

Métodos Estadísticos I - 2018

Visualización de datos en R

Ayudante alumna Eugenia Settecase

10 de Abril, 2018

- Importación de los Datos
- Análisis de datos
- Visualización gráfica paquete base
- Visualización gráfica con ggplot2
- Tablas

Importación de los Datos

Una de la formas más sencillas de cargar datos en R es utilizar un paquete que no está incluido dentro de los básicos del programa. En este caso se debe descargar el paquete **readxl** que permite importar datos desde Excel. Adicionalmente, se debe tener en cuenta que para este paquete en particular, el archivo debe tener extensión “.xls” para este paquete particular. Además se debe conocer dónde está almacenado, es decir la ruta del archivo.

Instalación del paquete

```
install.packages("readxl", repos = "http://cran.us.r-project.org")
```

Otra manera es utilizar la opción “*Import Dataset*” que está ubicada en la ventana superior derecha, en la pestaña de “*Environment*”. En ese caso, se debe seleccionar la forma de importación “**From Excel**”.

Una vez que el paquete ya fue instalado este debe ser cargado a partir de la sentencia **library()**. Se debe destacar que este último paso debe realizarse cada vez que se necesite hacer uso de un paquete que fue previamente instalado.

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.4.4
```

Antes de cargar los datos se puede indicar cuál es el directorio que se va a utilizar para que el programa lea los archivos y guarde las exportaciones en esa dirección. Esto se puede realizar a través de un comando o de manera manual en la pestaña “*Files*” que se encuentra en la ventana inferior derecha.

```
# setwd("C:/.../...../....")
```

Importante: A la hora de definir la ruta de un archivo o carpeta se debe tener en cuenta que R usa las barras con inclinación a la derecha: “/”

Importación

```
Base <- read_excel("Base_partos.xls")
```

```
View(Base)
```

Se puede visualizar la base de datos de dos maneras, desde la ventana “*Environment*”, que se encuentra arriba a la derecha o con el comando **View()**, indicando el nombre de la base de datos entre los paréntesis.

A destacar: R distingue entre mayúsculas y minúsculas, por lo que se debe tener cuidado a la hora de nombrar los objetos. Por ejemplo, el archivo “Base” es distinto a “base”.

Análisis de datos

Una vez cargada la base de datos se puede hacer uso de diferentes funciones para conocer algunas características sobre los datos.

```
summary(Base)
```

```
##      ID      Inicio del parto  Presentacion
## Min.   : 1.0   Length:1000      Length:1000
## 1st Qu.: 250.8 Class :character  Class :character
## Median : 500.5 Mode  :character  Mode  :character
## Mean   : 500.5
## 3rd Qu.: 750.2
## Max.   :1000.0
## Tamaño fetal acorde Terminación      Sexo del RN
## Length:1000      Length:1000      Length:1000
## Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character
##
##
##
##      Estudios      Estado Civil      Edad de la madre Partos previos
## Length:1000      Length:1000      Min.   :14.00      Min.   : 0.000
## Class :character  Class :character  1st Qu.:20.00      1st Qu.: 0.000
## Mode  :character  Mode  :character  Median :23.00      Median : 1.000
##                                     Mean  :24.48      Mean  : 1.264
##                                     3rd Qu.:29.00      3rd Qu.: 2.000
##                                     Max.   :42.00      Max.   :10.000
## Cantidad de consultas prenatales Edad Gestacional al parto Peso al nacer
## Min.   : 0.000      Min.   :21.00      Min.   : 560
## 1st Qu.: 5.000      1st Qu.:38.00      1st Qu.:2985
## Median : 6.000      Median :39.00      Median :3330
## Mean   : 6.019      Mean  :38.59      Mean  :3310
## 3rd Qu.: 8.000      3rd Qu.:40.00      3rd Qu.:3665
## Max.   :20.000      Max.   :42.00      Max.   :5180
```

```
summary(Base$`Edad de la madre`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 14.00  20.00  23.00  24.48  29.00  42.00
```

```
head(Base)
```

```
## # A tibble: 6 x 13
##      ID `Inicio del parto` Presentacion `Tamaño fetal acorde` Terminación
##   <dbl> <chr>           <chr>           <chr>           <chr>
## 1 1. Espontáneo      Cefálica      Si      Espontánea
## 2 2. Espontáneo      Cefálica      Si      Forceps
## 3 3. Espontáneo      Cefálica      Si      Forceps
## 4 4. Espontáneo      Cefálica      Si      Espontánea
## 5 5. Cesárea electiva Cefálica      Si      Cesárea
## 6 6. Espontáneo      Cefálica      Si      Espontánea
## # ... with 8 more variables: `Sexo del RN` <chr>, Estudios <chr>, `Estado
## #   Civil` <chr>, `Edad de la madre` <dbl>, `Partos previos` <dbl>,
## #   `Cantidad de consultas prenatales` <dbl>, `Edad Gestacional al
## #   parto` <dbl>, `Peso al nacer` <dbl>
```

```
str(Base)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   1000 obs. of  13 variables:
## $ ID : num  1 2 3 4 5 6 7 8 9 10 ...
## $ Inicio del parto : chr  "Espontáneo" "Espontáneo" "Espontáneo" "Espontáneo" ...
## $ Presentacion : chr  "Cefálica" "Cefálica" "Cefálica" "Cefálica" ...
## $ Tamaño fetal acorde : chr  "Si" "Si" "Si" "Si" ...
## $ Terminación : chr  "Espontánea" "Forceps" "Forceps" "Espontánea" ...
## $ Sexo del RN : chr  "Masculino" "Masculino" "Femenino" "Femenino" ...
## $ Estudios : chr  "Primaria" "Secundaria" "Secundaria" "Secundaria" ...
## $ Estado Civil : chr  "Soltero" "Soltero" "Unión estable" "Unión estable" ...
## $ Edad de la madre : num  20 18 23 21 25 21 21 30 21 18 ...
## $ Partos previos : num  1 0 0 1 1 0 0 2 2 1 ...
## $ Cantidad de consultas prenatales: num  2 8 7 10 5 5 5 10 6 5 ...
## $ Edad Gestacional al parto : num  39 40 39 41 41 40 39 39 39 36 ...
## $ Peso al nacer : num  3130 3640 3380 3420 4000 ...
```

Notas

- El signo \$ luego de nombrar un objeto indica que se quiere aplicar la función sobre una de las variables perteneciente al objeto, es decir que con `summary(Base$Edad de la madre)` se le pide un resumen de la variable “Edad de la madre”.
- Si se quiere saber para qué sirve una función en R se puede:
 - Buscar la función en la pestaña “Help” que se encuentra en la ventana inferior derecha, o
 - Escribir alguno de los siguiente comandos

```
?str
```

```
## starting httpd help server ... done
```

```
help("str")
```

- Una función que es útil cuando se trabaja con una sola base es **attach()**, así la base de datos se adjunta a la ruta de búsqueda de R. Esto significa que R realiza la búsqueda de la/s variable/s mencionada/s en la base de datos que fue “fijada”, de modo que se puede acceder a los objetos en la base de datos simplemente dando sus nombres.

```
attach(Base)
```

Medidas de posición y de dispersión

A continuación se ejemplifica el cálculo de algunas de las medidas dadas en clases.

```
mean(`Edad de la madre`)
```

```
## [1] 24.477
```

```
median(`Edad de la madre`)
```

```
## [1] 23
```

```
var(`Edad de la madre`)
```

```
## [1] 38.58205
```

```
sd(`Edad de la madre`)
```

```
## [1] 6.211445
```

```
min(`Peso al nacer`)
```

```
## [1] 560
```

```
max(`Peso al nacer`)
```

```
## [1] 5180
```

Visualización gráfica paquete base

Este apunte cubre principalmente la visualización gráfica a partir del paquete base de R. Sin embargo, al final de este documento se encuentran desarrollados los gráficos en un paquete más complejo llamado **ggplot2**.

Gráfico de barras

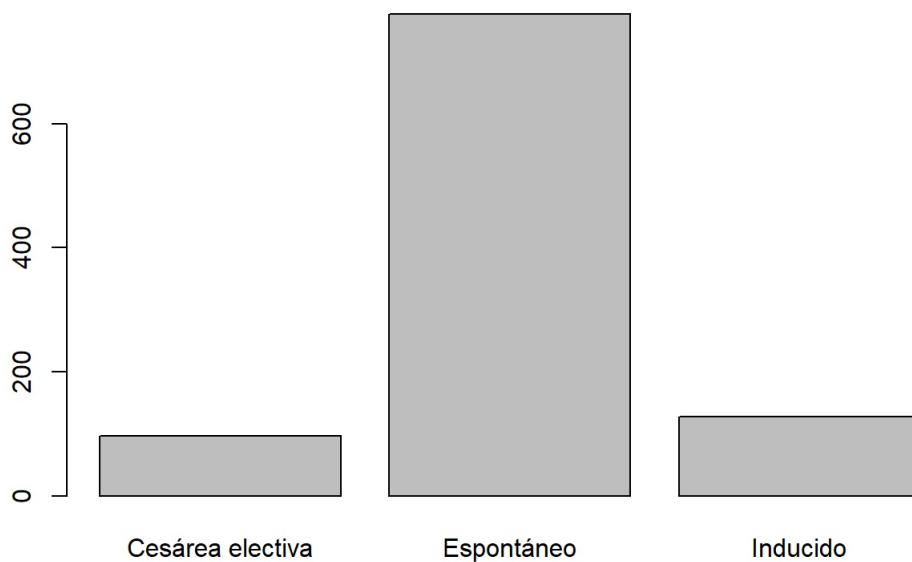
Para este gráfico se toma la variable “Inicio de Parto”, la cual es categórica e indica si el parto inició por cesárea, de manera espontánea o si fue inducido. Para graficar la variable se necesita tener los datos en formato de tabla, por lo que se construye una tabla que contenga el total de casos para cada una de las categorías, a esta tabla se la va a llamar “*tab.inicio*”.

