

Tare 8 - Manejando DF en R

Brigido Vicuna

8/14/2020

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

```
read.csv("http://winterolympicsmedals.com/medals.csv", header = TRUE, stringsAsFactors = TRUE) -> df
```

Questions for this assignment ¿Cuántos deportistas aparecen en el data frame?

```
df_Dep = aggregate(df$Medal, by = df["Year"], length)

str(df_Dep)
```

```
## 'data.frame':    20 obs. of  2 variables:
##  $ Year: int   1924 1928 1932 1936 1948 1952 1956 1960 1964 1968 ...
##  $ x   : int   49 41 42 51 68 67 72 81 103 106 ...
```

```
df_Dep = df_Dep[with(df_Dep, order(-df_Dep$Year)), ]
df_Dep
```

```
##      Year    x
## 20 2006 252
## 19 2002 234
## 18 1998 205
## 17 1994 183
## 16 1992 171
## 15 1988 138
## 14 1984 117
## 13 1980 115
## 12 1976 111
## 11 1972 105
## 10 1968 106
## 9  1964 103
## 8  1960  81
## 7  1956  72
## 6  1952  67
## 5  1948  68
## 4  1936  51
## 3  1932  42
## 2  1928  41
## 1  1924  49
```

```
sapply(df_Dep$x, cumsum)
```

```
## [1] 252 234 205 183 171 138 117 115 111 105 106 103 81 72 67 68 51 42 41
## [20] 49
```

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

```
attach(df)
gold = sum(df$Medal=="Gold")
silver = sum(df$Medal=="Silver")
bronze = sum(df$Medal=="Bronze")
gold
```

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

```
silver
```

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

```
bronze
```

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

```
total = gold+silver+bronze
total
```

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

¿En cuántos lugares se han hecho olimpiadas de invierno?

```
aggregate(df$Medal=="Gold", by = df["City"], length)
```

```
##           City    x
## 1    Albertville 171
## 2      Calgary 138
## 3    Chamonix   49
## 4  Cortina d'Ampezzo 72
## 5 Garmisch-Partenkirchen 51
## 6    Grenoble 106
## 7    Innsbruck 214
## 8    Lake Placid 157
## 9    Lillehammer 183
## 10    Nagano 205
## 11     Oslo    67
## 12  Salt Lake City 234
## 13    Sapporo 105
## 14    Sarajevo 117
## 15  Squaw Valley 81
## 16    St. Moritz 109
## 17     Turin 252
```

¿Cuántos hombres y cuántas mujeres hay?

```
x = sum(df$Event.gender=="X")
```

```
men = sum(df$Event.gender=="M") + x  
men
```

```
## [1] 1509
```

```
women = sum(df$Event.gender=="W")+ x  
women
```

```
## [1] 925
```

¿En qué año participaron más deportistas?

```
df_MedalYear = aggregate(df$Medal, by = df["Year"], length)  
  
str(df_MedalYear)
```

```
## 'data.frame':    20 obs. of  2 variables:  
##  $ Year: int  1924 1928 1932 1936 1948 1952 1956 1960 1964 1968 ...  
##  $ x   : int  49 41 42 51 68 67 72 81 103 106 ...
```

```
df_MedalYear = df_MedalYear[with(df_MedalYear, order(-df_MedalYear$Year)), ]  
df_MedalYear
```

```
##   Year  x  
## 20 2006 252  
## 19 2002 234  
## 18 1998 205  
## 17 1994 183  
## 16 1992 171  
## 15 1988 138  
## 14 1984 117  
## 13 1980 115  
## 12 1976 111  
## 11 1972 105  
## 10 1968 106  
## 9  1964 103  
## 8  1960 81  
## 7  1956 72  
## 6  1952 67  
## 5  1948 68  
## 4  1936 51  
## 3  1932 42  
## 2  1928 41  
## 1  1924 49
```

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_MedalNOC = df[df$Year<=1996 & df$Year>=1960, ]
tail(df_MedalNOC)
```

```
##      Year      City Sport  Discipline NOC   Event Event.gender Medal
## 1615 1994 Lillehammer Skiing  Alpine Skiing USA downhill      M   Gold
## 1616 1994 Lillehammer Skiing  Alpine Skiing USA downhill      W Silver
## 1617 1994 Lillehammer Skiing  Alpine Skiing USA  super-G      M Silver
## 1618 1994 Lillehammer Skiing  Alpine Skiing USA  super-G      W   Gold
## 1619 1994 Lillehammer Skiing Freestyle Ski. USA  moguls      W Silver
## 1620 1994 Lillehammer Skiing Freestyle Ski. UZB  aerals      W   Gold
```

```
df_MedalNOC = aggregate(df$Medal, by = df["NOC"], length)
df_MedalNOC
```

```
##      NOC      x
## 1  AUS      6
## 2  AUT    185
## 3  BEL      5
## 4  BLR      6
## 5  BUL      6
## 6  CAN    119
## 7  CHN     33
## 8  CRO      7
## 9  CZE     10
## 10 DEN      1
## 11 ESP      2
## 12 EST      6
## 13 EUA     19
## 14 EUN     23
## 15 FIN    151
## 16 FRA     83
## 17 FRG     41
## 18 GBR     21
## 19 GDR    110
## 20 GER    158
## 21 HUN      6
## 22 ITA    101
## 23 JPN     32
## 24 KAZ      5
## 25 KOR     31
## 26 LAT      1
## 27 LIE      9
## 28 LUX      2
## 29 NED     78
## 30 NOR    280
## 31 NZL      1
## 32 POL      8
## 33 PRK      2
## 34 ROU      1
## 35 RUS     76
## 36 SLO      4
## 37 SUI    118
## 38 SVK      1
```

```
## 39 SWE 118
## 40 TCH 25
## 41 UKR 5
## 42 URS 194
## 43 USA 216
## 44 UZB 1
## 45 YUG 4
```

```
df_MedalNOC = df_MedalNOC[with(df_MedalNOC, order(-df_MedalNOC$x)), ]
df_MedalNOC
```

```
##      NOC      x
## 30 NOR 280
## 43 USA 216
## 42 URS 194
## 2  AUT 185
## 20 GER 158
## 15 FIN 151
## 6  CAN 119
## 37 SUI 118
## 39 SWE 118
## 19 GDR 110
## 22 ITA 101
## 16 FRA 83
## 29 NED 78
## 35 RUS 76
## 17 FRG 41
## 7  CHN 33
## 23 JPN 32
## 25 KOR 31
## 40 TCH 25
## 14 EUN 23
## 18 GBR 21
## 13 EUA 19
## 9  CZE 10
## 27 LIE 9
## 32 POL 8
## 8  CRO 7
## 1  AUS 6
## 4  BLR 6
## 5  BUL 6
## 12 EST 6
## 21 HUN 6
## 3  BEL 5
## 24 KAZ 5
## 41 UKR 5
## 36 SLO 4
## 45 YUG 4
## 11 ESP 2
## 28 LUX 2
## 33 PRK 2
## 10 DEN 1
## 26 LAT 1
## 31 NZL 1
```

##	34	ROU	1
##	38	SVK	1
##	44	UZB	1