

PROGRAMACIÓN Y ESTADÍSTICA CON R

ANÁLISIS DE LOS EQUIPOS DE LA LIGA ESPAÑOLA DE FÚTBOL

CARLOS ALDAIR ORTÍZ SUÁREZ
DATA-ANALYSIS-21-12
27/01/2022
BEDU



Agenda

- Introducción a R y Software
- Programación y manipulación de datos
- Análisis Exploratorio de Datos
- Regresión lineal y clasificación
- Series de tiempo
- RStudio Cloud - Github, conexiones con BDs y lectura de datos externos
- Dashboards con Shiny - Entorno GUI

Se descarga la base de datos el día
03/Marzo/2022, desde la página:

<https://www.football-data.co.uk/spainm.php>

Introducción a R y Software

▲	Div	Date	Time	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	HS	AS	HST	AST	HF	AF	HC	AC	HY	AY
1	SP1	12/09/2020	15:00	Eibar	Celta	0	0	D	0	0	D	8	6	1	3	17	13	3	3	4	
2	SP1	12/09/2020	17:30	Granada	Ath Bilbao	2	0	H	0	0	D	4	9	2	2	8	10	1	2	3	
3	SP1	12/09/2020	20:00	Cadiz	Osasuna	0	2	A	0	1	A	12	6	3	4	14	21	6	2	2	
4	SP1	13/09/2020	13:00	Alaves	Betis	0	1	A	0	0	D	7	11	2	4	19	17	6	4	3	
5	SP1	13/09/2020	15:00	Valladolid	Sociedad	1	1	D	1	0	H	8	8	3	2	15	17	5	3	4	
6	SP1	13/09/2020	17:30	Villarreal	Huesca	1	1	D	0	1	A	15	10	4	3	10	22	6	6	1	
7	SP1	13/09/2020	20:00	Valencia	Levante	4	2	H	2	2	D	14	19	8	8	9	14	3	4	1	
8	SP1	19/09/2020	15:00	Villarreal	Eibar	2	1	H	0	0	D	14	12	8	4	5	12	4	3	1	
9	SP1	19/09/2020	17:30	Getafe	Osasuna	1	0	H	0	0	D	7	10	1	3	19	13	4	1	4	
10	SP1	19/09/2020	20:00	Celta	Valencia	2	1	H	1	0	H	23	9	6	3	12	13	6	6	3	
11	SP1	20/09/2020	15:00	Huesca	Cadiz	0	2	A	0	1	A	4	6	0	4	18	23	2	1	3	
12	SP1	20/09/2020	17:30	Betis	Valladolid	2	0	H	2	0	H	16	9	6	2	18	10	5	3	3	
13	SP1	20/09/2020	17:30	Granada	Alaves	2	1	H	1	1	D	10	5	3	3	9	12	4	4	0	
14	SP1	20/09/2020	20:00	Sociedad	Real Madrid	0	0	D	0	0	D	6	16	1	4	17	12	1	11	2	
15	SP1	26/09/2020	12:00	Alaves	Getafe	0	0	D	0	0	D	9	11	0	4	13	13	1	5	2	
16	SP1	26/09/2020	15:00	Valencia	Huesca	1	1	D	1	0	H	5	17	2	4	14	20	4	4	0	
17	SP1	26/09/2020	17:30	Elche	Sociedad	0	3	A	0	0	D	5	27	2	7	11	12	2	11	3	
18	SP1	26/09/2020	20:00	Betis	Real Madrid	2	3	A	2	1	H	15	11	5	4	9	10	8	3	5	
19	SP1	27/09/2020	11:00	Osasuna	Levante	1	3	A	1	1	D	9	11	4	5	15	16	3	7	2	

Showing 1 to 19 of 380 entries, 105 total columns

Datos de soccer de la temporada 2020/2021 de la
 primera división de la liga española.

	FTHG	FTAG
1	0	0
2	2	0
3	0	2
4	0	1
5	1	1
6	1	1
7	4	2
8	2	1
9	1	0
10	2	1
11	0	2
12	2	0
13	2	1
14	0	0
15	0	0
16	1	1
17	0	0

Showing 1 to 17 of 380 entries, 2 total columns

Datos de soccer de la temporada 2020/2021, goles de casa y visitantes.

