Fundamentos de estadistica para analitica de datos

Juan Salazar - Camilo Alarcon

31 de marzo de 2023

Resumen

El presente trabajo de estadísticas se enfoca en el análisis de los resultados de las pruebas ICFES en Colombia(Bogota) en los últimos años. A través de diversas técnicas estadísticas, se pretende examinar los datos y extraer conclusiones relevantes sobre el desempeño de los estudiantes en diferentes áreas de conocimiento, identificar patrones y tendencias, y explorar los factores que pueden influir en los resultados de las pruebas.

1. Introduction

Las pruebas ICFES son un instrumento de evaluación que se utiliza en Colombia para medir el conocimiento y las habilidades de los estudiantes de educación secundaria y para evaluar su capacidad para ingresar a la educación superior. Desde su creación en 1967, estas pruebas han sido objeto de diversos estudios y análisis estadísticos que han permitido conocer mejor los resultados obtenidos por los estudiantes, así como identificar fortalezas y debilidades del sistema educativo en Colombia.

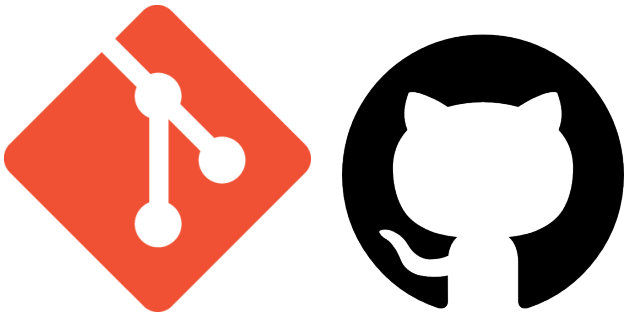
El presente trabajo de estadísticas tiene como objetivo principal analizar los resultados de las pruebas ICFES en Colombia en los últimos años, con el fin de identificar tendencias, patrones y posibles factores que puedan influir en el rendimiento de los estudiantes. Para ello, se aplicarán diversas técnicas estadísticas que permitirán examinar los datos y extraer conclusiones relevantes sobre el desempeño de los estudiantes en diferentes áreas de conocimiento.

Es importante destacar que el análisis de los resultados de las pruebas ICFES puede ofrecer información valiosa para la toma de decisiones en el ámbito educativo. A través de la identificación de las fortalezas y debilidades del sistema educativo, se pueden implementar estrategias que permitan mejorar la calidad de la educación en Colombia y, por ende, la formación de los estudiantes.

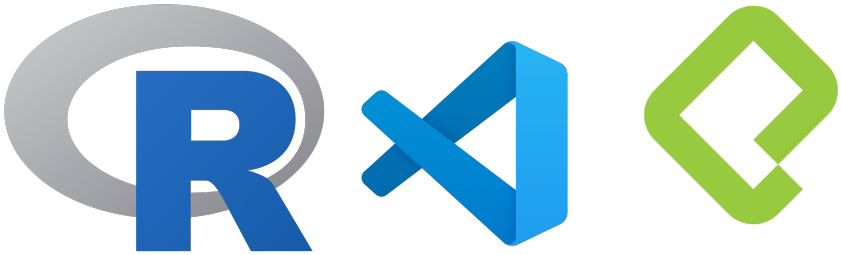
En resumen, el presente trabajo de estadísticas se enfoca en analizar los resultados de las pruebas ICFES en Colombia, con el objetivo de ofrecer información útil que permita mejorar la educación en el país. A través del análisis estadístico de los datos, se espera identificar las áreas de mejora y proponer soluciones que contribuyan a la formación de estudiantes más preparados y competitivos.

