre=1:10)%>%select(code,ordre)%>%arrange(code)
tout$code
```

```
## [1] "AUT" "CAN" "FIN" "GER" "ITA" "NOR" "SUI" "SWE" "URS" "USA"
```

```
tout$ordre
```

```
## [1] 4 7 6 1 10 2 9 8 5 3
```

```
or$ordre
```

```
## [1] 5 6 9 1 10 2 8 7 4 3
```

```
argent$ordre
```

```
## [1] 5 6 9 1 10 2 8 7 4 3
```

Encore une fois l'ordre du podium est conservé et comme pour les médailles d'or, il y a très peu de mouvement de rangs.

### 2.2.3 Bronze

Et enfin pour les médailles de bronze :

```
MedailleBronze<-donnees%>%group_by(code)%>%filter(medal=='bronze')%>%summarise(N=n()) %>%arrange(-N)%>%
MedailleBronze%>%mutate(ordre=1:10)
```

```
## # A tibble: 10 × 3
##   code      N ordre
##   <chr> <int> <int>
## 1   GER   103     1
## 2   NOR    90     2
## 3   AUT    76     3
## 4   USA    76     4
## 5   URS    58     5
## 6   FIN    55     6
```

```
## 7    CAN    47    7
## 8    SWE    47    8
## 9    SUI    46    9
## 10   FRA    42   10
```

Cette fois nous ne pouvons pas faire la même méthode car un pays a été remplacé : la France a plus gagné de médailles que l'Italie ce qui l'a fait sortir du top 10. L'Autriche a également fait sa place au détriment des Etats Unis

Ainsi, on peut dire que le top 10 des nations qui gagnent le plus de médailles en général, d'or, d'argent ou de bronze reste constant mis à part l'Italie et la France qui échangent la place numéro 10. Nous allons maintenant regarder pour chaque JO, quels ont été les pays qui ont le plus gagné.

### 3 Analyse par sport

Nous voulons sortir la liste des trois nations qui obtiennent le plus de médailles dans chaque sport.

```
donnees%>%group_by(sport,code)%>%summarise(NombreMed=n())%>%arrange(-NombreMed)%>%slice(1:3)
```

```
## Source: local data frame [45 x 3]
## Groups: sport [15]
##
##      sport  code NombreMed
##      <chr> <chr>      <int>
## 1 alpinesskiing  AUT        105
## 2 alpinesskiing  SUI         55
## 3 alpinesskiing  FRA         43
## 4    biathlon   GER         59
## 5    biathlon   NOR         29
## 6    biathlon   RUS         20
## 7    bobsleigh   GER         40
## 8    bobsleigh   SUI         30
## 9    bobsleigh   USA         20
## 10 crosscountry  NOR         96
## # ... with 35 more rows
```

Par exemple, pour le biathlon les trois meilleurs pays sont :

```
donnees%>%group_by(sport,code)%>%summarise(NombreMed=n())%>%arrange(-NombreMed)%>%slice(1:3)%>%filter(sport=="biathlon")
```

```
## Source: local data frame [3 x 3]
## Groups: sport [1]
##
##      sport  code NombreMed
##      <chr> <chr>      <int>
## 1 biathlon   GER         59
## 2 biathlon   NOR         29
## 3 biathlon   RUS         20
```

Ou encore pour le patinage artistique :

```
donnees%>%group_by(sport,code)%>%summarise(NombreMed=n())%>%arrange(-NombreMed)%>%slice(1:3)%>%filter(sport=="figure skating")
```

```
## Source: local data frame [3 x 3]
## Groups: sport [1]
##
##      sport  code NombreMed
```