Goles del Equipo de Casa

#Goles	Frecuencia Absoluta	Frecuencias Relativa	Probabilidad Marginal
1	0	98	0.257894737 25.79%
2	1	130	0.342105263 34.21%
3	2	98	0.257894737 25.79%
4	3	29	0.076315789 7.63%
5	4	19	0.050000000 5%
6	5	5	0.013157895 1.32%
7	6	1	0.002631579 0.26%

Tablas de Frecuencia Absoluta y Relativa y Probabilidad Marginal

Goles del Equipo de Visita

#Goles	Frecuencia Absoluta	Frecuencias Relativa	Probabilidad Marginal
1	0	115	0.302631579 30.26%
2	1	152	0.400000000 40%
3	2	73	0.192105263 19.21%
4	3	28	0.073684211 7.37%
5	4	10	0.026315789 2.63%
6	5	1	0.002631579 0.26%
7	6	1	0.002631579 0.26%

Tablas de Frecuencia Absoluta y Relativa y Probabilidad Marginal

Goles del Equipo de Casa y Visita

	Goles Casa	Goles Visita	Frecuencia Absoluta	Frecuencia Relativa	Probabilidad Conjunta
15	2	0	30	0.078947368	7.89%
3	0	2	25	0.065789474	6.58%
17	2	2	21	0.055263158	5.53%
10	1	2	17	0.044736842	4.47%
11	1	3	14	0.036842105	3.68%
23	3	1	11	0.028947368	2.89%
22	3	0	9	0.023684211	2.37%
30	4	1	8	0.021052632	2.11%
29	4	0	7	0.018421053	1.84%
4	0	3	6	0.015789474	1.58%
18	2	3	5	0.013157895	1.32%
24	3	2	5	0.013157895	1.32%
5	0	4	3	0.007894737	0.79%

Tablas de Frecuencia Absoluta y Relativa y Probabilidad
Conjunta

Programación y manipulación de datos



Datos de soccer de la temporada 2019/2020, 2020/2021 2021/2022 de la primera división de la liga española.

	Div	Date	Time	HomeTeam
1	SP1	16/08/2019	20:00	Ath Bilbao
2	SP1	17/08/2019	16:00	Celta
3	SP1	17/08/2019	18:00	Valencia
4	SP1	17/08/2019	19:00	Mallorca
5	SP1	17/08/2019	20:00	Leganes
6	SP1	17/08/2019	20:00	Villarreal
7	SP1	18/08/2019	16:00	Alaves
8	SP1	18/08/2019	18:00	Espanol
9	SP1	18/08/2019	20:00	Betis
10	SP1	18/08/2019	21:00	Ath Madrid
11	SP1	23/08/2019	19:00	Granada
12	SP1	23/08/2019	21:00	Levante
13	SP1	24/08/2019	16:00	Osasuna
14	SP1	24/08/2019	18:00	Real Madrid
15	SP1	24/08/2019	20:00	Celta

	Div	Date	Time	HomeTeam
1	SP1	12/09/2020	15:00	Eibar
2	SP1	12/09/2020	17:30	Granada
3	SP1	12/09/2020	20:00	Cadiz
4	SP1	13/09/2020	13:00	Alaves
5	SP1	13/09/2020	15:00	Valladolid
6	SP1	13/09/2020	17:30	Villarreal
7	SP1	13/09/2020	20:00	Valencia
8	SP1	19/09/2020	15:00	Villarreal
9	SP1	19/09/2020	17:30	Getafe
10	SP1	19/09/2020	20:00	Celta
11	SP1	20/09/2020	15:00	Huesca
12	SP1	20/09/2020	17:30	Betis
13	SP1	20/09/2020	17:30	Granada
14	SP1	20/09/2020	20:00	Sociedad
15	SP1	26/09/2020	12:00	Alaves

