

UNIVERSIDAD OBERTA DE CATALUNYA

VISUALIZACIÓN DE DATOS

PRIMER CICLO

AULA 2

NOMBRE: JOSHELYN ANDREA INTRIAGO PAREDES

PROF: MAR CANET SOLA

FECHA: 3 DE JUNIO DE 2021

A8: Creación de la visualización y entrega del proyecto (Práctica)

- 1) Título de la visualización donde se presentan la visualización realizada. URL de la visualización y del código. Y descripción corta del documento y del que se presenta**

Título: Dashboard: Diferencias de Género en los Jefes de hogar del Mercado Laboral Ecuatoriano sobre empleo y desempleo en Abril 2021.

Fuente: Enemdu - INEC

Periodo: Abril 2021

Autor: Joshelyn Intriago

Keywords: Empleo, desempleo, Ecuador, género.

La ENEMDU constituye la fuente oficial de los indicadores de mercado laboral del país. El marco conceptual y metodológico sigue recomendaciones internacionales de la Organización Internacional del Trabajo (OIT); organismo que periódicamente brinda apoyo técnico al INEC (Instituto Nacional de Estadísticas y Censo)

Este conjunto de datos corresponde al primer cuatrimestre del 2021, esta encuesta se la realiza 4 veces al año, sin embargo se realizan proyecciones mensuales. Es una encuesta que sirve para calcular los índices de empleo, desempleo y subempleo a nivel nacional, provincial, cantonal, por sexo, áreas, autoidentificación y demás permite confirmar brechas salariales entre hombres y mujeres y entre autoidentificación étnica. En esta ocasión el dashboard que se presentará tiene como

objetivo mostrar principales diferencias de género en los indicadores de mercado laboral ecuatoriano, el dashboard inicia con gráficos generales sobre distintos aspectos y luego se enfoca en la distinción entre género, es importante recalcar que una de las principales medidas establecida es el promedio del ingreso laboral medida en dólares americanos USD, además existe la medida de ingreso total donde se suman ingresos por remesas , jubilaciones, renta, donaciones, etc.

Así también se presenta una evolución histórica desde el diciembre de 2007 hasta abril de 2021 tanto del empleo y desempleo a nivel nacional , y por hombres y mujeres.

En el siguiente link se encuentra toda la información sobre la fuente de datos y de la encuesta mencionada: <https://www.ecuadorencifras.gob.ec/empleo-abril-2021/>

Descripción del informe:

Para facilitar el uso y procesamiento de la bases de datos de La Encuesta Nacional de Empleo, Desempleo y Subempleo - ENEMDU abril 2021 se hace uso del diccionario de variables . La base de datos consta de 30.414 casos y 145 variables, las mismas que son descritas en el archivo del diccionario de variables. Cabe recalcar que más adelante se explicará los pasos del procesamiento por lo cual una vez la data procesada ya no constará del mismo número de variables ni de observaciones.

En este informe se muestra la descripción y titulo de la visualización, así como las preguntas que responde la visualización y para qué y quien sirve el dashboard.

La descripción técnica de la visualización sobre qué librerías, software y pasos en el tratamiento de datos han sido utilizados para mejorar la calidad del conjunto de datos, y además tener una idea principal de las variables , problemas en los datos, distribución de los datos, etc.

- 1.1) El código es tiene que publicar en un repositorio público de GitHub con una licencia de código abierto. Tiene que contenga los archivos necesarios para correr la aplicación en local con la ayuda de un servidor web si fuera necesario.**

Link: https://github.com/jaintriago/practica_final_visualizacion

- 1.2) La URL de la visualización tiene que ser pública y poderse explorar sin login. Por ejemplo GitHub pages (<https://pages.github.com/>), surge.sh o subir la aplicación a un espacio web propio accesible o compartir un url accesible sin usuario y contraseña.**

Link:

https://public.tableau.com/app/profile/joshelyn2159/viz/final_16228779004940/Dashboard1

- 2) Explicar razonadamente qué preguntas responde la visualización presentada y qué uso puede tener por un usuario tipo.**

Las principales pregunta que responde la visualización son la siguientes:

Las respuestas a estas preguntas se encuentran en el dashboard elaborado en Tableau

- ¿Según la autoidentificación étnica quien gana más?

Respuesta: Las personas que se autoidentifican como blancas ganan más.

- ¿Quiénes ganan más según el sexo?

Respuesta: Ganan más los hombres su promedio de ingreso laboral es superior al de las mujeres.

- ¿Existe brecha salarial entre los hombres y las mujeres y de cuánto es esta diferencia?

Respuesta: Si existe brecha y es de aproximadamente 200 USD, y esta diferencia es significativa.

- ¿En qué tipo de sectores económicos o industriales, rama de actividad o grupos de ocupación reciben más salario las personas?

Respuesta: Ganan más los patronos, empleados de gobierno, personal adm/directivo de empresas públicas y privadas, defensa nacional y seguridad social, es decir el personal de jerarquía de las fuerzas armadas y policía, estas personas son del sector formal y poseen nombramientos.

- ¿A mayor nivel de instrucción mayores ingresos laborales?

Respuesta: Si, las personas con postgrados tienen un ingreso promedio superior a diferencia de las demás personas

- ¿El promedio de ingreso laboral depende del estado civil de las personas?

Respuesta: Si depende o por lo menos hay una diferencia entre los niveles, ya que hay una diferencia significativa entre las personas viudas y las casadas, esto puede deberse a que las personas viudas en su mayoría son personas de la tercera edad.

- ¿Existe una correlación entre la experiencia laboral (años de trabajo) y el sueldo que una persona recibe?

Respuesta: Si, a mayor experiencia laboral, mayor serán los ingresos percibidos.

- ¿En que sector, rama de actividad trabajan más las mujeres?

Respuesta: en el sector de comercio, del que pertenece al sector informal.

- ¿Hay más mujeres desempleadas que hombres?

Respuesta: Si, la tasa de desempleo es superior en las mujeres que en los hombres

- ¿Hubo una disminución de la tasa de empleo durante el último año en relación a los años anteriores?

Respuesta: No, a pesar de la pandemia la tasa de empleo no presenta un cambio radical o una diferencia significativa con los años anteriores.

- ¿Cuál es el principal motivo de desempleo?

Respuesta: El principal motivo de desempleo es porque el contrato de trabajo finalizó.

- ¿En qué rango de edad se encuentran las personas que ganan más?

Respuesta: Las personas entre 25 y 39 años ganan en promedio más que las demás personas de otras edades.

Explicación:

La razón por la que se puede responder a las preguntas anteriormente planteadas, es porque se han considerado esas variables dentro de la visualización y se han hecho los gráficos adecuados para que se pueda dar respuestas a esas preguntas planteadas.

Tipos de gráficos usados:

La mayoría de los gráficos son de barras simples son los que mejores representan la información que quería mostrar.

También hay de barras apilas, barras a los lados, barras horizontales, gráficos de líneas para las series de tiempo, gráfico de burbuja, treemaps, histogramas.

Utilidad:

Sirve para tener una idea de lo que ocurre en el mercado laboral ecuatoriano, en que sectores trabajan los ecuatorianos, quienes son los que ganan más, diferencias entre grupos, brechas salariales por género, niveles de desempleo, evolución de las tasas de empleo y desempleo desde el 2007 a la actualidad.

Esta visualización le puede servir, a personas que trabajen en distintos ministerios, personal del gobierno para la toma de decisiones, a fundaciones,

organizaciones no gubernamentales, a empresarios que deseen invertir en el país, al sector académico tanto estudiantes como profesores y usuarios en general que estén interesados en el tema.

3) Descripción técnica del proyecto: lenguajes, librerías, licencias, descripción técnica del proyecto.

Para este proyecto se uso Rstudio para realizar el tratamiento de los datos, y una primera visualización del conjunto de datos y para realizar el informe en Rmarkdown del proyecto, se usan librerías como dplyr ggplot2 que son las principales que se usaron. Más adelante se explicará el código y los pasos del tratamiento de datos que se realizó al conjunto de datos.

Y para realizar el dashboard se usa una licencia que otorga la universidad de Tableau, así mismo se genera un link del dashboard con ayuda de Tableau Public. Se ha utilizado este software ya que es fácil, interactivo y permite publicarlo de manera gratuita.

Explicación general del procesamiento en R:

Primero se carga la base de datos tipo .sav que es formato SPSS ya que ahí las variables ya estaban en factores con sus respectivas etiquetas, luego se inspeccionó las variables a ver cuáles poseían valores perdidos o valores centinelas o de no informa /no responde y para observar que variables se iban a seleccionar para el dashboard y también para observar en que tipo de variables se encontraban los datos. Uno de los primero pasos también fue filtrar la base por jefes de hogar, ya que la encuesta la metodología de muestreo se seleccionan hogares entonces en un hogar se encuestan a todos los miembros del hogar, por ello solo se filtra la información del jefe de hogar.

Luego se procede a transformar las variables que estaban como factores especialmente las distintas variables que había de ingresos pasarlas a numéricas como corresponda; luego a recodificar los valores centinelas (999999 o 999) o los valores de no informa/ no responde como NA, a recategorizar variables como la edad que era mejor tenerla en grupos de edad para la visualización. También se procede a calcular la variable ingresos_totales que es la suma del ingreso laboral más los demás ingresos (por remesas, transacciones de capitales, donaciones, jubilación ,pensión, renta). Además se procede a renombrar las variables para cuando se usen en el Tableau tengan un nombre entendible ya que en la pase de SPSS estaban codificadas como p01, p02 ... etc, por ello también se tuvo que hacer uso del diccionario de variables tanto para comprobar el nombre de la variable y el tipo de variable.

Finalmente Se procede hacer gráficos básicos, se hacen histogramas para las variables numéricas y gráfico de barras para las variables categóricas, esto solo para tener una idea básica de las variables, cual es la distribución de los datos, las etiquetas y demás.

A continuación se colocará el código de R y las salidas, cabe mencionar que el script estará disponible en el repositorio de github.

```
library(foreign); library(dplyr); library(purrr); library(ggplot2); library(
openxlsx)

#Leer el dataset

enemdu <- read.spss("enemdu_persona_2021_04.sav", to.data.frame = TRUE)

# ver el tipo de variables
str(enemdu)

## 'data.frame':      30414 obs. of      145 variables:
## $ area      : Factor w/ 2 levels " Urbana "," Rural ": 1 1 1 1 1 1 1 1 1
## $ upm       : Factor w/ 1274 levels "010150000802",...: 1 1 1 1 1 1 1 1 1
```

```

1
## $ p01      : Factor w/ 34 levels " Persona 1 ",...: 10 11 12 13 10 11 1
2              13              10              11              ...
## $ p02      : Factor w/ 2 levels " Hombre "," Mujer ": 1 2 2 2 1 2 1 2 2
2
## $ p03      : Factor w/ 100 levels "0","1","10","11",...: 65 66 27 27 54
42          21          10          80          63          ...
## $ p04      : Factor w/ 9 levels " Jefe "," Cónyuge ",...: 1 2 3 3 1 2 3
3              1              8              ...
## $ p05a     : Factor w/ 10 levels " IEES, seguro general ",...: 1 2 1 10
2              10              10              2              1              1              ...
## $ p05b     : Factor w/ 10 levels " IEES, seguro general ",...: 10 10 10
10           10           10           10           10           10           ...
## $ p06      : Factor w/ 6 levels " Casado(a) ",...: 1 1 6 6 1 1 6 6 4 6 .
..
## $ p07      : Factor w/ 2 levels " Si "," No ": 2 2 2 2 2 2 2 1 2 2 ...
## $ p08      : Factor w/ 6 levels " La mañana ",...: NA NA NA NA NA NA NA
1              NA              NA              ...
## $ p08a     : Factor w/ 4 levels " Aprendamos Juntos en casa ",...: NA N
A              NA              NA              NA              NA              1              NA              NA              ...
## $ p09      : Factor w/ 17 levels " Edad "," Terminó sus estudios ",...:
1              1              5              2              1              15              2              NA              1              1              ...
## $ p10a     : Factor w/ 10 levels " Ninguno "," Centro de alfabetizació
n              ",...: 9 9 9 8 9 6 9 7 6 4 ...
## $ p10b     : num 3 3 3 2 3 5 5 1 3 6 ...
## $ p11      : Factor w/ 2 levels " Si "," No ": NA NA NA NA NA NA NA NA
NA              1              ...
## $ p12a     : Factor w/ 2 levels " Si "," No ": 2 2 2 1 2 NA 1 NA NA NA
...
## $ p12b     : Factor w/ 460 levels "10110102","10110201",...: NA NA NA 3
4              NA              NA              258              NA              NA              NA              ...
## $ p15      : Factor w/ 8 levels " Indígena "," Afroecuatoriano ",...: 6
6              6              6              6              6              6              6              1              ...
## $ p15aa    : Factor w/ 3 levels " En esta ciudad o parroquia rural "
,...: 1 2 1 1 2 2 2 1 1 2 ...
## $ p15ab    : num NA 131450 NA NA 30351 ...
## $ cod_inf  : Factor w/ 25 levels " Persona 1 ",...: 4 4 4 4 1 1 1 1 1 1
...
## $ p20      : Factor w/ 2 levels " Si "," No ": 2 2 1 2 1 2 2 2 2 1 ...
## $ p21      : Factor w/ 12 levels " Atender negocio propio ",...: 12 12 N
A              12              12              12              12              NA              ...
## $ p22      : Factor w/ 2 levels " Si "," No ": 2 2 NA 2 NA 2 2 2 2 NA .
..
## $ p23      : Factor w/ 7 levels " Vacaciones ",...: NA NA NA NA NA NA N
A              NA              NA              NA              ...
## $ p231     : Factor w/ 2 levels " Si "," No ": NA NA NA NA NA NA NA NA
NA              NA              ...
## $ p232     : Factor w/ 2 levels " Si "," No ": NA NA NA NA NA NA NA NA
NA              NA              ...
## $ p233     : Factor w/ 3 levels " 3 meses o menos ",...: NA NA NA NA NA
NA              NA              NA              NA              ...

