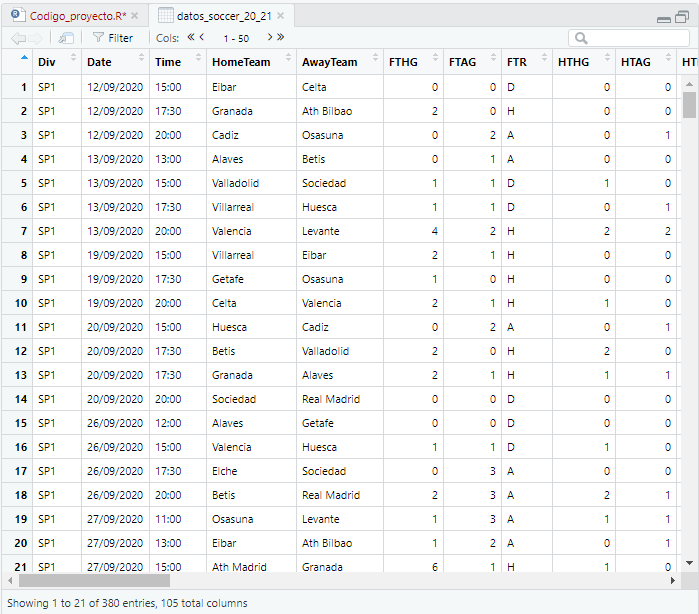
**ANÁLISIS DE LOS EQUIPOS DE LA LIGA ESPAÑOLA DE FÚTBOL**

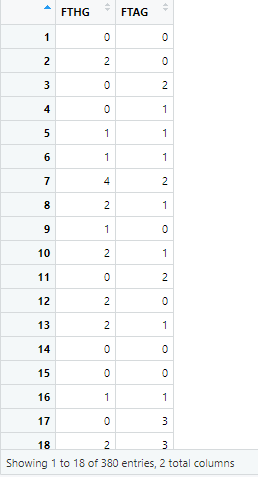
# Introducción a R y Software

1. Se importan los datos de soccer de la temporada 2020/2021 de la primera división de la liga española.
2. Se descarga la base de datos, el día 03 de marzo de 2022, desde la página: <https://www.football-data.co.uk/spainm.php>
3. Obtenidos los datos se proceden a abrir desde RStudio.



1. Se extraen las columnas que contienen los números de goles anotados por los equipos que jugaron en casa y los que jugaron como visitantes.





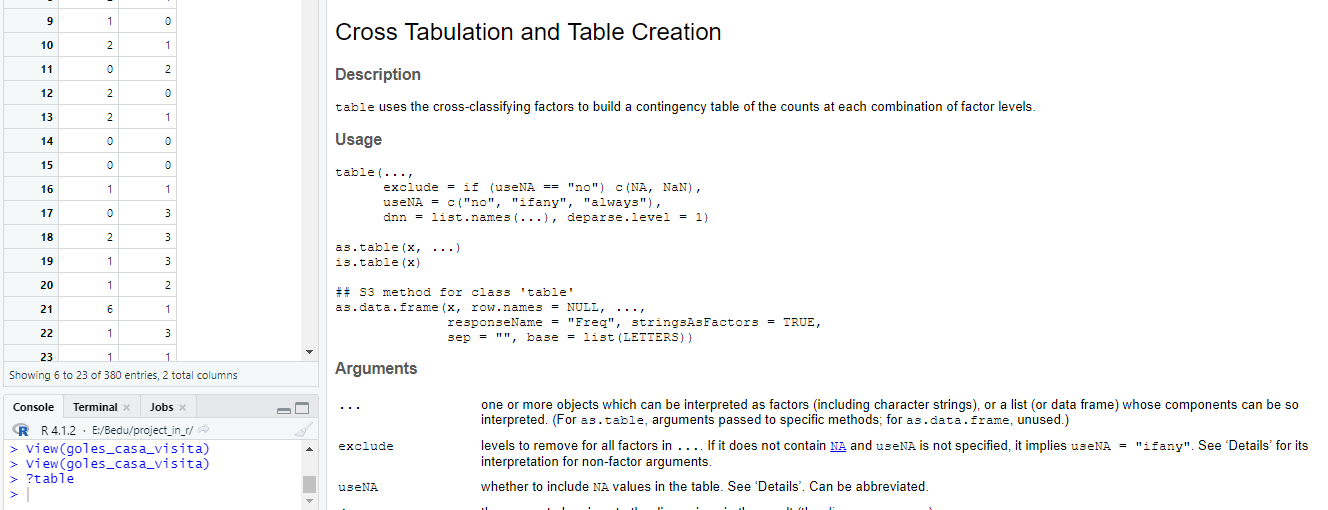
1. Consulta table en R
2. Se elabora tablas de frecuencias relativas para estimar las siguientes probabilidades:



Tabla de frecuencias Absoluta, Relativa y Marginal de los Goles de equipo de casa

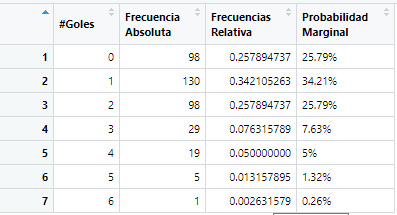
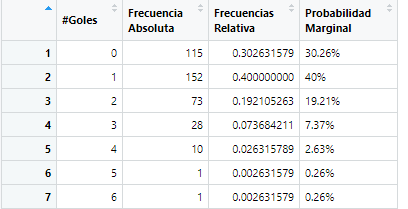
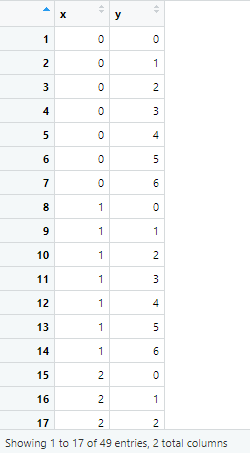


Tabla de frecuencias Absoluta, Relativa de los Goles de equipo de visita



1. Tablas de Probabilidad Conjunta de goles del equipo de casa y visitantes

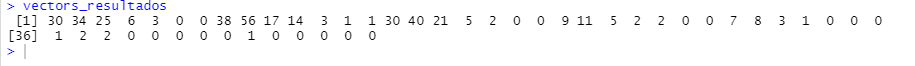
Primero se crea una tabla con los posibles resultados:



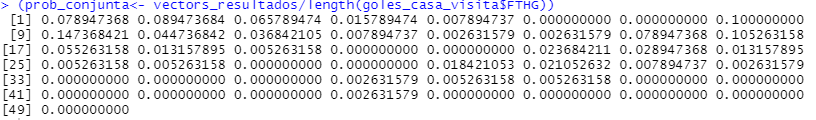
Luego se cuenta en número de casos favorables en las que se cumple ambos casos de la tabla anterior



El resultado es un vector que contiene el conteo de cada caso favorable

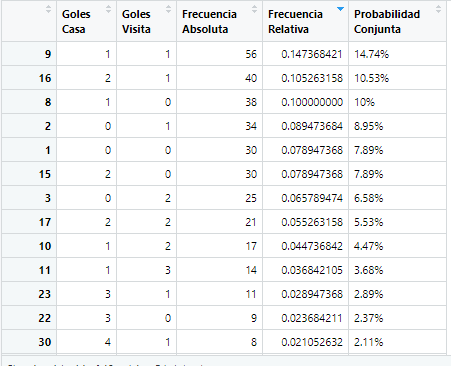


Se calcula la probabilidad de cada caso favorable en la que se cumplieron los casos propuestos, guardada en un vector.



