

**“AÑO DE LA RECUPERACIÓN Y CONSOLIDACIÓN DE LA ECONOMÍA
PERUANA”**
UNIVERSIDAD PRIVADA SAN JUAN BAUTISTA
FACULTAD DE CIENCIAS DE LA SALUD
ESCUELA PROFESIONAL DE MEDICINA HUMANA-EPMH



ASIGNATURA:
“SISTEMATIZACIÓN Y MÉTODOS ESTADÍSTICOS”

GRUPO: 5

INTEGRANTES:

MARÍA LUCIA JACOBO ATUNCAR

GAMBOA CANALES MARIPAZ

ARIANA ABIGIAL VIDAL ROMUCHO

FERNANDA GIANELLA CASTILLA SALVADOR

SEBASTIAN PALOMINO ROJAS

KRISTY STEFANY ALVAREZ PEVES

DOCENTE:

ING. SEGUNDO VICENTE CASTRO LOPEZ

SAN BORJA

2025-2

ÍNDICE:

- 1. Instalar paquetes:**
 - 2. Cargar paquetes:**
 - 3. Importacion de Datos**
 - 4. Resumen de variables categorical**
 - 5. uso de tidyverse**
 - 6. Valores perdidos en proporciones**
 - 7. Resumen de variables continuas**
 - 8. summary**
 - 9. ¿Mediana o media?**
- 10. GTSUMMARY:**

❖ Instalar paquetes: Funciones, paquetes e importe de datos

```
{r}
library(tidyverse)
library(rio)
library(here)
library(janitor)
```

❖ Cargar paquetes:

```
{r}
install.packages("tidyverse")
install.packages("rio")
install.packages("here")
install.packages("janitor")
install.packages("skimr")
install.packages("visdat")
```

Importacion de Datos:

Data	
cirrosis	418 obs. of 20 variables
cirrosis_1	418 obs. of 20 variables
cirrosis_2	418 obs. of 20 variables
cirrosis_3	418 obs. of 20 variables
cirrosis_4	418 obs. of 21 variables

1. Resumen de variables categoricas

```
table(cirrosis_4$estado, useNA = "ifany")
```

{r}		table(cirrosis_4\$estado, useNA = "ifany")	
Censurado	232	Censurado_trasplante	25

Fallecido	161
-----------	-----

{r}		
table(cirrosis_4\$medicamento, useNA = "ifany")		
D_penicilamina	Placebo	<NA>

```
table(cirrosis_4$medicamento, useNA = "ifany")
```

→ uso de tidyverse

```
{r}
cirrosis_4 |>
  dplyr::count(ascitis, sort = TRUE)
```

A tibble: 3 × 2

ascitis	n
No	288
NA	106
Sí	24

3 rows

→ Valores perdidos en proporciones

```
{r}
prop.table(table(cirrosis_4$ascitis, useNA = "ifany"))
```

No	Sí	<NA>
0.68899522	0.05741627	0.25358852

2. Resumen de variables continuas

```
{r}
mean(cirrosis_4$dias_seguimiento, na.rm = TRUE)
```

[1] 1917.782

```
{r}
sd(cirrosis_4$dias_seguimiento, na.rm = TRUE)
```

[1] 1104.673

Otras funciones que puedes encontrar útil son **min()**, **max()**, **median()** y **IQR()**

```
{r}
min(cirrosis_4$edad, na.rm = TRUE) # Proporciona el valor mínimo
```

```
IQR(cirrosis_4$edad, na.rm = TRUE) # Calcula el rango intercuartilico
```

```
median(cirrosis_4$edad, na.rm = TRUE) # Calcula la mediana
```

[1] 9598

[1] 5628

[1] 18628

- ❖ **La función mean():** se utiliza para calcular la media aritmética de un vector de datos.
- ❖ **La función sd():** se emplea para calcular la desviación estándar de un conjunto de datos.
- ❖ **min():** Esta función devuelve el valor mínimo de un vector.
- ❖ **median():** Para encontrar la mediana de un conjunto de datos.
- ❖ **IQR():** Para calcular el rango intercuartil, que es la diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1).