```



```

## $ p24      : Factor w/ 85 levels "1","10","102",...: NA NA 38 NA 38 NA N
A            NA            NA            38            ...
## $ p25      : Factor w/ 9 levels " Su trabajo así lo exige ",...: NA NA N
A            NA            NA            NA            NA            NA            ...
## $ p251     : Factor w/ 2 levels " Si "," No ": NA NA NA NA NA NA NA NA NA
NA            NA            ...
## $ p26      : Factor w/ 5 levels " Horario normal ",...: NA NA 1 NA 1 NA
NA            NA            NA            1            ...
## $ p27      : Factor w/ 4 levels " Trabajar más horas en su trabajo act
ual          ",...: NA NA 4 NA 1 NA NA NA NA 4 ...
## $ p28      : Factor w/ 2 levels " Si "," No ": NA NA NA NA 1 NA NA NA N
A            NA            ...
## $ p29      : Factor w/ 44 levels "1","10","11",...: NA NA NA NA 2 NA NA
NA            NA            NA            ...
## $ p30      : Factor w/ 7 levels " Buscar más trabajo ",...: NA NA NA NA
7            NA            NA            NA            NA            ...
## $ p31      : Factor w/ 5 levels " Por que no trabaja jornada completa "
,...: NA NA NA NA NA NA NA NA NA NA NA ...
## $ p32      : Factor w/ 11 levels " Acudir a sitios de contratación tem
poral        ",...: 11 11 NA 11 NA 11 8 11 11 NA ...
## $ p33      : Factor w/ 42 levels "1","10","11",...: NA NA NA NA NA NA 8
NA            NA            NA            ...
## $ p34      : Factor w/ 12 levels " Trabajo ocasional ",...: 12 12 NA 3 N
A            9            NA            9            12            NA            ...
## $ p341     : Factor w/ 2 levels " Si "," No ": NA NA NA 1 NA NA NA NA N
A            NA            ...
## $ p35      : Factor w/ 2 levels " Si "," No ": NA NA NA 1 NA NA NA NA N
A            NA            ...
## $ p351     : Factor w/ 2 levels " Si "," No ": 2 2 NA NA NA 2 NA 2 2 N
A            ...
## $ p36      : Factor w/ 6 levels " Rentista "," Jubilado pensión ",...: 2
4            NA            NA            4            NA            3            1            NA            ...
## $ p37      : Factor w/ 2 levels " Si "," No ": NA NA NA 1 NA NA 1 NA N
A            NA            ...
## $ p38      : Factor w/ 9 levels " Liquidación empresa ",...: NA NA NA 3
NA            NA            5            NA            NA            NA            ...
## $ p381     : Factor w/ 2 levels " Si "," No ": NA NA NA 1 NA NA 2 NA NA
NA            ...
## $ p39      : Factor w/ 65 levels "1","10","100",...: NA NA NA 33 NA NA 1
2            NA            NA            NA            ...
## $ p40      : Factor w/ 314 levels "1010","1020",...: NA NA 207 207 283 N
A            274            NA            NA            311            ...
## $ p41      : Factor w/ 374 levels "110","1112","1113",...: NA NA 146 14
6            103            NA            56            NA            NA            344            ...
## $ p42      : Factor w/ 10 levels " Empleado de gobierno ",...: NA NA 2 2
6            NA            1            NA            NA            10            ...
## $ p42a     : Factor w/ 2 levels " Si"," No ": NA NA NA NA 2 NA NA NA N
A            NA            ...
## $ p43      : Factor w/ 6 levels "Con nombramiento ",...: NA NA 2 NA NA N
A            NA            NA            NA            3            ...
## $ p44a     : Factor w/ 2 levels " Si "," No ": NA NA 2 NA NA NA NA NA N

```

A			1	...
##	\$ p44b	: Factor w/ 2 levels	" Si ", " No ": NA NA 2 NA NA NA NA NA N	
A			1	...
##	\$ p44c	: Factor w/ 2 levels	" Si ", " No ": NA NA 2 NA NA NA NA NA N	
A			2	...
##	\$ p44d	: Factor w/ 2 levels	" Si ", " No ": NA NA 1 NA NA NA NA NA N	
A			2	...
##	\$ p44e	: Factor w/ 2 levels	" Si ", " No ": NA NA 1 NA NA NA NA NA N	
A			2	...
##	\$ p44f	: Factor w/ 2 levels	" Si ", " No ": NA NA 1 NA NA NA NA NA N	
A			2	...
##	\$ p44g	: Factor w/ 2 levels	" Si ", " No ": NA NA 2 NA NA NA NA NA N	
A			2	...
##	\$ p44h	: Factor w/ 2 levels	" Si ", " No ": NA NA 2 NA NA NA NA NA N	
A			2	...
##	\$ p44i	: Factor w/ 2 levels	" Si ", " No ": NA NA 2 NA NA NA NA NA N	
A			2	...
##	\$ p44j	: Factor w/ 2 levels	" Si ", " No ": NA NA 1 NA NA NA NA NA N	
A			2	...
##	\$ p44k	: Factor w/ 2 levels	" Si ", " No ": NA NA 1 NA NA NA NA NA N	
A			2	...
##	\$ p45	: Factor w/ 81 levels	"0","1","10","11",...: NA NA 6 NA 13 N	
A	NA	NA	NA	47
##	\$ p46	: Factor w/ 12 levels	" Local patrono ",...: NA NA 1 NA 9 NA	
NA	NA	NA	9	...
##	\$ p47a	: Factor w/ 2 levels	" Menos de 100 ",...: NA NA 2 NA 1 NA N	
A	NA	NA	1	...
##	\$ p47b	: num	NA NA NA NA 1 NA NA NA NA 1	...
##	\$ p48	: Factor w/ 4 levels	" Registros contables ",...: NA NA NA N	
A	3	NA	NA	NA
##	\$ p49	: Factor w/ 3 levels	" Si ", " No ",...: NA NA NA NA 1 NA NA N	
A	NA	NA	NA	...
##	\$ p50	: Factor w/ 2 levels	" Uno ", " Más de uno ": NA NA 1 NA 1 NA	
NA	NA	NA	1	...
##	\$ p51a	: Factor w/ 82 levels	"1","10","102",...: NA NA 36 NA 36 NA N	
A	NA	NA	36	...
##	\$ p51b	: Factor w/ 27 levels	"1","10","11",...: NA NA NA NA NA NA N	
A	NA	NA	NA	...
##	\$ p51c	: Factor w/ 7 levels	"0","10","2","4",...: NA NA NA NA NA NA	
NA	NA	NA	NA	...
##	\$ p52	: Factor w/ 117 levels	"1010","1020",...: NA NA NA NA NA NA N	
A	NA	NA	NA	...
##	\$ p53	: Factor w/ 125 levels	"1321","1323",...: NA NA NA NA NA NA N	
A	NA	NA	NA	...
##	\$ p54	: Factor w/ 10 levels	" Empleado de gobierno ",...: NA NA NA	
NA	NA	NA	NA	NA
##	\$ p54a	: Factor w/ 2 levels	" Si ", " No ": NA NA NA NA NA NA NA	
NA	NA	NA	NA	...
##	\$ p55	: Factor w/ 12 levels	" Local patrono ",...: NA NA NA NA NA N	
A	NA	NA	NA	NA
##	\$ p56a	: Factor w/ 2 levels	" Menos de 100 ",...: NA NA NA NA NA NA	

```

NA      NA      NA      NA      ...
## $ p56b      : num      NA NA NA NA NA NA NA NA NA NA ...
## $ p57      : Factor w/ 4 levels " Registros contables ",...: NA NA NA N
A      NA      NA      NA      NA      NA      NA      ...
## $ p58      : Factor w/ 3 levels " Si "," No ",...: NA NA NA NA NA NA NA
NA      NA      NA      NA      ...
## $ p59      : Factor w/ 5 levels " Contento"," Poco contenido ",...: NA N
A      1      NA      2      NA      NA      NA      NA      1      ...
## $ p60a     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 1 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
## $ p60b     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 2 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
## $ p60c     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 2 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
## $ p60d     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 2 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
## $ p60e     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 1 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
## $ p60f     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 2 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
## $ p60g     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 2 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
## $ p60h     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 2 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
## $ p60i     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 2 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
## $ p60j     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 2 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
## $ p60k     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 2 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
## $ p61b1    : Factor w/ 6 levels " IESS General ",...: 5 2 1 5 2 5 5 5 5
5      ...
## $ p63      : num      NA NA NA NA 0 NA NA NA NA NA NA ...
## $ p64a     : Factor w/ 2 levels " Si "," No ": NA NA NA NA 2 NA NA NA N
A      NA      NA      NA      NA      NA      NA      NA      NA      ...
##      [list      output      truncated]
## - attr(*, "variable.labels")= Named chr      "Area" "Unidad primaria de mu
estreo"      "Persona"      "Sexo"      ...
##      ..- attr(*, "names")= chr      "area" "upm" "p01" "p02" ...
## - attr(*, "codepage")= int 65001

#      seleccionar      las      variables      mas      relevantes
nmdu <- enemdu %>% select(c(area, p01, p02, p03, p04, p05a, p06, p07, p10a
,      p11,      p12a,      p15,
p20, p21, p23, p231, p24, p36, p42, p43, p45, p
50,      p51a,      p61b1,
p71b, p72b, p73b, p74b, p75, p76, p77, p78, ing
r1,      sd03,      conduct,
empleo, desempleo, secemp, grupo1, rama1, nnivi
ns))

```

```
#      filtrar      por      los      jefes      de      hogar
nmdujh      <-      nmdu      %>%      filter(p01=="01")

#      procesamiento-----
#      transformar a numericas las variables que corresponden
nmdujh$p24      <-      as.numeric(nmdujh$p24)
nmdujh$p45      <-      as.numeric(nmdujh$p45)
nmdujh$p51a      <-      as.numeric(nmdujh$p51a)

nmdujh$p71b[nmdujh$p71b      ==      999999]      <-      NA
nmdujh$p72b[nmdujh$p72b      ==      999999]      <-      NA
nmdujh$p73b[nmdujh$p73b      ==      999999]      <-      NA
nmdujh$p76      <-      as.character(nmdujh$p76)
nmdujh$p76      <-      as.numeric(nmdujh$p76)
nmdujh$p78      <-      as.character(nmdujh$p78)
nmdujh$p78      <-      as.numeric(nmdujh$p78)

###      recategorizar      la      edad      en      grupos----
nmdujh$p03      <-      as.character(nmdujh$p03)
nmdujh$p03[nmdujh$p03      ==      "98      y      más"]      <-      NA
nmdujh$p03 <- as.numeric(nmdujh$p03)

nmdujh$p03_rec <- cut(nmdujh$p03, breaks = c(15, 24, 39, 64, 97), labels =
c("18-24", "25-39", "40-64", "65      y      más"))

###      ingreso      Laboral
# transformarLa a numerica y eliminar valores centinelas 999 , no informa,
nmdujh$ingrl_rec      <-      as.character(nmdujh$ingrl)
nmdujh$ingrl_rec[nmdujh$ingrl      ==      "No      informa"]      <-      NA
nmdujh$ingrl_rec[nmdujh$ingrl      ==      "999"]      <-      NA
nmdujh$ingrl_rec[nmdujh$ingrl      ==      "Gasta mas de lo que gana"] <- NA
nmdujh$ingrl_rec      <-      as.numeric(nmdujh$ingrl_rec)

####      ingreso      total      sumar      los      distintos      ingresos----
nmdujh$ingrl_tot <- colSums(nmdujh [ ,c(43, 25:28, 30, 32)], na.rm = F)

ene <- read.csv("enemdu_proc2.csv", sep=";", encoding = "UTF-8")
summary(ene$ingrl_tot)

##      Min.      1st      Qu.      Median      Mean      3rd      Qu.      Max.
##      0      150      330      489      600      13520

ene$ingrl_tot[ene$ingrl_tot      ==      0]      <-      NA

###      renombrar      Las      variables
names(ene) <- c("area", "persona", "Sexo", "Edad", "RelacionParentezco", "
Seguro_social",
```

```

"Estado_civil", "Asiste_clases", "Nivelinstruccion", "Sabe_
leer_escribir",
"Obtuvo_titulo", "Autoidentificacion", "Trabja", "Actividad
_hogar", "Razon_notrabajo",
"Razon_covid", "Horas_trabajo_semanal", "Condicion_inactivi
dad", "Categoria_ocupacion",
"Tipo_trabajo", "Años_trabajo", "Numero_trabajo", "Horas_tr
abajo_principal",
"Forma_Seguridad_social", "Ingreso_transacciones_capital", "
Ingresos_jubilacion",
"Donaciones", "Remesas", "Recibe_Bono_DDHH", "Monto_Bono_DD
HH", "Recibe_Bono_Discapacidad",
"Monto_Bono_Discapacidad", "IngresoLaboral", "Principal_mot
ivo_desempleo", "Condicion_actividad",
"Empleo", "Desempleo", "Sectoros_empleados", "Grupo_ocupaci
on", "Rama_actividad",
"Nivel_instruccion", "Categorias_edad", "Ingreso_laboral", "
Ingresos_totales")

# guardar los datos para usar en tableau

#write.csv(ene, "enemdu_procesada_n.csv")
#write.xlsx(ene, "enemdu_proc_n.xlsx")
#sacar variables no relevantes
ene2 <- ene %>% select(-c(2, 5, 15, 16))

### graficos básicos
#histogramas y #graficos de barras

grafica_por_tipo <- function(datos, var) {

  vector_var <- datos[,var]

  if (is.factor(vector_var)) {
    g <- datos %>% filter(!is.na(var)) %>%
      ggplot(aes_string(x = var, fill= var)) +
      geom_bar()+coord_flip()
  }

  else {
    g <- datos %>% filter(!is.na(var)) %>%
      ggplot(aes_string(x = var, fill = var)) +
      geom_histogram(color= "deeppink", fill = "cyan", bins = 30)
  }

  g + labs( fill = "", title = var) + # etiquetas de ejes y titulo

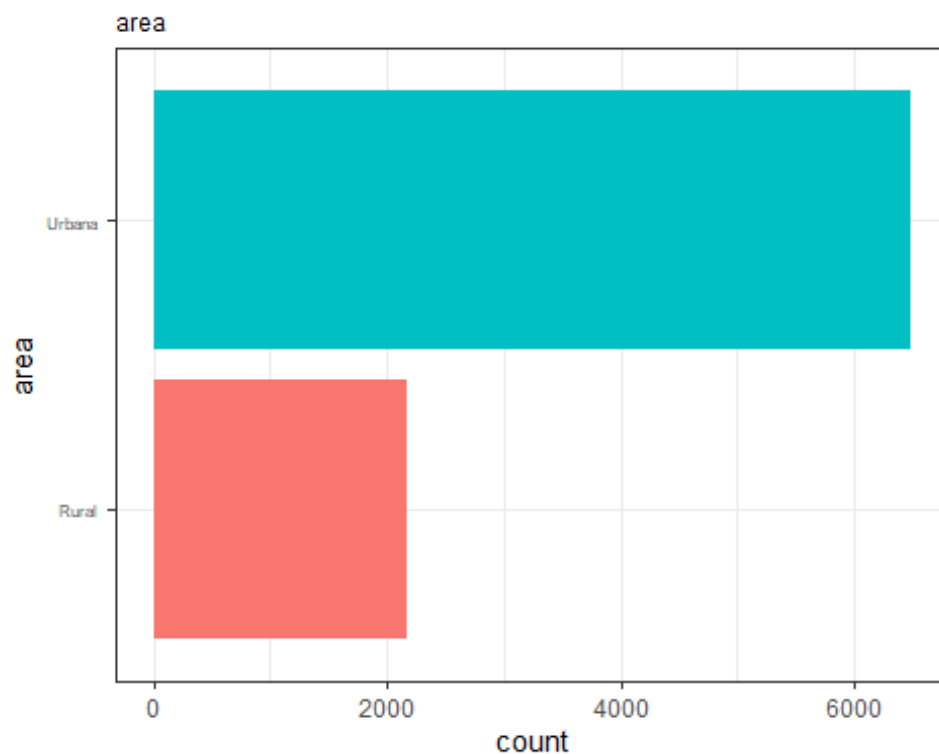
```

