

Programación y estadística con R

| Equipo 16

Ana Elizabeth Guzmán Jiménez

Carlos Paz

Fred Jordi Miramontes Arias

Luis Arturo Rosas León

Ludim Sánchez

Norma Arrazola Herrera

Contenido

PostWork 1: Introducción a R y Software	4
Ejercicio 1: Importa los datos de soccer de la temporada 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	4
Ejercicio 2: Del data frame que resulta de importar los datos a R, extrae las columnas que contienen los números de goles anotados por los equipos que jugaron en casa (FTHG) y los goles anotados por los equipos que jugaron como visitante (FTAG)	4
Ejercicio 3: Consulta cómo funciona la función table en R al ejecutar en la consola ?table	5
Ejercicio 4: Posteriormente elabora tablas de frecuencias relativas para estimar las siguientes probabilidades:	6
Parte 4.1: La probabilidad (marginal) de que el equipo que juega en casa anote x goles ($x = 0, 1, 2, \dots$)	6
Parte 4.2: La probabilidad (marginal) de que el equipo que juega como visitante anote y goles ($y = 0, 1, 2, \dots$)	9
Parte 4.3: La probabilidad (conjunta) de que el equipo que juega en casa anote x goles y el equipo que juega como visitante anote y goles ($x = 0, 1, 2, \dots, y = 0, 1, 2, \dots$)	9
PostWork 2: Programación y manipulación de datos en R	10
Ejercicio 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	10
Ejercicio 2: Revisa la estructura de de los data frames al usar las funciones: str, head, View y summary	11
Ejercicio 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)	11
Ejercicio 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)	12
PostWork 3: Análisis Exploratorio de Datos (AED o EDA) con R	13
Ejercicio 1: Con el último data frame obtenido en el postwork de la sesión 2, elabora tablas de frecuencias relativas para estimar las siguientes probabilidades:	13
Parte 1.1: La probabilidad (marginal) de que el equipo que juega en casa anote x goles ($x=0,1,2,$)	14
Parte 1.2: La probabilidad (marginal) de que el equipo que juega como visitante anote y goles ($y=0,1,2,$)	14

Parte 1.3: La probabilidad (conjunta) de que el equipo que juega en casa anote x goles y el equipo que juega como visitante anote y goles ($x=0,1,2,, y=0,1,2,$)	15
Ejercicio 2: Realiza lo siguiente:	15
Parte 2.1: Un gráfico de barras para las probabilidades marginales estimadas del número de goles que anota el equipo de casa.....	15
Parte 2.2: Un gráfico de barras para las probabilidades marginales estimadas del número de goles que anota el equipo visitante.	17
Parte 2.3: Un HeatMap para las probabilidades conjuntas estimadas de los números de goles que anotan el equipo de casa y el equipo visitante en un partido.....	18

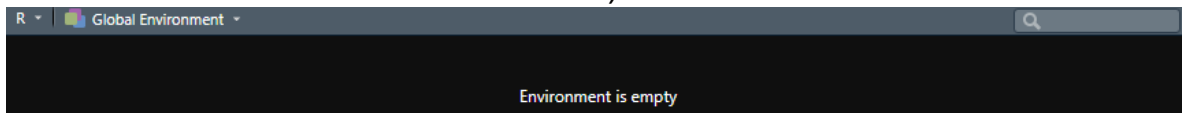
PostWork 1: Introducción a R y Software

Ejercicio 1: Importa los datos de soccer de la temporada 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>

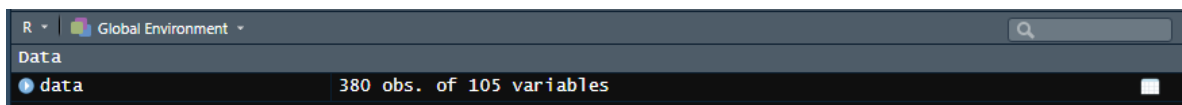
```
#<variable> <- read.csv("enlace o ubicación donde se encuentra el  
archivo csv")
```

```
#La variable almacenara los datos que contenga el archivo el cual se  
#obtendrá mediante la lectura del método read.csv que recibe como  
#parámetro tanto como el enlace del csv como la dirección local donde  
#se encuentre el archivo csv.
```

```
data <- read.csv("https://www.football-  
data.co.uk/mmz4281/1920/SP1.csv")
```



```
data <- read.csv("https://www.football-data.co.uk/mmz4281/1920/SP1.csv")  
> data <- read.csv("https://www.football-data.co.uk/mmz4281/1920/SP1.csv")  
> |
```



Ejercicio 2: Del data frame que resulta de importar los datos a R, extrae las columnas que contienen los números de goles anotados por los equipos que jugaron en casa (FTHG) y los goles anotados por los equipos que jugaron como visitante (FTAG)

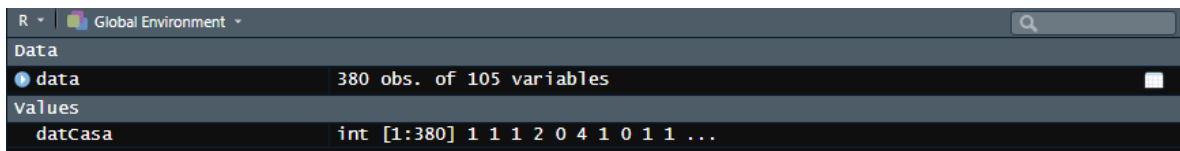
```
#<variable> <- <DataFrame>$<columna>
```