```
##           <chr> <chr>      <int>
## 1 figureskating  USA        45
## 2 figureskating  URS        24
## 3 figureskating  CAN        22
```

## 4 Evolution dans le temps

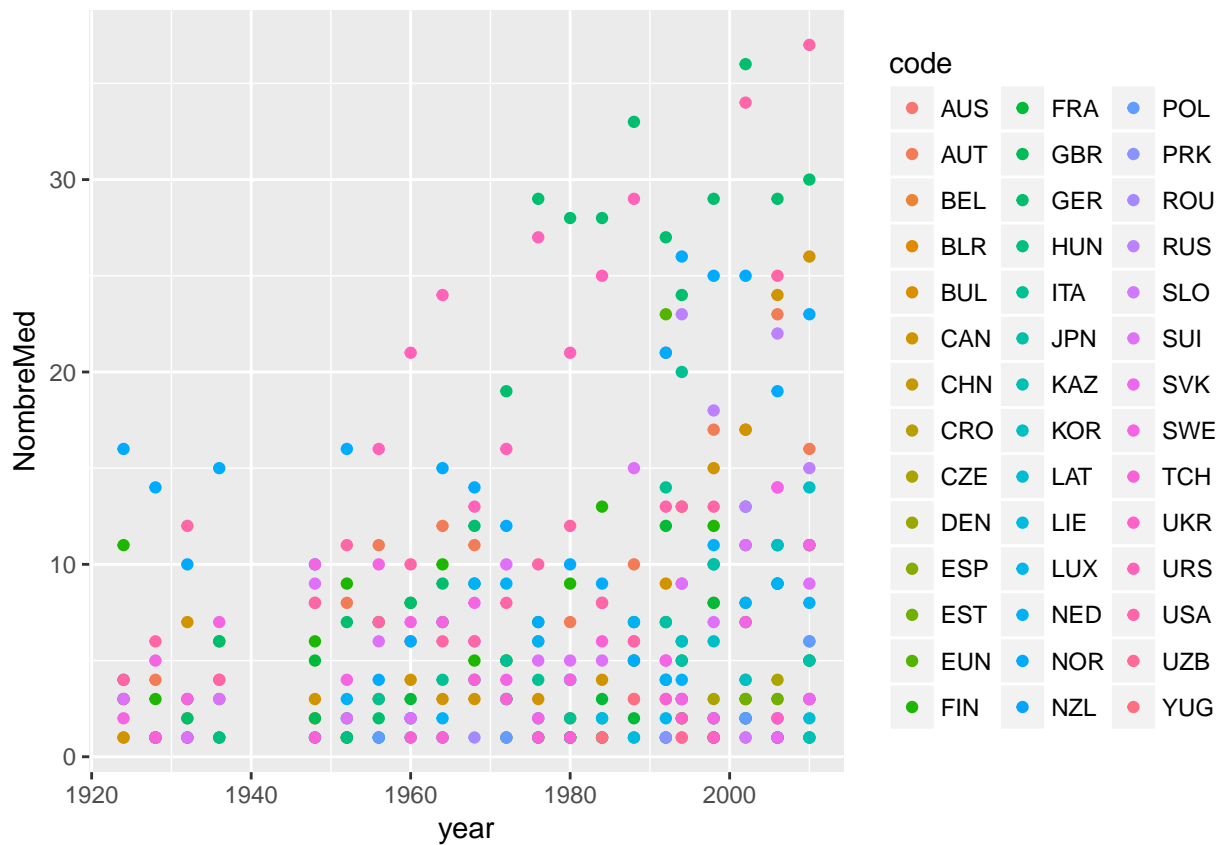
L'objectif de cette partie est de se concentrer sur l'évolution de l'obtention des médailles des pays. Pour commencer, nous avons choisi de ne conserver que les trois nations qui ont gagné le plus de médailles par année de JO.

```
donnees%>%group_by(year,code)%>%summarise(NombreMed=n())%>%arrange(-NombreMed)%>%slice(1:3)
```