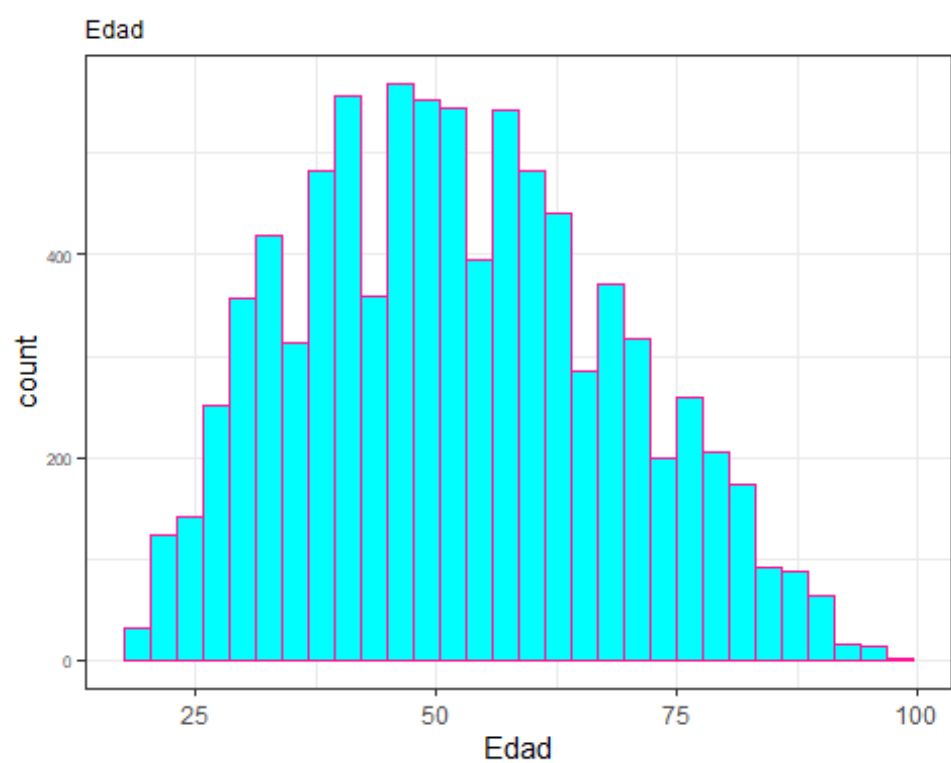
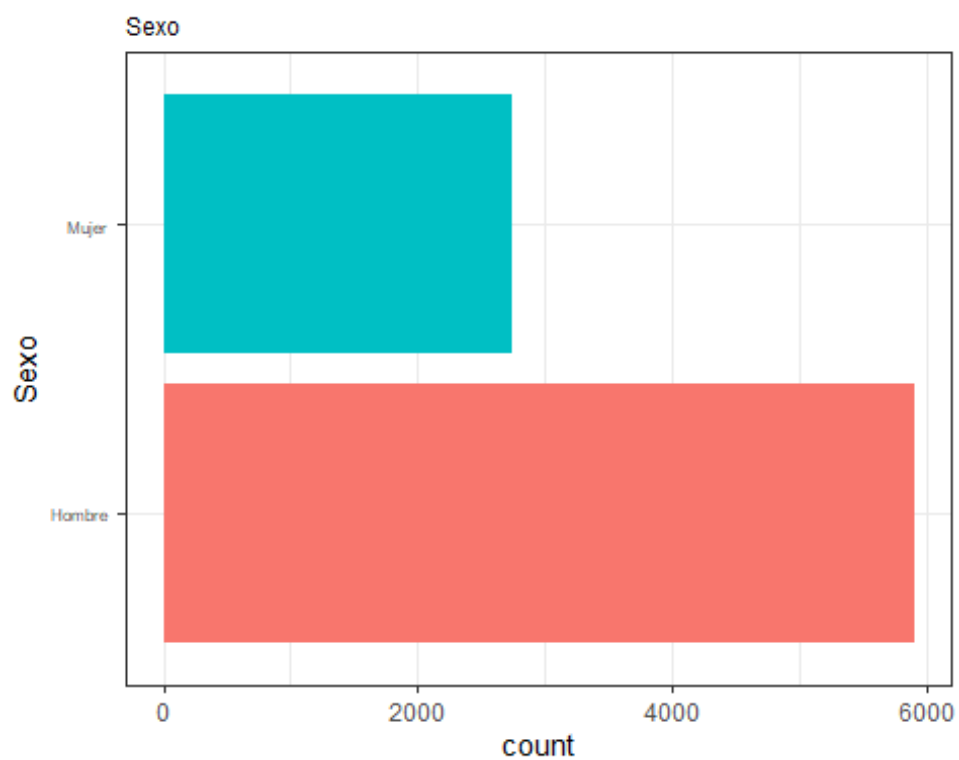
```
theme_bw() + theme(plot.title = element_text(size = 9),
  axis.text.y = element_text(size = 6), legend.position = "None")
}

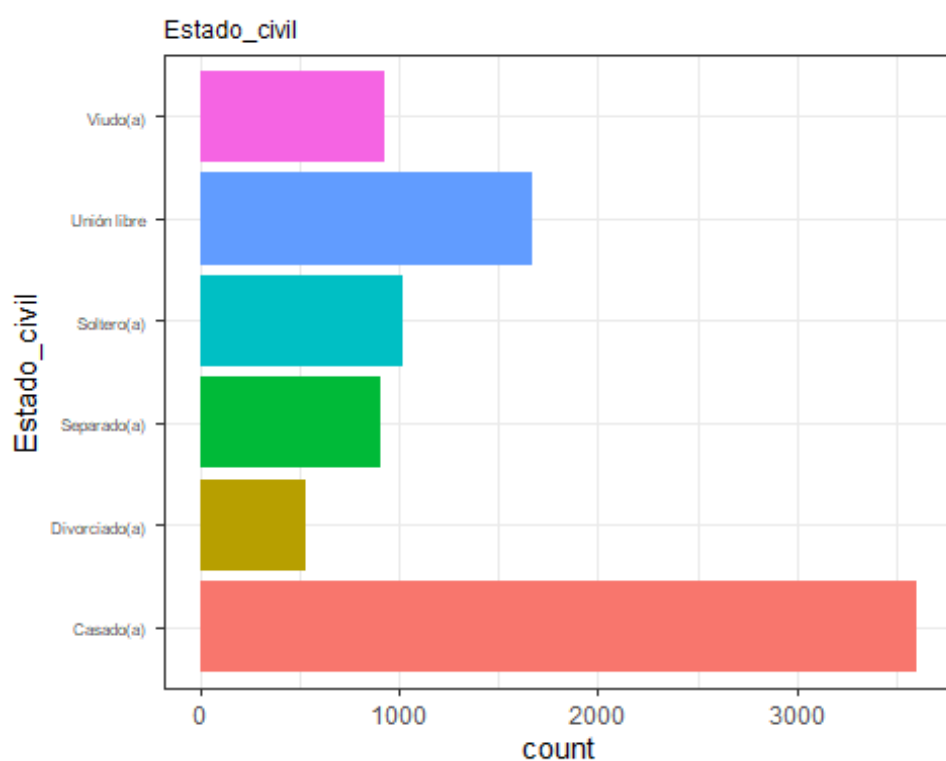
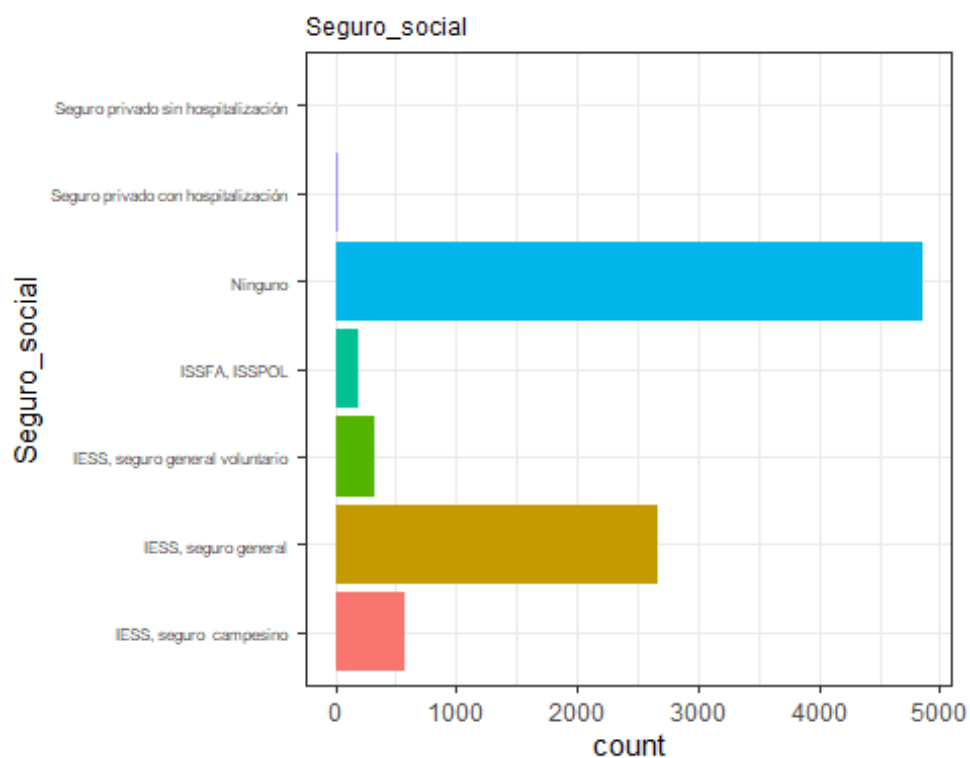
## pasar a double algunas variables que estaban como integer
ene2[c(3, 13, 17, 19, 21:24, 26, 28, 39, 40)] <- map(ene2[c(3, 13, 17, 19,
21:24, 26, 28, 39, 40)], ~as.numeric(.x))

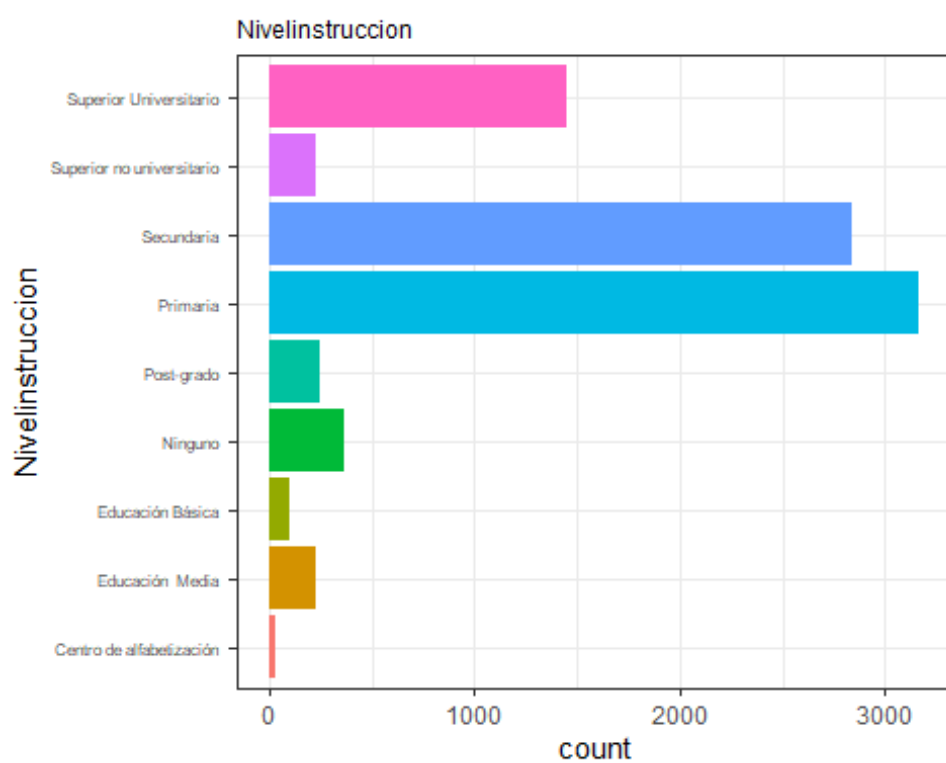
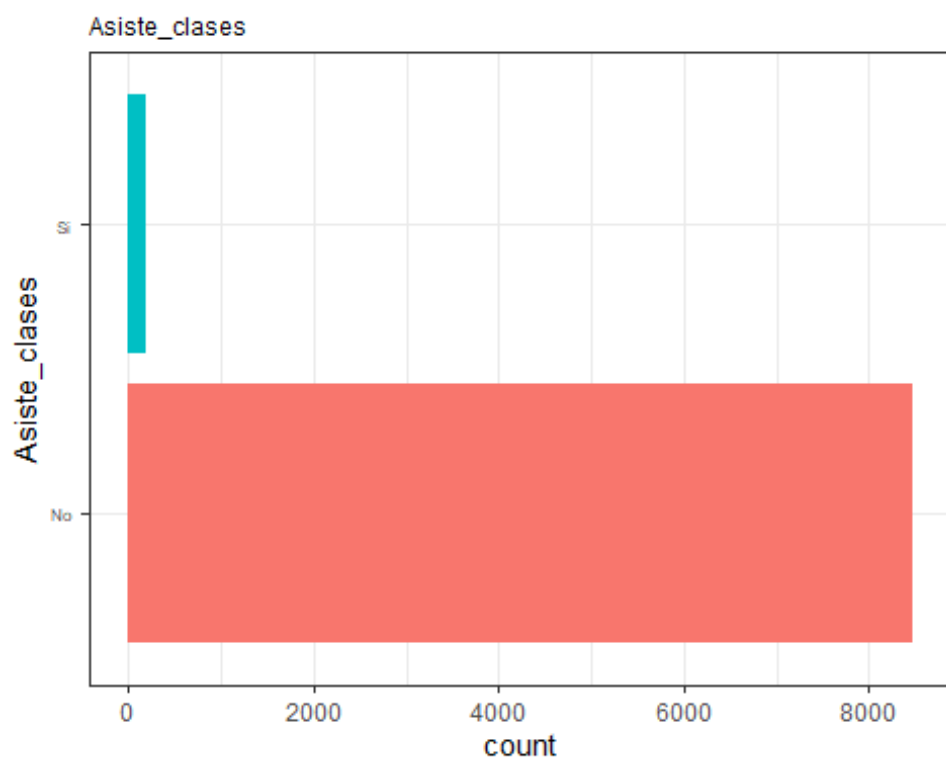
variables <- colnames(ene2) ## obtener los nombres de las variables

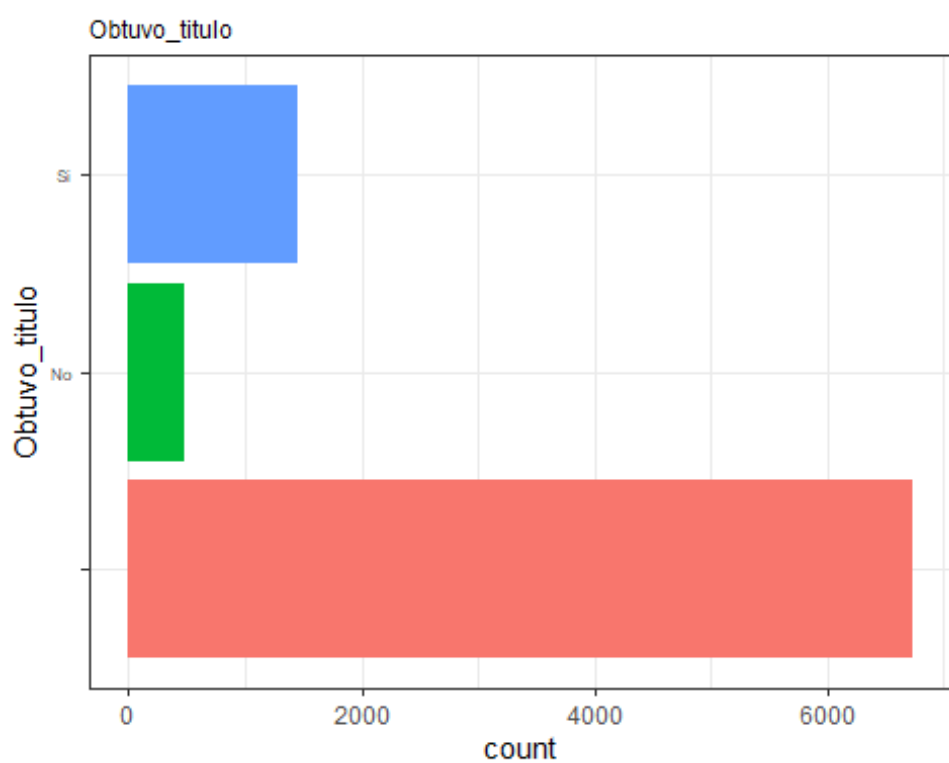
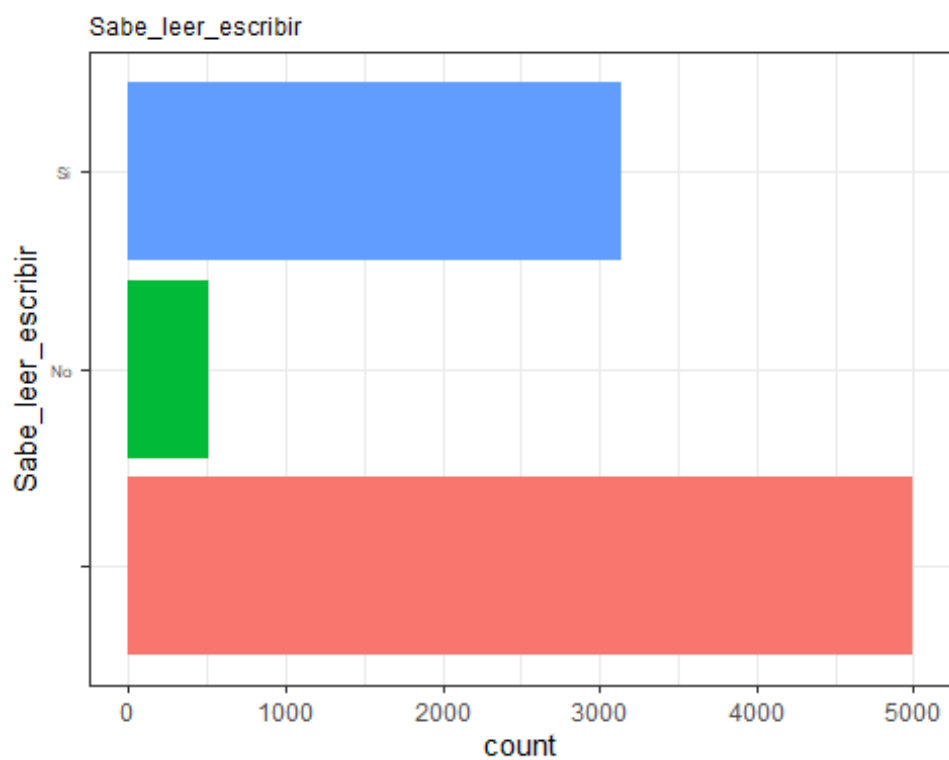
##aplicar la funcion de graficos para los graficos
lista_graficos <- map(variables, ~grafica_por_tipo(ene2, .x))
lista_graficos
```

