

# Postwork - Sesión 2

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

*11/2021*

1.Importa los datos de soccer de las temporadas 2017/2018, 2018/2019 y 2019/2020 de la primera división de la liga española a R, los datos los puedes encontrar en el siguiente enlace: <https://www.football-data.co.uk/spainm.php>

```
df.1718 <- read.csv('https://www.football-data.co.uk/mmz4281/1718/SP1.csv')
df.1819 <- read.csv('https://www.football-data.co.uk/mmz4281/1819/SP1.csv')
df.1920 <- read.csv('https://www.football-data.co.uk/mmz4281/1920/SP1.csv')
```

2.Obten una mejor idea de las características de los data frames al usar las funciones: str, head, View y summary

```
str(df.1718)
head(df.1718)
View(df.1718)
summary(df.1718)

str(df.1819)
head(df.1819)
View(df.1819)
summary(df.1819)

str(df.1920)
head(df.1920)
View(df.1920)
summary(df.1920)
```

3.Con la función select del paquete dplyr selecciona únicamente las columnas Date, HomeTeam, AwayTeam, FTHG, FTAG y FTR; esto para cada uno de los data frames. (Hint: también puedes usar lapply).

```
suppressMessages(library(dplyr))
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
df.1920<-df.1920[,-3] # Eliminamos la columna Tiempo de este data frame
df<-list(df.1718,df.1819,df.1920)
df<-lapply(df,select,Date:FTR)
```

4.Asegúrate de que los elementos de las columnas correspondientes de los nuevos data frames sean del mismo tipo (Hint 1: usa as.Date y mutate para arreglar las fechas). Con ayuda de la función rbind forma un único data frame que contenga las seis columnas mencionadas en el punto 3 (Hint 2: la función do.call podría ser utilizada).

```
head(as.Date(df.1718$Date,format="%d/%m/%Y"))
```

```
## [1] "0017-08-18" "0017-08-18" "0017-08-19" "0017-08-19" "0017-08-19"
## [6] "0017-08-20"
```

Observamos que al utilizar la función `as.Date`, nos convierte la fecha en formato extraño.

Lo corregimos reemplazando el string “00” con el string “20”, y convertimos de nuevo a Fecha:

```
df<-lapply(df,mutate,Date=as.Date(sub("00","20",as.character(as.Date(Date,format="%d/%m/%Y")))))
str(df)
```

```
## List of 3
## $ : 'data.frame': 380 obs. of 6 variables:
## ..$ Date : Date[1:380], format: "2017-08-18" "2017-08-18" ...
## ..$ HomeTeam: Factor w/ 20 levels "Alaves","Ath Bilbao",...: 13 19 6 10 17 2 4 11 14 15 ...
## ..$ AwayTeam: Factor w/ 20 levels "Alaves","Ath Bilbao",...: 1 12 18 3 8 9 5 16 20 7 ...
## ..$ FTHG : int [1:380] 1 1 2 2 1 0 2 0 1 0 ...
## ..$ FTAG : int [1:380] 0 0 3 2 1 0 0 3 0 1 ...
## ..$ FTR : Factor w/ 3 levels "A","D","H": 3 3 1 2 2 2 3 1 3 1 ...
## $ : 'data.frame': 380 obs. of 6 variables:
## ..$ Date : Date[1:380], format: "2018-08-17" "2018-08-17" ...
## ..$ HomeTeam: Factor w/ 20 levels "Alaves","Ath Bilbao",...: 5 10 4 6 20 7 14 19 2 17 ...
## ..$ AwayTeam: Factor w/ 20 levels "Alaves","Ath Bilbao",...: 13 18 1 8 16 11 9 15 12 3 ...
## ..$ FTHG : int [1:380] 0 0 3 1 1 1 2 1 2 1 ...
## ..$ FTAG : int [1:380] 3 0 0 1 2 2 0 4 1 1 ...
## ..$ FTR : Factor w/ 3 levels "A","D","H": 1 2 3 2 1 1 3 1 3 2 ...
## $ : 'data.frame': 380 obs. of 6 variables:
## ..$ Date : Date[1:380], format: "2019-08-16" "2019-08-17" ...
## ..$ HomeTeam: Factor w/ 20 levels "Alaves","Ath Bilbao",...: 2 6 18 13 11 20 1 8 5 3 ...
## ..$ AwayTeam: Factor w/ 20 levels "Alaves","Ath Bilbao",...: 4 15 17 7 14 10 12 16 19 9 ...
## ..$ FTHG : int [1:380] 1 1 1 2 0 4 1 0 1 1 ...
## ..$ FTAG : int [1:380] 0 3 1 1 1 4 0 2 2 0 ...
## ..$ FTR : Factor w/ 3 levels "A","D","H": 3 1 2 3 1 2 3 1 1 3 ...
```

```
#
df<-do.call(rbind,df)
head(df)
```

```
##      Date      HomeTeam   AwayTeam FTHG FTAG FTR
## 1 2017-08-18    Leganes     Alaves     1     0   H
## 2 2017-08-18 Valencia Las Palmas     1     0   H
## 3 2017-08-19      Celta   Sociedad     2     3   A
## 4 2017-08-19      Girona Ath Madrid     2     2   D
## 5 2017-08-19     Sevilla   Espanol     1     1   D
## 6 2017-08-20 Ath Bilbao     Getafe     0     0   D
```