Introducción al software estadístico

Módulo VII

Nicolás Schmidt

nschmidt@cienciassociales.edu.uy

Departamento de Ciencia Política
Unidad de Métodos y Acceso a Datos
Facultad de Ciencias Sociales
Universidad de la República

Estructura de la presentación

"Un gráfico puede valer más que mil palabras, pero puede tomar muchas palabras para hacerlo"

John Tukey

Estructura de la presentación

Estructura de la presentación

Gráficos del paquete `graphics`

Una de las fortalezas de R es su capacidad gráfica. Si bien el paquete `graphics` que viene por defecto con R es muy completo, por fuera de este paquete hay grandes desarrollos en términos de visualización de datos. Los paquetes `ggplot2` y `plotly` son un buen ejemplo de ello.

Hay dos tipos de ordenes gráficas:

- **Alto Nivel**: Son funciones que crean un gráfico.
  - `plot()`, `hist()`, `boxplot()`, `pairs()`,...
- **Bajo Nivel**: Son funciones que añaden información a un gráfico existente.
  - `points`, `lines`, `text`, `axis`, `legend`,...

Estructura de la presentación

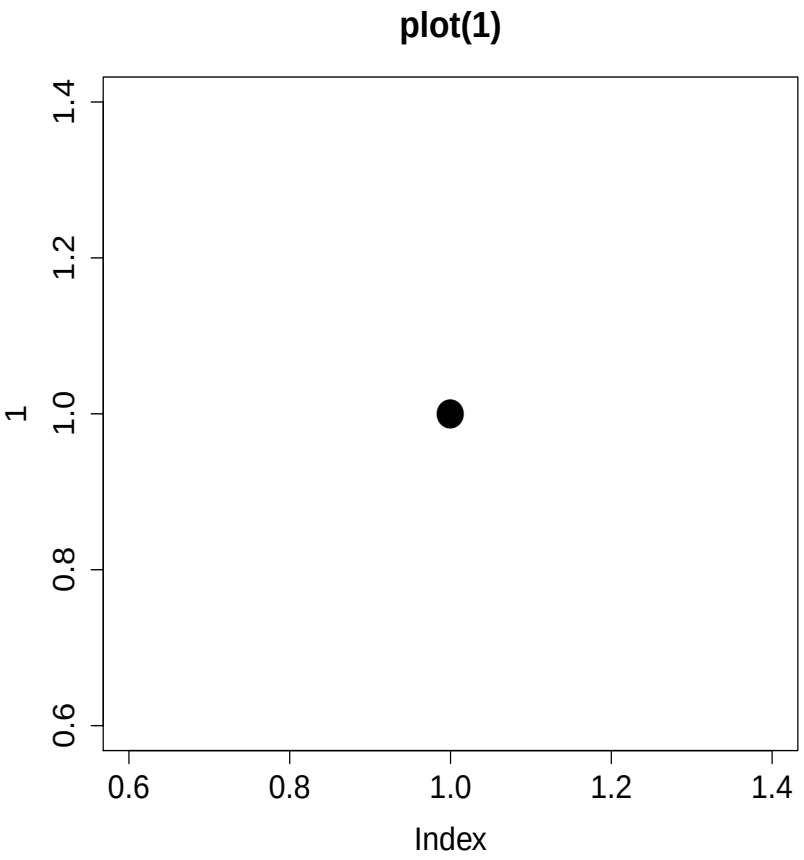
Funciones de Alto Nivel

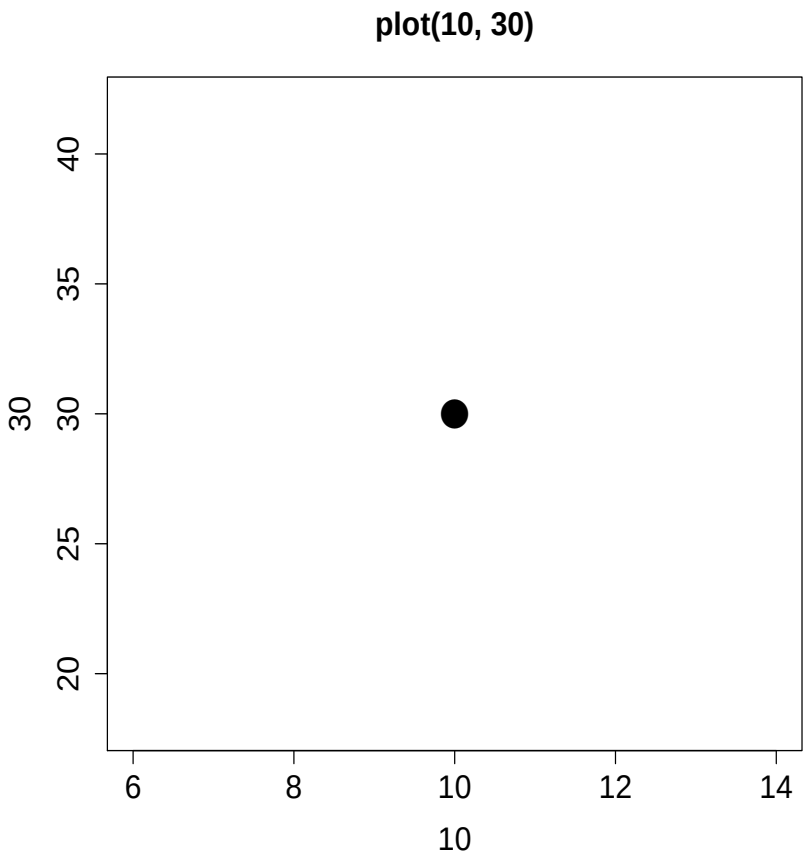
La función genérica para realizar un gráfico en R es `plot()`. Esta función genérica genera un gráfico en función del tipo o tipos de datos de entrada.