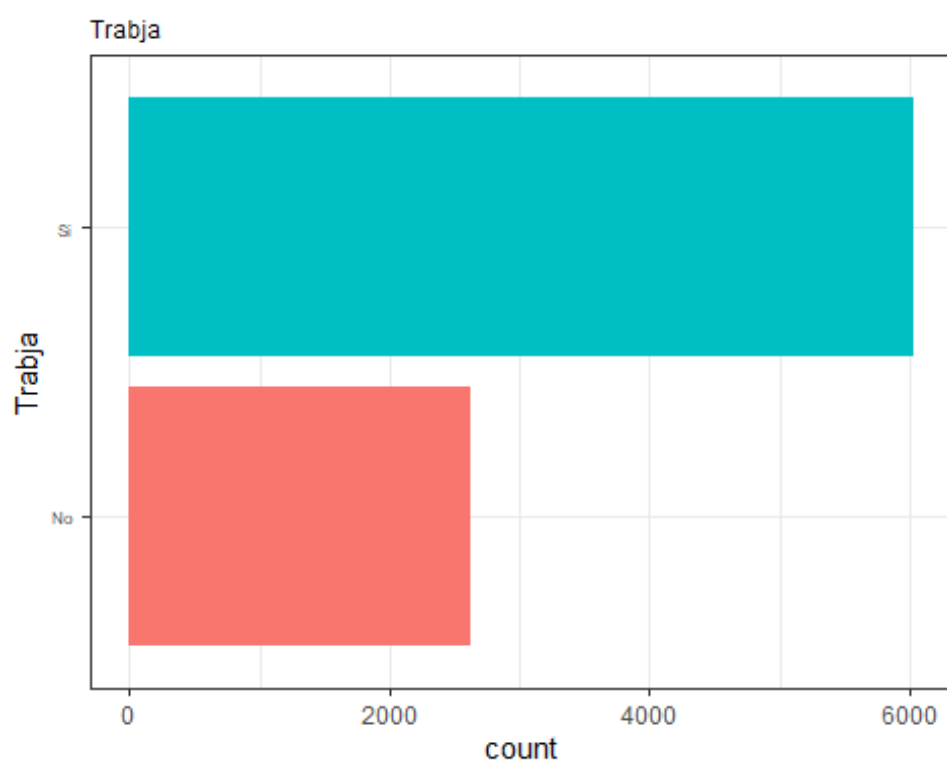
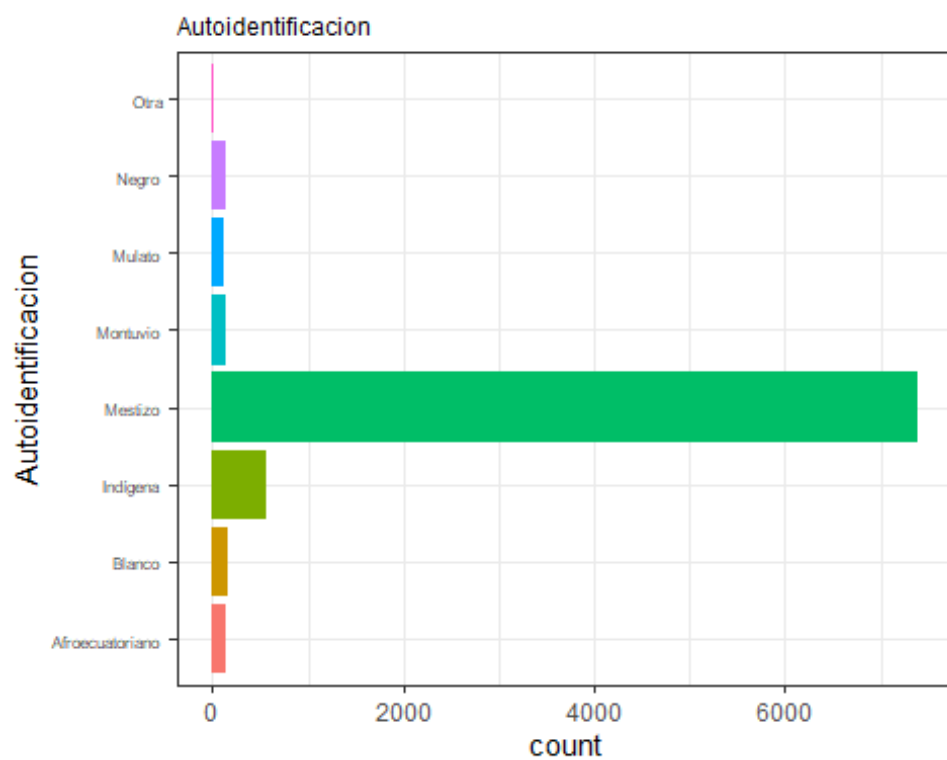


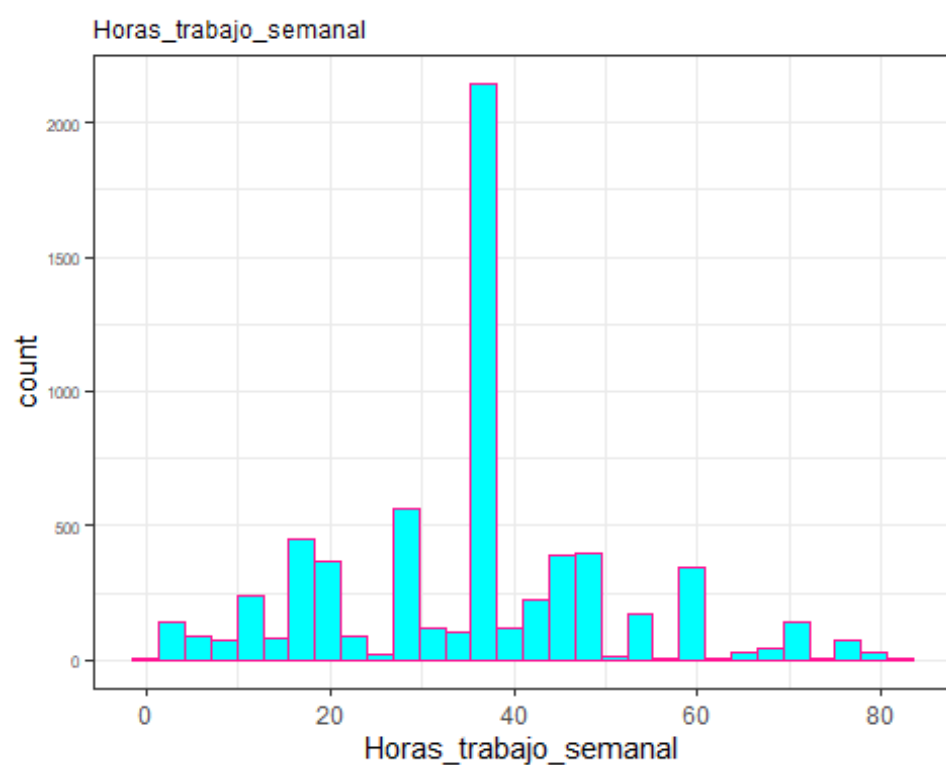
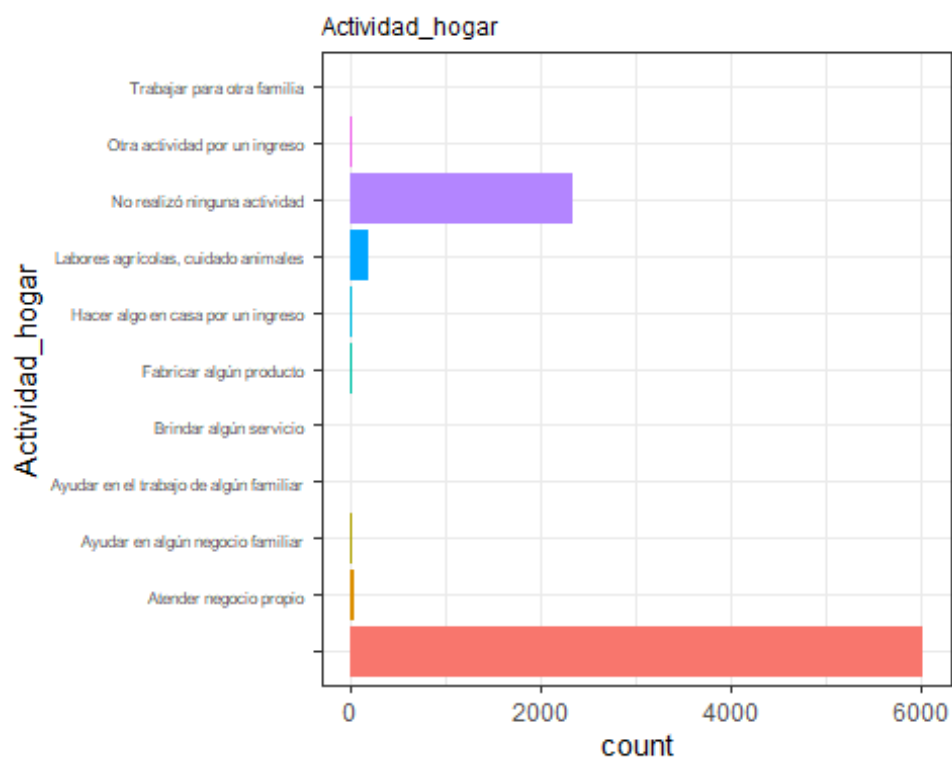