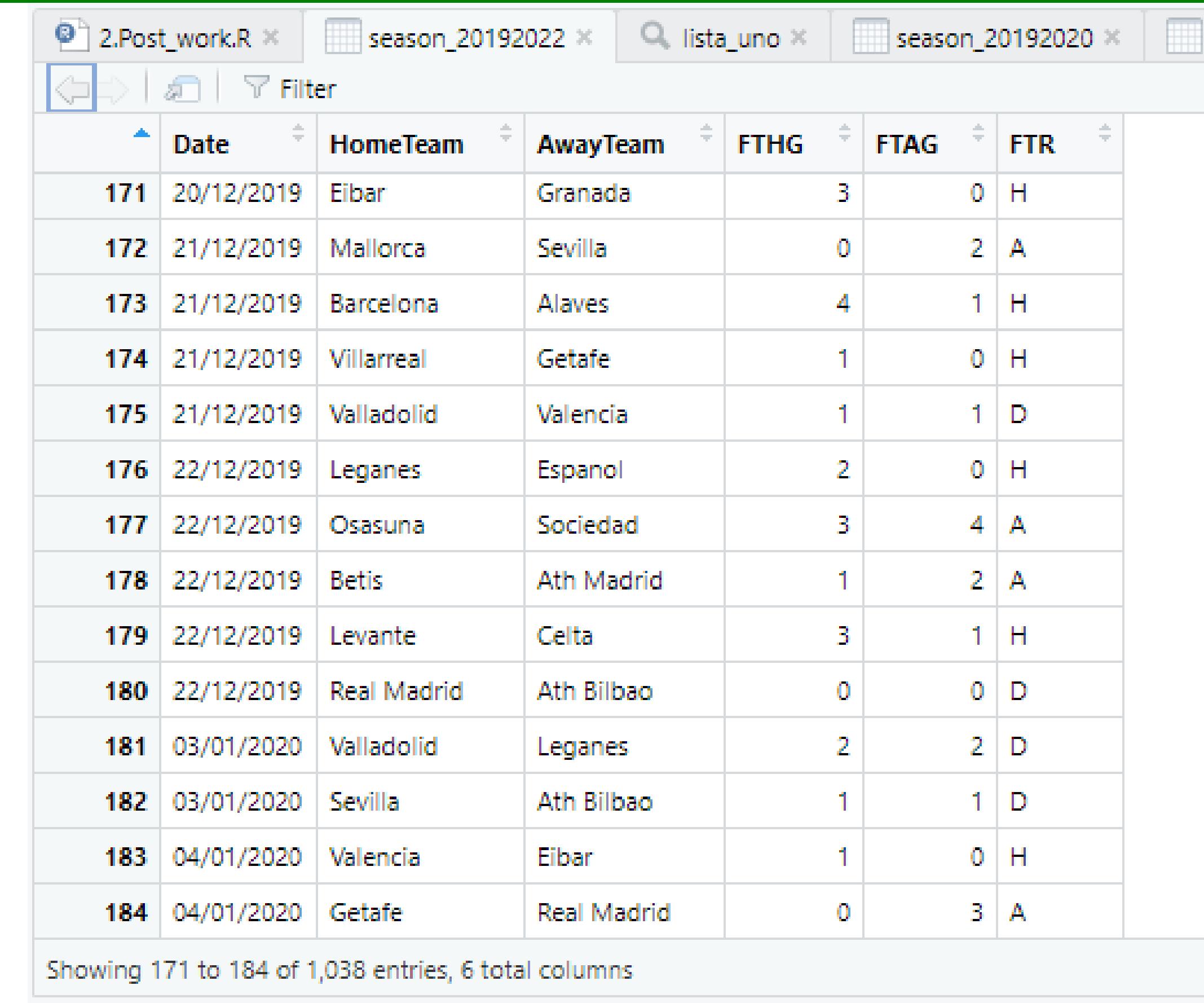
	Div	Date	Time	HomeTeam
1	SP1	13/08/2021	20:00	Valencia
2	SP1	14/08/2021	18:30	Cadiz
3	SP1	14/08/2021	18:30	Mallorca
4	SP1	14/08/2021	21:00	Alaves
5	SP1	14/08/2021	21:00	Osasuna
6	SP1	15/08/2021	16:30	Celta
7	SP1	15/08/2021	19:00	Barcelona
8	SP1	15/08/2021	21:15	Sevilla
9	SP1	16/08/2021	19:00	Villarreal
10	SP1	16/08/2021	21:00	Elche
11	SP1	20/08/2021	20:00	Betis
12	SP1	21/08/2021	16:00	Alaves
13	SP1	21/08/2021	18:30	Espanol
14	SP1	21/08/2021	18:30	Granada
15	SP1	21/08/2021	21:00	Ath Bilbao

Estructura de los Datos

La Liga 2020/21 Fixtures															
	Div	Date	Time	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	HS	AS		
1	SP1	12/09/2020	15:00	Eibar	Celta	0	0	D	0	0	D	8	6		
2	SP1	12/09/2020	17:30	Granada	Ath Bilbao	2	0	H	0	0	D	4	9		
3	SP1	12/09/2020	20:00	Cadiz	Osasuna	0	2	A	0	1	A	12	6		
4	SP1	13/09/2020	13:00	Alaves	Betis	0	1	A	0	0	D	7	11		
5	SP1	13/09/2020	15:00	valladolid	Sociedad	1	1	D	1	0	H	8	8		
6	SP1	13/09/2020	17:30	villarreal	Huesca	1	1	D	0	1	A	15	10		

```
> str(season_20202021)
'data.frame': 380 obs. of 105 variables:
 $ Div      : chr  "SP1" "SP1" "SP1" "SP1" ...
 $ Date     : chr  "12/09/2020" "12/09/2020" "12/09/2020" "13/09/2020"
 $ Time     : chr  "15:00" "17:30" "20:00" "13:00" ...
 $ HomeTeam : chr  "Eibar" "Granada" "Cadiz" "Alaves" ...
 $ AwayTeam : chr  "celta" "Ath Bilbao" "Osasuna" "Betis" ...
 $ FTHG    : int  0 2 0 0 1 1 4 2 1 2 ...
 $ FTAG    : int  0 0 2 1 1 1 2 1 0 1 ...
 $ FTR     : chr  "D" "H" "A" "A" ...
 $ HTHG    : int  0 0 0 0 1 0 2 0 0 1 ...
 $ HTAG    : int  0 0 1 0 0 1 2 0 0 0 ...
 $ HTR     : chr  "D" "D" "A" "D" ...
```

Se juntan los tres conjuntos de datos en una sola tabla,
con las columnas relevantes



The screenshot shows a data grid in RStudio with the following columns: Date, HomeTeam, AwayTeam, FTHG, FTAG, and FTR. The data consists of 14 rows, numbered 171 to 184. The columns are ordered as Date, HomeTeam, AwayTeam, FTHG, FTAG, and FTR.