```
table(`Inicio del parto`)
```

```
## Inicio del parto
## Cesárea electiva      Espontáneo      Inducido
##              96              777              127
```

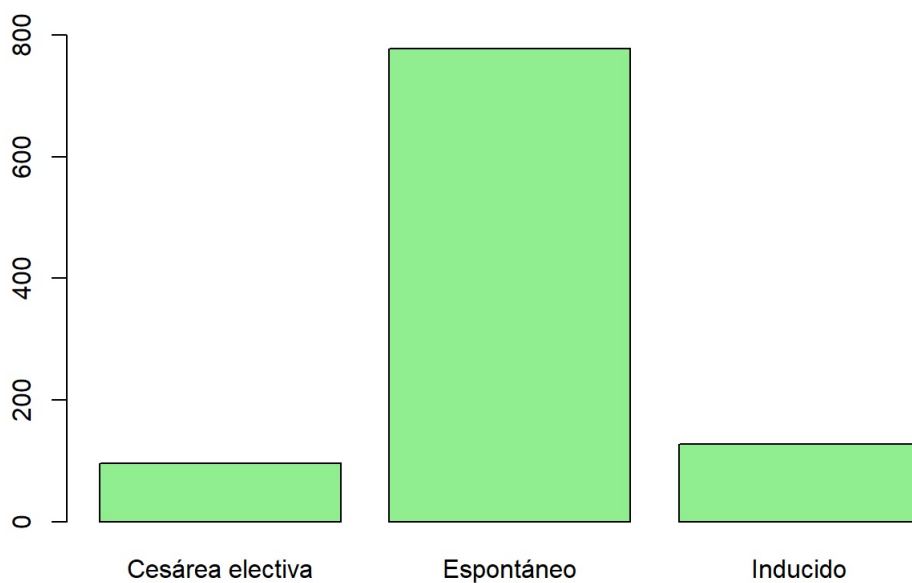
```
tab.inicio<-table(`Inicio del parto`)
```

```
barplot(tab.inicio)
```



El gráfico obtenido puede ser útil para tener una primera idea acerca de la variable pero no sería correcto presentarlo en un informe ¿Cómo se podría mejorar?

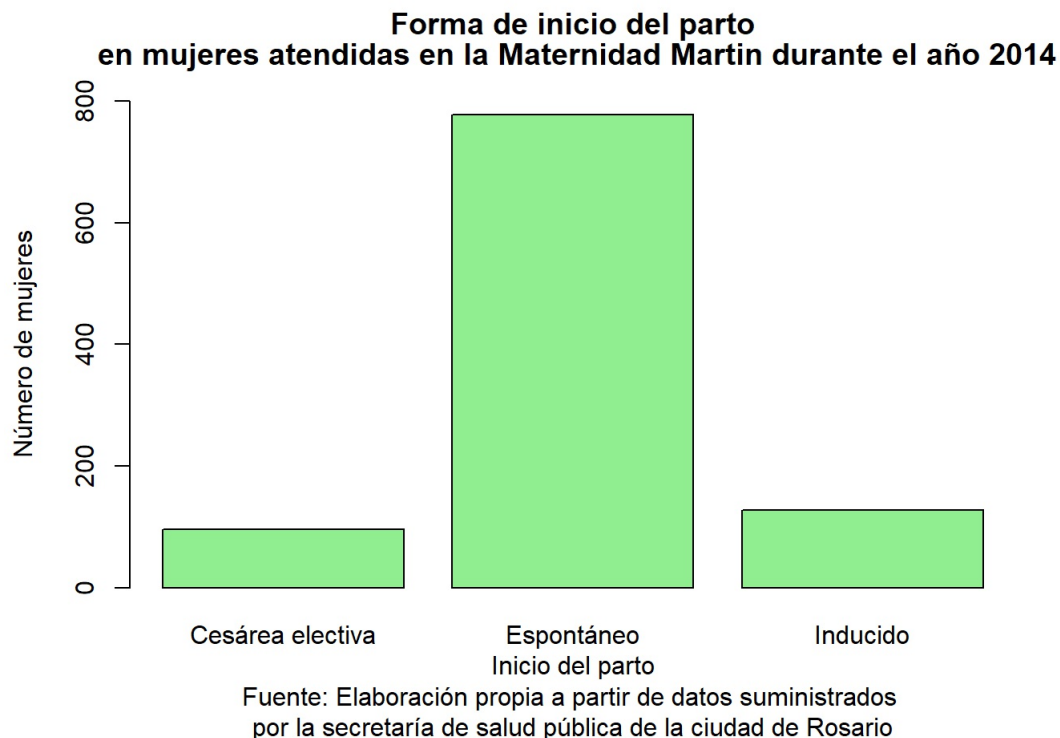
```
barplot(tab.inicio, ylim =c(0,800) , col = "lightgreen")
```



En el paso anterior se cambió el color de las barras y el rango del eje Y, mientras que en el siguiente código se le agrega el título, los nombre de cada uno de los ejes y la fuente.

```
barplot(tab.inicio, ylim =c(0,800) , col = "lightgreen", main = strsplit(c("Forma de inicio del parto; en
mujeres atendidas en la Maternidad Martin durante el año 2014"), split = ";"),ylab = "Número de mujeres"
, sub="Fuente: Elaboración propia a partir de datos suministrados \npor la secretaría de salud pública de
la ciudad de Rosario")

mtext("Inicio del parto", side=1 , line=2)
```



En este gráfico se agrega otra opción para centrar el título utilizando la función **strsplit()**, donde se indica en **split= “;”** que el texto luego del punto y coma irá en el siguiente renglón. Se debe notar que en **barplot()** se indica con **“ylab=”** el nombre de eje Y, pero no así para el eje de las X. Esto se debe a que si se hiciera de esa manera se solaparía con la fuente especificada. Por lo tanto se utiliza la función **mtext()** donde:

- **“side=1”** indica que el texto irá debajo del gráfico
- **“line=2”** indica que será colocado en la línea del margen n°2 (ver imagen)

Nota: la función **mtext()** debe ser corrida juntp con **barplot()** para que se ejecute.

Gráfico de Barras compuesto

En este caso se cruzan las variables **“Inicio de Parto”** y **“Sexo del recién nacido”**. Nuevamente se crea una tabla que contenga la información resumida de ambas variables para poder ser graficadas.

```
tab.in.sexorn<- table(`Sexo del RN`, `Inicio del parto`)

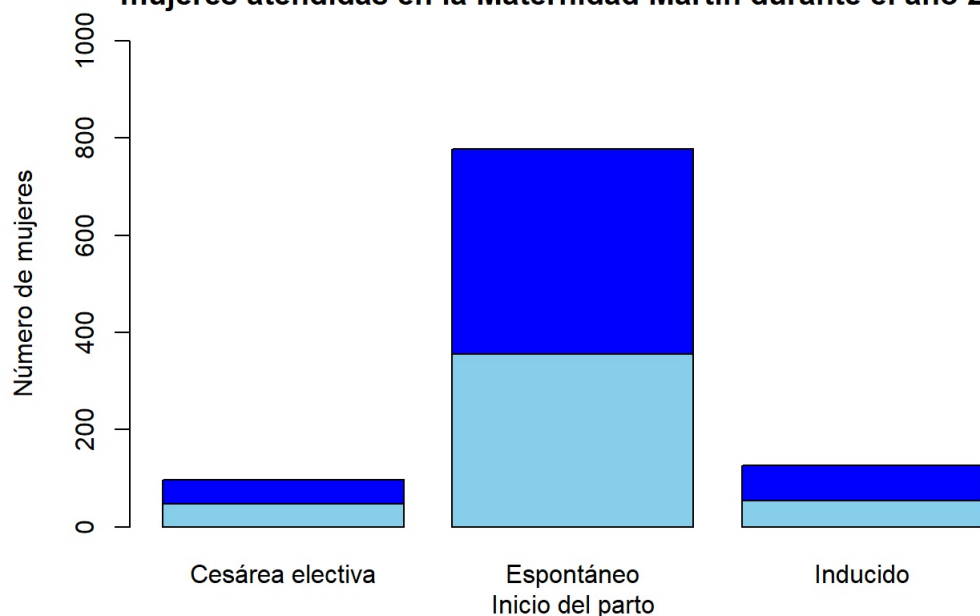
tab.in.sexorn
```

```
##          Inicio del parto
## Sexo del RN Cesárea electiva Espontáneo Inducido
## Femenino          48          355          54
## Masculino          48          422          73
```

```
barplot(tab.in.sexorn, col=c("skyblue","blue"), ylim =c(0,1000), main = strsplit(c("Forma de inicio del p
arto según el sexo del recién nacido para; mujeres atendidas en la Maternidad Martin durante el año 2014"
), split = ";"), sub="Fuente: Elaboración propia a partir de datos suministrados \npor la secretaría de s
alud pública de la ciudad de Rosario" ,ylab = "Número de mujeres" )

mtext("Inicio del parto", side=1 , line=2)
```

Forma de inicio del parto según el sexo del recién nacido para mujeres atendidas en la Maternidad Martin durante el año 2014



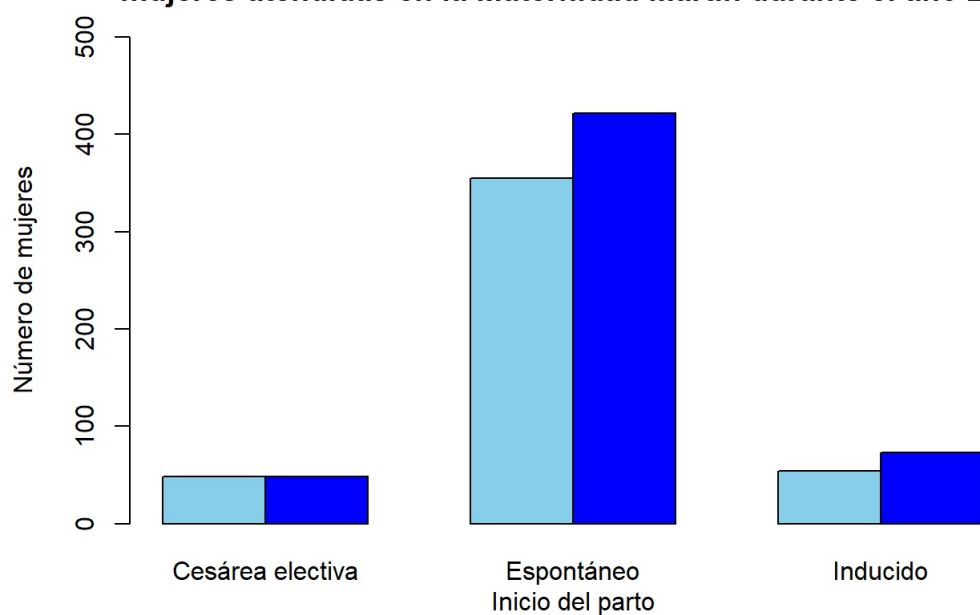
Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

Otro ejemplo de este gráfico utiliza la sentencia **"beside="**, la cual indica que se quiere dividir a las barras según el *"Sexo del recién nacido"* para cada categoría de *"Inicio de parto"*.