Si bien la función `plot` ayuda a visualizar la información lo que debe determinar el tipo de gráfico a utilizar es la estructura de los datos que se desea visualizar y el objetivo que se persigue con lo que se desea comunicar con los datos.

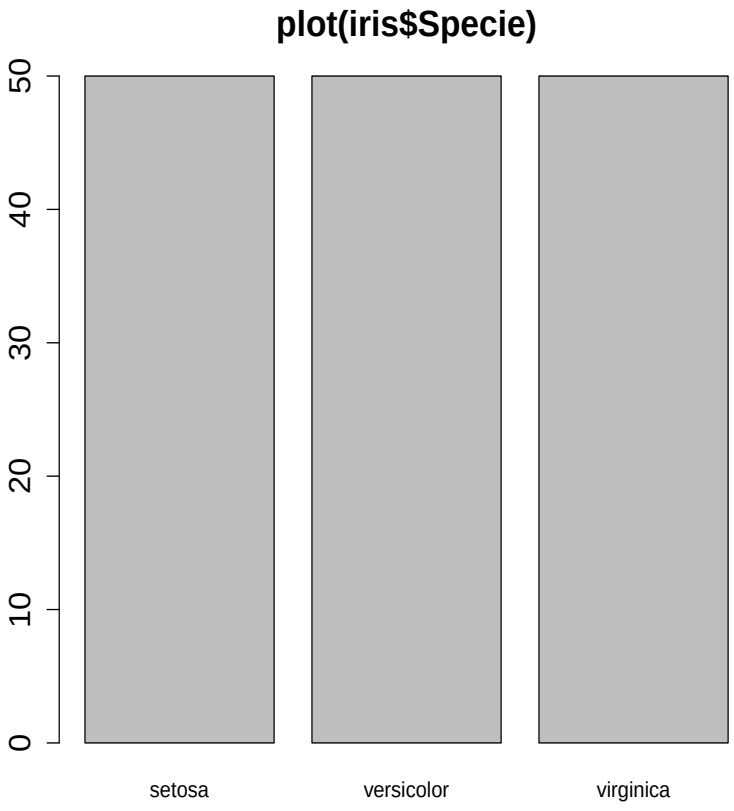
Los gráficos deben estar guiados **estadísticamente**, y luego, si es posible **estéticamente**. Nunca a la inversa!.

