EjercicioMICE

Silvia Pineda

Lectura y carga de librerías

```
library(mice)
```

```
Attaching package: 'mice'

The following object is masked from 'package:stats':

    filter

The following objects are masked from 'package:base':

    cbind, rbind
```

```
data<-nhanes
```

1. Inspecciona si las variables están bien declaradas

```
str(data)
```

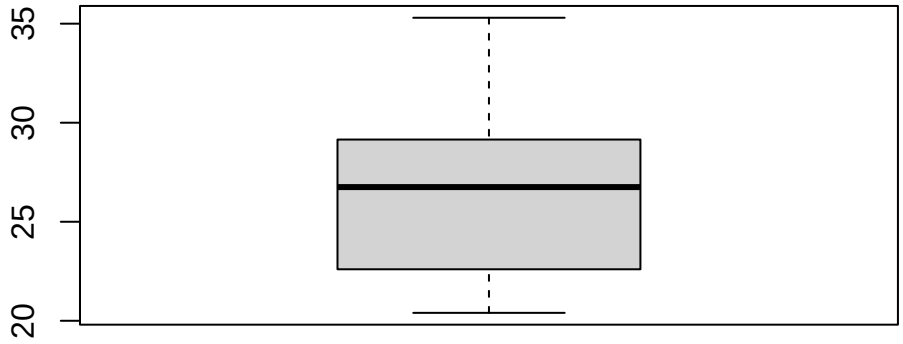
```
'data.frame':   25 obs. of  4 variables:
 $ age: num  1 2 1 3 1 3 1 1 2 2 ...
 $ bmi: num  NA 22.7 NA NA 20.4 NA 22.5 30.1 22 NA ...
 $ hyp: num  NA 1 1 NA 1 NA 1 1 1 NA ...
 $ chl: num  NA 187 187 NA 113 184 118 187 238 NA ...
```

```r
data$age<-as.factor(data$age)
data$hyp<-as.factor(data$hyp)
summary(data)
```

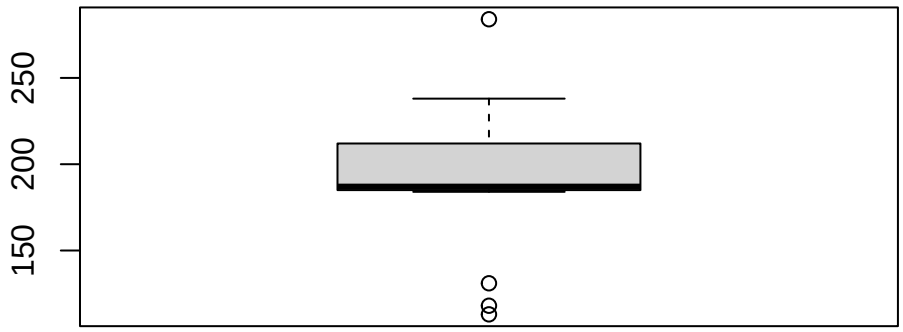
```
 age         bmi           hyp          chl       
 1:12   Min.   :20.40   1   :13   Min.   :113.0  
 2: 7   1st Qu.:22.65   2   : 4   1st Qu.:185.0  
 3: 6   Median :26.75   NA's: 8   Median :187.0  
        Mean   :26.56             Mean   :191.4  
        3rd Qu.:28.93             3rd Qu.:212.0  
        Max.   :35.30             Max.   :284.0  
        NA's   :9                 NA's   :10     
```