➤ **La función `summary()` para calcular todas las medidas, a la vez.**

summary(): Es una función nativa de R que permite calcular todas las medidas estadísticas a la vez. Abajo, un ejemplo con la variable hepatomegalia

```
{r}
summary(cirrosis_4$hepatomegalia)

Length     Class      Mode
 418 character character
```

➤ **¿Mediana o media?:** Una de las cosas que queremos saber al realizar un análisis descriptivo es la distribución de los datos. La media y mediana puede informarnos al respecto. Abajo, un ejemplo para la variable EDAD

```
{r}
mean(cirrosis_4$edad, na.rm = T)

[1] 18533.35
```

```
{r}
median(cirrosis_4$edad, na.rm = T)

[1] 18628
```

La media y mediana de la variable insulina son diferentes.

Estos sugiere que los datos tienen una distribución asimétrica:

`par(mfrow=c(1,2)) # Configuramos el número de elementos por Figura`

- `hist(cirrosis_4$edad)`
- `hist(cirrosis_4$sgot)`

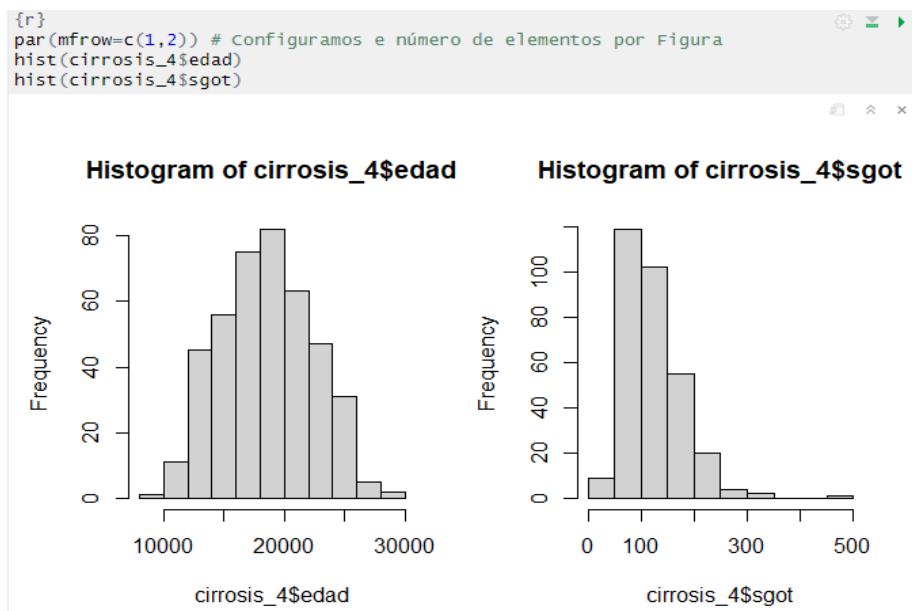


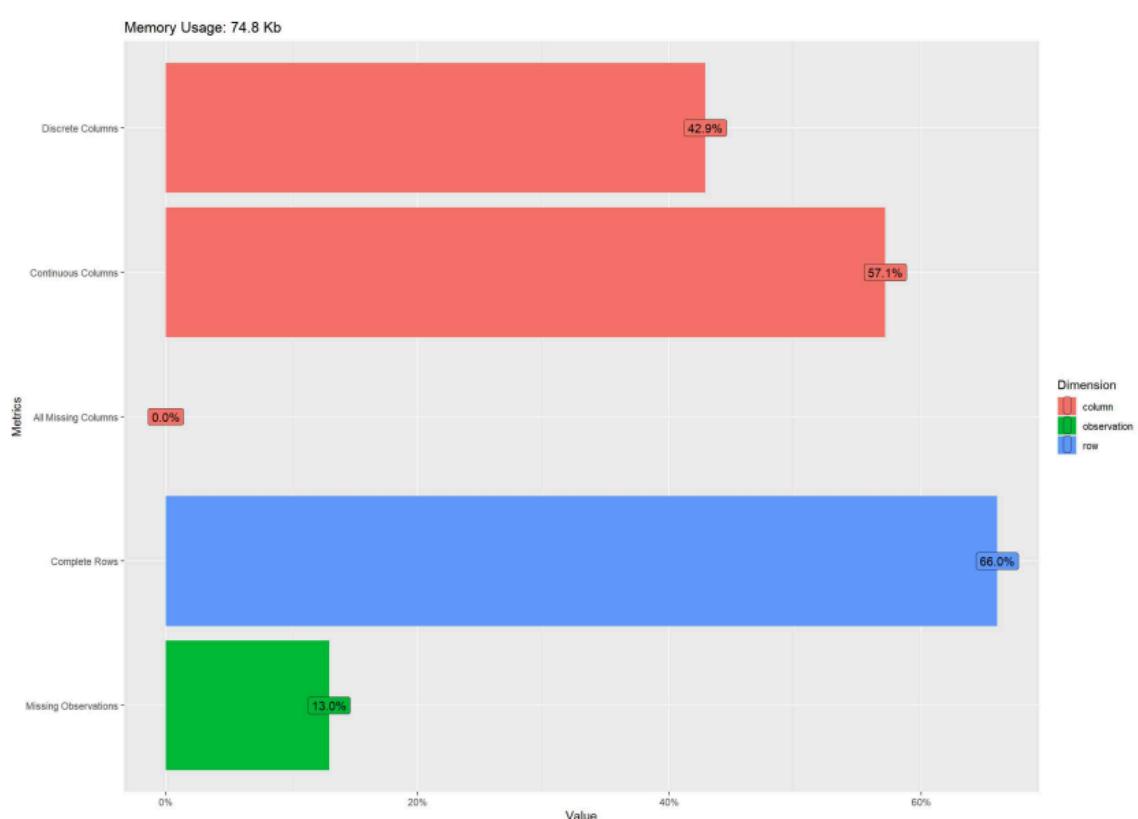
TABLA 1 - GTSUMMARY: facilita la creación de tablas analíticas y de resumen listas para publicar. Permite resumir conjuntos de datos, modelos de regresión y otros, ofreciendo valores predeterminados y opciones altamente personalizables.

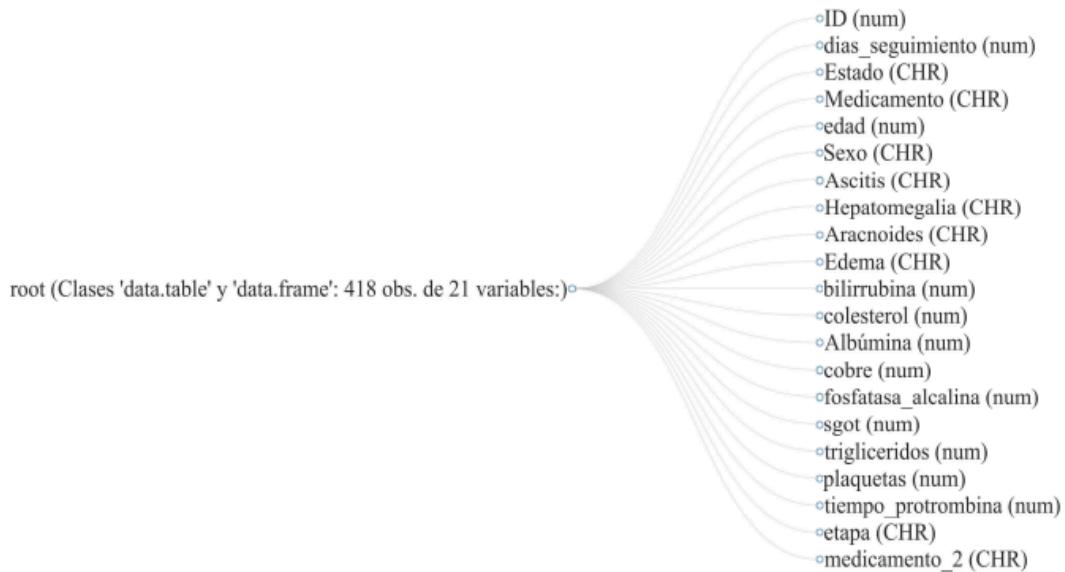
```
{r}
cirrosis_4 |>
  select(edad, estado, medicamento_2, medicamento, etapa,
         plaquetas, trigliceridos) |>
 tbl_summary()
```

