Etapa 1. El contexto

Blanca Ruiz

2025-09-24

# Exploración de la base de datos

library(here)

here() starts at /home/raul/Dropbox/Mathematics/Teaching/Tec/Learning/R Projects/1001B

M19 <- read.csv(here("data", "Datos\_molec\_2019.csv"))  
M20 <- read.csv(here("data", "Datos\_molec\_2020.csv"))  
str(M19)

'data.frame': 1980 obs. of 108 variables:  
 $ cd : int 14 14 14 14 14 14 14 14 14 14 ...  
 $ periodo : int 219 219 219 219 219 219 219 219 219 219 ...  
 $ folio : chr "11B167" "11B167" "11B167" "11B167" ...  
 $ entidad : int 1 1 1 1 1 1 1 1 1 1 ...  
 $ control : int 40006 40006 40006 40006 40131 40131 40131 40131 40132 40132 ...  
 $ viv\_sel : int 3 2 1 4 3 4 2 1 1 2 ...  
 $ num\_hog : int 1 1 1 1 1 1 1 1 1 1 ...  
 $ hog\_mud : int 0 0 0 0 0 0 0 0 0 0 ...  
 $ n\_ren\_ele : int 2 1 1 2 1 1 1 1 1 2 ...  
 $ p1 : int 1 1 1 1 1 1 1 1 1 1 ...  
 $ p2 : int 2 2 2 2 2 1 1 1 2 1 ...  
 $ p3\_1 : int 2 2 2 2 2 2 1 2 2 2 ...  
 $ p3\_2 : int 2 2 2 2 2 1 2 2 2 2 ...  
 $ p3\_3 : int 2 2 1 2 2 1 2 1 2 1 ...  
 $ p3\_4 : int 2 2 2 2 2 2 2 2 2 2 ...  
 $ p3\_5 : int 2 2 2 2 2 1 1 2 2 2 ...  
 $ p4 : int 0 0 0 0 0 0 1 0 0 0 ...  
 $ p5 : int 0 0 0 0 0 0 4 0 0 0 ...  
 $ p5\_6esp : chr " " " " " " " " ...  
 $ p6\_1 : int 0 0 0 0 0 0 2 0 0 0 ...  
 $ p6\_2 : int 0 0 0 0 0 0 2 0 0 0 ...  
 $ p6\_3 : int 0 0 0 0 0 0 1 0 0 0 ...  
 $ p6\_4 : int 0 0 0 0 0 0 2 0 0 0 ...  
 $ p6\_5 : int 0 0 0 0 0 0 2 0 0 0 ...  
 $ p6\_6 : int 0 0 0 0 0 0 2 0 0 0 ...  
 $ p6\_6esp : chr " " " " " " " " ...  
 $ p7 : int 0 0 0 0 0 0 2 0 0 0 ...  
 $ p7\_3 : int 0 0 0 0 0 0 0 0 0 0 ...  
 $ p8\_1 : int 0 0 0 0 0 0 2 0 0 0 ...  
 $ p8\_2 : int 0 0 0 0 0 0 1 0 0 0 ...  
 $ p9 : int 0 0 0 0 0 0 1 0 0 0 ...  
 $ p9\_5esp : chr " " " " " " " " ...  
 $ p10 : int 0 0 0 0 0 2 0 0 0 0 ...  
 $ p11 : int 0 0 0 0 0 4 0 0 0 0 ...  
 $ p11\_6esp : chr " " " " " " " " ...  
 $ p12\_1 : int 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_2 : int 0 0 0 0 0 1 0 0 0 0 ...  
 $ p12\_3 : int 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_4 : int 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_5 : int 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_6 : int 0 0 0 0 0 1 0 0 0 0 ...  
 $ p12\_7 : int 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_8 : int 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_9 : int 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_9esp : chr " " " " " " " " ...  
 $ p13 : int 0 0 0 0 0 2 0 0 0 0 ...  
