***RESPUESTAS A LOS EJERCICIOS CIENCIA DE DATOS EN R***

***Todos los ejercicios se realizarán utilizando el archivo walmart\_ecommerce.csv, el cuál contiene el identificador del producto vendido, precio de lista, precio con descuento, Marca, Categoría y si el producto está disponible o no.***

**Pregunta 1.-** Obtén el diagrama de Pareto de las marcas existentes en la tienda. Recuerda ampliar el panel en el que se despliegan los gráficos, de lo contrario, R te arrojará un error

**RESPUESTA:**

#Indicando el directorio en donde se encuentra el archivo a cargar

setwd("C:/Users/rocio/Escritorio2/Pavillion-Rocio/Cursos impartidos en linea/SQL para Ciencia de Datos/Bases de datos")

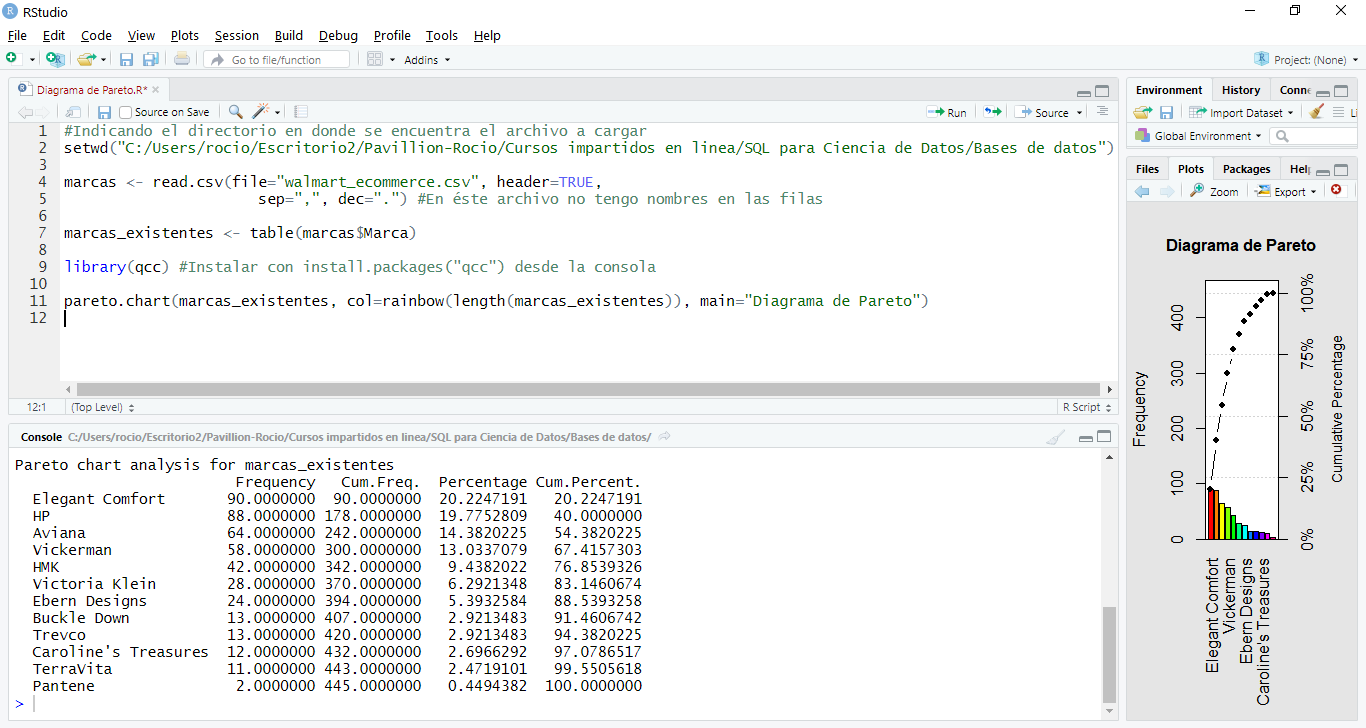
marcas <- read.csv(file="walmart\_ecommerce.csv", header=TRUE,

