

# Encuesta Demográfica Retrospectiva 2019: Tutorial para el procesamiento de datos N° 1

*Medalla de oro a la calidad  
en la Gestión Pública*



IRAM-ISO 9001:2015

# ¿Qué encontrará en este tutorial?

A través de este material buscamos facilitar la explotación de la primera encuesta demográfica retrospectiva de estadística pública del país: Encuesta Demográfica Retrospectiva 2019 (EDER-CABA)<sup>1</sup>.

Dado su carácter retrospectivo, la EDER-CABA cuenta con una estructura de datos diferente al modo en que se presenta la información de las encuestas transversales, requiriendo un tratamiento particular de sus bases de datos.

En este documento se desarrollan los alcances del uso de la EDER-CABA en cuanto a su potencial de explotación longitudinal y una descripción de la estructura de datos de las bases de usuarios del operativo 2019.

A la vez, se facilita un paso a paso con los comandos de programación para procesar los datos, a partir de un ejemplo de aplicación.

Estos comandos pueden ser modificados según la selección de variables que resulten de su interés. Si bien se presentan en RStudio, también pueden adaptarse al lenguaje de programación de su *software* de uso habitual.

<sup>1</sup> Documento elaborado por el equipo Ubacyt 20020190200298BA (dirección de Sabrina Ferraris, Instituto de Investigaciones Gino Germani, Universidad Nacional de Buenos Aires). Autoría del documento: Mariela Giacoponello, con coordinación/supervisión de Mario Martínez Salgado y Sabrina Ferraris, y colaboración de investigadores asociados al proyecto.

## Índice

|  |   |
|--|---|
| 1. ¿Qué preguntas podemos responder con la EDER-CABA?    | 3 |
| 2. ¿Qué información contiene la EDER-CABA?               | 4 |
| 3. ¿Cómo procesar los datos de la EDER-CABA?             | 5 |
| 4. Tutorial para procesar los datos de la EDER-CABA 2019 | 8 |

# 1. ¿Qué preguntas podemos responder con la EDER-CABA?

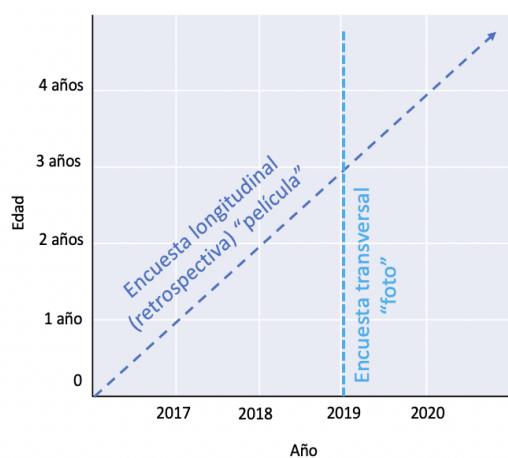
## Potencial de explotación longitudinal

La EDER-CABA 2019 es la primera encuesta de estadística pública del país que capta información sociodemográfica longitudinal, es decir, a lo largo de la trayectoria biográfica de las personas.

Releva las distintas historias de vida en temas como los cambios en el lugar de residencia, la trayectoria educativa y laboral, las uniones conyugales y la tenencia de hijo/as en una muestra de 1.220 personas residentes en la Ciudad de Buenos Aires de tres cohortes de nacimiento: 1948-1952, 1968-1972 y 1978-1982.

A diferencia de los operativos transversales que recolectan datos al momento de la encuesta –una “foto” de la situación–, la EDER-CABA reconstruye la historia de los eventos y sus cambios a lo largo del tiempo –la “película” desde el nacimiento hasta el momento de la encuesta.

Fig. n°1. Encuesta longitudinal y retrospectiva



El potencial de explotación longitudinal de la EDER-CABA favorece el estudio diacrónico de los procesos sociodemográficos a partir de tomar en consideración la dimensión temporal.