2. Inspecciona si las variables numéricas tienen datos atípicos

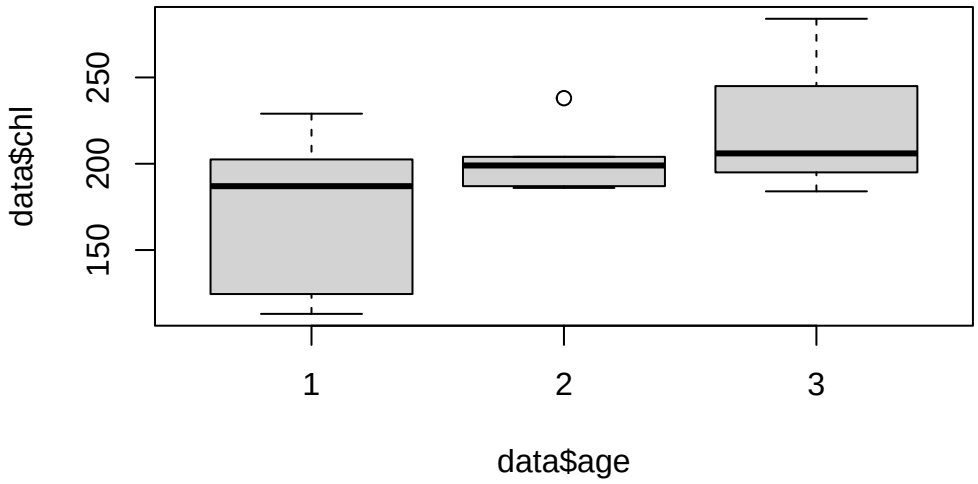
```r
boxplot(data$bmi)
```



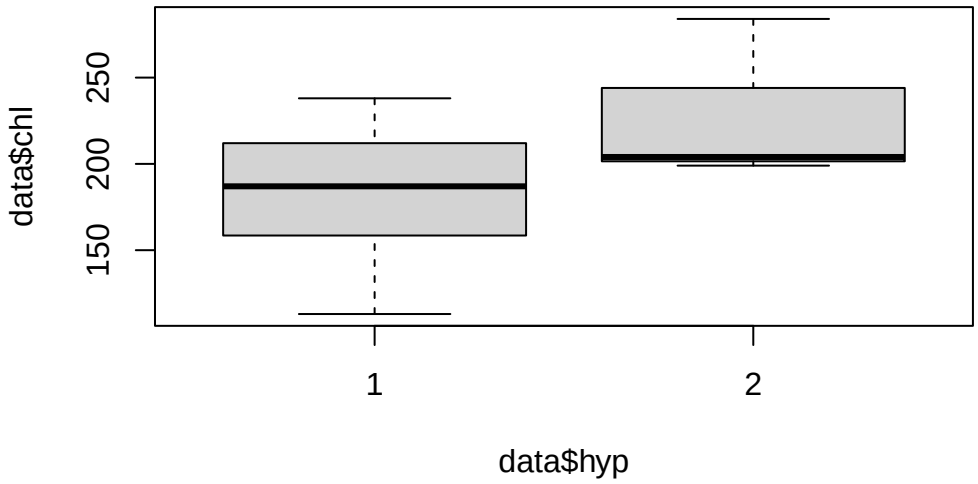

```r
boxplot(data$chl)
```




```
boxplot(data$chl~data$age)
```