	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR
171	20/12/2019	Eibar	Granada	3	0	H
172	21/12/2019	Mallorca	Sevilla	0	2	A
173	21/12/2019	Barcelona	Alaves	4	1	H
174	21/12/2019	Villarreal	Getafe	1	0	H
175	21/12/2019	Valladolid	Valencia	1	1	D
176	22/12/2019	Leganes	Espanol	2	0	H
177	22/12/2019	Osasuna	Sociedad	3	4	A
178	22/12/2019	Betis	Ath Madrid	1	2	A
179	22/12/2019	Levante	Celta	3	1	H
180	22/12/2019	Real Madrid	Ath Bilbao	0	0	D
181	03/01/2020	Valladolid	Leganes	2	2	D
182	03/01/2020	Sevilla	Ath Bilbao	1	1	D
183	04/01/2020	Valencia	Eibar	1	0	H
184	04/01/2020	Getafe	Real Madrid	0	3	A

Showing 171 to 184 of 1,038 entries, 6 total columns

Ajuste en tipo de Datos (Date)

	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR
1	16/08/2019	Ath Bilbao	Barcelona	1	0	H
2	17/08/2019	Celta	Real Madrid	1	3	A
3	17/08/2019	Valencia	Sociedad	1	1	D
4	17/08/2019	Mallorca	Eibar	2	1	H
5	17/08/2019	Leganes	Olasuna	0	1	A
6	17/08/2019	Villarreal	Granada	4	4	D
7	18/08/2019	Alaves	Levante	1	0	H
8	18/08/2019	Espanol	Sevilla	0	2	A
9	18/08/2019	Betis	Valladolid	1	2	A
10	18/08/2019	Ath Madrid	Getafe	1	0	H
11	23/08/2019	Granada	Sevilla	0	1	A
12	23/08/2019	Levante	Villarreal	2	1	H
13	24/08/2019	Olasuna	Eibar	0	0	D
14	24/08/2019	Real Madrid	Valladolid	1	1	D

Showing 1 to 14 of 1,038 entries, 6 total columns

Console Terminal × Jobs ×

R 4.1.2 · E:/Bedu/project_in_r/data/ ↵

```
> View(season_20192022)
> str(season_20192022)
'data.frame': 1038 obs. of 6 variables:
 $ Date    : chr "16/08/2019" "17/08/2019" "17/08/2019" "17/08/2019" ...
 $ HomeTeam: chr "Ath Bilbao" "Celta" "Valencia" "Mallorca" ...
 $ AwayTeam: chr "Barcelona" "Real Madrid" "Sociedad" "Eibar" ...
 $ FTHG    : int 1 1 1 2 0 4 1 0 1 1 ...
 $ FTAG    : int 0 3 1 1 1 4 0 2 2 0 ...
 $ FTR     : chr "H" "A" "D" "H" ...
```