```
## Source: local data frame [63 x 3]
## Groups: year [21]
##
##   year  code NombreMed
##   <int> <chr>      <int>
## 1  1924  NOR        16
## 2  1924  FIN        11
## 3  1924  GBR         4
## 4  1928  NOR        14
## 5  1928  USA         6
## 6  1928  SWE         5
## 7  1932  USA        12
## 8  1932  NOR        10
## 9  1932  CAN         7
## 10 1936  NOR        15
## # ... with 53 more rows
```

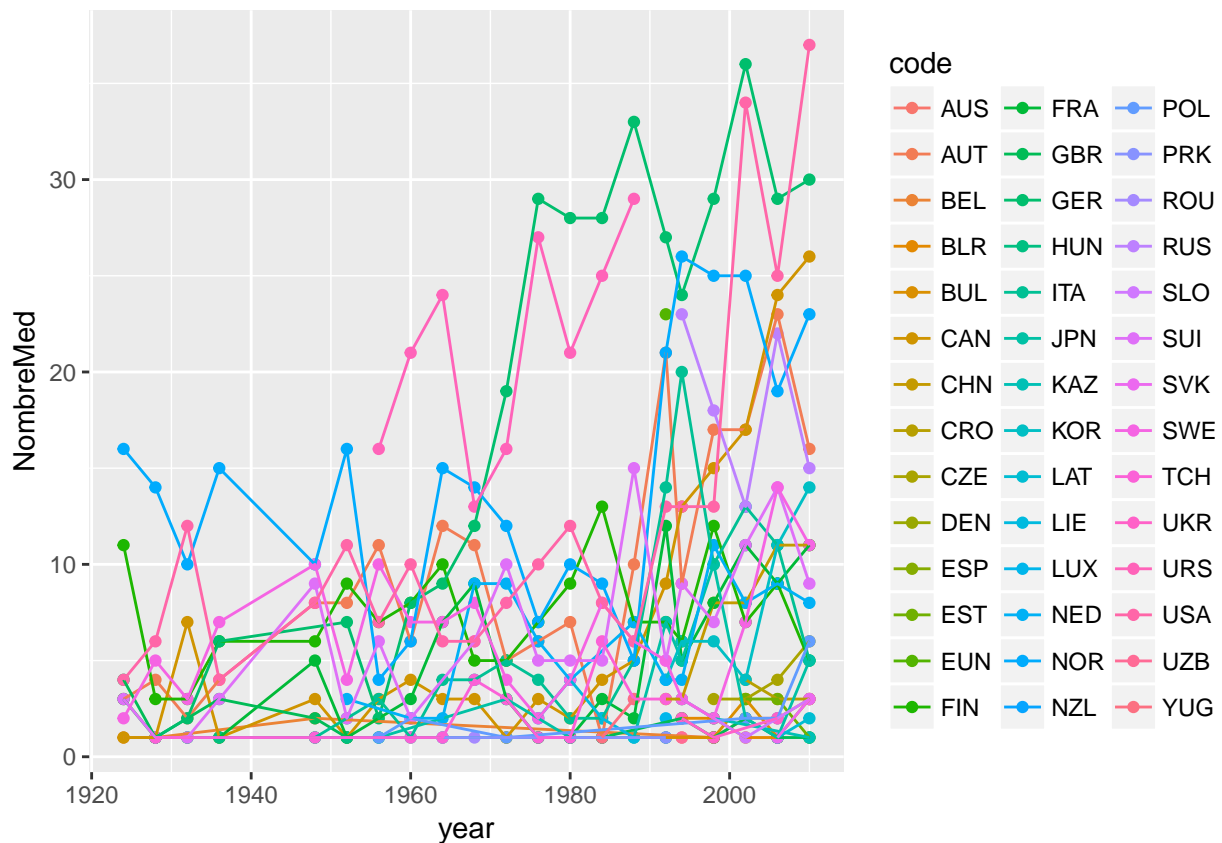
Ces données sont difficilement analysables et visualisables, nous les traiterons donc autrement. Nous avons donc décidé de construire un graphique qui affiche le nombre de médailles obtenues (toutes couleurs confondues) pour chaque nation et pour chaque JO.