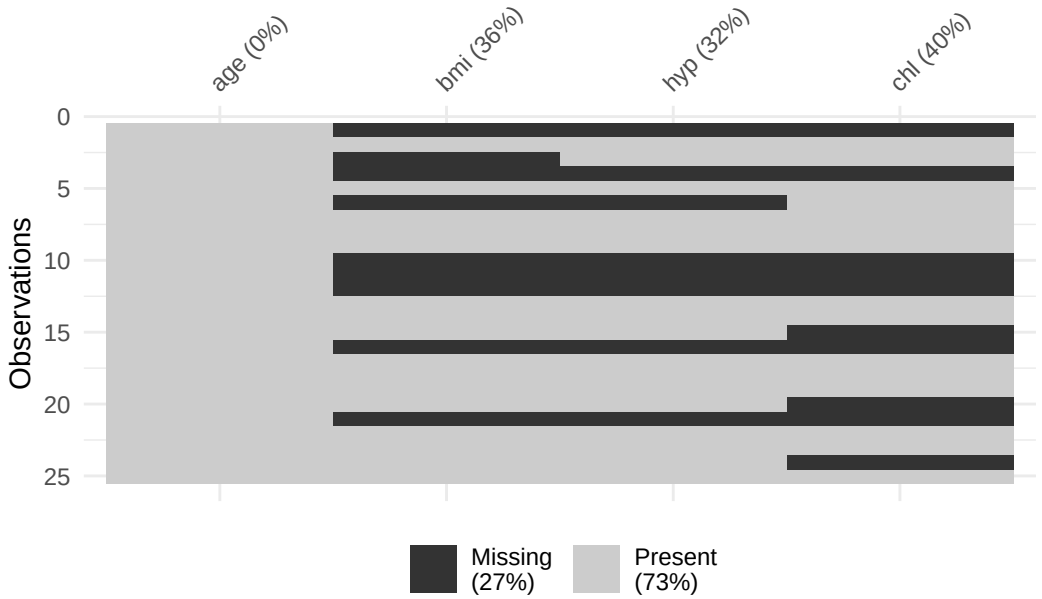
```
boxplot(data$chl~data$hyp)
```



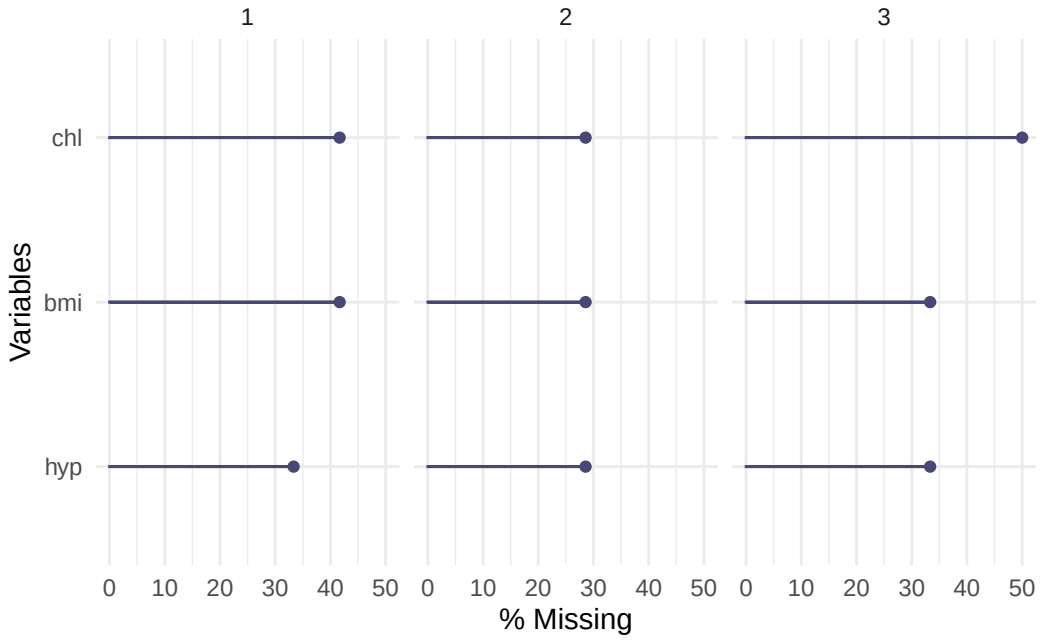
Los datos "outliers" que se observan en la variable chl son parte de la asociación observada con la variable "hyp" por tanto no son outliers y no hay que borrarlos.

3. Visualiza y cuantifica los datos missing. ¿Qué observas?

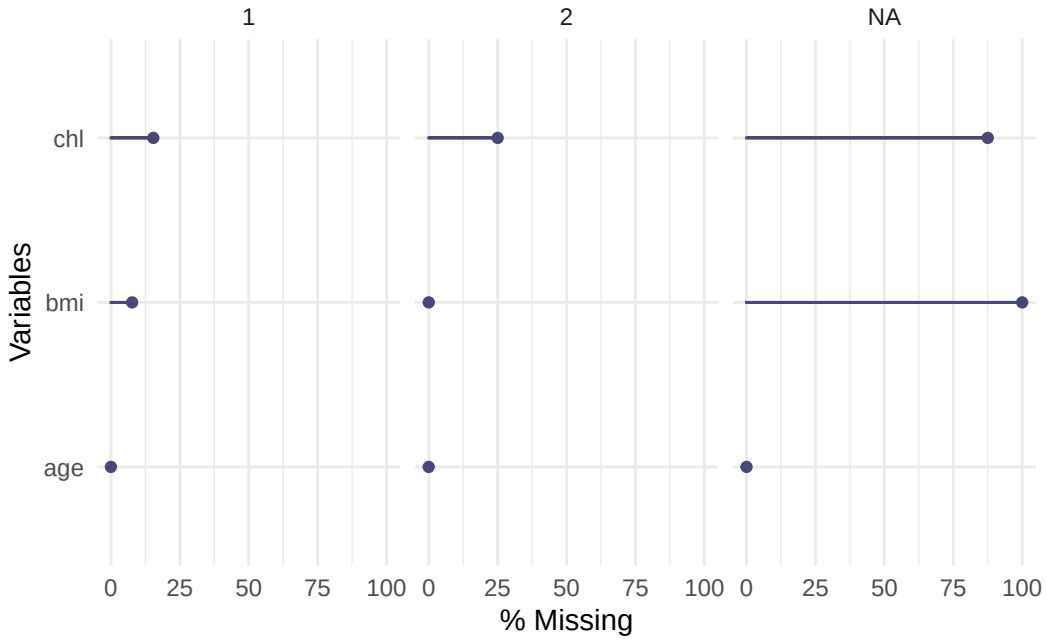
```
library(naniar)
vis_miss(data)
```




```
##Variables cualitativas
gg_miss_var(data, show_pct = TRUE,facet = age)
```