#La variable almacenara los datos los cuales se extraen de la
#<columna> del <DataFrame> que se obtienen mediante el símbolo \$.

```
datCasa <- data$FTHG
```

```
datCasa<-data$FTHG
```

```
> datCasa<-data$FTHG  
> |
```



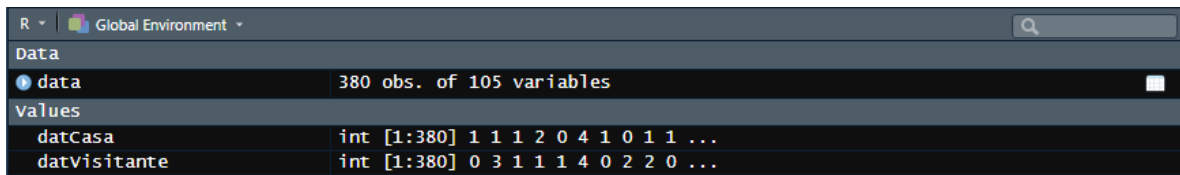
R Global Environment

Data	
data	380 obs. of 105 variables
Values	
datCasa	int [1:380] 1 1 1 2 0 4 1 0 1 1 ...

```
datVisitante <- data$FTAG
```

```
datVisitante<-data$FTAG
```

```
> datVisitante<-data$FTAG  
> |
```



R Global Environment

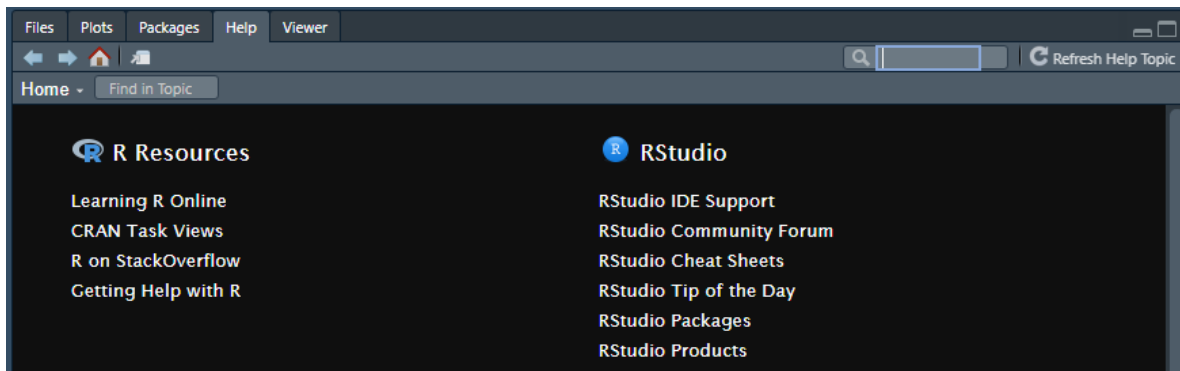
Data	
data	380 obs. of 105 variables
Values	
datCasa	int [1:380] 1 1 1 2 0 4 1 0 1 1 ...
datVisitante	int [1:380] 0 3 1 1 1 4 0 2 2 0 ...

Ejercicio 3: Consulta cómo funciona la función table en R al ejecutar
en la consola ?table

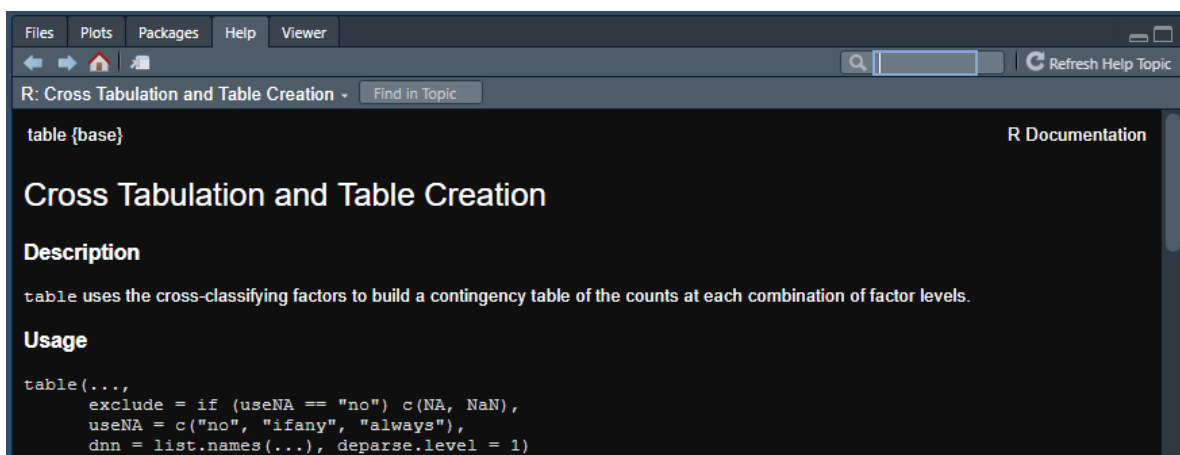
#?<método>

#Esta es una de las formas de poder acceder a la ayuda que se ofrece
#para saber el funcionamiento del <método> tanto como lo que realiza
#como los parámetros que se requieran para utilizarlo.

