

SISTEMATIZACION Y MÉTODOS ESTADÍSTICOS

 SAN JUAN BAUTISTA



UNIVERSIDAD PRIVADA

ESTUDIANTES:

- MARÍA LUCIA JACOBO ATUNCAR
- GAMBOA CANALES MARIPAZ
- ARIANA ABIGIAL VIDAL ROMUCHO
- FERNANDA GIANELLA CASTILLA SALVADOR
- SEBASTIAN PALOMINO ROJAS
- KRISTY STEFANY ALVAREZ PEVES

CARGAR PAQUETES Y IMPORTAR LA DATA



Cargar los paquetes

```
{r}
library(tidyverse)
library(here)
library(rio)
library(gtsummary)
library(car)
library(survival)
```

pacientes total
de : 418

```
{r}
cirrosis_4 <- import(here("cirrosi.csv"))
```

Un vistazo a los datos

```
{r}
head(cirrosis_4)
```

A tibble: 6 × 21

	id	dias_seguimie...	estado	medicamento	edad
	<dbl>	<dbl>	<chr>	<chr>	<dbl>
1	1	400	Fallecido	D_penicilamina	21464
2	2	4500	Censurado	D_penicilamina	20617
3	3	1012	Fallecido	D_penicilamina	25594
4	4	1925	Fallecido	D_penicilamina	19994
5	5	1504	Censurado_tras...	Placebo	13918
6	6	2503	Fallecido	Placebo	24201