```
donnees%>%group_by(year,code)%>%summarise(NombreMed=n())%>%ggplot(aes(x=year, y=NombreMed, color=code))
```



Ce graphique étant illisible, nous allons essayer de relier chaque année de chaque nation entre eux. Cependant il faut faire attention car cela fausse légèrement les données car ce ne sont pas des données continues (JO seulement tous les 4 ans). Toutefois, on peut bien voir que les JO de 1940 et 1944 n'ont pas eu lieu à cause de la 2ème guerre mondiale.

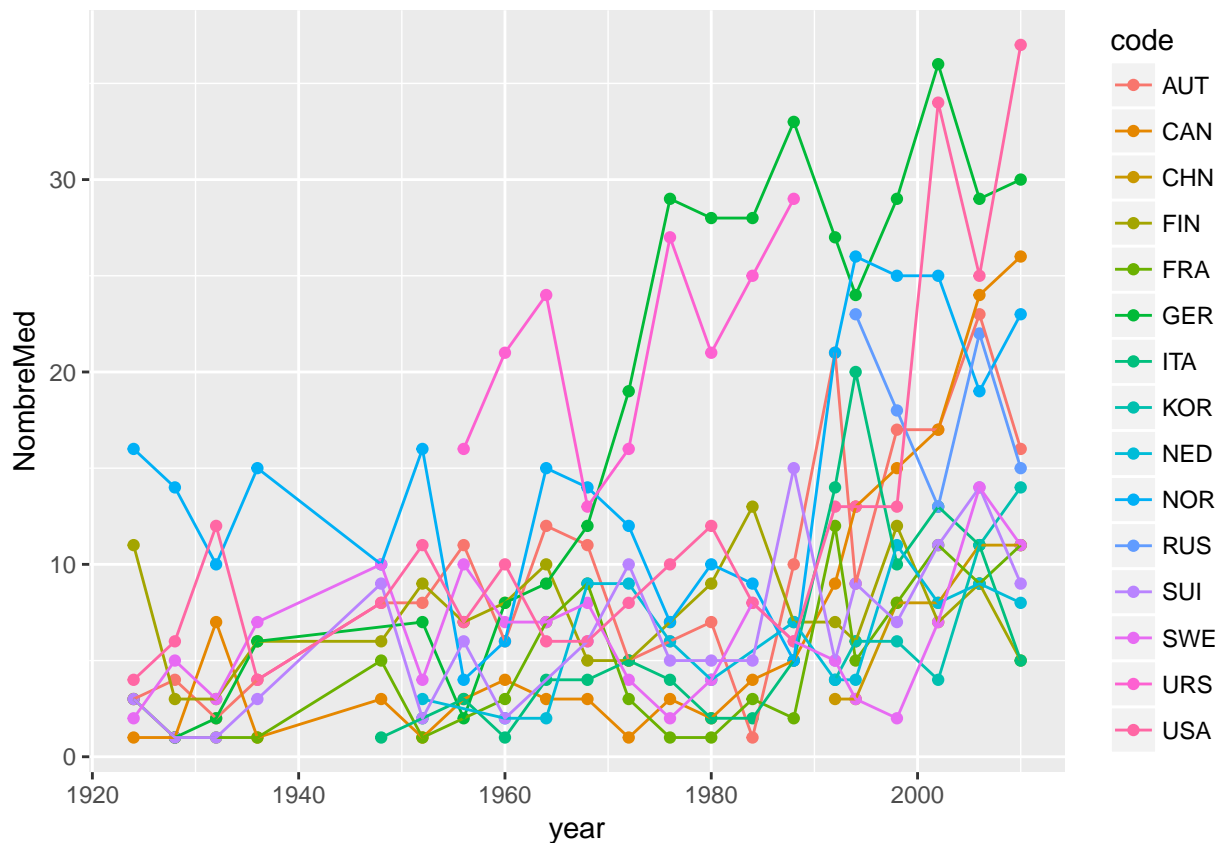
```
donnees%>%group_by(year,code)%>%summarise(NombreMed=n())%>%ggplot(aes(x=year, y=NombreMed, color=code))
```



Il est difficile de voir exactement quelle couleur correspond à quelle nation mais il reste possible de suivre l'évolution du nombre de médailles d'un pays bien précis (exemple : Norvège en bleu que l'on peut suivre de 1924 à 2010). Nous pouvons également observer que depuis les années 1960-1970, plusieurs nations obtiennent beaucoup de médailles : par exemple l'Allemagne a gagné trois fois plus.

Pour plus de lisibilité, nous allons afficher seulement les pays qui ont obtenu un nombre total de médailles sur l'ensemble de tous les jeux supérieurs à un certain nombre. Nous commençons avec 43 pour faire rentrer la Chine et la Corée du Sud.

