

# Ejercicios - Data Frame R

Ramon Ceballos

23/1/2021

Carga en R el CSV de la web <http://winterolympicsmedals.com/medals.csv> y responde a las preguntas de la tarea.

```
df_medals = read.csv("http://winterolympicsmedals.com/medals.csv", header=TRUE, stringsAsFactors = FALSE)
str(df_medals)
```

```
## 'data.frame':    2311 obs. of  8 variables:
## $ Year          : int  1924 1924 1924 1924 1924 1924 1924 1924 1924 1924 ...
## $ City          : chr  "Chamonix" "Chamonix" "Chamonix" "Chamonix" ...
## $ Sport         : chr  "Skating" "Skating" "Skating" "Bobsleigh" ...
## $ Discipline    : chr  "Figure skating" "Figure skating" "Figure skating" "Bobsleigh" ...
## $ NOC           : chr  "AUT" "AUT" "AUT" "BEL" ...
## $ Event         : chr  "individual" "individual" "pairs" "four-man" ...
## $ Event.gender  : chr  "M" "W" "X" "M" ...
## $ Medal         : chr  "Silver" "Gold" "Gold" "Bronze" ...
```

```
head(df_medals,10)
```

```
##   Year   City      Sport   Discipline NOC      Event Event.gender
## 1  1924 Chamonix   Skating Figure skating AUT    individual      M
## 2  1924 Chamonix   Skating Figure skating AUT    individual      W
## 3  1924 Chamonix   Skating Figure skating AUT      pairs      X
## 4  1924 Chamonix Bobsleigh   Bobsleigh BEL    four-man      M
## 5  1924 Chamonix Ice Hockey   Ice Hockey CAN    ice hockey      M
## 6  1924 Chamonix Biathlon     Biathlon FIN military patrol      M
## 7  1924 Chamonix   Skating Figure skating FIN      pairs      X
## 8  1924 Chamonix   Skating Speed skating FIN    10000m      M
## 9  1924 Chamonix   Skating Speed skating FIN    10000m      M
## 10 1924 Chamonix   Skating Speed skating FIN    1500m      M
##   Medal
## 1 Silver
## 2   Gold
## 3   Gold
## 4 Bronze
## 5   Gold
## 6 Silver
## 7 Silver
## 8   Gold
## 9 Silver
## 10  Gold
```

## Preguntas de esta tarea

1.¿Cuántos deportistas aparecen en el data frame?

```
#Para saber el número de deportistas tenemos que averiguar el nº de filas que hay  
deportistas = nrow(df_medals)  
deportistas
```

```
## [1] 2311
```

2.¿Cuántos han ganado medallas de oro, cuántos de plata y cuántos de bronce?

```
#Selecciono las medallas de oro  
gold = nrow(subset(df_medals, Medal == "Gold", select = "Medal"))  
gold
```

```
## [1] 774
```

```
#Selecciono las medallas de plata  
silver = nrow(subset(df_medals, Medal == "Silver", select = "Medal"))  
silver
```

```
## [1] 773
```

```
#Selecciono las medallas de bronce  
bronze = nrow(subset(df_medals, Medal == "Bronze", select = "Medal"))  
bronze
```

```
## [1] 764
```

3.¿En cuántos lugares se han hecho olimpiadas de invierno?

```
#Convierto a factor la columna de las ciudades y cuento cuanto niveles tiene dicho factor  
df_medals$City = as.factor(df_medals$City)  
cities = length(levels(df_medals$City))  
cities
```

```
## [1] 17
```

4.¿Cuántos hombres y cuántas mujeres hay?

```
#Calculo los hombres  
hombres = nrow(subset(df_medals, Event.gender == "M", select = "Event.gender"))  
hombres
```

```
## [1] 1386
```

```
#Calculo las mujeres
mujeres = nrow(subset(df_medals, Event.gender == "W", select = "Event.gender"))
mujeres
```

```
## [1] 802
```