2. Aplicaciones utilizadas

2.1. Control de versiones



2.2. hernel + entorno



2.3. Limpieza de entorno y cargue de la base para analisis

```
rm(list = ls())

library(readxl)

url <- "https://raw.githubusercontent.com/jusalazar2/R/main/Tarea_1/tablas_dept/BOGOT%C3%81.csv"
destfile <- "Bogota.csv"
curl::curl_download(url, destfile)
icfes <- read.csv2(destfile, sep = ",", dec=".",stringsAsFactors = TRUE)
#View(icfes)

# Ver variable de un dataframe
str(icfes)
```

2.4. Describimos nuestra base para entender que tenemos y como podemos comenzar el analisis

```
## 'data.frame':    83600 obs. of  25 variables:
##  $ ESTU_DEPTO_RESIDE           : Factor w/ 1 level "BOGOTÁ": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FAMI_ESTRATOVIVIENDA        : Factor w/ 8 levels "-","Estrato 1",..: 4 4 4 5 4 3 3 5 3 3 ...
##  $ FAMI_PERSONASHOGAR          : Factor w/ 7 levels "","-","1 a 2",..: 7 4 4 4 4 4 4 4 4 2 ...
##  $ FAMI_EDUCACIONPADRE         : Factor w/ 14 levels "","-","Educación profesional completa",..: 13 10 11 7 11 4 13 8 11
12 ...
##  $ FAMI_EDUCACIONMADRE         : Factor w/ 14 levels "","-","Educación profesional completa",..: 13 11 12 7 8 3 11 5 11
14 ...
##  $ FAMI_TIENEINTERNET          : Factor w/ 4 levels "","-","No","Si": 4 3 4 4 4 4 4 3 4 3 ...
##  $ FAMI_TIENECOMPUTADOR        : Factor w/ 4 levels "","-","No","Si": 4 4 4 4 4 4 4 3 4 2 ...
##  $ FAMI_COMELECHEDERIVADOS     : Factor w/ 6 levels "","-","1 o 2 veces por semana",..: 6 6 6 6 6 6 4 3 4 4 ...
##  $ FAMI_COMECARNEPESCADOHUEVO  : Factor w/ 6 levels "","-","1 o 2 veces por semana",..: 4 6 5 6 6 6 4 5 3 4 ...
##  $ FAMI_COMECEREALFRUTOSLEGUMBRE: Factor w/ 6 levels "","-","1 o 2 veces por semana",..: 3 4 6 6 6 4 3 6 4 3 ...
##  $ ESTU_DEDICACIONINTERNET     : Factor w/ 7 levels "","-","30 minutos o menos",..: 5 5 6 5 4 6 4 3 5 3 ...
##  $ COLE_GENERO                 : Factor w/ 3 levels "FEMENINO","MASCULINO",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ COLE_NATURALEZA             : Factor w/ 2 levels "NO OFICIAL","OFICIAL": 1 2 2 1 1 1 2 2 1 2 ...
##  $ COLE_CALENDARIO             : Factor w/ 3 levels "A","B","OTRO": 1 1 1 1 1 1 1 1 1 1 ...
##  $ COLE_CARACTER               : Factor w/ 6 levels "","-","ACADÉMICO",..: 3 3 6 3 3 3 3 3 5 6 ...
##  $ COLE_JORNADA                : Factor w/ 6 levels "COMPLETA","MAÑANA",..: 1 2 2 2 1 1 2 3 1 6 ...
##  $ PUNT_LECTURA_CRITICA        : int  60 62 63 64 52 55 57 75 53 55 ...
##  $ PUNT_MATEMATICAS            : int  65 54 57 56 66 64 51 73 48 62 ...
##  $ PUNT_C_NATURALES            : int  54 61 55 59 54 57 45 59 43 74 ...
##  $ PUNT_SOCIALES_CIUDADANAS    : int  59 73 57 60 52 57 38 75 46 59 ...
##  $ PUNT_INGLES                 : num  63 53 52 68 58 59 56 64 46 64 ...
##  $ PUNT_GLOBAL                 : int  299 309 288 302 281 292 242 350 237 313 ...
##  $ ESTU_FECHANACIMIENTO        : Factor w/ 4572 levels "01/01/1900 12:00:00 AM",..: 2283 1495 478 2160 1581 4079 1016 26
75 3054 3750 ...
##  $ ESTU_GENERO                 : Factor w/ 3 levels "-","F","M": 3 3 3 3 3 3 3 3 3 3 ...
##  $ ESTU_GENERACION.E           : Factor w/ 4 levels "GENERACION E - EXCELENCIA DEPARTAMENTAL",..: 4 3 4 4 4 4 4 2 4 3
...
```