Se crea la tabla de resultados y se le da formato adecuado



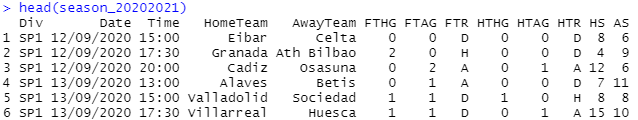


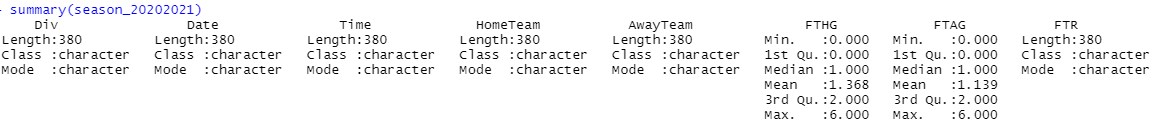
# Programación y manipulación de datos en R

1. Se importan los datos de soccer de la temporada 2019/2020, 2020/2021 y 2021/2022 de la primera división de la liga española.
2. Se descarga la base de datos, el día 03 de marzo de 2022, desde la página: <https://www.football-data.co.uk/spainm.php>
3. Obtenidos los datos se proceden a abrir desde RStudio y se revisan las estructuras de las tablas creadas con las funciones: funciones: str, head, View y summary.

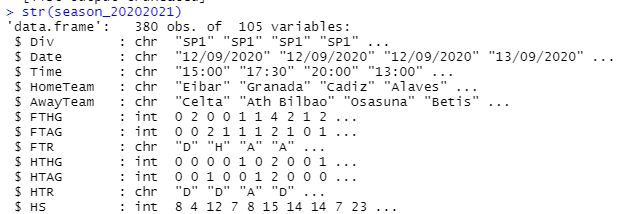


Vistas (View ()):

Primero 6 datos (head ()):

Resumen (sumary()):

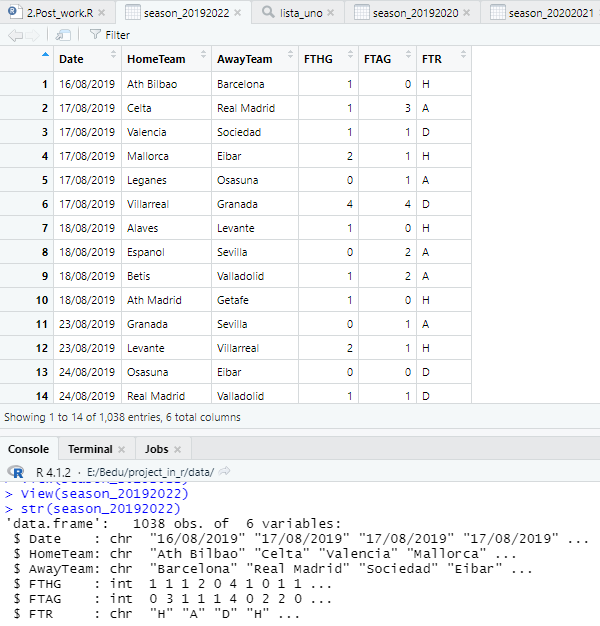
Estructura (str())



1. Tabla con las columnas Date, HomeTeam, AwayTeam, FTHG, FTAG y FTR. Se juntan las tres tablas.



1. Se Ajusta el tipo de dato de la columna fecha, para tener el tipo correcto.



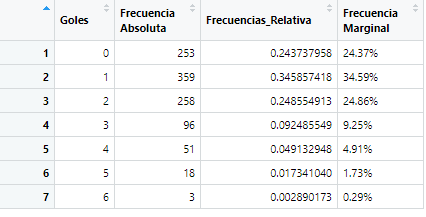
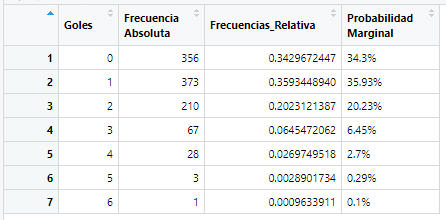
Nota: Con el propósito de documentar la forma rápida de hacerlo se agrega el código usando las funciones resumidas, con la intención de mostrar las dos formas de realizar el código.

