Etapa 1. El contexto

Blanca Ruiz

2025-09-26

# Exploración de la base de datos

library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.1 ✔ stringr 1.5.2  
✔ ggplot2 4.0.0 ✔ tibble 3.3.0  
✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
✔ purrr 1.1.0   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(here)

here() starts at C:/Users/Oscar Flores/Documents/GitHub/MA1001B

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

Warning: One or more parsing issues, call `problems()` on your data frame for details,  
e.g.:  
 dat <- vroom(...)  
 problems(dat)

Rows: 1980 Columns: 108  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (15): cd, periodo, folio, entidad, n\_ren\_ele, p5\_6esp, p9\_5esp, p12\_9esp...  
dbl (91): control, viv\_sel, num\_hog, hog\_mud, p1, p2, p3\_1, p3\_2, p3\_3, p3\_4...  
lgl (2): p6\_6esp, p11\_6esp  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

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

Warning: One or more parsing issues, call `problems()` on your data frame for details,  
e.g.:  
 dat <- vroom(...)  
 problems(dat)

Rows: 2010 Columns: 108  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (11): folio, p5\_6esp, p6\_6esp, p11\_6esp, p12\_9esp, p15\_5esp, p17\_6esp, p...  
dbl (96): entidad, control, viv\_sel, num\_hog, hog\_mud, n\_ren\_el, cd, periodo...  
lgl (1): p9\_5esp  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

str(M19)