 $ p13\_3 : int 0 0 0 0 0 0 0 0 0 0 ...  
 $ p14\_1 : int 0 0 0 0 0 2 0 0 0 0 ...  
 $ p14\_2 : int 0 0 0 0 0 1 0 0 0 0 ...  
 $ p15 : int 0 0 0 0 0 1 0 0 0 0 ...  
 $ p15\_5esp : chr " " " " " " " " ...  
 $ p16 : int 0 0 1 0 0 1 0 3 0 2 ...  
 $ p17 : int 0 0 4 0 0 4 0 3 0 3 ...  
 $ p17\_6esp : chr " " " " " " " " ...  
 $ p18\_1 : int 0 0 1 0 0 1 0 2 0 2 ...  
 $ p18\_2 : int 0 0 1 0 0 1 0 1 0 1 ...  
 $ p18\_3 : int 0 0 1 0 0 2 0 2 0 2 ...  
 $ p18\_4 : int 0 0 1 0 0 2 0 2 0 2 ...  
 $ p18\_5 : int 0 0 1 0 0 1 0 1 0 1 ...  
 $ p19 : int 0 0 2 0 0 2 0 2 0 2 ...  
 $ p19\_3 : int 0 0 0 0 0 0 0 0 0 0 ...  
 $ p20\_1 : int 0 0 1 0 0 2 0 2 0 2 ...  
 $ p20\_2 : int 0 0 2 0 0 1 0 1 0 1 ...  
 $ p21 : int 0 0 1 0 0 2 0 2 0 2 ...  
 $ p21\_5esp : chr " " " " " " " " ...  
 $ p22 : int 0 0 0 0 0 0 0 0 0 0 ...  
 $ p23\_1 : int 0 0 0 0 0 0 0 0 0 0 ...  
 $ p23\_2 : int 0 0 0 0 0 0 0 0 0 0 ...  
 $ p24 : int 0 0 0 0 0 5 2 0 0 0 ...  
 $ p25 : int 0 0 0 0 0 3 3 0 0 0 ...  
 $ p25\_6esp : chr " " " " " " " " ...  
 $ p26 : int 0 0 15 0 0 10 60 15 0 20 ...  
 $ p27 : int 0 0 2 0 0 2 2 2 0 2 ...  
 $ p28 : int 0 0 0 0 0 0 0 0 0 0 ...  
 $ p28\_7esp : chr " " " " " " " " ...  
 $ p29 : int 0 0 2 0 0 4 3 2 0 3 ...  
 $ p30 : int 0 0 2 0 0 4 3 3 0 3 ...  
 $ p31 : int 0 0 2 0 0 2 1 2 0 2 ...  
 $ p32 : int 3 5 0 5 2 0 0 0 1 0 ...  
 $ p32\_6esp : chr " " " " " " " " ...  
 $ p33\_1 : int 2 2 2 2 2 2 2 2 2 2 ...  
 $ p33\_2 : int 2 2 2 2 2 2 2 2 2 2 ...  
 $ p33\_3 : int 2 2 2 2 2 2 2 2 2 2 ...  
 $ p33\_4 : int 2 2 2 2 2 2 2 2 2 2 ...  
 $ p34\_1 : int 2 2 2 2 3 1 1 3 3 3 ...  
 $ p34\_2 : int 2 1 1 1 3 1 1 3 3 1 ...  
 $ p34\_3 : int 2 2 2 1 3 1 1 3 3 1 ...  
 $ p34\_3\_1 : int 0 0 0 1 0 2 2 0 0 1 ...  
 $ p34\_4 : int 2 1 1 1 1 1 1 1 3 1 ...  
 $ p34\_4\_1 : int 0 1 1 1 1 1 1 1 0 1 ...  
 $ p35 : int 1 1 1 1 1 1 1 1 1 1 ...  
 $ p36\_1 : int 2 1 1 1 3 1 1 3 3 1 ...  
 $ p36\_2 : int 2 1 2 1 3 1 1 3 3 1 ...  
 $ p36\_3 : int 2 1 2 1 3 1 1 3 3 1 ...  
 $ p36\_4 : int 2 2 2 2 3 1 1 3 3 1 ...  
 $ factor : int 18795 15188 22781 14041 16030 21866 33063 14578 24608 15907 ...  
 $ h\_lec : int 4 4 3 4 4 1 1 1 4 1 ...  
 $ mat\_lec : int 4 4 3 4 4 3 2 3 4 3 ...  
 $ perslec : int 2 2 1 2 2 1 1 1 2 1 ...  
 [list output truncated]

1980 observaciones con 108 variables. La mayoría de las variables son cualitativas, pero están insertas como números. Se tiene que revisar en el contexto del problema cuáles son realmente numéricas.