```
minMed<-donnees%>%group_by(code)%>%summarise(NombreMed=n())%>%filter(NombreMed>43)
donnees%>%group_by(year,code)%>%summarise(NombreMed=n())%>%filter(code %in% minMed$code)%>%ggplot(aes(x=
```



Il est déjà beaucoup plus aisé de distinguer les pays. On peut ainsi en un rapide coup d'oeil voir la classement des “meilleurs” nations pour chaque JO. Par exemple, en 2010 le classement est : Etats Unis, Allemagne, Canada, Norvège, Autriche, Russie, Corée du Sud, Chine-France-Suède. . . On remarque aussi que de nouveaux pays emergent tels que le Canada, la Chine ou la Corée du Sud : ils ont réciproquement gagné seulement une quarentaire de médailles en tout mais 14 et 13 aux derniers JO. De plus, il faut rester vigilant car des pays n'existent plus comme l'URSS, ce que l'on peut facilement distinguer(rose).

Afin d'être certain de ne pas oublier de nouveaux pays émergents, nous allons sortir le classement des 20 premières nations des JO de 2006 et 2010.

```
donnees%>%filter(year==2006)%>%group_by(code)%>%summarise(NombreMed=n())%>%arrange(-NombreMed)%>%slice(
```

```
## # A tibble: 20 × 2
##   code NombreMed
##   <chr>      <int>
## 1   GER         29
## 2   USA         25
## 3   CAN         24
## 4   AUT         23
## 5   RUS         22
## 6   NOR         19
## 7   SUI         14
## 8   SWE         14
## 9   CHN         11
## 10  ITA         11
## 11  KOR         11
## 12  FIN          9
## 13  FRA          9
```

```
## 14 NED 9
## 15 CZE 4
## 16 CRO 3
## 17 EST 3
## 18 AUS 2
## 19 POL 2
## 20 UKR 2
```

```
donnees%>%filter(year==2010)%>%group_by(code)%>%summarise(NombreMed=n())%>%arrange(-NombreMed)%>%slice(
```

