

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Laboratorio de Seminario de Sistemas II
Auxiliar de cátedra: Glen Calel Robledo

DOCUMENTACIÓN

Práctica 2

Danilo Urías Coc
201314565

Guatemala, junio de 2020

Contenido

Requerimientos técnicos	1
Requisitos del sistema:	1
Herramientas y versiones utilizadas para la solución.....	1
Descripción de las acciones utilizadas.....	2
Resultados obtenidos.....	4

Requerimientos técnicos

Requisitos del sistema:

Hardware utilizado:

Memoria RAM 8GB
Procesador Intel core i5 2.20 GHz

Software adicional utilizado:

Sistema operativo Windows 10 professional
Microsoft Office Excel
Notepad ++

Herramientas y versiones utilizadas para la solución

- **R version 4.0.2**

R es un entorno de software libre para computación estadística y gráficos. Compila y se ejecuta en una amplia variedad de plataformas UNIX, Windows y MacOS.

Documentación y descarga de R: <https://www.r-project.org/>

- **R studio Free - Open Source License**

RStudio es un IDE para el lenguaje de programación R, dedicado a la computación estadística y gráficos. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo.

Enlace de herramienta: <https://rstudio.com/products/rstudio/>

- **Paquetes R que requirieron instalación**

gridExtra: Proporciona una serie de funciones de nivel de usuario para trabajar con gráficos de "cuadrícula", en particular para organizar múltiples diagramas basados en cuadrículas en una página y dibujar tablas.

Grid: En conjunto con gridExtra permite la generación de tablas.

ggplot2: Es un paquete de visualización de datos para el lenguaje de programación estadística R.

Descripción de las acciones utilizadas

Manipulación de datos

Data frame: es una tabla o una estructura tipo matriz bidimensional en la que cada columna contiene valores de una variable y cada fila contiene un conjunto de valores de cada columna.

Función Read: Lee el contenido de un archivo de datos especificado a un data frame. De forma predeterminada, el separador de datos del archivo son comas.

subset.data.frame: Recibe un dataframe origen y mediante condiciones devuelve un nuevo dataframe con las columnas acordes a la condición.

Función aggregate: Divide una muestra en subconjuntos, calcula los estadísticos deseados por cada subconjunto y devuelve el resultado en una forma conveniente, esto significa que el resultado puede ser un vector, un data.frame o una lista según sea más conveniente.

Función C: Permite devolver un dataframe o un vector a partir n parámetros que recibe en su llamada.

Función Paste: une todos los vectores de caracteres que se le suministran y construye una sola cadena de caracteres.

Reportes

- **Pie(Parámetros):** Dibuja un gráfico de pie.

Parámetros:

x: Especifica el vector numérico con los datos a representar.

col: Especifica un vector de colores que serán utilizados en el gráfico.

cex: Describe el tamaño de texto de las etiquetas de datos.

Labels: Vector con los valores de las etiquetas de datos del gráfico.

Main: Título del gráfico.

- **Barplot(Parámetros):** Dibuja un gráfico de barras.

Parámetros:

x: Especifica el vector numérico con los datos a representar.

col: Especifica un vector de colores que serán utilizados en el gráfico.

ylab: Vector con la etiqueta del eje y.

xlab: Vector con la etiqueta del eje x.

xlim: Limite de datos del ejex.

ylim: Limite de datos del ejey.

Main: Título del gráfico.

- **hist(Parámetros):** Dibuja un histograma

Parámetros:

x: Especifica el vector numérico con los datos a representar.

col: Especifica un vector de colores que serán utilizados en el gráfico.

ylab: Vector con la etiqueta del eje y.

xlab: Vector con la etiqueta del eje x.

xlim: Limite de datos del ejex.

ylim: Limite de datos del ejey.

Main: Título del gráfico.

- **pairs(Parámetros):** Dibuja una matriz de dispersión de los valores que son pasados en el parámetro x.

Parámetros:

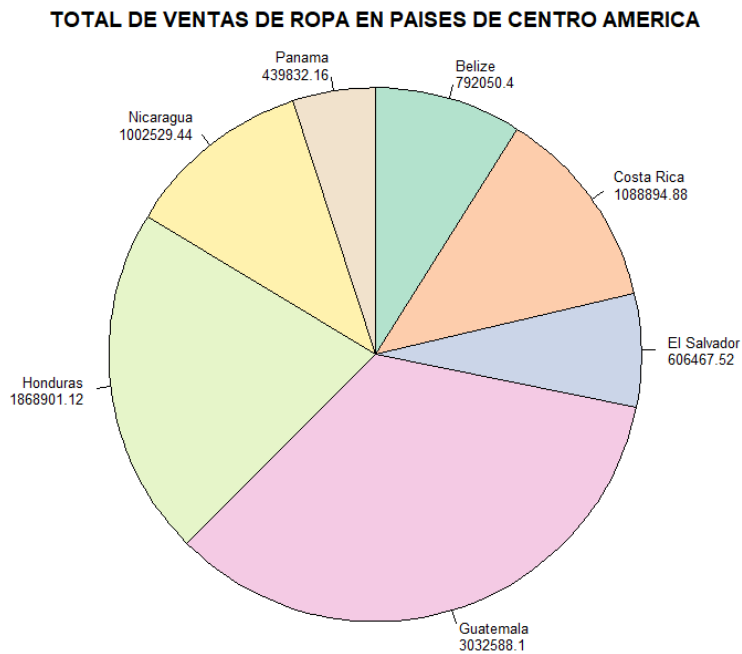
x: Especifica el vector numérico o dataframe con los datos a representar.