## Diccionarios

Dicc19 <- read.csv(here("data", "Diccionario\_molec\_2019.csv"))

El diccionario indica qué significan las variables en el contexto del problema, pero no a qué equivalen sus categorías. Eso se debe leer en el archivo de descriptor de las preguntas y en el cuestionario.

## Resumen de las variables

summary(M19)

cd periodo folio entidad   
 Min. : 1.00 Min. :219 Length:1980 Min. : 1.00   
 1st Qu.: 2.00 1st Qu.:219 Class :character 1st Qu.: 9.00   
 Median : 9.00 Median :219 Mode :character Median :15.00   
 Mean :13.96 Mean :219 Mean :15.74   
 3rd Qu.:25.00 3rd Qu.:219 3rd Qu.:21.00   
 Max. :43.00 Max. :219 Max. :32.00   
   
 control viv\_sel num\_hog hog\_mud n\_ren\_ele   
 Min. :40001 Min. :1.000 Min. :1 Min. :0.00000 Min. :1.000   
 1st Qu.:40138 1st Qu.:2.000 1st Qu.:1 1st Qu.:0.00000 1st Qu.:1.000   
 Median :40222 Median :3.000 Median :1 Median :0.00000 Median :1.000   
 Mean :40291 Mean :2.502 Mean :1 Mean :0.04697 Mean :1.732   
 3rd Qu.:40390 3rd Qu.:3.000 3rd Qu.:1 3rd Qu.:0.00000 3rd Qu.:2.000   
 Max. :41420 Max. :4.000 Max. :1 Max. :2.00000 Max. :7.000   
   
 p1 p2 p3\_1 p3\_2 p3\_3   
 Min. :1.00 Min. :0.000 Min. :0.00 Min. :0.000 Min. :0.0   
 1st Qu.:1.00 1st Qu.:1.000 1st Qu.:1.00 1st Qu.:1.000 1st Qu.:1.0   
 Median :1.00 Median :1.000 Median :2.00 Median :2.000 Median :2.0   
 Mean :1.03 Mean :1.351 Mean :1.53 Mean :1.612 Mean :1.6   
 3rd Qu.:1.00 3rd Qu.:2.000 3rd Qu.:2.00 3rd Qu.:2.000 3rd Qu.:2.0   
 Max. :2.00 Max. :2.000 Max. :2.00 Max. :2.000 Max. :2.0   
   
 p3\_4 p3\_5 p4 p5   
 Min. :0.000 Min. :0.000 Min. : 0.000 Min. :0.00   
 1st Qu.:2.000 1st Qu.:1.000 1st Qu.: 0.000 1st Qu.:0.00   
 Median :2.000 Median :2.000 Median : 0.000 Median :0.00   
 Mean :1.897 Mean :1.582 Mean : 1.335 Mean :1.37   
 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.: 2.000 3rd Qu.:3.00   
 Max. :2.000 Max. :2.000 Max. :50.000 Max. :6.00   
   
 p5\_6esp p6\_1 p6\_2 p6\_3   
 Length:1980 Min. :0.0000 Min. :0.0000 Min. :0.0000   
 Class :character 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
 Mode :character Median :0.0000 Median :0.0000 Median :0.0000   
 Mean :0.7848 Mean :0.6889 Mean :0.6985   
 3rd Qu.:2.0000 3rd Qu.:2.0000 3rd Qu.:2.0000   
 Max. :2.0000 Max. :2.0000 Max. :2.0000   
   