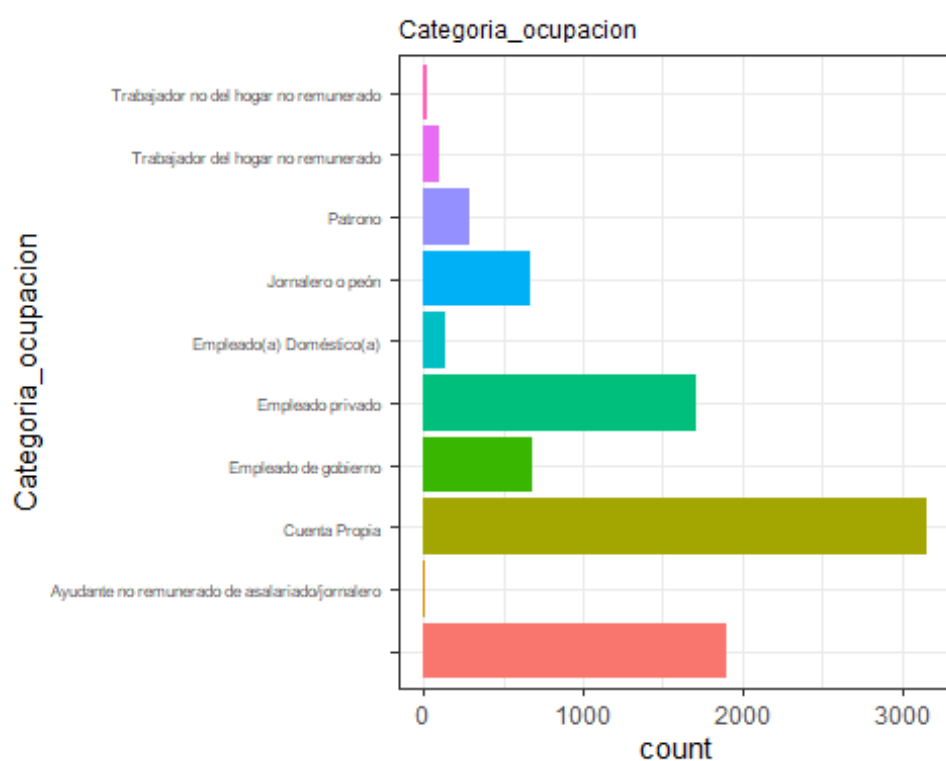
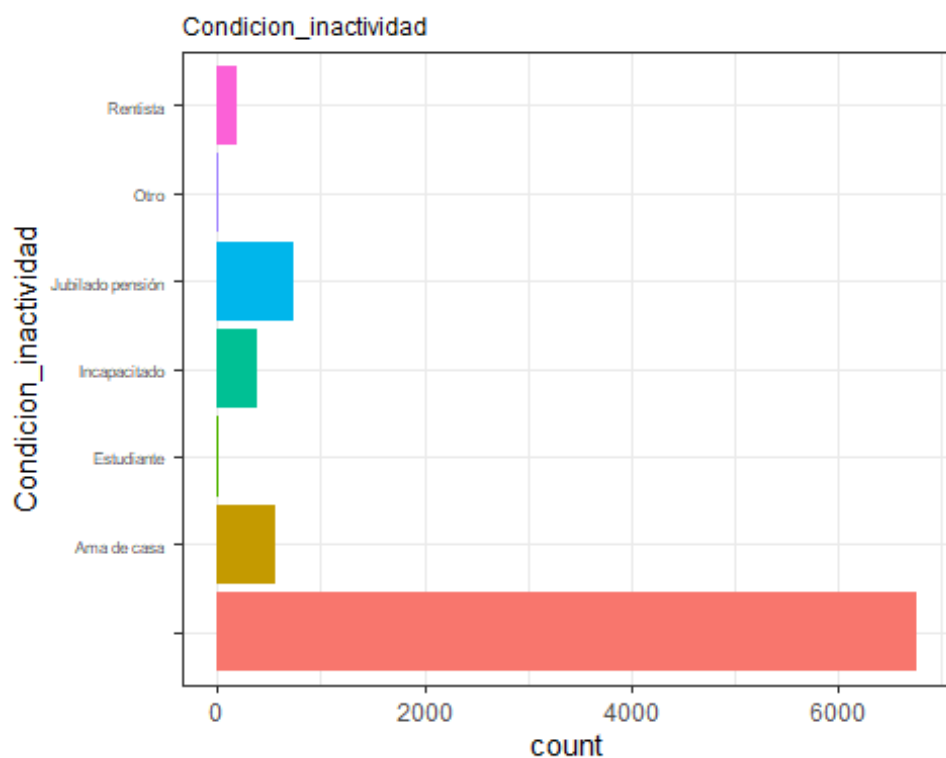


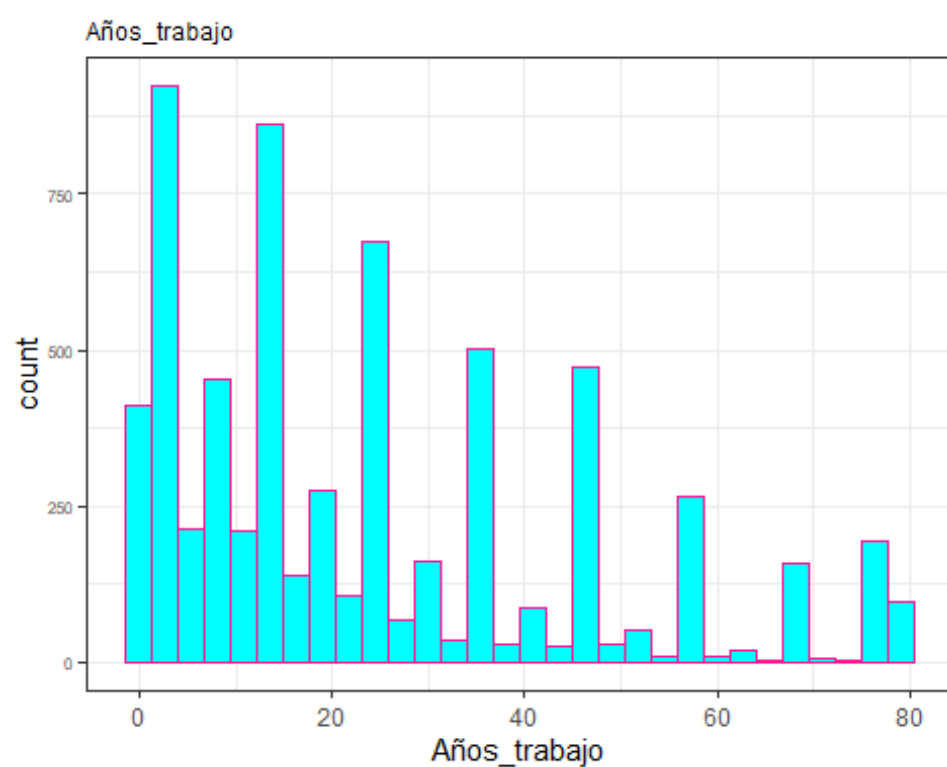
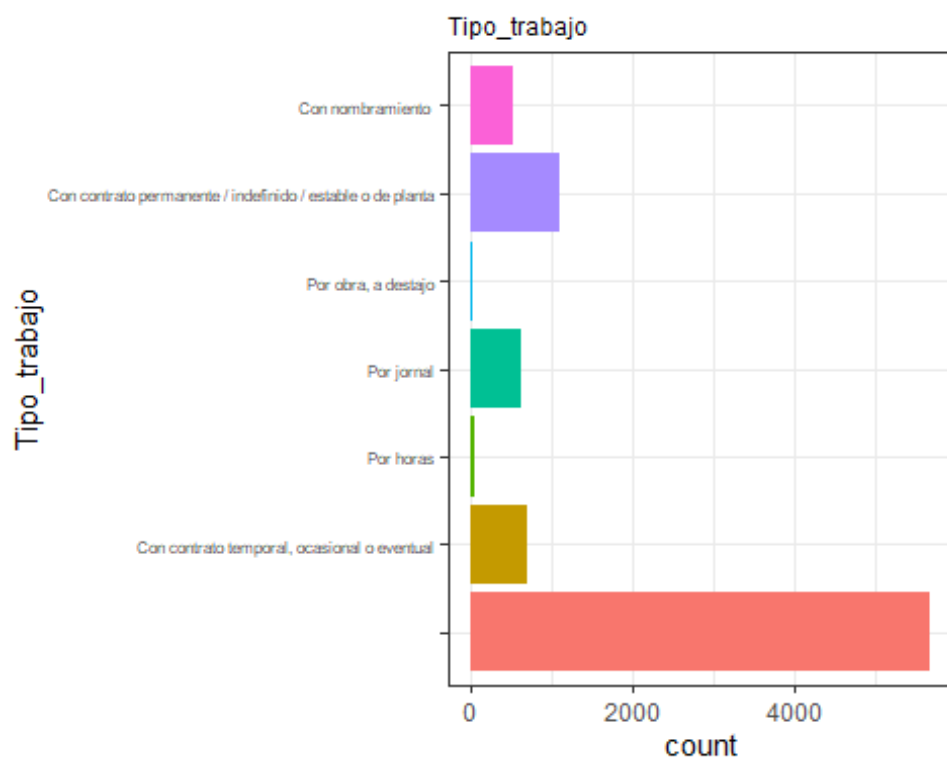


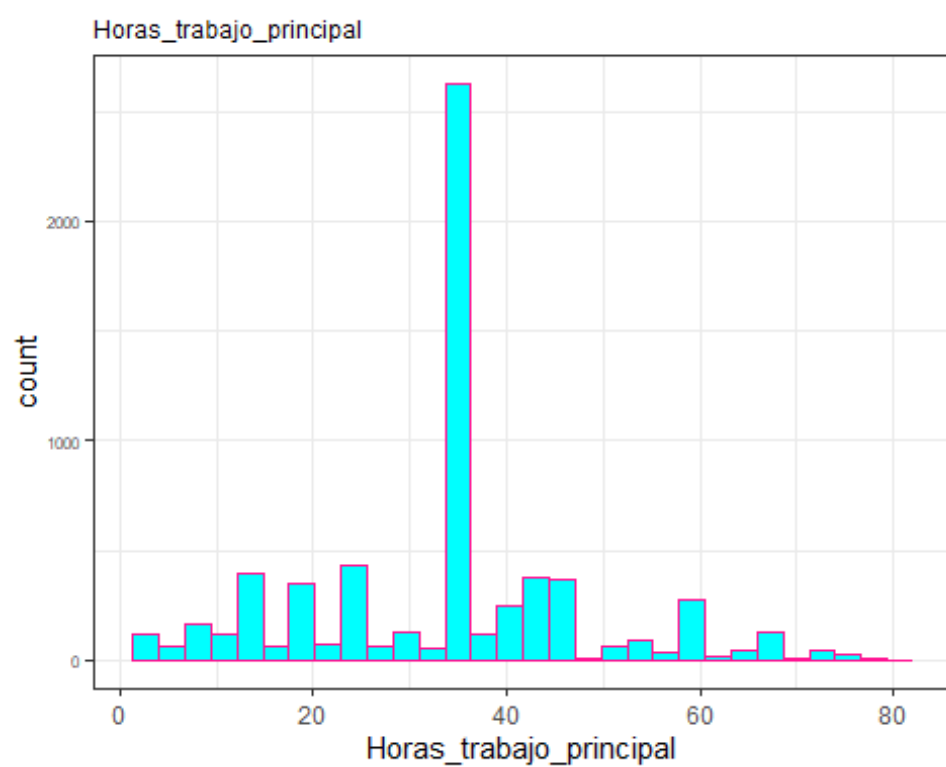
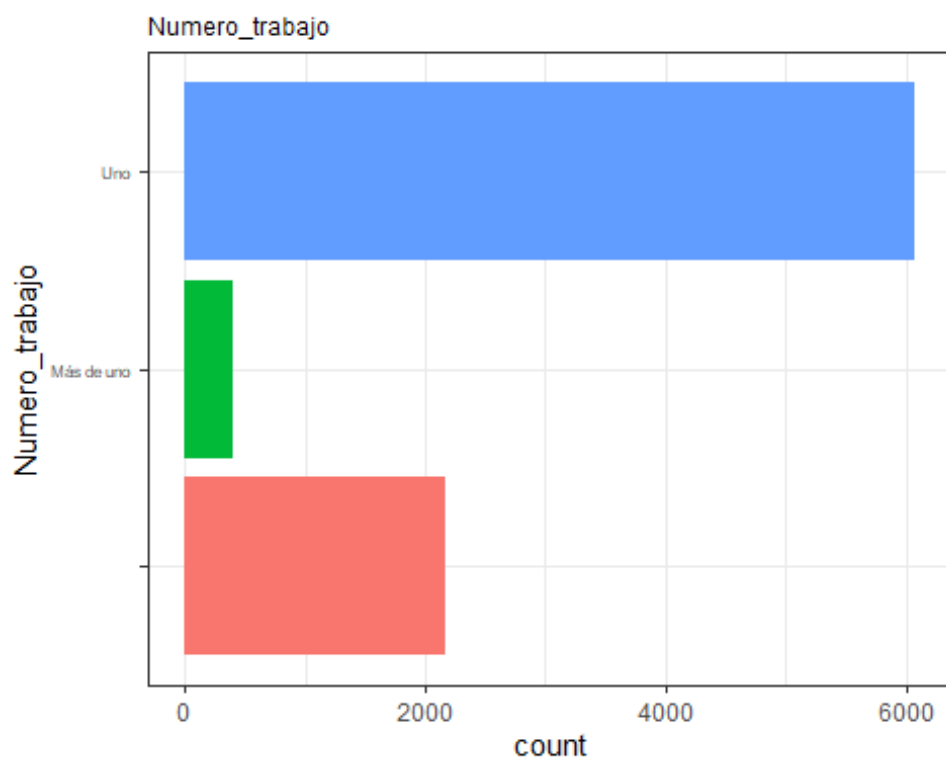