2.5. Validamos el nombre de los campos para comenzar a trabajar con ellos

```
colnames(icfes)
```

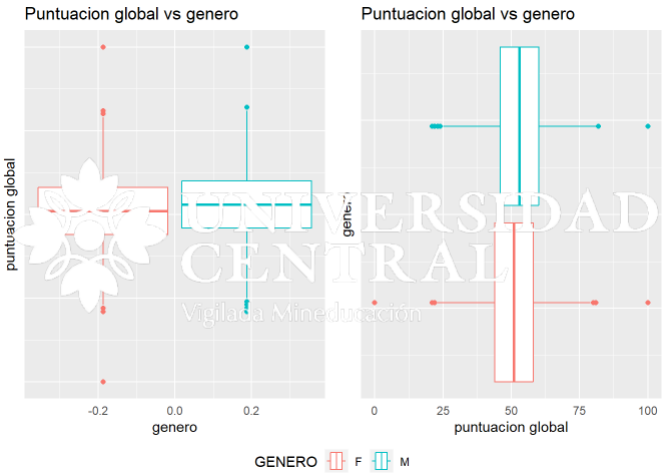
```
##  [1] "ESTU_DEPTO_RESIDE"            "FAMI_ESTRATOVIVIENDA"
##  [3] "FAMI_PERSONASHOGAR"           "FAMI_EDUCACIONPADRE"
##  [5] "FAMI_EDUCACIONMADRE"          "FAMI_TIENEINTERNET"
##  [7] "FAMI_TIENECOMPUTADOR"         "FAMI_COMELECHEDERIVADOS"
##  [9] "FAMI_COMECARNEPESCADOHUEVO"   "FAMI_COMECEREALFRUTOSLEGUMBRE"
## [11] "ESTU_DEDICACIONINTERNET"      "COLE_GENERO"
## [13] "COLE_NATURALEZA"              "COLE_CALENDARIO"
## [15] "COLE_CARACTER"                "COLE_JORNADA"
## [17] "PUNT_LECTURA_CRITICA"         "PUNT_MATEMATICAS"
## [19] "PUNT_C_NATURALES"             "PUNT_SOCIALES_CIUDADANAS"
## [21] "PUNT_INGLES"                  "PUNT_GLOBAL"
## [23] "ESTU_FECHANACIMIENTO"         "ESTU_GENERO"
## [25] "ESTU_GENERACION.E"
```

2.6. Mejor puntuacion por genero

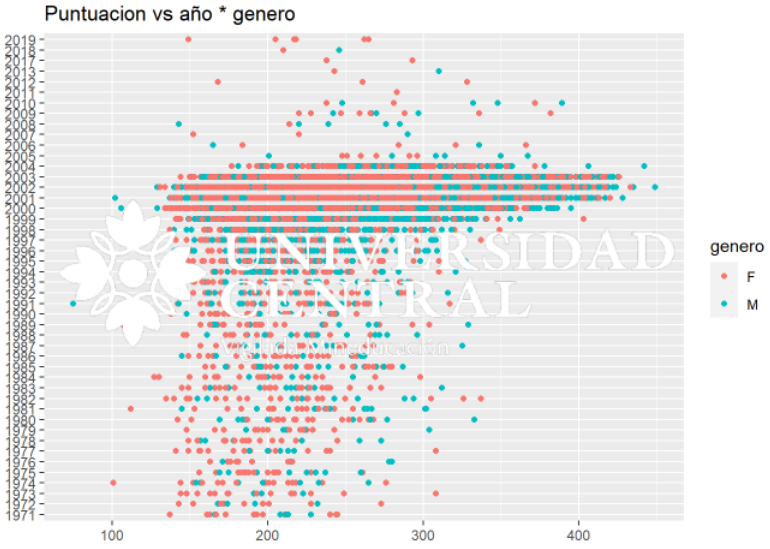
seleccionamos la data

```
#Selecionamos la data

pgrafica <- subset(icfes, ESTU_GENERO == c("M","F"),
                        select = c( `PUNT_C_NATURALES`,`PUNT_GLOBAL`,`PUNT_INGLES`,`PUNT_LECTURA_CRITICA`,`PUNT_MATEMATICAS
`,`PUNT_SOCIALES_CIUDADANAS`,`ESTU_GENERO` ))

#Cambiamos el nombre de los encabezados

columnitas <- c("GLOBAL","INGLES","MATEMATICAS","SOCIALES_CIUDADANAS","LECTURA_CRITICA","C_NATURALES","GENERO")

colnames(pgrafica) <- columnitas
```