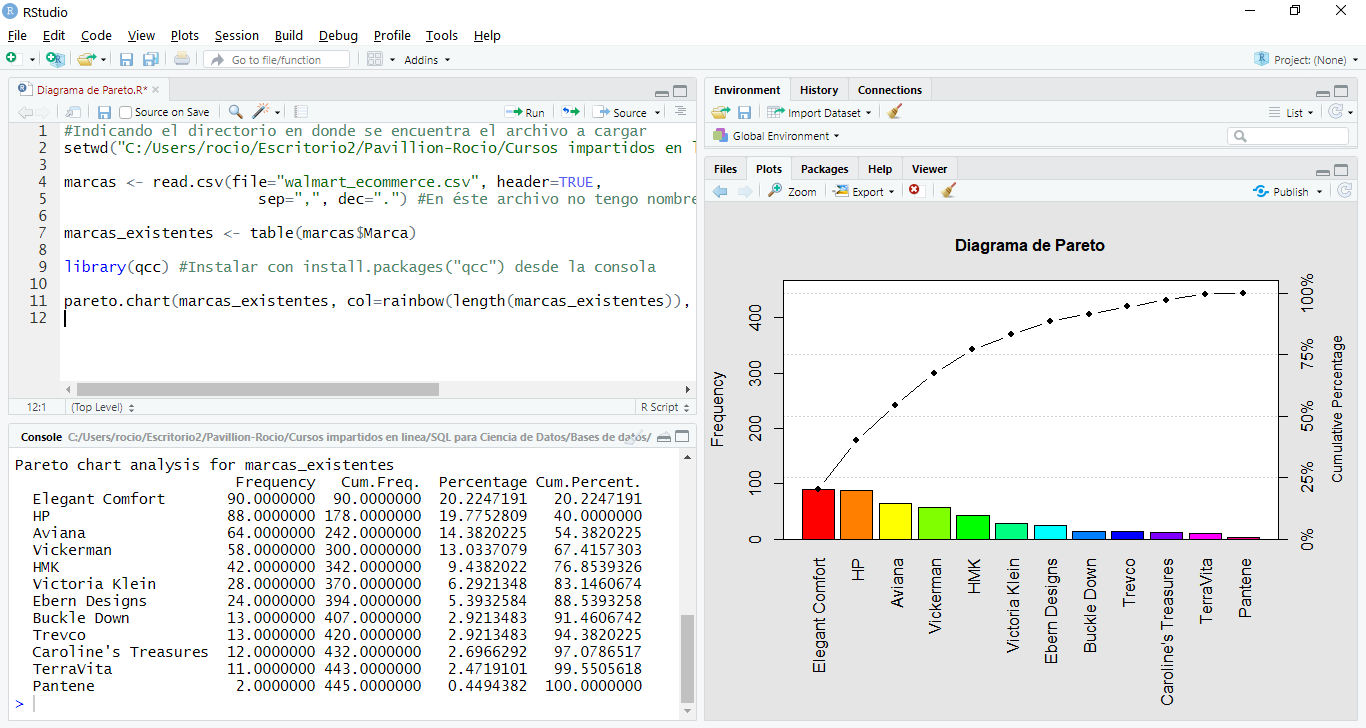
sep=",", dec=".") #En éste archivo no tengo nombres en las filas

marcas\_existentes <- table(marcas$Marca)

library(qcc) #Instalar con install.packages("qcc") desde la consola

pareto.chart(marcas\_existentes, col=rainbow(length(marcas\_existentes)), main="Diagrama de Pareto")