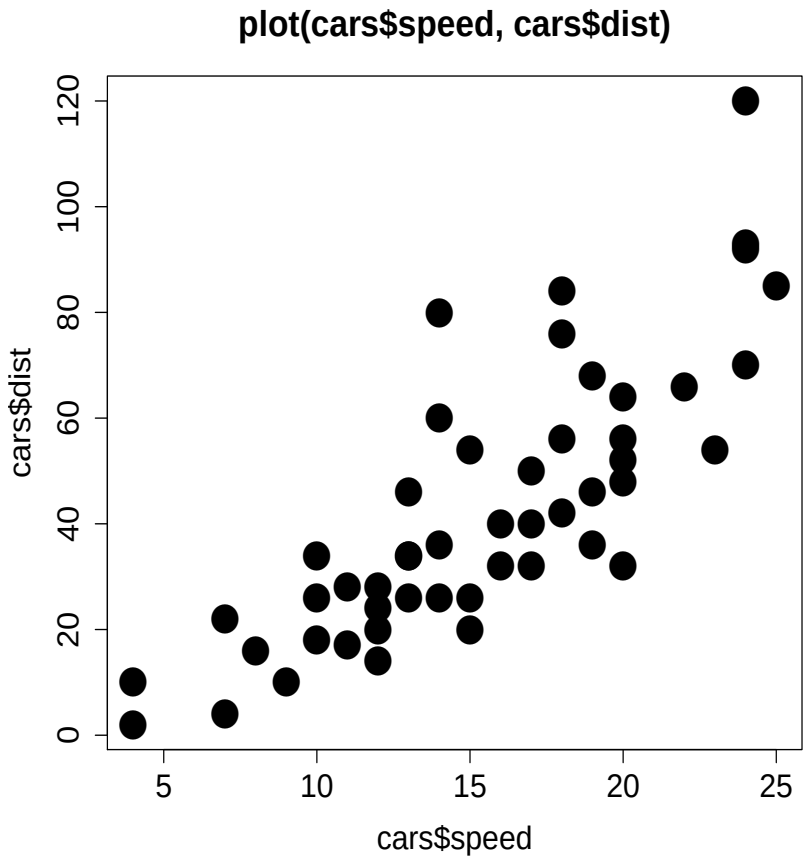
Funciones de Alto Nivel



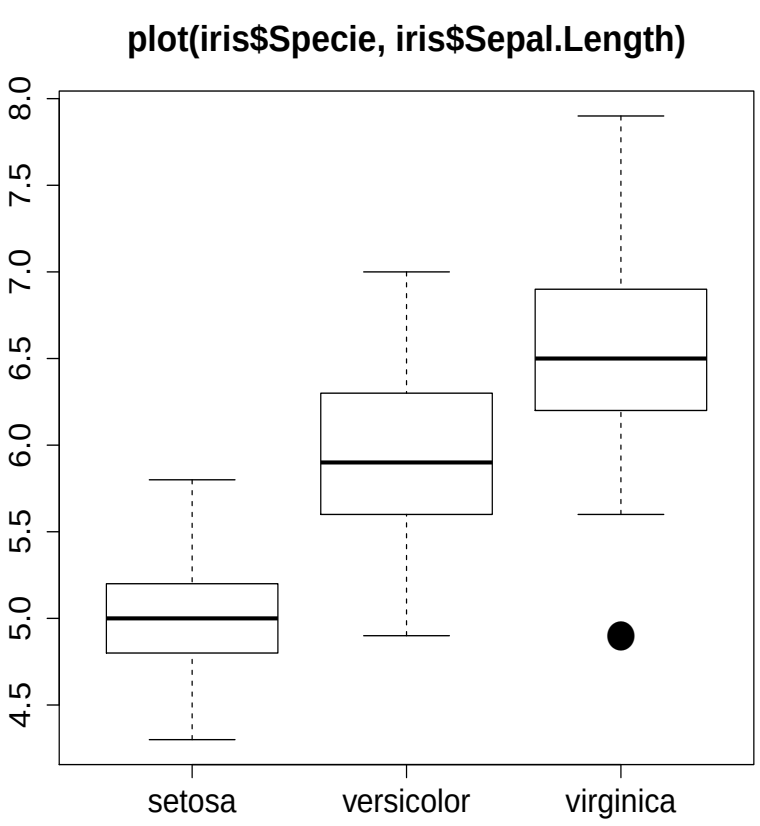


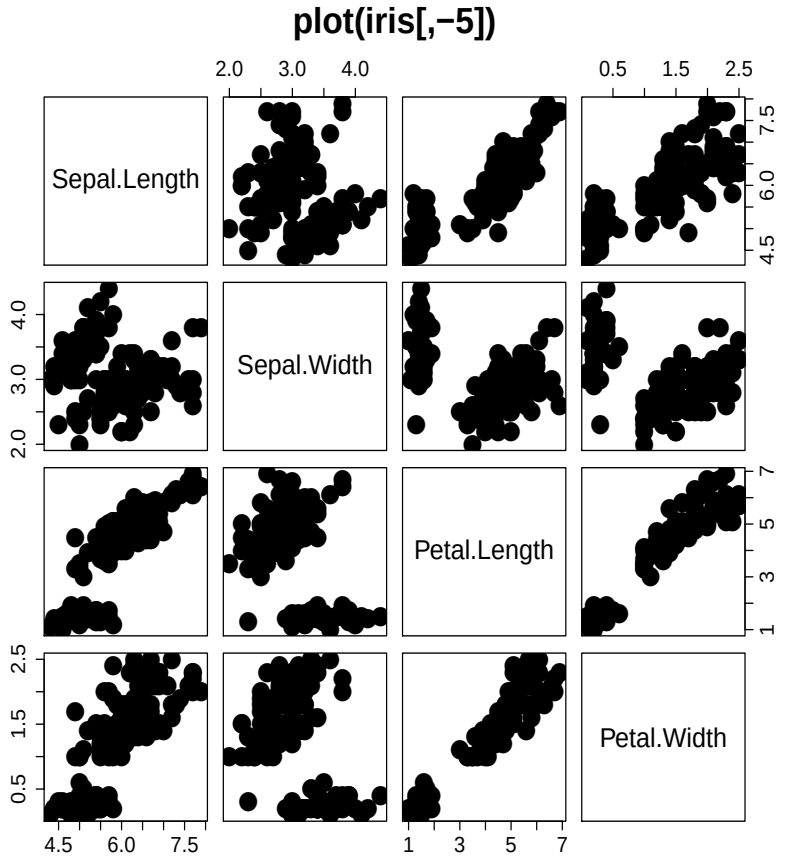
Funciones de Alto Nivel





Funciones de Alto Nivel



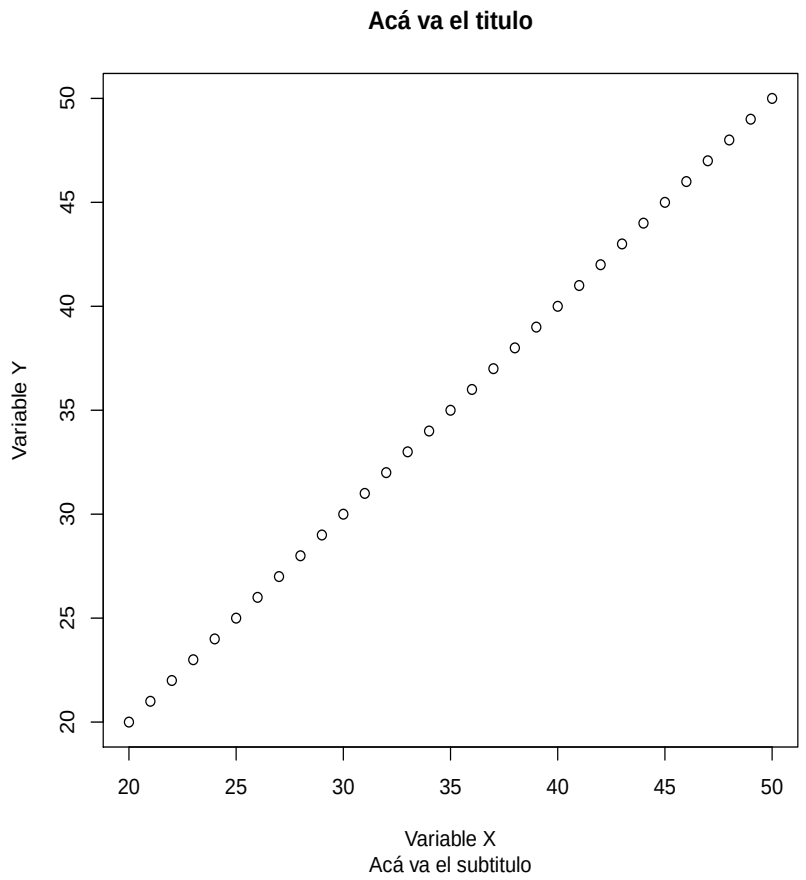


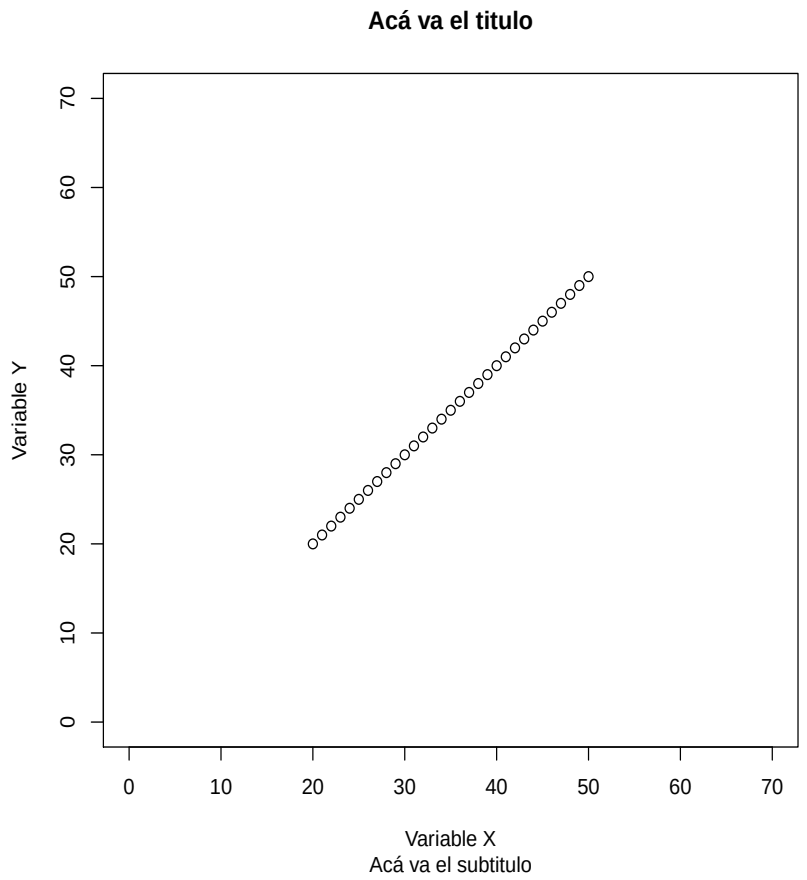
Argumentos de funciones de alto nivel

Argumento	Descripción	Ejemplo
`add = TRUE`	Si es `TRUE` añade un gráfico al existente	
`axes = TRUE`	Si es `FALSE` no incorpora ejes ni caja del gráfico	
`log = 'x'`	Logaritmo del eje x	
`log = 'y'`	Logaritmo del eje y	