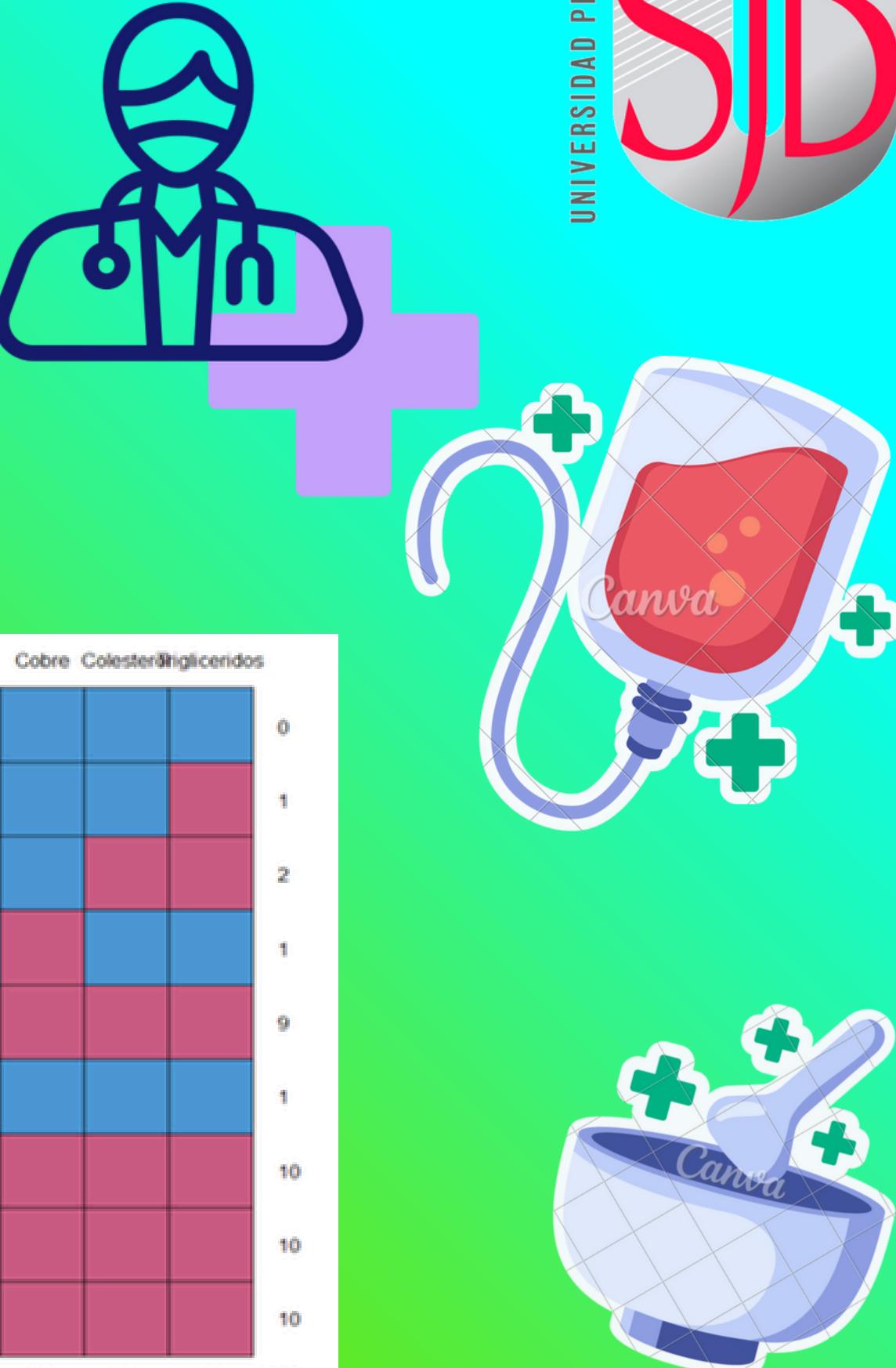
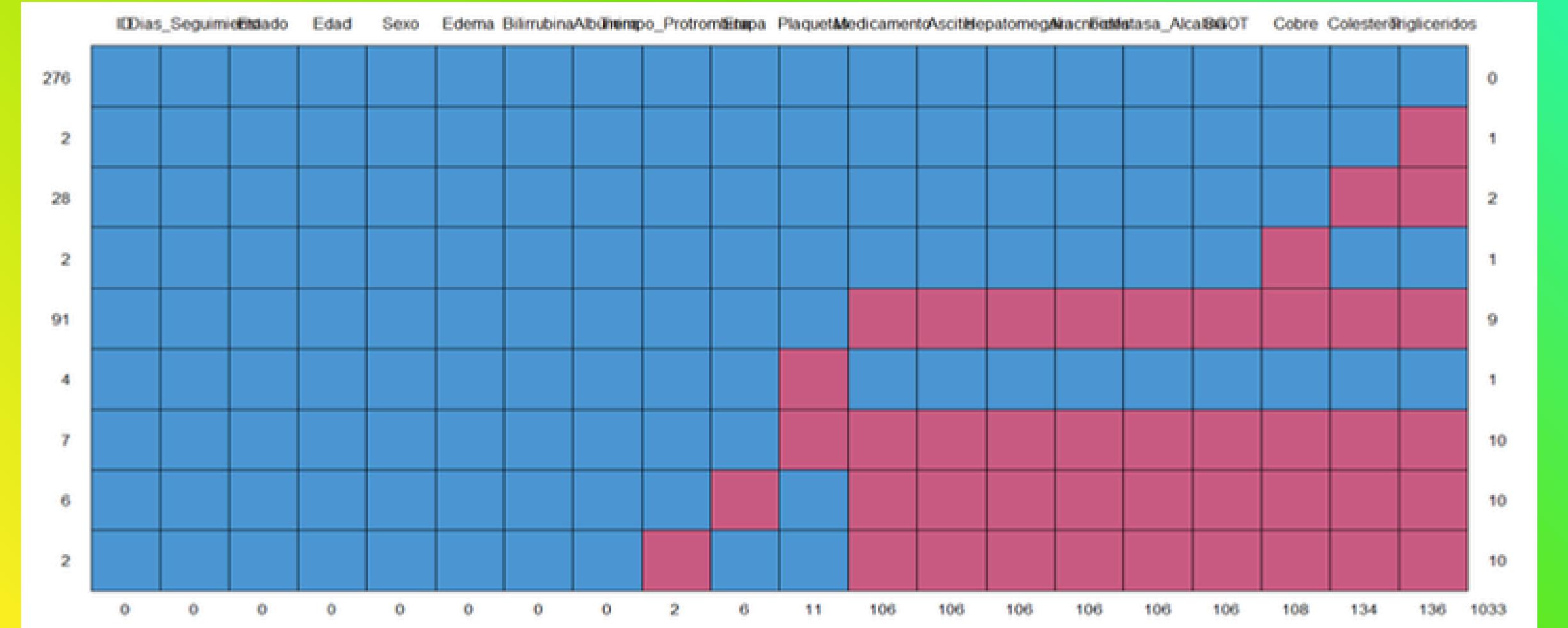
6 rows | 1-5 of 21 columns