 p6\_4 p6\_5 p6\_6 p6\_6esp   
 Min. :0.0000 Min. :0.0000 Min. :0.0000 Length:1980   
 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 Class :character   
 Median :0.0000 Median :0.0000 Median :0.0000 Mode :character   
 Mean :0.6465 Mean :0.7131 Mean :0.8177   
 3rd Qu.:1.0000 3rd Qu.:2.0000 3rd Qu.:2.0000   
 Max. :2.0000 Max. :2.0000 Max. :2.0000   
   
 p7 p7\_3 p8\_1 p8\_2   
 Min. :0.0000 Min. : 0.0 Min. :0.0000 Min. :0.0000   
 1st Qu.:0.0000 1st Qu.: 0.0 1st Qu.:0.0000 1st Qu.:0.0000   
 Median :0.0000 Median : 0.0 Median :0.0000 Median :0.0000   
 Mean :0.9364 Mean : 195.5 Mean :0.7545 Mean :0.4561   
 3rd Qu.:2.0000 3rd Qu.: 0.0 3rd Qu.:2.0000 3rd Qu.:1.0000   
 Max. :3.0000 Max. :99999.0 Max. :2.0000 Max. :2.0000   
   
 p9 p9\_5esp p10 p11   
 Min. :0.0000 Length:1980 Min. : 0.000 Min. :0.000   
 1st Qu.:0.0000 Class :character 1st Qu.: 0.000 1st Qu.:0.000   
 Median :0.0000 Mode :character Median : 0.000 Median :0.000   
 Mean :0.4848 Mean : 1.363 Mean :1.156   
 3rd Qu.:1.0000 3rd Qu.: 2.000 3rd Qu.:3.000   
 Max. :5.0000 Max. :99.000 Max. :6.000   
   
 p11\_6esp p12\_1 p12\_2 p12\_3   
 Length:1980 Min. :0.0000 Min. :0.0000 Min. :0.0000   
 Class :character 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
 Mode :character Median :0.0000 Median :0.0000 Median :0.0000   
 Mean :0.6197 Mean :0.5955 Mean :0.5949   
 3rd Qu.:2.0000 3rd Qu.:2.0000 3rd Qu.:2.0000   
 Max. :2.0000 Max. :2.0000 Max. :2.0000   
   
 p12\_4 p12\_5 p12\_6 p12\_7   
 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
 Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000   
 Mean :0.5742 Mean :0.6172 Mean :0.5818 Mean :0.5773   
 3rd Qu.:1.0000 3rd Qu.:2.0000 3rd Qu.:2.0000 3rd Qu.:1.2500   
 Max. :2.0000 Max. :2.0000 Max. :2.0000 Max. :2.0000   
   
 p12\_8 p12\_9 p12\_9esp p13   
 Min. :0.000 Min. :0.000 Length:1980 Min. :0.000   
 1st Qu.:0.000 1st Qu.:0.000 Class :character 1st Qu.:0.000   
 Median :0.000 Median :0.000 Mode :character Median :0.000   
 Mean :0.554 Mean :0.651 Mean :0.752   
 3rd Qu.:1.000 3rd Qu.:2.000 3rd Qu.:2.000   
 Max. :2.000 Max. :2.000 Max. :3.000   
   
 p13\_3 p14\_1 p14\_2 p15   
 Min. : 0.00 Min. :0.0000 Min. :0.0000 Min. :0.0000   
 1st Qu.: 0.00 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
 Median : 0.00 Median :0.0000 Median :0.0000 Median :0.0000   
 Mean : 18.22 Mean :0.6197 Mean :0.3551 Mean :0.4096   
 3rd Qu.: 0.00 3rd Qu.:2.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
 Max. :1000.00 Max. :2.0000 Max. :2.0000 Max. :5.0000   
   
 p15\_5esp p16 p17 p17\_6esp   
 Length:1980 Min. : 0.000 Min. :0.00 Length:1980   
 Class :character 1st Qu.: 0.000 1st Qu.:0.00 Class :character   
 Mode :character Median : 0.000 Median :0.00 Mode :character   
 Mean : 1.092 Mean :1.11   
 3rd Qu.: 1.000 3rd Qu.:3.00   
 Max. :45.000 Max. :6.00   
   