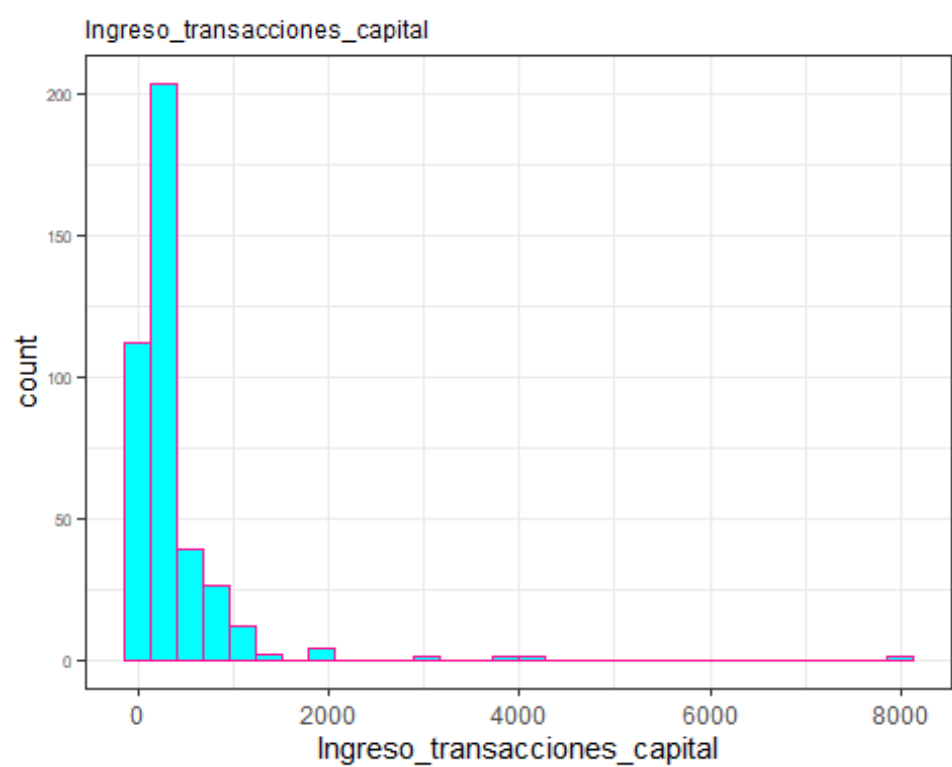
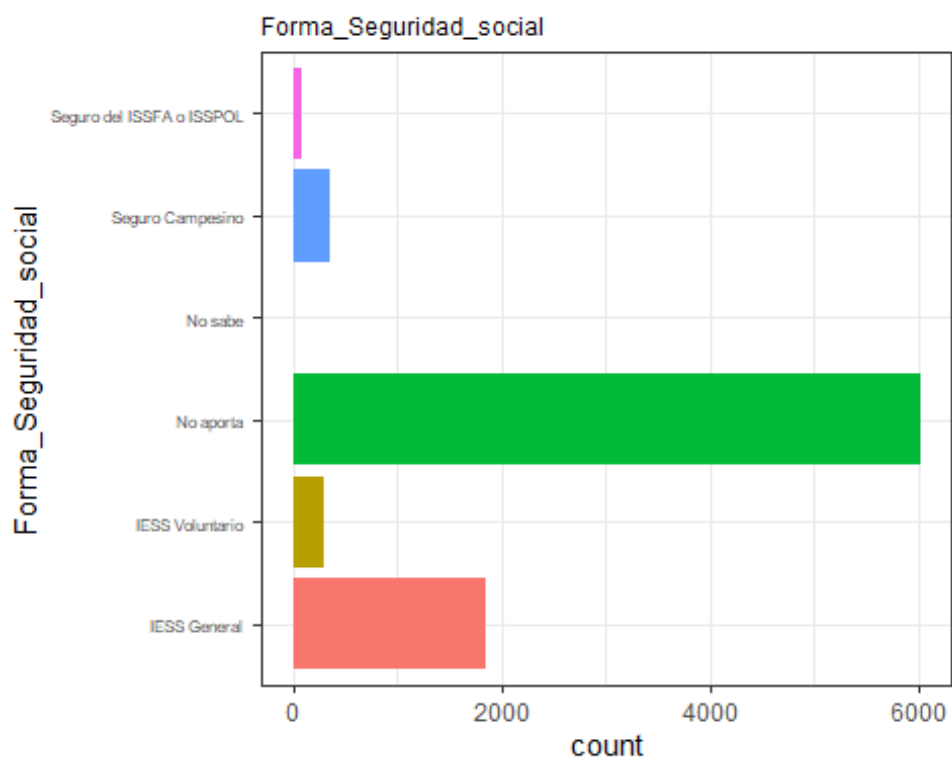


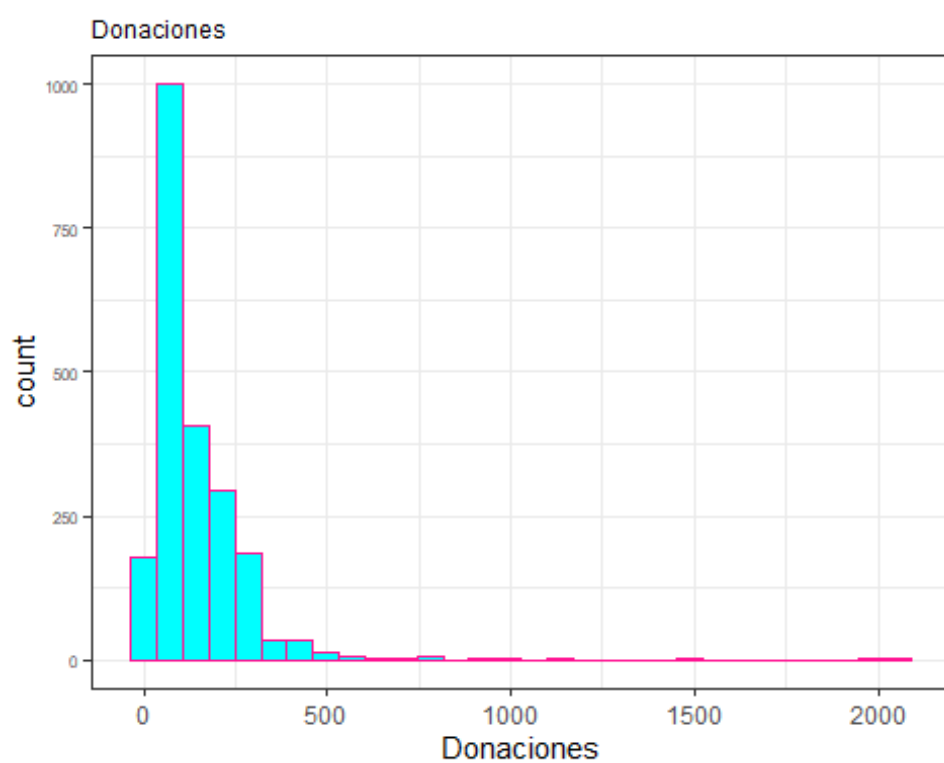
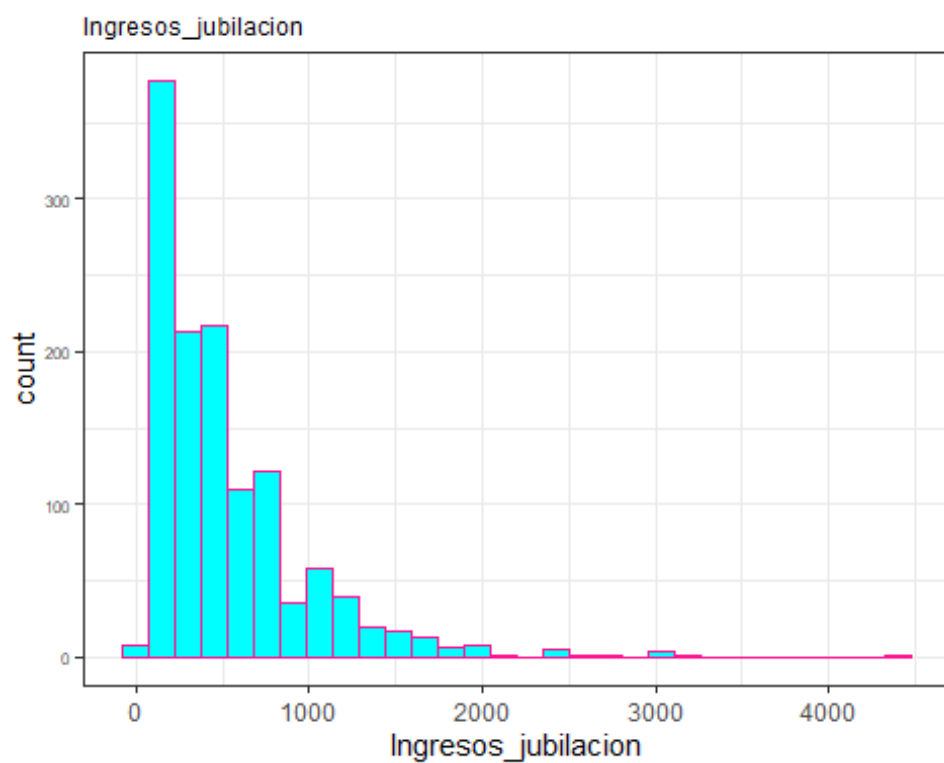


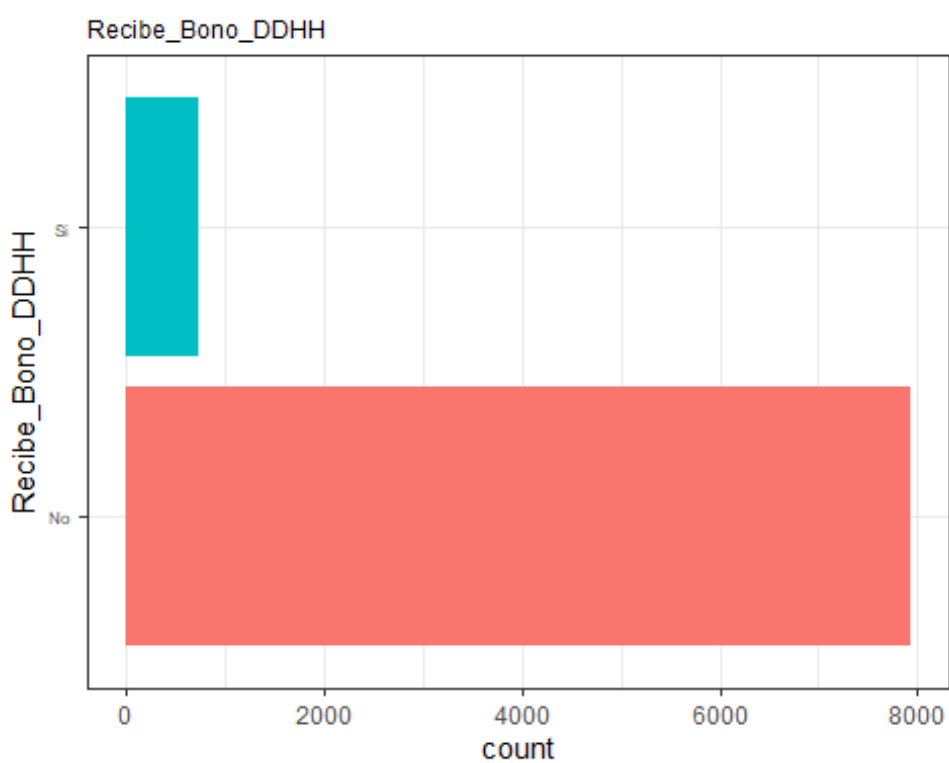
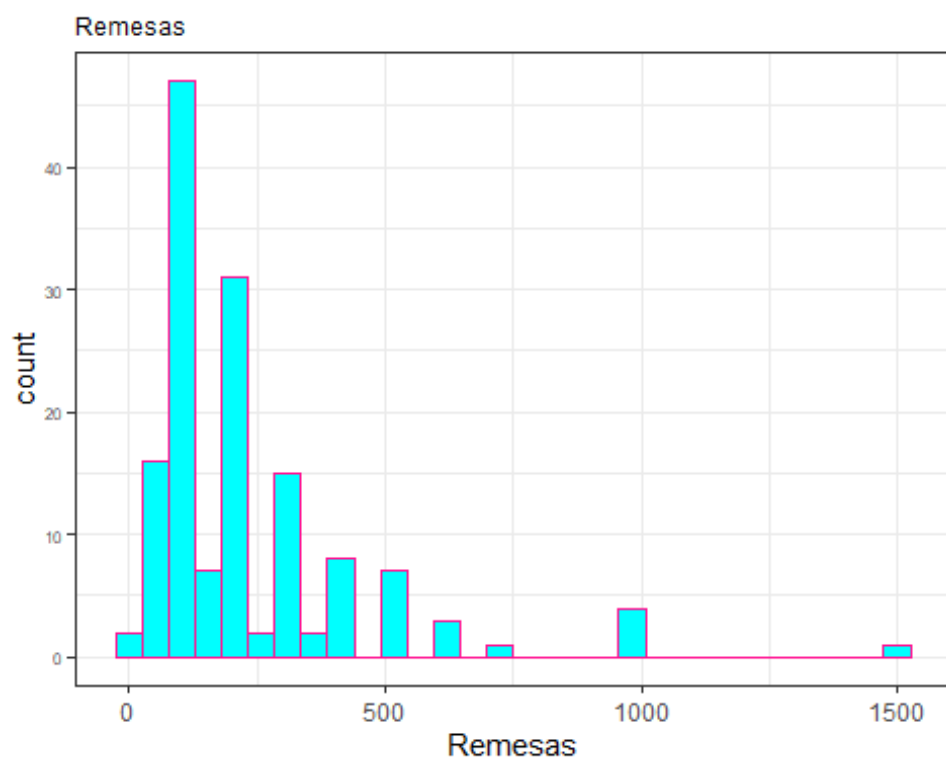