?table



```
> ?table
> |
```



Ejercicio 4: Posteriormente elabora tablas de frecuencias relativas para estimar las siguientes probabilidades:

Parte 4.1: La probabilidad (marginal) de que el equipo que juega en casa anote x goles ($x = 0, 1, 2, \dots$)

```
#(<variable> <- table(<vector>))
```

```
#El método table va a convertir los datos del vector en una matriz la se
#termina almacenando en la variable para su posterior manipulación
#/y/o consulta, al estar encerrados entre paréntesis al final de
#almacenar los datos en la variable se terminarán consultando.
```

```
(tablaCasa<-table(datCasa))
```

```
(tablaCasa<-table(datCasa))
```

```
> (tablaCasa<-table(datCasa))
datCasa
  0    1    2    3    4    5    6
88 132  99  38  14    8    1
> |
```

Environment	History	Connections	Tutorial
Import Dataset 112 MiB			
R Global Environment			
Data			
data 380 obs. of 105 variables			
Values			
datCasa	int [1:380]	1 1 1 2 0 4 1 0 1 1 ...	
datVisitante	int [1:380]	0 3 1 1 1 4 0 2 2 0 ...	
tablaCasa	'table' int [1:7(1d)]	88 132 99 38 14 8 1	

```
(totalGoles<-table(datVisitante,datCasa))
```

```
(totalGoles<-table(datvisitante,datCasa))
```

```
> (totalGoles<-table(datvisitante,datCasa))
      datCasa
datvisitante  0    1    2    3    4    5    6
  0  33  43  39  14    4    2    1
  1  28  49  35  14    5    3    0
  2  15  32  20    7    4    3    0
  3   8   5   3   2   0   0   0
  4   2   3   2   1   1   0   0
  5   2   0   0   0   0   0   0
> |
```

Environment	History	Connections	Tutorial
Import Dataset 112 MiB			
R Global Environment			
Data			
data 380 obs. of 105 variables			
Values			
datCasa	int [1:380]	1 1 1 2 0 4 1 0 1 1 ...	
datvisitante	int [1:380]	0 3 1 1 1 4 0 2 2 0 ...	
tablaCasa	'table' int [1:7(1d)]	88 132 99 38 14 8 1	
totalGoles	'table' int [1:6, 1:7]	33 28 15 8 2 2 43 49 32 5 ...	

```
#(<variable> <- sum(<table>))
```

#El método sum va retornar la suma de todos los elementos
#contenidos en la <table>.

```
(totalFrecAbsoluta <- sum(totalGoles))
```

```
(totalFrecAbsoluta <- sum(totalGoles))
```

```
> (totalFrecAbsoluta <- sum(totalGoles))  
[1] 380  
>
```

R Global Environment	
Data	
data	380 obs. of 105 variables
values	
datCasa	int [1:380] 1 1 1 2 0 4 1 0 1 1 ...
datVisitante	int [1:380] 0 3 1 1 1 4 0 2 2 0 ...
tablaCasa	'table' int [1:7(1d)] 88 132 99 38 14 8 1
totalFrecAbsoluta	380L
totalGoles	'table' int [1:6, 1:7] 33 28 15 8 2 2 43 49 32 5 ...

```
#(<variable> <- round(<table>/<int>,<int>))
```

#El método round recibe 2 parámetros el primero es una tabla que
#contiene la cantidad de goles por columna el cual se divide en el
#número total de goles y el segundo parámetro es para limitar el
#número de decimales a imprimir por lo que round retornara una tabla
#con los resultados por columna limitados al numero de decimales
#colocados en el segundo parámetro.

```
(FrecRelCasa <- round (tablaCasa/totalFrecAbsoluta,4))
```

```
(FrecRelCasa<-round (tablaCasa/totalFrecAbsoluta,4))
```

```
> (FrecRelCasa<-round (tablaCasa/totalFrecAbsoluta,4))  
datCasa  
      0      1      2      3      4      5      6  
0.2316 0.3474 0.2605 0.1000 0.0368 0.0211 0.0026  
>
```

R Global Environment	
Data	
data	380 obs. of 105 variables
values	
datCasa	int [1:380] 1 1 1 2 0 4 1 0 1 1 ...
datVisitante	int [1:380] 0 3 1 1 1 4 0 2 2 0 ...
FrecRelCasa	'table' num [1:7(1d)] 0.2316 0.3474 0.2605 0.1 0.0368 ...
tablaCasa	'table' int [1:7(1d)] 88 132 99 38 14 8 1
totalFrecAbsoluta	380L
totalGoles	'table' int [1:6, 1:7] 33 28 15 8 2 2 43 49 32 5 ...

