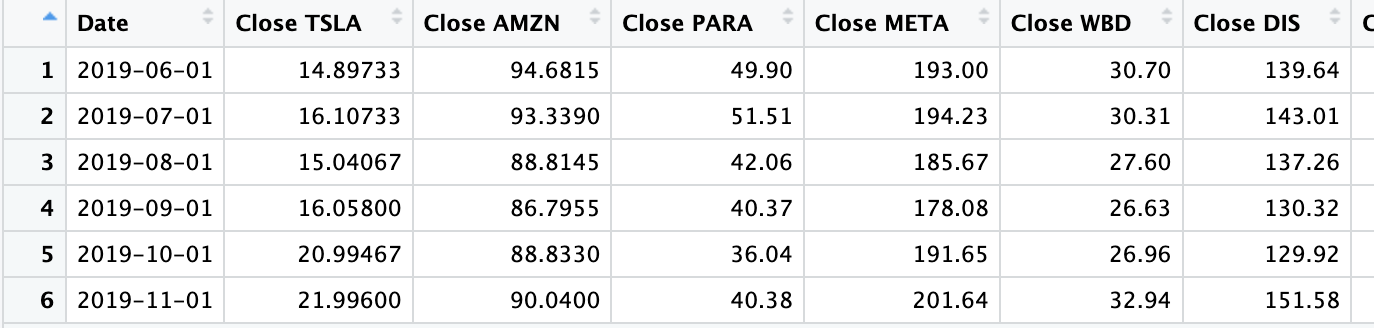
La figura 1.1 muestra reporte donde se han resaltado los valores faltantes, la calificación de servicio para la observación 2 y la calificación de limpieza para la observación 13 están marcados como TRUE porque no están presentes, El resto de los valores en la figura están etiquetados como FALSE porque están presentes

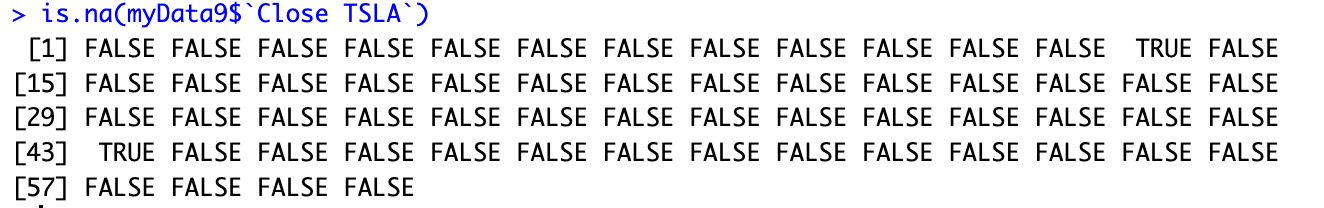
Base de datos cargado en R para llevar a cabo la limpieza de datos



El comando para llevar a cabo en R la detección de valores faltantes

is.na(myData9$`Close TSLA`)