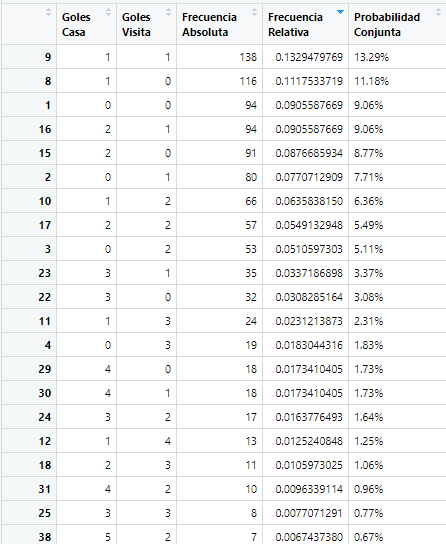


# Análisis Exploratorio de Datos (AED o EDA) con R

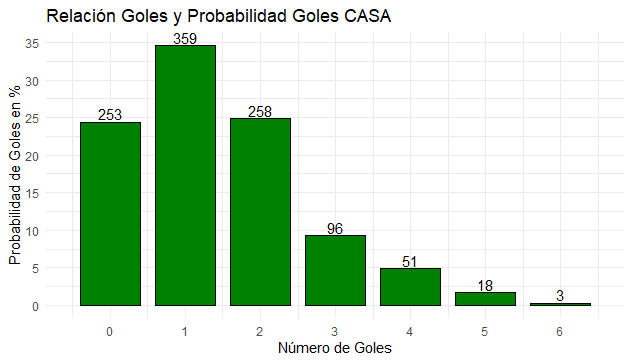
1. Código



1. Tabla de la probabilidad marginal de que el equipo que juega en casa anote goles
2. Tabla de la probabilidad marginal de que el equipo que juega de visita anote goles

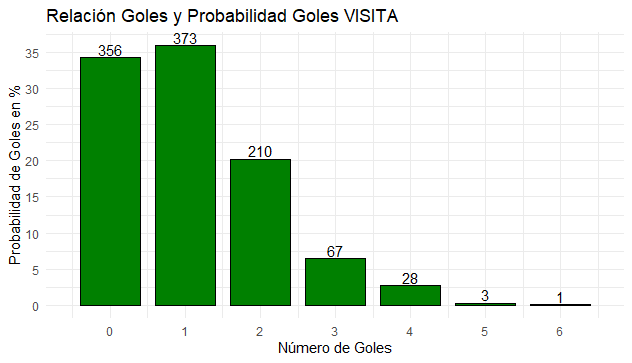
1. La probabilidad conjunta de que el equipo que juega en casa anote x goles y el equipo que juega como visitante anote goles
2. Un gráfico de barras para las probabilidades marginales estimadas del número de goles que anota el equipo de casa





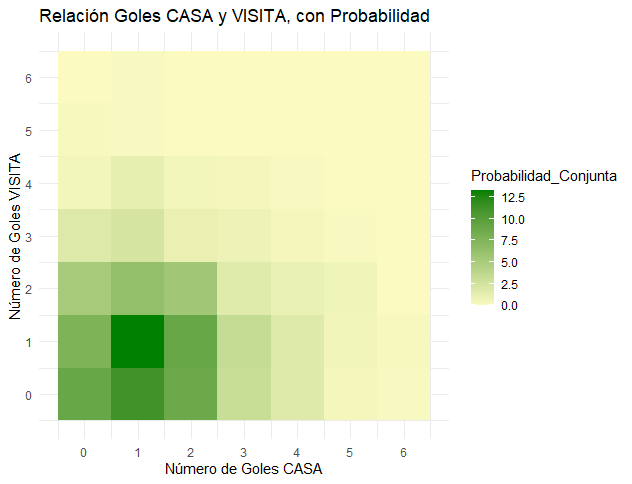
1. Un gráfico de barras para las probabilidades marginales estimadas del número de goles que anota el equipo visitante.