```
barplot(tab.in.sexorn, col=c("skyblue","blue"), ylim=c(0,500), beside=TRUE,
       main = strsplit(c("Forma de inicio del parto según el sexo del recién nacido para; mujeres atendi
das en la Maternidad Martin durante el año 2014"), split = ";"), sub="Fuente: Elaboración propia a partir
de datos suministrados \npor la secretaría de salud pública de la ciudad de Rosario" ,ylab = "Número de
mujeres")

mtext("Inicio del parto", side=1 , line=2)
```

Forma de inicio del parto según el sexo del recién nacido para mujeres atendidas en la Maternidad Martin durante el año 2014



Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

En la siguiente sentencia se muestra cómo agregar una "leyenda", es decir , una referencia para la variable *"Sexo del Recién nacido"*.

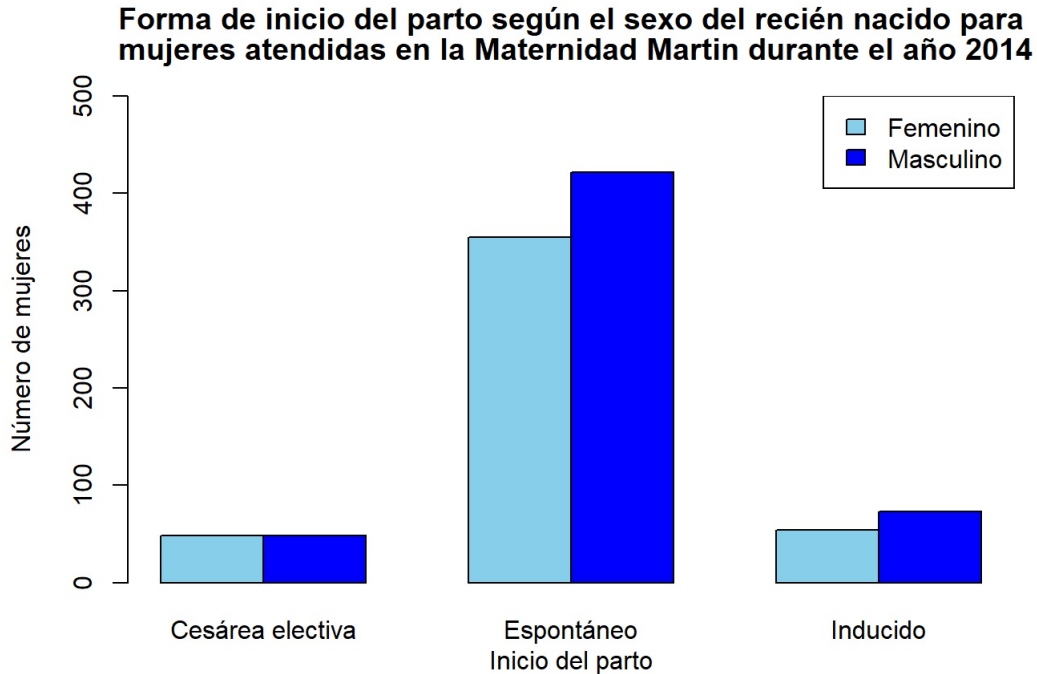
```

barplot(tab.in.sexorn, col=c("skyblue","blue"), ylim=c(0,500), beside=TRUE,
        main = strsplit(c("Forma de inicio del parto según el sexo del recién nacido para; mujeres atendi-
das en la Maternidad Martin durante el año 2014"), split = ";"), sub="Fuente: Elaboración propia a partir
de datos suministrados \npor la secretaria de salud pública de la ciudad de Rosario" ,ylab = "Número de
mujeres")

mtext("Inicio del parto", side=1 , line=2)

legend("topright", legend = c("Femenino", "Masculino"), fill = c("skyblue","blue"))

```



Fuente: Elaboración propia a partir de datos suministrados por la secretaria de salud pública de la ciudad de Rosario

Nota: Igual que en el caso de **mtext()**, para que la función sea ejecutada **legend()** esta se debe correr junto con el gráfico correspondiente.

Gráfico de barras porcentual

Para este gráfico se construye una tabla cruzada de porcentajes de las variables con las que se venía trabajando. Para ello se toma la tabla cruzada *"tab.in.sexorn"* y se la convierte en una tabla de proporciones utilizando la sentencia **prop.table()**, luego multiplicando por 100 se obtienen los porcentajes.

```

tab.porc <- prop.table(tab.in.sexorn)*100

tab.porc

```

```

##           Inicio del parto
## Sexo del RN Cesárea electiva Espontáneo Inducido
## Femenino           4.8           35.5           5.4
## Masculino          4.8           42.2           7.3

```

```

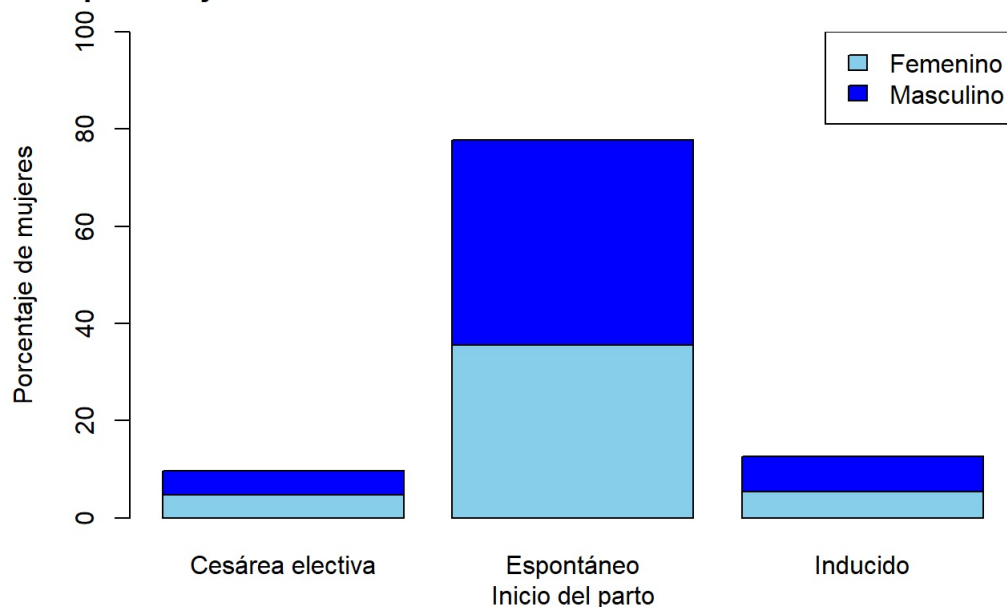
barplot(tab.porc, col=c("skyblue","blue"), ylim=c(0,100), main = strsplit(c("Forma de inicio del parto se-
gún el sexo del recién nacido; para mujeres atendidas en la Maternidad Martin durante el año 2014"), spli-
t = ";"), ylab = "Porcentaje de mujeres", sub="Fuente: Elaboración propia a partir de datos suministrados
\npor la secretaria de salud pública de la ciudad de Rosario")

mtext("Inicio del parto", side=1 , line=2)

legend("topright", legend = c("Femenino", "Masculino"), fill = c("skyblue","blue"))

```

Forma de inicio del parto según el sexo del recién nacido para mujeres atendidas en la Maternidad Martín durante el año 2014



Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

Gráfico de Bastones

A continuación se utiliza la variable “Partos previos” que es una variable cuantitativa discreta con 11 valores. De manera similar a la que se venía trabajando se contruye una tabla para la variable.

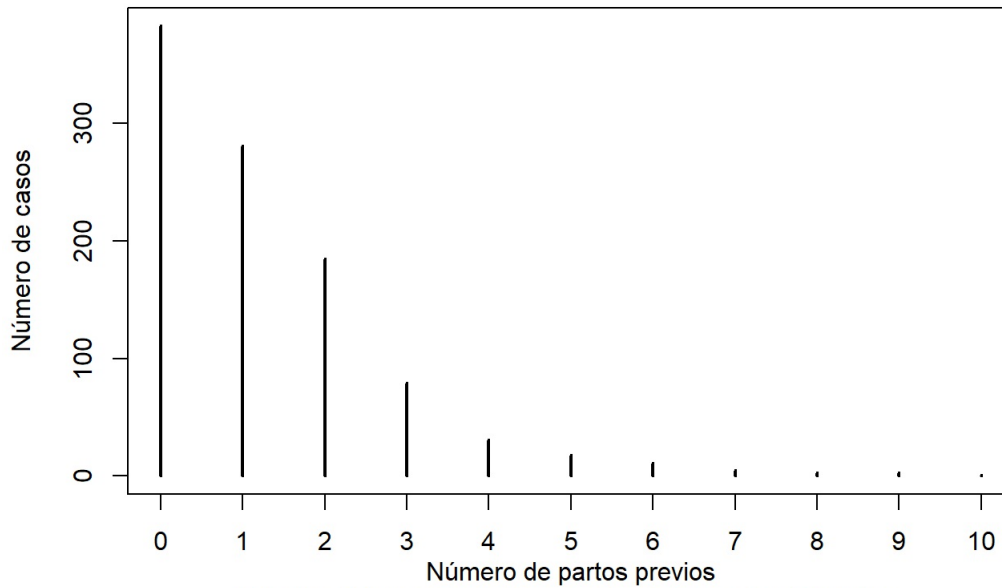
```
tab.partos <- table(`Partos previos`)
tab.partos
```

```
## Partos previos
##  0  1  2  3  4  5  6  7  8  9 10
## 383 281 185 79 31 18 11 5 3 3 1
```

```
plot(tab.partos, type="h", main = strsplit(c("Número de partos previos en mujeres atendidas; en la Maternidad Martín durante el año 2014"), split = ";"), xlab=NA, ylab = "Número de casos", sub="Fuente: Elaboración propia a partir de datos suministrados \npor la secretaría de salud pública de la ciudad de Rosario")
```

```
mtext("Número de partos previos", side=1 , line=2)
```


Número de partos previos en mujeres atendidas en la Maternidad Martín durante el año 2014



Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

Si se hace uso de `mtext()` para el caso de la función `plot()` se va a tener que aclarar dentro del argumento de esta última que el "label de x es igual a NA", es decir "`xlab=NA`". Si no se realiza este paso el programa colocará el nombre de la variable utilizada.

Nota 1: Recordar que la función `mtext()` se utiliza en este caso para agregar el "label" del eje x para que este no se solape con la fuente. **Nota 2:** R denomina con "NA"-abreviatura de "Not Available"- a los valores perdidos.

Gráfico Escalonado

A continuación, se grafica la distribución acumulada para la variable "Partos Previos". Para este caso se necesita que los datos estén en formato matriz, para ello se utiliza la función "`as.matrix()`" que tendrá como argumento a la tabla "`tab.partos`" y "`ncol= 1`", esto indica el número de columnas de la matriz.

```
matriz.partos<-as.matrix(tab.partos, ncol = 1)

matriz.partos
```

```
##      [,1]
## 0    383
## 1    281
## 2    185
## 3     79
## 4     31
## 5     18
## 6     11
## 7      5
## 8      3
## 9      3
## 10     1
```

Por lo tanto, como se puede observar la matriz construida pasa a ser un vector de dimensión 11x1.

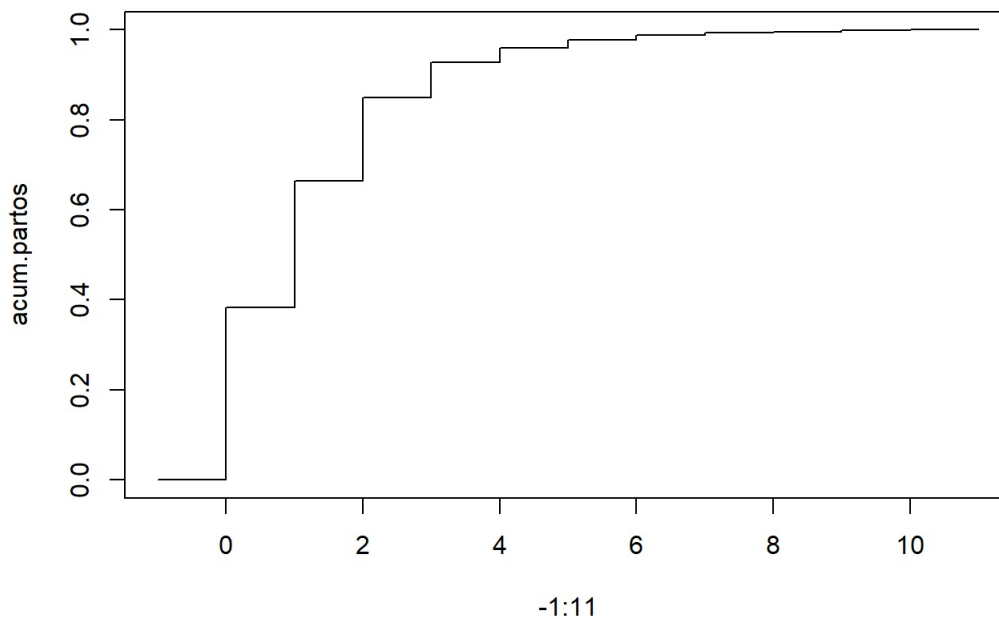
En el paso siguiente se calculan las proporciones de los valores presentes en la matriz y sus valores acumulados.