## 5.¿En qué año participaron más deportistas?

```
athletes_year = table(df_medals$Year)
athletes_year
```

```
##
## 1924 1928 1932 1936 1948 1952 1956 1960 1964 1968 1972 1976 1980 1984 1988 1992
##   49   41   42   51   68   67   72   81  103  106  105  111  115  117  138  171
## 1994 1998 2002 2006
##  183  205  234  252
```

```
#El año que más atletas participaron fue el año 2006 con 252 atletas.
```

## 6.El campo NOC indica el país del ganador de la medalla. ¿Qué país puede presumir de haber ganado más medallas de oro en los juegos de invierno entre 1960 y 1996?

```
df_medals$Year=as.integer(df_medals$Year)
```

```
#Selecciono una submatriz de los años comprendidos entre 1960 y 1996, con los atletas que ganaron medal
df2 = subset(df_medals, 1960 <= Year & Year <= 1996 & Medal=="Gold")
```

```
#Introduzco una nueva variable que le de el valor 1 a cada medalla de oro
str(df2)
```

```
## 'data.frame':   411 obs. of  8 variables:
##  $ Year      : int   1960 1960 1960 1960 1960 1960 1960 1960 1960 1960 ...
##  $ City      : Factor w/ 17 levels "Albertville",...: 15 15 15 15 15 15 15 15 15 15 ...
##  $ Sport     : chr   "Skiing" "Skating" "Skiing" "Skating" ...
##  $ Discipline : chr   "Alpine Skiing" "Figure skating" "Alpine Skiing" "Speed skating" ...
##  $ NOC       : chr   "AUT" "CAN" "CAN" "EUA" ...
##  $ Event     : chr   "slalom" "pairs" "slalom" "500m" ...
##  $ Event.gender: chr   "M" "X" "W" "W" ...
##  $ Medal     : chr   "Gold" "Gold" "Gold" "Gold" ...
```

```
new_column = c(rep(1,411))
df2$Number_gold = new_column
str(df2)
```

```
## 'data.frame':   411 obs. of  9 variables:
##  $ Year      : int   1960 1960 1960 1960 1960 1960 1960 1960 1960 1960 ...
```

```
## $ City      : Factor w/ 17 levels "Albertville",...: 15 15 15 15 15 15 15 15 15 15 ...
## $ Sport     : chr  "Skiing" "Skating" "Skiing" "Skating" ...
## $ Discipline : chr  "Alpine Skiing" "Figure skating" "Alpine Skiing" "Speed skating" ...
## $ NOC       : chr  "AUT" "CAN" "CAN" "EUA" ...
## $ Event     : chr  "slalom" "pairs" "slalom" "500m" ...
## $ Event.gender: chr  "M" "X" "W" "W" ...
## $ Medal     : chr  "Gold" "Gold" "Gold" "Gold" ...
## $ Number_gold : num  1 1 1 1 1 1 1 1 1 1 ...
```

*#Medallas de oro obtenidas por cada país*

```
Gold_NOC= aggregate(Number_gold~NOC, data =df2, FUN = sum)
Gold_NOC
```

```
##      NOC Number_gold
## 1  AUT          25
## 2  CAN          12
## 3  ESP           1
## 4  EUA           7
## 5  EUN           9
## 6  FIN          20
## 7  FRA          12
## 8  FRG          11
## 9  GBR           4
## 10 GDR          39
## 11 GER          19
## 12 ITA          22
## 13 JPN           3
## 14 KAZ           1
## 15 KOR           6
## 16 LIE           2
## 17 NED          14
## 18 NOR          40
## 19 POL           1
## 20 RUS          11
## 21 SUI          19
## 22 SWE          24
## 23 TCH           2
## 24 UKR           1
## 25 URS          71
## 26 USA          34
## 27 UZB           1
```

```
Max_Gold_NOC =subset(Gold_NOC, Number_gold == max(Gold_NOC$Number_gold), select = "NOC")
Max_Gold_NOC
```

```
##      NOC
## 25 URS
```

*#La URS es el país con más medallas de oro entre los años 1960 y 1996*