Characteristic	N = 418 ¹
edad	18,628 (15,628, 21,281)
estado	
Censurado	232 (56%)
Censurado_trasplante	25 (6.0%)
Fallecido	161 (39%)
medicamento_2	
medicado	312 (100%)
Unknown	106
medicamento	

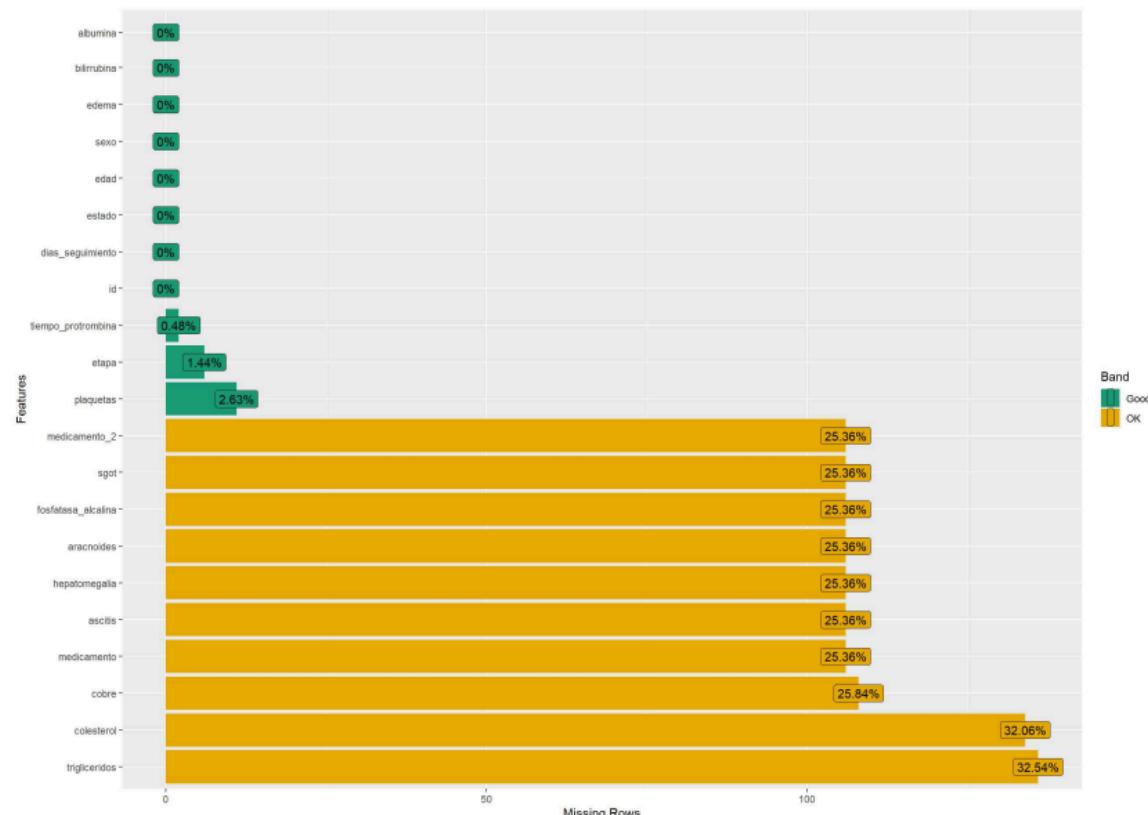
TABLA 2 - FLEXTABLE: Es un paquete y una función que facilitan la creación de informes tabulares.