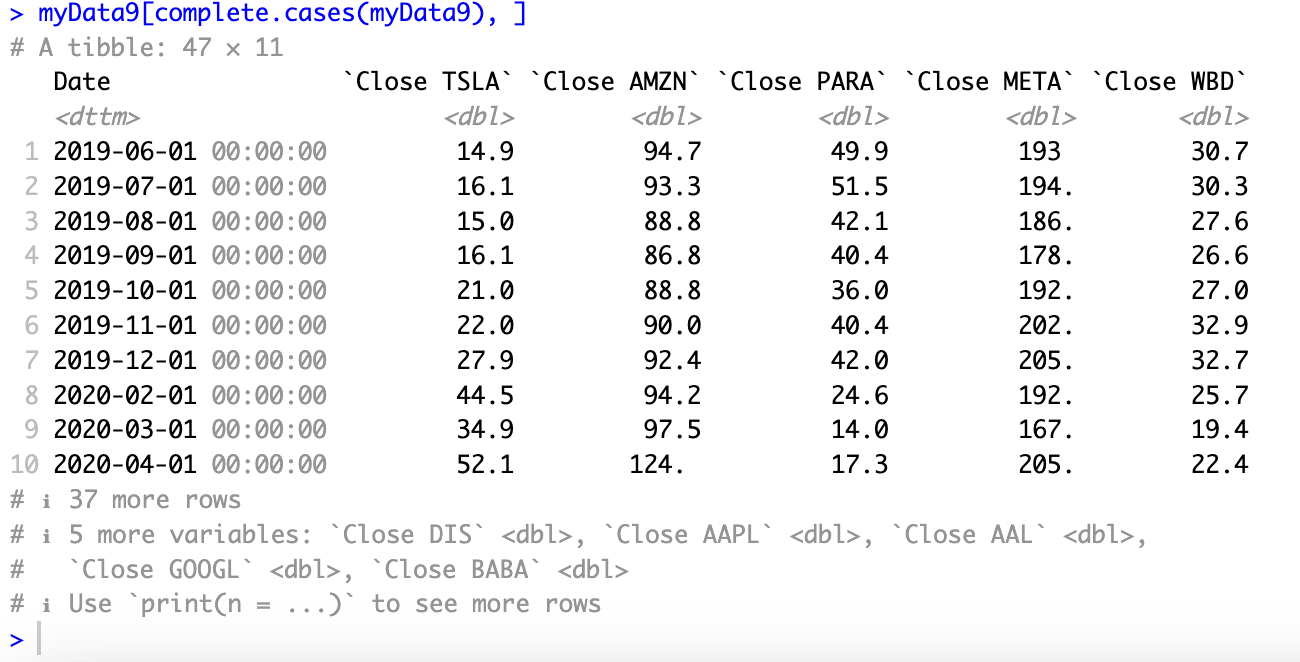
FIGURA 1



Para identificar las filas en el marco de datos o los casos que están completos podemos usar la función:

myData9[complete.cases(myData9), ]

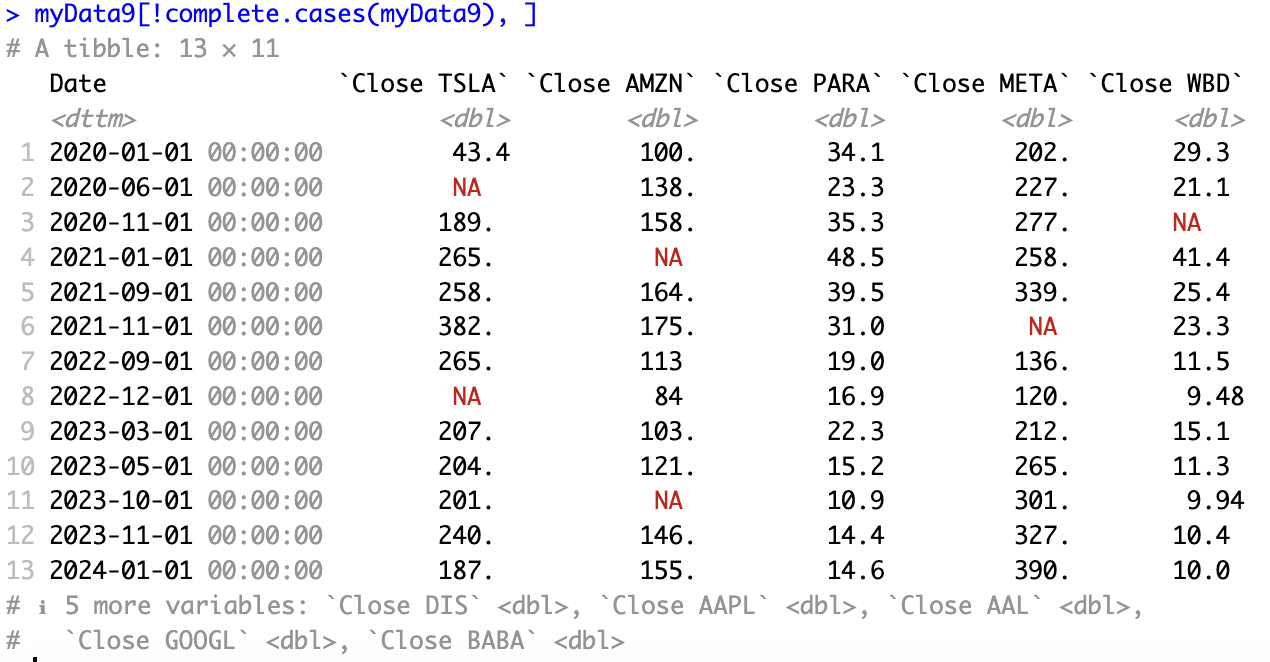
FIGURA 2



Cuando los datos están casi completos, es posible que no sea conveniente enlistar todos lo casos completos, para ello podemos usar el operador not (signo ¡) con la función complete.cases que servirá para identificar observaciones con valores faltantes, la función será así:

myData9[!complete.cases(myData9), ]