Uso de MICE para observar el patrón de datos faltante

El paquete proporciona una buena función para obtener una mejor comprensión del patrón de datos faltantes [micemd.pattern\(\)](#).

```
{r}
md.pattern(cirrosis) #Para poder visualiza el patrón de datos faltantes
```

	ID	Dias_Seguimiento	Estado	Edad
276	1		1	1
2	1		1	1
28	1		1	1
2	1		1	1
91	1		1	1
4	1		1	1
7	1		1	1
6	1		1	1
2	1		1	1
0	0	0	0	0
	Sexo	Edema	Bilirrubina	Albumina
276	1	1	1	1
2	1	1	1	1
28	1	1	1	1
2	1	1	1	1
91	1	1	1	1
4	1	1	1	1
7	1	1	1	1
6	1	1	1	1
2	1	1	1	1
0	0	0	0	0
	Tiempo_Protrombina	Etapa		
276		1	1	
2		1	1	
28		1	1	
2		1	1	
91		1	1	
4		1	1	
7		1	1	
6		1	0	
2		0	1	
		2	6	
	Plaquetas	Medicamento	Ascitis	
276	1	1	1	
2	1	1	1	
28	1	1	1	
2	1	1	1	
91	1	0	0	
4	0	1	1	
7	0	0	0	
6	1	0	0	
2	1	0	0	
11	106	106	106	



Imputación de los datos que faltan

La función se encarga del proceso de imputación ***mice()***

```
{r}  
imputaciones <- mice(cirrosis, m = 5, method = "pmm", maxit = 10, seed = 123)
```

```
{r}
data_completo <- complete(imputaciones, action = "long") # o usa 1, 2, ..., m para
elegir una imputación específica
```