Parte 4.2: La probabilidad (marginal) de que el equipo que juega como visitante anote y goles ($y = 0, 1, 2, \dots$)

```
(tablaCasa <- table(datVisitante))
```

```
(tablaCasa<-table(datvisitante))
datvisitante
  0   1   2   3   4   5
136 134  81  18   9   2
>
```

R - Global Environment	
Data	
data	380 obs. of 105 variables
Values	
datCasa	int [1:380] 1 1 1 2 0 4 1 0 1 1 ...
datVisitante	int [1:380] 0 3 1 1 1 4 0 2 2 0 ...
FrecRelCasa	'table' num [1:7(1d)] 0.2316 0.3474 0.2605 0.1 0.0368 ...
tablaCasa	'table' int [1:6(1d)] 136 134 81 18 9 2
totalFrecAbsoluta	380L
totalGoles	'table' int [1:6, 1:7] 33 28 15 8 2 2 43 49 32 5 ...

```
(FrecRelVisitante <- round(tablaCasa/totalFrecAbsoluta,4))
```

```
(FrecRelVisitante<-round (tablaCasa/totalFrecAbsoluta,4))
> (FrecRelVisitante<-round (tablaCasa/totalFrecAbsoluta,4))
datvisitante
  0   1   2   3   4   5
0.3579 0.3526 0.2132 0.0474 0.0237 0.0053
>
```

R - Global Environment	
Data	
data	380 obs. of 105 variables
Values	
datCasa	int [1:380] 1 1 1 2 0 4 1 0 1 1 ...
datVisitante	int [1:380] 0 3 1 1 1 4 0 2 2 0 ...
FrecRelCasa	'table' num [1:7(1d)] 0.2316 0.3474 0.2605 0.1 0.0368 ...
FrecRelVisitante	'table' num [1:6(1d)] 0.3579 0.3526 0.2132 0.0474 0.0237 ...
tablaCasa	'table' int [1:6(1d)] 136 134 81 18 9 2
totalFrecAbsoluta	380L
totalGoles	'table' int [1:6, 1:7] 33 28 15 8 2 2 43 49 32 5 ...

Parte 4.3: La probabilidad (conjunta) de que el equipo que juega en casa anote x goles y el equipo que juega como visitante anote y goles ($x = 0, 1, 2, \dots, y = 0, 1, 2, \dots$)

```
(FrecRelCon <- round(totalGoles/totalFrecAbsoluta,4))
```

```
(FrecRelCon<-round(totalGoles/totalFrecAbsoluta,4))
```

```
> (FrecRelCon<-round(totalGoles/totalFrecAbsoluta,4))
      datCasa
datvisitante  0    1    2    3    4    5    6
      0 0.0868 0.1132 0.1026 0.0368 0.0105 0.0053 0.0026
      1 0.0737 0.1289 0.0921 0.0368 0.0132 0.0079 0.0000
      2 0.0395 0.0842 0.0526 0.0184 0.0105 0.0079 0.0000
      3 0.0211 0.0132 0.0079 0.0053 0.0000 0.0000 0.0000
      4 0.0053 0.0079 0.0053 0.0026 0.0026 0.0000 0.0000
      5 0.0053 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
> |
```

R Global Environment	
Data	
data	380 obs. of 105 variables
Values	
datCasa	int [1:380] 1 1 1 2 0 4 1 0 1 1 ...
datVisitante	int [1:380] 0 3 1 1 1 4 0 2 2 0 ...
FrecRelCasa	'table' num [1:7(1d)] 0.2316 0.3474 0.2605 0.1 0.0368 ...
FrecRelCon	'table' num [1:6, 1:7] 0.0868 0.0737 0.0395 0.0211 0.0053 ...
FrecRelVisitante	'table' num [1:6(1d)] 0.3579 0.3526 0.2132 0.0474 0.0237 ...
tablacasa	'table' int [1:6(1d)] 136 134 81 18 9 2
totalFrecAbsoluta	380L
totalGoles	'table' int [1:6, 1:7] 33 28 15 8 2 2 43 49 32 5 ...

PostWork 2: Programación y manipulación de datos en R

Ejercicio 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>

```
#Importamos los datos de soccer de las temporadas 2017/2018, 2018/2019 y
2019/2020 de la primera división de la liga española a R, desde
https://www.football-data.co.uk/spainm.php"
setwd ("C:/Users/arraz/Documents/Bedu_statisticsWithR/spainleague")
#Cambia el directorio al tuyo
```

```
e11920 <- "https://www.football-data.co.uk/mmz4281/1920/SP1.csv"
e11819 <- "https://www.football-data.co.uk/mmz4281/1819/SP1.csv"
e11718 <- "https://www.football-data.co.uk/mmz4281/1718/SP1.csv"
```

```
download.file(url = e11920, destfile = "e1-1920.csv", mode = "wb")
download.file(url = e11819, destfile = "e1-1829.csv", mode = "wb")
download.file(url = e11718, destfile = "e1-1719.csv", mode = "wb")
```

```
dir()
```

```
ligaesp <- lapply(dir(), read.csv) # leemos los archivos descargados
usando la funcion lapply y guardandolos en un dataframe
```

Ejercicio 2: Revisa la estructura de de los data frames al usar las funciones: str, head, View y summary

#Obtenemos una mejor idea de las características de los data frames al usar las funciones: str, head, View y summary

```
str(ligaesp[[1]]); str(ligaesp[[2]]); str(ligaesp[[3]])
head(ligaesp[[1]]); head(ligaesp[[2]]); head(ligaesp[[3]])
View(ligaesp[[1]]); View(ligaesp[[2]]); View(ligaesp[[3]])
summary(ligaesp[[1]]); summary(ligaesp[[2]]); summary(ligaesp[[3]])
```