FIGURA 3

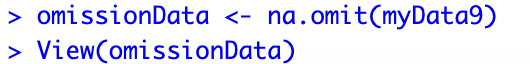


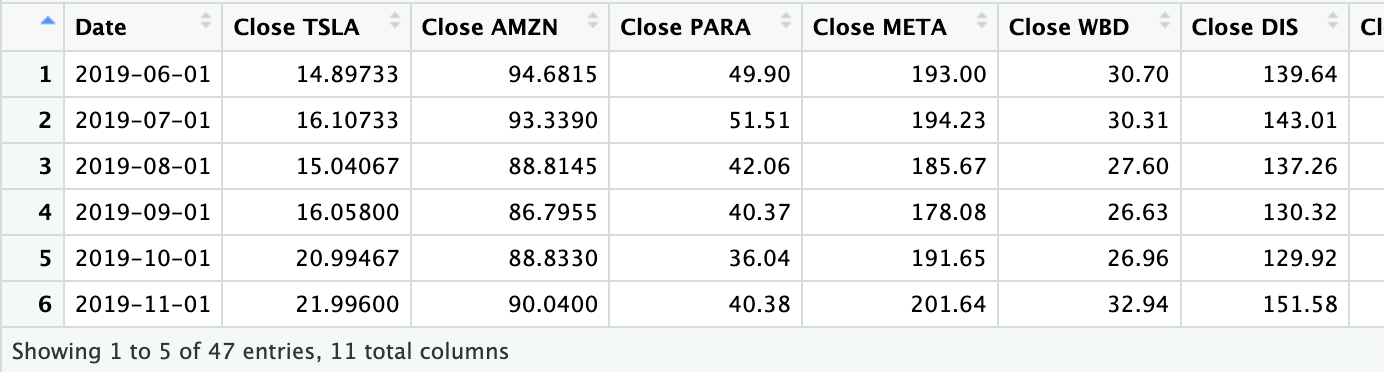
Para implementar la estrategia de omisión utilizaremos la función na.omit con el objetivo de eliminar las observaciones con valores faltantes y almacenar el conjunto de datos resultante en el marco de datos omissionData, para ello la función será así:

omissionData <- na.omit(myData9)

> View(omissionData)

FIGURA 4



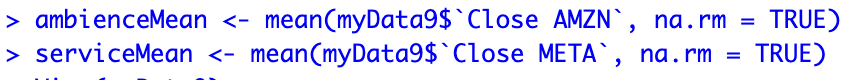


R ha creado un nuevo marco de datos, este conjunto de datos está listo para el análisis de datos.

Para implementar la estrategia de imputación de la media simple, se iniciará con el marco original, se calcula el valor promedio utilizando mean, para calcular los valores promedio de las variables AMZN y META ingresar:

ambienceMean <- mean(myData9$`Close AMZN`, na.rm = TRUE)

> serviceMean <- mean(myData9$`Close META`, na.rm = TRUE)





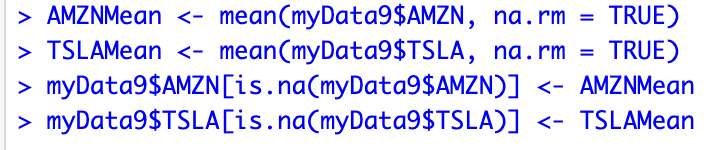
Para imputar los valores faltantes en las variables AMZN y META, de nuevo utilizamos la función is.na para identificar los valores faltantes y reemplazarlos con las medias calculadas en el paso, ingresar:

AMZNMean <- mean(myData9$AMZN, na.rm = TRUE)

> TSLAMean <- mean(myData9$TSLA, na.rm = TRUE)

> myData9$AMZN[is.na(myData9$AMZN)] <- AMZNMean

> myData9$TSLA[is.na(myData9$TSLA)] <- TSLAMean





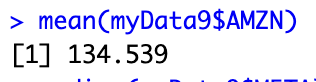


Medidas de Posición en R

Para obtener la mediana y el promedio en R, es necesario ejecutar estos comandos

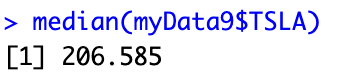
Promedio:

mean(myData9$AMZN)