- **plot(Parámetros):** Dibuja un gráfico en coordenadas x,y que son pasadas como parámetros.

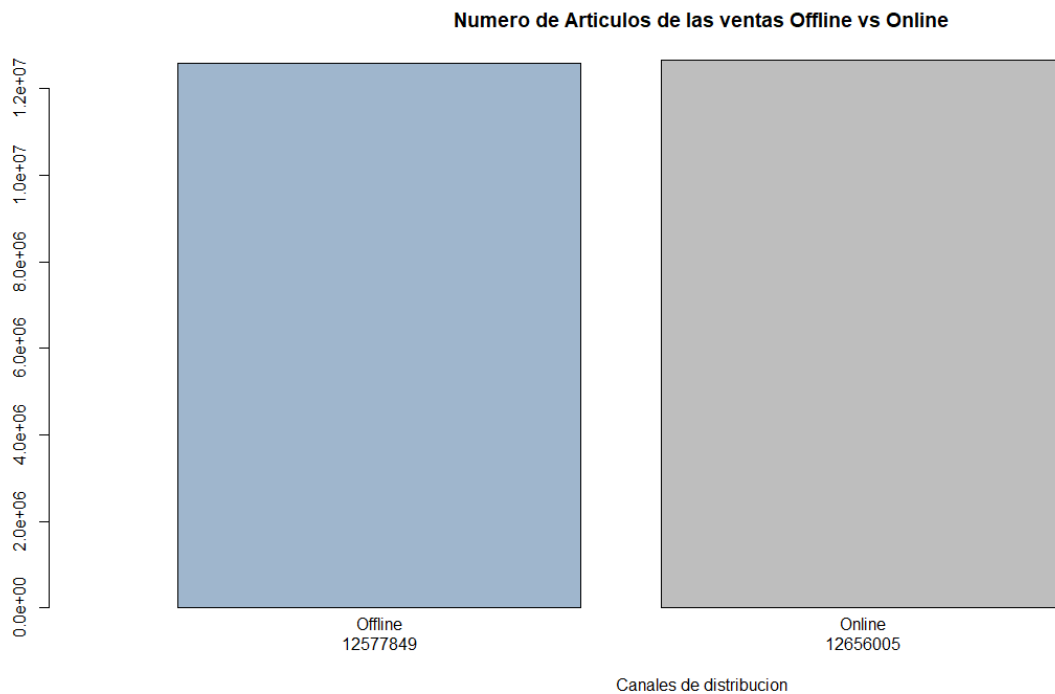
Resultados obtenidos

1. Archivo *sales*:

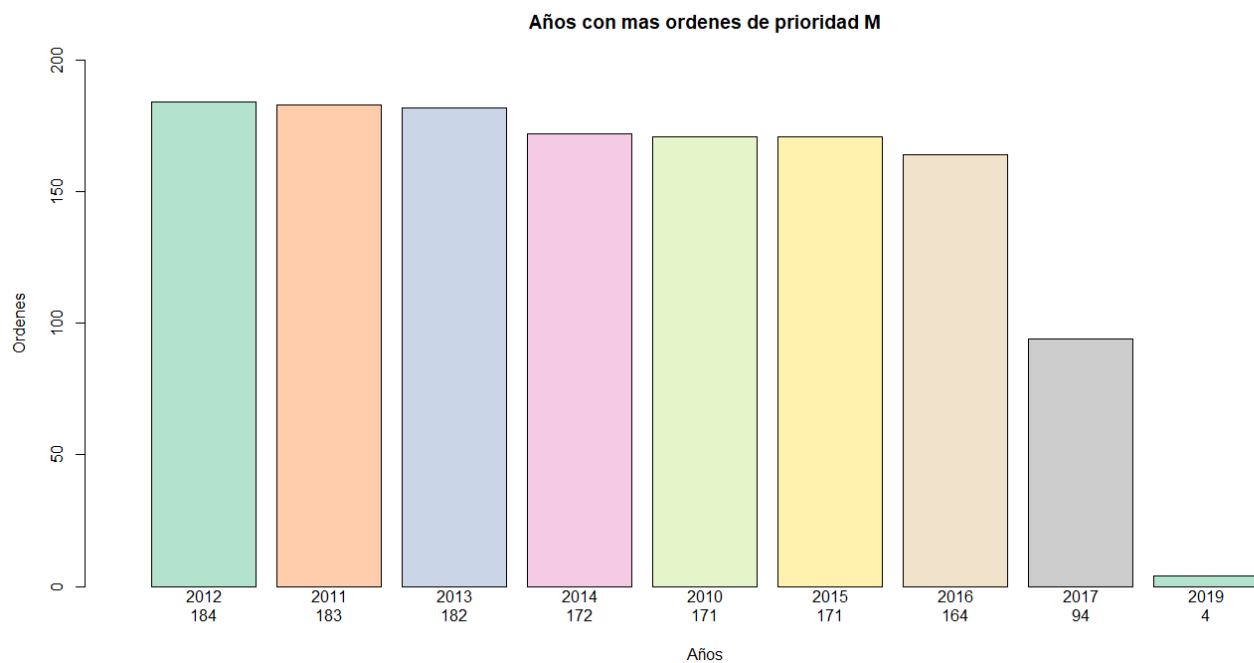
a. Gráfica de pie que muestre el total de ganancias de los países de Centroamérica a partir de las ventas de ropa.



b. Gráfica de barras a su elección, que compare el número de artículos de las ventas en línea contra las ventas que no son en línea.