```
matriz.prop<-prop.table(matriz.partos)

acum.partos <- c(0, cumsum(matriz.prop), 1)
```

Se debe notar que en "`acum.partos`" se concatenan dos valores, el cero y el uno, esto se debe a la forma en que R construye el gráfico escalonado.

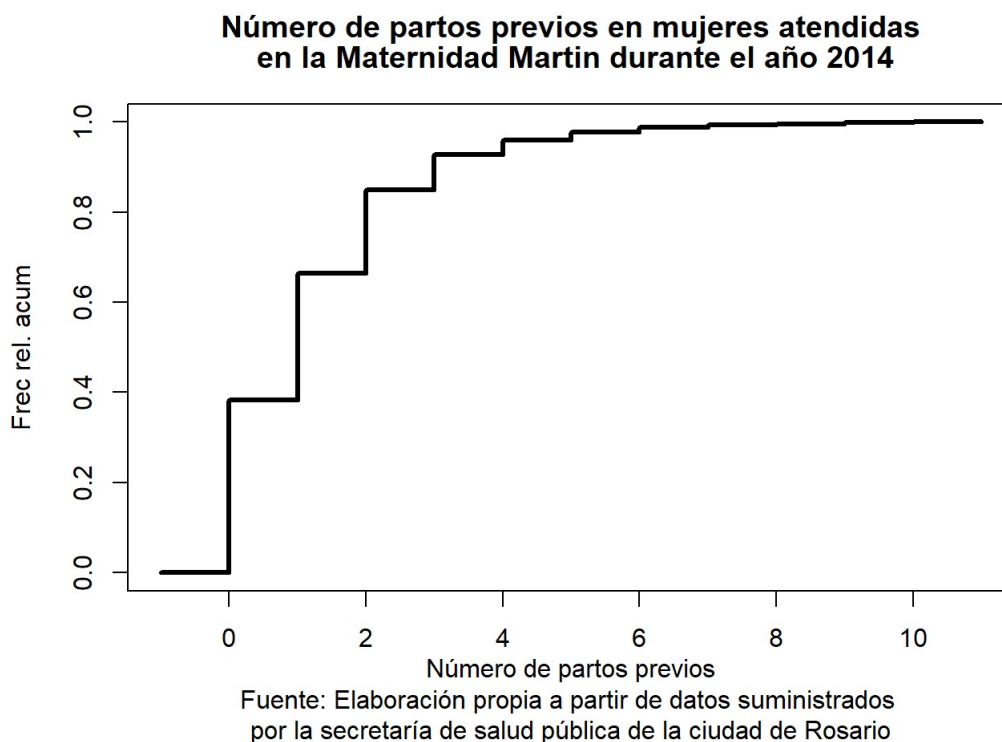
```
plot(-1:11, acum.partos, type = "s")
```



Para este gráfico se especifica que el rango del eje X es de (-1) a 11, se define de esta manera para que el gráfico coincida con la teoría dada. Además, con “**type=**”s (s minúscula) se le pide que el gráfico sea escalonado.

```
plot(-1:11, acum.partos, type = "s", lwd=3,
     main = strsplit(c("Número de partos previos en mujeres atendidas; en la Maternidad Martin durante el
año 2014"), split = ";"), xlab=NA, ylab = "Frec rel. acum", sub="Fuente: Elaboración propia a partir de d
atos suministrados \npor la secretaria de salud pública de la ciudad de Rosario")

mtext("Número de partos previos", side=1, line=2)
```



Con la sentencia “*lwd=3*” cambiamos el grosor de la línea graficada.

Diagrama de sectores

En este caso se toma la variable *Sexo del recién nacido* para construir un gráfico de sectores. Como en los casos anteriores se construye una tabla que resume los porcentajes para cada categoría.

```
tab.sexorn <- prop.table(table(`Sexo del RN`))*100
```

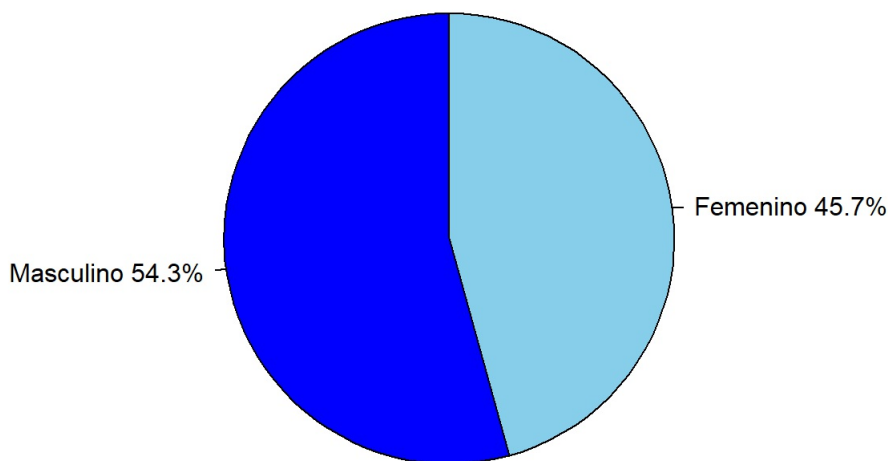
```
tab.sexorn
```

```
## Sexo del RN
## Femenino Masculino
##      45.7      54.3
```

```
etiquetas<- c("Femenino", "Masculino")
etiquetas <- paste(etiquetas, tab.sexorn)
etiquetas <- paste(etiquetas,"%", sep="")
```

```
pie(tab.sexorn,labels=etiquetas, edges=100, radius=1,2 ,angle=45, col = c("skyblue","blue"),
    main = strsplit(c("Sexo del recién nacido para hijos de mujeres atendidas; en la Maternidad Martin du
rante el año 2014"), split = ";"), sub="Fuente: Elaboración propia a partir de datos suministrados \npor
la secretaria de salud pública de la ciudad de Rosario")
```

Sexo del recién nacido para hijos de mujeres atendidas en la Maternidad Martin durante el año 2014



Fuente: Elaboración propia a partir de datos suministrados por la secretaria de salud pública de la ciudad de Rosario

En la sentencia “**labels=**” se indican los nombres que se les van a asignar a cada porción del círculo. En primera instancia, dichos nombres fueron guardados en un objeto denominado *etiquetas*, al cuál también se le asignan los porcentajes correspondientes a “Femenino” y “Masculino” utilizando la función **paste()** que sirve para concatenar vectores.

Histograma

Para este ejemplo se utiliza la variable “*Edad de la madre*”. Antes de realizar el gráfico se deben tener en cuenta las reglas de construcción para determinar el número de barras. Para ello se calcula el rango de la variable y se obtiene la diferencia entre los límites del rango.

```
range(`Edad de la madre`)
```

```
## [1] 14 42
```

```
Clases= 42-14
Clases
```

```
## [1] 28
```

Al contar con 28 edades diferentes (clases) y tomando 10 barras, se obtiene que la amplitud de cada intervalo será aproximadamente 3, como se muestra a continuación.

```
Amplitud= 28/10
Amplitud
```

```
## [1] 2.8
```

Antes de graficar los datos se debe presentar una tabla de frecuencias pero como R no cuenta con una sentencia que realice una tabla similar a la vista en clase se crea una función para poder obtenerla. En el siguiente comando se construye una función denominada **Hist_tabla**, esta se podrá usar en cualquier otro código siempre que sea definida antes, es decir que se corra el siguiente comando en el programa.

```
Hist_tabla<- function (variable,min,max,amplitud){
  factorx <- factor(cut(variable, breaks=seq(min,max,amplitud), right = FALSE))
  df_tabla <- as.data.frame(table(factorx))
  relative <- prop.table(df_tabla$Freq)
  df_tabla <- cbind(transform(df_tabla, cumFreq = cumsum(Freq), cum.prop=cumsum(relative)),relative)
  df_tabla <- df_tabla[,c(1,2,3,5,4)]

  colnames(df_tabla)<-c("Intervalos","Frecuencia absoluta","Frecuencia acumulada","Proporción","Acumulada")

  return (df_tabla)
}
```

Para construir la tabla se debe completar el argumento de la función. En primer lugar la variable, luego los valores mínimo y el máximo que esta toma y por último la amplitud del intervalo:*Hist_tabla(variable,min,max,amplitud)*

```
tab.histograma <-Hist_tabla(`Edad de la madre`, 14, 44, 3)
tab.histograma
```

```
##      Intervalos Frecuencia absoluta Frecuencia acumulada Proporción
## 1      [14,17)             61             61      0.061
## 2      [17,20)            178            239      0.178
## 3      [20,23)            218            457      0.218
## 4      [23,26)            164            621      0.164
## 5      [26,29)            119            740      0.119
## 6      [29,32)            104            844      0.104
## 7      [32,35)             75            919      0.075
## 8      [35,38)             50            969      0.050
## 9      [38,41)             19            988      0.019
## 10     [41,44)             12           1000      0.012
##      Acumulada
## 1      0.061
## 2      0.239
## 3      0.457
## 4      0.621
## 5      0.740
## 6      0.844
## 7      0.919
## 8      0.969
## 9      0.988
## 10     1.000
```

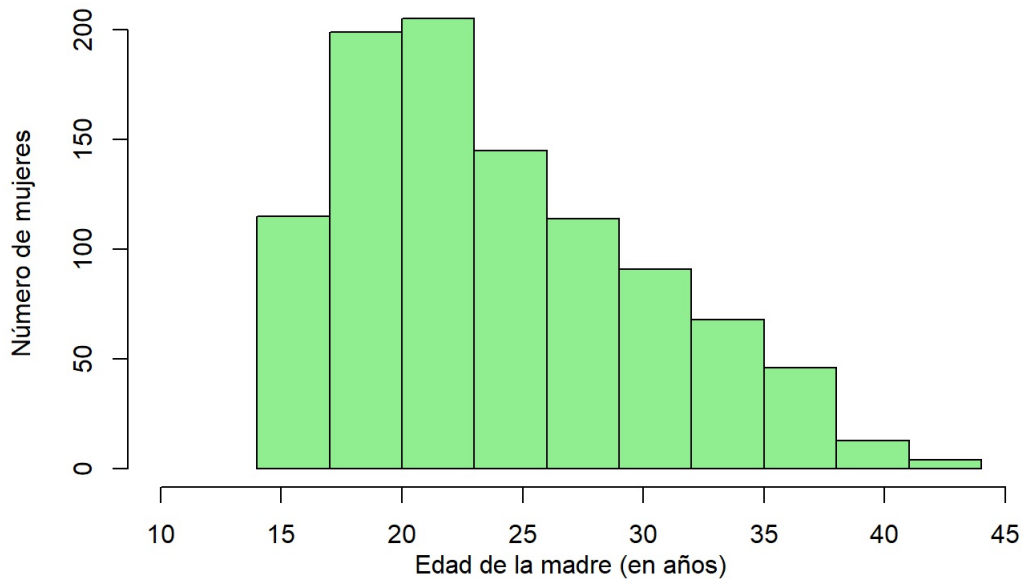
IMPORTANTE Por como se construye la función los intervalos se definen como abiertos por derecha, es decir que son del estilo “[a,b)”, debido a esto se debe indicar un valor “max” superior al máximo de la variable para que todos los valores sean incluidos en la tabla.

Construcción del histograma