id	dias_seguimiento	estado	medicamento	edad	sexo	ascitis	hep
1	400	Fallecido	D_penicilamina	21,464	Mujer	Sí	Sí
2	4,500	Censurado	D_penicilamina	20,617	Mujer	No	Sí
3	1,012	Fallecido	D_penicilamina	25,594	Hombre	No	No
4	1,925	Fallecido	D_penicilamina	19,994	Mujer	No	Sí
5	1,504	Censurado_trasplante	Placebo	13,918	Mujer	No	Sí
6	2,503	Fallecido	Placebo	24,201	Mujer	No	Sí
7	1,832	Censurado	Placebo	20,284	Mujer	No	Sí
8	2,466	Fallecido	Placebo	19,379	Mujer	No	No





Perfil de datos faltante



Distribución univariante

Histograma

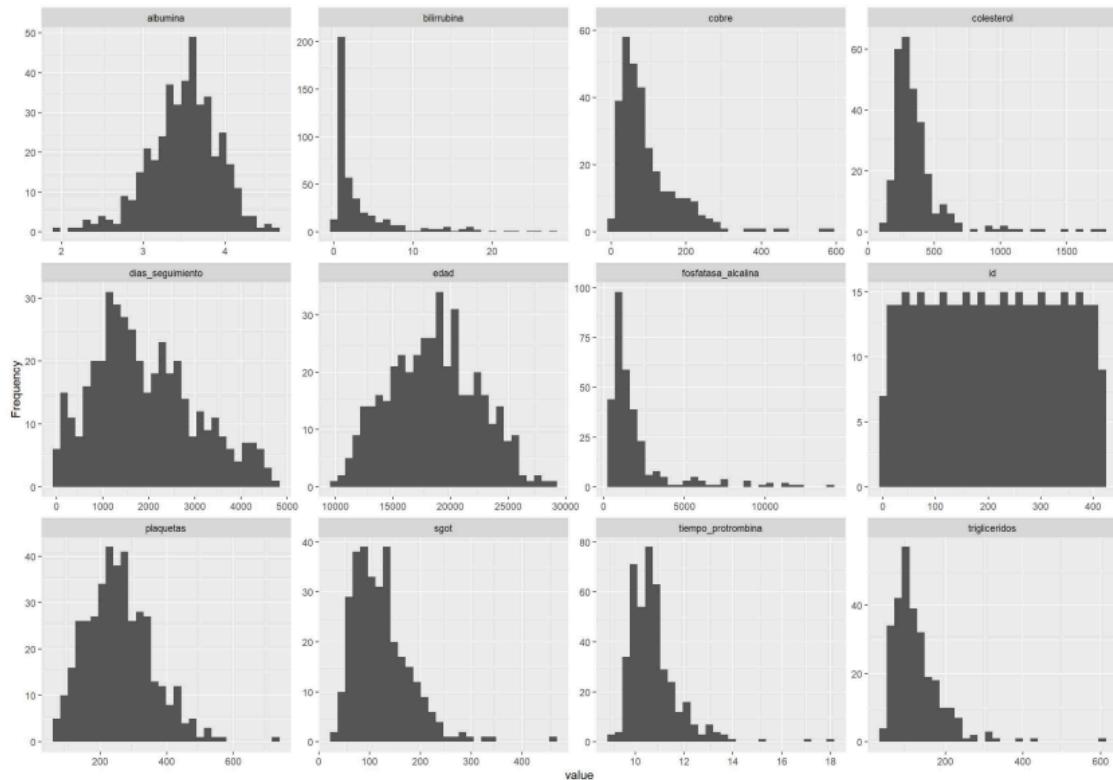


Gráfico de barras (con frecuencia)