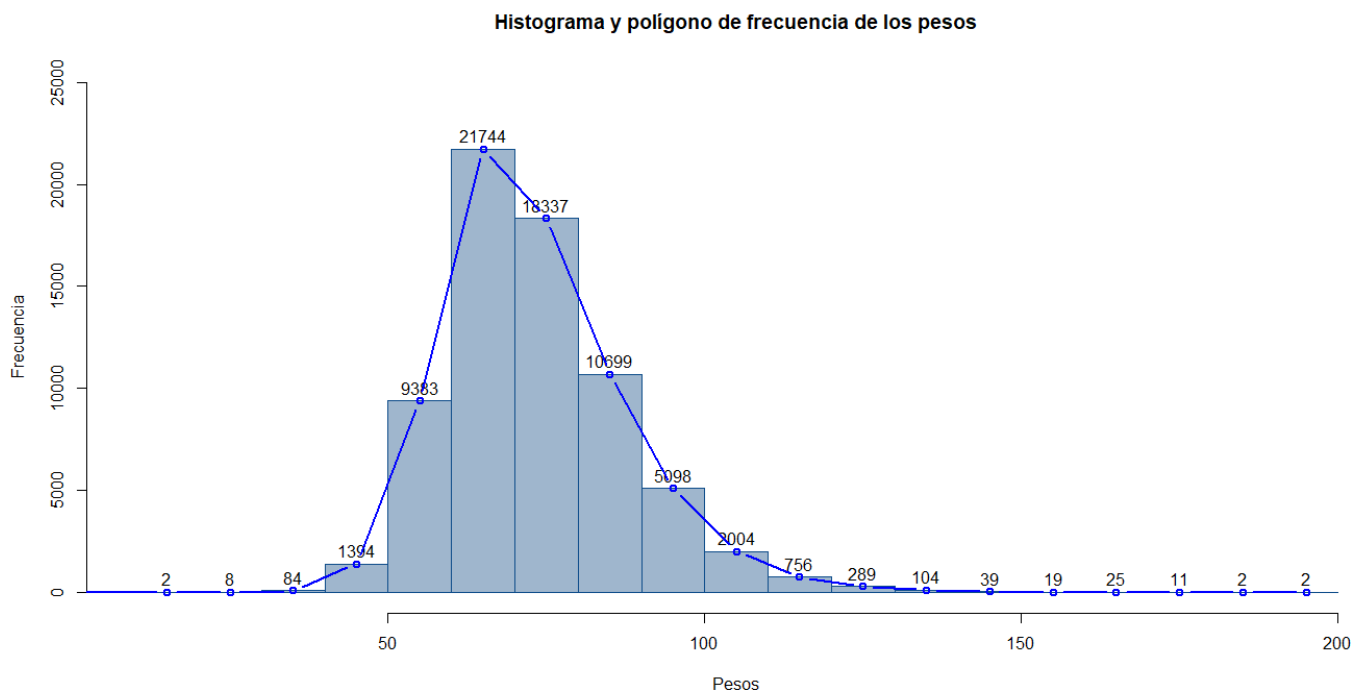


c. Gráfica a su elección que muestre el año con más ordenes de prioridad M.



2. Archivo *cardio_train*:

a. Genere el histograma y polígono de frecuencia (en la misma gráfica) del peso.

