

Manejando Data Frames con R

Yimmy Eman

2022-07-01

```
df <- read.csv("../data/medals.csv", header = T)
head(df,5)
```

```
##   X Year      City      Sport      Discipline NOC      Event Event.gender Medal
## 1 1 1924 Chamonix      Skating Figure skating AUT individual          M Silver
## 2 2 1924 Chamonix      Skating Figure skating AUT individual          W   Gold
## 3 3 1924 Chamonix      Skating Figure skating AUT      pairs          X   Gold
## 4 4 1924 Chamonix Bobsleigh      Bobsleigh BEL    four-man          M Bronze
## 5 5 1924 Chamonix Ice Hockey      Ice Hockey CAN    ice hockey          M   Gold
```

```
names(df)
```

```
## [1] "X"           "Year"        "City"        "Sport"       "Discipline"
## [6] "NOC"        "Event"       "Event.gender" "Medal"
```

```
str(df)
```

```
## 'data.frame':    2311 obs. of  9 variables:
## $ X              : int  1 2 3 4 5 6 7 8 9 10 ...
## $ 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" ...
```

Pregunta 1

¿Cuántos deportistas aparecen en el data frame?

```
length(df$X)
```

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

Pregunta 2

¿Cuántos deportistas han ganado medallas de oro? ¿Cuántos han ganado medallas de plata? ¿Y cuántos de bronce?

```
gold_meadals <- length(which(df$Medal == "Gold"))
gold_meadals
```

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

```
silver_medals <- length(which(df$Medal == "Silver"))
silver_medals
```

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

Pregunta 3

¿En cuántos lugares se han hecho olimpiadas de invierno?

```
df$City <- as.factor(df$City)
length(levels(df$City))
```

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

Pregunta 4

¿Cuántos hombres y cuántas mujeres hay?

```
woman_count <- length(df$Event.gender[which(df$Event.gender == "W")])
woman_count
```

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

```
men_count <- length(df$Event.gender[which(df$Event.gender == "M")])
men_count
```

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

Pregunta 5

¿En qué año participaron más deportistas?

```
as.numeric(names(which.max(table(df$Year))))
```

```
## [1] 2006
```

Pregunta 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?

```
df2 <- df[df$Medal == "Gold", ]  
df2 <- df2[which(df2$Year >= 1960 & df2$Year <= 1996), ]  
names(which.max(table(df2[, "NOC"])))
```

```
## [1] "URS"
```