 p18\_1 p18\_2 p18\_3 p18\_4   
 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
 Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000   
 Mean :0.4268 Mean :0.4692 Mean :0.4611 Mean :0.4975   
 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
 Max. :2.0000 Max. :2.0000 Max. :2.0000 Max. :2.0000   
   
 p18\_5 p19 p19\_3 p20\_1   
 Min. :0.000 Min. :0.0000 Min. : 0.0 Min. :0.0000   
 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.: 0.0 1st Qu.:0.0000   
 Median :0.000 Median :0.0000 Median : 0.0 Median :0.0000   
 Mean :0.403 Mean :0.8177 Mean : 512.7 Mean :0.6475   
 3rd Qu.:1.000 3rd Qu.:2.0000 3rd Qu.: 0.0 3rd Qu.:2.0000   
 Max. :2.000 Max. :3.0000 Max. :999999.0 Max. :2.0000   
   
 p20\_2 p21 p21\_5esp p22   
 Min. :0.0000 Min. :0.0000 Length:1980 Min. :0.0000   
 1st Qu.:0.0000 1st Qu.:0.0000 Class :character 1st Qu.:0.0000   
 Median :0.0000 Median :0.0000 Mode :character Median :0.0000   
 Mean :0.3672 Mean :0.4879 Mean :0.1374   
 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000   
 Max. :2.0000 Max. :5.0000 Max. :5.0000   
   
 p23\_1 p23\_2 p24 p25   
 Min. :0.00000 Min. :0.00000 Min. :0.0000 Min. :0.000   
 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.000   
 Median :0.00000 Median :0.00000 Median :0.0000 Median :0.000   
 Mean :0.07121 Mean :0.05253 Mean :0.7631 Mean :1.066   
 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:3.000   
 Max. :2.00000 Max. :2.00000 Max. :5.0000 Max. :6.000   
   
 p25\_6esp p26 p27 p28   
 Length:1980 Min. : 0.00 Min. :0.000 Min. :0.0000   
 Class :character 1st Qu.: 0.00 1st Qu.:0.000 1st Qu.:0.0000   
 Mode :character Median : 20.00 Median :2.000 Median :0.0000   
 Mean : 27.54 Mean :1.322 Mean :0.5808   
 3rd Qu.: 40.00 3rd Qu.:2.000 3rd Qu.:0.0000   
 Max. :360.00 Max. :2.000 Max. :7.0000   
   
 p28\_7esp p29 p30 p31   
 Length:1980 Min. :0.000 Min. :0.000 Min. :0.000   
 Class :character 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000   
 Mode :character Median :2.000 Median :3.000 Median :1.000   
 Mean :1.748 Mean :2.111 Mean :1.134   
 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:2.000   
 Max. :4.000 Max. :4.000 Max. :2.000   
   
 p32 p32\_6esp p33\_1 p33\_2   
 Min. :0.0000 Length:1980 Min. :0.000 Min. :0.000   
 1st Qu.:0.0000 Class :character 1st Qu.:1.000 1st Qu.:2.000   
 Median :0.0000 Mode :character Median :2.000 Median :2.000   
 Mean :0.7106 Mean :1.712 Mean :1.801   
 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:2.000   
 Max. :6.0000 Max. :2.000 Max. :2.000   
   
 p33\_3 p33\_4 p34\_1 p34\_2 p34\_3   
 Min. :0.000 Min. :0.000 Min. :0.00 Min. :0.00 Min. :0.000   
 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:1.00 1st Qu.:1.00 1st Qu.:1.000   
 Median :2.000 Median :2.000 Median :2.00 Median :1.00 Median :2.000   
 Mean :1.856 Mean :1.767 Mean :1.71 Mean :1.47 Mean :1.639   
 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.00 3rd Qu.:2.00 3rd Qu.:2.000   
 Max. :2.000 Max. :2.000 Max. :3.00 Max. :3.00 Max. :3.000   
   
 p34\_3\_1 p34\_4 p34\_4\_1 p35   
 Min. :0.0000 Min. :0.000 Min. :0.0000 Min. :0.0000   
 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:1.0000   
 Median :0.0000 Median :1.000 Median :1.0000 Median :1.0000   
 Mean :0.5323 Mean :1.416 Mean :0.9621 Mean :0.9818   
 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:1.0000   
 Max. :3.0000 Max. :3.000 Max. :6.0000 Max. :2.0000   
   