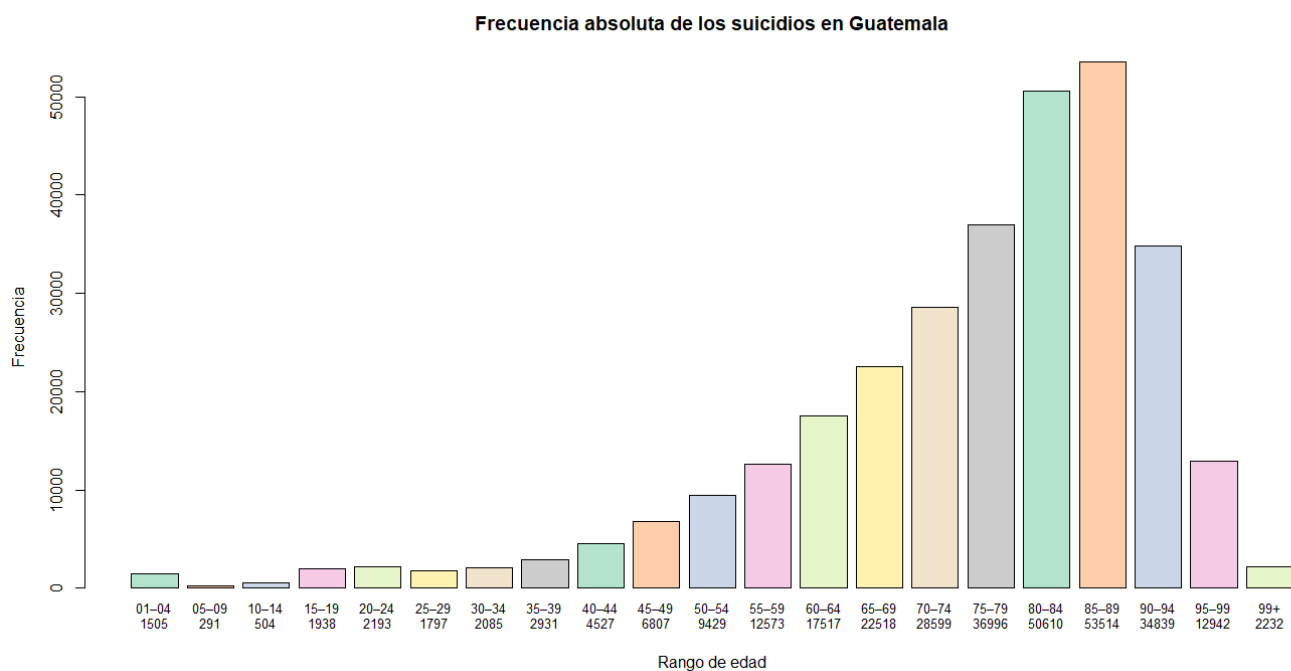


3. Archivo *suicide*:

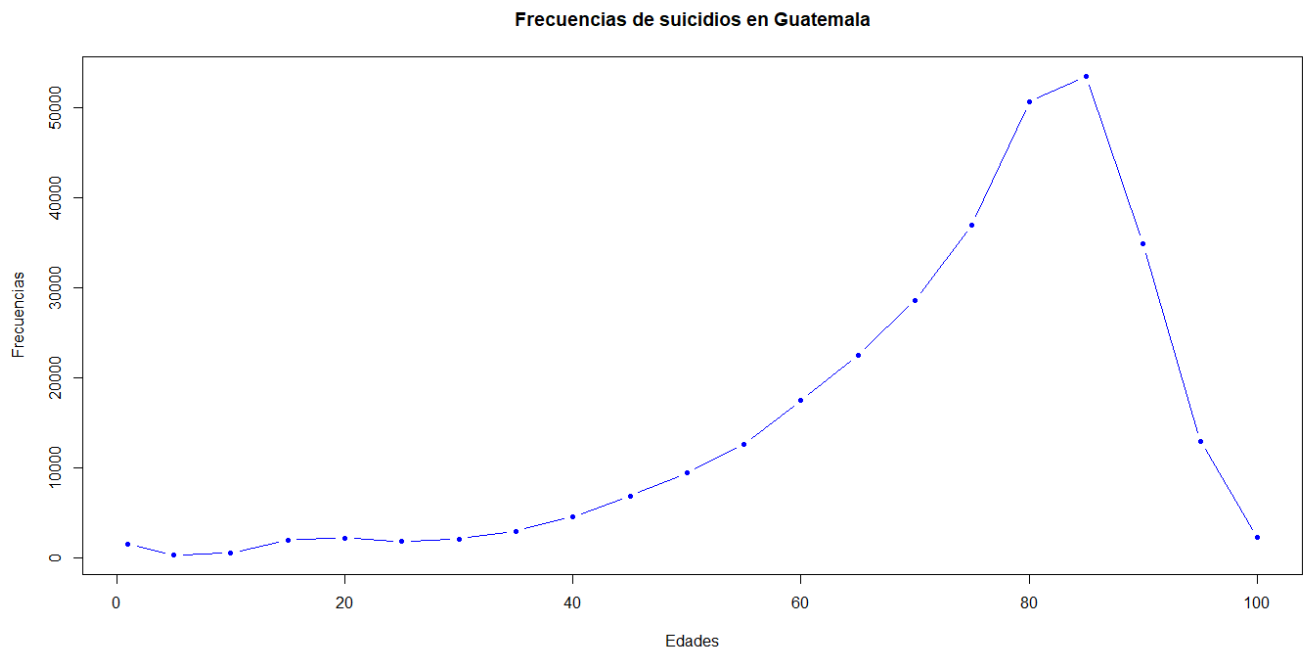
a. Tabla de frecuencias de los rangos de los suicidios en Guatemala (todas las edades)

	Edades	f	fa	fr	fra
1	01-04	1505	1505	0.00491273	0.00491273
2	05-09	291	1796	0.00094990	0.00586263
3	10-14	504	2300	0.00164519	0.00750783
4	15-19	1938	4238	0.00632616	0.01383399
5	20-24	2193	6431	0.00715855	0.02099253
6	25-29	1797	8228	0.00586590	0.02685843
7	30-34	2085	10313	0.00680601	0.03366444
8	35-39	2931	13244	0.00956758	0.04323202
9	40-44	4527	17771	0.01477736	0.05800938
10	45-49	6807	24578	0.02221990	0.08022928
11	50-54	9429	34007	0.03077882	0.11100811
12	55-59	12573	46580	0.04104169	0.15204980
13	60-64	17517	64097	0.05718026	0.20923006
14	65-69	22518	86615	0.07350488	0.28273494
15	70-74	28599	115214	0.09335492	0.37608986
16	75-79	36996	152210	0.12076501	0.49685487
17	80-84	50610	202820	0.16520482	0.66205969
18	85-89	53514	256334	0.17468426	0.83674395
19	90-94	34839	291173	0.11372398	0.95046793
20	95-99	12942	304115	0.04224621	0.99271414
21	99+	2232	306347	0.00728586	1.00000000