A partir de preguntas que historizan los eventos en las biografías de las distintas generaciones, incorpora la dimensión temporal que permite analizar los procesos sociodemográficos en la población de la Ciudad de

Buenos Aires durante la segunda mitad del siglo XX y el inicio del siglo XXI. A su vez, también facilita el estudio de las interrelaciones que dichos eventos demográficos guardan entre sí en las trayectorias de las personas.

El componente diacrónico de la encuesta arroja información sobre el tiempo histórico y el tiempo biográfico. El tiempo histórico da cuenta de los cambios en los patrones culturales entre generaciones o cohortes de nacimiento; y el tiempo biográfico muestra cómo se relacionan los eventos entre sí a lo largo del tiempo o bien en una dimensión particular de la vida (por ejemplo, la residencial), a partir de sus cambios de estado en cada año de vida de las personas.

El hecho de poder calendarizar los eventos por las edades de ocurrencia permite controlar el sesgo de endogeneidad (es decir, conocer qué fenómeno es el que afecta al otro) que suele ocurrir cuando sólo se tiene la información de momento –la “foto”– de las encuestas transversales.

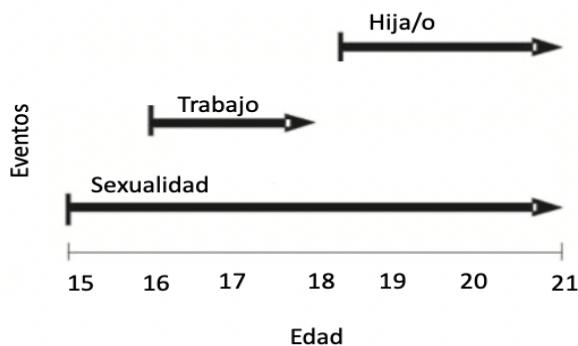
¿Cómo son las trayectorias laborales de las personas y cómo han cambiado entre generaciones y según género? ¿Cómo se relacionan estas trayectorias con la mater/paternidad y los cuidados? ¿Cuántos cambios de residencia han tenido a lo largo de sus vidas? ¿Cómo ha variado en el curso de vida la conformación familiar? ¿Ha sido la misma según el origen social? Estas son algunas de las preguntas que podemos responder a partir de la EDER-CABA.

Por ejemplo, podemos estudiar ciertos eventos que funcionan como hitos o acontecimientos biográficos en el paso de la juventud a la adultez (el ingreso al mercado laboral o el inicio de la mater/paternidad), sus cambios de estado en el tiempo y analizarlos con arreglo a ciertas variables de origen como la generación, el grupo social o el género.

Siguiendo este ejemplo, en la figura 2 puede verse la edad de inicio de tres eventos “de adultez” de una persona encuestada. A los 15 años da inicio a su vida sexual, a los 16 años ingresa al mundo del trabajo y a los 18 años se inicia en la mater/paternidad.

Así, a través de analizar en conjunto la temporalidad de estos tres eventos puede observarse cómo la trayectoria laboral es interrumpida cuando inicia su maternidad.

Fig. n°2. Ejemplo de registro retrospectivo en tres eventos de transición a la adultez



Para ampliar la información sobre los alcances de la EDER-CABA 2019, consulte el [Informe metodológico](#) disponible en el sitio web de la DGEyC.

## 2. ¿Qué información contiene la EDER-CABA?

### Estructura de las bases de datos

La EDER-CABA tiene una estructura de datos diferente a la que suele emplearse para presentar la información de las encuestas de carácter trasversal (una fila, un caso), requiriendo un tratamiento específico de sus bases de datos.

A través de las preguntas retrospectivas, se construye una matriz por cada año de vida de las personas encuestadas. Se ubican los eventos en la matriz que tiene como filas los años o la edad de las personas (años-persona) y como columnas los diferentes eventos y sus cambios de estado a lo largo del tiempo.

Algunas variables identifican la ocurrencia o no de los eventos en consideración (fijas) y otras registran los cambios de estado de dichos eventos a lo largo del tiempo (móviles).