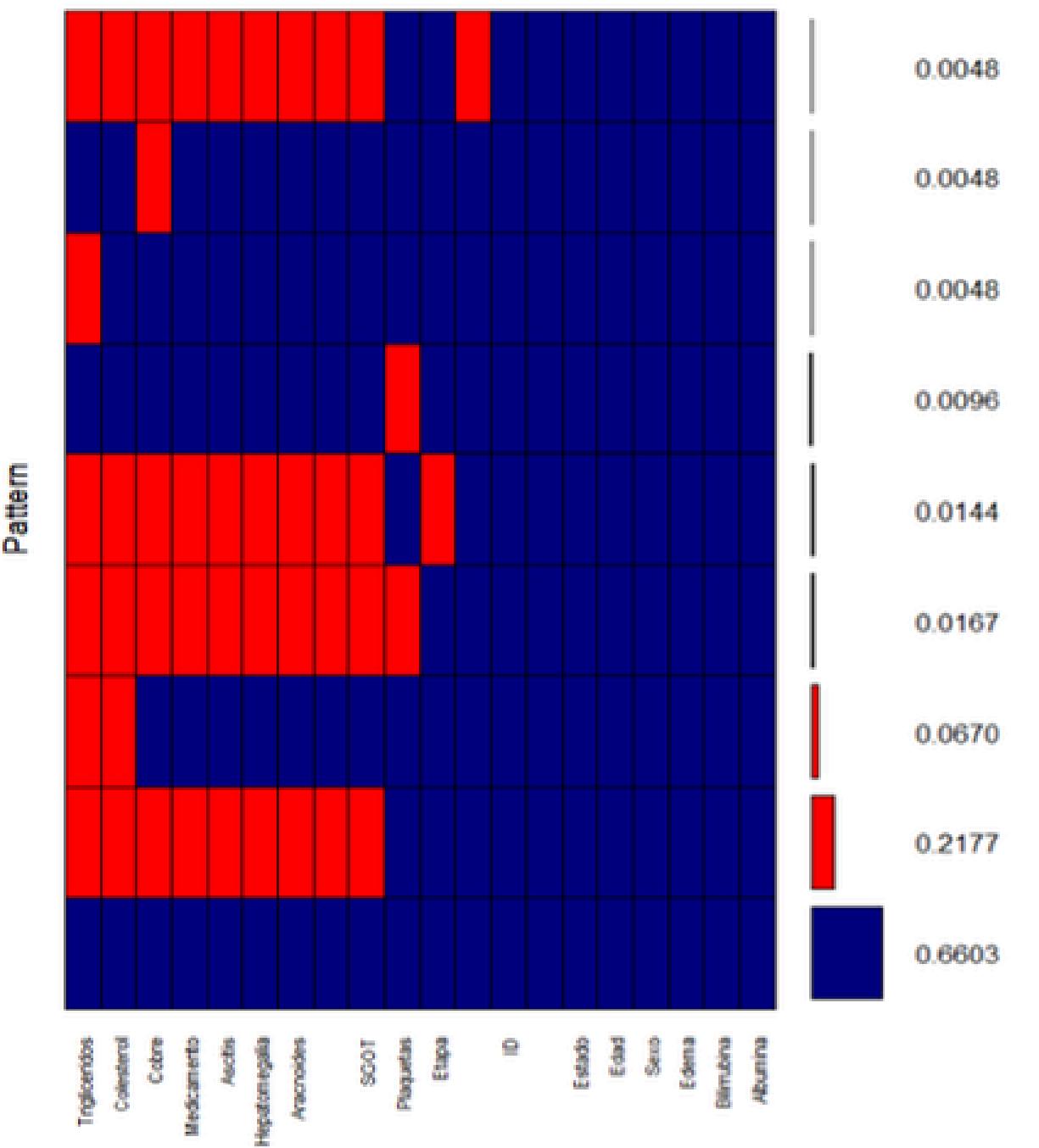
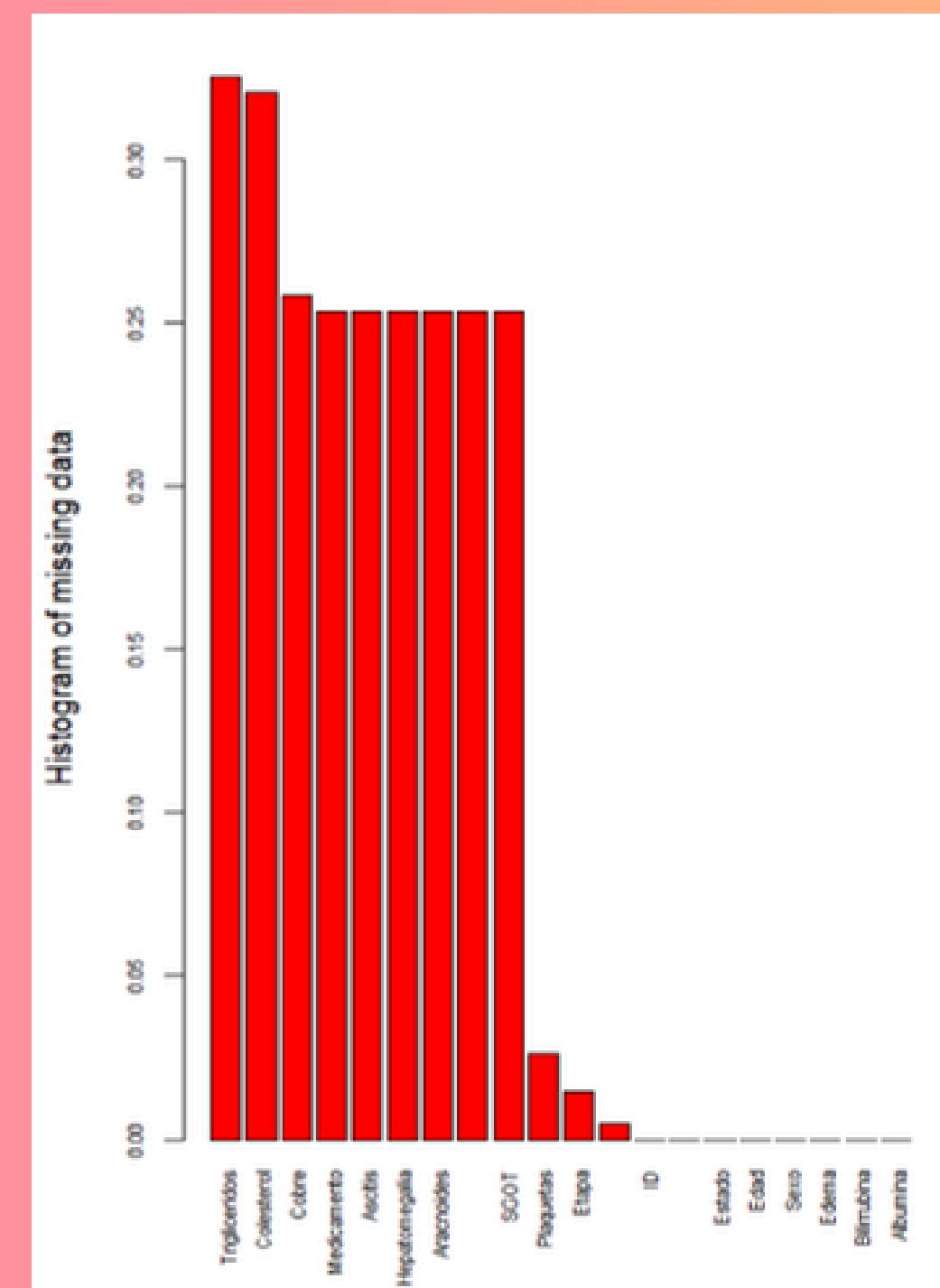
```
{r}
library(vIM)
aggr_plot <- aggr(cirrosis, col=c('navyblue','red'), numbers=TRUE, sortVars=TRUE,
labels=names(cirrosis), cex.axis=.7, gap=3, ylab=c("Histogram of missing data",
"Pattern"))
```