```
hist(`Edad de la madre`,border="black", col="lightgreen",xlim=c(10, 44), breaks=c(seq(14, 44, 3)),
     xlab=NA, ylab="Número de mujeres",main = strsplit(c("Edad (en años) de las mujeres atendidas en la
Maternidad Martin; durante el año 2014"), split = ";"), sub="Fuente: Elaboración propia a partir de datos
suministrados \npor la secretaría de salud pública de la ciudad de Rosario" )

mtext("Edad de la madre (en años)", side=1 , line=2)
```

Edad (en años) de las mujeres atendidas en la Maternidad Martin durante el año 2014



Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

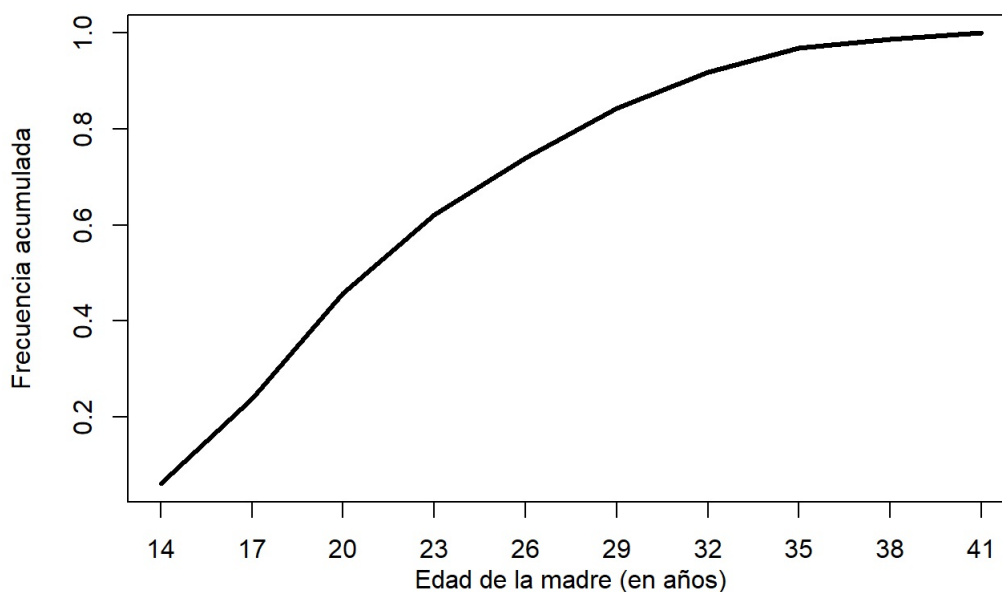
Poligono acumulativo (frecuencias relativas o absolutas acumuladas)

Para este gráfico se necesita la Frecuencia Relativa Acumulada de la variable que está guardada en la columna "Acumulada" de la tabla construida a partir de la función "Hist_tabla".

```
plot(tab.histograma$Acumulada, type="l", xaxt="n", lwd=3, xlab=NA, ylab="Frecuencia acumulada", main = str
split(c("Edad (en años) de las mujeres atendidas en la Maternidad Martin; durante el año 2014"), split =
";"), sub="Fuente: Elaboración propia a partir de datos suministrados \npor la secretaría de salud públic
a de la ciudad de Rosario" )

mtext("Edad de la madre (en años)", side=1, line=2)
valores_eje_X=c(14, 17, 20, 23, 26, 29, 32, 35, 38, 41)
axis(1, at=1:10, labels =valores_eje_X)
```

Edad (en años) de las mujeres atendidas en la Maternidad Martin durante el año 2014



Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

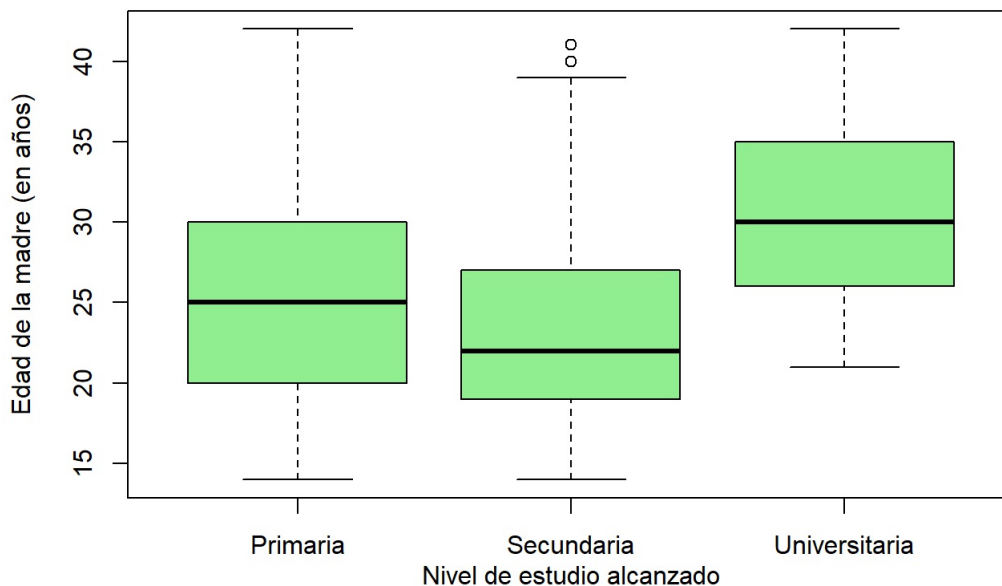
Box-plot

En este caso se usan las variables *Edad de la madre* y *Nivel de estudios*.

```
boxplot(`Edad de la madre`~Estudios, data=Base, col="lightgreen", main=strsplit("Edad (en años) de las mu  
jeres atendidas en la Maternidad Martin; según nivel de estudio alcanzado durante año 2014", split = ";")  
, xlab=NA, ylab="Edad de la madre (en años)", sub="Fuente: Elaboración propia a partir de datos suministr  
ados \npor la secretaría de salud pública de la ciudad de Rosario")
```

```
mtext("Nivel de estudio alcanzado", side=1 , line=2)
```

Edad (en años) de las mujeres atendidas en la Maternidad Martin según nivel de estudio alcanzado durante año 2014



Fuente: Elaboración propia a partir de datos suministrados
por la secretaría de salud pública de la ciudad de Rosario

Visualización gráfica con ggplot2

El paquete **ggplot2** fue creado para diseñar gráficos más complejos, los cuales definen un *Objeto geométrico* que se indica con *geoms* y es donde que especifica que gráfico se va a construir.

Instalación y carga del paquete

```
install.packages("ggplot2", repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/Eugenia/Documents/R/win-library/3.4'  
## (as 'lib' is unspecified)
```

```
## package 'ggplot2' successfully unpacked and MD5 sums checked  
##  
## The downloaded binary packages are in  
## C:\Users\Eugenia\AppData\Local\Temp\Rtmpa0evqf\downloaded_packages
```

```
library(ggplot2)
```

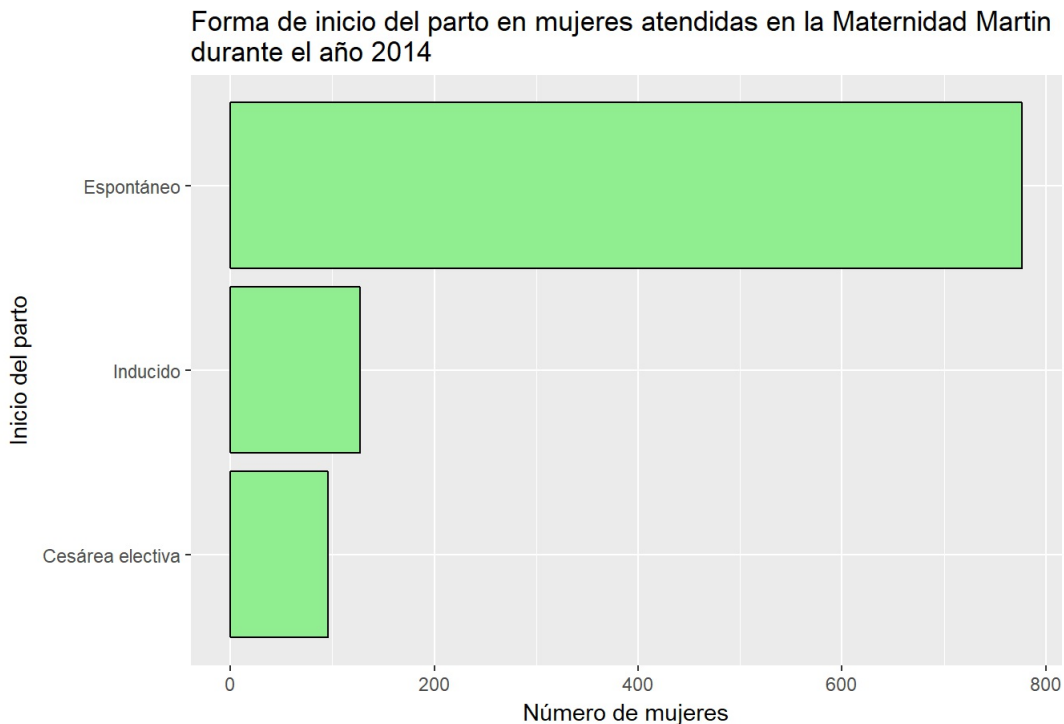
```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

Construcción del gráfico

```
gr1 <- ggplot(Base, aes(x=`Inicio del parto`))

gr1_1 <- gr1 + geom_bar(fill="light green", colour="black")
gr1_2 <- gr1_1 + xlab("Inicio del parto") + ylab("Número de mujeres")
gr1_3 <- gr1_2 + ggtitle("Forma de inicio del parto en mujeres atendidas en la Maternidad Martin \ndurant
e el año 2014") +
  labs(caption=("Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pú
blica de la ciudad de Rosario"))
gr1_4 <- gr1_3 + scale_x_discrete(limits= c("Cesárea electiva", "Inducido", "Espontáneo")) + coord_flip()

gr1_4
```