### Glosario

**Evento:** Acontecimiento que ocurre en un momento determinado, tal como el matrimonio, nacimiento de un hijo, cambio de lugar de residencia.

Los eventos registrados en la EDER-CABA son los de duración de al menos 12 meses (un año calendario).

**Estado:** Alude a los cambios en la situación resultante de la ocurrencia de un evento, ser madre, padre, unido, ocupado, entre otros.

En la figura 3 puede observarse un ejemplo del formato matriz años-persona. La persona encuestada 1 ha transitado el evento del trabajo mientras que la persona 2 no ha experimentado el fenómeno al momento de la encuesta. Los cambios en relación con el trabajo de la persona 1 varían entre los 28 y los 32 años. A los 28 años no trabajaba (categoría '0') y a los 29 comienza un nuevo trabajo (categoría '1') en el que cambia de posición a los 31 (categoría '2'). Al año siguiente no trabaja (vuelve a la categoría '0').

Fig. n°3. Estructura de la matriz de datos años-persona

| Persona | Edad | Evento = Trabajo | Estado = Trabajo                        |
|---------|------|------------------|---|
| 1       | 28   | 1                | 0 No trabaja o interrupción del trabajo |
| 1       | 29   | 1                | 1 Comienzo en un nuevo trabajo          |
| 1       | 30   | 1                | 1                                       |
| 1       | 31   | 1                | 2 Cambios dentro del mismo trabajo      |
| 1       | 32   | 1                | 0                                       |
| 2       | 19   | 2                | 0                                       |
| 2       | 20   | 2                | 0                                       |

El hecho de conocer la sucesión temporal de los eventos permite avanzar desde análisis de carácter descriptivo hacia modelos explicativos.

Los datos de la EDER-CABA se distribuyen en cinco bases de usuarios, tres con información retrospectiva con registros años-persona y dos con información donde cada registro es una persona (presentación "tradicional").

En la tabla a continuación se presenta un resumen sobre cada base de datos según el contenido (descripción), la estructura de datos ("matriz años-persona" o "registro por persona"), la cantidad de casos disponibles (personas o años-persona) y el cuestionario que le da origen.

Tabla nº1. Estructura de las bases de datos usuarios EDER-CABA

| Nombre BD                       | Descripción   | Estructura           | Cantidad de registros | Cuestionario que le da origen                |
|---------------------------------|---|----------------------|-----------------------|--|
| eder2019_usuarios_retro         | Cada registro representa un año calendario en que suceden los distintos eventos biográficos de al menos un año de duración.   | Matriz años-persona  | 65.200                | Trayectoria Individual y Antecedentes (TI-A) |
| eder2019_usuarios_retro_conyuge | Presenta información de todos lo/as cónyuges hasta el momento de la encuesta. Cada año calendario se repite según la cantidad de cónyuges.  |                      | 77.524                |  |
| eder2019_usuarios_retro_hijo    | Presenta información de todos lo/as hijo/as hasta el momento de la encuesta. Cada año calendario se repite según la cantidad de hijo/as.  |                      | 122.405               |  |
| eder2019_IOS                    | Cada registro es una persona (población objetivo). Brinda información sobre las condiciones socioeconómicas de origen de la persona encuestada.   | Registro por persona | 1.220                 | Composición del Hogar y Seguimiento (CoH-S)  |
| eder2019_usuarios_personas      | Cada registro es una persona (miembro del hogar) Brinda información por cada hogar identificado en la vivienda encuestada y realiza una caracterización sociodemográfica de los miembros del hogar. | Registro por persona | 3.519                 |  |

Para ampliar la información sobre el desarrollo de las bases de datos, explore el Diseño de registro que acompaña las bases de datos disponibles en el sitio web de la DGEyC.

### 3. ¿Cómo procesar los datos de la EDER-CABA?

#### Ejemplo de aplicación

En este apartado se muestra cómo procesar los datos retrospectivos de la EDER-CABA 2019 en base a un ejemplo.