R Console

Description: df [20 x 2]

Variable	Count
Trigliceridos	0.325358852
Colesterol	0.320574163
Cobre	0.258373206
Medicamento	0.253588517
Ascitis	0.253588517
Hepatomegalia	0.253588517
Aracnoides	0.253588517
Fosfatasa_Alcalina	0.253588517
SGOT	0.253588517
Plaquetas	0.026315789
Etapa	0.014354067
Tiempo_Protrombina	0.004784689
ID	0.000000000
Dias_Seguimiento	0.000000000
Estado	0.000000000
Edad	0.000000000
Sexo	0.000000000
Edema	0.000000000
Bilirrubina	0.000000000
Albumina	0.000000000

20 rows





La función se encarga del proceso de imputación mice().

```
{r}  
tempData <- mice(cirrosis,m=5,maxit=50,meth='pmm',seed=500)  
summary(tempData)
```



```
iter imp variable  
1 1 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
1 2 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
1 3 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
1 4 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
1 5 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
2 1 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
2 2 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
2 3 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
2 4 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
2 5 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
3 1 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
3 2 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
3 3 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
3 4 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
3 5 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
4 1 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
4 2 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
4 3 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
4 4 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
4 5 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
5 1 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
5 2 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
5 3 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
5 4 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
5 5 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
6 1 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
6 2 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
6 3 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
6 4 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
6 5 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
7 1 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
7 2 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
7 3 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
7 4 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
7 5 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
8 1 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
8 2 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
8 3 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
8 4 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
8 5 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
9 1 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
9 2 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
9 3 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
9 4 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina  
9 5 Colesterol Cobre Fosfatasa_Alcalina SGOT Trigliceridos Plaquetas Tiempo_Protrombina
```