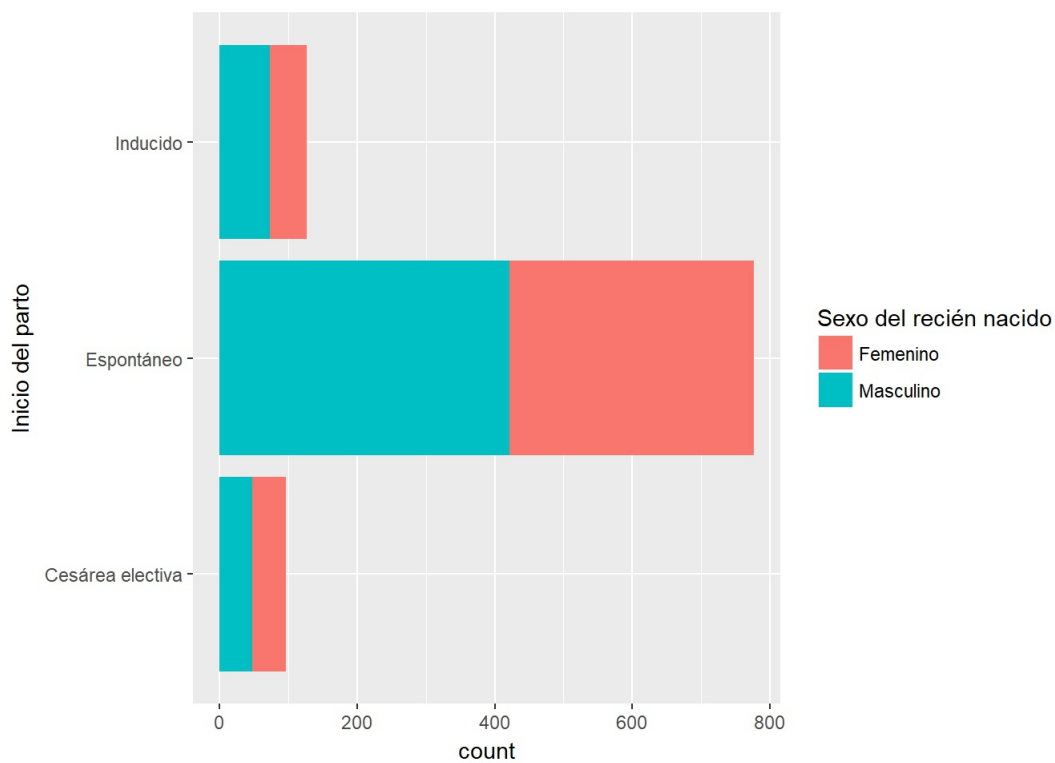
Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

Gráfico de barras compuesto

En este ítem se construye un gráfico de barras compuesto con las variables “*Inicio de parto*” y “*Sexo del recién nacido*”.

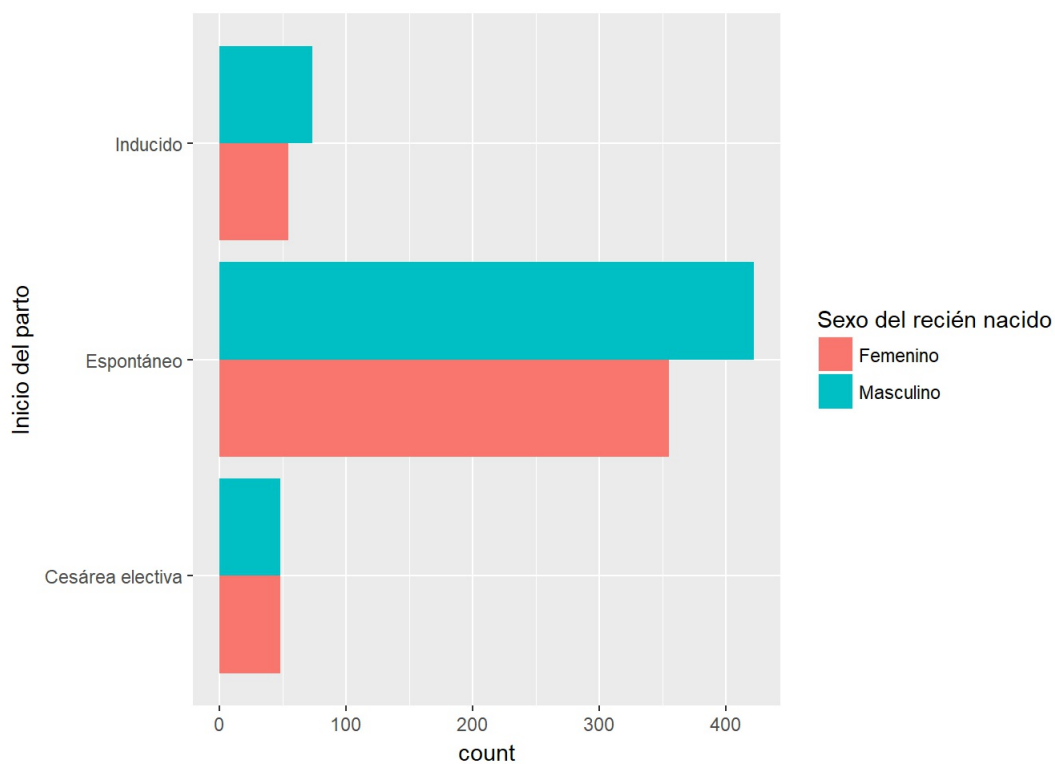
```
gr2<- ggplot(Base, aes(x=`Inicio del parto`, fill=`Sexo del RN`)) + coord_flip() + geom_bar() + scale_fi
ll_discrete(name="Sexo del recién nacido")

gr2
```



```
gr2_1<- ggplot(Base, aes(x=`Inicio del parto`, fill=`Sexo del RN`)) + coord_flip() + geom_bar(position = "dodge") + scale_fill_discrete(name="Sexo del recién nacido")
```

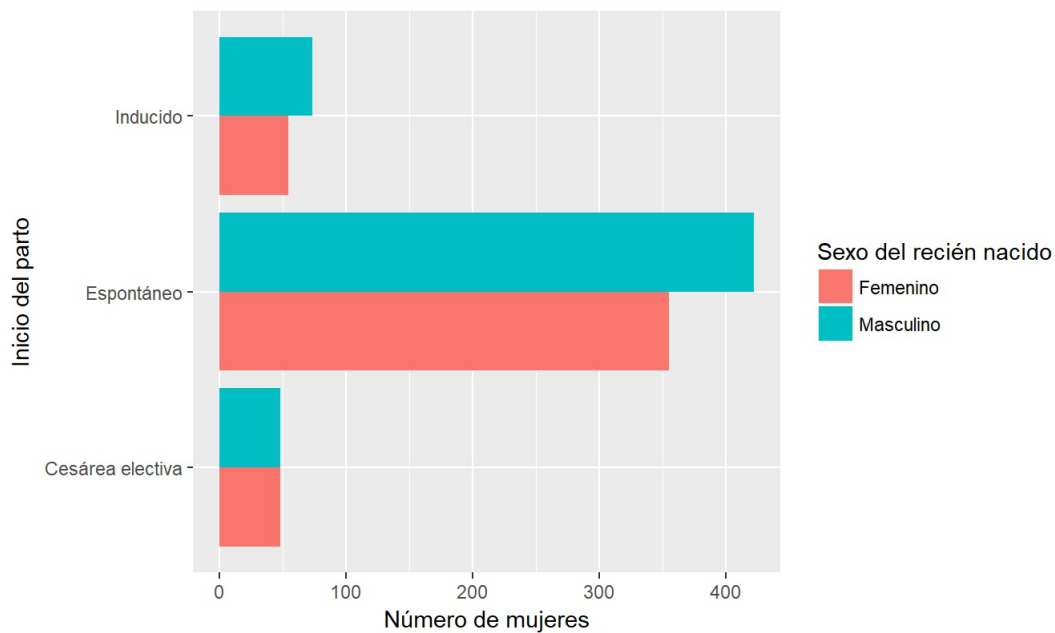
gr2_1



```
gr2_2 <- gr2_1+ xlab("Inicio del parto")+ ylab("Número de mujeres") +
  ggtitle("Forma de inicio del parto según el sexo del recién nacido \npara mujeres atendidas en la Maternidad Martin durante el año 2014") +
  labs(caption=("Fuente: Elaboración propia a partir de datos \nsuministrados por la secretaría de salud pública de la ciudad de Rosario"))
```

gr2_2

Forma de inicio del parto según el sexo del recién nacido para mujeres atendidas en la Maternidad Martin durante el año 2014



Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

Otro ejemplo

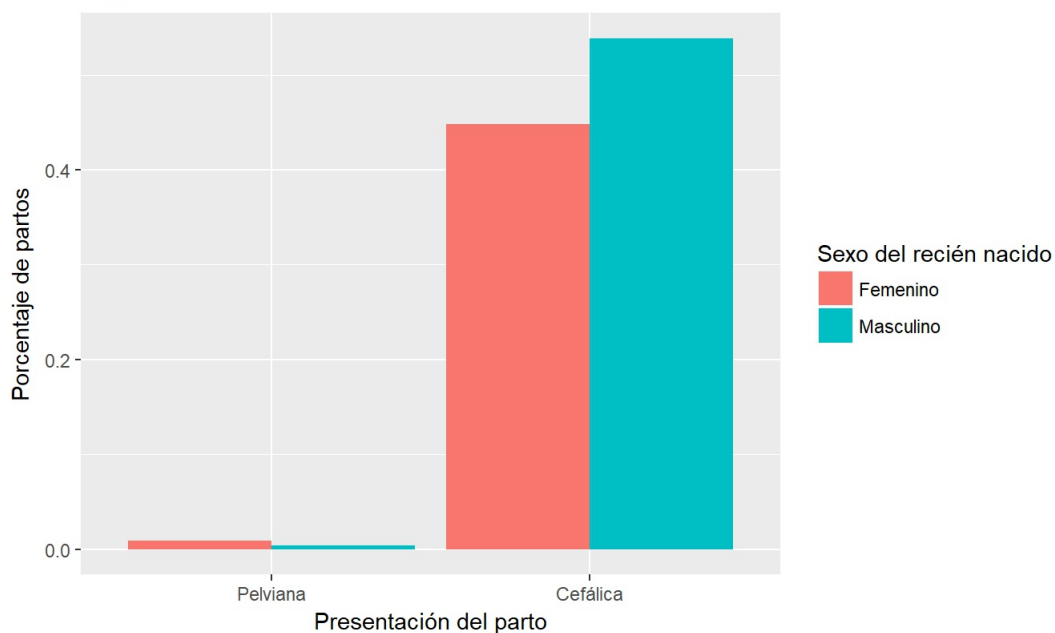
En este caso se grafica la variable "Presentación" según "Sexo del recién nacido"

```
gr3 <-ggplot(Base, aes(Presentacion, fill=`Sexo del RN`)) +
  geom_bar(aes(Presentacion, (..count../sum(..count..)), position="dodge"))

gr3_1 <- gr3 + xlab("Presentación del parto") + ylab("Porcentaje de partos") +
  ggtitle("Presentación del parto para mujeres atendidas en la Maternidad Martin \nsegún el sexo del recién nacido durante el año 2014")+
  labs(caption=("Fuente: Elaboración propia a partir de datos suministrados \npor la secretaría de salud pública de la ciudad de Rosario"))+
  scale_fill_discrete(name="Sexo del recién nacido")+
  scale_x_discrete(limits=c("Pelviana", "Cefálica"))

gr3_1
```

Presentación del parto para mujeres atendidas en la Maternidad Martin según el sexo del recién nacido durante el año 2014

