

# Postwork6

Orlando Antonio Aguilar Puerto      Mario Alberto Encinas Cardona  
Emanuel Flores Martínez      Andrés Benjamín Sánchez Alvarado

23/01/2021

## Desarrollo

Importa el conjunto de datos `match.data.csv` a R y realiza lo siguiente: Librerías

```
library(knitr)
library(ggplot2)
library(dplyr)
library(ggfortify)
```

```
df <- read.csv("https://raw.githubusercontent.com/beduExpert/Programacion-con-R-Santander/master/Sesion
```

```
df <- mutate(df, sumagoles = home.score + away.score)
df <- mutate(df, date = as.Date(date, "%Y-%m-%d"))
```

Agrega una nueva columna `sumagoles` que contenga la suma de goles por partido.

date	home.team	home.score	away.team	away.score	sumagoles
2010-08-28	Hercules	0	Ath Bilbao	1	1
2010-08-28	Levante	1	Sevilla	4	5
2010-08-28	Malaga	1	Valencia	3	4
2010-08-29	Espanol	3	Getafe	1	4
2010-08-29	La Coruna	0	Zaragoza	0	0
2010-08-29	Mallorca	0	Real Madrid	0	0

**Obtén el promedio por mes de la suma de goles.** Se crea otro dataframe con el rango completo de los datos de `match.data.csv`, incluyendo aquellos meses en los que no hay partidos (junio y julio).

```
date.full <- data.frame( date = seq(min(df$date), max(df$date),
                                   by = "1 month"))
```

Se restan los caracteres de los días en los observables de la variable “date”. Esto convierte a los valores en strings.

```
df$date <- substr(df$date, 1, 7)
date.full$date <- substr(date.full$date, 1, 7)
```

Se agrupan los datos:

```
promedios <- df %>%
  select(date, sumagoles) %>%
  group_by(date) %>%
  summarise(Promedio = mean(sumagoles))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

Table 2: Promedio de goles por partido agrupados por mes

date	Promedio
2010-08	2.200000
2010-09	2.425000
2010-10	3.025641
2010-11	2.902439
2010-12	2.733333
2011-01	3.000000

**Crea la serie de tiempo del promedio por mes de la suma de goles hasta diciembre de 2019.**  
Se mezclan los dataframes de los promedios agrupados y las fechas completas, de tal forma que se generan renglones para los meses faltantes (junio y julio) con valores de promedio NA

```
promedios2 <- merge(date.full, promedios, by = "date", all.x = TRUE)
ts.prom <- ts(promedios2$Promedio, start = c(2010,8), end = c(2019,12), frequency = 12)
```

```
autoplot(ts.prom, ts.colour = "blue1", ts.linetype = "solid") + labs(title = "Promedio de goles por partido")
```

Grafica la serie de tiempo.

### Promedio de goles por partido