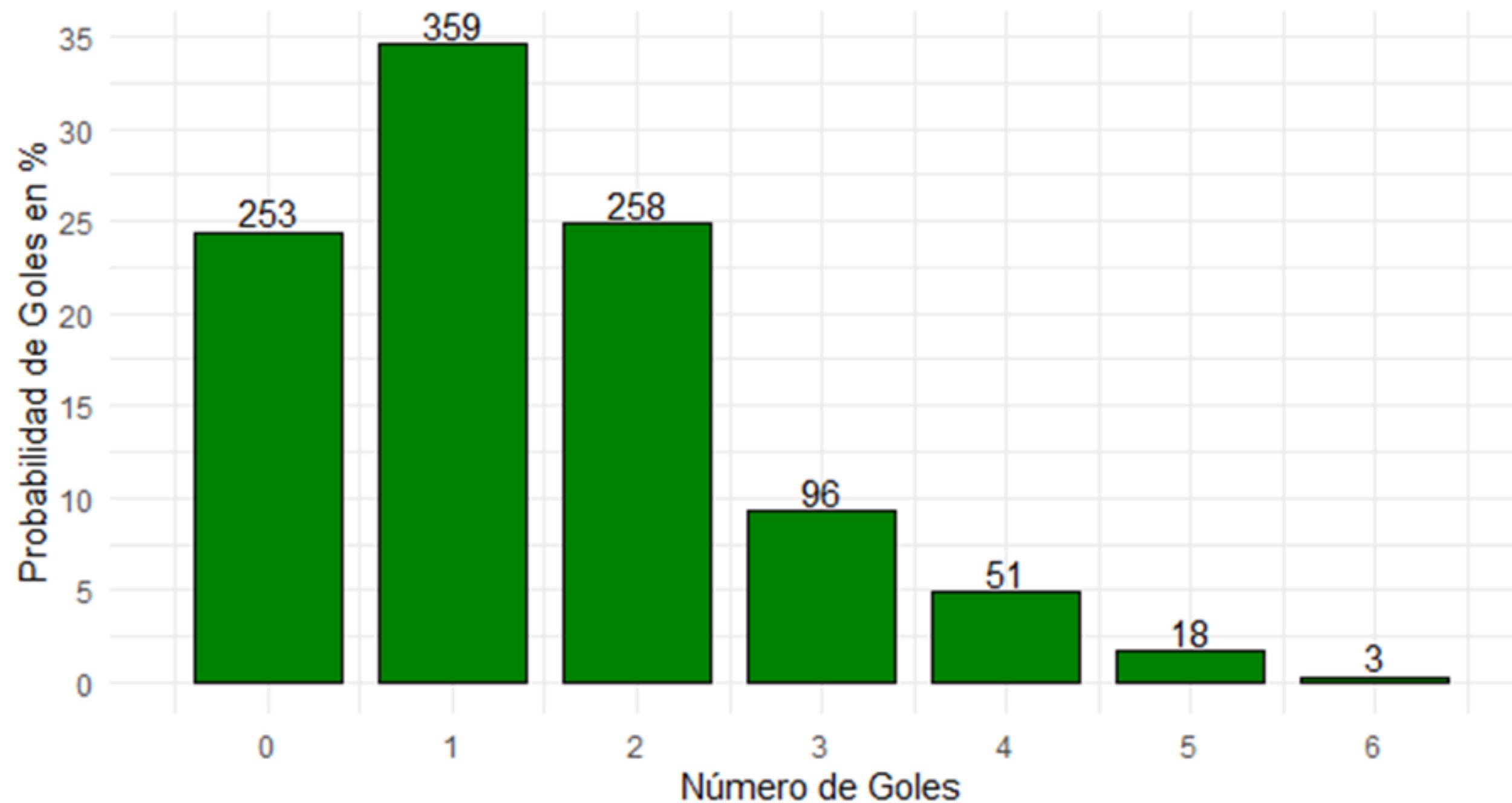
Análisis Exploratorio de Datos

Goles del Equipo de Casa

	Goles	Frecuencia Absoluta	Frecuencias_Relativa	Frecuencia Marginal
1	0	253	0.243737958	24.37%
2	1	359	0.345857418	34.59%
3	2	258	0.248554913	24.86%
4	3	96	0.092485549	9.25%
5	4	51	0.049132948	4.91%
6	5	18	0.017341040	1.73%
7	6	3	0.002890173	0.29%

Tablas de Frecuencia Absoluta y Relativa y Probabilidad Marginal . Datos del 2019-2022

Relación Goles y Probabilidad Goles CASA



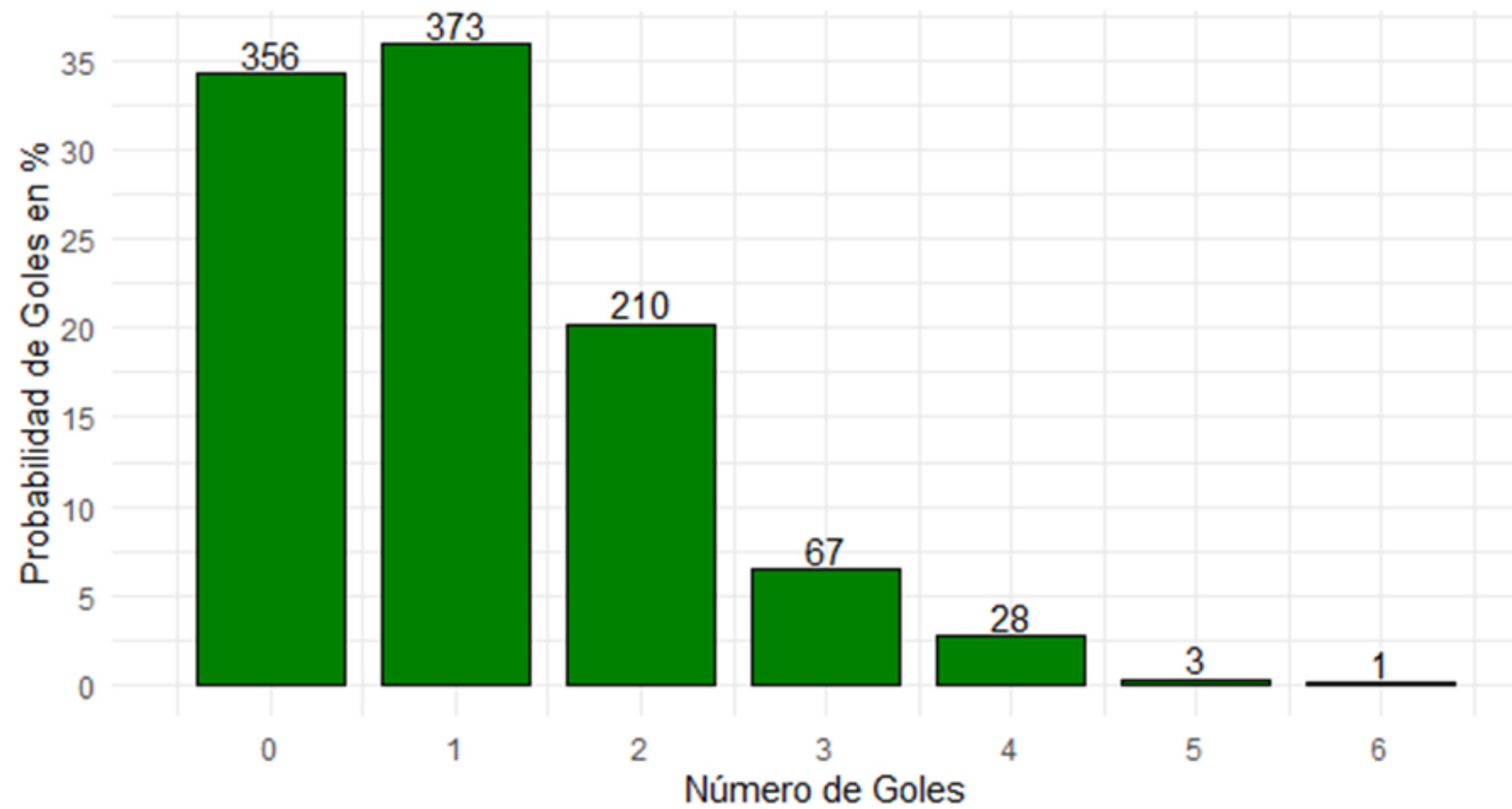
Visualización Probabilidad de Goles del Equipo de Casa.
Datos del 2019-2022