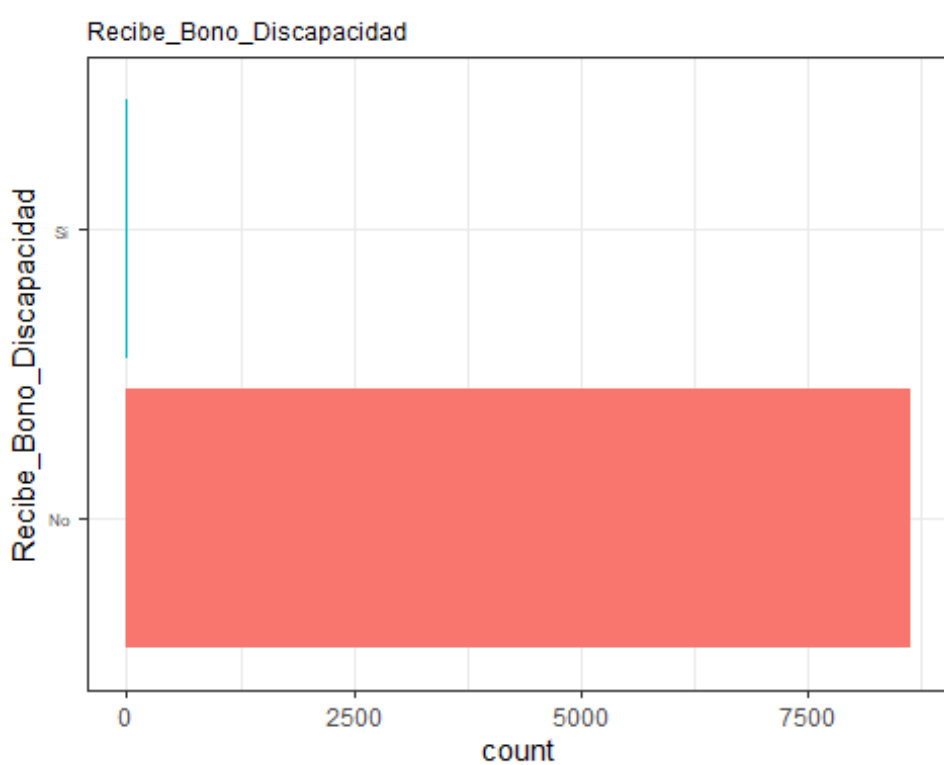
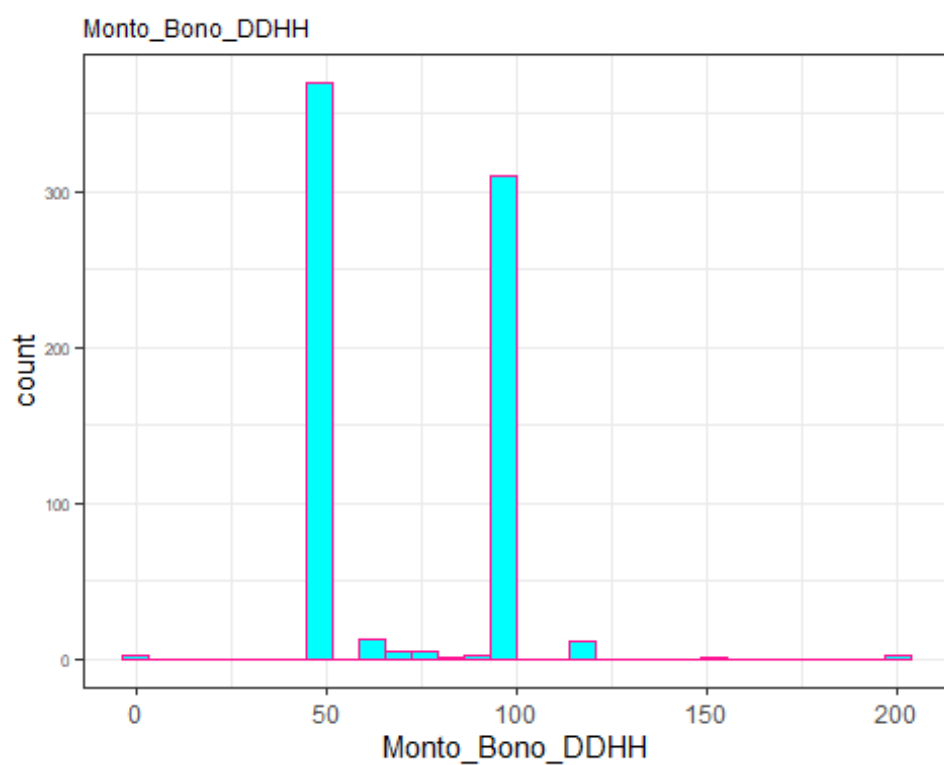


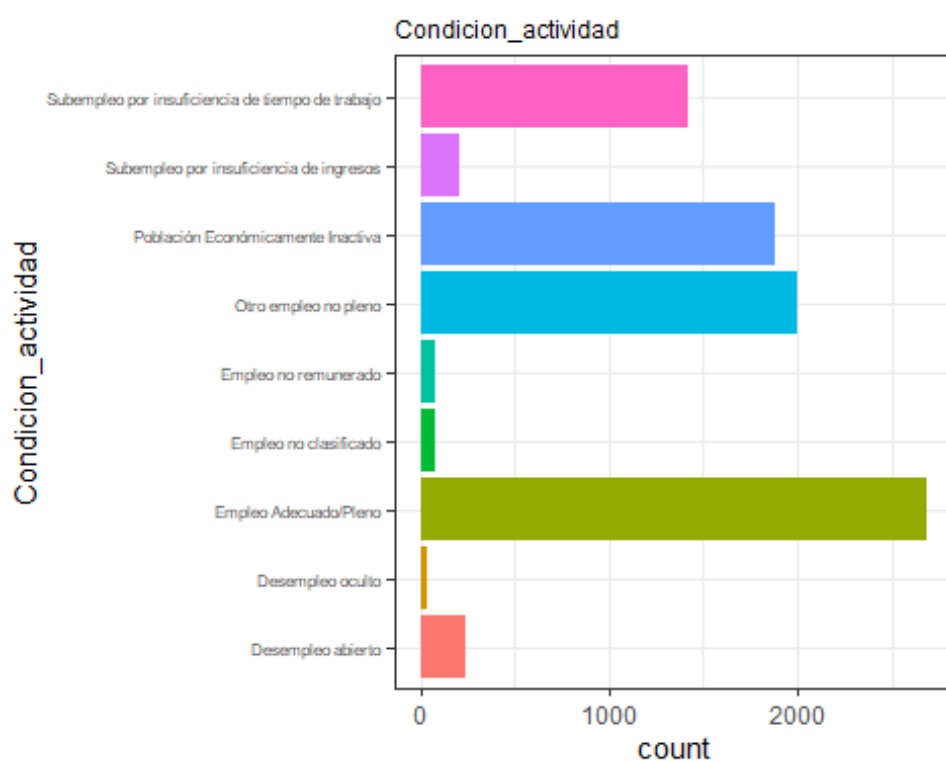
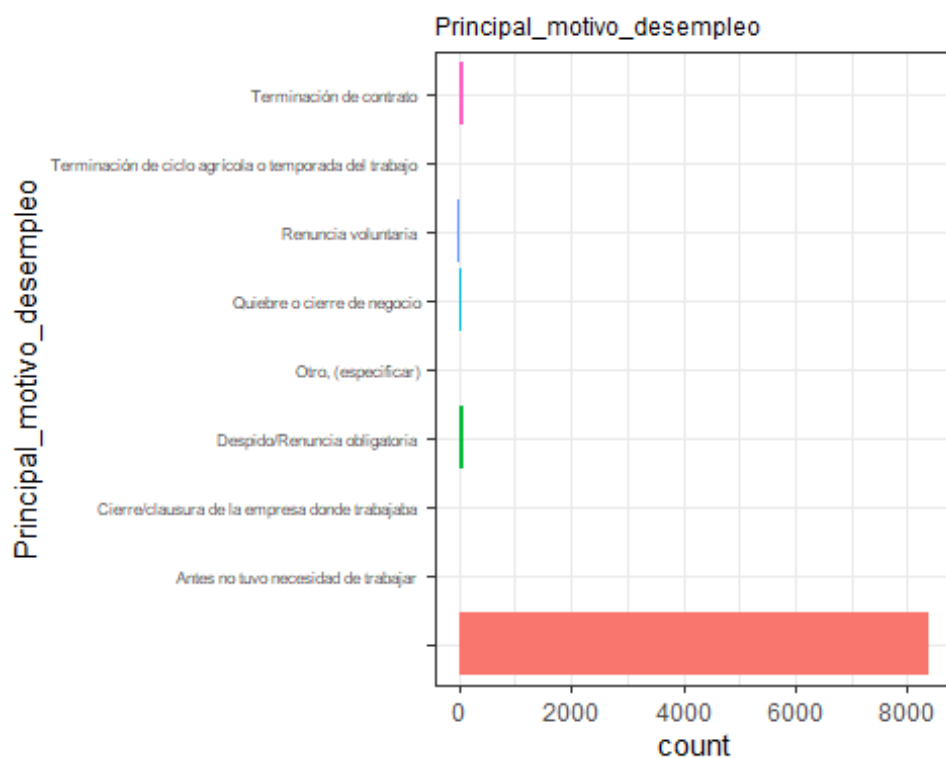


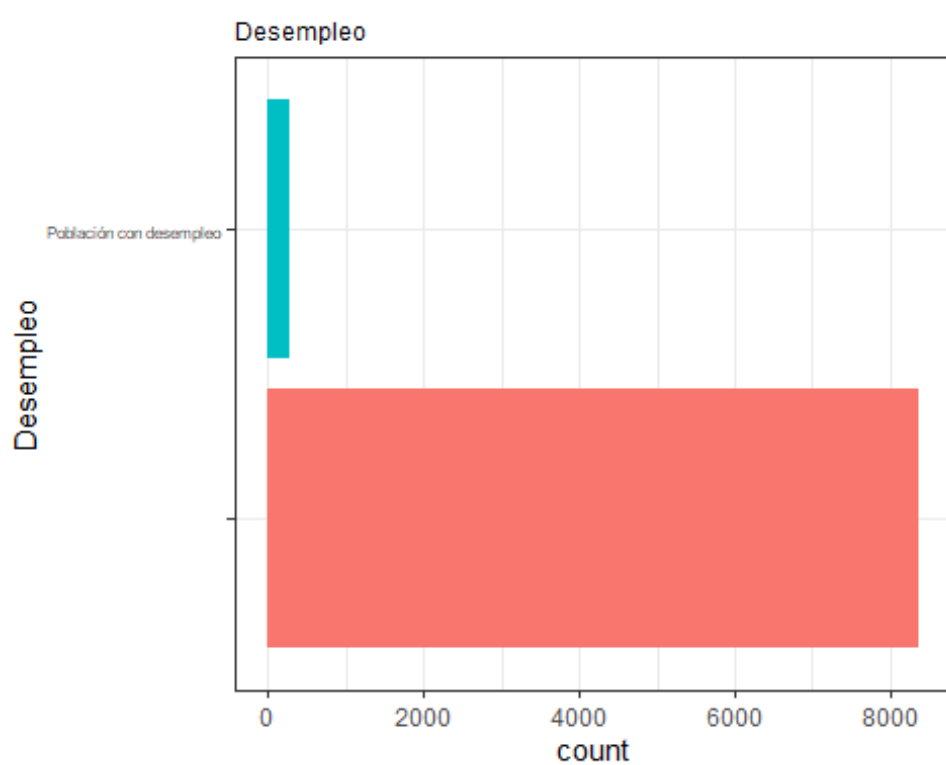
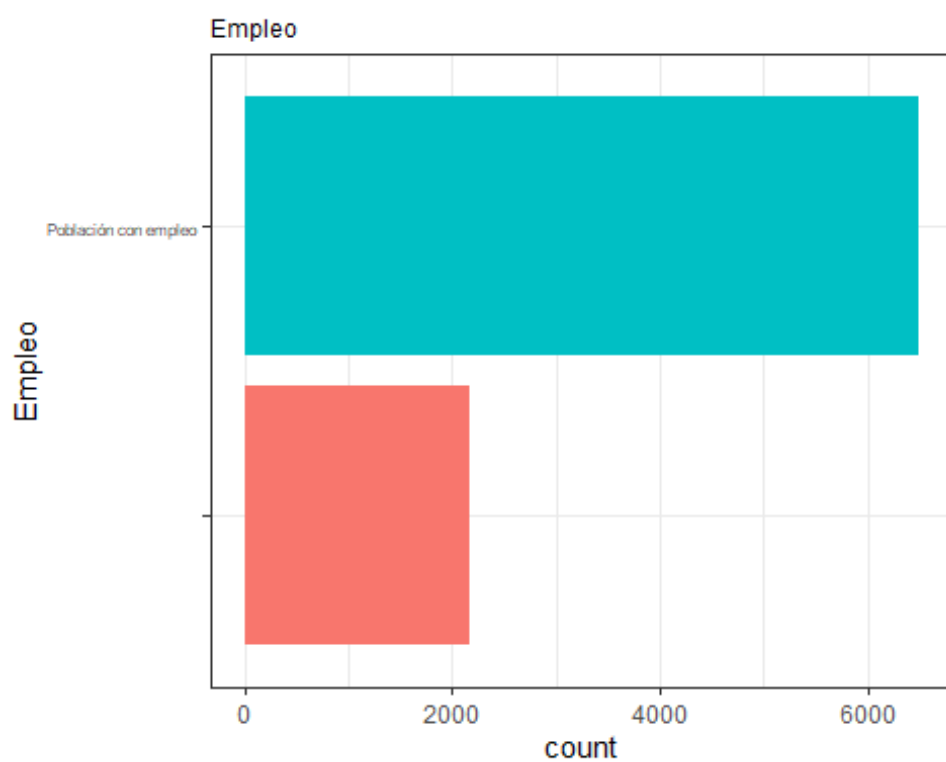


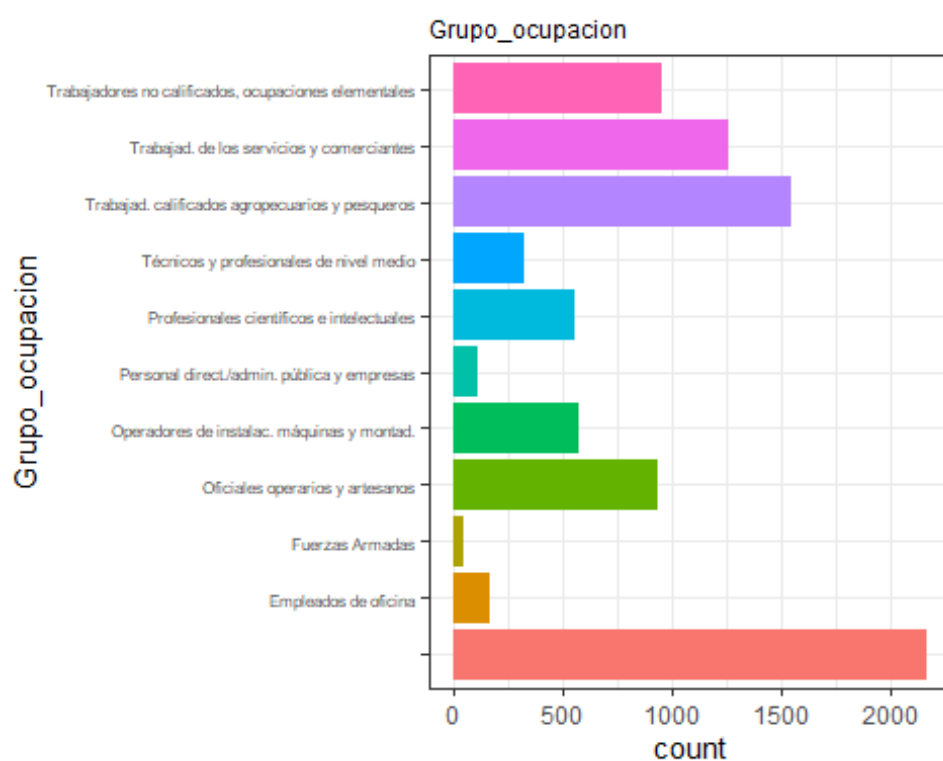
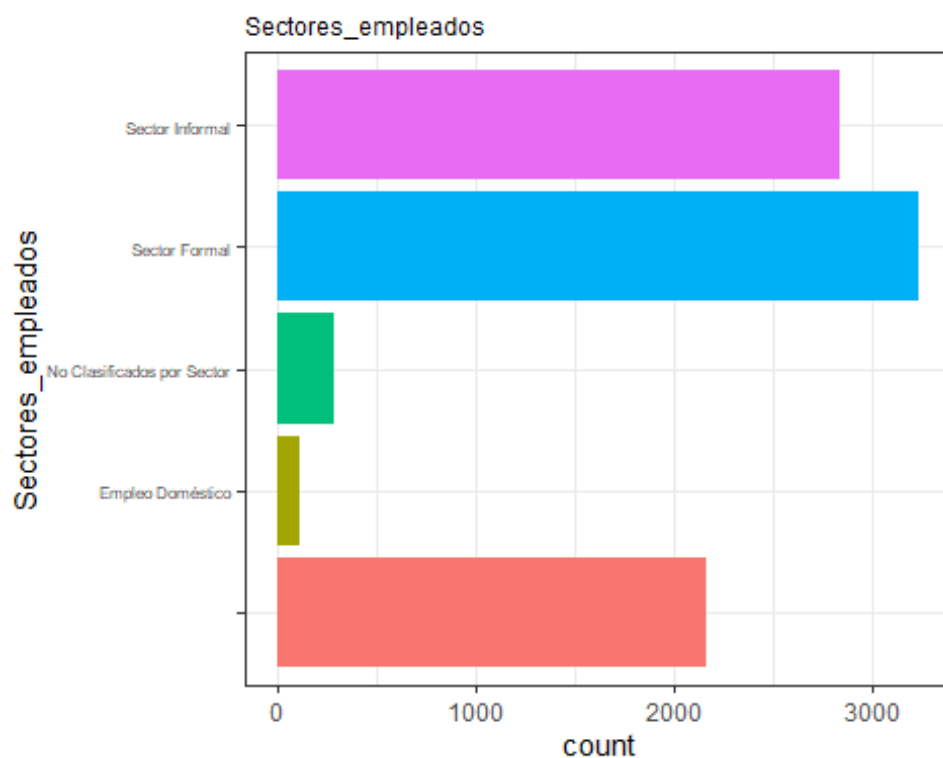