`log = 'xy'`	Logaritmo de ambos ejes	
`xlab`	Títulos de los eje x	`xlab = 'Eje x'`
`ylab`	Títulos de los eje y	`ylab = 'Eje y'`
`xlim`	Limite inferior y superior del eje y	`xlim = c(1, 10)`
`ylim`	Limite inferior y superior del eje x	`ylim = c(1, 10)`
`main`	Titulo principal	`main = 'Titulo'`
`sub`	Subtitulo	`sub = 'subtitulo'`

Argumentos de funciones de alto nivel

```
plot(20:50, 20:50, main = "Acá va el titulo", sub = "Acá va el subtitulo",
     xlab ="Variable X", ylab = "Variable Y")
plot(20:50, 20:50, main = "Acá va el titulo", sub = "Acá va el subtitulo",
     xlab ="Variable X", ylab = "Variable Y", xlim = c(0, 70), ylim = c(0, 70))
```



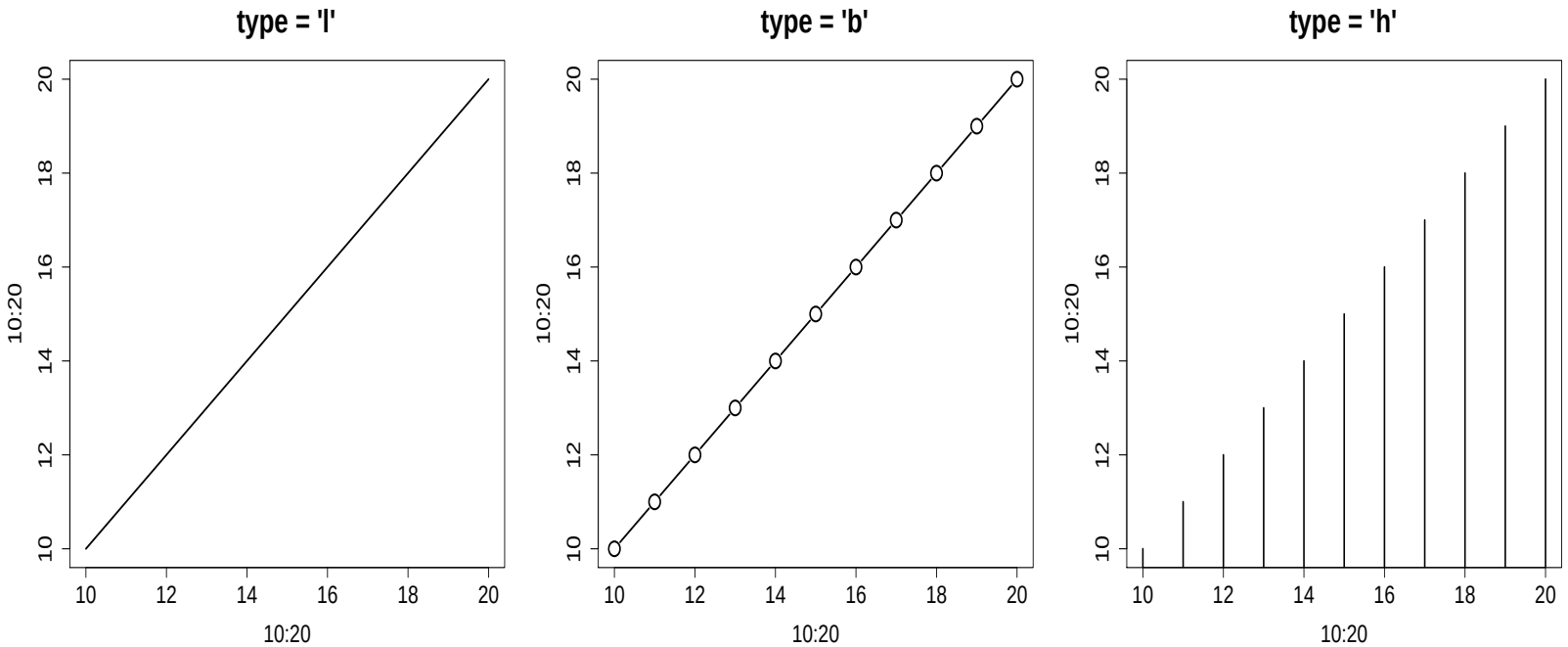


Argumento `type = "p"`

Tipo	Descripción
`type = "p"`	Puntos (valor por defecto)
`type = "l"`	Líneas
`type = "b"`	Puntos y lineas con espacio
`type = ".º"`	Puntos y lineas sin espacio
`type = "h"`	Lineas verticales
`type = "s"`	Escalera. Inicia hacia la derecha.
`type = "S"`	Escalera Inicia hacia arriba.
`type = "n"`	No se dibuja ningún gráfico, solo los ejes.