Goles del Equipo de Visita

	Goles	Frecuencia Absoluta	Frecuencias_Relativa	Probabilidad Marginal
1	0	356	0.3429672447	34.3%
2	1	373	0.3593448940	35.93%
3	2	210	0.2023121387	20.23%
4	3	67	0.0645472062	6.45%
5	4	28	0.0269749518	2.7%
6	5	3	0.0028901734	0.29%
7	6	1	0.0009633911	0.1%

Tablas de Frecuencia Absoluta y Relativa y Probabilidad Marginal . Datos del 2019-2022

Relación Goles y Probabilidad Goles VISITA



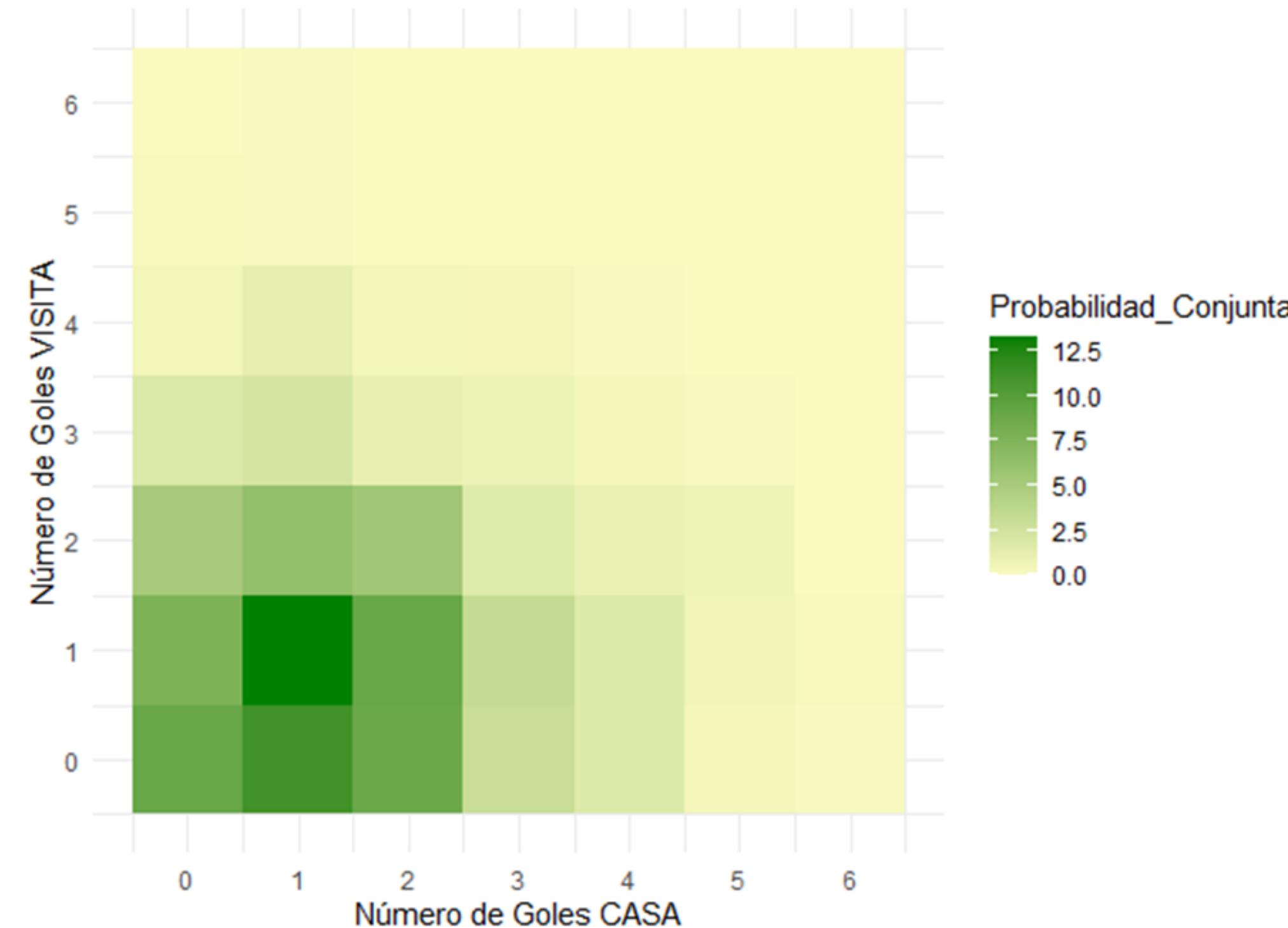
Visualización Probabilidad de Goles del Equipo de Visita.
Datos del 2019-2022