Description: df [6 x 5]

	it	im	dep	meth	out
	<dbl>	<dbl>	<chr>	<chr>	<chr>
1	0	0		constant	Estado
2	0	0		constant	Medicamento
3	0	0		constant	Sexo
4	0	0		constant	Ascitis
5	0	0		constant	Hepatomegalia
6	0	0		constant	Aracnoides

6 rows

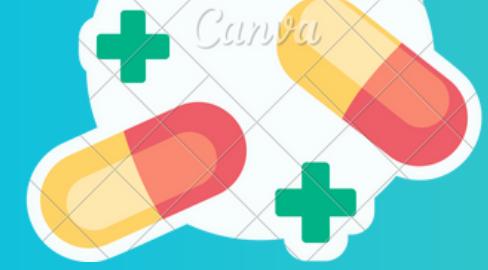




comprobar los datos imputados, por ejemplo, para la variable Ozono, debe introducir la siguiente línea de código

```
{r}  
tempData$imp$Ozone  
  
NULL
```





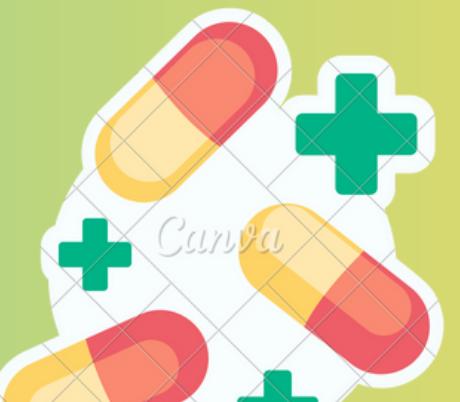
Verificar el método de imputación utilizado para cada variable, es muy fácil de hacer mice

```
{r}  
tempData$meth
```

	ID	Dias_Seguimiento	Estado	Medicamento	Edad
	""	""	""	""	""
	Sexo	Ascitis	Hepatomegalia	Aracnoides	Edema
Bilirrubina		Colesterol	Albumina	Cobre	Fosfatasa_Alcalina
	"pmm"	"pmm"	"pmm"	"pmm"	"pmm"
SGOT		Trigliceridos	Plaquetas	Tiempo_Protrombina	Etapa
"pmm"		"pmm"	"pmm"	"pmm"	"pmm"

```
{r}  
modelFit1 <- with(tempData, lm(ID ~ Dias_Seguimiento + Edad + Colesterol ))  
summary(pool(modelFit1))
```

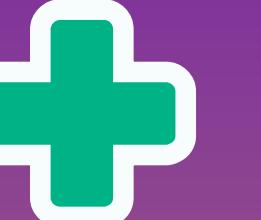
	term	estimate	std.error	statistic	df	p.value
1	(Intercept)	3.113582e+02	35.377093811	8.8011233	297.09657	1.135239e-16
2	Dias_Seguimiento	-4.021995e-02	0.005163621	-7.7890980	407.22354	5.659849e-14
3	Edad	-6.512596e-04	0.001497965	-0.4347628	404.25044	6.639666e-01
4	Colesterol	-3.460544e-02	0.027584315	-1.2545331	55.29269	2.149256e-01



{r}

```
tempData2 <- mice(cirrosis,m=50,seed=245435)
modelFit2 <- with(tempData2,lm(ID ~ Dias_Seguimiento+ Edad +Colesterol))
summary(pool(modelFit2))
```

	term	estimate	std.error	statistic	df	p.value
1	(Intercept)	3.000504e+02	37.479680204	8.0056828	289.5026	2.905240e-14
2	Dias_Seguimiento	-3.952033e-02	0.005218742	-7.5727707	397.2463	2.585144e-13
3	Edad	-4.472936e-04	0.001513843	-0.2954690	395.8267	7.677906e-01
4	Colesterol	-1.791796e-02	0.033446018	-0.5357277	105.6486	5.932728e-01



MUCHAS GRACIAS

