

# Tarea 8: Manejando Data Frames con R

Jhan Kevin

5/5/2020

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

```
medals <- read.csv('../tareass/medals.csv')
str(medals)
```

```
## 'data.frame':    2311 obs. of  8 variables:
## $ Year           : int  1924 1924 1924 1924 1924 1924 1924 1924 1924 1924 ...
## $ City           : Factor w/ 17 levels "Albertville",...: 3 3 3 3 3 3 3 3 3 ...
## $ Sport          : Factor w/ 7 levels "Biathlon","Bobsleigh",...: 6 6 6 2 4 1 6 6 6 ...
## $ Discipline     : Factor w/ 15 levels "Alpine Skiing",...: 6 6 6 3 8 2 6 15 15 ...
## $ NOC            : Factor w/ 45 levels "AUS","AUT","BEL",...: 2 2 2 3 6 15 15 15 15 ...
## $ Event          : Factor w/ 67 levels "10000m","1000m",...: 48 48 56 41 47 54 56 1 1 7 ...
## $ Event.gender   : Factor w/ 3 levels "M","W","X": 1 2 3 1 1 1 3 1 1 ...
## $ Medal          : Factor w/ 3 levels "Bronze","Gold",...: 3 2 2 1 2 3 3 2 3 ...
```

El data frame contiene información sobre las medallas ganadas en los juegos olímpicos de invierno.

## ¿Cuántos deportistas aparecen en el data frame?

Los datos no discriminan por deportista, no es posible establecer si hay deportistas repeditos, por lo cual se asume que cada medalla fue ganada por un deportista diferente. Así el número de filas será el número de deportistas.

```
n_dep <- nrow(medals)
```

Aparecen 2311 deportistas.

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

```
# Usando subset
n_gold <- dim(subset(medals, Medal == "Gold"))[1]

# Otra manera
silver <- medals[medals$Medal=='Silver',]
n_silver <- nrow(silver)

bronze <- medals[medals$Medal=='Bronze',]
n_bronze <- nrow(bronze)
```

- 774 han ganado medallas de oro.
- 773 han ganado medallas de plata.
- 764 han ganado medallas de bronce.

### ¿En cuantos lugares se han hecho olimpiadas de invierno?

El data frame contiene una columna de ciudades, almacenada como factor. Se debe contar la cantidad de niveles.

```
n_city <- length(unique(medals$City))
```

En 17 ciudades.

### ¿Cuantos hombres y cuantas mujeres hay?

El df contiene una columna que clasifica los deportistas por género.

```
n_men <- dim(subset(medals, Event.gender == "M"))[1]
n_women <- dim(subset(medals, Event.gender == "W"))[1]
n_ind <- dim(subset(medals, Event.gender == "X"))[1]
```

Se tiene registro de 1386 hombres, 802 mujeres y 123 indecisos.

### ¿En qué año participaron más deportistas?

```
max_year <- max(aggregate(cbind(Year)~Year,
                           data = medals,
                           function(x){max(NROW(x))}
))
```

En el año 2006

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?

```
med_1960_1996 <- subset(medals, Year <= '1996' & Year >= '1960' & Medal == 'Gold')
med_1960_1996 <- table(med_1960_1996$NOC)
med_1960_1996[med_1960_1996 == max(med_1960_1996)]
```

```
## URS
## 71
```