spc\_tbl\_ [1,980 × 108] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
 $ cd : chr [1:1980] "14" "14" "14" "14" ...  
 $ periodo : chr [1:1980] "0219" "0219" "0219" "0219" ...  
 $ folio : chr [1:1980] "11B167" "11B167" "11B167" "11B167" ...  
 $ entidad : chr [1:1980] "01" "01" "01" "01" ...  
 $ control : num [1:1980] 40006 40006 40006 40006 40131 ...  
 $ viv\_sel : num [1:1980] 3 2 1 4 3 4 2 1 1 2 ...  
 $ num\_hog : num [1:1980] 1 1 1 1 1 1 1 1 1 1 ...  
 $ hog\_mud : num [1:1980] 0 0 0 0 0 0 0 0 0 0 ...  
 $ n\_ren\_ele : chr [1:1980] "02" "01" "01" "02" ...  
 $ p1 : num [1:1980] 1 1 1 1 1 1 1 1 1 1 ...  
 $ p2 : num [1:1980] 2 2 2 2 2 1 1 1 2 1 ...  
 $ p3\_1 : num [1:1980] 2 2 2 2 2 2 1 2 2 2 ...  
 $ p3\_2 : num [1:1980] 2 2 2 2 2 1 2 2 2 2 ...  
 $ p3\_3 : num [1:1980] 2 2 1 2 2 1 2 1 2 1 ...  
 $ p3\_4 : num [1:1980] 2 2 2 2 2 2 2 2 2 2 ...  
 $ p3\_5 : num [1:1980] 2 2 2 2 2 1 1 2 2 2 ...  
 $ p4 : num [1:1980] 0 0 0 0 0 0 1 0 0 0 ...  
 $ p5 : num [1:1980] 0 0 0 0 0 0 4 0 0 0 ...  
 $ p5\_6esp : chr [1:1980] NA NA NA NA ...  
 $ p6\_1 : num [1:1980] 0 0 0 0 0 0 2 0 0 0 ...  
 $ p6\_2 : num [1:1980] 0 0 0 0 0 0 2 0 0 0 ...  
 $ p6\_3 : num [1:1980] 0 0 0 0 0 0 1 0 0 0 ...  
 $ p6\_4 : num [1:1980] 0 0 0 0 0 0 2 0 0 0 ...  
 $ p6\_5 : num [1:1980] 0 0 0 0 0 0 2 0 0 0 ...  
 $ p6\_6 : num [1:1980] 0 0 0 0 0 0 2 0 0 0 ...  
 $ p6\_6esp : logi [1:1980] NA NA NA NA NA NA ...  
 $ p7 : num [1:1980] 0 0 0 0 0 0 2 0 0 0 ...  
 $ p7\_3 : num [1:1980] 0 0 0 0 0 0 0 0 0 0 ...  
 $ p8\_1 : num [1:1980] 0 0 0 0 0 0 2 0 0 0 ...  
 $ p8\_2 : num [1:1980] 0 0 0 0 0 0 1 0 0 0 ...  
 $ p9 : num [1:1980] 0 0 0 0 0 0 1 0 0 0 ...  
 $ p9\_5esp : chr [1:1980] NA NA NA NA ...  
 $ p10 : num [1:1980] 0 0 0 0 0 2 0 0 0 0 ...  
 $ p11 : num [1:1980] 0 0 0 0 0 4 0 0 0 0 ...  
 $ p11\_6esp : logi [1:1980] NA NA NA NA NA NA ...  
 $ p12\_1 : num [1:1980] 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_2 : num [1:1980] 0 0 0 0 0 1 0 0 0 0 ...  
 $ p12\_3 : num [1:1980] 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_4 : num [1:1980] 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_5 : num [1:1980] 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_6 : num [1:1980] 0 0 0 0 0 1 0 0 0 0 ...  
 $ p12\_7 : num [1:1980] 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_8 : num [1:1980] 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_9 : num [1:1980] 0 0 0 0 0 2 0 0 0 0 ...  
 $ p12\_9esp : chr [1:1980] NA NA NA NA ...  
 $ p13 : num [1:1980] 0 0 0 0 0 2 0 0 0 0 ...  
 $ p13\_3 : num [1:1980] 0 0 0 0 0 0 0 0 0 0 ...  
 $ p14\_1 : num [1:1980] 0 0 0 0 0 2 0 0 0 0 ...  
 $ p14\_2 : num [1:1980] 0 0 0 0 0 1 0 0 0 0 ...  
 $ p15 : num [1:1980] 0 0 0 0 0 1 0 0 0 0 ...  
 $ p15\_5esp : chr [1:1980] NA NA NA NA ...  
 $ p16 : num [1:1980] 0 0 1 0 0 1 0 3 0 2 ...  
 $ p17 : num [1:1980] 0 0 4 0 0 4 0 3 0 3 ...  