```
## # A tibble: 20 × 2
##   code NombreMed
##   <chr>      <int>
## 1 USA        37
## 2 GER        30
## 3 CAN        26
## 4 NOR        23
## 5 AUT        16
## 6 RUS        15
## 7 KOR        14
## 8 CHN        11
## 9 FRA        11
## 10 SWE        11
## 11 SUI         9
## 12 NED         8
## 13 CZE         6
## 14 POL         6
## 15 FIN         5
## 16 ITA         5
## 17 JPN         5
## 18 AUS         3
## 19 BLR         3
## 20 CRO         3
```

Il n'y a pas d'autres pays qui gagnent plus les deux derniers JO pour lesquels nous possédons des données sauf la République Tchèque (15ème en 2010). Cette dernière, résultante de l'ancienne Tchécoslovaquie, nous allons rapidement regarder le classement de cet ancien pays lors de ses derniers JO : son classement était du même ordre.

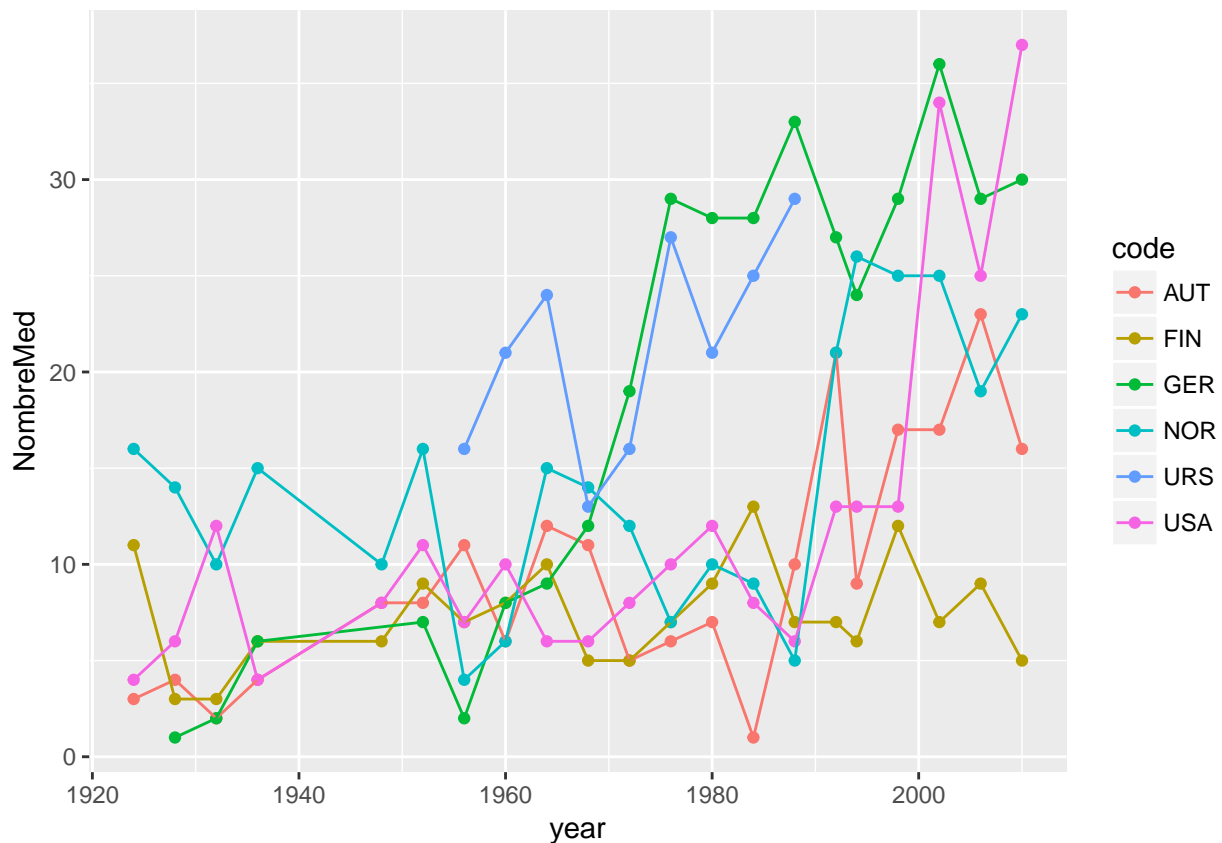
```
donnees%>%filter(year==1992)%>%group_by(code)%>%summarise(NombreMed=n())%>%arrange(-NombreMed)%>%mutate
```

```
## # A tibble: 1 × 3
##   code NombreMed ordre
##   <chr>      <int> <int>
## 1 TCH         3     16
```

Voici l'évolution du nombre de médailles pour les pays ayant obtenu plus de 150 médailles sur les 86 années.

```
minMed<-donnees%>%group_by(code)%>%summarise(NombreMed=n())%>%filter(NombreMed>150)
donnees%>%group_by(year,code)%>%summarise(NombreMed=n())%>%filter(code %in% minMed$code)%>%ggplot(aes(x=
```





On note que lors des derniers JO (2010), les Etats Unis ont dépassé l'Allemagne, première depuis les années 1970. Globalement, ces pays gagnent plus de médailles de nos jours qu'avant les années 1970.

Grâce aux graphiques nous avons pu voir que les pays gagnants le plus de médailles à chaque JO sont globalement les mêmes (Allemagne -détrônée en 2010-, Norvège...). Cependant nous commençons à voir de nouvelles nations qui atteignent le sommet à leur rythme (Chine, Corée du Sud, Canada...). Parmi ces derniers, se trouvent les Etats Unis qui se sont grandement améliorés depuis les JO des années 1990 et qui prennent finalement la tête du classement.

## 5 Position de la France

Cette partie est une partie légèrement à part qui va traiter le cas de la France. Etant curieuses de savoir à quel niveau se situe la France, nous avons décidé d'y consacrer quelques lignes.

Nous savons déjà que sur l'ensemble des jeux la France ne fait pas partie des 10 nations à avoir gagné le plus de médailles sauf si on ne regarde que les médailles de bronze où elle se classe à la 10ème place.