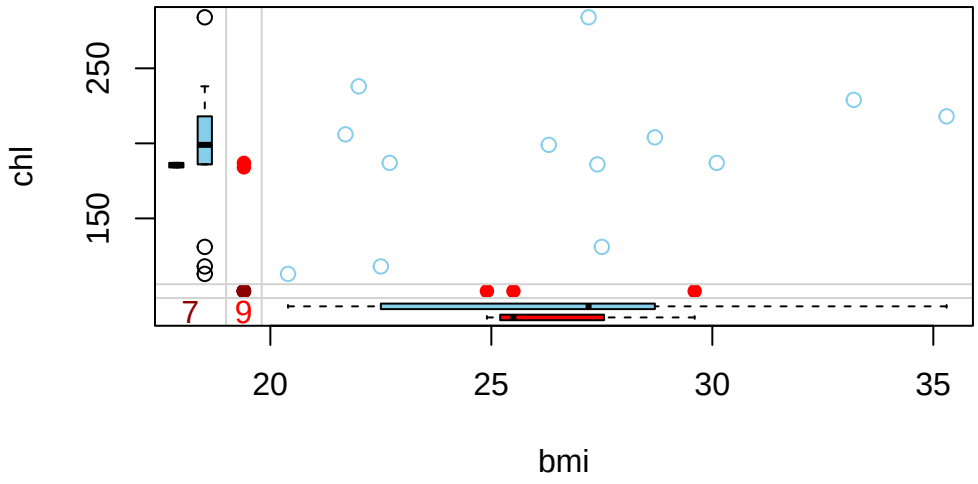
```
gg_miss_var(data, show_pct = TRUE,facet = hyp)
```




```
library(VIM)
```

```
Loading required package: colorspace

Loading required package: grid

VIM is ready to use.

Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues


Attaching package: 'VIM'

The following object is masked from 'package:datasets':

    sleep
```

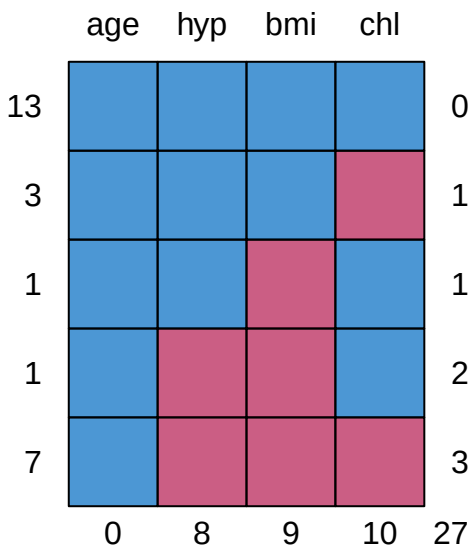
```
##Variables cuantitativas
marginplot(data[,c(2,4)])
```



No vemos ninguna asociación ente los datos missing y el resto de variables o patrones raros, así que asumimos que son de tipo MCAR.

4. Inspecciona el patrón de datos missing. ¿Qué observas? ¿Cuántas observaciones hay con todos las variables missing? ¿Y con 3 de las 4 variables?

```
library(mice)

md.pattern(data)
```




```
   age hyp bmi chl
13   1   1   1   1  0
3    1   1   1   0  1
1    1   1   0   1  1
1    1   0   0   1  2
7    1   0   0   0  3
     0   8   9  10 27
```

No hay ninguna observación con todas las variables missing, pero si hay 7 observaciones que tienen 3 variables missing.