Argumentos de funciones de alto nivel

```r
plot(10:20, 10:20, type = "l", main = "type = 'l'")
plot(10:20, 10:20, type = "b", main = "type = 'b'")
plot(10:20, 10:20, type = "h", main = "type = 'h'")
```

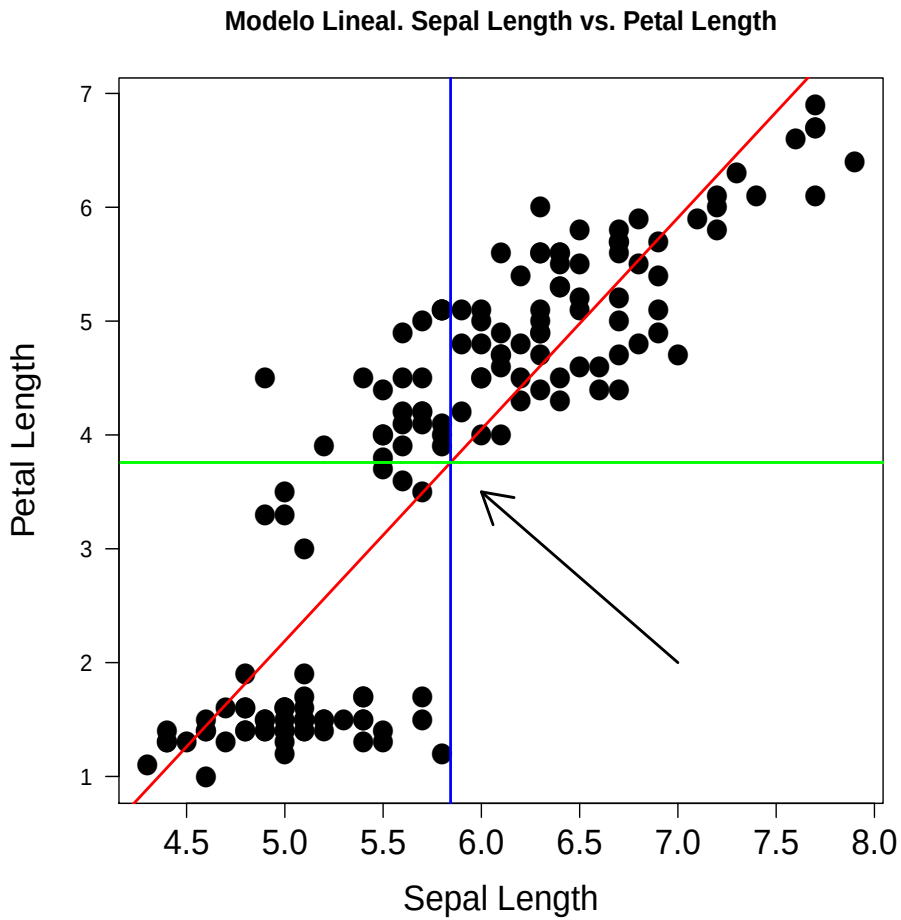


Estructura de la presentación

Argumentos de funciones de alto nivel

Función	Descripción
`points(x, y)`	Añade puntos
`lines(x, y)`	Añade lineas
`text(x, y, etiquetas, ...)`	Añade la etiqueta en la coordenada xy
`abline(a,b)`	Recta de pendiente *b* y ordenada en el origen *a*
`abline(h = y)`	Linea horizontal de altura y
`abline(v = x)`	Linea vertical
`abline(objeto lm)`	Ajusta una recta de regresión
`legend(x, y, leyeda...)`	Añade leyenda al gráfico
`title(main, sub)`	Añade titulo y subtitulo
`axis(side, ...)`	Añada un eje, side puede ser 1, 2, 3, o 4