#### Tema

Nos interesa conocer la relación entre el estudio y el trabajo en la juventud de tres cohortes de nacimiento (1948-1952, 1968-1972 y 1978-1982), sus diferencias según género y origen social.

#### Información disponible

Como se describe en el apartado anterior, la información de la EDER-CABA 2019 se encuentra distribuida en cinco bases de datos.

Para este ejemplo utilizaremos únicamente las bases 'eder2019\_usuarios\_retro' y 'eder2019\_usuarios\_

IOS'. La base de datos 'eder2019\_usuarios\_retro' cuenta con variables fijas como cohortes y sexo<sup>2</sup> y con variables móviles sobre los estados en cada año calendario de las personas encuestadas respecto a su situación educativa y laboral.

La BD 'eder2019\_usuarios\_IOS' contiene el Índice de Origen Social (IOS) que integra en un indicador las condiciones socioeconómicas de origen de las personas encuestadas<sup>3</sup>.

2 Para este ejercicio se toma la variable 'sexo' como proxy de género ya que su estructura dicotómica (varón/mujer) no categoriza las disidencias.

3 El IOS es una medida relativa que indica la posición de la persona encuestada en la estratificación social a los 14 años, en relación con las personas que integran su cohorte de nacimiento. Para ampliar la información sobre el desarrollo del IOS, consulte el documento metodológico disponible en el sitio web de la DGEyC.

Tabla nº2. Distribución de las variables de interés en las bases de usuarios EDER-CABA

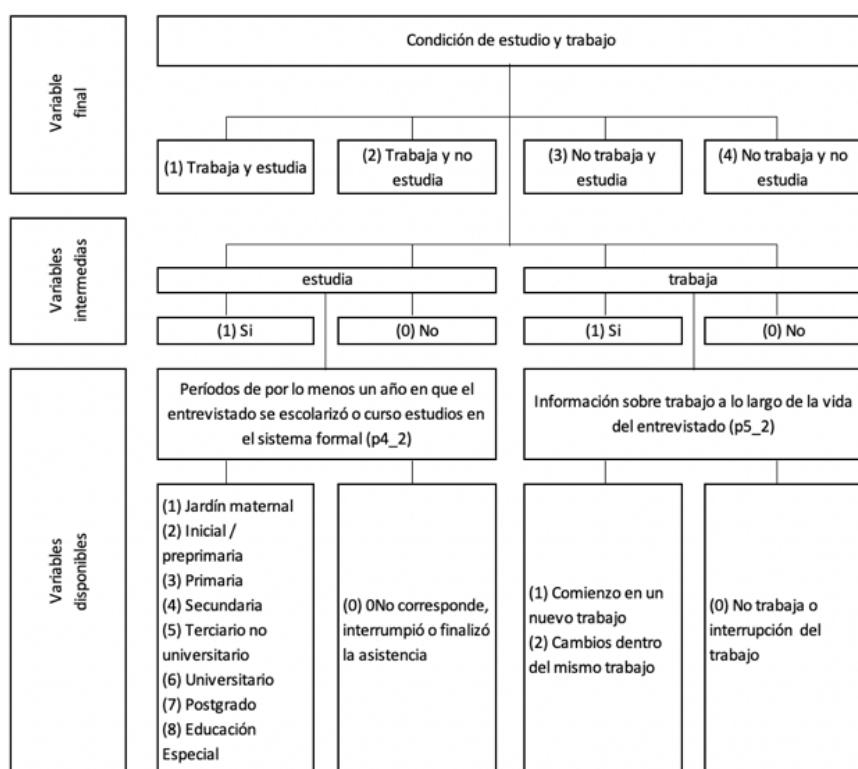
| Nombre base de datos    | Variables seleccionadas |  |  |
|-------------------------|-------------------------|--|--|
|                         | Código                  | Nombre   | Categorías   |
| eder2019_usuarios_retro | p2                      | Sexo   | 1 Varón<br>2 Mujer   |
|                         |                         | cohorte  | 1 1948-1952<br>2 1968-1972<br>3 1978-1982  |
|                         |                         | p5_2   | Información sobre trabajo a lo largo de la vida del entrevistado<br>0 No trabaja o interrupción del trabajo<br>1 Comienzo en un nuevo trabajo<br>2 Cambios dentro del mismo trabajo  |
|                         | p4_2                    | Períodos de por lo menos un año en que el entrevistado se escolarizó o curso estudios en el sistema formal | 0 No corresponde, interrumpió o finalizó la asistencia<br>1 Jardín maternal<br>2 Inicial/preprimaria<br>3 Primaria<br>4 Secundaria<br>5 Terciario no universitario<br>6 Universitario<br>7 Postgrado<br>8 Educación Especial |
|                         |                         |  |  |
|                         |                         |  |  |
|                         |                         |  |  |
|                         |                         |  |  |
| eder2019_usuarios_IOS   | losterciles             | Información en terciles del índice de origen social  | 1 Bajo<br>2 Medio<br>3 Alto  |