5. Haz una imputación múltiple para rellenar los datos missing, después contesta a las siguientes preguntas:

1. ¿Qué métodos has usado para cada variable?
2. Guarda los datos completos de alguna de las bases de datos imputadas
3. Visualiza las imputaciones. ¿Cómo visualizarías la variable cualitativa?

```
impData <- mice(data,m=5,maxit=50,seed=500)
```

```
iter imp variable
 1   1  bmi  hyp  chl
 1   2  bmi  hyp  chl
 1   3  bmi  hyp  chl
 1   4  bmi  hyp  chl
 1   5  bmi  hyp  chl
 2   1  bmi  hyp  chl
 2   2  bmi  hyp  chl
 2   3  bmi  hyp  chl
 2   4  bmi  hyp  chl
 2   5  bmi  hyp  chl
 3   1  bmi  hyp  chl
 3   2  bmi  hyp  chl
 3   3  bmi  hyp  chl
 3   4  bmi  hyp  chl
 3   5  bmi  hyp  chl
 4   1  bmi  hyp  chl
 4   2  bmi  hyp  chl
 4   3  bmi  hyp  chl
 4   4  bmi  hyp  chl
 4   5  bmi  hyp  chl
 5   1  bmi  hyp  chl
 5   2  bmi  hyp  chl
 5   3  bmi  hyp  chl
 5   4  bmi  hyp  chl
 5   5  bmi  hyp  chl
 6   1  bmi  hyp  chl
 6   2  bmi  hyp  chl
 6   3  bmi  hyp  chl
 6   4  bmi  hyp  chl
 6   5  bmi  hyp  chl
 7   1  bmi  hyp  chl
 7   2  bmi  hyp  chl
 7   3  bmi  hyp  chl
 7   4  bmi  hyp  chl
 7   5  bmi  hyp  chl
 8   1  bmi  hyp  chl
 8   2  bmi  hyp  chl
 8   3  bmi  hyp  chl
 8   4  bmi  hyp  chl
 8   5  bmi  hyp  chl
 9   1  bmi  hyp  chl
```

```
9   2  bmi  hyp  chl
9   3  bmi  hyp  chl
9   4  bmi  hyp  chl
9   5  bmi  hyp  chl
10   1  bmi  hyp  chl
10   2  bmi  hyp  chl
10   3  bmi  hyp  chl
10   4  bmi  hyp  chl
10   5  bmi  hyp  chl
11   1  bmi  hyp  chl
11   2  bmi  hyp  chl
11   3  bmi  hyp  chl
11   4  bmi  hyp  chl
11   5  bmi  hyp  chl
12   1  bmi  hyp  chl
12   2  bmi  hyp  chl
12   3  bmi  hyp  chl
12   4  bmi  hyp  chl
12   5  bmi  hyp  chl
13   1  bmi  hyp  chl
13   2  bmi  hyp  chl
13   3  bmi  hyp  chl
13   4  bmi  hyp  chl
13   5  bmi  hyp  chl
14   1  bmi  hyp  chl
14   2  bmi  hyp  chl
14   3  bmi  hyp  chl
14   4  bmi  hyp  chl
14   5  bmi  hyp  chl
15   1  bmi  hyp  chl
15   2  bmi  hyp  chl
15   3  bmi  hyp  chl
15   4  bmi  hyp  chl
15   5  bmi  hyp  chl
16   1  bmi  hyp  chl
16   2  bmi  hyp  chl
16   3  bmi  hyp  chl
16   4  bmi  hyp  chl
16   5  bmi  hyp  chl
17   1  bmi  hyp  chl
17   2  bmi  hyp  chl
17   3  bmi  hyp  chl
17   4  bmi  hyp  chl
```

```
17   5  bmi  hyp  chl
18   1  bmi  hyp  chl
18   2  bmi  hyp  chl
18   3  bmi  hyp  chl
18   4  bmi  hyp  chl
18   5  bmi  hyp  chl
19   1  bmi  hyp  chl
19   2  bmi  hyp  chl
19   3  bmi  hyp  chl
19   4  bmi  hyp  chl
19   5  bmi  hyp  chl
20   1  bmi  hyp  chl
20   2  bmi  hyp  chl
20   3  bmi  hyp  chl
20   4  bmi  hyp  chl
20   5  bmi  hyp  chl
21   1  bmi  hyp  chl
21   2  bmi  hyp  chl
21   3  bmi  hyp  chl
21   4  bmi  hyp  chl
21   5  bmi  hyp  chl
22   1  bmi  hyp  chl
22   2  bmi  hyp  chl
22   3  bmi  hyp  chl
22   4  bmi  hyp  chl
22   5  bmi  hyp  chl
23   1  bmi  hyp  chl
23   2  bmi  hyp  chl
23   3  bmi  hyp  chl
23   4  bmi  hyp  chl
23   5  bmi  hyp  chl
24   1  bmi  hyp  chl
24   2  bmi  hyp  chl
24   3  bmi  hyp  chl
24   4  bmi  hyp  chl
24   5  bmi  hyp  chl
25   1  bmi  hyp  chl
25   2  bmi  hyp  chl
25   3  bmi  hyp  chl
25   4  bmi  hyp  chl
25   5  bmi  hyp  chl
26   1  bmi  hyp  chl
26   2  bmi  hyp  chl
```

```
26   3  bmi  hyp  chl
26   4  bmi  hyp  chl
26   5  bmi  hyp  chl
27   1  bmi  hyp  chl
27   2  bmi  hyp  chl
27   3  bmi  hyp  chl
27   4  bmi  hyp  chl
27   5  bmi  hyp  chl
28   1  bmi  hyp  chl
28   2  bmi  hyp  chl
28   3  bmi  hyp  chl
28   4  bmi  hyp  chl
28   5  bmi  hyp  chl
29   1  bmi  hyp  chl
29   2  bmi  hyp  chl
29   3  bmi  hyp  chl
29   4  bmi  hyp  chl
29   5  bmi  hyp  chl
30   1  bmi  hyp  chl
30   2  bmi  hyp  chl
30   3  bmi  hyp  chl
30   4  bmi  hyp  chl
30   5  bmi  hyp  chl
31   1  bmi  hyp  chl
31   2  bmi  hyp  chl
31   3  bmi  hyp  chl
31   4  bmi  hyp  chl
31   5  bmi  hyp  chl
32   1  bmi  hyp  chl
32   2  bmi  hyp  chl
32   3  bmi  hyp  chl
32   4  bmi  hyp  chl
32   5  bmi  hyp  chl
33   1  bmi  hyp  chl
33   2  bmi  hyp  chl
33   3  bmi  hyp  chl
33   4  bmi  hyp  chl
33   5  bmi  hyp  chl
34   1  bmi  hyp  chl
34   2  bmi  hyp  chl
34   3  bmi  hyp  chl
34   4  bmi  hyp  chl
34   5  bmi  hyp  chl
```

```
35   1  bmi  hyp  chl
35   2  bmi  hyp  chl
35   3  bmi  hyp  chl
35   4  bmi  hyp  chl
35   5  bmi  hyp  chl
36   1  bmi  hyp  chl
36   2  bmi  hyp  chl
36   3  bmi  hyp  chl
36   4  bmi  hyp  chl
36   5  bmi  hyp  chl
37   1  bmi  hyp  chl
37   2  bmi  hyp  chl
37   3  bmi  hyp  chl
37   4  bmi  hyp  chl
37   5  bmi  hyp  chl
38   1  bmi  hyp  chl
38   2  bmi  hyp  chl
38   3  bmi  hyp  chl
38   4  bmi  hyp  chl
38   5  bmi  hyp  chl
39   1  bmi  hyp  chl
39   2  bmi  hyp  chl
39   3  bmi  hyp  chl
39   4  bmi  hyp  chl
39   5  bmi  hyp  chl
40   1  bmi  hyp  chl
40   2  bmi  hyp  chl
40   3  bmi  hyp  chl
40   4  bmi  hyp  chl
40   5  bmi  hyp  chl
41   1  bmi  hyp  chl
41   2  bmi  hyp  chl
41   3  bmi  hyp  chl
41   4  bmi  hyp  chl
41   5  bmi  hyp  chl
42   1  bmi  hyp  chl
42   2  bmi  hyp  chl
42   3  bmi  hyp  chl
42   4  bmi  hyp  chl
42   5  bmi  hyp  chl
43   1  bmi  hyp  chl
43   2  bmi  hyp  chl
43   3  bmi  hyp  chl
```