Showing 1 to 9 of 380 entries, 64 total columns

Div	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	HS	AS	HST	AST	HF	AF	HC	AC	HY	AY	HR	AR
1	SP1	18/08/17	Leganes	Alaves	1	0	H				16	6	9	3	14	18	4	2	0	1	0
2	SP1	18/08/17	Valencia	Las Palmas	1	0	H				22	5	6	4	25	13	5	2	3	3	0
3	SP1	19/08/17	Celta	Sociedad	2	3	A		1	1	16	13	5	6	12	11	5	4	3	1	0
4	SP1	19/08/17	Girona	Ath Madrid	2	2	D				13	9	6	3	15	15	6	0	2	4	0
5	SP1	19/08/17	Sevilla	Espanol	1	1	D		1	1	9	9	4	6	14	12	7	3	2	4	1
6	SP1	20/08/17	Ath Bilbao	Getafe	0	0	D		0	0	12	8	2	2	16	15	7	6	1	3	0
7	SP1	20/08/17	Barcelona	Betis	2	0	H				15	3	2	0	16	15	8	0	2	1	0
8	SP1	20/08/17	La Coruna	Real Madrid	0	3	A		0	2	12	16	6	8	16	12	4	4	5	1	0
9	SP1	21/08/17	Levante	Villarreal	1	0	H		0	0	14	9	3	1	18	14	11	6	1	3	0

Console output:

```
R 4.1.0 - ~/Bedu_statisticsWithR/spainleague/
> View(ligaesp[[1]]); View(ligaesp[[2]]); View(ligaesp[[3]])
> summary(ligaesp[[1]]); summary(ligaesp[[2]]); summary(ligaesp[[3]])
```

Summary statistics for the first three data frames:

Div	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	HS	AS	HST	AST	HF	AF	HC	AC	HY	AY	HR	AR
Length:380	Length:380	Length:380	Length:380	Min.:0.000	Min.:0.000	Length:380	Min.:0.0000	Min.:0.0000	Length:380	Min.:2.00	Min.:1.00	Min.:0.000	Min.:0.000	Min.:4.00	Min.:0.00	Min.:0.000	Min.:0.000	Min.:0.000	Min.:0.000	Min.:0.000	Min.:1.050
Class:character	Class:character	Class:character	Class:character	1st Qu.:0.750	1st Qu.:0.000	Class:character	1st Qu.:0.0000	1st Qu.:0.0000	Class:character	1st Qu.:10.00	1st Qu.:8.00	1st Qu.:3.000	1st Qu.:2.000	1st Qu.:11.00	1st Qu.:11.00	1st Qu.:4.000	1st Qu.:2.000	1st Qu.:1.000	1st Qu.:2.000	1st Qu.:0.0000	1st Qu.:1.617
Mode:character	Mode:character	Mode:character	Mode:character	Median:1.000	Median:1.000	Mode:character	Median:0.0000	Median:0.0000	Mode:character	Median:13.00	Median:10.00	Median:4.500	Median:3.000	Median:13.00	Median:13.00	Median:5.000	Median:3.000	Median:1.000	Median:3.000	Median:0.0000	Median:2.075
				Mean:1.547	Mean:1.147		Mean:0.6605	Mean:0.4868		Mean:13.53	Mean:10.47	Mean:4.758	Mean:3.805	Mean:13.73	Mean:13.95	Mean:5.613	Mean:4.192	Mean:2.339	Mean:2.676	Mean:0.1105	Mean:2.777
				3rd Qu.:2.000	3rd Qu.:2.000		3rd Qu.:1.0000	3rd Qu.:1.0000		3rd Qu.:16.00	3rd Qu.:13.00	3rd Qu.:6.000	3rd Qu.:5.000	3rd Qu.:17.00							
				Max.:7.000	Max.:6.000		Max.:5.0000	Max.:3.0000		Max.:30.00	Max.:24.00	Max.:14.000	Max.:13.000	Max.:29.00							

Ejercicio 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).

#Con la función select del paquete dplyr seleccionamos las columnas Date, HomeTeam, AwayTeam, FTHG, FTAG y FTR para cada data frame.

```
ligaesp <- lapply(ligaesp, select, c("Date", "HomeTeam", "AwayTeam",
"FTHG", "FTAG", "FTR"))
```

	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR
1	18/08/17	Leganes	Alaves	1	0	H
2	18/08/17	Valencia	Las Palmas	1	0	H
3	19/08/17	Celta	Sociedad	2	3	A
4	19/08/17	Girona	Ath Madrid	2	2	D
5	19/08/17	Sevilla	Espanol	1	1	D
6	20/08/17	Ath Bilbao	Getafe	0	0	D
7	20/08/17	Barcelona	Betis	2	0	H
8	20/08/17	La Coruna	Real Madrid	0	3	A
9	21/08/17	Levante	Villarreal	1	0	H
10	21/08/17	Malaga	Eibar	0	1	A
11	25/08/17	Betis	Celta	2	1	H
12	25/08/17	Sociedad	Villarreal	3	0	H
13	26/08/17	Alaves	Barcelona	0	2	A

Ejercicio 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).