 $ p17\_6esp : chr [1:1980] NA NA NA NA ...  
 $ p18\_1 : num [1:1980] 0 0 1 0 0 1 0 2 0 2 ...  
 $ p18\_2 : num [1:1980] 0 0 1 0 0 1 0 1 0 1 ...  
 $ p18\_3 : num [1:1980] 0 0 1 0 0 2 0 2 0 2 ...  
 $ p18\_4 : num [1:1980] 0 0 1 0 0 2 0 2 0 2 ...  
 $ p18\_5 : num [1:1980] 0 0 1 0 0 1 0 1 0 1 ...  
 $ p19 : num [1:1980] 0 0 2 0 0 2 0 2 0 2 ...  
 $ p19\_3 : num [1:1980] 0 0 0 0 0 0 0 0 0 0 ...  
 $ p20\_1 : num [1:1980] 0 0 1 0 0 2 0 2 0 2 ...  
 $ p20\_2 : num [1:1980] 0 0 2 0 0 1 0 1 0 1 ...  
 $ p21 : num [1:1980] 0 0 1 0 0 2 0 2 0 2 ...  
 $ p21\_5esp : chr [1:1980] NA NA NA NA ...  
 $ p22 : num [1:1980] 0 0 0 0 0 0 0 0 0 0 ...  
 $ p23\_1 : num [1:1980] 0 0 0 0 0 0 0 0 0 0 ...  
 $ p23\_2 : num [1:1980] 0 0 0 0 0 0 0 0 0 0 ...  
 $ p24 : num [1:1980] 0 0 0 0 0 5 2 0 0 0 ...  
 $ p25 : num [1:1980] 0 0 0 0 0 3 3 0 0 0 ...  
 $ p25\_6esp : chr [1:1980] NA NA NA NA ...  
 $ p26 : num [1:1980] 0 0 15 0 0 10 60 15 0 20 ...  
 $ p27 : num [1:1980] 0 0 2 0 0 2 2 2 0 2 ...  
 $ p28 : num [1:1980] 0 0 0 0 0 0 0 0 0 0 ...  
 $ p28\_7esp : chr [1:1980] NA NA NA NA ...  
 $ p29 : num [1:1980] 0 0 2 0 0 4 3 2 0 3 ...  
 $ p30 : num [1:1980] 0 0 2 0 0 4 3 3 0 3 ...  
 $ p31 : num [1:1980] 0 0 2 0 0 2 1 2 0 2 ...  
 $ p32 : num [1:1980] 3 5 0 5 2 0 0 0 1 0 ...  
 $ p32\_6esp : chr [1:1980] NA NA NA NA ...  
 $ p33\_1 : num [1:1980] 2 2 2 2 2 2 2 2 2 2 ...  
 $ p33\_2 : num [1:1980] 2 2 2 2 2 2 2 2 2 2 ...  
 $ p33\_3 : num [1:1980] 2 2 2 2 2 2 2 2 2 2 ...  
 $ p33\_4 : num [1:1980] 2 2 2 2 2 2 2 2 2 2 ...  
 $ p34\_1 : num [1:1980] 2 2 2 2 3 1 1 3 3 3 ...  
 $ p34\_2 : num [1:1980] 2 1 1 1 3 1 1 3 3 1 ...  
 $ p34\_3 : num [1:1980] 2 2 2 1 3 1 1 3 3 1 ...  
 $ p34\_3\_1 : num [1:1980] 0 0 0 1 0 2 2 0 0 1 ...  
 $ p34\_4 : num [1:1980] 2 1 1 1 1 1 1 1 3 1 ...  
 $ p34\_4\_1 : num [1:1980] 0 1 1 1 1 1 1 1 0 1 ...  
 $ p35 : num [1:1980] 1 1 1 1 1 1 1 1 1 1 ...  
 $ p36\_1 : num [1:1980] 2 1 1 1 3 1 1 3 3 1 ...  
 $ p36\_2 : num [1:1980] 2 1 2 1 3 1 1 3 3 1 ...  
 $ p36\_3 : num [1:1980] 2 1 2 1 3 1 1 3 3 1 ...  
 $ p36\_4 : num [1:1980] 2 2 2 2 3 1 1 3 3 1 ...  
 $ factor : num [1:1980] 18795 15188 22781 14041 16030 ...  
 $ h\_lec : num [1:1980] 4 4 3 4 4 1 1 1 4 1 ...  
 $ mat\_lec : num [1:1980] 4 4 3 4 4 3 2 3 4 3 ...  
 $ perslec : num [1:1980] 2 2 1 2 2 1 1 1 2 1 ...  
 [list output truncated]  
 - attr(\*, "spec")=  
 .. cols(  
 .. cd = col\_character(),  
 .. periodo = col\_character(),  
 .. folio = col\_character(),  
 .. entidad = col\_character(),  
 .. control = col\_double(),  
 .. viv\_sel = col\_double(),  
 .. num\_hog = col\_double(),  
 .. hog\_mud = col\_double(),  
 .. n\_ren\_ele = col\_character(),  
 .. p1 = col\_double(),  
 .. p2 = col\_double(),  
 .. p3\_1 = col\_double(),  
 .. p3\_2 = col\_double(),  
 .. p3\_3 = col\_double(),  
 .. p3\_4 = col\_double(),  