**Pregunta 2.-** Obtén el promedio, la mediana y la moda de los precios de lista de los productos

**RESPUESTA:**

#Indicando el directorio en donde se encuentra el archivo a cargar

setwd("C:/Users/rocio/Escritorio2/Pavillion-Rocio/Cursos impartidos en linea/SQL para Ciencia de Datos/Bases de datos")

marcas <- read.csv(file="walmart\_ecommerce.csv", header=TRUE,

sep=",", dec=".") #En éste archivo no tengo nombres en las filas

#Obtenemos el Promedio

mean(marcas$Precio\_de\_lista)

#Obtenemos la Mediana

median(marcas$Precio\_de\_lista)

#Obtenemos la Moda

table(marcas$Precio\_de\_lista)

frecuencias <- data.frame(table(marcas$Precio\_de\_lista))

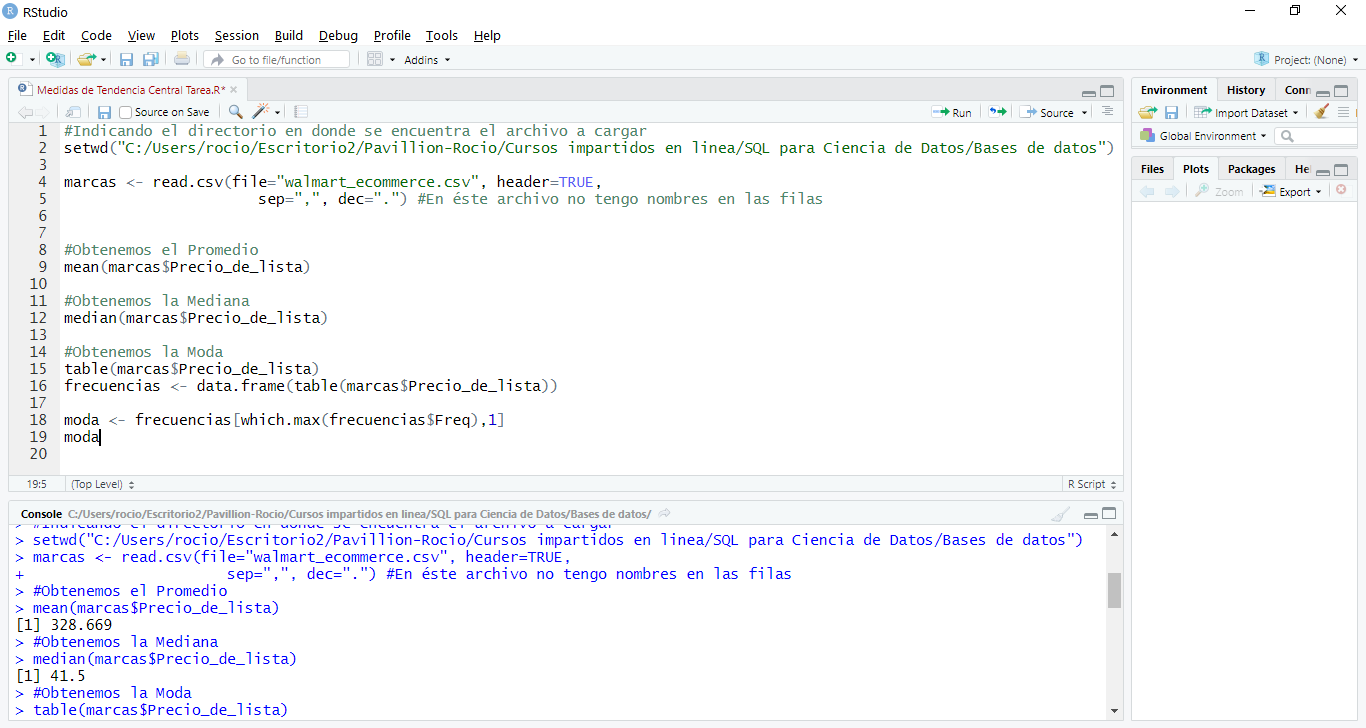
moda <- frecuencias[which.max(frecuencias$Freq),1]

