

PW02

Equipo 22

1/29/2021

Postwork Sesión 2

Objetivo

- Importar múltiples archivos csv a R
- Observar algunas características y manipular los data frames
- Combinar múltiples data frames en un único data frame

Requisitos

1. R, RStudio
2. Haber realizado el prework y seguir el curso de los ejemplos de la sesión

Desarrollo Ahora vamos a generar un cúmulo de datos mayor al que se tenía, esta es una situación habitual que se puede presentar para complementar un análisis, siempre es importante estar revisando las características o tipos de datos que tenemos, por si es necesario realizar alguna transformación en las variables y poder hacer operaciones aritméticas si es el caso, además de sólo tener presente algunas de las variables, no siempre se requiere el uso de todas para ciertos procesamiento. 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>

```
lfp1920 <- "https://www.football-data.co.uk/mmz4281/1920/SP1.csv"
lfp1819 <- "https://www.football-data.co.uk/mmz4281/1819/SP1.csv"
lfp1718 <- "https://www.football-data.co.uk/mmz4281/1718/SP1.csv"

download.file(url = lfp1920, destfile = "lfp1920.csv", mode = "wb")
download.file(url = lfp1819, destfile = "lfp1819.csv", mode = "wb")
download.file(url = lfp1718, destfile = "lfp1718.csv", mode = "wb")

liga <- lapply(list.files(pattern="*.csv"), read.csv)
```

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

```
str(liga)
head(liga)
summary(liga)
```

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`).

```
library(dplyr)
lista <- lapply(liga,select,c("Date","HomeTeam","AwayTeam","FTHG","HTAG","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).

```
data <- do.call(rbind, lista)
data <- mutate(data,Date = as.Date(Date, "%d/%m/%y"))
head(data)
```

##	Date	HomeTeam	AwayTeam	FTHG	HTAG	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	1	A
## 4	2017-08-19	Girona	Ath Madrid	2	0	D
## 5	2017-08-19	Sevilla	Espanol	1	1	D
## 6	2017-08-20	Ath Bilbao	Getafe	0	0	D

Se escribe un archivo csv que se ocupará para el postwork de la sesión 3

```
write.csv(data, file = "temporadas.csv", row.names = FALSE)
```