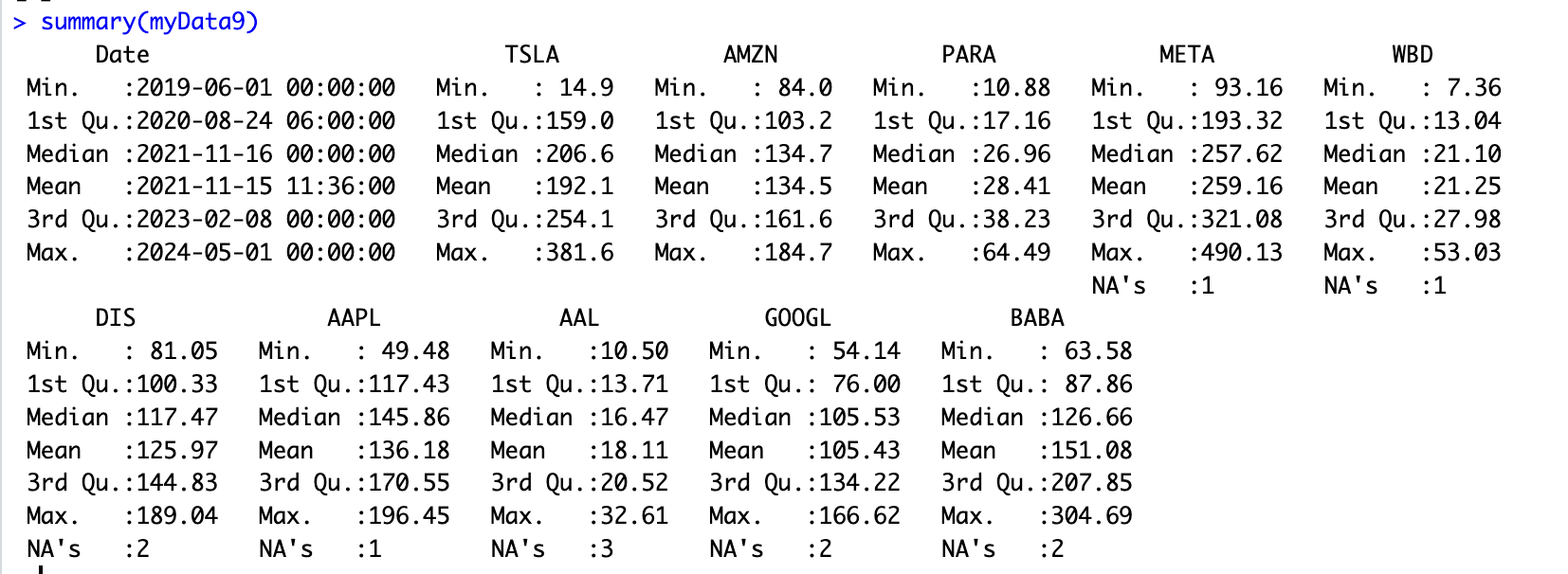
Mediana:

median(myData9$TSLA)