"Aseguramos de que los elementos de las columnas correspondientes de los nuevos data frames sean del mismo tipo
usamos `as.Date` y `mutate` para arreglar las fechas"

```
ligaesp[[1]] <- mutate(ligaesp[[1]], Date = as.Date(Date,
format="%d/%m/%y"))
ligaesp[[2]] <- mutate(ligaesp[[2]], Date = as.Date(Date, "%d/%m/%Y"))
ligaesp[[3]] <- mutate(ligaesp[[3]], Date = as.Date(Date, "%d/%m/%Y"))
```

"Con la función `rbind` y `do.call` formamos un único data frame que contenga las seis columnas mencionadas en el punto 3"

```
data <- do.call(rbind, ligaesp)
head(data)
dim(data)
```

```
> head(data)
      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
> dim(data)
[1] 1140    6
```

PostWork 3: Análisis Exploratorio de Datos (AED o EDA) con R

Ejercicio 1: Con el último data frame obtenido en el postwork de la sesión 2, elabora tablas de frecuencias relativas para estimar las siguientes probabilidades:

```
df <- read.csv("https://github.com/sh4rkd/Equipo-16-R/raw/master/PostWork-2/csv/total.csv")
```

```
datCasa <- df$FTHG
```

```
datVis <- df$FTAG
```

```
FrecAbs <- table(datVis, datCasa)
```

```
sumaFrecAbs <- sum(FrecAbs)
```

```
df <- read.csv("https://github.com/sh4rkd/Equipo-16-R/raw/master/PostWork-2/csv/total.csv")
datCasa<-df$FTHG
datVis<-df$FTAG
FrecAbs<-table(datVis,datCasa)
sumaFrecAbs<-sum(FrecAbs)
```

```
> df <- read.csv("https://github.com/sh4rkd/Equipo-16-R/raw/master/Postwork-2/csv/total.csv")
> datCasa<-df$FTHG
> datVis<-df$FTAG
> FrecAbs<-table(datVis,datCasa)
> sumaFrecAbs<-sum(FrecAbs)
> |
```

R Global Environment	
Data	
df	1140 obs. of 6 variables
Values	
datCasa	int [1:1140] 1 1 2 2 1 0 2 0 1 0 ...
datVis	int [1:1140] 0 0 3 2 1 0 0 3 0 1 ...
FrecAbs	'table' int [1:7, 1:9] 89 92 52 21 6 5 0 132 131 78 ...
sumaFrecAbs	1140L

Parte 1.1: La probabilidad (marginal) de que el equipo que juega en casa anote x goles (x=0,1,2,)

(ProbCasa <- round(table(datCasa)/sumaFrecAbs,4))

```
(ProbCasa<-round(table(datCasa)/sumaFrecAbs,4))
```

```
> (ProbCasa<-round(table(datCasa)/sumaFrecAbs,4))
datCasa
  0      1      2      3      4      5      6      7      8
0.2325 0.3272 0.2667 0.1123 0.0351 0.0193 0.0053 0.0009 0.0009
> |
```

R Global Environment	
Data	
df	1140 obs. of 6 variables
Values	
datCasa	int [1:1140] 1 1 2 2 1 0 2 0 1 0 ...
datVis	int [1:1140] 0 0 3 2 1 0 0 3 0 1 ...
FrecAbs	'table' int [1:7, 1:9] 89 92 52 21 6 5 0 132 131 78 ...
ProbCasa	'table' num [1:9(1d)] 0.2325 0.3272 0.2667 0.1123 0.0351 ...
sumaFrecAbs	1140L

Parte 1.2: La probabilidad (marginal) de que el equipo que juega como visitante anote y goles (y=0,1,2,)

(ProbVis<-round(table(datVis)/sumaFrecAbs,4))

```
(ProbVis<-round(table(datVis)/sumaFrecAbs,4))
```

```
> (ProbVis<-round(table(datVis)/sumaFrecAbs,4))
datVis
  0      1      2      3      4      5      6
0.3518 0.3404 0.2123 0.0544 0.0289 0.0096 0.0026
> |
```

R Global Environment	
Data	
df	1140 obs. of 6 variables
Values	
datCasa	int [1:1140] 1 1 2 2 1 0 2 0 1 0 ...
datVis	int [1:1140] 0 0 3 2 1 0 0 3 0 1 ...
FrecAbs	'table' int [1:7, 1:9] 89 92 52 21 6 5 0 132 131 78 ...
ProbCasa	'table' num [1:9(1d)] 0.2325 0.3272 0.2667 0.1123 0.0351 ...
ProbVis	'table' num [1:7(1d)] 0.3518 0.3404 0.2123 0.0544 0.0289 ...
sumaFrecAbs	1140L

Parte 1.3: La probabilidad (conjunta) de que el equipo que juega en casa anote x goles y el equipo que juega como visitante anote y goles (x=0,1,2,, y=0,1,2,)

(FrecRel<-round(FrecAbs/sumaFrecAbs,4))

```
(FrecRel<-round(FrecAbs/sumaFrecAbs,4))
```

```
> (FrecRel<-round(FrecAbs/sumaFrecAbs,4))
      datCasa
datvis  0      1      2      3      4      5      6      7      8
0 0.0781 0.1158 0.0877 0.0447 0.0140 0.0088 0.0026 0.0000 0.0000
1 0.0807 0.1149 0.0939 0.0325 0.0105 0.0053 0.0018 0.0009 0.0000
2 0.0456 0.0684 0.0614 0.0246 0.0070 0.0044 0.0000 0.0000 0.0009
3 0.0184 0.0175 0.0114 0.0061 0.0000 0.0000 0.0009 0.0000 0.0000
4 0.0053 0.0088 0.0088 0.0018 0.0035 0.0009 0.0000 0.0000 0.0000
5 0.0044 0.0018 0.0018 0.0018 0.0000 0.0000 0.0000 0.0000 0.0000
6 0.0000 0.0000 0.0018 0.0009 0.0000 0.0000 0.0000 0.0000 0.0000
>
```