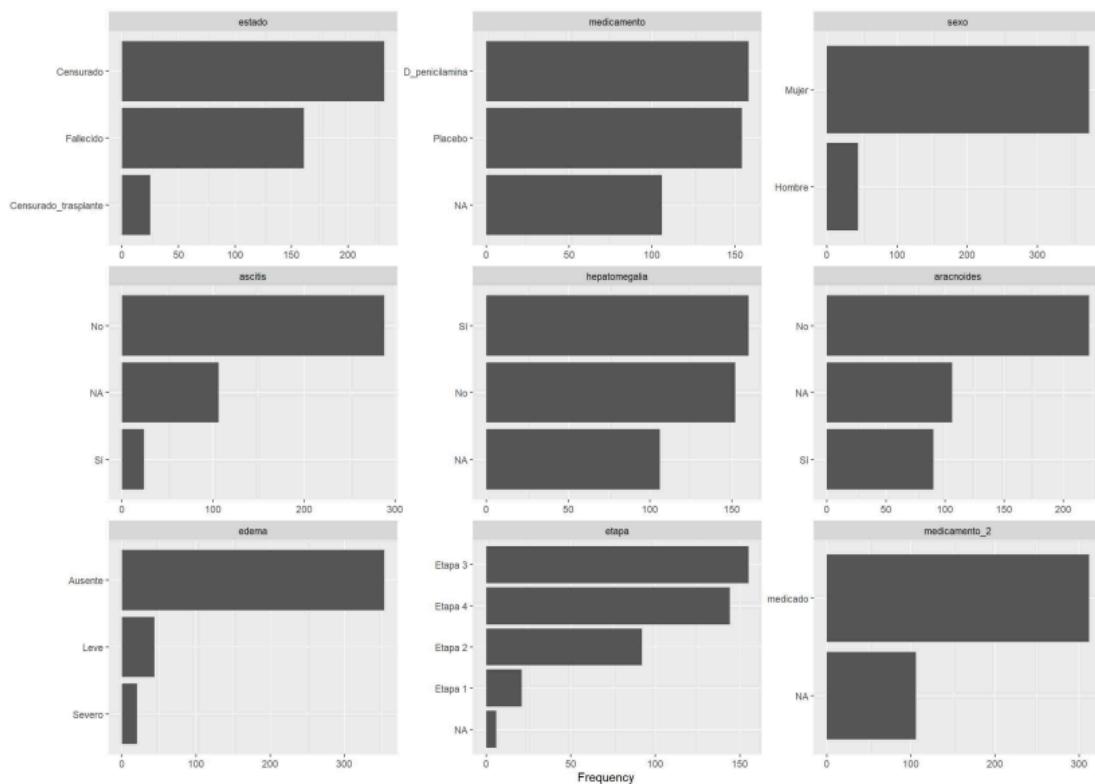
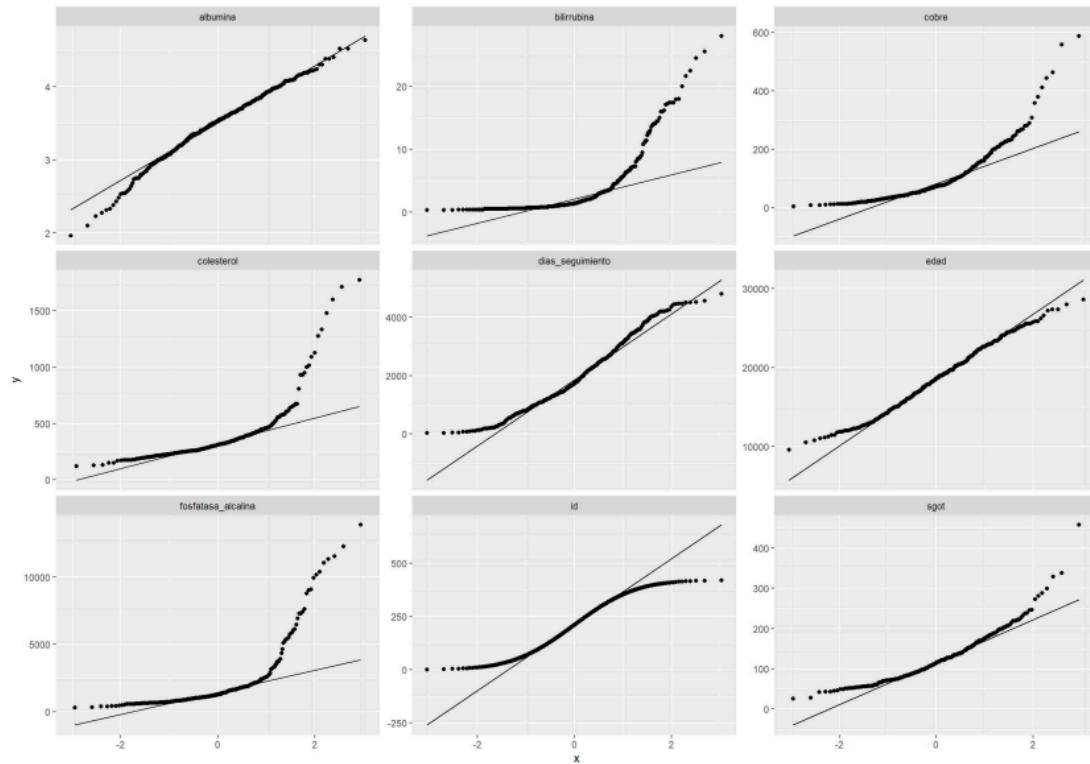