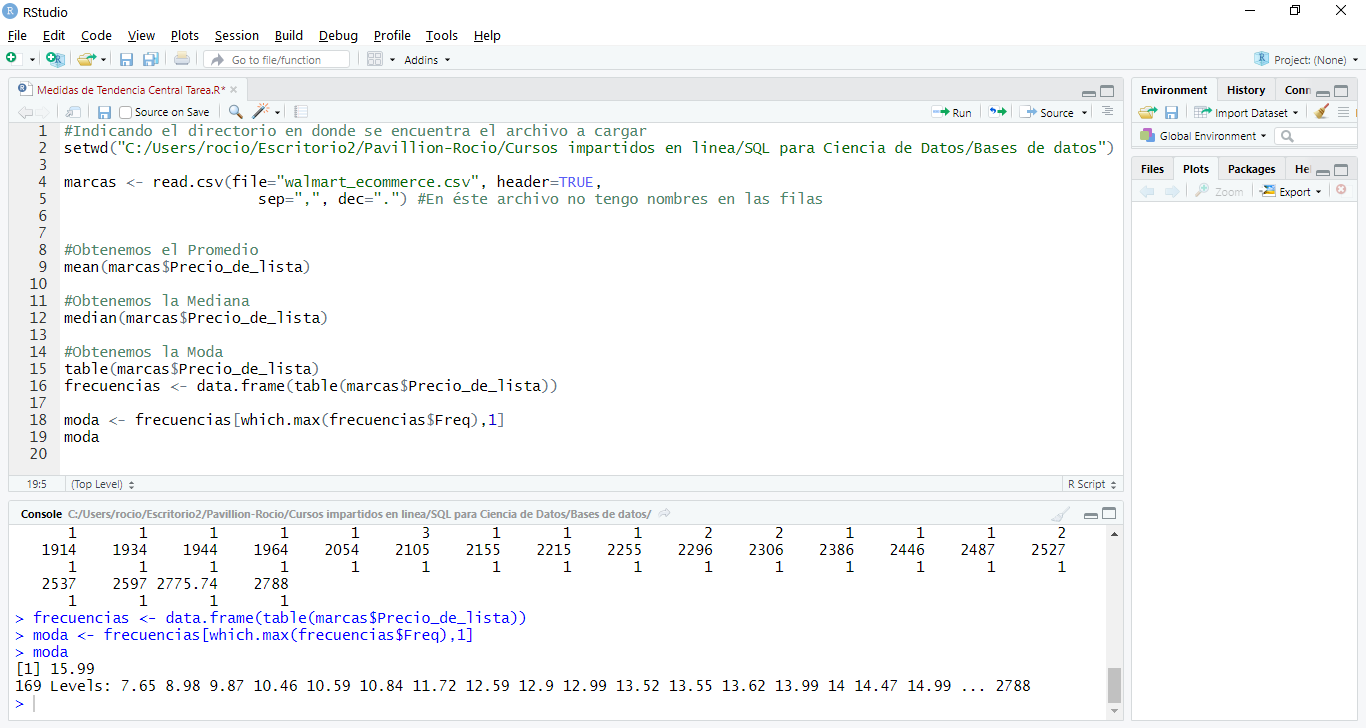
moda

**Promedio: 328.66**

**Mediana: 41.5**

**Moda: 15.99**





**Pregunta 3.-** Obtén el rango, la varianza y la desviación estándar de los precios de lista de los productos

**RESPUESTA:**

#Indicando el directorio en donde se encuentra el archivo a cargar

setwd("C:/Users/rocio/Escritorio2/Pavillion-Rocio/Cursos impartidos en linea/SQL para Ciencia de Datos/Bases de datos")

marcas <- read.csv(file="walmart\_ecommerce.csv", header=TRUE,

sep=",", dec=".") #En éste archivo no tengo nombres en las filas

#Obtenemos el Rango

rango = max(marcas$Precio\_de\_lista) - min(marcas$Precio\_de\_lista)

rango

#Varianza

var(marcas$Precio\_de\_lista, na.rm = TRUE)