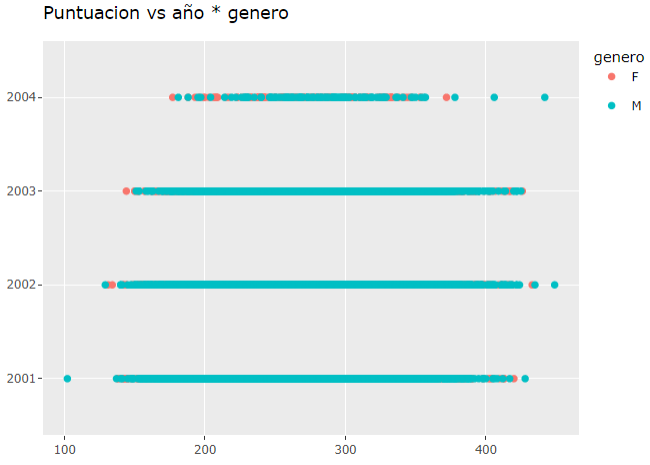
graficamos el resultado



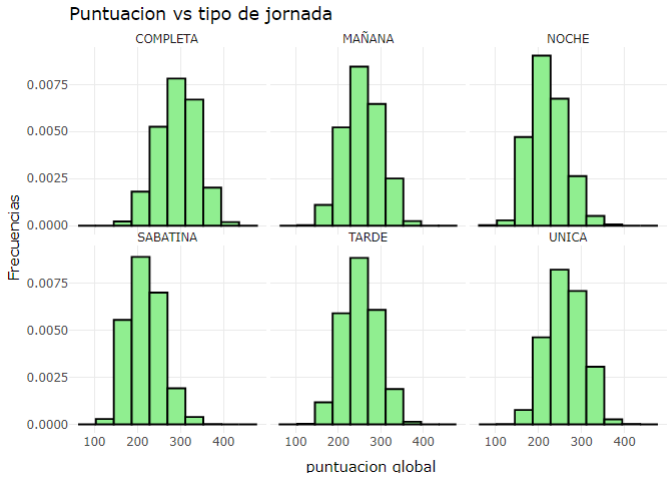
2.7. Generamos una grafica de punto para ver la distribucion de los participantes de la presentacion del examen



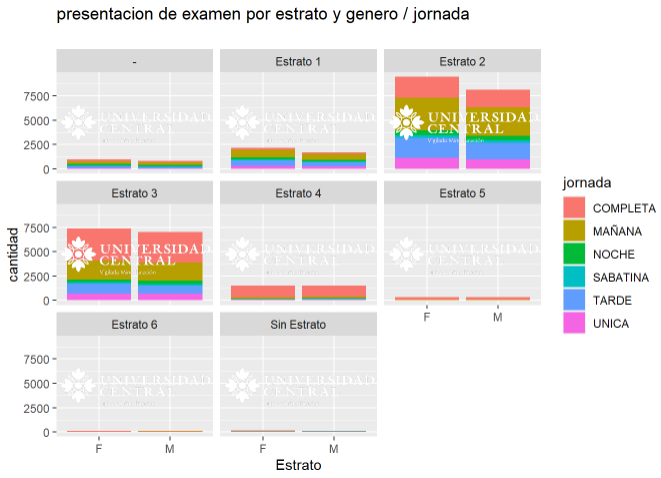
Si hacemos un poco de zoom para ver el mejor resultado de todos, podemos observar que en el año 2002 una mujer lo obtuvo