### Condición de estudio y trabajo

Construiremos la variable ‘Condición de estudio y trabajo’ para identificar todas las combinaciones posibles a cada edad de las personas: trabajan y estudian, trabajan y no estudian, no trabajan y

estudian, o no trabajan y no estudian. Para ello, elaboramos dos variables intermedias dicotómicas sobre el estudio y el trabajo.

Fig. nº4. Proceso de transformación de las variables de origen a la variable de interés



Nos preguntamos qué porcentaje continúa (20 años) y cuál es la proporción que combina estudiando a una edad determinada dicha dimensión con el trabajo. Interesa conocer las desigualdades de género en la condición frente al estudio y el trabajo, así como su variación entre generaciones y grupos sociales.

Estudiaremos este proceso en un tramo de las biografías de las personas (desde los 6 a los 35 años) con el fin de identificar las variaciones que ocurren en estas dos dimensiones en simultáneo en ese período temporal.

Se señala que, con el objeto de seguir evidenciando las potencialidades de esta encuesta, podríamos profundizar el estudio de las desigualdades de género y ampliar el análisis, estudiando la relación entre “no estudiar y no trabajar” con las trayectorias de trabajo no remunerado (información que esta encuesta también provee).

A partir de estos datos elaboramos tablas de contingencia y gráficas como ejemplos de muestra sobre cómo procesar los datos para analizar las diferencias en la combinación de estas dos dimensiones de la vida entre hombres y mujeres, cohortes de nacimiento y el índice de origen social.

# 4. Tutorial para procesar los datos de la EDER-CABA 2019

## Objetivo

Mostrar algunos elementos básicos del Lenguaje R para el manejo de las bases de datos de la Encuesta Demográfica Retrospectiva (EDER-CABA). Con algunos ejemplos se ilustran algunas de las posibilidades de análisis que ofrece la EDER-CABA 2019.

## Preliminares

Antes de comenzar, limpiar el área de trabajo (Environment)

```
rm(list = ls())
```

Definir directorio de trabajo. Por defecto la carpeta *Documentos* cumple esta función. Para cambiar de sitio debemos seguir los siguientes pasos:

1. En el menú principal pulsar la opción **Session**
2. Desplegar las opciones de **Set Working Directory**
3. Elegimos la opción **Choose Directory...**
4. En la ventana emergente seleccionar la carpeta de interés

Otra opción es utilizar la función `setwd()`

```
setwd("Ruta donde aloja las bases de datos")
```

## Descargar y fusionar las bases de datos

Los datos de la EDER-CABA se encuentran alojados en el sitio web de la Dirección General de Estadística y Censos. Al pulsar en el siguiente [enlace](#) se puede acceder a ellos. Descargar el archivo y poner el contenido en el directorio de trabajo.