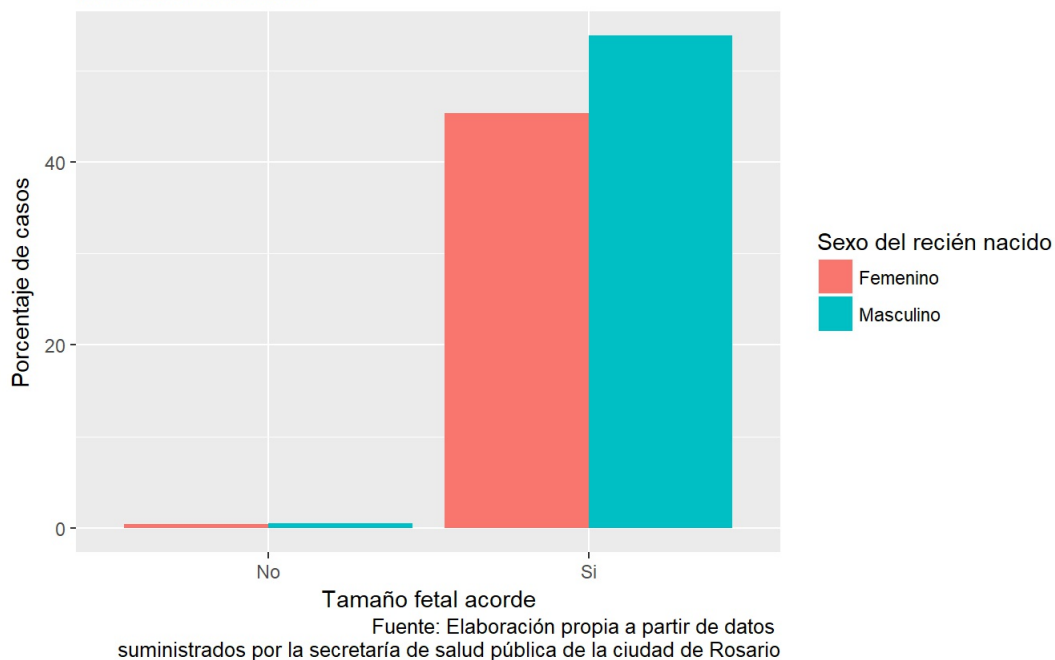


Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

Gráfico de barras porcentuales

```
gr4 <-ggplot(Base, aes(`Tamaño fetal acorde`, fill=`Sexo del RN`)) +  
  geom_bar(aes(`Tamaño fetal acorde`, 100*(..count..)/sum(..count..)), position="dodge")  
  
gr4_1 <- gr4 + xlab("Tamaño fetal acorde") + ylab("Porcentaje de casos") +  
  ggtitle("Tamaño fetal acorde según el sexo del recién nacido \npara los bebes de mujeres atendidas en l  
a Maternidad Martin \ndurante el año 2014")+  
  labs(caption=("Fuente: Elaboración propia a partir de datos \nsuministrados por la secretaría de salud  
pública de la ciudad de Rosario"))+  
  scale_fill_discrete(name="Sexo del recién nacido")  
  
gr4_1
```

**Tamaño fetal acorde según el sexo del recién nacido
para los bebes de mujeres atendidas en la Maternidad Martin
durante el año 2014**

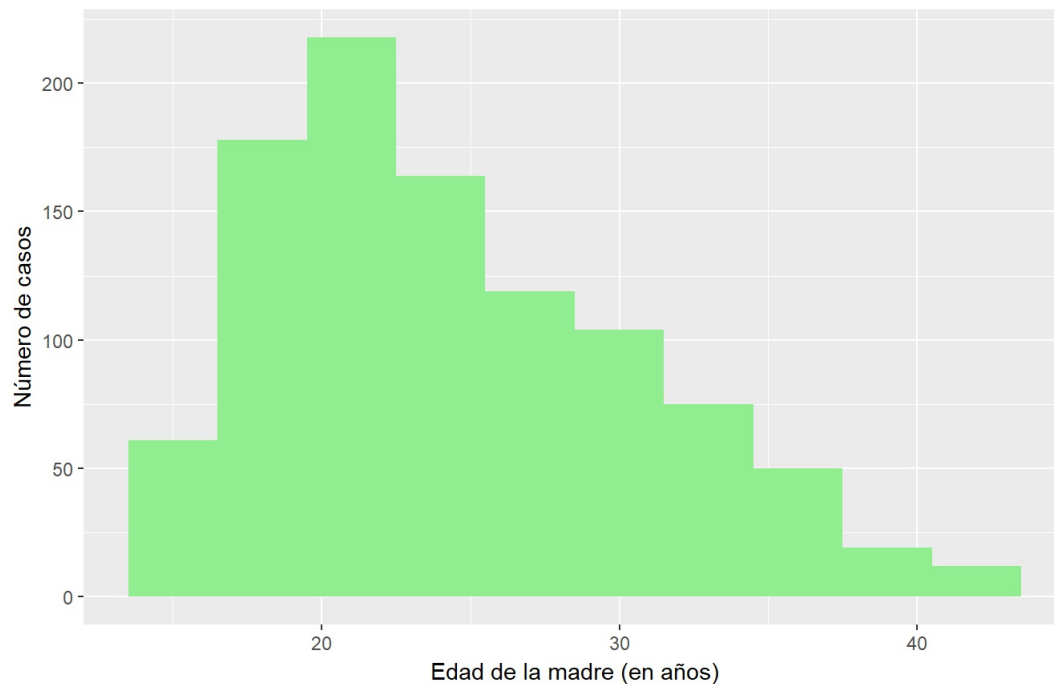


Histograma

Teniendo en cuenta la cantidad de intervalos y la amplitud previamente que se calculo anteriormente

```
gr6<-ggplot(Base, aes(x=`Edad de la madre`)) +  
  geom_histogram(binwidth = 3,fill="lightgreen")+ xlab("Edad de la madre (en años)") +  
  ylab("Número de casos") + ggtitle("Edad (en años) de las mujeres atendidas en la Maternidad Martin; dur  
ante el año 2014") +  
  labs(caption=("Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pú  
blica de la ciudad de Rosario"))  
  
gr6
```

Edad (en años) de las mujeres atendidas en la Maternidad Martin; durante el año 2014



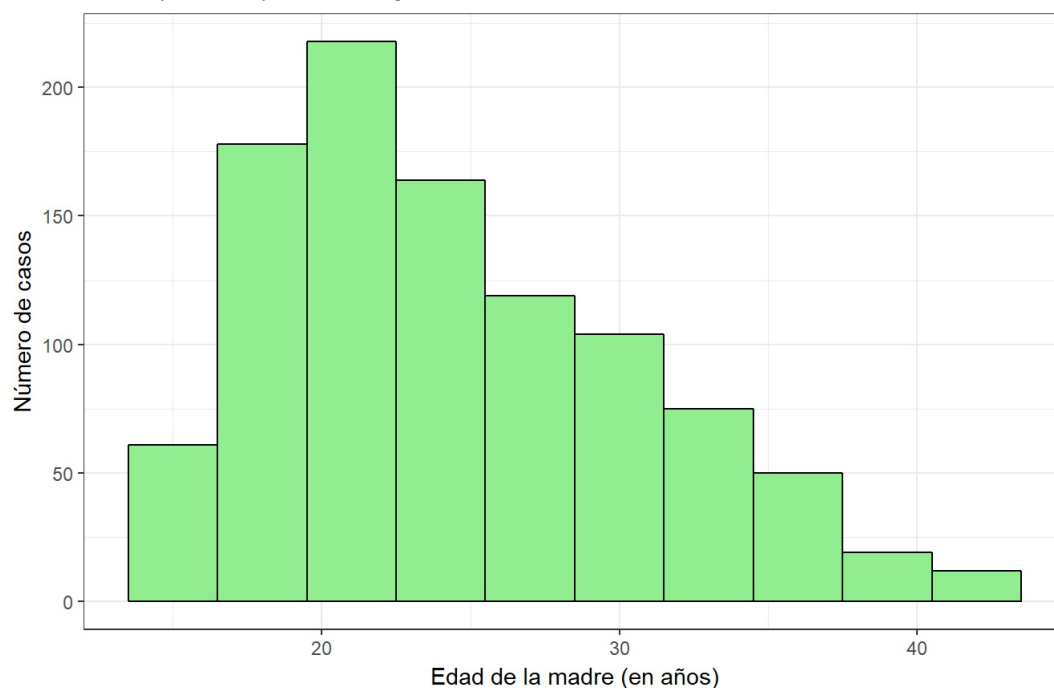
Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

Los bordes de las barras en *gr4_1* no se encuentran marcados, por lo que en *gr4_2* se procede a cambiar esto utilizando la sentencia **colour="black"** dentro de **geom_histogram()**.

```
gr6_1<- ggplot(Base, aes(x=`Edad de la madre`)) +
  geom_histogram(binwidth = 3,fill="lightgreen", colour="black")+ xlab("Edad de la madre (en años)") +
  ylab("Número de casos") + ggtitle("Edad (en años) de las mujeres atendidas en la Maternidad Martin; durante el año 2014") +
  labs(caption=("Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario")) + theme_bw()

gr6_1
```

Edad (en años) de las mujeres atendidas en la Maternidad Martin; durante el año 2014



Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

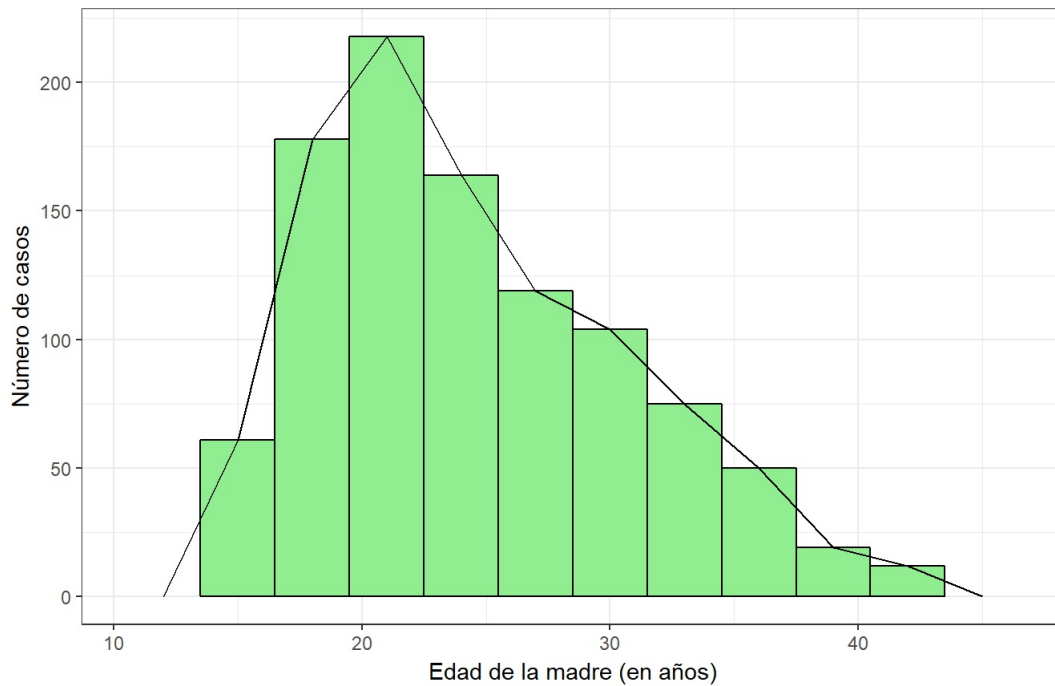
Además, se agregó la sentencia **theme_bw()**, que permite cambiar el color del fondo del gráfico a blanco.

Histograma con polígono de frecuencias

```
gr7<- gr6_1 + geom_freqpoly(binwidth=3)
```

```
gr7
```

Edad (en años) de las mujeres atendidas en la Maternidad Martin; durante el año 2014



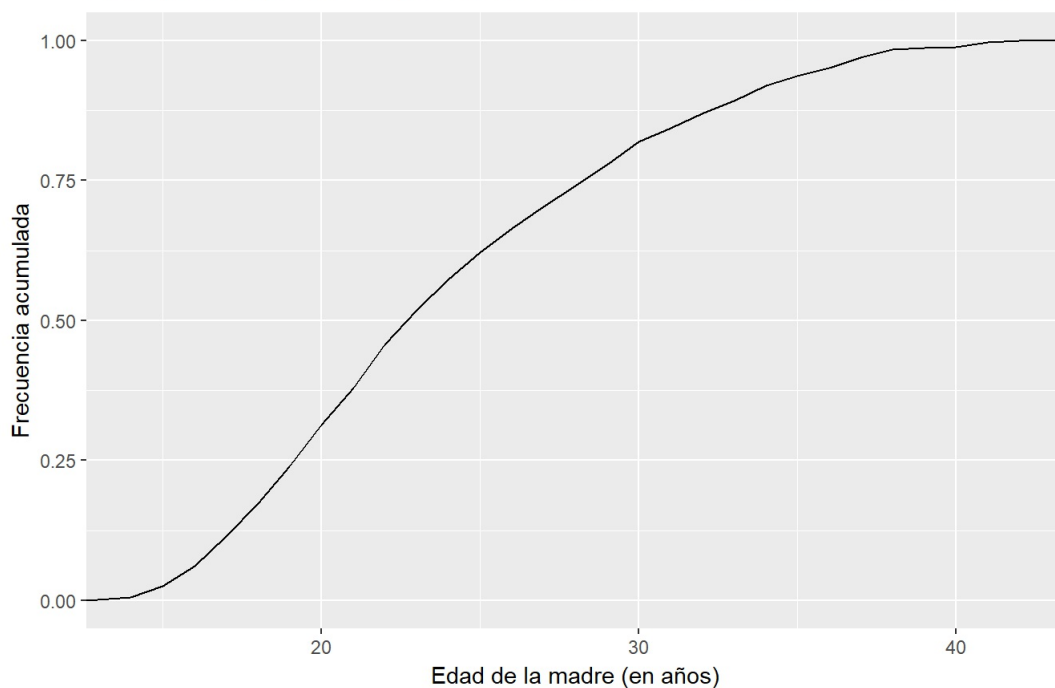
Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

Polígono acumulativo (frecuencias relativas o absolutas acumuladas)

```
gr8 <- ggplot(Base, aes(x=`Edad de la madre`)) + geom_line(stat = "ecdf") + xlab("Edad de la madre (en años)") +
  ylab("Frecuencia acumulada") + ggtitle("Edad (en años) de las mujeres atendidas en la Maternidad Martin ; durante el año 2014") +
  labs(caption=("Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario"))
```

```
gr8
```

Edad (en años) de las mujeres atendidas en la Maternidad Martin; durante el año 2014



Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

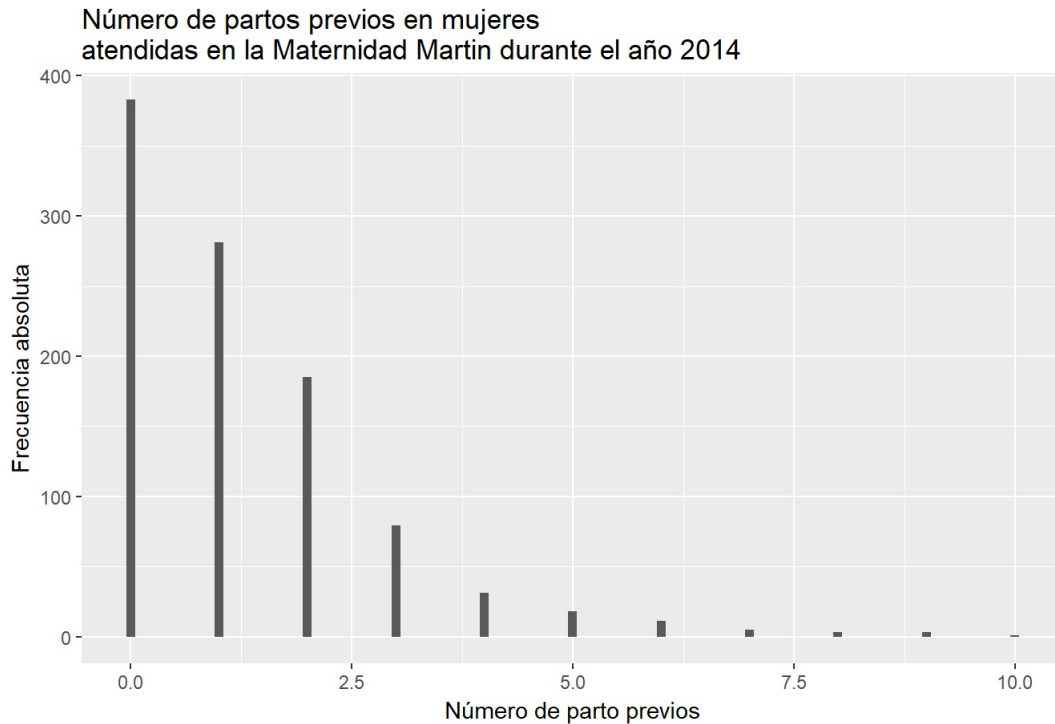
Gráfico de bastones

Para ello se vuelve a utilizar la variable *Partos previos*, se debe recordar que no es correcta la representación esta variable a través del gráfico de bastones, como fue aclarado en la sección anterior del apunte.

En este paquete los gráficos de bastones se pueden construir de la misma manera que los gráficos de barras modificando el ancho de las mismas de la siguiente manera **geom_bar(width = 0.1)**

```
gr9<-ggplot(Base, aes(`Partos previos`)) + geom_bar(width = 0.1)+
  xlab("Número de parto previos")+ ylab("Frecuencia absoluta") +
  ggtitle("Número de partos previos en mujeres \natendidas en la Maternidad Martin durante el año 2014")
+ labs(caption=("Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario"))
```

gr9



Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

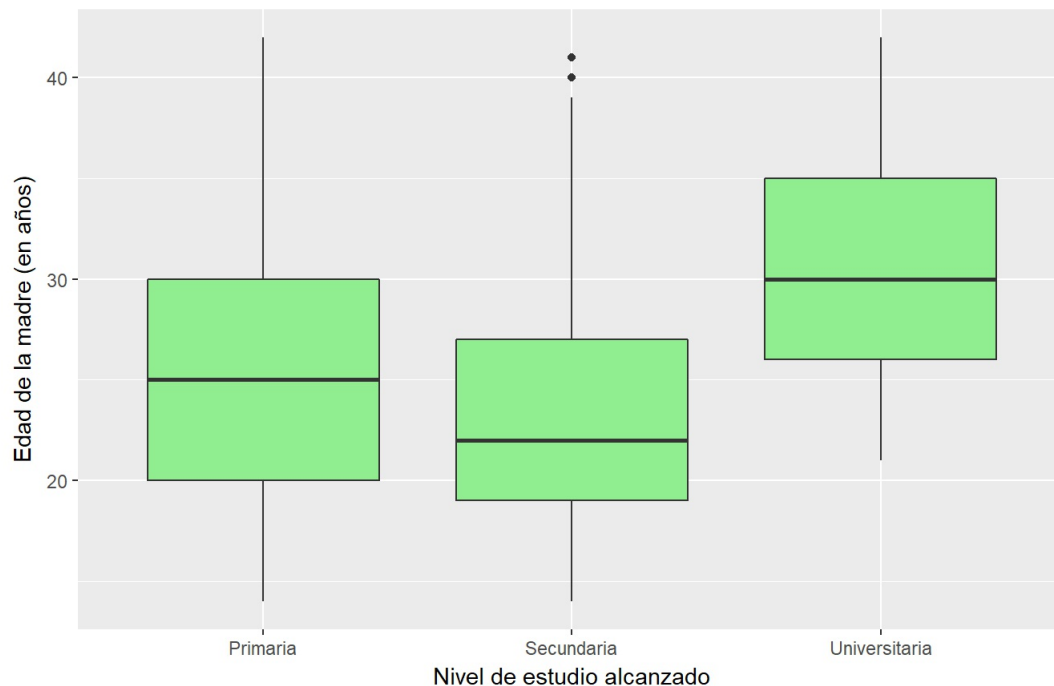
Gráfico Boxplot

Para este ejemplo tomamos las variables *Edad de la madre* y *Nivel de estudios*. En este caso es necesario especificar en **geom_boxplot()** la variable correspondiente a cada eje.

```
gr10<-ggplot(Base, aes(x=Estudios, y=`Edad de la madre`)) + geom_boxplot(fill="lightgreen") +
  xlab("Nivel de estudio alcanzado") + ylab("Edad de la madre (en años)") +
  ggtitle("Edad(en años) de las mujeres atendidas en la Maternidad Martin; según nivel de estudio alcanza  
do durante año 2014")+
  labs(caption=("Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario"))
```

gr10

Edad(en años) de las mujeres atendidas en la Maternidad Martin; según nivel de e:



Fuente: Elaboración propia a partir de datos suministrados por la secretaría de salud pública de la ciudad de Rosario

Tablas

A continuación se muestran algunos ejemplos de como construir tablas cruzadas porcentuales.

- Tomando las variables *Sexo del recién nacido* y *Presentación del parto*

```
tab.cruzada1<- as.data.frame(prop.table(table(`Sexo del RN`,Presentacion))*100)
names(tab.cruzada1)<-c("Sexo", "Presentación", "Porcentaje")
tab.cruzada1
```

```
##      Sexo Presentación Porcentaje
## 1  Femenino      Cefálica      44.8
## 2  Masculino      Cefálica      53.9
## 3  Femenino      Pelviana       0.9
## 4  Masculino      Pelviana       0.4
```

```
tab.present<- as.data.frame(prop.table(table(Base$Presentacion))*100)
names(tab.present)<-c("Presentación", "Porcentaje")
tab.present
```

```
##      Presentación Porcentaje
## 1      Cefálica      98.7
## 2      Pelviana       1.3
```

- Tomando las variables *Sexo del recién nacido* y *Tamaño fetal*

```
tab.cruzada2 <- as.data.frame(prop.table(table(`Sexo del RN`,`Tamaño fetal acorde`))*100)
names(tab.cruzada2)<-c("Sexo", "Tamaño fetal acorde", "Porcentaje")
tab.cruzada2
```

```
##      Sexo Tamaño fetal acorde Porcentaje
## 1  Femenino                No       0.4
## 2  Masculino                No       0.5
## 3  Femenino                Si      45.3
## 4  Masculino                Si      53.8
```

```
tab.tam<- as.data.frame(prop.table(table(`Tamaño fetal acorde`))*100)
names(tab.tam)<-c("Tamaño fetal acorde","Porcentaje")
tab.tam
```

##	Tamaño fetal acorde	Porcentaje
## 1	No	0.9
## 2	Si	99.1