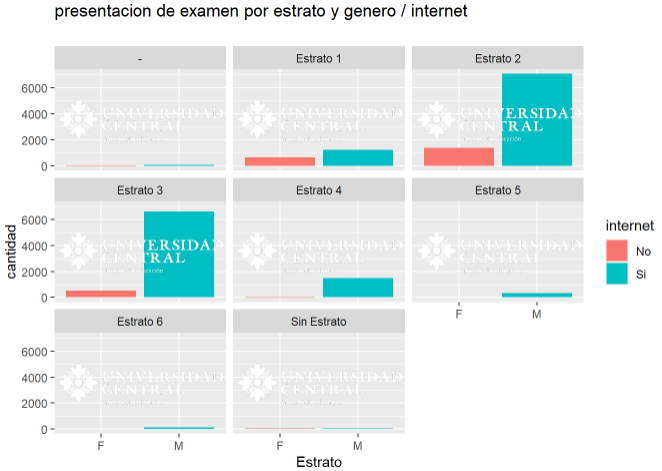
2.8. Verificaremos que grupo obtuvieron el mejor puntaje



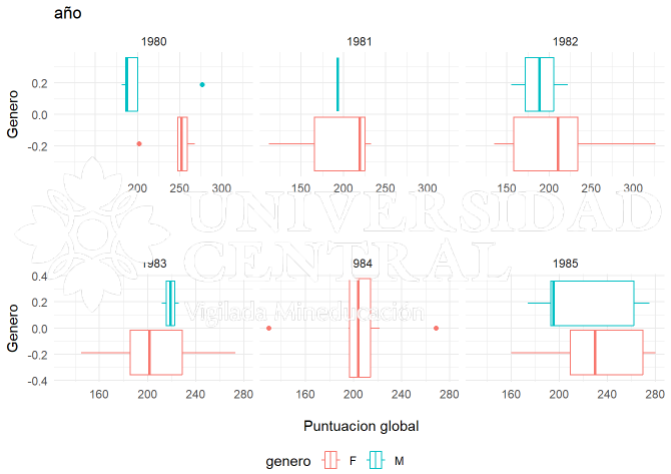
2.9. Validaremos el estrato con mas participantes

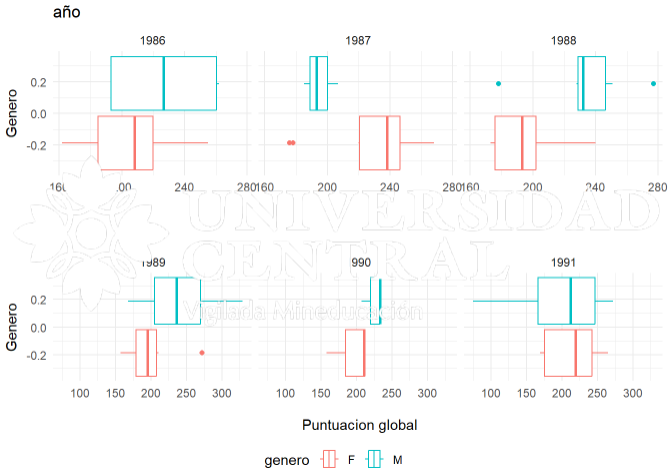


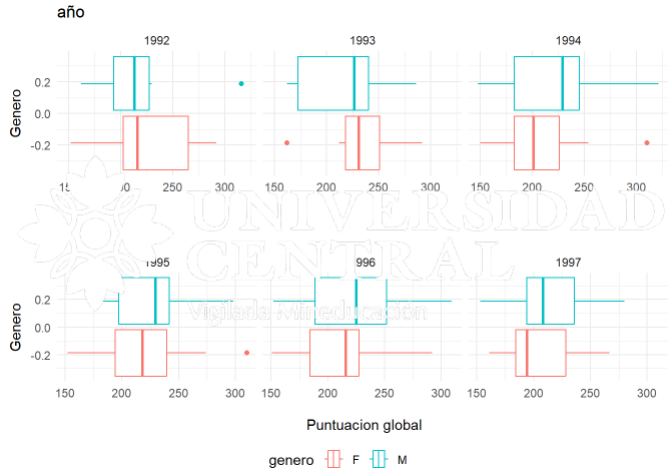
2.10. Validaremos el acceso a internet de los participantes

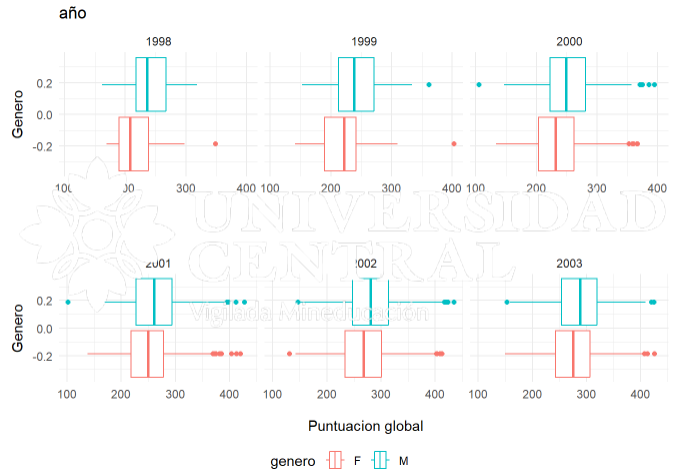


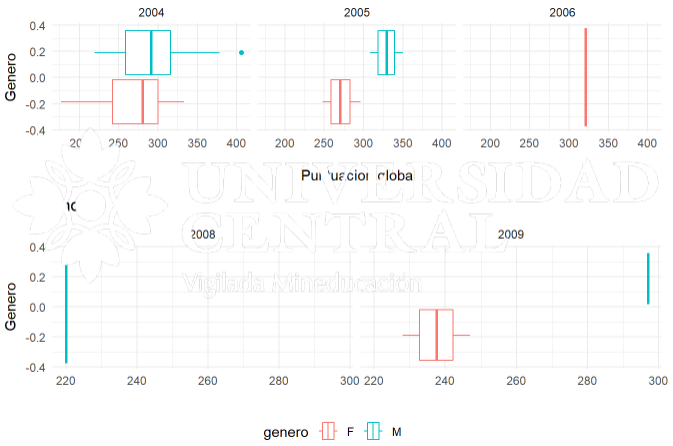
2.11. Validaremos la distribucion de los participantes por año de nacimiento

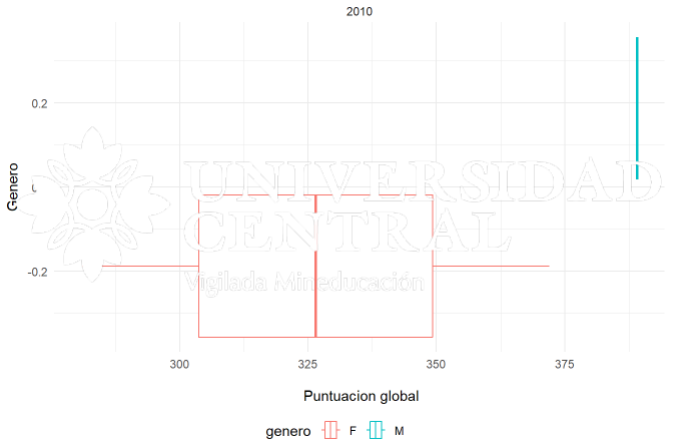


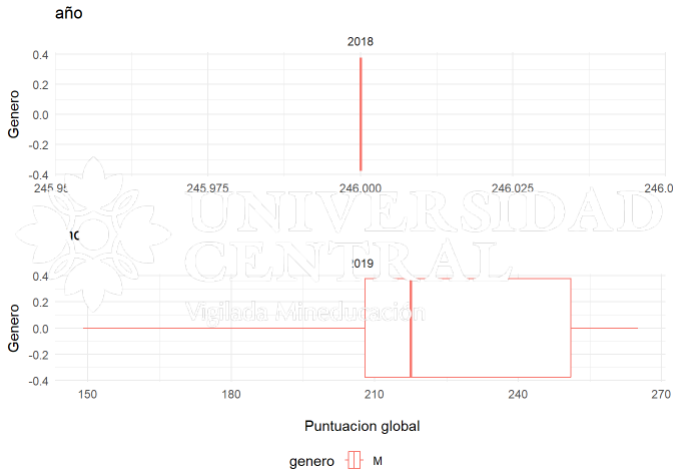




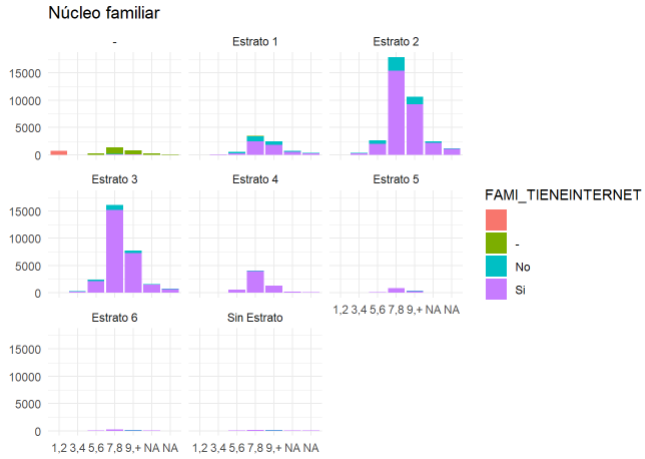

año
1998
1999
2000
2001
2002
2003
Genero
0.2
0.0
-0.2
100
200
300
400
Puntuacion global
genero
F
M


2004
2005
2006
2008
2009
Genero
0.4
0.2
0.0
-0.2
-0.4
220
240
260
280
300
350
400
genero
F
M

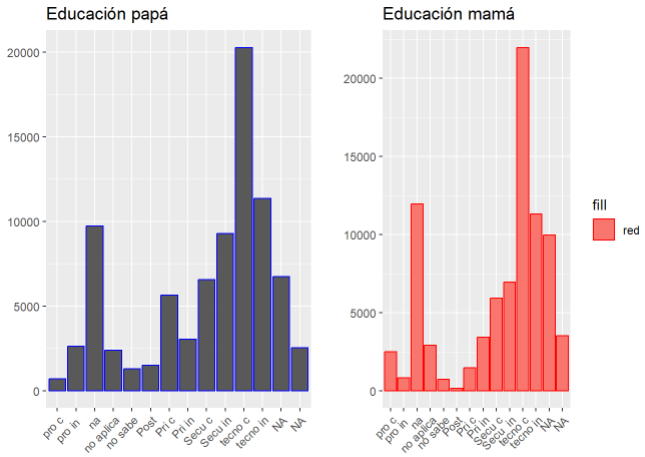

2010
Genero
0.2
-0.2
300
325
350
375
Puntuacion global
genero
F
M



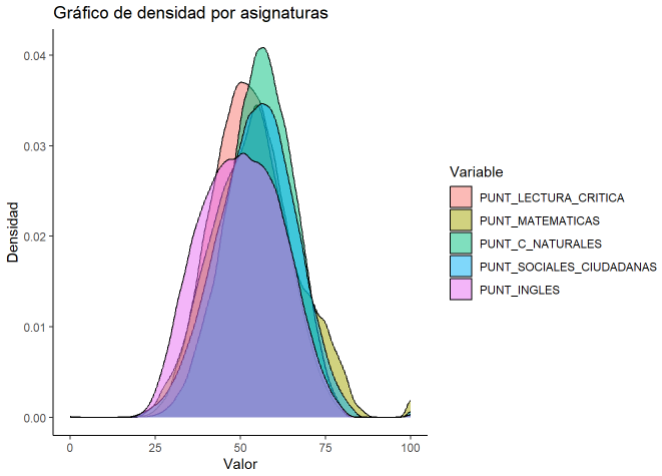
2.12. Validaremos el nucleo familiar vs estrato / con acceso a internet de los participantes



2.13. Educacion de los padres



2.14. Graficaremos la densidad de las pruebas para medir el mejor resultado



3. Conclusiones

3.1. La base de bogota esta rara

Despues de ver las distribuciones de las graficas pordemos concluir que, los estratos que mas participan de este examen son el 2 y 3, en su mayoria mujeres, pero mucha de la poblacion tiene acceso al internet, pero sin duda las personas que mas le meten ganas son las de jornada norturna

Referencias