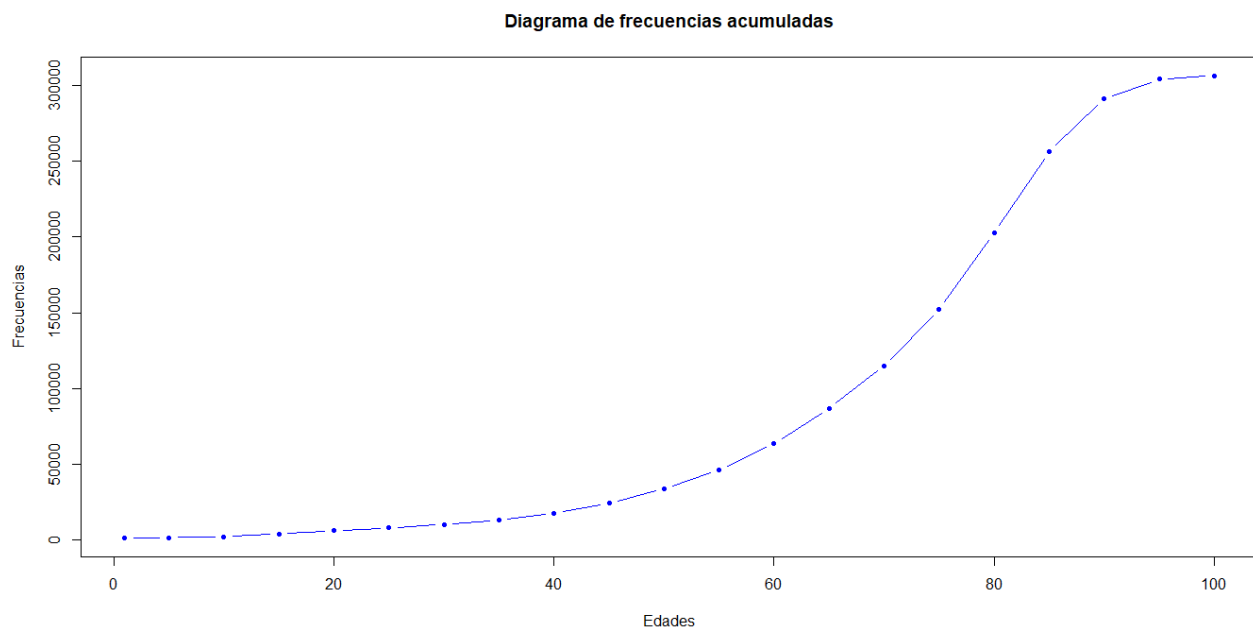
b. Gráfica de barras de la frecuencia absoluta



c. Polígono de frecuencias.

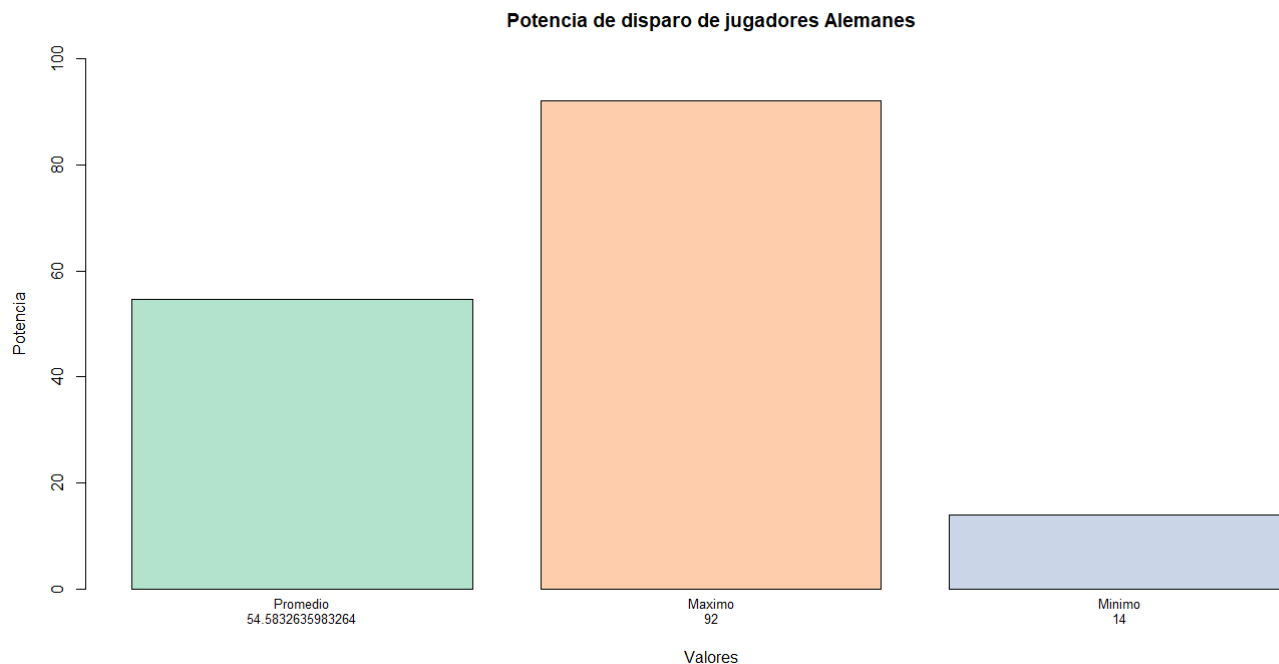


d. Diagrama de frecuencia acumulada.