 .. p3\_5 = col\_double(),  
 .. p4 = col\_double(),  
 .. p5 = col\_double(),  
 .. p5\_6esp = col\_character(),  
 .. p6\_1 = col\_double(),  
 .. p6\_2 = col\_double(),  
 .. p6\_3 = col\_double(),  
 .. p6\_4 = col\_double(),  
 .. p6\_5 = col\_double(),  
 .. p6\_6 = col\_double(),  
 .. p6\_6esp = col\_logical(),  
 .. p7 = col\_double(),  
 .. p7\_3 = col\_double(),  
 .. p8\_1 = col\_double(),  
 .. p8\_2 = col\_double(),  
 .. p9 = col\_double(),  
 .. p9\_5esp = col\_character(),  
 .. p10 = col\_double(),  
 .. p11 = col\_double(),  
 .. p11\_6esp = col\_logical(),  
 .. p12\_1 = col\_double(),  
 .. p12\_2 = col\_double(),  
 .. p12\_3 = col\_double(),  
 .. p12\_4 = col\_double(),  
 .. p12\_5 = col\_double(),  
 .. p12\_6 = col\_double(),  
 .. p12\_7 = col\_double(),  
 .. p12\_8 = col\_double(),  
 .. p12\_9 = col\_double(),  
 .. p12\_9esp = col\_character(),  
 .. p13 = col\_double(),  
 .. p13\_3 = col\_double(),  
 .. p14\_1 = col\_double(),  
 .. p14\_2 = col\_double(),  
 .. p15 = col\_double(),  
 .. p15\_5esp = col\_character(),  
 .. p16 = col\_double(),  
 .. p17 = col\_double(),  
 .. p17\_6esp = col\_character(),  
 .. p18\_1 = col\_double(),  
 .. p18\_2 = col\_double(),  
 .. p18\_3 = col\_double(),  
 .. p18\_4 = col\_double(),  
 .. p18\_5 = col\_double(),  
 .. p19 = col\_double(),  
 .. p19\_3 = col\_double(),  
 .. p20\_1 = col\_double(),  
 .. p20\_2 = col\_double(),  
 .. p21 = col\_double(),  
 .. p21\_5esp = col\_character(),  
 .. p22 = col\_double(),  
 .. p23\_1 = col\_double(),  
 .. p23\_2 = col\_double(),  
 .. p24 = col\_double(),  
 .. p25 = col\_double(),  
 .. p25\_6esp = col\_character(),  
 .. p26 = col\_double(),  
 .. p27 = col\_double(),  
 .. p28 = col\_double(),  
 .. p28\_7esp = col\_character(),  
 .. p29 = col\_double(),  
 .. p30 = col\_double(),  
 .. p31 = col\_double(),  
 .. p32 = col\_double(),  
 .. p32\_6esp = col\_character(),  
 .. p33\_1 = col\_double(),  
 .. p33\_2 = col\_double(),  
 .. p33\_3 = col\_double(),  
 .. p33\_4 = col\_double(),  
 .. p34\_1 = col\_double(),  
 .. p34\_2 = col\_double(),  
 .. p34\_3 = col\_double(),  
 .. p34\_3\_1 = col\_double(),  
 .. p34\_4 = col\_double(),  
 .. p34\_4\_1 = col\_double(),  
 .. p35 = col\_double(),  
 .. p36\_1 = col\_double(),  
 .. p36\_2 = col\_double(),  
 .. p36\_3 = col\_double(),  
 .. p36\_4 = col\_double(),  
 .. factor = col\_double(),  
 .. h\_lec = col\_double(),  
 .. mat\_lec = col\_double(),  
 .. perslec = col\_double(),  
 .. l\_formato = col\_double(),  
 .. r\_formato = col\_double(),  
 .. p\_formato = col\_double(),  
 .. persleclrp = col\_double(),  
 .. edad = col\_double(),  
 .. sexo = col\_double(),  
 .. anio = col\_double(),  
 .. nivel = col\_double(),  
 .. cond\_activ = col\_character()  
 .. )  
 - attr(\*, "problems")=<externalptr>