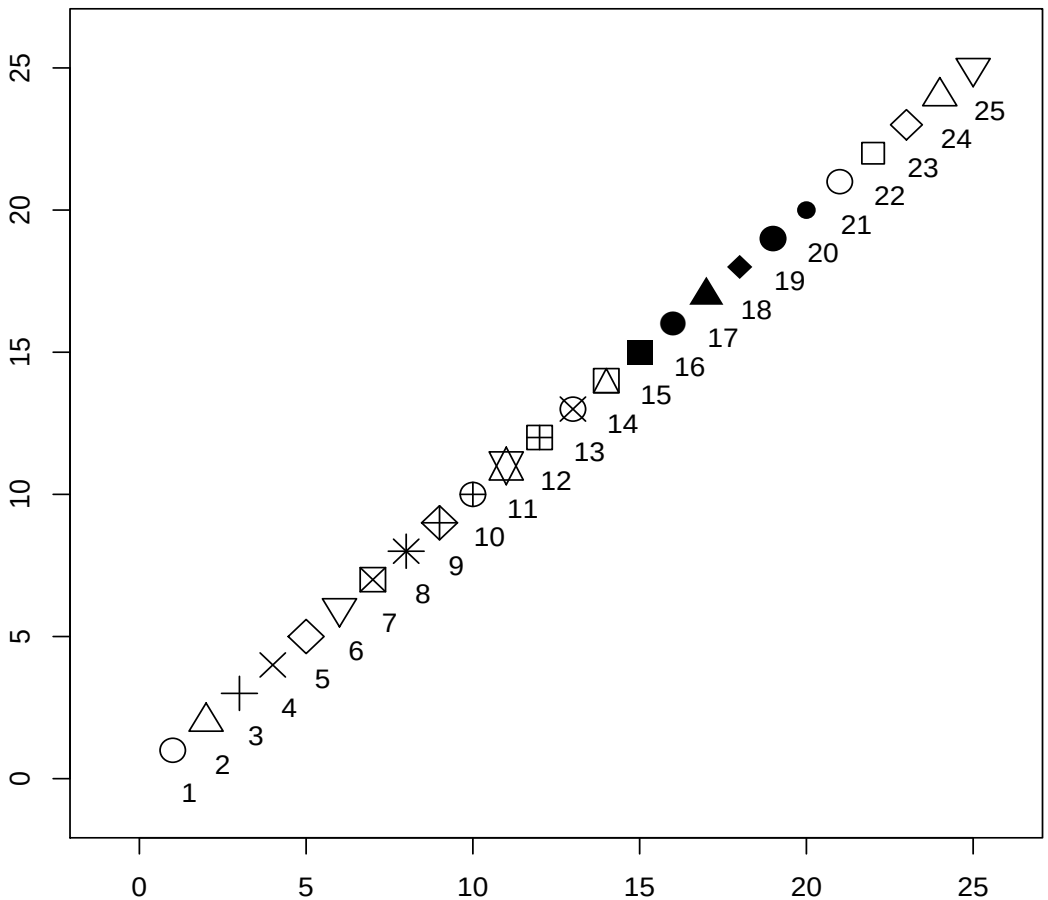
```r
plot(iris[,1], iris[,3], pch = 16, cex = 2, yaxt="n", xlab="Sepal Length",
     ylab="Petal Length", cex.axis = 1.5, cex.lab = 1.5)
abline(lm(iris[,3] ~ iris[,1]), col = "red", lwd = 2)
abline(v=mean(iris[,1]), col="blue",lwd = 2 )
abline(h=mean(iris[,3]), col="green", lwd = 2)
arrows(7, 2, 6, 3.5, lwd = 2)
axis(side = 2, las = 2)
title(main="Modelo Lineal. Sepal Length vs. Petal Length",cex=1.5)
```



Estructura de la presentación

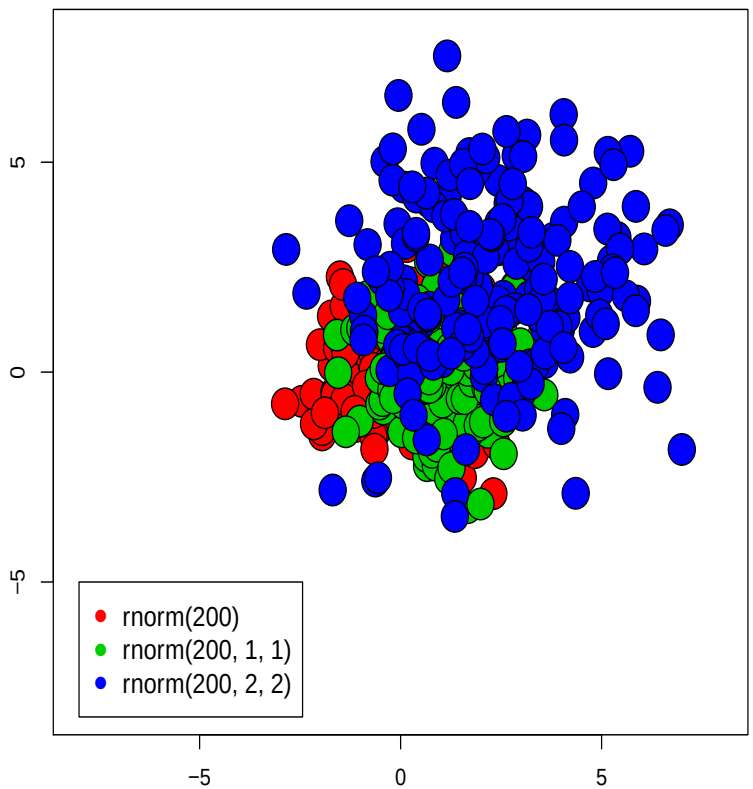
Argumento pch()

```
plot(1:25, pch = 1:25, cex = 2, xlim = c(-1,26), ylim = c(-1,26), xlab = "",
     ylab = "")
text(1:25 + 0.5, 1:25 - 1.5, 1:25 )
```