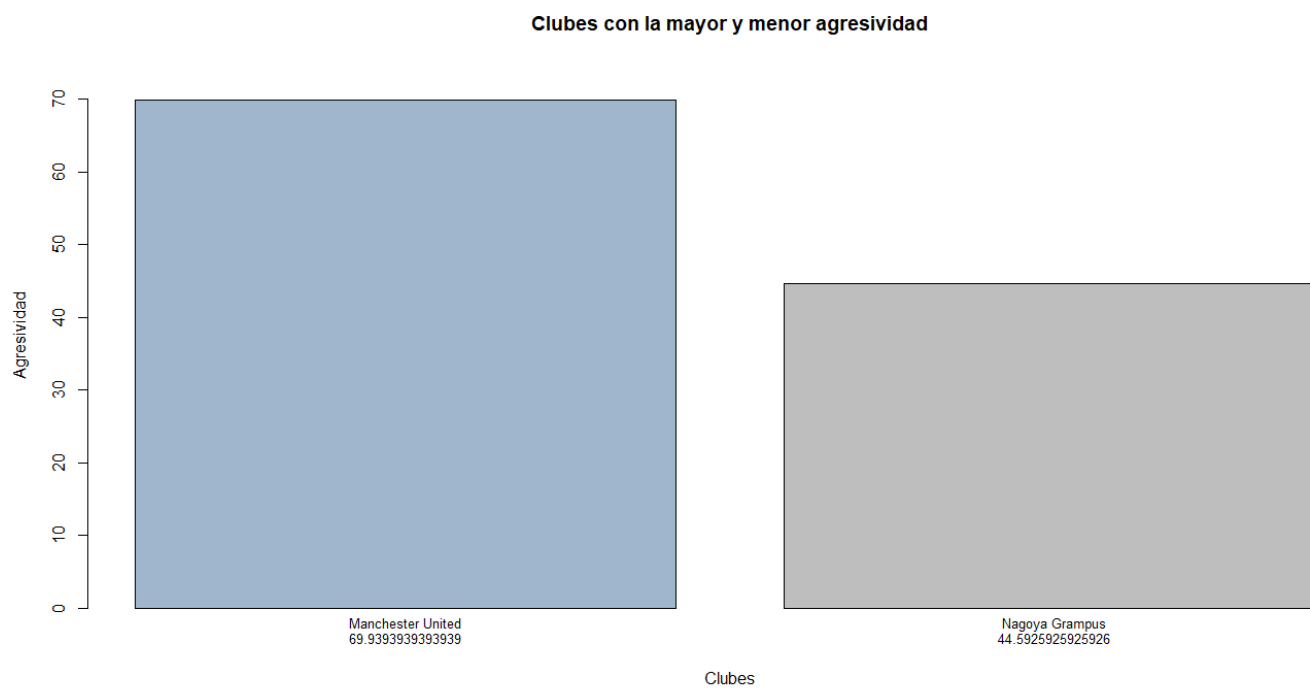


4. Archivo *fifa*:

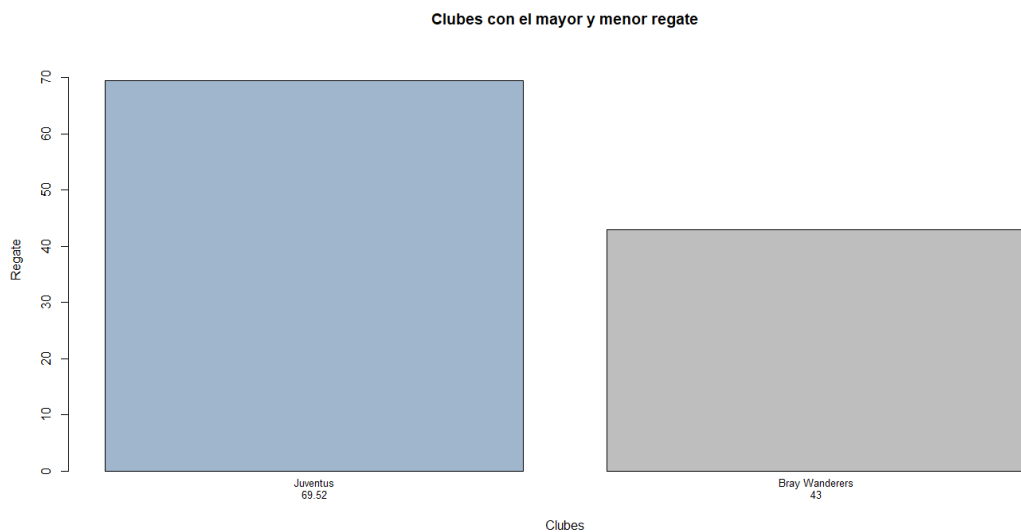
a. Gráfica de barras de la potencia de disparo promedio, máximo y mínimo de los jugadores con nacionalidad alemana.



b. Gráfica que muestre al club con la mayor agresividad promedio y al club con la menor agresividad promedio.