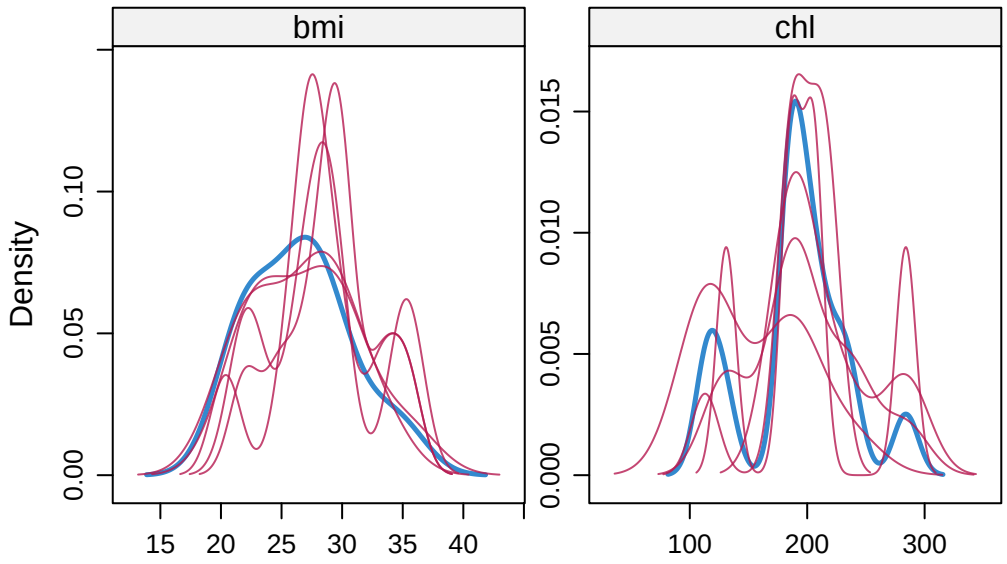
```
 43   4  bmi  hyp  chl
 43   5  bmi  hyp  chl
 44   1  bmi  hyp  chl
 44   2  bmi  hyp  chl
 44   3  bmi  hyp  chl
 44   4  bmi  hyp  chl
 44   5  bmi  hyp  chl
 45   1  bmi  hyp  chl
 45   2  bmi  hyp  chl
 45   3  bmi  hyp  chl
 45   4  bmi  hyp  chl
 45   5  bmi  hyp  chl
 46   1  bmi  hyp  chl
 46   2  bmi  hyp  chl
 46   3  bmi  hyp  chl
 46   4  bmi  hyp  chl
 46   5  bmi  hyp  chl
 47   1  bmi  hyp  chl
 47   2  bmi  hyp  chl
 47   3  bmi  hyp  chl
 47   4  bmi  hyp  chl
 47   5  bmi  hyp  chl
 48   1  bmi  hyp  chl
 48   2  bmi  hyp  chl
 48   3  bmi  hyp  chl
 48   4  bmi  hyp  chl
 48   5  bmi  hyp  chl
 49   1  bmi  hyp  chl
 49   2  bmi  hyp  chl
 49   3  bmi  hyp  chl
 49   4  bmi  hyp  chl
 49   5  bmi  hyp  chl
 50   1  bmi  hyp  chl
 50   2  bmi  hyp  chl
 50   3  bmi  hyp  chl
 50   4  bmi  hyp  chl
 50   5  bmi  hyp  chl
```

```
##Métodos
summary(impData)
```