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"))

Rows: 108 Columns: 6  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (5): NOMBRE\_CAMPO, TIPO, MNEMÓNICO, CATÁLOGO, RANGO\_CLAVES  
dbl (1): LONGITUD  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

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   
 Length:1980 Length:1980 Length:1980 Length:1980   
 Class :character Class :character Class :character Class :character   
 Mode :character Mode :character Mode :character Mode :character   
   
   
   
   
 control viv\_sel num\_hog hog\_mud   
 Min. :40001 Min. :1.000 Min. :1 Min. :0.00000   
 1st Qu.:40138 1st Qu.:2.000 1st Qu.:1 1st Qu.:0.00000   
 Median :40222 Median :3.000 Median :1 Median :0.00000   
 Mean :40291 Mean :2.502 Mean :1 Mean :0.04697   
 3rd Qu.:40390 3rd Qu.:3.000 3rd Qu.:1 3rd Qu.:0.00000   
 Max. :41420 Max. :4.000 Max. :1 Max. :2.00000   
   
 n\_ren\_ele p1 p2 p3\_1   
 Length:1980 Min. :1.00 Min. :0.000 Min. :0.00   
 Class :character 1st Qu.:1.00 1st Qu.:1.000 1st Qu.:1.00   
 Mode :character Median :1.00 Median :1.000 Median :2.00   
 Mean :1.03 Mean :1.351 Mean :1.53   
 3rd Qu.:1.00 3rd Qu.:2.000 3rd Qu.:2.00   
 Max. :2.00 Max. :2.000 Max. :2.00   
   
 p3\_2 p3\_3 p3\_4 p3\_5 p4   
 Min. :0.000 Min. :0.0 Min. :0.000 Min. :0.000 Min. : 0.000   
 1st Qu.:1.000 1st Qu.:1.0 1st Qu.:2.000 1st Qu.:1.000 1st Qu.: 0.000   
 Median :2.000 Median :2.0 Median :2.000 Median :2.000 Median : 0.000   
 Mean :1.612 Mean :1.6 Mean :1.897 Mean :1.582 Mean : 1.335   
 3rd Qu.:2.000 3rd Qu.:2.0 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.: 2.000   
 Max. :2.000 Max. :2.0 Max. :2.000 Max. :2.000 Max. :50.000   
   
 p5 p5\_6esp p6\_1 p6\_2   
 Min. :0.00 Length:1980 Min. :0.0000 Min. :0.0000   
 1st Qu.:0.00 Class :character 1st Qu.:0.0000 1st Qu.:0.0000   
 Median :0.00 Mode :character Median :0.0000 Median :0.0000   
 Mean :1.37 Mean :0.7848 Mean :0.6889   
 3rd Qu.:3.00 3rd Qu.:2.0000 3rd Qu.:2.0000   
 Max. :6.00 Max. :2.0000 Max. :2.0000   
   
 p6\_3 p6\_4 p6\_5 p6\_6   
 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 :0.6985 Mean :0.6465 Mean :0.7131 Mean :0.8177   
 3rd Qu.:2.0000 3rd Qu.:1.0000 3rd Qu.:2.0000 3rd Qu.:2.0000   
 Max. :2.0000 Max. :2.0000 Max. :2.0000 Max. :2.0000   
   
 p6\_6esp p7 p7\_3 p8\_1   
 Mode:logical Min. :0.0000 Min. : 0.0 Min. :0.0000   
 NA's:1980 1st Qu.:0.0000 1st Qu.: 0.0 1st Qu.:0.0000   
 Median :0.0000 Median : 0.0 Median :0.0000   
 Mean :0.9364 Mean : 195.5 Mean :0.7545   
 3rd Qu.:2.0000 3rd Qu.: 0.0 3rd Qu.:2.0000   
 Max. :3.0000 Max. :99999.0 Max. :2.0000   
   
 p8\_2 p9 p9\_5esp p10   
 Min. :0.0000 Min. :0.0000 Length:1980 Min. : 0.000   
 1st Qu.:0.0000 1st Qu.:0.0000 Class :character 1st Qu.: 0.000   
 Median :0.0000 Median :0.0000 Mode :character Median : 0.000   
 Mean :0.4561 Mean :0.4848 Mean : 1.363   
 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.: 2.000   
 Max. :2.0000 Max. :5.0000 Max. :99.000   
   