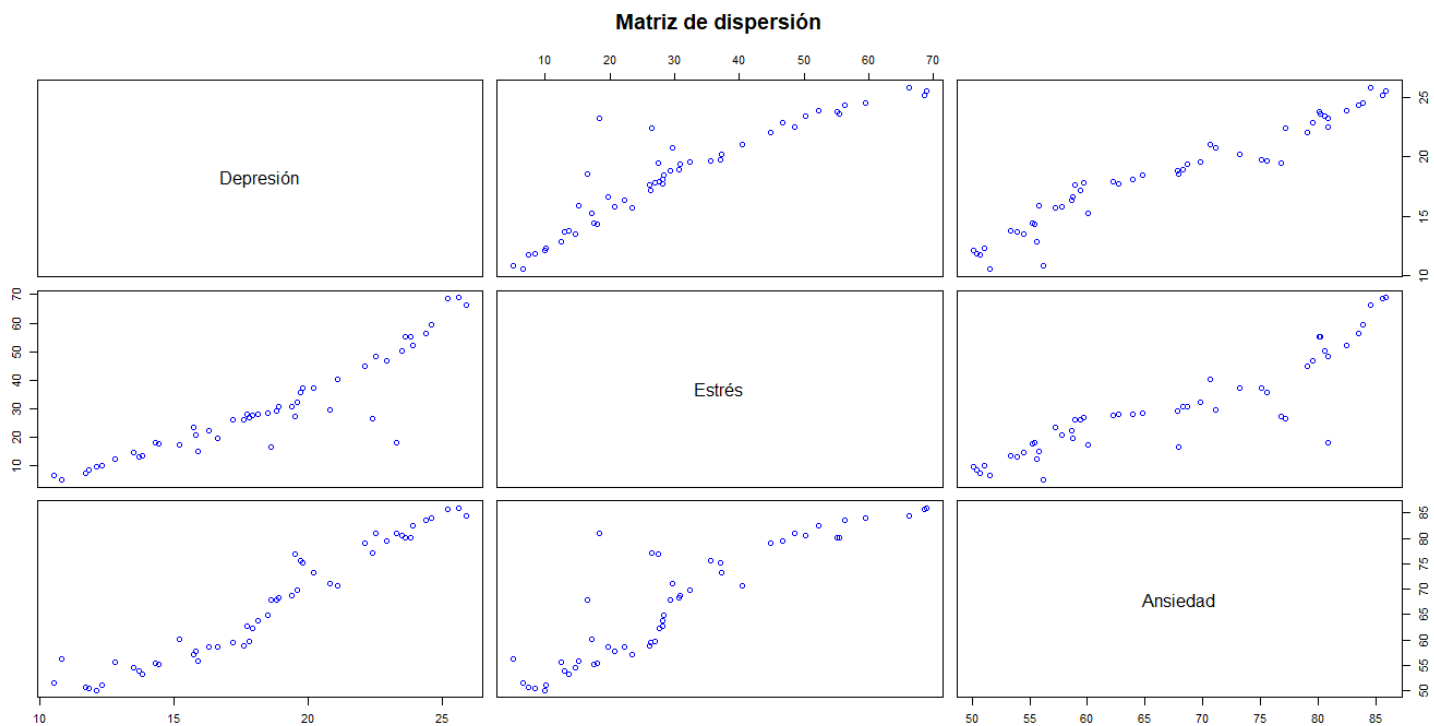


c. Gráfica que muestre al club con el mayor regate promedio y al club con el menor regate promedio.



5. Archivo *regresion*.

a. Análisis exploratorio de datos para comprobar si existe relación entre estrés, depresión y ansiedad.