1. 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.





# Regresión lineal y clasificación

1. Del conjunto de datos de soccer de la liga española de las temporadas 2019/2020, 2020/2021 y 2021/2022. Se crea el data frame “SmallData”, contiene las columnas: date, home.team, home.score, away.team y away.score;.



1. Se crea el archivo soccer.csv



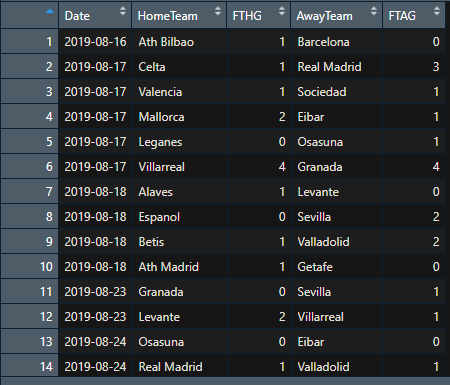
1. Con la fcreate.fbRanks.dataframes importa soccer.csv guardado en listasoccer. Asigna estos data frames a variables llamadas anotaciones y equipos.

Se crea un vector de fechas que correspondan a las fechas en las que se jugaron partidos.

Crea una variable llamada n que contenga el número de fechas diferentes.

Con la función rank.teams y “anotaciones” y “equipos”, crea un ranking de equipos usando únicamente datos desde la fecha inicial y hasta la penúltima fecha en la que se jugaron partidos, estas fechas las deberá especificar en max.date y min.date. Guarda los resultados con el nombre ranking.

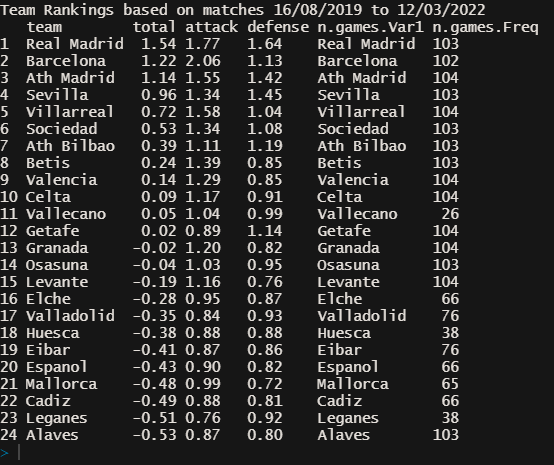


Lista anotaciones y equipos

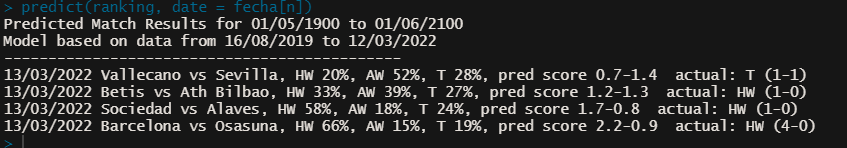


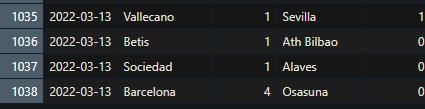


Ranking



1. Finalmente estima las probabilidades de los eventos, el equipo de casa gana, el equipo visitante gana o el resultado es un empate para los partidos que se jugaron en la última fecha del vector de fechas fecha.





# Series de tiempo

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

1. Alojar el fichero data.csv en una base de datos llamada match\_games, nombrando al collection como match
2. Una vez hecho esto, realizar un count para conocer el número de registros que se tiene en la base



1. Consulta, en la base de datos para conocer los partidos como casa y como visitante cuales fueron los dos mejores partidos



1. Agrega el dataset de mtcars a la misma BDD



# Dashboards con Shiny - Entorno GUI