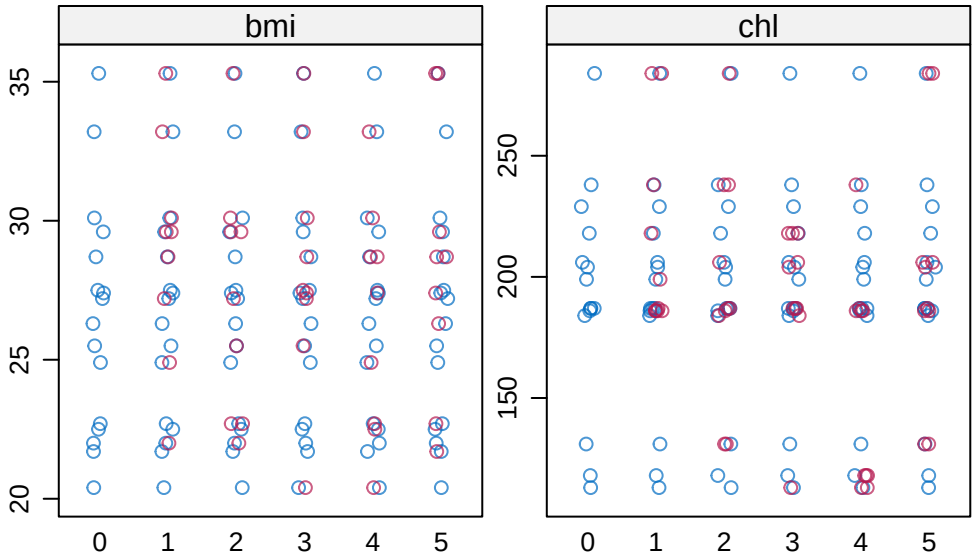
```
Class: mids
Number of multiple imputations:  5
```

```
Imputation methods:
     age       bmi       hyp       chl
      ""    "pmm" "logreg"    "pmm"
PredictorMatrix:
    age bmi hyp chl
age   0   1   1   1
bmi   1   0   1   1
hyp   1   1   0   1
chl   1   1   1   0
```

```r
##Complete Data
completeData <- complete(impData, 1)



##Visualzacion para las variables cuantitativas

densityplot(impData)
```




```r
stripplot(impData)
```