```
donnees %>% group_by(code) %>% summarise(NombreMed = n()) %>% arrange(-NombreMed) %>% mutate(ordre = 1:42) %>% filter(ordre == 11)
```

```
## # A tibble: 1 × 3
##   code NombreMed ordre
##   <chr>      <int> <int>
## 1  FRA         97     11
```

Dans le classement général la France est 11ème sur 42 avec un total de 97 médailles (or, argent et bronze) en 86 ans.

Voici le classement de la France pour chaque JO et le nombre de médailles obtenues:

```
donnees%>%group_by(year,code)%>%summarise(NombreMed=n())%>%mutate(ordre=1:length(NombreMed))%>%filter(c
```

```
## Source: local data frame [20 x 4]
## Groups: year [20]
##
##   year  code NombreMed ordre
##   <int> <chr>      <int> <int>
## 1  1924  FRA         3     5
## 2  1928  FRA         1     5
## 3  1932  FRA         1     4
## 4  1936  FRA         1     4
## 5  1948  FRA         5     5
## 6  1952  FRA         1     4
## 7  1960  FRA         3     4
## 8  1964  FRA         7     4
## 9  1968  FRA         9     4
## 10 1972  FRA         3     5
## 11 1976  FRA         1     4
## 12 1980  FRA         1     5
## 13 1984  FRA         3     4
## 14 1988  FRA         2     4
## 15 1992  FRA        12     7
## 16 1994  FRA         5     7
## 17 1998  FRA         8    11
## 18 2002  FRA        11    11
## 19 2006  FRA         9    11
## 20 2010  FRA        11    10
```

La France se maintenait vers la 4ème ou 5ème place jusqu'en 1992 où elle est descendue jusqu'à la 7ème place et est à la limite du top 10 à partir de 1998.

On va maintenant regarder les sports dans lesquels la France est placée numéro un en terme de nombre de médaille durant un JO.

```
donnees%>%group_by(year,sport,code)%>%summarise(NombreMed=n())%>%arrange(-NombreMed)%>%slice(1) %>%filt
```

```
## Source: local data frame [4 x 4]
## Groups: year, sport [4]
##
##   year      sport  code NombreMed
##   <int>      <chr> <chr>      <int>
## 1  1924      curling  FRA         1
## 2  1968    alpinesskiing  FRA         8
## 3  1992 freestyleskiing  FRA         5
## 4  2010      biathlon  FRA         6
```

Ainsi, quatres sports différents ressortent et comme nous nous y attendions, la France gagne beaucoup au biathlon en 2010.

Voici la liste des sports (avec les années de JO associées) pour lesquels la France fait partie du top 3 des nations qui ont gagné le plus de médailles en général.

```
donnees%>%group_by(year,sport,code)%>%summarise(NombreMed=n())%>%arrange(-NombreMed)%>%slice(1:3) %>%fi
```

```
## Source: local data frame [22 x 4]
## Groups: year, sport [22]
##
```

```
##      year      sport code NombreMed
##      <int>      <chr> <chr>      <int>
## 1    1924      biathlon   FRA         1
## 2    1924      curling    FRA         1
## 3    1924 figureskating   FRA         1
## 4    1928 figureskating   FRA         1
## 5    1936 alpinesskiing   FRA         1
## 6    1948 alpinesskiing   FRA         5
## 7    1952 figureskating   FRA         1
## 8    1960 alpinesskiing   FRA         3
## 9    1964 alpinesskiing   FRA         6
## 10   1968 alpinesskiing   FRA         8
## # ... with 12 more rows

donnees%>%group_by(year,sport,code)%>%summarise(NombreMed=n())%>%arrange(-NombreMed)%>%slice(1:3) %>%fi