R Global Environment	
Data	
df	1140 obs. of 6 variables
Values	
datCasa	int [1:1140] 1 1 2 2 1 0 2 0 1 0 ...
datVis	int [1:1140] 0 0 3 2 1 0 0 3 0 1 ...
FrecAbs	'table' int [1:7, 1:9] 89 92 52 21 6 5 0 132 131 78 ...
FrecRel	'table' num [1:7, 1:9] 0.0781 0.0807 0.0456 0.0184 0.0053 ...
ProbCasa	'table' num [1:9(1d)] 0.2325 0.3272 0.2667 0.1123 0.0351 ...
ProbVis	'table' num [1:7(1d)] 0.3518 0.3404 0.2123 0.0544 0.0289 ...
sumaFrecAbs	1140L

Ejercicio 2: Realiza lo siguiente:

library(ggplot2)

```
library(ggplot2)
```

```
> library(ggplot2)
>
```

	Name	Description	Version	
<input checked="" type="checkbox"/>	ggplot2	Create Elegant Data Visualisations Using the Grammar of Graphics	3.3.4	

Parte 2.1: Un gráfico de barras para las probabilidades marginales estimadas del número de goles que anota el equipo de casa.

#barplot(<table>,<mensaje en eje x>,<mensaje en eje y>,<mensaje en #la cabecera>,<vector con valor inicial en la primera posición y valor

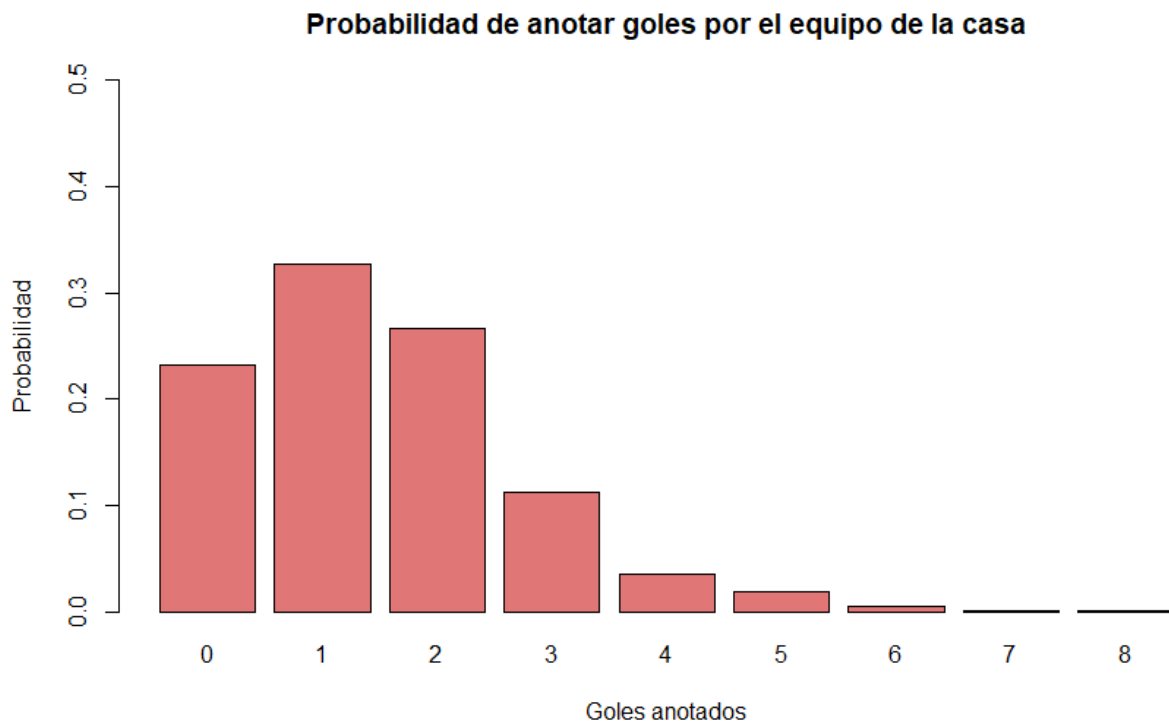
#final en la segunda posición, para determinar el tamaño de altura
#máximo para las barras>, <color en valor rgba>)

#el método barplot recibirá los parámetros anteriormente mencionados
#para graficar en forma de barra del color los elementos y medidas
#recibidas por parámetro.

```
barplot(ProbCasa,xlab="Goles anotados", ylab="Probabilidad",  
        main = "Probabilidad de anotar goles por el equipo de la casa",  
        ylim =c(0,0.5),  
        col = rgb(0.8,0.1,0.1,0.6))
```

```
barplot(ProbCasa,xlab="Goles anotados", ylab="Probabilidad",  
        main = "Probabilidad de anotar goles por el equipo de la casa",  
        ylim =c(0,0.5),  
        col = rgb(0.8,0.1,0.1,0.6))
```

```
> barplot(ProbCasa,xlab="Goles anotados", ylab="Probabilidad",  
+         main = "Probabilidad de anotar goles por el equipo de la casa",  
+         ylim =c(0,0.5),  
+         col = rgb(0.8,0.1,0.1,0.6))  
>
```