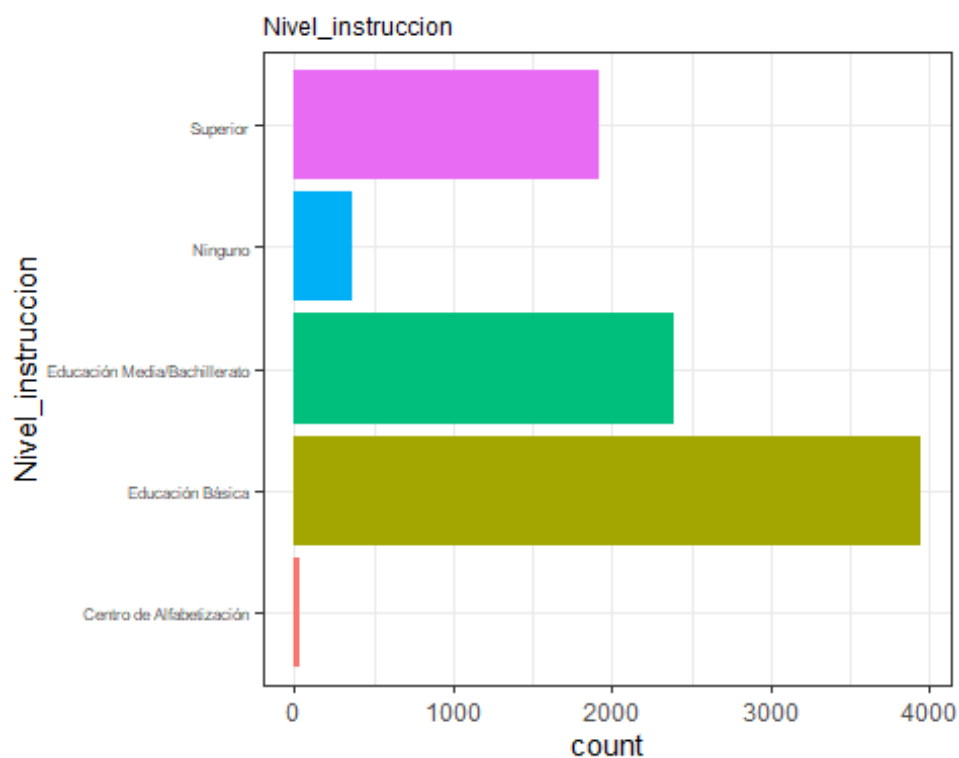
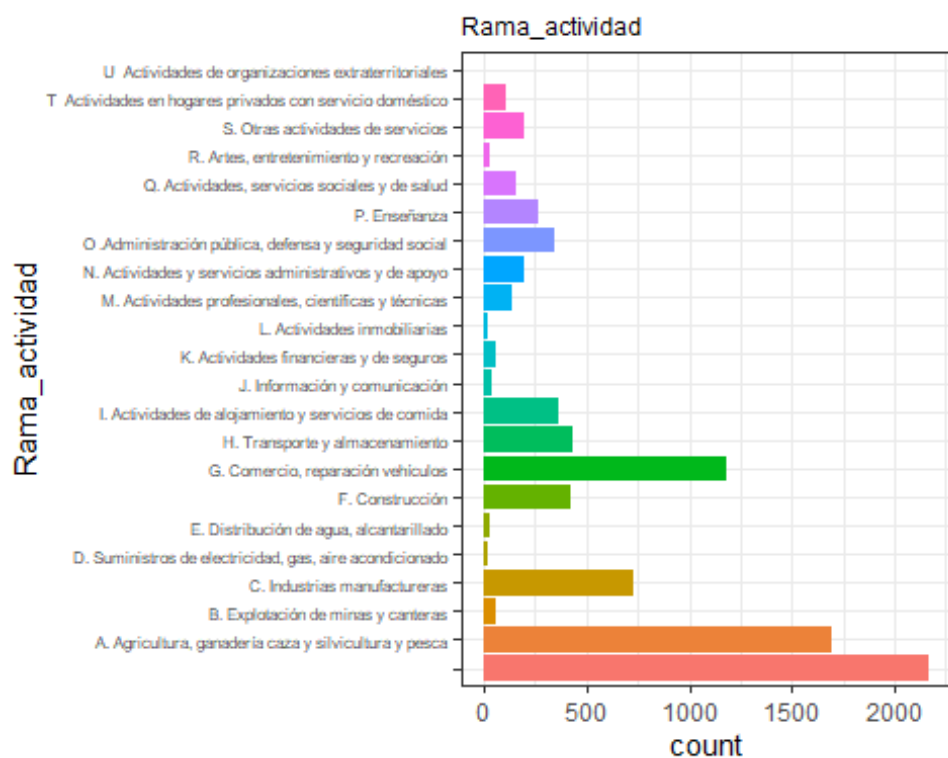


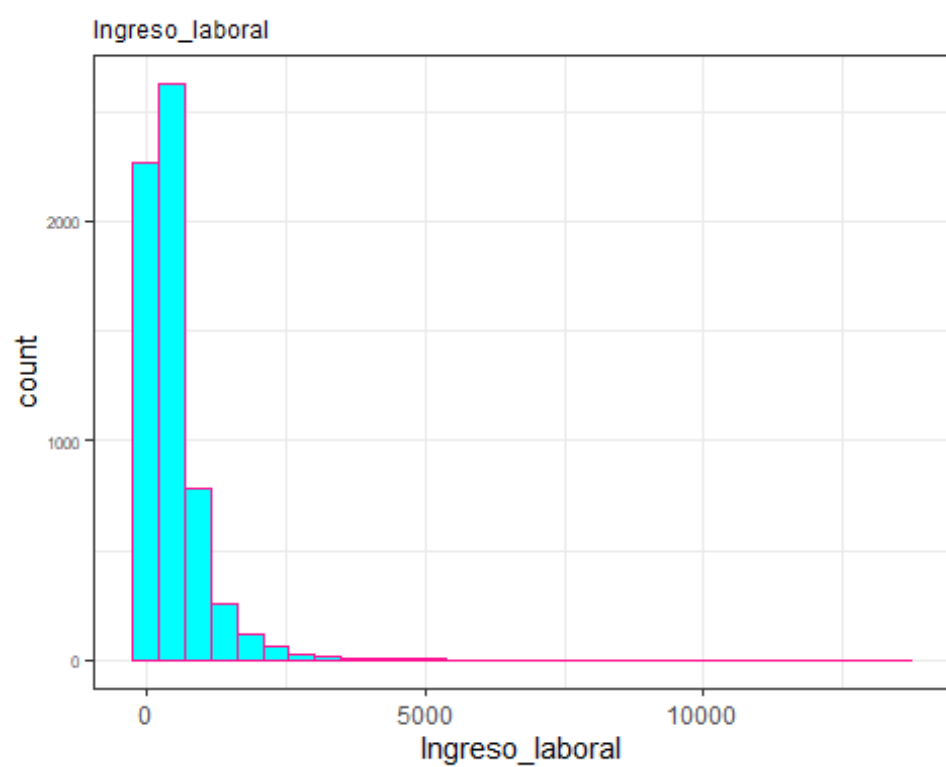
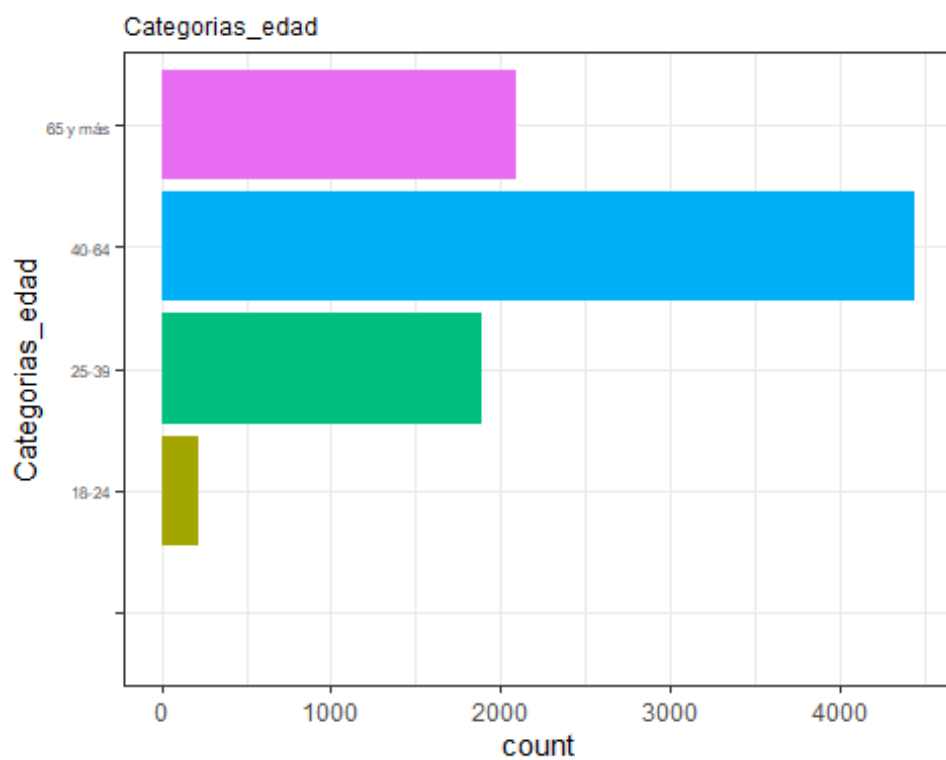


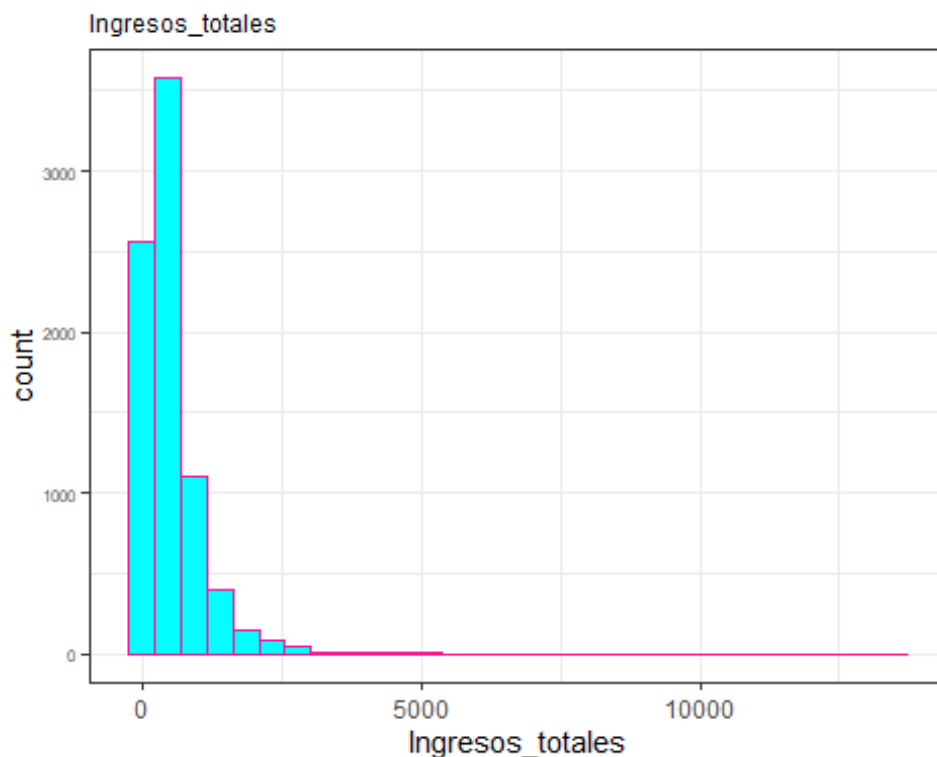












Interpretación Gráficos: Se observa que los datos tienen asimetría en la distribución, los datos de las variables numéricas no se distribuyen de manera normal. En los gráficos de barras se puede observar que unas categorías predominan más que otras y en otros casos las distribuciones son similares.

4) La/s visualizaciones realizadas

Filtros: De fecha (periodo) para los gráficos de series de tiempo.

Filtros de sexo, área, nivel de instrucción, autoidentificación, estado civil, son filtros universales y generales que se aplican al resto de los gráficos.

En los details on demand or tooltips se muestran generalmente en todos los gráficos detalles como el conteo, el ingreso laboral promedio, y el ingreso total promedio, ya que son datos importantes para presentar.

Link de la visualización:
https://public.tableau.com/app/profile/joshelyn2159/viz/final_16228779004940/Dashboard1

Licencia del repositorio:

Creative Commons Zero v1.0 Universal