Con base en los datos del archivo `eder2019_usuarios_retro.txt` crear el objeto `eder_retro`.

```
eder_retro <- read.table("eder2019_usuarios_retro.txt", header = TRUE, sep = "\t", fileEncoding = "UCS-2LE")
```

Utilizamos la función `dim()` para conocer la estructura de las bases de datos (cantidad de registros y variables)

```
dim(eder_retro)
## [1] 65200 115
```

Para conocer el nombre de las variables se puede usar la función `names()`

```
names(eder_retro)
```

Para cargar el resto de bases de datos repetimos el paso donde se emplea la función `read.table()`

```
eder_hijo <- read.table("eder2019_usuarios_retro_hijo.txt",
                         header = TRUE, sep =
"\t", fileEncoding = "UCS-2LE")
eder_conyu <- read.table("eder2019_usuarios_retro_conyuge.txt",
                         header = TRUE, sep =
"\t", fileEncoding = "UCS-2LE")
eder_ios <- read.delim2("eder2019_usuarios_ios.txt")
eder_pers <- read.table("eder2019_usuarios_personas.txt",
                         header = TRUE, sep =
"\t", fileEncoding = "UCS-2LE")
```

Al tener los datos en el *Environment* podemos crear el objeto `eder` fusionando los objetos `eder_retro` y `eder_ios`. Para ello utilizamos la función `merge()` y las variables identificador: `id`, `nhogar` y `miembro`

```
eder <- merge(eder_retro, eder_ios,
               by = c("id", "nhogar", "miembro"),
               all.x = TRUE)
```

Podemos guardar la base de datos unificada en el formato `*.RData` para un uso posterior con la función `save()`. Este archivo quedará alojado en el directorio de trabajo.

```
save (eder, file = "eder19_caba.RData")
```

Para mostrar cómo podemos cargar los datos recién guardados limpiamos el área de trabajo (*Environment*)

```
rm(list = ls())
```

Y cargamos los datos con la función `load()`

```
load("eder19_caba.RData")
```

## Ejemplo: condición de trabajo y estudio

Ahora creamos el objeto `caba` a partir del objeto `eder`. Para ello seleccionamos algunas variables que nos permitirán ilustrar algunas capacidades con un ejemplo sobre la condición de trabajo y estudio: trabajo (`p5_1, p5_2`), estudio (`p4_1, p4_2`); adicionalmente seleccionamos las variables sexo (`p2`), generación (`cohorte`), ios (`iosterciles`) y las variables identificador: `id`, `annio_retro`, `edad_retro` y `p1_1_annio`

```
caba <- eder[, c("id", "annio_retro", "edad_retro", "p1_1_annio", "cohorte", "iosterciles", "p2", "p5_1", "p5_2", "p4_1", "p4_2")]
dim(caba)
## [1] 65200     11
```

Creamos las variables dicotómicas sobre el trabajo: trabaja y no trabaja, y sobre la educación: estudia y no estudia

```
caba$trabaja <- ifelse(caba$p5_2 %in% c(1:2), 1, 0)
caba$estudia <- ifelse(caba$p4_2 %in% c(1:8), 1, 0)
```

Ordenamos los registros de la base de acuerdo con las variables `id` y `edad_retro`

```
caba <- caba[order(caba$id, caba$edad_retro),]
```

Ver el resultado en la base de datos

```
View(caba)
```

Creamos la variable categórica que reúne las combinaciones entre estudio y trabajo en cada año de la vida

```
caba$est.trab <- 0
caba$est.trab[caba$trabaja==1 &
```

```
caba$estudia==1] <- 1 # Trabaja y estudia
caba$est.trab[caba$trabaja==1 &
caba$estudia==0] <- 2 # Trabaja y no estudia
caba$est.trab[caba$trabaja==0 &
caba$estudia==1] <- 3 # No trabaja y estudia
caba$est.trab[caba$trabaja==0 &
caba$estudia==0] <- 4 # No trabaja y no estudia
```

Con la función `table()` obtenemos la distribución de la variable recién creada

```
table(caba$est.trab, useNA = "always")
```

```
##          1      2      3      4    <NA>
## 4286 31313 15805 13796      0
```