## Source: local data frame [22 x 4]
## Groups: year, sport [22]
##
##      year      sport code NombreMed
##      <int>      <chr> <chr>      <int>
## 1    2010      biathlon   FRA         6
## 2    2010 nordiccombined   FRA         1
## 3    2010      snowboard   FRA         3
## 4    2002      alpinesskiing   FRA         4
## 5    2002      snowboard   FRA         3
## 6    1998      bobsleigh    FRA         1
## 7    1998 figureskating   FRA         2
## 8    1998 nordiccombined   FRA         1
## 9    1994      biathlon   FRA         3
## 10   1992 freestyleskiing   FRA         5
## # ... with 12 more rows
```

On compte donc 22 sports et années pour lesquelles la France fait partie des trois “meilleures” nations mais nous n’affichons que les extrêmes c’est à dire les informations provenant des plus anciennes données et les plus récentes.

Nous pouvons donc en conclure que la France est bien classée dans le classement de l’ensemble des pays et souvent dans certains sports (par exemple en biathlon, saut alpin ‘alpinesskiing’, patinage artistique ‘figureskating’, snowboard...). Cependant nous regrettons de ne pas posséder les données des derniers JO, à savoir ceux de Sotchi en 2014.

## 6 Conclusion

Grâce à ces données nous avons pu voir que les pays qui gagnent le plus de médailles évoluent peu entre les années 1924 et 2010 : les 10 premiers pays restent globalement constants (à quelques rangs près), par exemple l’Allemagne et la Norvège. Cependant on peut noter l’apparition de nouveaux pays qui ont récemment grandement augmenté leur nombre de médailles (Etats Unis, Canada, Chine...). Pour ce qui est de la France, on peut dire qu’elle est globalement assez bien classée, même si elle s’est faite dépassée par plusieurs pays depuis les années 1990, ne se situant plus forcément dans le top 10 des nations qui gagnent le plus de médailles.

Ce projet nous a permis d’étudier des données qui nous intéressait mais également de mieux prendre en main Rstudio et différents packages tels que knitr, dplyr, etc.

## 7 Annexe : Correspondance code et nom de nations

```
donnees%>%select(code, country)%>%unique()%>%arrange(code)%>%slice(1:10)
```

```
## # A tibble: 10 × 2
##   code          country
##   <chr>         <chr>
## 1 AUS      Australia
## 2 AUT      Austria
## 3 BEL      Belgium
## 4 BLR      Belarus
## 5 BUL      Bulgaria
## 6 CAN      Canada
## 7 CHN People'sRepublicofChina
## 8 CRO      Croatia
## 9 CZE      CzechRepublic
## 10 DEN      Denmark
```

```
donnees%>%select(code, country)%>%unique()%>%arrange(code)%>%slice(11:20)
```

```
## # A tibble: 10 × 2
##   code          country
##   <chr>         <chr>
## 1 ESP      Spain
## 2 EST      Estonia
## 3 EUN      UnifiedTeam(exUSSR)
## 4 FIN      Finland
## 5 FRA      France
## 6 GBR      GreatBritain
## 7 GER      Germany
## 8 GER FederalRepublicofGermany(1950-1990)
## 9 GER      UnitedTeamofGermany(1956-1960)
## 10 GER GermanDemocraticRepublic(1955-1990)
```

```
donnees%>%select(code, country)%>%unique()%>%arrange(code)%>%slice(21:30)
```

```
## # A tibble: 10 × 2
##   code          country
##   <chr>         <chr>
## 1 HUN      Hungary
## 2 ITA      Italy
## 3 JPN      Japan
## 4 KAZ      Kazakhstan
## 5 KOR RepublicofKorea
## 6 LAT      Latvia
## 7 LIE      Liechtenstein
## 8 LUX      Luxembourg
## 9 NED      Netherlands
## 10 NOR      Norway
```

```
donnees%>%select(code, country)%>%unique()%>%arrange(code)%>%slice(31:40)
```

```
## # A tibble: 10 × 2
##   code          country
##   <chr>         <chr>
```

```
## 1    NZL                NewZealand
## 2    POL                Poland
## 3    PRK DemocraticPeople'sRepublicofKorea
## 4    ROU                Romania
## 5    RUS                RussianFederation
## 6    SLO                Slovenia
## 7    SUI                Switzerland
## 8    SVK                Slovakia
## 9    SWE                Sweden
## 10   TCH                Czechoslovakia
```

```
donnees%>%select(code,country)%>%unique()%>%arrange(code)%>%slice(41:50)
```

```
## # A tibble: 5 × 2
##   code      country
##   <chr>    <chr>
## 1   UKR      Ukraine
## 2   URS      USSR
## 3   USA UnitedStatesofAmerica
## 4   UZB      Uzbekistan
## 5   YUG      Yugoslavia
```