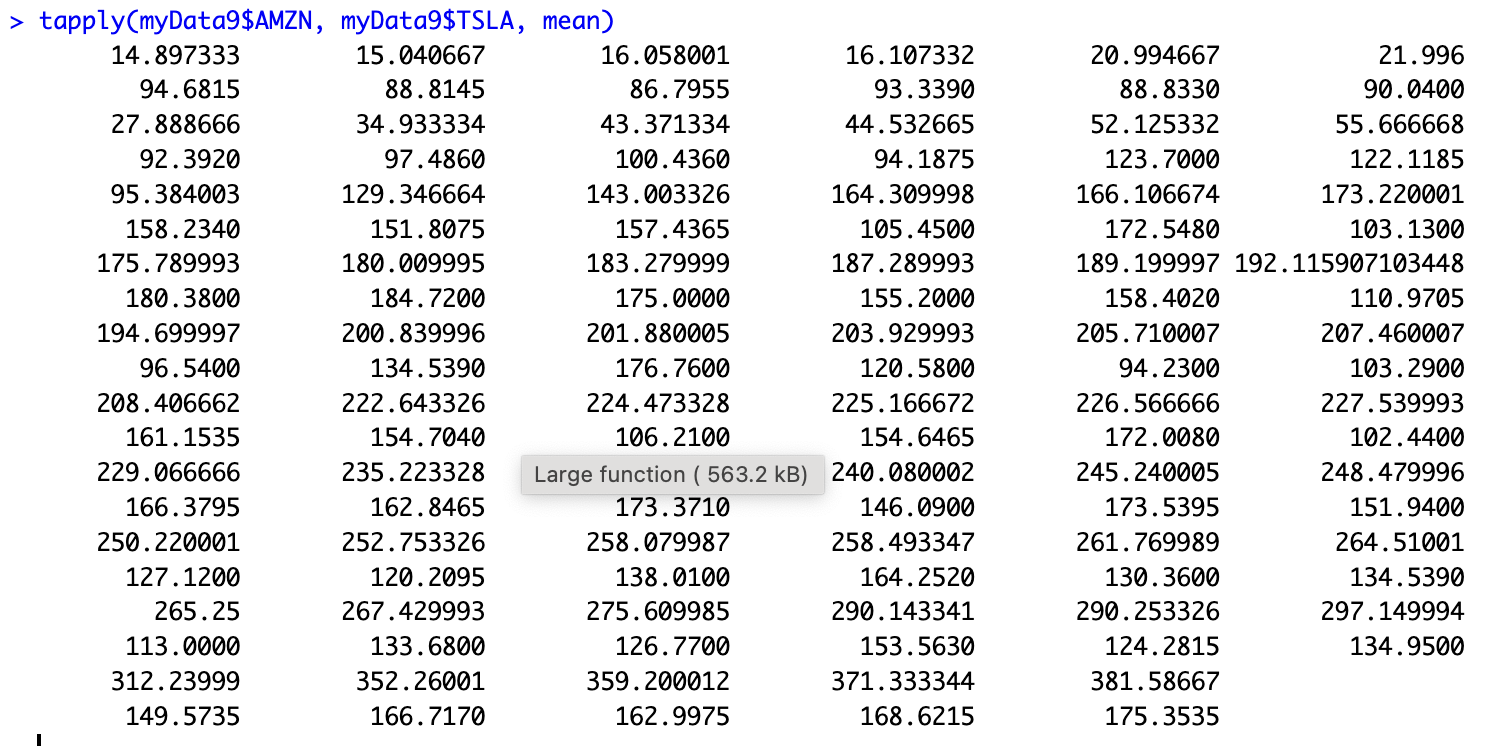


En el caso de la función summary nos permitirá obtener los valores: mìnimo, cuartil, mediana, media, tercel cuartíl y máximo para cada variable de la base de datos que se está trabajando

La función sería así:



En el caso de la función tapply nos servirá para encontrar medias o desviaciones estándar



En el caso de necesitar un percentil que no se sea un cuartil utilizamos la función:



Para calcular mínimos y máximos utilizamos la funciòn:

max(myData9$AMZN) - min(myData9$AMZN)

Para calcular el rango intercuartilico utilizamos la función:

quantile(myData9$AMZN, 0.75) - quantile(myData9$AMZN, 0.25)

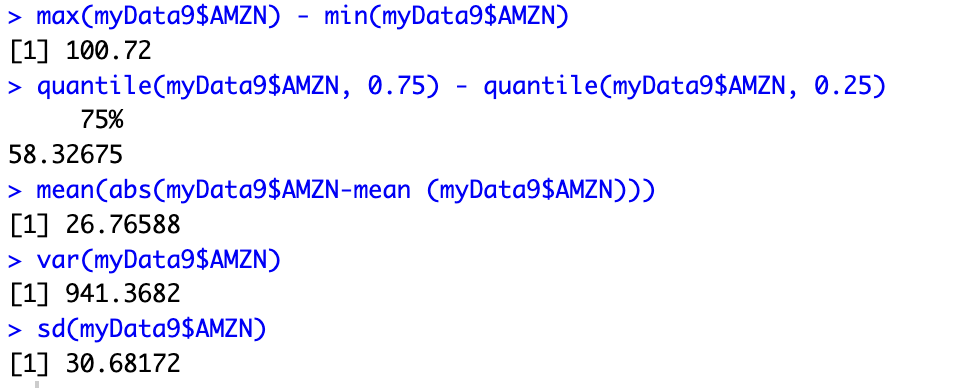
Para calcular la desviación mediana absoluta utilizamos la función:

mean(abs(myData9$AMZN-mean (myData9$AMZN)))