Ahora etiquetamos las variables `est.trab`, `sexo`, `cohorte` e `ios`

```
caba$est.trab_e <- factor(caba$est.trab, levels = c(1:4),
                           labels = c("Trabaja y estudia",
                                     "Trabaja y no estudia",
                                     "No trabaja y estudia",
                                     "No trabaja y no estudia"))
```

```
caba$sexo_e <- factor(caba$p2, levels = c(1,2),
                        labels = c("Hombre", "Mujer"))
```

```
caba$cohorte_e <- factor(caba$cohorte, levels = c(1:3),
                           labels = c("1948-1952", "1968-1972", "1978-1982"))
```

```
caba$ios_e <- factor(caba$ios, levels = c(1:3),
                      labels = c("Bajo", "Medio", "Alto"))
```

Distribución de los registros en la variable `est.trab_e`

```
table(caba$est.trab_e, useNA = "always")
```

```
##          Trabaja y estudia      Trabaja y no
##          estudia      No trabaja y estudia
##          4286
##          31313      15805
##          No trabaja y no estudia
##          13796
##          0
```

Resultado en un momento determinado (a los 20 años)

```
table(caba$est.trab_e[caba$edad_retro == 20],  
useNA = "always")  
gits = 1)
```

```
##  
## Trabaja y estudia Trabaja y no  
estudia No trabaja y estudia  
## 310  
489 235  
## No trabaja y no estudia  
<NA>  
## 186  
0
```

Tabla de contingencia que combina la condición de estudio y el trabajo con el sexo a los 20 años

```
table(caba$est.trab_e[caba$edad_retro == 20],  
caba$sexo_e[caba$edad_retro == 20],  
useNA = "always")
```

```
##  
## Hombre Mujer  
<NA>  
## Trabaja y estudia 164 146  
0  
## Trabaja y no estudia 266 223  
0  
## No trabaja y estudia 110 125  
0  
## No trabaja y no estudia 64 122  
0  
## <NA> 0 0
```

Resultados en %

```
tabla <- table(caba$est.trab_e[caba$edad_re-  
tro == 20],  
caba$sexo_e[caba$edad_retro == 20])
```

% de fila

```
round(prop.table(tabla, margin = 1)*100, di-  
gits = 1)
```

```
##  
## Hombre Mujer  
## Trabaja y estudia 52.9 47.1  
## Trabaja y no estudia 54.4 45.6  
## No trabaja y estudia 46.8 53.2  
## No trabaja y no estudia 34.4 65.6
```

% de columna

```
round(prop.table(tabla, margin = 2)*100, di-
```

```
##  
## Hombre Mujer  
## Trabaja y estudia 27.2 23.7  
## Trabaja y no estudia 44.0 36.2  
## No trabaja y estudia 18.2 20.3  
## No trabaja y no estudia 10.6 19.8
```

Tabla de contingencia que combina la condición de estudio y trabajo con la cohorte de nacimiento a los 20 años

```
table(caba$est.trab_e[caba$edad_retro == 20],  
caba$cohorte_e[caba$edad_retro == 20],  
useNA = "always")
```

```
##  
## 1948-1952 1968-  
1972 1978-1982 <NA>  
## Trabaja y estudia 85  
112 113 0  
## Trabaja y no estudia 189  
161 139 0  
## No trabaja y estudia 61  
77 97 0  
## No trabaja y no estudia 72  
56 58 0  
## <NA> 0  
0 0 0
```

Conjunto de tablas de contingencia que combinan la condición de estudio y el trabajo con el ios y el sexo a los 20 años

```
table(caba$est.trab_e[caba$edad_retro == 20],  
caba$ios_e[caba$edad_retro == 20],  
caba$sexo_e[caba$edad_retro == 20])
```

```
## , , = Hombre  
##  
## Bajo Medio Alto  
## Trabaja y estudia 22 69 73  
## Trabaja y no estudia 134 91 36  
## No trabaja y estudia 16 31 63  
## No trabaja y no estudia 27 23 14  
##  
## , , = Mujer  
##  
## Bajo Medio Alto  
## Trabaja y estudia 23 58 64  
## Trabaja y no estudia 124 57 39  
## No trabaja y estudia 22 41 62  
## No trabaja y no estudia 61 39 20
```