 p11 p11\_6esp p12\_1 p12\_2   
 Min. :0.000 Mode:logical Min. :0.0000 Min. :0.0000   
 1st Qu.:0.000 NA's:1980 1st Qu.:0.0000 1st Qu.:0.0000   
 Median :0.000 Median :0.0000 Median :0.0000   
 Mean :1.156 Mean :0.6197 Mean :0.5955   
 3rd Qu.:3.000 3rd Qu.:2.0000 3rd Qu.:2.0000   
 Max. :6.000 Max. :2.0000 Max. :2.0000   
   
 p12\_3 p12\_4 p12\_5 p12\_6   
 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 :0.5949 Mean :0.5742 Mean :0.6172 Mean :0.5818   
 3rd Qu.:2.0000 3rd Qu.:1.0000 3rd Qu.:2.0000 3rd Qu.:2.0000   
 Max. :2.0000 Max. :2.0000 Max. :2.0000 Max. :2.0000   
   
 p12\_7 p12\_8 p12\_9 p12\_9esp   
 Min. :0.0000 Min. :0.000 Min. :0.000 Length:1980   
 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:0.000 Class :character   
 Median :0.0000 Median :0.000 Median :0.000 Mode :character   
 Mean :0.5773 Mean :0.554 Mean :0.651   
 3rd Qu.:1.2500 3rd Qu.:1.000 3rd Qu.:2.000   
 Max. :2.0000 Max. :2.000 Max. :2.000   
   
 p13 p13\_3 p14\_1 p14\_2   
 Min. :0.000 Min. : 0.00 Min. :0.0000 Min. :0.0000   
 1st Qu.:0.000 1st Qu.: 0.00 1st Qu.:0.0000 1st Qu.:0.0000   
 Median :0.000 Median : 0.00 Median :0.0000 Median :0.0000   
 Mean :0.752 Mean : 18.22 Mean :0.6197 Mean :0.3551   
 3rd Qu.:2.000 3rd Qu.: 0.00 3rd Qu.:2.0000 3rd Qu.:1.0000   
 Max. :3.000 Max. :1000.00 Max. :2.0000 Max. :2.0000   
   
 p15 p15\_5esp p16 p17   
 Min. :0.0000 Length:1980 Min. : 0.000 Min. :0.00   
 1st Qu.:0.0000 Class :character 1st Qu.: 0.000 1st Qu.:0.00   
 Median :0.0000 Mode :character Median : 0.000 Median :0.00   
 Mean :0.4096 Mean : 1.092 Mean :1.11   
 3rd Qu.:1.0000 3rd Qu.: 1.000 3rd Qu.:3.00   
 Max. :5.0000 Max. :45.000 Max. :6.00   
   
 p17\_6esp p18\_1 p18\_2 p18\_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 :0.4268 Mean :0.4692 Mean :0.4611   
 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
 Max. :2.0000 Max. :2.0000 Max. :2.0000   
   
 p18\_4 p18\_5 p19 p19\_3   
 Min. :0.0000 Min. :0.000 Min. :0.0000 Min. : 0.0   
 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.: 0.0   
 Median :0.0000 Median :0.000 Median :0.0000 Median : 0.0   
 Mean :0.4975 Mean :0.403 Mean :0.8177 Mean : 512.7   
 3rd Qu.:1.0000 3rd Qu.:1.000 3rd Qu.:2.0000 3rd Qu.: 0.0   
 Max. :2.0000 Max. :2.000 Max. :3.0000 Max. :999999.0   
   
 p20\_1 p20\_2 p21 p21\_5esp   
 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 :0.6475 Mean :0.3672 Mean :0.4879   
 3rd Qu.:2.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
 Max. :2.0000 Max. :2.0000 Max. :5.0000   
   