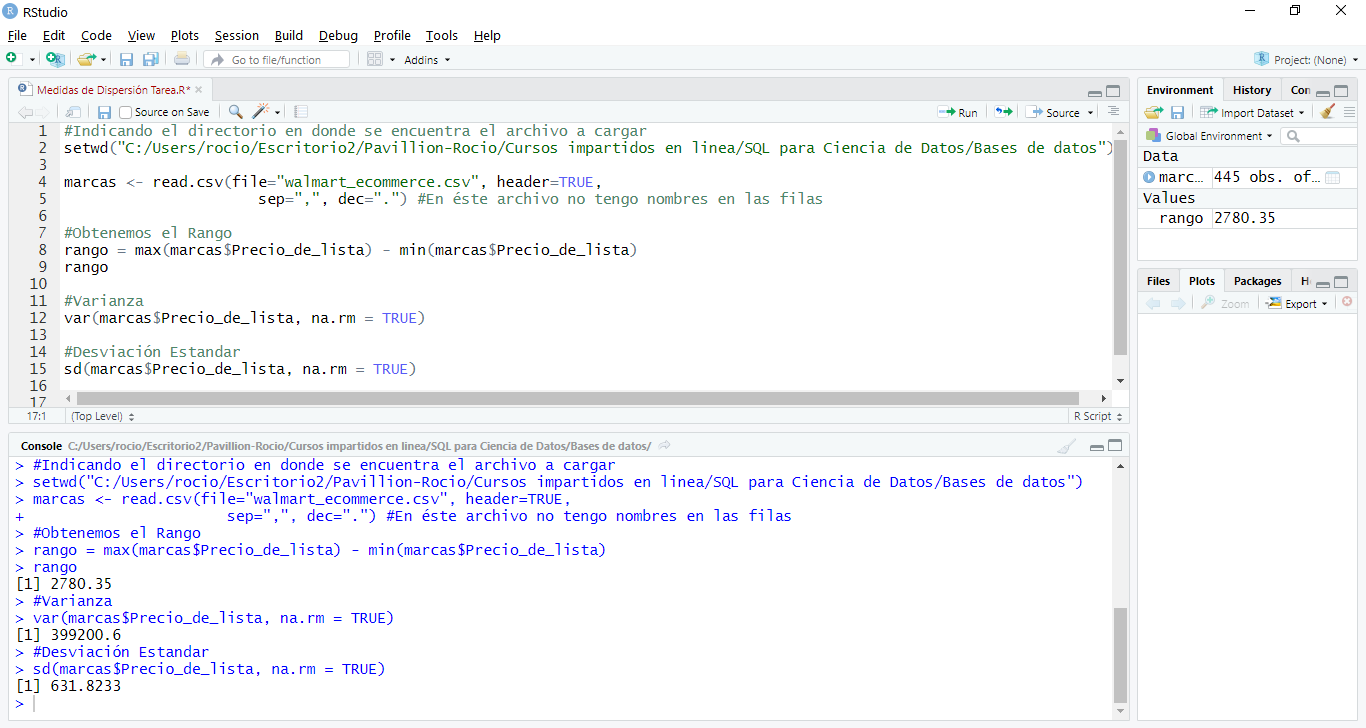
#Desviación Estandar

sd(marcas$Precio\_de\_lista, na.rm = TRUE)

**Rango: 2780.35**

**Varianza: 399200.6**

**Desviación Estándar: 631.8233**



**Pregunta 4.-** Crea un histograma de los precios de lista, que contenga 10 divisiones, utilizando la librería ggplot2

**RESPUESTA:**

#Indicando el directorio en donde se encuentra el archivo a cargar

setwd("C:/Users/rocio/Escritorio2/Pavillion-Rocio/Cursos impartidos en linea/SQL para Ciencia de Datos/Bases de datos")

marcas <- read.csv(file="walmart\_ecommerce.csv", header=TRUE,

sep=",", dec=".") #En éste archivo no tengo nombres en las filas

#Histograma con graficos de ggplot2

library(ggplot2)

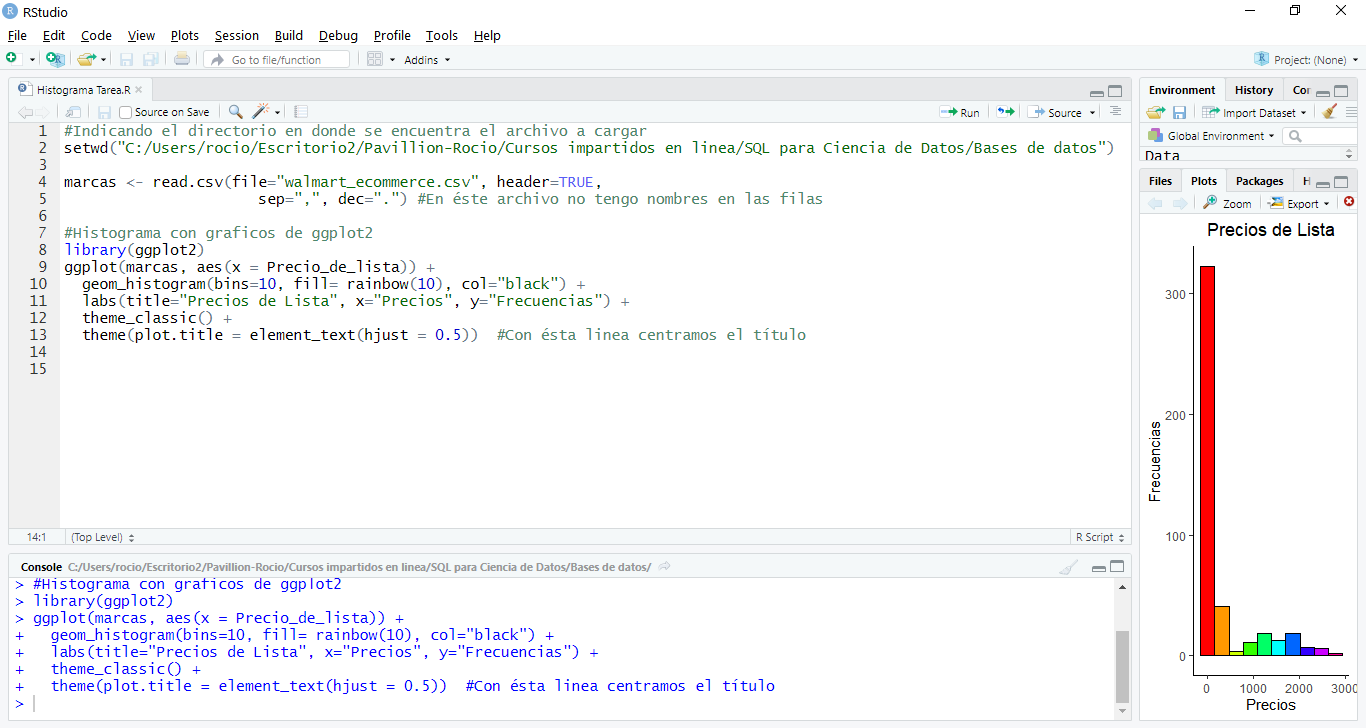
ggplot(marcas, aes(x = Precio\_de\_lista)) +