Para generar un gráfico de la distribución de la condición de estudio y trabajo por sexo y edad (6 a 35 años) usamos la librería ggplot2

```
install.packages("ggplot2")
library(ggplot2)

"Trabaja y no estudia",
"No trabaja y estudia",
"No trabaja y no estudia"))
```

Para hacer la gráfica antes debemos preparar los datos. Creamos en primer lugar una variable auxiliar para el conteo de casos

```
caba$aux <- 1
```

Después creamos el objeto datos que agrega la suma de casos por cada categoría de la variable sobre estudio y trabajo para cada sexo y cada año retrospectivo

```
datos <- aggregate(list(caba$aux),
by = list(caba$sexo,
caba$edad_retro, caba$est.trab),
FUN = sum,
na.rm = TRUE)
```

Cambiamos los nombres de las columnas

```
names(datos) <- c("Sexo", "Edad", "Categoria",
"Suma")
```

Creamos el objeto 'datos2' que agrega la cantidad de casos para cada sexo y año retrospectivo

```
datos2 <- aggregate(list(caba$aux),
by = list(caba$sexo,
caba$edad_retro),
FUN = sum,
na.rm = TRUE)
```

```
names(datos2) <- c("Sexo", "Edad", "Total")
```

Fusionamos los objetos datos y datos2

```
datos <- merge(datos, datos2, by = c("Sexo",
"Edad"), all = TRUE)
```

Calculamos el porcentaje

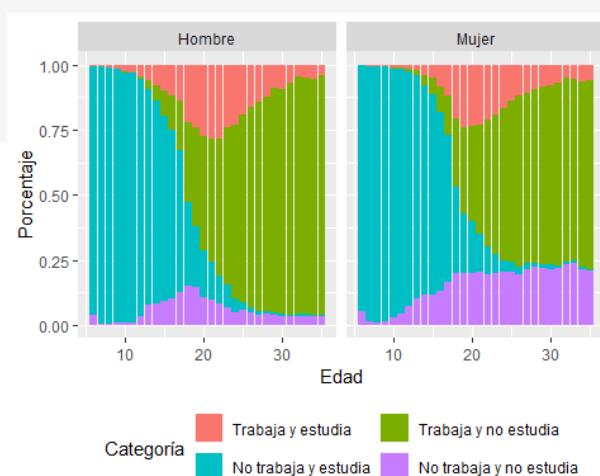
```
datos$Porcentaje <- datos$Suma/
datos$Total*100
```

Etiquetamos la variable Categoria

```
datos$Categoria <- factor(datos$Categoria,
levels = c(1:4),
labels = c("Trabaja y estudia",
```

Ahora generamos un gráfico de barras apiladas

```
ggplot(datos[datos$Edad %in% c(6:35),],
aes(fill = Categoria, y = Porcentaje, x
= Edad)) +
geom_bar(position = "fill", stat = "identi-
ty") +
facet_wrap(~Sexo) +
theme(legend.position = "bottom") +
guides(fill = guide_legend(title="Cate-
goria", nrow = 2, byrow = TRUE))
```



En el gráfico cada barra representa la distribución de las categorías (Categoria) en diferentes edades (Edad), y las alturas de las barras indican los valores de porcentaje (Porcentaje).

Para guardar el gráfico en la carpeta de trabajo usamos la función ggsave()

```
ggsave("est_tra_sexo.png", plot = last_
plot(), device = "png", dpi = 300)
```

Por último, siguiendo con el ejemplo, adaptando estos comandos puede reemplazar la variable 'sexo' por la de generaciones ('cohorte') o bien la de origen social ('iosterciles').

Para más información estadística podés seguirnos en nuestras redes,  
visitarnos en nuestra web o consultarnos por mail.