 p36\_1 p36\_2 p36\_3 p36\_4   
 Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000   
 Median :1.000 Median :1.000 Median :1.000 Median :1.000   
 Mean :1.424 Mean :1.207 Mean :1.334 Mean :1.422   
 3rd Qu.:2.000 3rd Qu.:1.000 3rd Qu.:2.000 3rd Qu.:2.000   
 Max. :3.000 Max. :3.000 Max. :3.000 Max. :3.000   
   
 factor h\_lec mat\_lec perslec   
 Min. : 2852 Min. :1.000 Min. :1.000 Min. :1.000   
 1st Qu.: 12880 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000   
 Median : 16764 Median :1.000 Median :3.000 Median :1.000   
 Mean : 19297 Mean :2.044 Mean :2.756 Mean :1.258   
 3rd Qu.: 24724 3rd Qu.:3.000 3rd Qu.:4.000 3rd Qu.:2.000   
 Max. :103721 Max. :4.000 Max. :4.000 Max. :2.000   
 NA's :60 NA's :60 NA's :60   
 l\_formato r\_formato p\_formato persleclrp edad   
 Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.0 Min. :18.00   
 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:1.0 1st Qu.:31.00   
 Median :2.000 Median :2.000 Median :2.000 Median :1.0 Median :44.00   
 Mean :1.926 Mean :1.937 Mean :1.929 Mean :1.3 Mean :45.07   
 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.0 3rd Qu.:57.00   
 Max. :3.000 Max. :3.000 Max. :3.000 Max. :2.0 Max. :95.00   
 NA's :1170 NA's :1332 NA's :1308 NA's :60   
 sexo anio nivel cond\_activ   
 Min. :1.000 Min. :1.000 Min. : 0.000 Min. : 1.000   
 1st Qu.:1.000 1st Qu.:3.000 1st Qu.: 3.000 1st Qu.: 1.000   
 Median :2.000 Median :3.000 Median : 4.000 Median : 1.000   
 Mean :1.559 Mean :3.463 Mean : 4.368 Mean : 3.388   
 3rd Qu.:2.000 3rd Qu.:4.000 3rd Qu.: 7.000 3rd Qu.: 7.000   
 Max. :2.000 Max. :6.000 Max. :99.000 Max. :99.000   
 NA's :59

Aunque en la mayor parte de las variables, las medidas calculadas (media, mediana, mínimo, máximo), en muchas no tienen sentido porque son categóricas. El summary permite ver si la variable tiene datos faltantes (NA´s), por ejemplo, la variable anio tiene valores faltantes.

# Selección de las variables

Por ejemplo, si me interesa analizar las variables 3 a 34 (se revisa en la base de datos el número de columna que tiene la variable de interés):

M19selec <- M19[3:34]  
M20selec <- M20[3:34]

## Datos perdidos o faltantes

Visualizo si las variables de interés tienen datos perdidos:

summary(M20selec)

control viv\_sel num\_hog hog\_mud n\_ren\_el   
 Min. :40025 Min. :1.000 Min. :1 Min. :0.00000 Min. : 1.000   
 1st Qu.:40153 1st Qu.:2.000 1st Qu.:1 1st Qu.:0.00000 1st Qu.: 1.000   
 Median :40248 Median :3.000 Median :1 Median :0.00000 Median : 1.000   
 Mean :40307 Mean :2.517 Mean :1 Mean :0.04726 Mean : 1.734   
 3rd Qu.:40398 3rd Qu.:4.000 3rd Qu.:1 3rd Qu.:0.00000 3rd Qu.: 2.000   
 Max. :41420 Max. :4.000 Max. :1 Max. :2.00000 Max. :10.000   
   