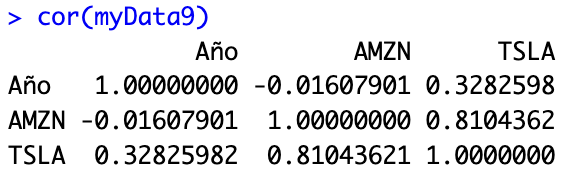
Para obtener la varianza y la desviación estándar muestrales utilizamos la función:

var(myData9$AMZN)

sd(myData9$AMZN)

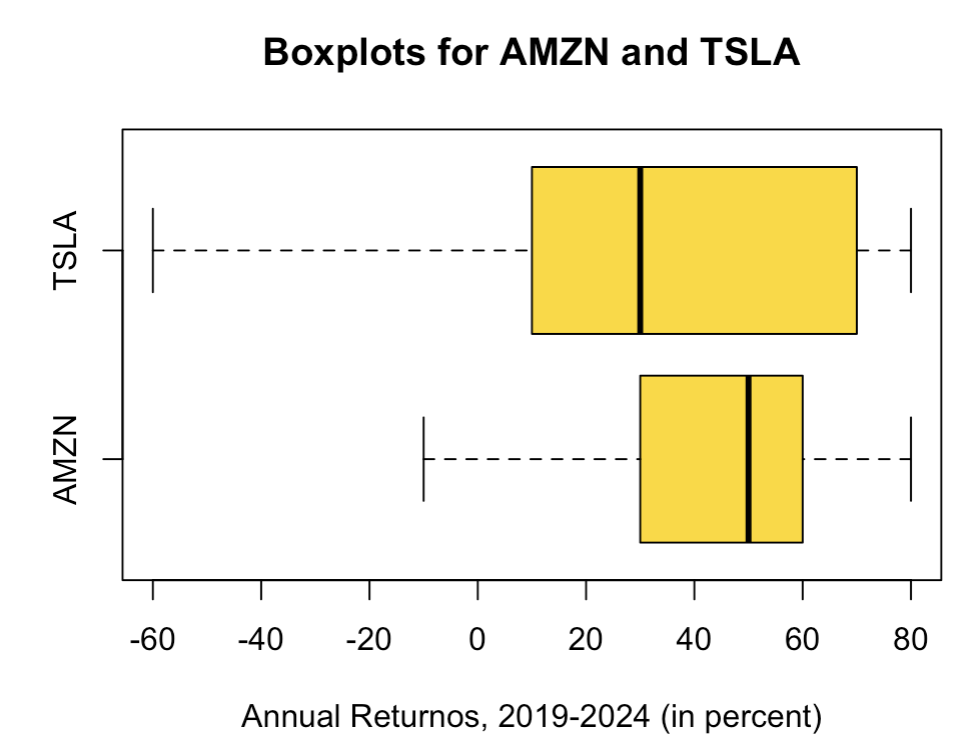


Para calcular el coeficiente de correlación entre las variables AMZN y TSLA se debe utilizar la función:



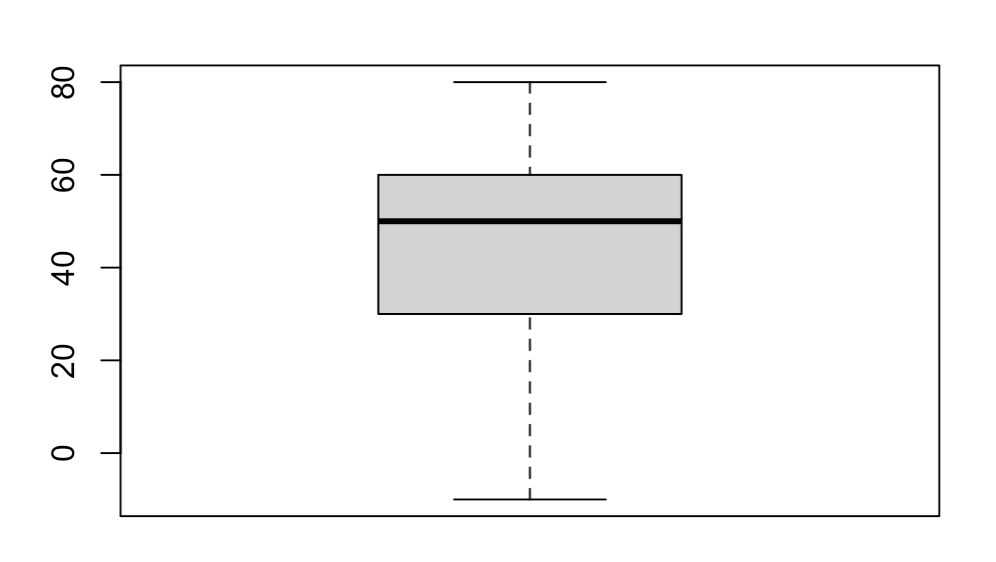
Para diagramas de caja utilizando nuestra información, esta en la función:

boxplot(myData9$AMZN, myData9$TSLA, main= "Boxplots for AMZN and TSLA" , xlab="Annual Returnos, 2019-2024 (in percent)", names=c("AMZN","TSLA"), horizontal = TRUE, col="gold")



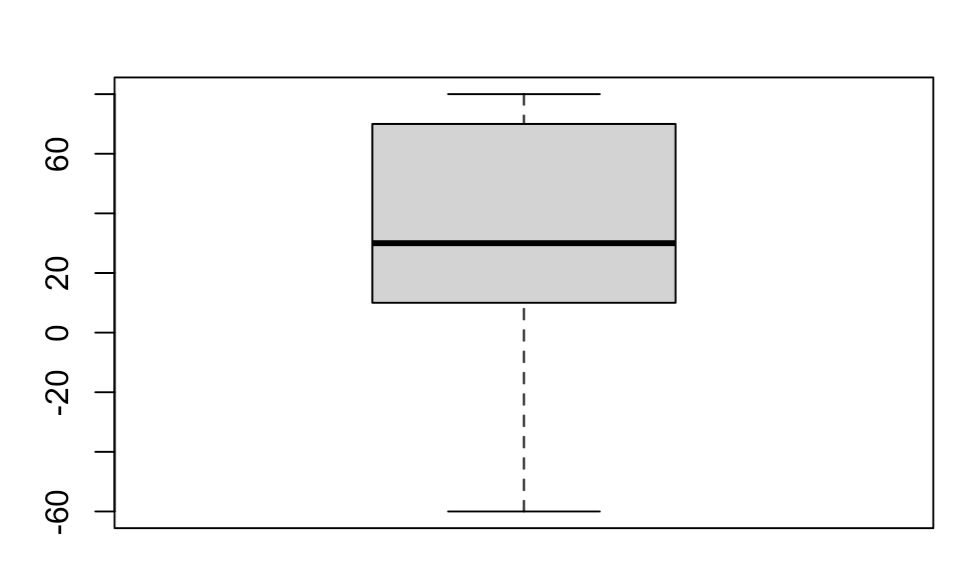
Para tartar los valores atìpicos de las 2 variables que estamos trabajando, las funciones son las siguientes

outliersAMZN <- boxplot(myData9$AMZN)$out; outliersAMZN



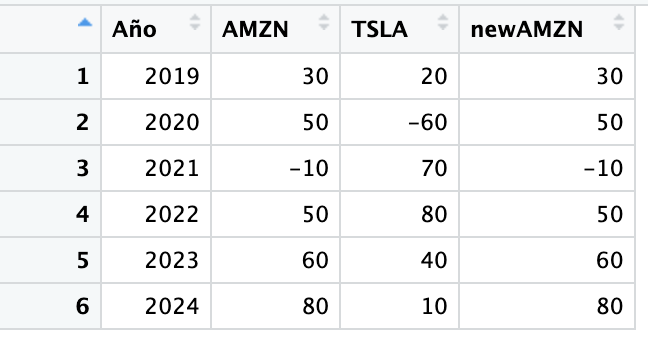
Y esta otra:

outliersTSLA <- boxplot(myData9$TSLA)$out; outliersTSLA



Para tratar los valores atípicos utilizamos esta función:

myData9$newAMZN <- ifelse(myData9$AMZN %in% outliersAMZN, NA, myData9$AMZN)



Al haber reemplazado el valor atípico podemos de nuevo ejecutar la función de summary:

