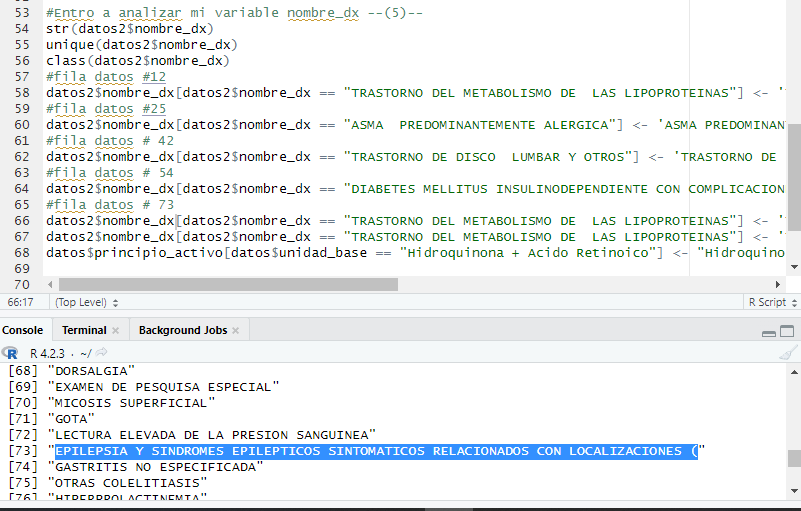
Aquí se puede identificar con la línea unique los diferentes tipos de datos existentes en mi columna nombre\_dX, en los cuales eh reemplazado por datos lógicos y conocidos Como se puede apreciar algunos datos tiene espacios dobles o caracteres espaciales.



Aquí podemos identificar los tipos de datos existente en la variable fecha atención



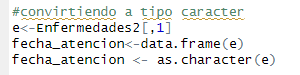
En esta parte reemplazo todos los espacios existentes



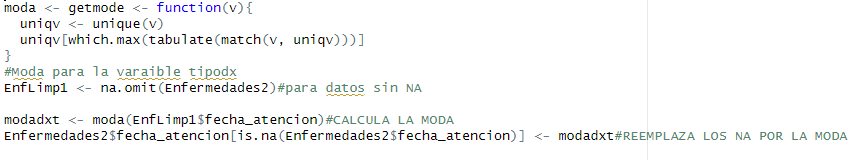
Aquí se puede identificar el tipo de carácter de la variable



Aquí podemos convertir la variable a tipo de datos carácter.



Aquí creamos la función Moda



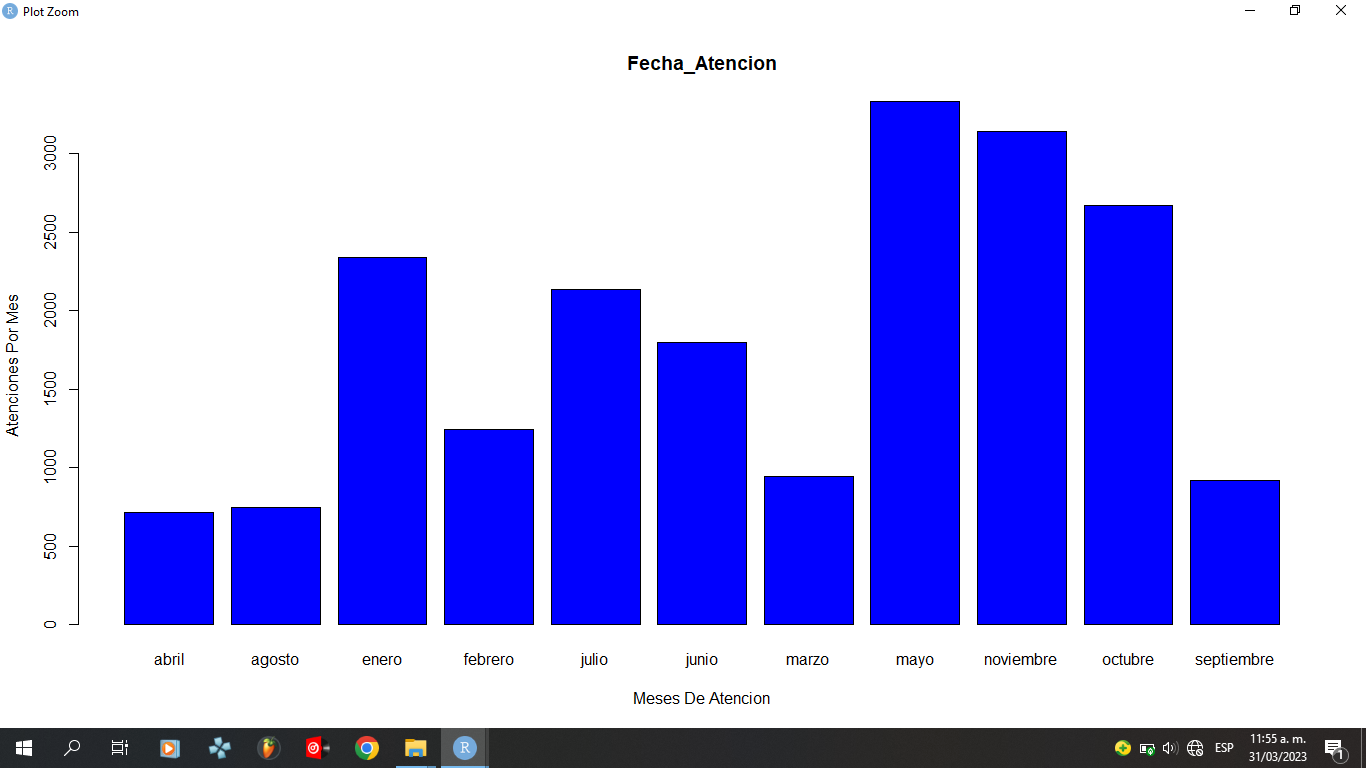
Aquí vuelvo a mostrar los datos de la nueva data creado y observo que se encuentre de una forma adecuada



Aquí con el comando barplot puedo graficar la variable y visualizar las diferentes iteraciones que están sucediendo



Gráfico de en barplot de variable Fecha atención sin datos anómalos

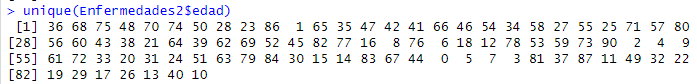


**TABLA EDAD**

En esta parte puedo observar la tabla o variable edad y puedo identificar que tiene todos los datos correctamente

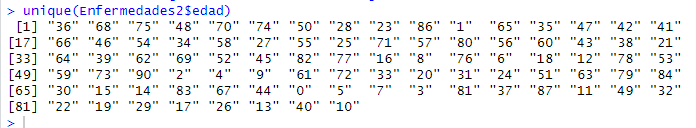






Lo que se procedo a realizar es reemplazar espacios si lo existirán y se puede identificar que si existían algunos espacios porque si se cambió el tipo de formato como estaban presentados

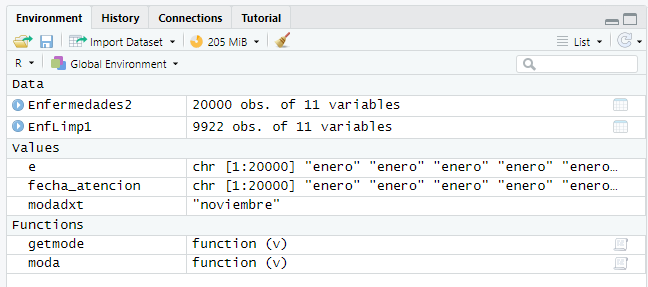




Con este comando puedo visualizar si existen NA’s y em muestra la suma total



Observo cambio cuando se saca la moda

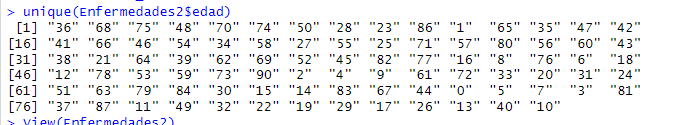


**VARIABLE EDAD**

Aquí reviso los datos anómalos de mi variable edad



Aquí veo los datos de mi variable

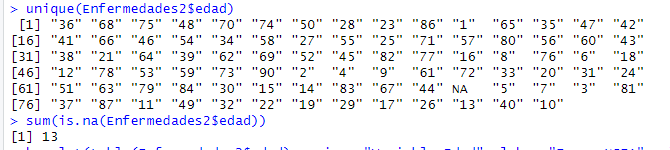


Aquí asigno NA mi dato anómalo identificado en este caso es 0 por que no existe alguien que tengo 0 años



Vuelvo y ejecuto el unique para observar NA asignado y se observa que ha sido correcta

A la misma vez ejecuto l asuma de todos los NA y me da un total de 13 NA, como se puede visualizar



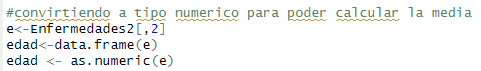
Ahora procedo imputar datos, lo hago calculando la media, pero primero verifico de que tipo es mi variable, lo hago con str()



Aquí puedo observar que mi variable es de tipo carácter



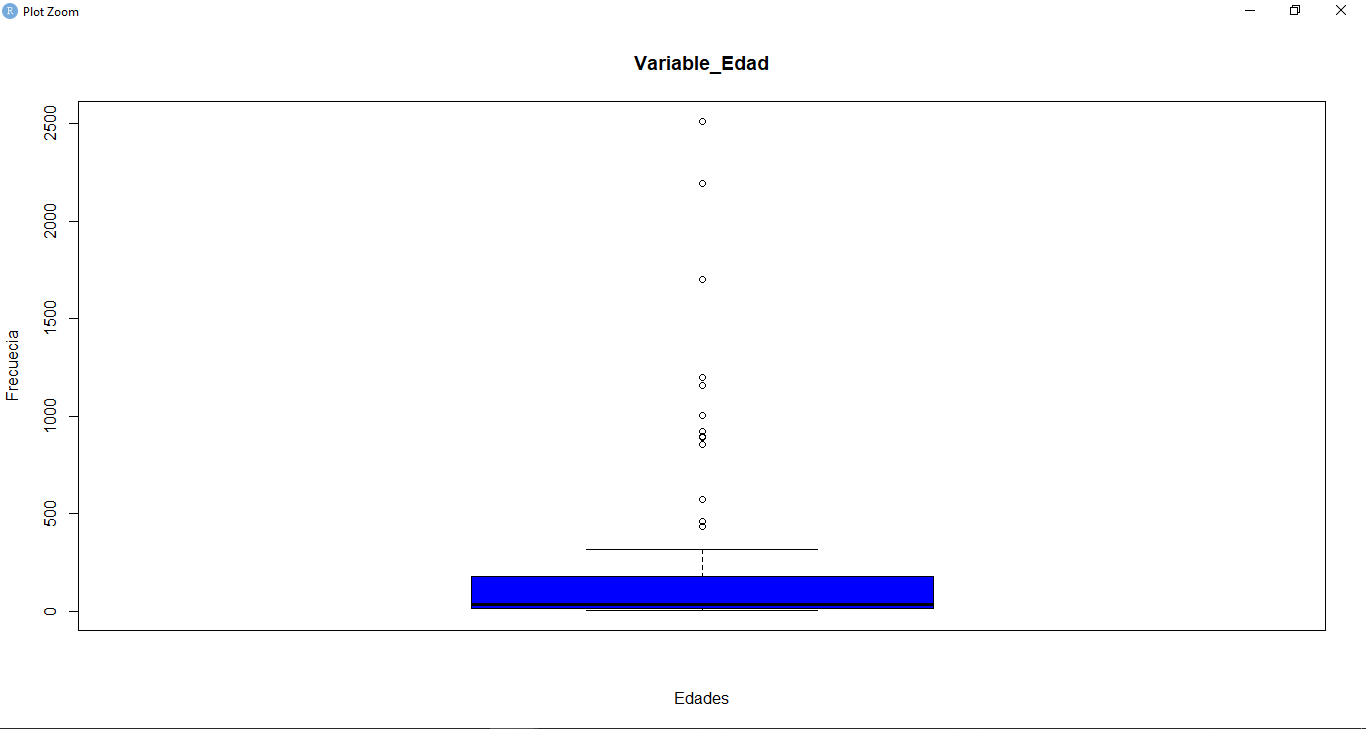
Ahora procedo a convertirlo en una variable de tipo numérica para poder calcular la media, lo hago con el siguiente código



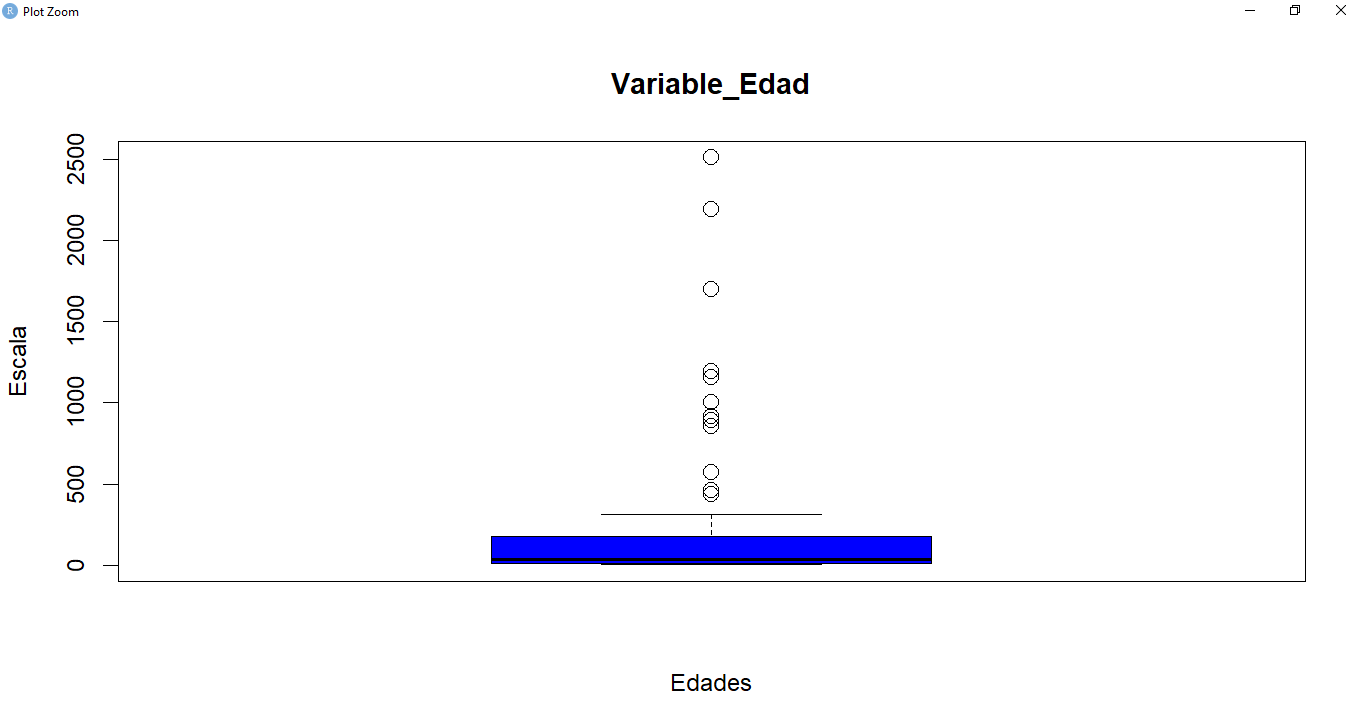
Aquí puedo observar que mi variable ya es de tipo numérica



Después procedo a graficar mi variable, con un boxplot debido a que es una variable numérica, en la cual puedo identificar que la caja de texto esta correcta con la mediana, pero se identifican otros datos adicionales esto se debe a que esos son datos muy altos o muy bajos



Mi grafico final me queda de la siguiente manera, se pueden observar unos puntos por fuera del diagrama de bigote, esto se debe porque son valores o datos atípicos es decir son valores que se encuentran fuera del rango de los más comunes.



Una vez hecho esto procedo a calcular la media de mi variable, lo hago con el presente código



Aquí puedo visualizar la media de mi variable



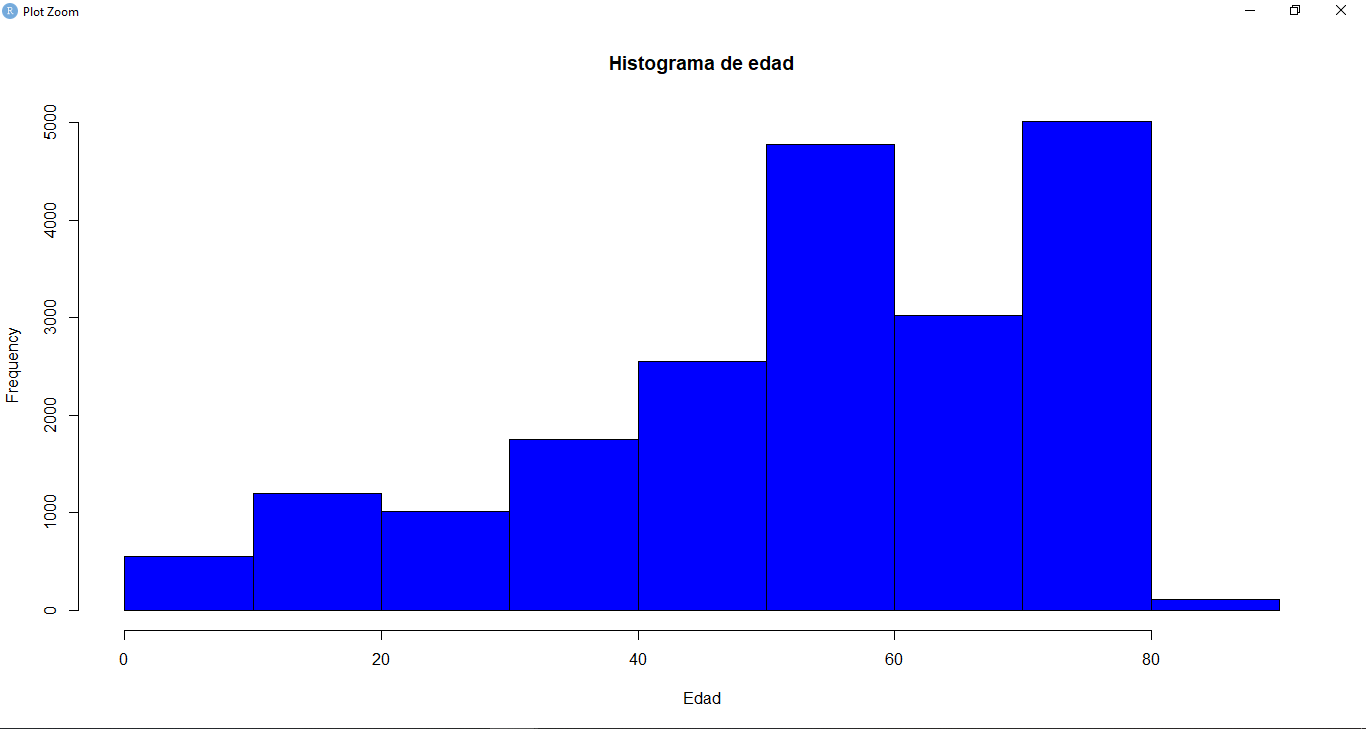
Ahora procedo a imputar datos faltantes, lo hago con el presente código



Ahora hago un sum con el fin de observar que ya no existen NA y se imputaron con éxito

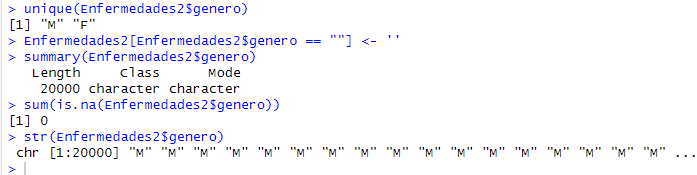


También puedo graficar un gráfico de tipo His, con el fin de analizar otros cambios

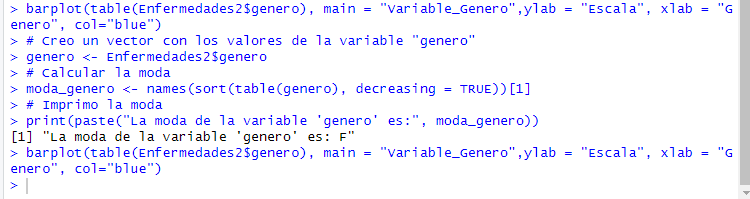


**VARIABLE GENERO**

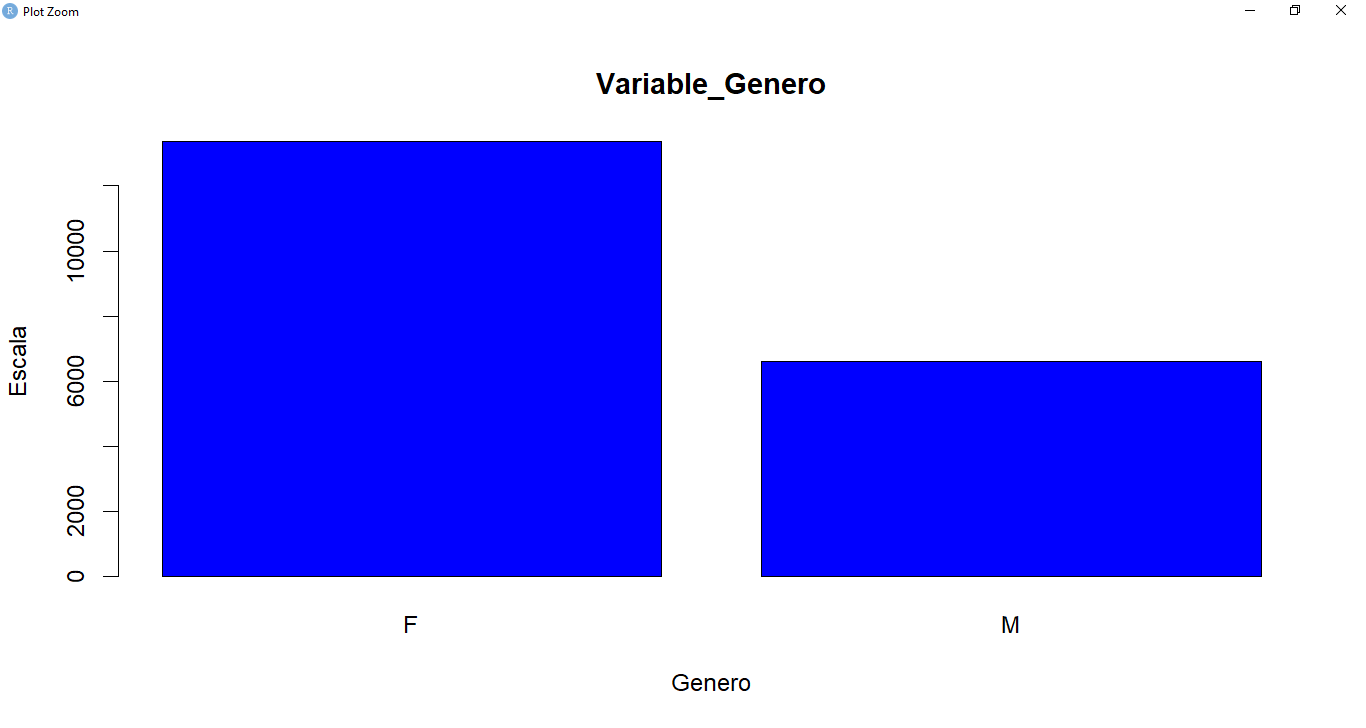
Inicio el mismo proceso que eh realizado con las anteriores variables, y puedo identificar que solo tiene 2 tipos de datos los cuales son M Y F además puedo identificar que no existen datos anómalos ni nada por el estilo. También puedo observar que el tipo de variable es de tipo carácter, entonces procedo a sacarle la MODA para poder visualizar el valor que más se repite u otro tipo de análisis



Luego procedo a sacar la MODA, con el fin de identificar el valor que más se repite



Y seguidamente grafico con el fin de observarlo visualmente



Aquí también puedo observar la moda



**VARIABLE NOMBRE\_DX**

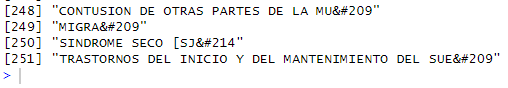
Una vez realizado el unique se puede identificar que existe un error de sintaxis en las descripciones o el otro error puede ser por ser codifico mal el data set, aquí se muestra el error que se visualiza no es un dato anómalo solo es un dato de error de digitación o como se había mencionado anteriormente, lo que se procede es a cambiarbiarlo por un nombre completo y real, según lo investigado en páginas we de sal, al parecer debe decir CONTROL DE SALUD DE RUTINA DEL NIÑO



En la fila 184 también se identificó que se encuentra vacía por lo tanto se procederá a reemplazar por un dato NA anómalo



También en otras columnas se identificó otros datos raros y se procederá a realizar lo más conveniente



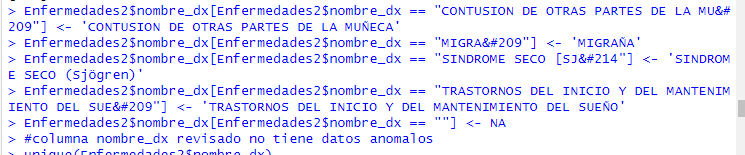
Con esta línea reemplazo por el nuevo dato que voy a asignar



Vuelvo y reviso el dataset e idnetifco que ya se cmabio de forma correcta y ahora mi dato ya es mas valedero

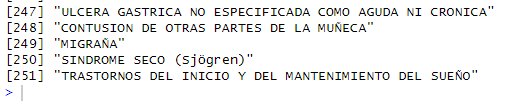


Aquí se puede observar que ya se realizaron los respectivos cambios y sustitución por NA donde había vacíos



Procedo a graficar nuevamente la variable y observo que ya se realizaron los cambio de un amanera exitosa

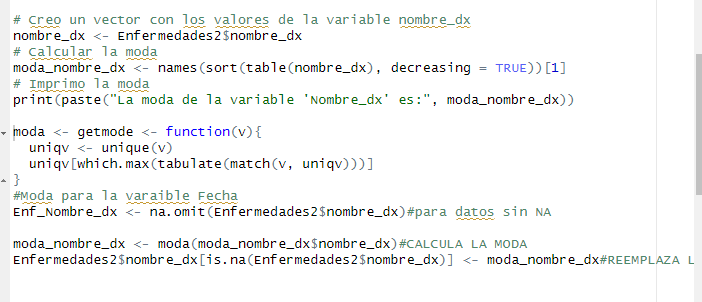




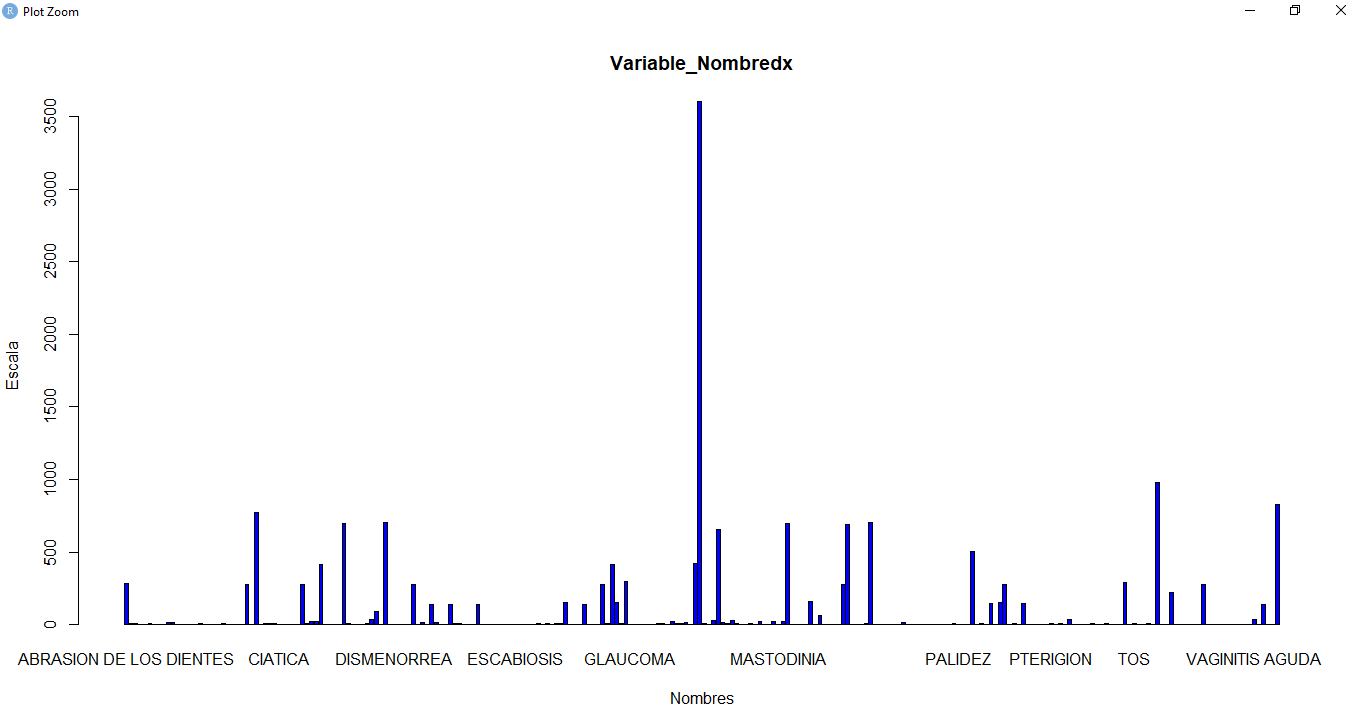
Verificando lo valores vacíos puedo identificar que hay 2036 datos NA



Hacemos el reemplazo por la moda y volvemos a comprobar si existen datos anómalos

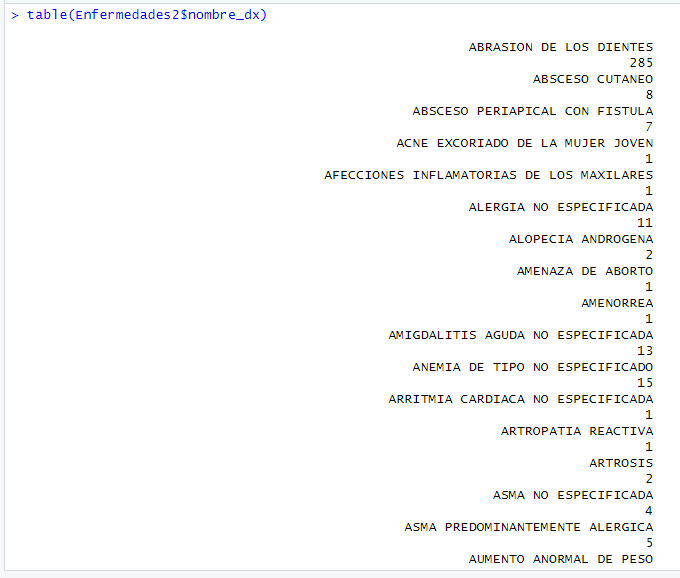


Ahora pasamos a graficar para poder observar el valor más representativo



Ahora en la gráfica puedo observar que como hay muchos datos se distorsiona, y no se puede visualizar con exactitud el dato exacto, para ello primero imprimo el toda la tabla o variable para observar el número de veces que repite cada dato, seguidamente paso a sacar la MODA debido a que mi variable es de tipo CHARACTER, el fin es saber cual es el valor mas representativo para poder graficar.

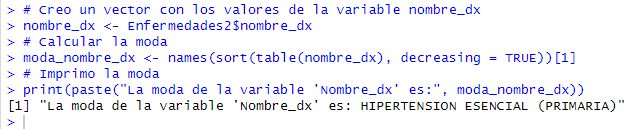






aquí puedo observar que el valor mas representativo es el presente dato

Sabiendo lo anterior procedo a sacar la MODA

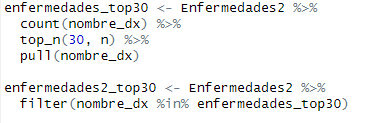


Aquí se puede visualizar que el resultado de la MODA si me arroja que HIPERTENSION ESENCIAL (PRIMARIA)

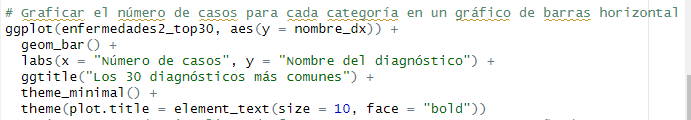
ese es el valor más representativo, ahora procedo a graficarlo de una forma mas reducida para verificar el valor mas representativo y lo hago con categorías mas comunes en este caso escojo un rango de 30 primero llamo las librerías



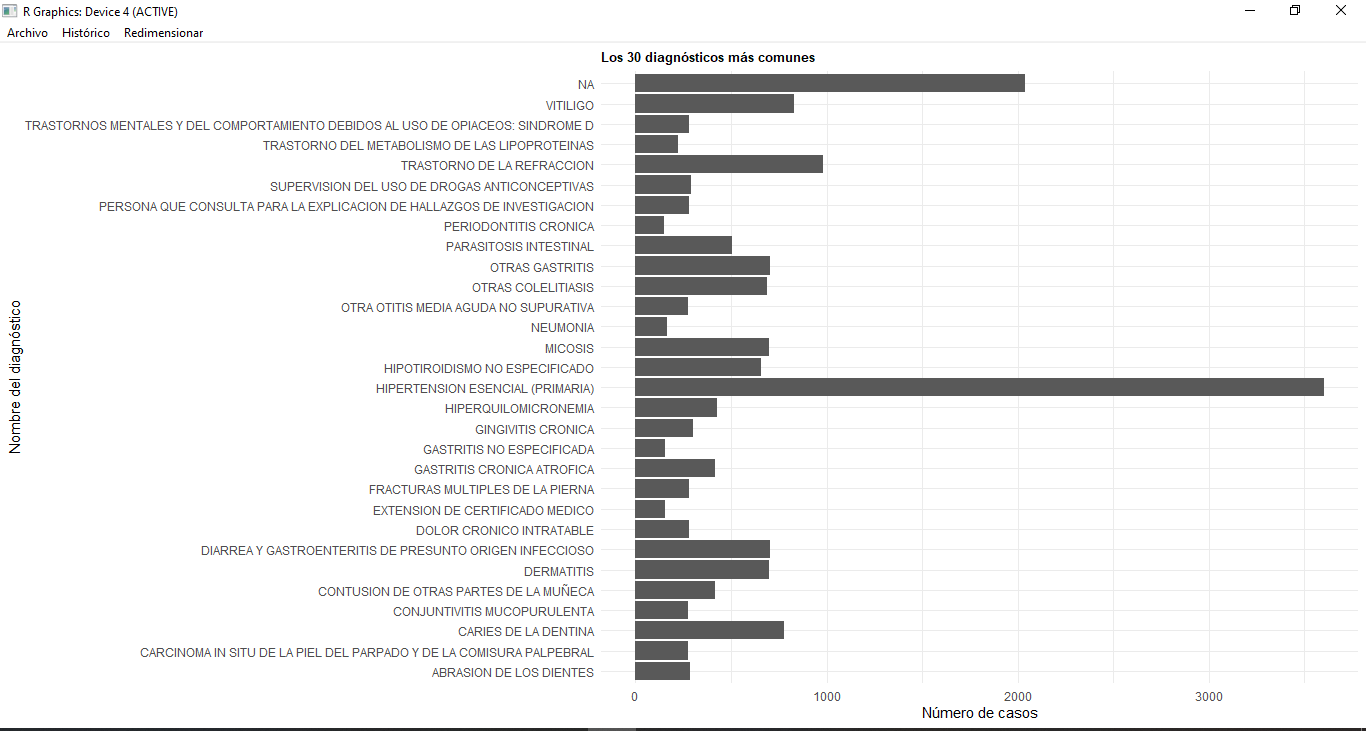
Seguidamente Creo un data set solo con las 30 categorías más comunes, lo hago con el siguiente código



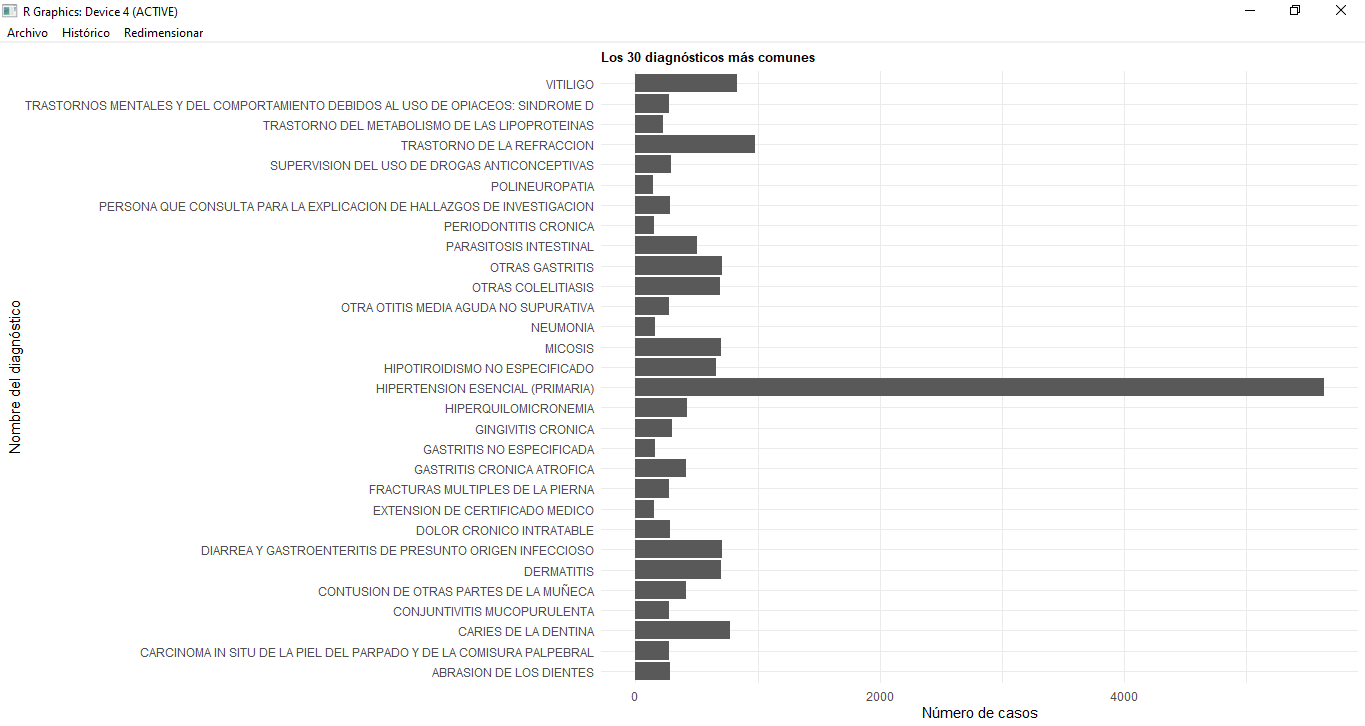
Con estas líneas imprimo la grafica discriminado los 30 datos mas comunes y con mayor frecuencia como se puede observar lo hago de forma horizontal con el fin de visualizar adecuadamente



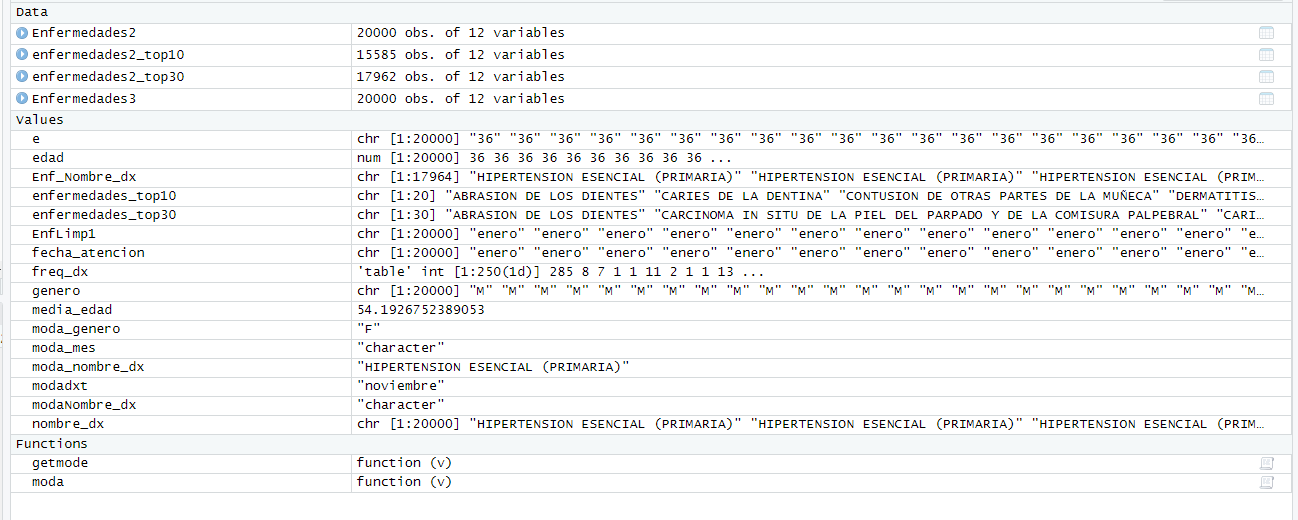
Esta es la gráfica generada y efectivamente se puede visualizar que es valor más representativo es HIPERTENSION ESENCIAL (PRIMARIA), esta es la grafica incluyendo datos NA



Esta es una grafica ya reemplazado los datos NA por la MODA sigue siendo la misma solo que esta vez ya no representa nada los datos NA

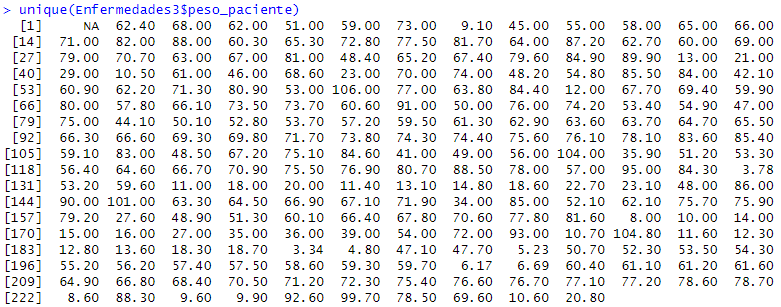


Anexo de panel entorno de variables

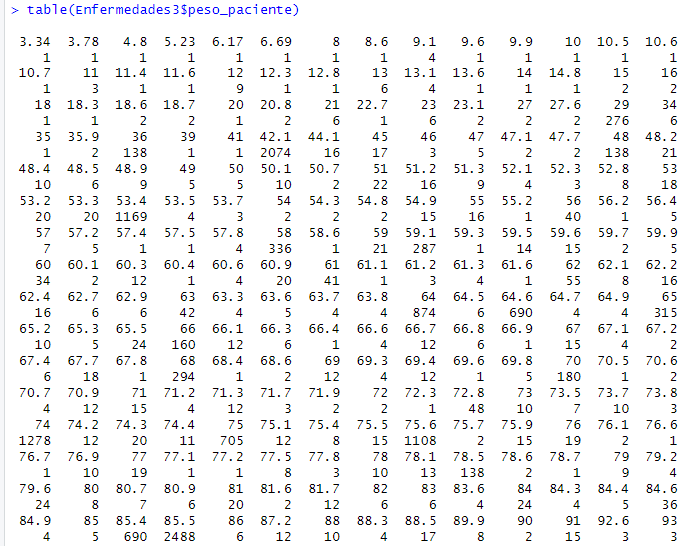


**VARIABLE PESO\_PACIENTE**

Aquí procedo a cargar mi nuevo data set llamado Enfermedades3 y procedo a hacer un unique con el fin de visualizar datos diferentes a los de la variable,



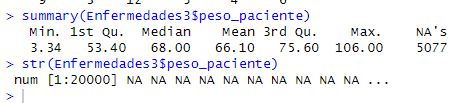
Con el siguiente código puedo verificar el numero de veces que se repite los datos



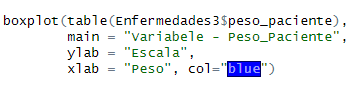
Seguidamente puedo observar los datos NA uqe esta presentes los cuales me arrojan un total de 5077

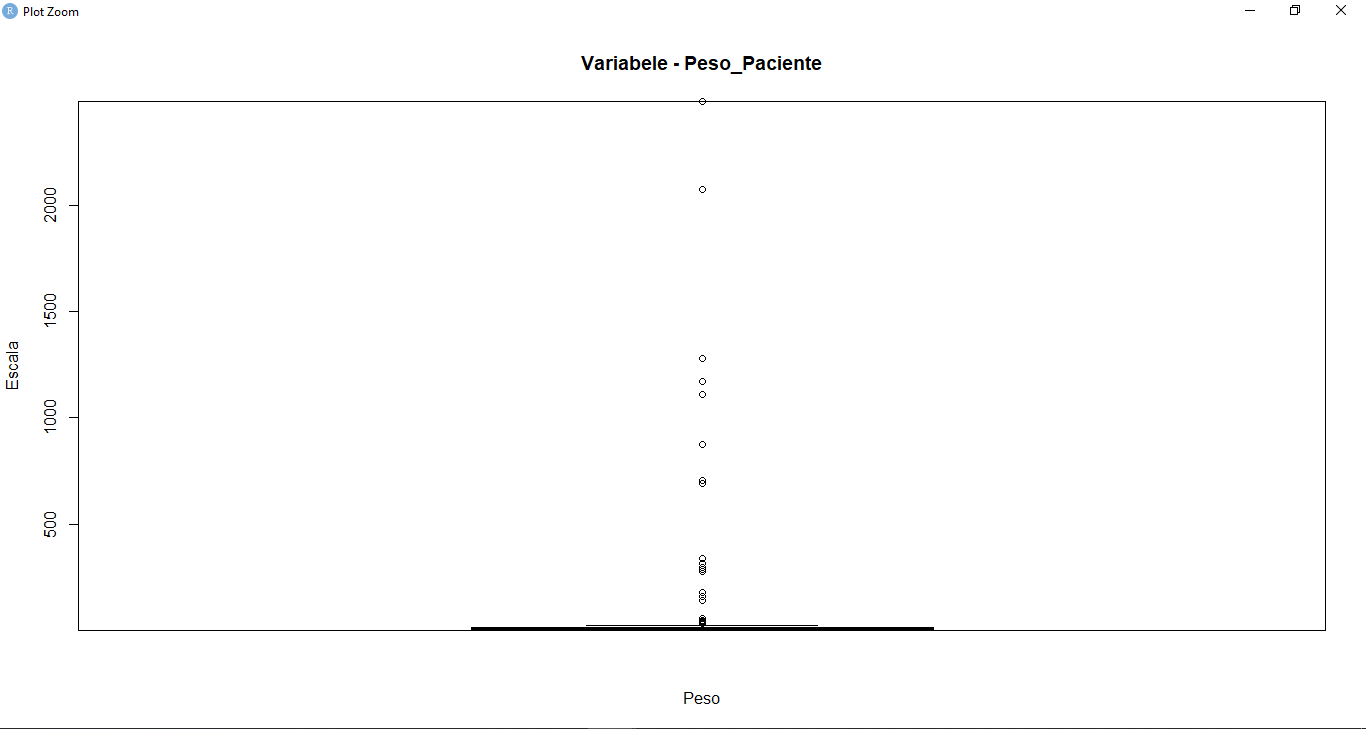


Procedo a reemplazarlos por la media debido a que mi variable es de tipo Numérico pero además de ello vuelvo y verifico con un sumary para mirar más detalles de los datos contenidos, seguidamente ejecuto un str para poder corroborar que los datos son de tipo numerico.



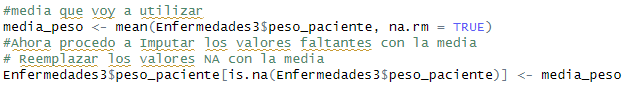
Ahora procedo a graficar con el siguiente codigo





Puedo observar que la variable tiene muchos datos y al parecer hay muchos anómalos

Procedo a remplazarlos por la media debido a que mi variable es de tipo numérica, lo hago con el siguiente código



Puedo observar que la media es igual a:







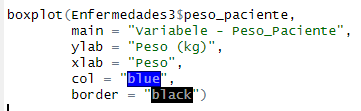
Seguidamente vuelvo a comprobar si fue efectivo, y como se puede observar ya no hay datos NA



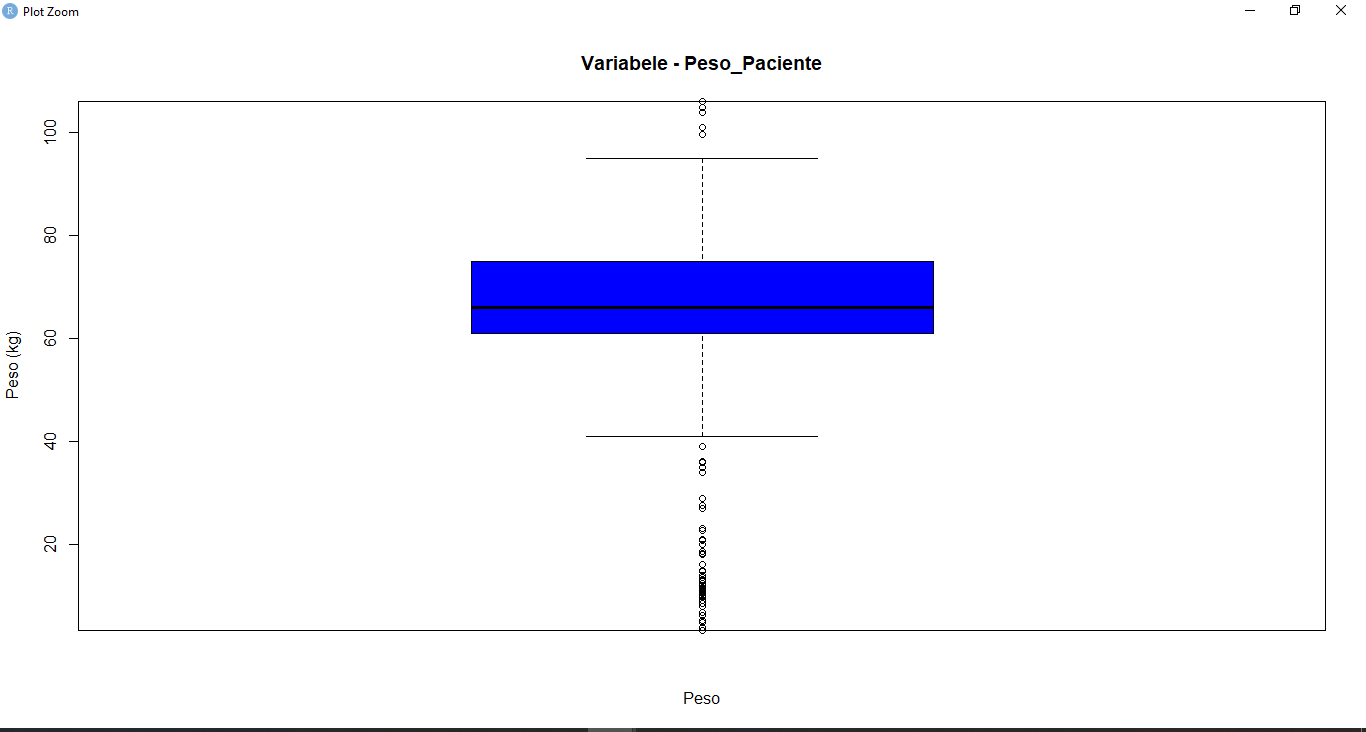
Con un sumary observa queya no existen NA’s



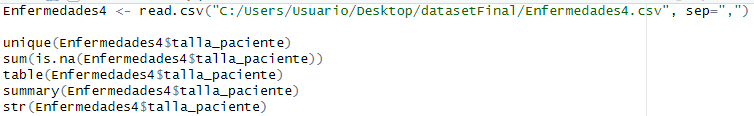
Vuelvo a graficar para analizar como quedo el grafico nuevamente, lo hago directamente a la variable



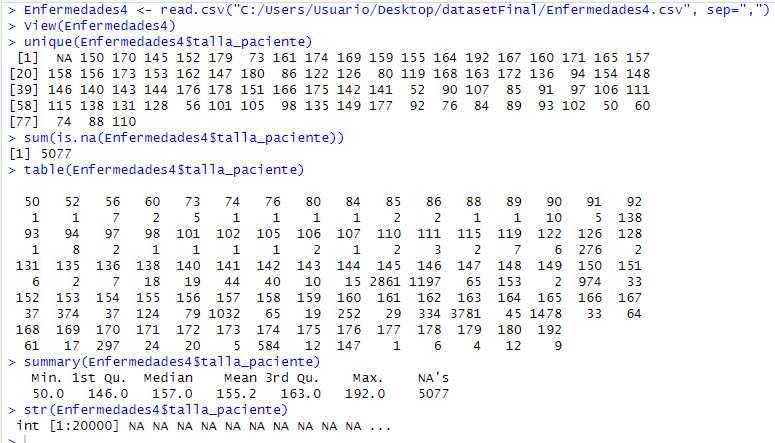
Ahí puedo visualizar que existen datos atípicos que hacen que se representen por fuera, cabe resaltar que no son NA



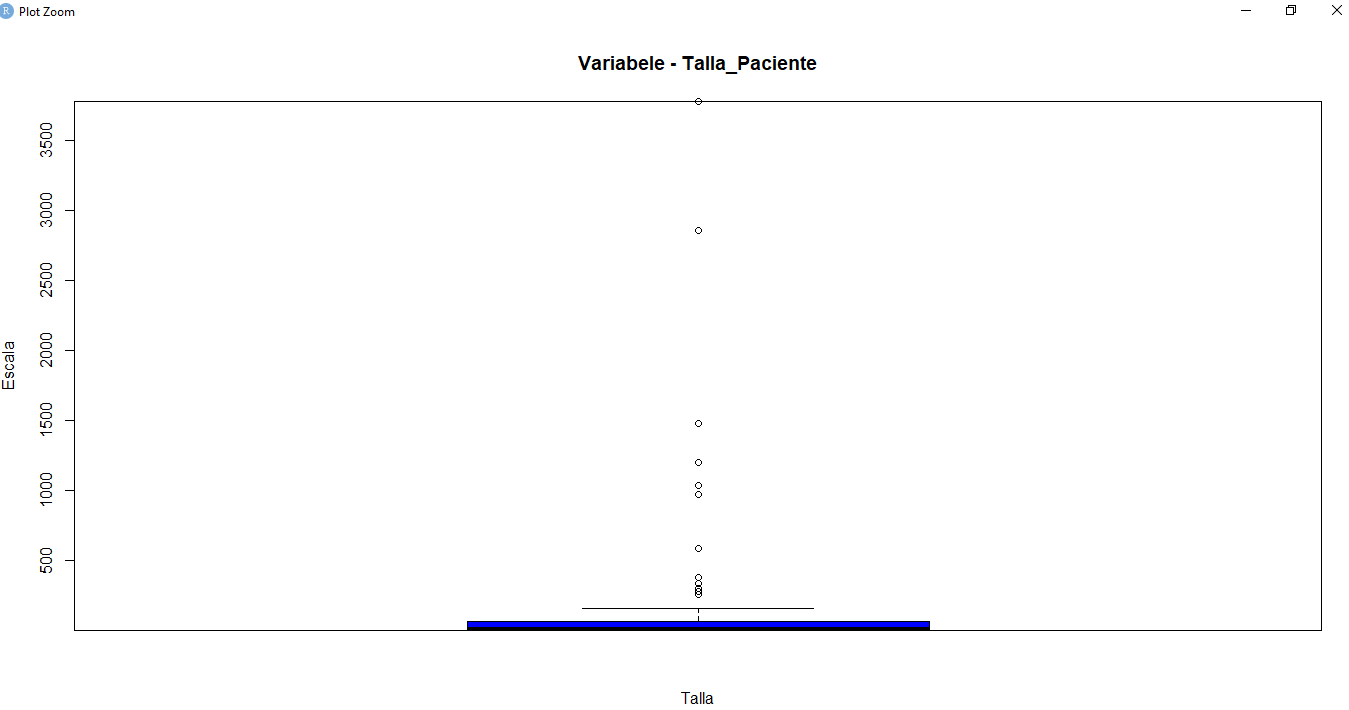
**VARIABLE TALLA\_PACIENTE**

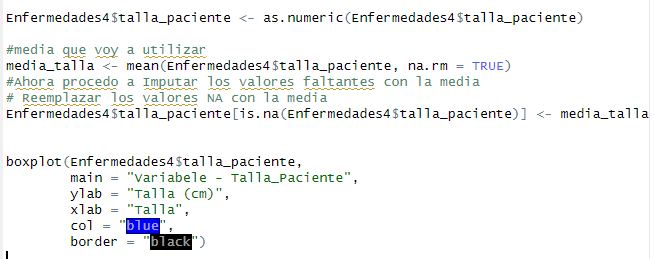


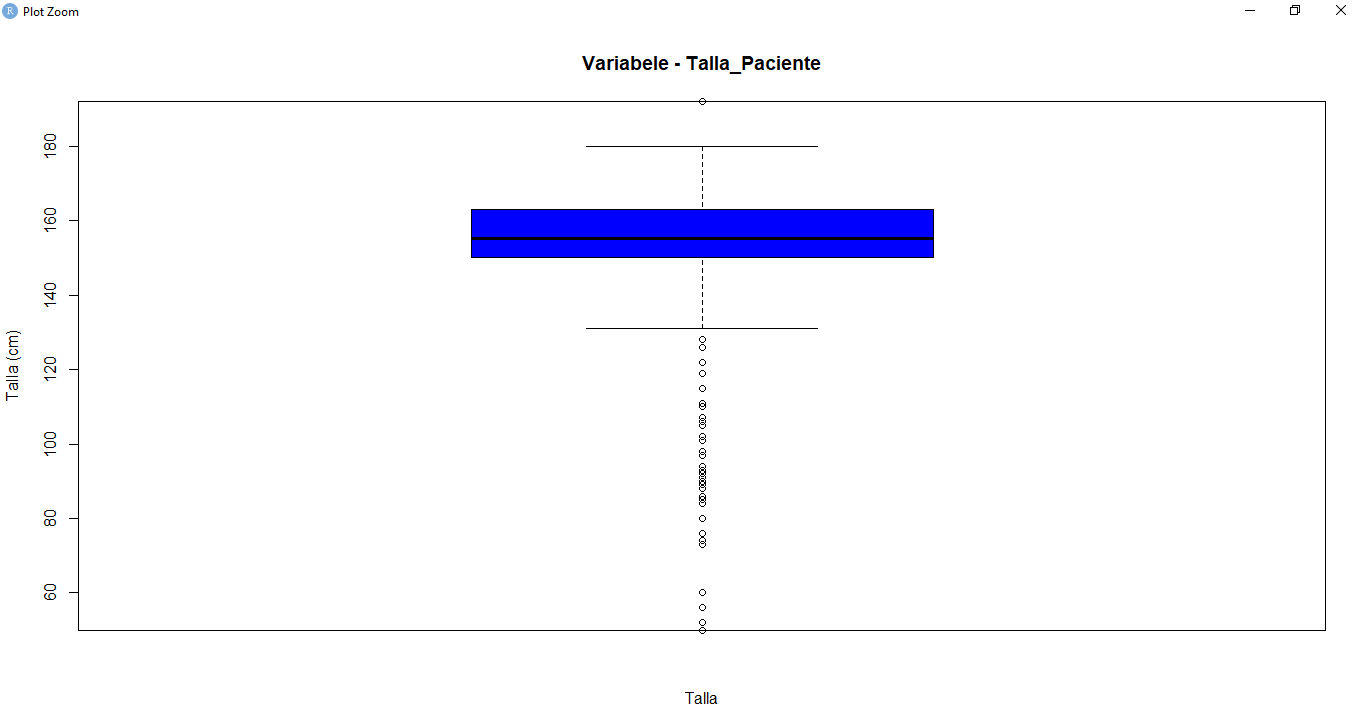
Consola



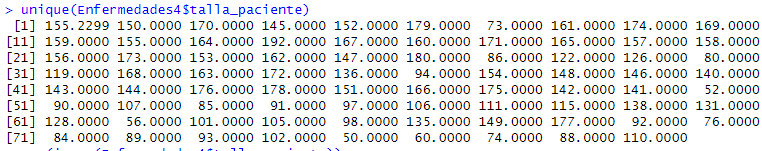
Aquí grafico



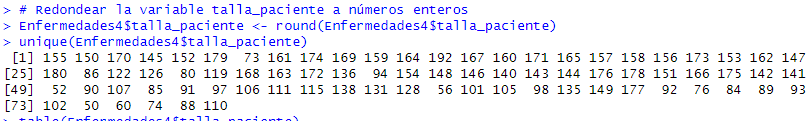




Si los valores se visualizan de la siguiente manera, es por la media, esto no afecta el valor de los resultados



Pero si se quisieran quitar estos decimales se le aplica un round() y quedaran nuevamente asi



**VARIABLE NOMBRE\_MEDICAMENTO**

Aquí ya se ve el dato anomale que asigne

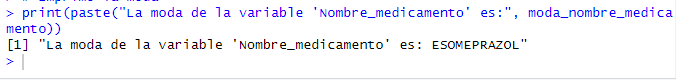


Aquí puedo visualizar el total de datos NA

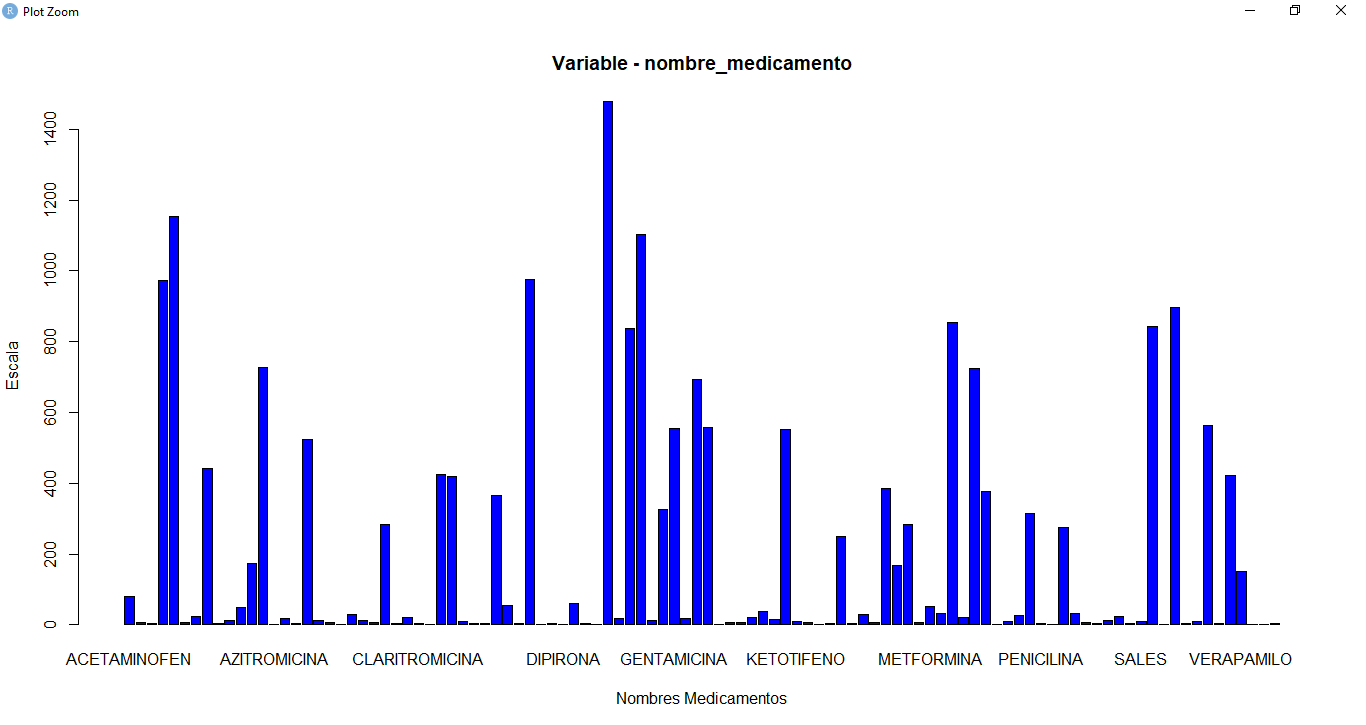


Ahora procedo a sacar la MODA para mi caso es

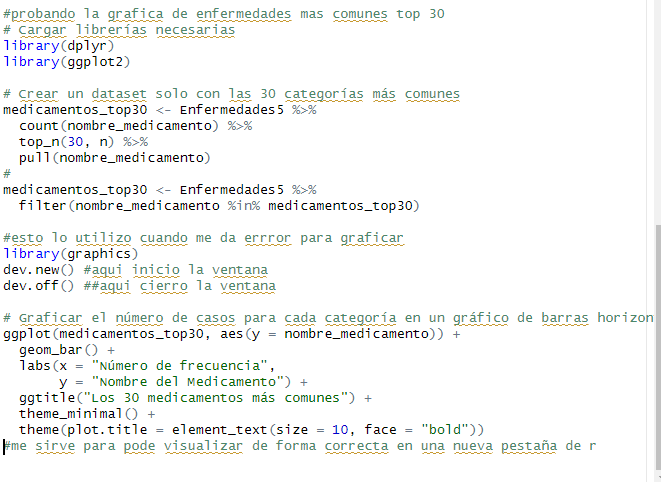




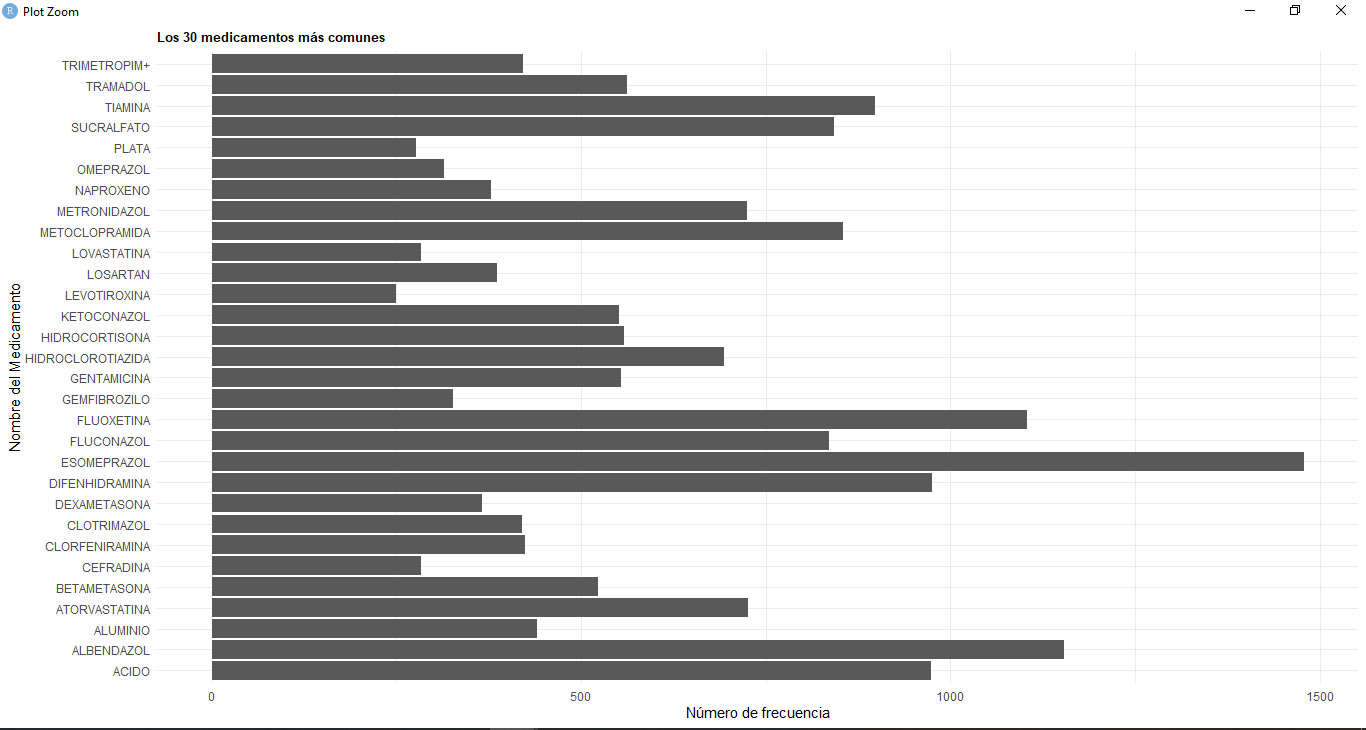
Aquí se puede aprecia visualmente



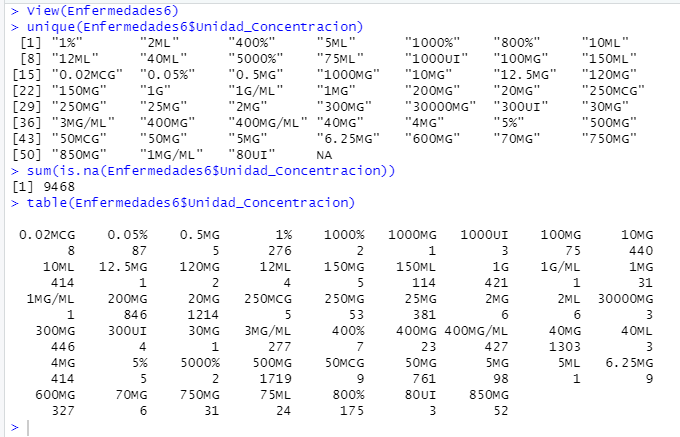
Luego procedo a sacar los top 30 de los medicamentos mas frecuentes con el fin de acercar mas la visualización en el lugar donde esta sucediedo mas cantidad de frecuencia, lo hago con el siguiente codigo.



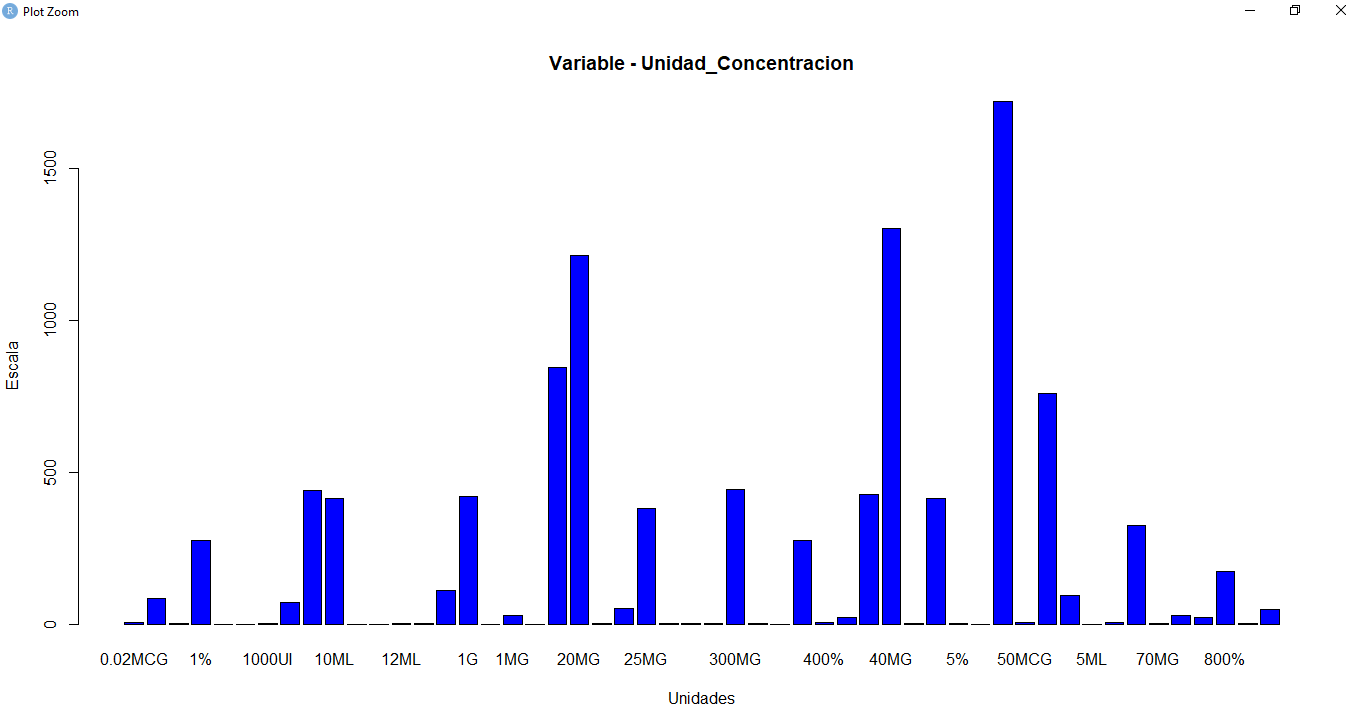
La grafica quedaría de la siguiente manera



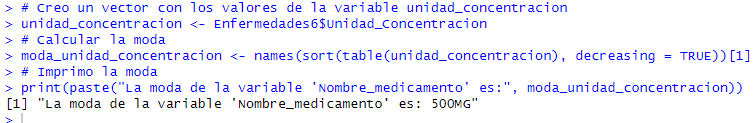
**VARIABLE UNIDAD\_CONCENTRACION**



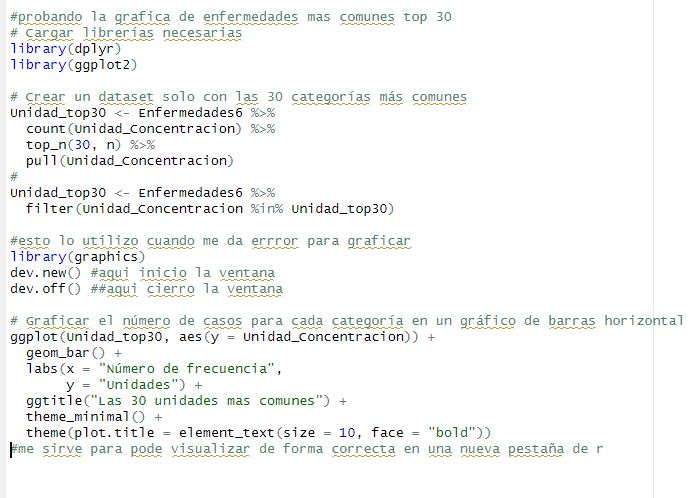
Ahora graficamos para visualizar los cambios



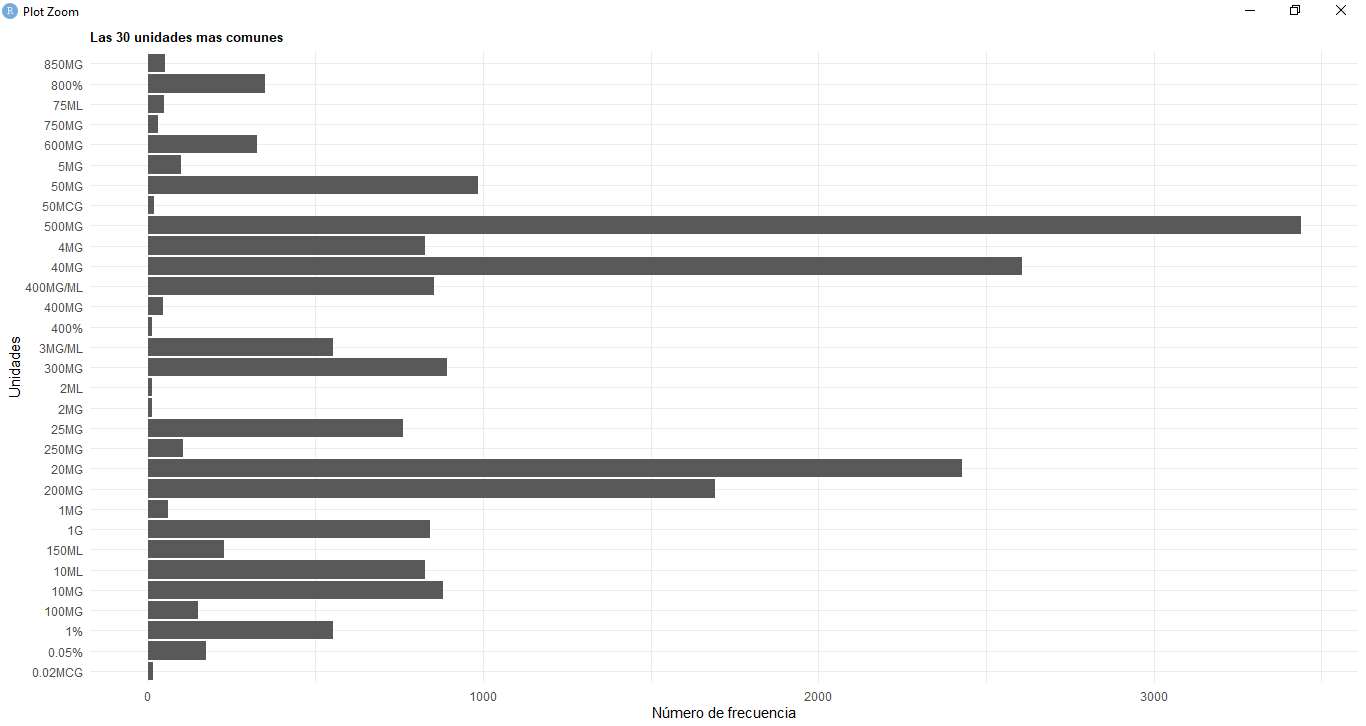
Calculo la moda y la imprimo



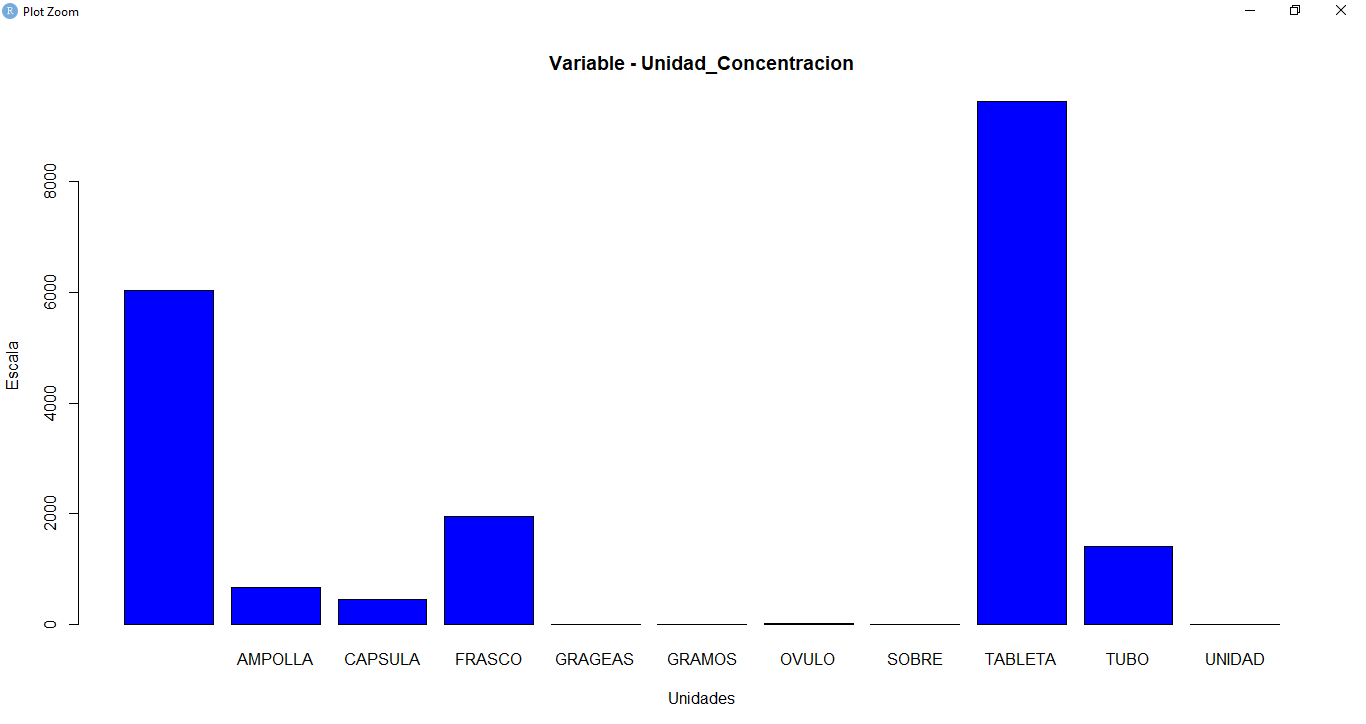
Ahora grafico de una sacando las 30 unidades mas comunes



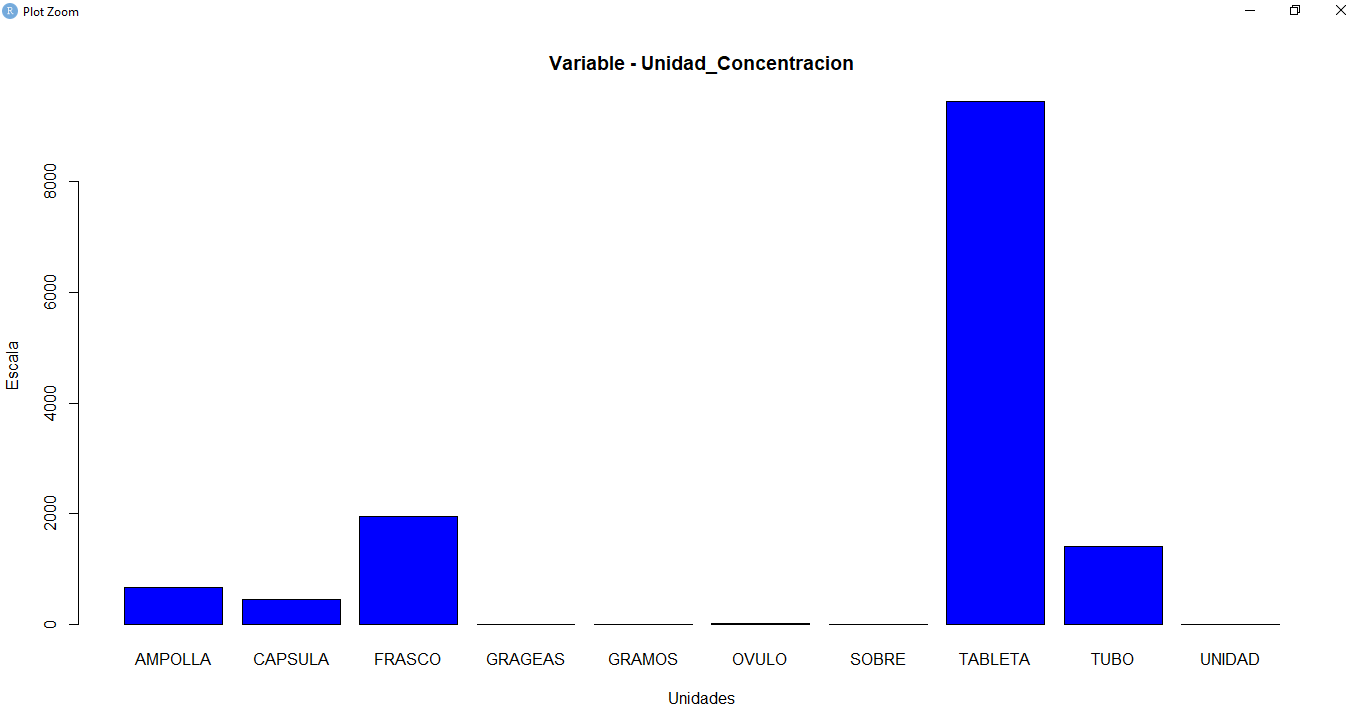
Aquí puedo visualizar que la unidad mas común es 500MG

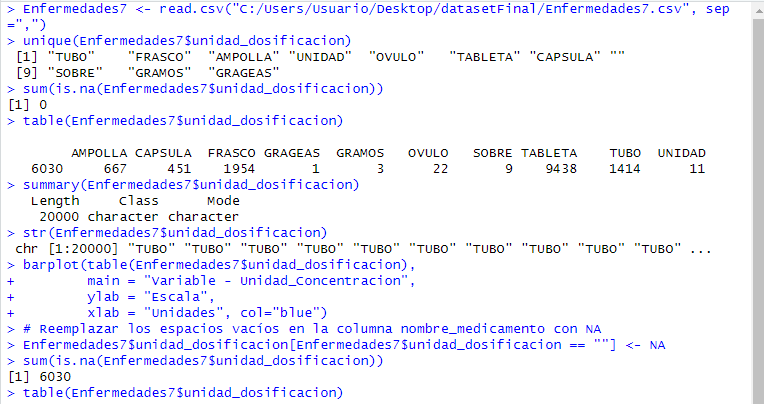


**VARIABLE UNIDAD\_DOSIFICACION**

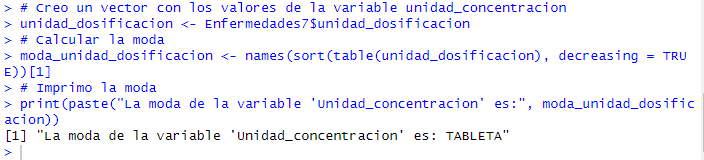


Ahora veo que hay valores vacíos entonces procedo a reemplazarlos por NA





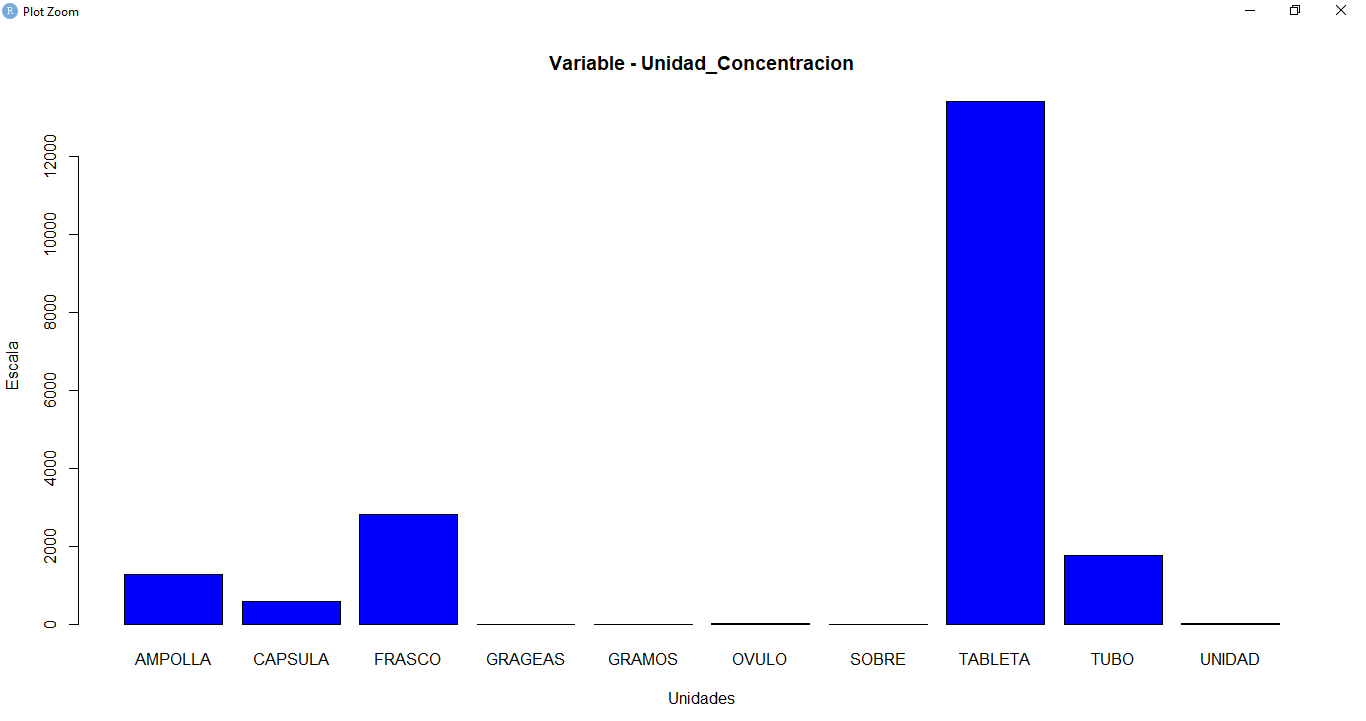
Ahora miro la moda



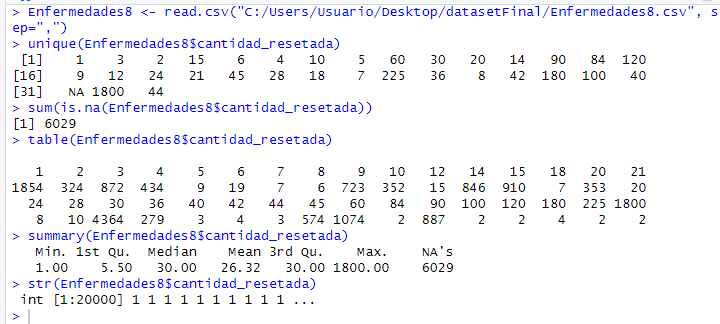




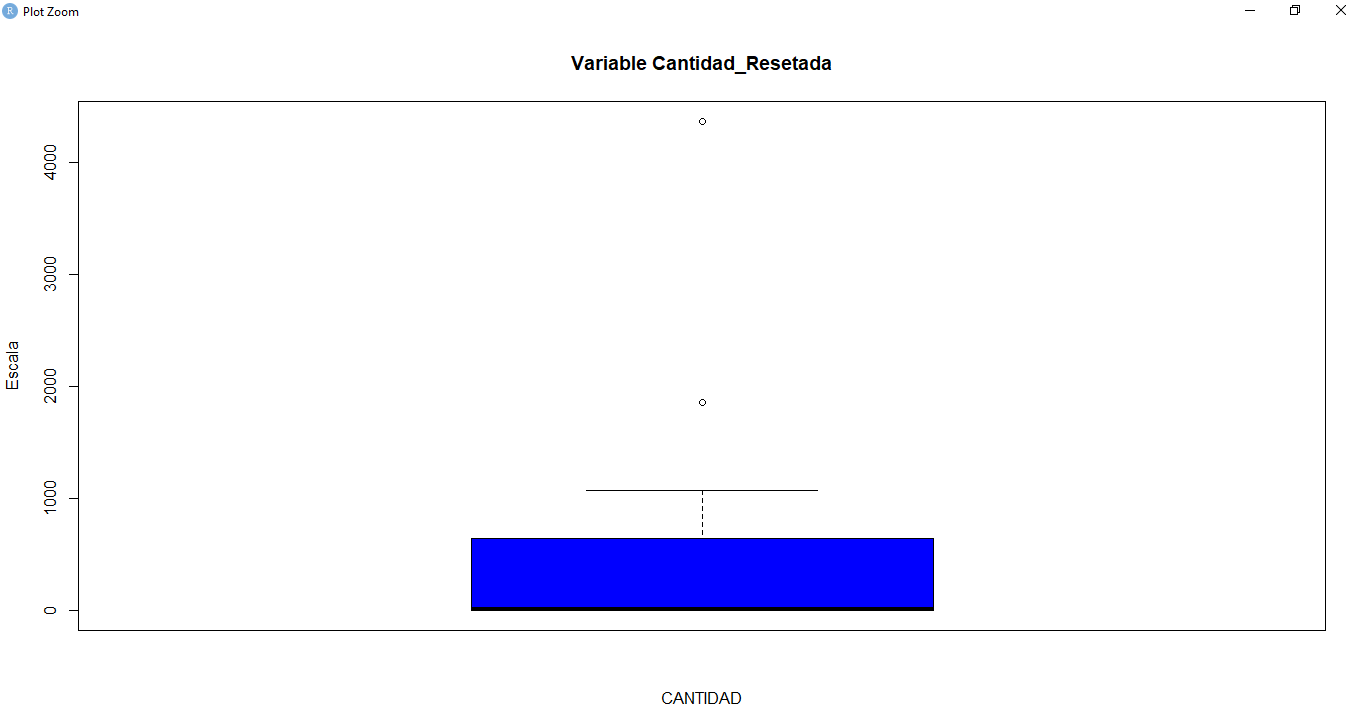
Aquí grafico sin los NA

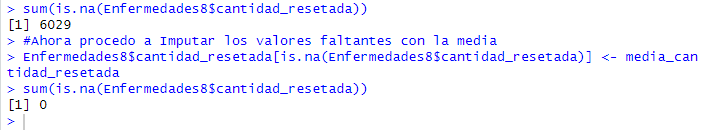


**VARIABLE CANTIDAD\_RESETADA**



Ahora grafico



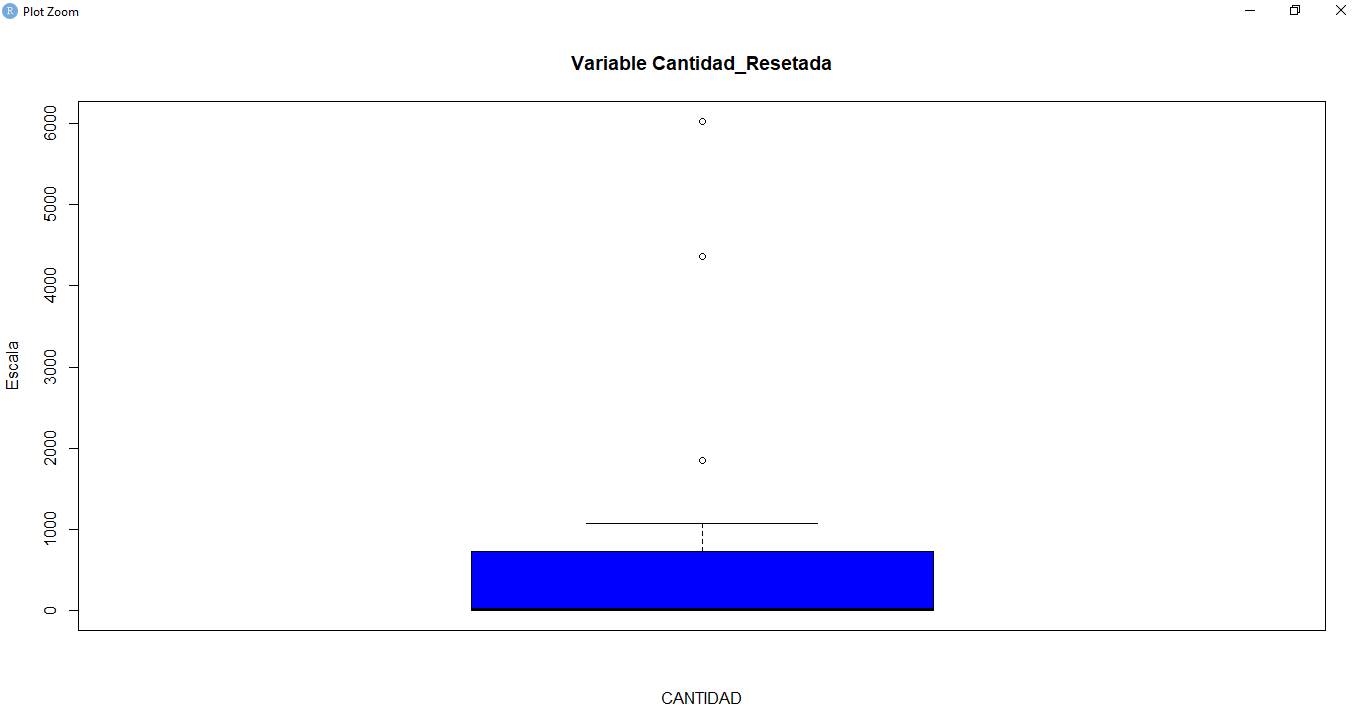


La media



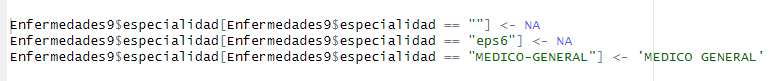


Gráficamente



**VARIABLE ESPECIALIDADES**







Ahora grafico para poder visualizar