 cd periodo sexo edad anio   
 Min. : 1.00 Min. :220 Min. :1.000 Min. :18.00 Min. :1.00   
 1st Qu.: 2.00 1st Qu.:220 1st Qu.:1.000 1st Qu.:31.00 1st Qu.:3.00   
 Median : 9.00 Median :220 Median :2.000 Median :44.00 Median :3.00   
 Mean :14.01 Mean :220 Mean :1.552 Mean :45.49 Mean :3.47   
 3rd Qu.:25.00 3rd Qu.:220 3rd Qu.:2.000 3rd Qu.:58.00 3rd Qu.:4.00   
 Max. :43.00 Max. :220 Max. :2.000 Max. :94.00 Max. :6.00   
 NA's :65   
 nivel cond\_act p1 p2   
 Min. : 0.000 Min. : 1.000 Min. :1.000 Min. :0.000   
 1st Qu.: 3.000 1st Qu.: 1.000 1st Qu.:1.000 1st Qu.:1.000   
 Median : 4.000 Median : 1.000 Median :1.000 Median :1.000   
 Mean : 4.418 Mean : 3.289 Mean :1.026 Mean :1.367   
 3rd Qu.: 7.000 3rd Qu.: 7.000 3rd Qu.:1.000 3rd Qu.:2.000   
 Max. :99.000 Max. :99.000 Max. :2.000 Max. :2.000   
   
 p3\_1 p3\_2 p3\_3 p3\_4   
 Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:2.000   
 Median :2.000 Median :2.000 Median :2.000 Median :2.000   
 Mean :1.545 Mean :1.638 Mean :1.655 Mean :1.904   
 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000   
 Max. :2.000 Max. :2.000 Max. :2.000 Max. :2.000   
   
 p3\_5 p4 p5 p5\_6esp   
 Min. :0.000 Min. : 0.000 Min. :0.00 Length:2010   
 1st Qu.:1.000 1st Qu.: 0.000 1st Qu.:0.00 Class :character   
 Median :2.000 Median : 0.000 Median :0.00 Mode :character   
 Mean :1.588 Mean : 1.428 Mean :1.33   
 3rd Qu.:2.000 3rd Qu.: 2.000 3rd Qu.:3.00   
 Max. :2.000 Max. :60.000 Max. :6.00   
   
 p6\_1 p6\_2 p6\_3 p6\_4   
 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
 Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000   
 Mean :0.7781 Mean :0.6801 Mean :0.6861 Mean :0.6363   
 3rd Qu.:2.0000 3rd Qu.:2.0000 3rd Qu.:2.0000 3rd Qu.:1.0000   
 Max. :2.0000 Max. :2.0000 Max. :2.0000 Max. :2.0000   
   
 p6\_5 p6\_6 p6\_6esp p7   
 Min. :0.000 Min. :0.000 Length:2010 Min. :0.0000   
 1st Qu.:0.000 1st Qu.:0.000 Class :character 1st Qu.:0.0000   
 Median :0.000 Median :0.000 Mode :character Median :0.0000   
 Mean :0.696 Mean :0.805 Mean :0.9224   
 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.0000   
 Max. :2.000 Max. :2.000 Max. :3.0000   
   
 p7\_3 p8\_1   
 Min. : 0 Min. :0.0000   
 1st Qu.: 0 1st Qu.:0.0000   
 Median : 0 Median :0.0000   
 Mean : 1631 Mean :0.7473   
 3rd Qu.: 0 3rd Qu.:2.0000   
 Max. :999999 Max. :2.0000

La variable anio tiene 39 datos faltantes. Anio está en la columna 10 en la base de datos de las variables de interés.

Para calcular el porcentaje de datos faltantes que tiene y que se los vamos a quitar:

sum(is.na(M20selec[, 10])) / length(M20selec[, 10])