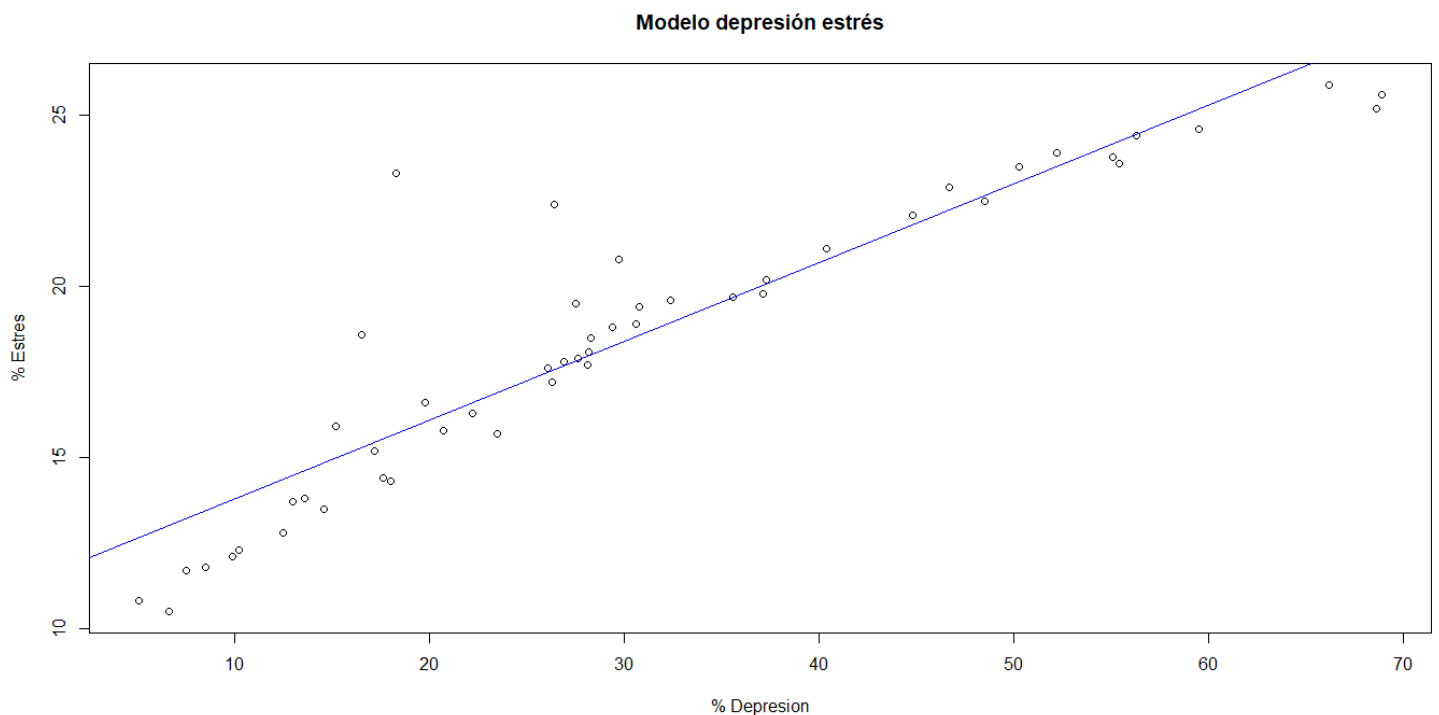
La diagonal Depresión – Estrés – ansiedad, permite ubicar el diagrama de dispersión respectivo, por ejemplo en la fila 1, columna 2 se encuentra la intersección Depresión-Estrés, en la fila 3, columna 2 se encuentra el diagrama Estrés Ansiedad.

b. De existir relación, realice una regresión lineal entre los datos y demuestre que el modelo produce datos válidos.

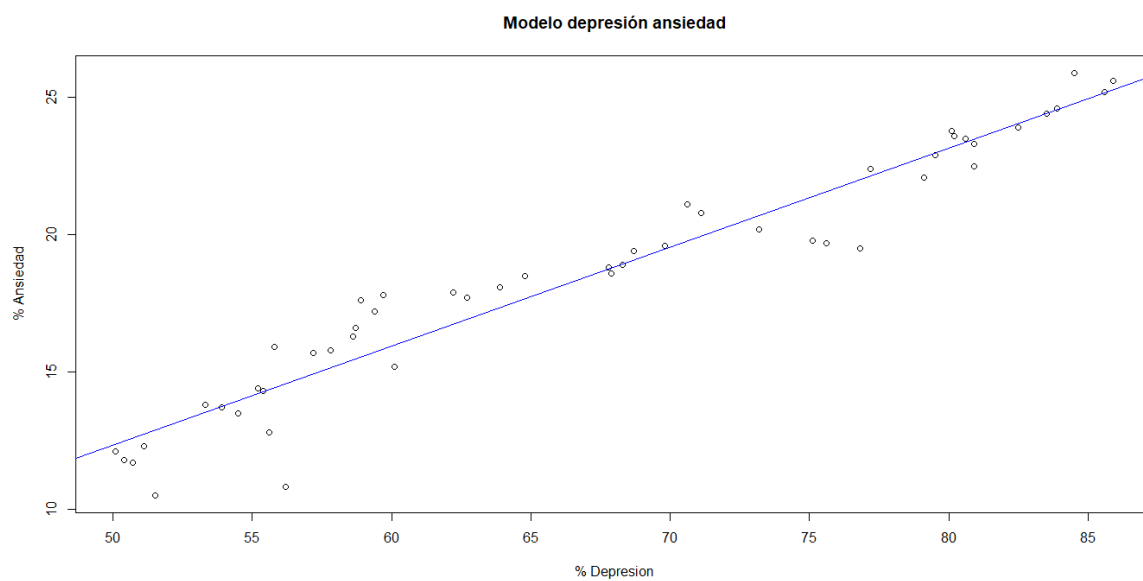
Los diagramas de dispersión del inciso anterior, evidencian que existe una correlación fuerte entre las variables, en las siguientes gráficas, la línea representa un modelo lineal obtenido con el comando "lm".

Se puede observar como varios de los datos reales coinciden con los datos del modelo, de manera que es posible validar el modelo.

Variables: **Depresión, Estrés**



Variables: **Depresión, Ansiedad**



Variables: **Estrés, Ansiedad**