Goles del Equipo de Casa y Visita

	Goles Casa	Goles Visita	Frecuencia Absoluta	Frecuencia Relativa	Probabilidad Conjunta
9	1	1	138	0.1329479769	13.29%
8	1	0	116	0.1117533719	11.18%
1	0	0	94	0.0905587669	9.06%
16	2	1	94	0.0905587669	9.06%
15	2	0	91	0.0876685934	8.77%
2	0	1	80	0.0770712909	7.71%
10	1	2	66	0.0635838150	6.36%
17	2	2	57	0.0549132948	5.49%
3	0	2	53	0.0510597303	5.11%
23	3	1	35	0.0337186898	3.37%
22	3	0	32	0.0308285164	3.08%
11	1	3	24	0.0231213873	2.31%
4	0	3	19	0.0183044316	1.83%
29	4	0	18	0.0173410405	1.73%
30	4	1	18	0.0173410405	1.73%
24	3	2	17	0.0163776493	1.64%
12	1	4	13	0.0125240848	1.25%

Tablas de Frecuencia Absoluta y Relativa y Probabilidad
Conjunta. Datos del 2019-2022

Relación Goles CASA y VISITA, con Probabilidad



Visualización Probabilidad de Goles del Equipo de Visita
y Casa, relacionados con su Probabilidad Conjunta. Datos
del 2019-2022

Regresión lineal y clasificación

Regresión lineal y clasificación

	Date	HomeTeam	FTHG	AwayTeam	FTAG	
1	2019-08-16	Ath Bilbao		1	Barcelona	0
2	2019-08-17	Celta		1	Real Madrid	3
3	2019-08-17	Valencia		1	Sociedad	1
4	2019-08-17	Mallorca		2	Eibar	1
5	2019-08-17	Leganes		0	Osasuna	1
6	2019-08-17	Villarreal		4	Granada	4
7	2019-08-18	Alaves		1	Levante	0
8	2019-08-18	Espanol		0	Sevilla	2
9	2019-08-18	Betis		1	Valladolid	2
10	2019-08-18	Ath Madrid		1	Getafe	0
11	2019-08-23	Granada		0	Sevilla	1
12	2019-08-23	Levante		2	Villarreal	1
13	2019-08-24	Osasuna		0	Eibar	0
14	2019-08-24	Real Madrid		1	Valladolid	1
15	2019-08-24	Celta		1	Valencia	0
16	2019-08-24	Getafe		1	Ath Bilbao	1
17	2019-08-25	Alaves		0	Espanol	0
18	2019-08-25	Mallorca		0	Sociedad	1