[1] 0.03233831

## Para quitar los NA’s

M20selec <- na.omit(M20selec)  
summary(M20selec)

control viv\_sel num\_hog hog\_mud n\_ren\_el   
 Min. :40025 Min. :1.000 Min. :1 Min. :0.0000 Min. : 1.00   
 1st Qu.:40153 1st Qu.:2.000 1st Qu.:1 1st Qu.:0.0000 1st Qu.: 1.00   
 Median :40248 Median :3.000 Median :1 Median :0.0000 Median : 1.00   
 Mean :40305 Mean :2.523 Mean :1 Mean :0.0473 Mean : 1.74   
 3rd Qu.:40398 3rd Qu.:4.000 3rd Qu.:1 3rd Qu.:0.0000 3rd Qu.: 2.00   
 Max. :41420 Max. :4.000 Max. :1 Max. :2.0000 Max. :10.00   
 cd periodo sexo edad anio   
 Min. : 1.00 Min. :220 Min. :1.000 Min. :18.00 Min. :1.00   
 1st Qu.: 2.00 1st Qu.:220 1st Qu.:1.000 1st Qu.:31.00 1st Qu.:3.00   
 Median : 9.00 Median :220 Median :2.000 Median :44.00 Median :3.00   
 Mean :14.02 Mean :220 Mean :1.549 Mean :44.89 Mean :3.47   
 3rd Qu.:25.00 3rd Qu.:220 3rd Qu.:2.000 3rd Qu.:57.00 3rd Qu.:4.00   
 Max. :43.00 Max. :220 Max. :2.000 Max. :93.00 Max. :6.00   
 nivel cond\_act p1 p2   
 Min. :1.000 Min. : 1.000 Min. :1.000 Min. :0.000   
 1st Qu.:3.000 1st Qu.: 1.000 1st Qu.:1.000 1st Qu.:1.000   
 Median :4.000 Median : 1.000 Median :1.000 Median :1.000   
 Mean :4.413 Mean : 3.245 Mean :1.008 Mean :1.389   
 3rd Qu.:7.000 3rd Qu.: 7.000 3rd Qu.:1.000 3rd Qu.:2.000   
 Max. :9.000 Max. :99.000 Max. :2.000 Max. :2.000   
 p3\_1 p3\_2 p3\_3 p3\_4   
 Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:2.000   
 Median :2.000 Median :2.000 Median :2.000 Median :2.000   
 Mean :1.569 Mean :1.664 Mean :1.683 Mean :1.939   
 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000   
 Max. :2.000 Max. :2.000 Max. :2.000 Max. :2.000   
 p3\_5 p4 p5 p5\_6esp   
 Min. :0.000 Min. : 0.000 Min. :0.000 Length:1945   
 1st Qu.:1.000 1st Qu.: 0.000 1st Qu.:0.000 Class :character   
 Median :2.000 Median : 0.000 Median :0.000 Mode :character   
 Mean :1.614 Mean : 1.475 Mean :1.367   
 3rd Qu.:2.000 3rd Qu.: 2.000 3rd Qu.:3.000   
 Max. :2.000 Max. :60.000 Max. :6.000   
 p6\_1 p6\_2 p6\_3 p6\_4   
 Min. :0.000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
 Median :0.000 Median :0.0000 Median :0.0000 Median :0.0000   
 Mean :0.801 Mean :0.6997 Mean :0.7075 Mean :0.6545   
 3rd Qu.:2.000 3rd Qu.:2.0000 3rd Qu.:2.0000 3rd Qu.:1.0000   
 Max. :2.000 Max. :2.0000 Max. :2.0000 Max. :2.0000   
 p6\_5 p6\_6 p6\_6esp p7   
 Min. :0.0000 Min. :0.0000 Length:1945 Min. :0.0000   
 1st Qu.:0.0000 1st Qu.:0.0000 Class :character 1st Qu.:0.0000   
 Median :0.0000 Median :0.0000 Mode :character Median :0.0000   
 Mean :0.7162 Mean :0.8288 Mean :0.9501   
 3rd Qu.:2.0000 3rd Qu.:2.0000 3rd Qu.:2.0000   
 Max. :2.0000 Max. :2.0000 Max. :3.0000   
 p7\_3 p8\_1   
 Min. : 0 Min. :0.0000   
 1st Qu.: 0 1st Qu.:0.0000   
 Median : 0 Median :0.0000   
 Mean : 1685 Mean :0.7692   
 3rd Qu.: 0 3rd Qu.:2.0000   
 Max. :999999 Max. :2.0000

Se verifica que ya no hay datos perdidos en mi data set de interés.