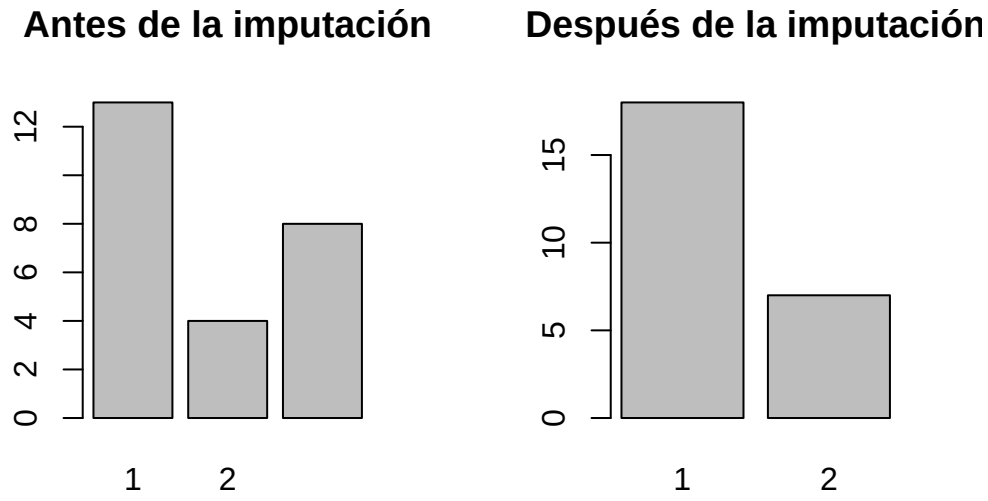

```
# Cuantificar la distribución de la variable categórica antes y después de la imputación
table(data$hyp, useNA = "ifany")
```

```
   1    2 <NA>
  13    4    8
```

```
table(completeData$hyp)
```

```
 1  2
18  7
```

```
# Graficar la distribución de la variable categórica antes y después de la imputación
par(mfrow = c(1, 2))  # Organizar las gráficas en una fila de 2 columnas
barplot(table(data$hyp, useNA = "ifany"), main = "Antes de la imputación")
barplot(table(completeData$hyp), main = "Después de la imputación")
```



6. Ajusta un modelo de regresión lineal para predecir el colesterol en función de las demás variables y calcula el R^2. Comenta los resultados.

```
modelFit1 <- with(impData,lm(chl~age+bmi+hyp))
modelFit1
```

```
call :
with.mids(data = impData, expr = lm(chl ~ age + bmi + hyp))

call1 :
mice(data = data, m = 5, maxit = 50, seed = 500)

nmis :
age bmi hyp chl
  0   9   8  10

analyses :
[[1]]

Call:
lm(formula = chl ~ age + bmi + hyp)
```

```
Coefficients:
(Intercept)         age2         age3          bmi         hyp2
     43.316       36.705       77.535        4.804      -10.568


[[2]]

Call:
lm(formula = chl ~ age + bmi + hyp)

Coefficients:
(Intercept)         age2         age3          bmi         hyp2
     -8.623       66.924       82.107        6.246      -17.090


[[3]]

Call:
lm(formula = chl ~ age + bmi + hyp)

Coefficients:
(Intercept)         age2         age3          bmi         hyp2
    -12.393       68.633       89.189        6.334      -27.689


[[4]]

Call:
lm(formula = chl ~ age + bmi + hyp)

Coefficients:
(Intercept)         age2         age3          bmi         hyp2
       8.00        58.51        85.22         5.13       -12.99


[[5]]

Call:
lm(formula = chl ~ age + bmi + hyp)

Coefficients:
(Intercept)         age2         age3          bmi         hyp2
     -6.572       60.617       83.189        5.876       14.650
```

```r
pool(modelFit1)
```

```
Class: mipo    m = 5
         term m   estimate        ubar           b           t dfcom        df
1 (Intercept) 5   4.745482 3385.620541 524.7105583 4015.273211    20 14.065859
2        age2 5  58.277340  377.182973 163.2226748  573.050182    20  8.896354
3        age3 5  83.447143  473.304451  18.2304435  495.180983    20 17.306730
4         bmi 5   5.677918    4.007194   0.4638056    4.563761    20 15.131786
5        hyp2 5 -10.737873  352.434596 244.3389422  645.641327    20  6.584147
         riv     lambda       fmi
1 0.18597851 0.15681440 0.2556299
2 0.51928964 0.34179766 0.4524538
3 0.04622085 0.04417886 0.1383172
4 0.13889186 0.12195351 0.2188051
5 0.83194651 0.45413253 0.5680430
```

```r
summary(pool(modelFit1))
```

```
         term   estimate std.error   statistic        df     p.value
1 (Intercept)   4.745482 63.366183  0.07488982 14.065859 0.941357058
2        age2  58.277340 23.938467  2.43446420  8.896354 0.038002238
3        age3  83.447143 22.252662  3.74998470 17.306730 0.001553129
4         bmi   5.677918  2.136296  2.65783279 15.131786 0.017808304
5        hyp2 -10.737873 25.409473 -0.42259329  6.584147 0.686048578
```

```r
# Calcular el pseudo R cuadrado

pool.r.squared(pool(modelFit1), adjusted = FALSE)
```

```
          est     lo 95    hi 95       fmi
R^2 0.5315259 0.1577559 0.796209 0.3513097
```