Gráfico QQ

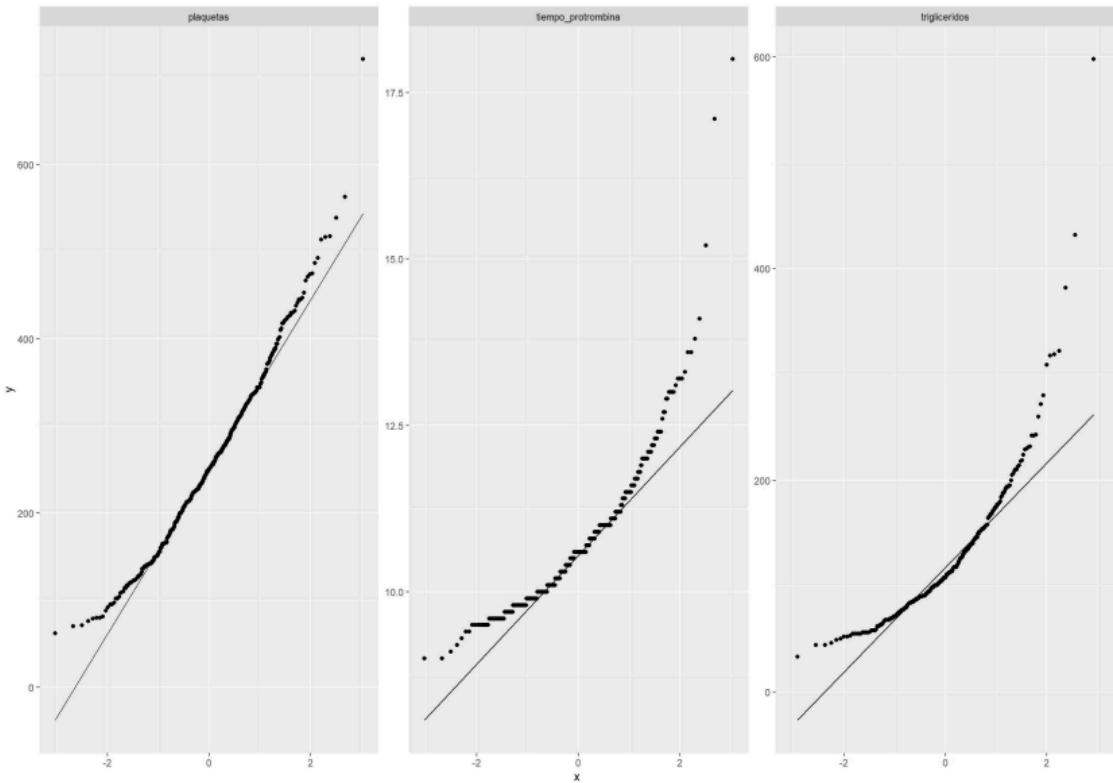
```
## Warning: Removed 454 rows containing non-finite outside the
## scale range (`stat_qq()`).
```

```
## Warning: Removed 454 rows containing non-finite outside the
## scale range (`stat_qq_line()`).
```



```
## Warning: Removed 149 rows containing non-finite outside the
## scale range (`stat_qq()`).
```