Showing 1 to 18 of 1,038 entries, 5 total columns

Data Frame: "SmallData"

```
> (anotaciones <- listasoccer$scores)
```

	date	home.team	home.score	away.team	away.score
1	2019-08-16	Ath Bilbao	1	Barcelona	0
2	2019-08-17	Celta	1	Real Madrid	3
3	2019-08-17	Valencia	1	Sociedad	1
4	2019-08-17	Mallorca	2	Eibar	1
5	2019-08-17	Leganes	0	Osasuna	1
6	2019-08-17	Villarreal	4	Granada	4
7	2019-08-18	Alaves	1	Levante	0
8	2019-08-18	Espanol	0	Sevilla	2
9	2019-08-18	Betis	1	Valladolid	2
10	2019-08-18	Ath Madrid	1	Getafe	0
11	2019-08-23	Granada	0	Sevilla	1
12	2019-08-23	Levante	2	Villarreal	1
13	2019-08-24	Osasuna	0	Eibar	0
14	2019-08-24	Real Madrid	1	Valladolid	1
15	2019-08-24	Celta	1	Valencia	0
16	2019-08-24	Getafe	1	Ath Bilbao	1
17	2019-08-25	Alaves	0	Espanol	0
18	2019-08-25	Mallorca	0	Sociedad	1
19	2019-08-25	Leganes	0	Ath Madrid	1
20	2019-08-25	Barcelona	5	Betis	2
21	2019-08-30	Sevilla	1	Celta	1
22	2019-08-30	Ath Bilbao	2	Sociedad	0
23	2019-08-31	Osasuna	2	Barcelona	2
24	2019-08-31	Getafe	1	Alaves	1
25	2019-08-31	Levante	2	Valladolid	0
26	2019-08-31	Betis	2	Leganes	1

Lista Anotaciones

```
> equipos <- listasoccer$teams
      name
 1    Alaves
 2 Ath Bilbao
 3 Ath Madrid
 4 Barcelona
 5    Betis
 6    Cadiz
 7    Celta
 8    Eibar
 9    Elche
10   Espanol
11   Getafe
12   Granada
13   Huesca
14   Leganes
15   Levante
16   Mallorca
17   Osasuna
18 Real Madrid
19   Sevilla
20   Sociedad
21   Valencia
22 Valladolid
```

Lista Equipos

Team Rankings based on matches 16/08/2019 to 12/03/2022

	team	total	attack	defense	n.games.Var1	n.games.Freq
1	Real Madrid	1.54	1.77	1.64	Real Madrid	103
2	Barcelona	1.22	2.06	1.13	Barcelona	102
3	Ath Madrid	1.14	1.55	1.42	Ath Madrid	104
4	Sevilla	0.96	1.34	1.45	Sevilla	103
5	Villarreal	0.72	1.58	1.04	Villarreal	104
6	Sociedad	0.53	1.34	1.08	Sociedad	103
7	Ath Bilbao	0.39	1.11	1.19	Ath Bilbao	103
8	Betis	0.24	1.39	0.85	Betis	103
9	Valencia	0.14	1.29	0.85	Valencia	104
10	Celta	0.09	1.17	0.91	Celta	104
11	Vallecano	0.05	1.04	0.99	Vallecano	26
12	Getafe	0.02	0.89	1.14	Getafe	104
13	Granada	-0.02	1.20	0.82	Granada	104
14	Osasuna	-0.04	1.03	0.95	Osasuna	103
15	Levante	-0.19	1.16	0.76	Levante	104
16	Elche	-0.28	0.95	0.87	Elche	66
17	Valladolid	-0.35	0.84	0.93	Valladolid	76
18	-----	-----	-----	-----	-----	20

Tabla Ranking

```
> predict(ranking, date = fecha[n])
Predicted Match Results for 01/05/1900 to 01/06/2100
Model based on data from 16/08/2019 to 12/03/2022
-----
13/03/2022 Vallecano vs Sevilla, HW 20%, AW 52%, T 28%, pred score 0.7-1.4 actual: T (1-1)
13/03/2022 Betis vs Ath Bilbao, HW 33%, AW 39%, T 27%, pred score 1.2-1.3 actual: HW (1-0)
13/03/2022 Sociedad vs Alaves, HW 58%, AW 18%, T 24%, pred score 1.7-0.8 actual: HW (1-0)
13/03/2022 Barcelona vs Osasuna, HW 66%, AW 15%, T 19%, pred score 2.2-0.9 actual: HW (4-0)
```

Predicciones

1035	2022-03-13	Vallecano	1	Sevilla	1
1036	2022-03-13	Betis	1	Ath Bilbao	0
1037	2022-03-13	Sociedad	1	Alaves	0
1038	2022-03-13	Barcelona	4	Osasuna	0

Predicciones

1035	2022-03-13	Vallecano	1	Sevilla	1
1036	2022-03-13	Betis	1	Ath Bilbao	0
1037	2022-03-13	Sociedad	1	Alaves	0
1038	2022-03-13	Barcelona	4	Osasuna	0

Predicciones

RStudio Cloud - Github, conexiones con BDs y lectura de datos externos



	COUNT(*)
▶	1140

Base de Datos, Consulta, Tablas

	Date	HomeTeam	AwayTeam	FTHG	FTAG
▶	2020-06-14	Real Madrid	Eibar	3	1

	Date	HomeTeam	AwayTeam	FTHG	FTAG
▶	2020-02-09	Osasuna	Real Madrid	1	4

Partidos Con mejores Resultados 2020