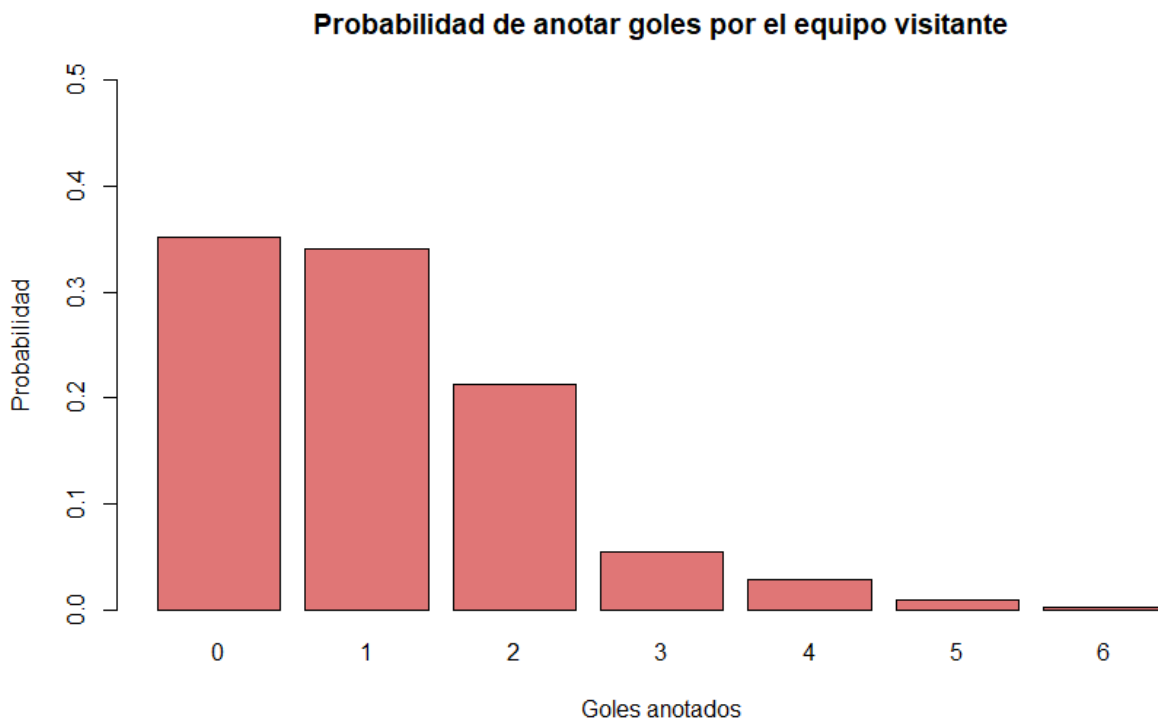


Parte 2.2: Un gráfico de barras para las probabilidades marginales estimadas del número de goles que anota el equipo visitante.

```
barplot(ProbVis,xlab="Goles anotados", ylab="Probabilidad",  
        main = "Probabilidad de anotar goles por el equipo visitante",  
        ylim =c(0,0.5),  
        col = rgb(0.8,0.1,0.1,0.6))
```

```
barplot(ProbVis,xlab="Goles anotados", ylab="Probabilidad",  
        main = "Probabilidad de anotar goles por el equipo visitante",  
        ylim =c(0,0.5),  
        col = rgb(0.8,0.1,0.1,0.6))
```

```
> barplot(ProbVis,xlab="Goles anotados", ylab="Probabilidad",  
+         main = "Probabilidad de anotar goles por el equipo visitante",  
+         ylim =c(0,0.5),  
+         col = rgb(0.8,0.1,0.1,0.6))  
>
```



Parte 2.3: Un HeatMap para las probabilidades conjuntas estimadas de los números de goles que anotan el equipo de casa y el equipo visitante en un partido.

```
ggplot(as.data.frame(FrecRel), aes(x=datCasa, y=datVis, fill = Freq)) +
  geom_tile()+
  geom_text(aes(label=round(Freq,4))) +
  scale_fill_gradient(low="pink", high="blue") + labs(x="Numero de Goles
de la Casa") + labs(y="Numero de Goles del Visitante")+
  labs(fill="Probabilidad")
```

```
ggplot(as.data.frame(FrecRel), aes(x=datCasa, y=datVis, fill = Freq))+
  geom_tile()+
  geom_text(aes(label=round(Freq,4)))+
  scale_fill_gradient(low="pink", high="blue") +
  labs(x="Numero de Goles de la Casa")+
  labs(y="Numero de Goles del Visitante")+
  labs(fill="Probabilidad")
```

```
> ggplot(as.data.frame(FrecRel), aes(x=datCasa, y=datVis, fill = Freq))+
+   geom_tile()+
+   geom_text(aes(label=round(Freq,4)))+
+   scale_fill_gradient(low="pink", high="blue") +
+   labs(x="Numero de Goles de la Casa")+
+   labs(y="Numero de Goles del visitante")+
+   labs(fill="Probabilidad")
>
```