 p22 p23\_1 p23\_2 p24   
 Min. :0.0000 Min. :0.00000 Min. :0.00000 Min. :0.0000   
 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.0000   
 Median :0.0000 Median :0.00000 Median :0.00000 Median :0.0000   
 Mean :0.1374 Mean :0.07121 Mean :0.05253 Mean :0.7631   
 3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:1.0000   
 Max. :5.0000 Max. :2.00000 Max. :2.00000 Max. :5.0000   
   
 p25 p25\_6esp p26 p27   
 Min. :0.000 Length:1980 Min. : 0.00 Min. :0.000   
 1st Qu.:0.000 Class :character 1st Qu.: 0.00 1st Qu.:0.000   
 Median :0.000 Mode :character Median : 20.00 Median :2.000   
 Mean :1.066 Mean : 27.54 Mean :1.322   
 3rd Qu.:3.000 3rd Qu.: 40.00 3rd Qu.:2.000   
 Max. :6.000 Max. :360.00 Max. :2.000   
   
 p28 p28\_7esp p29 p30   
 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 :2.000 Median :3.000   
 Mean :0.5808 Mean :1.748 Mean :2.111   
 3rd Qu.:0.0000 3rd Qu.:3.000 3rd Qu.:3.000   
 Max. :7.0000 Max. :4.000 Max. :4.000   
   
 p31 p32 p32\_6esp p33\_1   
 Min. :0.000 Min. :0.0000 Length:1980 Min. :0.000   
 1st Qu.:0.000 1st Qu.:0.0000 Class :character 1st Qu.:1.000   
 Median :1.000 Median :0.0000 Mode :character Median :2.000   
 Mean :1.134 Mean :0.7106 Mean :1.712   
 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:2.000   
 Max. :2.000 Max. :6.0000 Max. :2.000   
   
 p33\_2 p33\_3 p33\_4 p34\_1 p34\_2   
 Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.00 Min. :0.00   
 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:1.00 1st Qu.:1.00   
 Median :2.000 Median :2.000 Median :2.000 Median :2.00 Median :1.00   
 Mean :1.801 Mean :1.856 Mean :1.767 Mean :1.71 Mean :1.47   
 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.00 3rd Qu.:2.00   
 Max. :2.000 Max. :2.000 Max. :2.000 Max. :3.00 Max. :3.00   
   
 p34\_3 p34\_3\_1 p34\_4 p34\_4\_1   
 Min. :0.000 Min. :0.0000 Min. :0.000 Min. :0.0000   
 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:0.0000   
 Median :2.000 Median :0.0000 Median :1.000 Median :1.0000   
 Mean :1.639 Mean :0.5323 Mean :1.416 Mean :0.9621   
 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:1.0000   
 Max. :3.000 Max. :3.0000 Max. :3.000 Max. :6.0000   
   
 p35 p36\_1 p36\_2 p36\_3   
 Min. :0.0000 Min. :0.000 Min. :0.000 Min. :0.000   
 1st Qu.:1.0000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000   
 Median :1.0000 Median :1.000 Median :1.000 Median :1.000   
 Mean :0.9818 Mean :1.424 Mean :1.207 Mean :1.334   
 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:1.000 3rd Qu.:2.000   
 Max. :2.0000 Max. :3.000 Max. :3.000 Max. :3.000   
   
 p36\_4 factor h\_lec mat\_lec   
 Min. :0.000 Min. : 2852 Min. :1.000 Min. :1.000   
 1st Qu.:1.000 1st Qu.: 12880 1st Qu.:1.000 1st Qu.:2.000   
 Median :1.000 Median : 16765 Median :1.000 Median :3.000   
 Mean :1.422 Mean : 19297 Mean :2.044 Mean :2.756   
 3rd Qu.:2.000 3rd Qu.: 24724 3rd Qu.:3.000 3rd Qu.:4.000   
 Max. :3.000 Max. :103721 Max. :4.000 Max. :4.000   
 NA's :60 NA's :60   
 perslec l\_formato r\_formato p\_formato persleclrp   
 Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.0   
 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:1.0   
 Median :1.000 Median :2.000 Median :2.000 Median :2.000 Median :1.0   
 Mean :1.258 Mean :1.926 Mean :1.937 Mean :1.929 Mean :1.3   
 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.0   
 Max. :2.000 Max. :3.000 Max. :3.000 Max. :3.000 Max. :2.0   
 NA's :60 NA's :1170 NA's :1332 NA's :1308 NA's :60   
 edad sexo anio nivel   
 Min. :18.00 Min. :1.000 Min. :1.000 Min. : 0.000   
 1st Qu.:31.00 1st Qu.:1.000 1st Qu.:3.000 1st Qu.: 3.000   
 Median :44.00 Median :2.000 Median :3.000 Median : 4.000   
 Mean :45.07 Mean :1.559 Mean :3.463 Mean : 4.368   
 3rd Qu.:57.00 3rd Qu.:2.000 3rd Qu.:4.000 3rd Qu.: 7.000   
 Max. :95.00 Max. :2.000 Max. :6.000 Max. :99.000   
 NA's :59   
 cond\_activ   
 Length:1980   
 Class :character   
 Mode :character

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] 65

## Para quitar los NA’s

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

control viv\_sel num\_hog hog\_mud n\_ren\_el   
 Min. : NA Min. : NA Min. : NA Min. : NA Min. : NA   
 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA   
 Median : NA Median : NA Median : NA Median : NA Median : NA   
 Mean :NaN Mean :NaN Mean :NaN Mean :NaN Mean :NaN   
 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA   
 Max. : NA Max. : NA Max. : NA Max. : NA Max. : NA   
 cd periodo sexo edad anio   
 Min. : NA Min. : NA Min. : NA Min. : NA Min. : NA   
 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA   
 Median : NA Median : NA Median : NA Median : NA Median : NA   
 Mean :NaN Mean :NaN Mean :NaN Mean :NaN Mean :NaN   
 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA   
 Max. : NA Max. : NA Max. : NA Max. : NA Max. : NA   
 nivel cond\_act p1 p2 p3\_1   
 Min. : NA Min. : NA Min. : NA Min. : NA Min. : NA   
 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA   
 Median : NA Median : NA Median : NA Median : NA Median : NA   
 Mean :NaN Mean :NaN Mean :NaN Mean :NaN Mean :NaN   
 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA   
 Max. : NA Max. : NA Max. : NA Max. : NA Max. : NA   
 p3\_2 p3\_3 p3\_4 p3\_5 p4   
 Min. : NA Min. : NA Min. : NA Min. : NA Min. : NA   
 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA   
 Median : NA Median : NA Median : NA Median : NA Median : NA   
 Mean :NaN Mean :NaN Mean :NaN Mean :NaN Mean :NaN   
 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA   
 Max. : NA Max. : NA Max. : NA Max. : NA Max. : NA   
 p5 p5\_6esp p6\_1 p6\_2 p6\_3   
 Min. : NA Length:0 Min. : NA Min. : NA Min. : NA   
 1st Qu.: NA Class :character 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA   
 Median : NA Mode :character Median : NA Median : NA Median : NA   
 Mean :NaN Mean :NaN Mean :NaN Mean :NaN   
 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA   
 Max. : NA Max. : NA Max. : NA Max. : NA   
 p6\_4 p6\_5 p6\_6 p6\_6esp p7   
 Min. : NA Min. : NA Min. : NA Length:0 Min. : NA   
 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA Class :character 1st Qu.: NA   
 Median : NA Median : NA Median : NA Mode :character Median : NA   
 Mean :NaN Mean :NaN Mean :NaN Mean :NaN   
 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA   
 Max. : NA Max. : NA Max. : NA Max. : NA   
 p7\_3 p8\_1   
 Min. : NA Min. : NA   
 1st Qu.: NA 1st Qu.: NA   
 Median : NA Median : NA   
 Mean :NaN Mean :NaN   
 3rd Qu.: NA 3rd Qu.: NA   
 Max. : NA Max. : NA

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

prueba