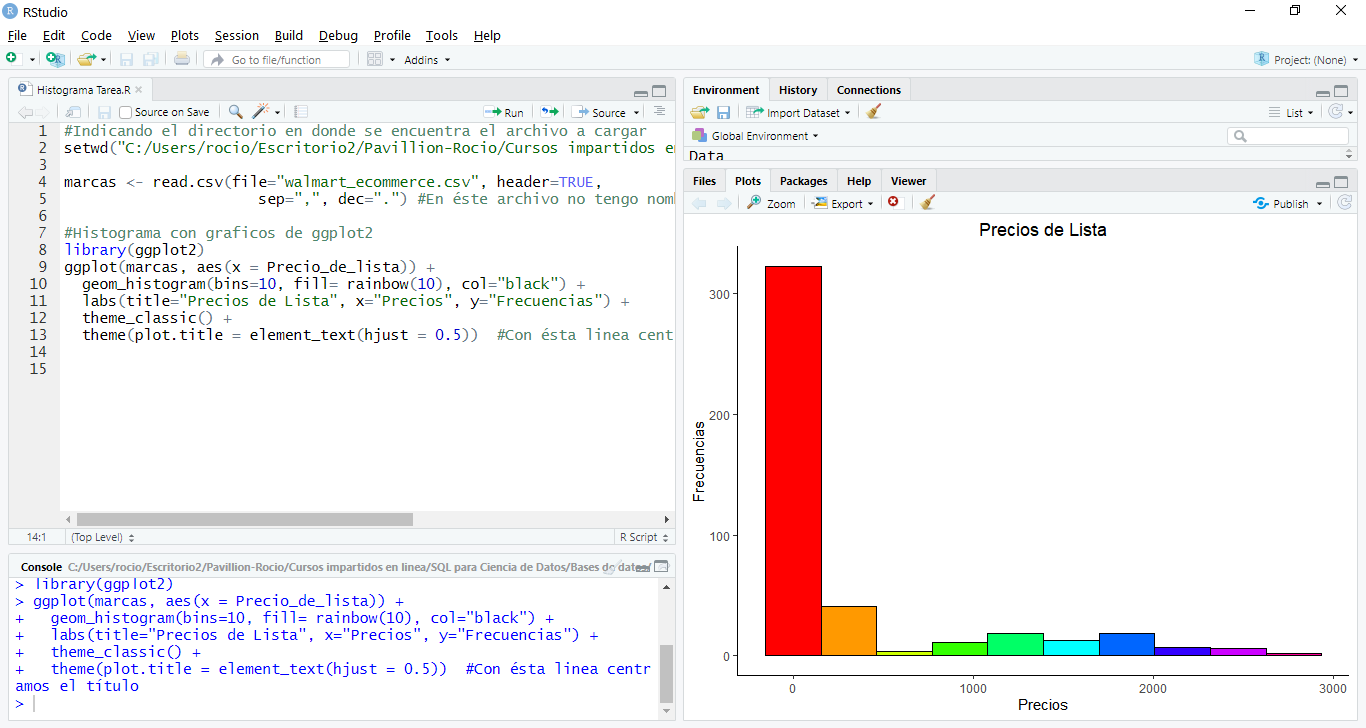
geom\_histogram(bins=10, fill= rainbow(10), col="black") +

labs(title="Precios de Lista", x="Precios", y="Frecuencias") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5)) #Con ésta linea centramos el título





**Pregunta 5.-** Genera un gráfico de caja y bigotes de los precios de lista. En caso de que existan outliers, crea una lista ordenada de los mismos

**RESPUESTA:**

#Indicando el directorio en donde se encuentra el archivo a cargar

setwd("C:/Users/rocio/Escritorio2/Pavillion-Rocio/Cursos impartidos en linea/SQL para Ciencia de Datos/Bases de datos")

marcas <- read.csv(file="walmart\_ecommerce.csv", header=TRUE,

sep=",", dec=".")

#Creamos el gráfico de Caja y Bigotes

boxplot(marcas$Precio\_de\_lista, horizontal=TRUE)

#Obtenemos la lista de los valores que aparecen como outliers

boxplot.stats(marcas$Precio\_de\_lista)

precios\_stats <- boxplot.stats(marcas$Precio\_de\_lista) #Guardas los estadísticos en precios\_stats

ordered\_outliers <- sort(precios\_stats$out) #Ordenas los outliers en forma ascendente

ordered\_outliers #Le pides que te los muestre