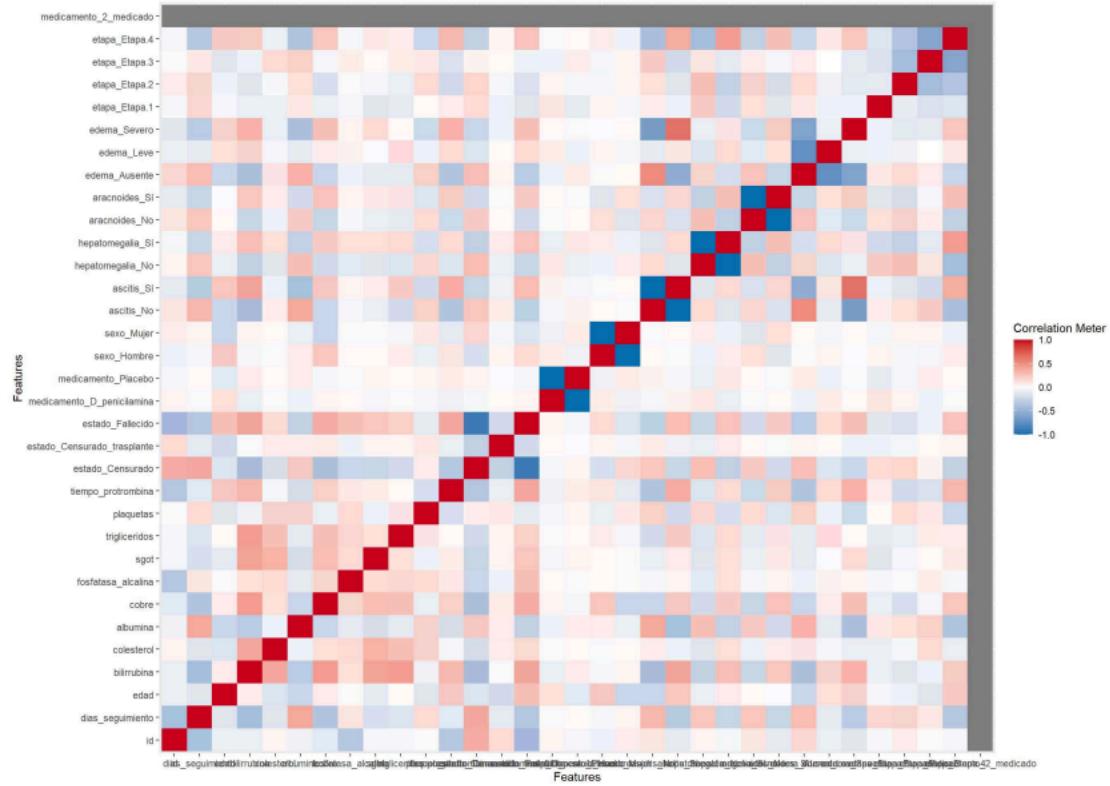
```
## Warning: Removed 149 rows containing non-finite outside the
## scale range (`stat_qq_line()`).
```



Correlation Analysis

```

## Warning in cor(x = structure(list(id = c(1, 2, 3, 4, 5,
## 7, 8, 9, 10, 11, : La desviación estándar es cero
  
```



Principal Component Analysis

```
## Warning in plot_prcmp(data = structure(list(id = c(1, 2, 3, 4, 5, 7, 8, : The following features are dropped due to zero variance:
## * medicamento_2_medicado
```