Argumento pch()

```
plot(-8:8, -8:8, type = "n", xlab = "", ylab = "", main = "")
points(rnorm(200), rnorm(200), pch = 21, bg = 2, cex = 3)
points(rnorm(200, 1, 1), rnorm(200), pch = 21, bg = 3, cex = 3)
points(rnorm(200, 2, 2), rnorm(200, 2, 2), pch = 21, bg = 4, cex = 3)
legend(-8,-5, c("rnorm(200)", "rnorm(200, 1, 1)", "rnorm(200, 2, 2)"),
        cex = 1.2, col = c(2,3,4), pch = 16)
```



Estructura de la presentación

Colores en R

R cuenta con una paleta amplia de colores o se pueden establecen colores con `rgb()` o con `hsv()`. A su vez hay una gran cantidad de paquetes que contienen paletas de colores.

Ejemplo:

```
palette()
```

```
## [1] "black"   "red"     "green3"  "blue"    "cyan"    "magenta" "yellow"
## [8] "gray"
```

```
colors()[sample(1:length(colors()), 20)]
```

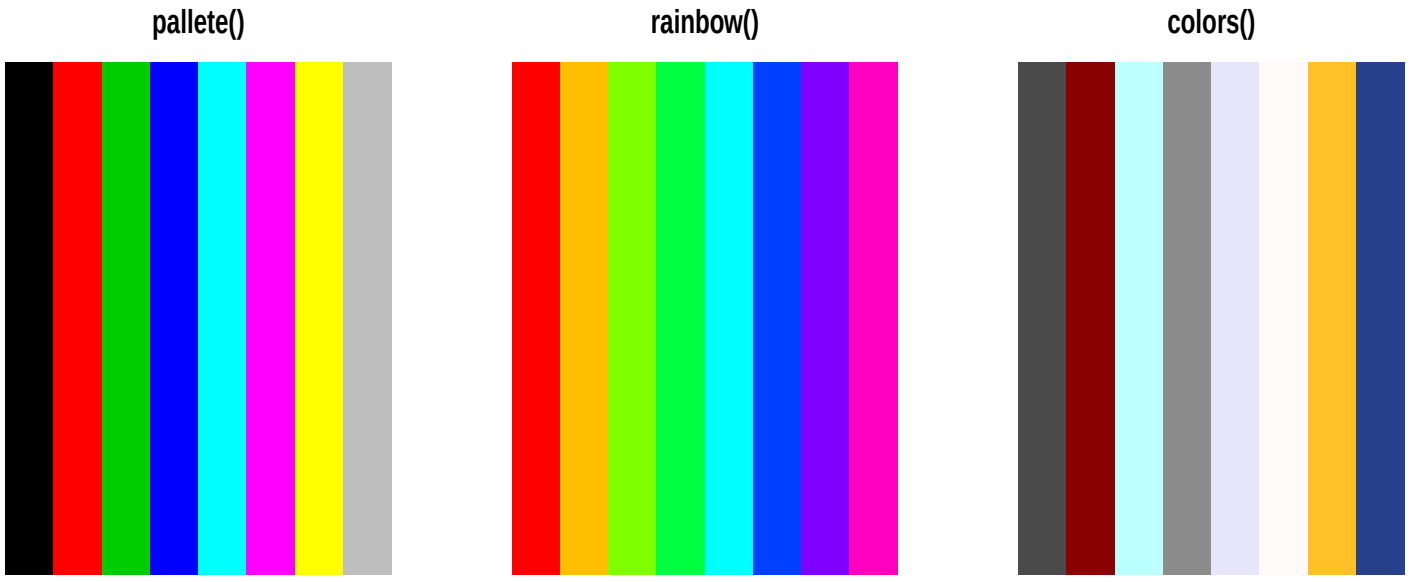
```
##  [1] "dodgerblue3"     "hotpink2"        "deeppink3"
##  [4] "pink3"           "darkolivegreen3" "gray63"
##  [7] "greenyellow"     "darkorchid4"     "seagreen"
## [10] "ivory2"          "green1"          "orchid1"
## [13] "gray49"          "lightblue4"      "salmon"
## [16] "lemonchiffon1"   "lightgoldenrod3" "gray52"
## [19] "green3"          "lavenderblush3"
```

```
length(colors())
```

```
## [1] 657
```

Colores en R: paquete grDevices

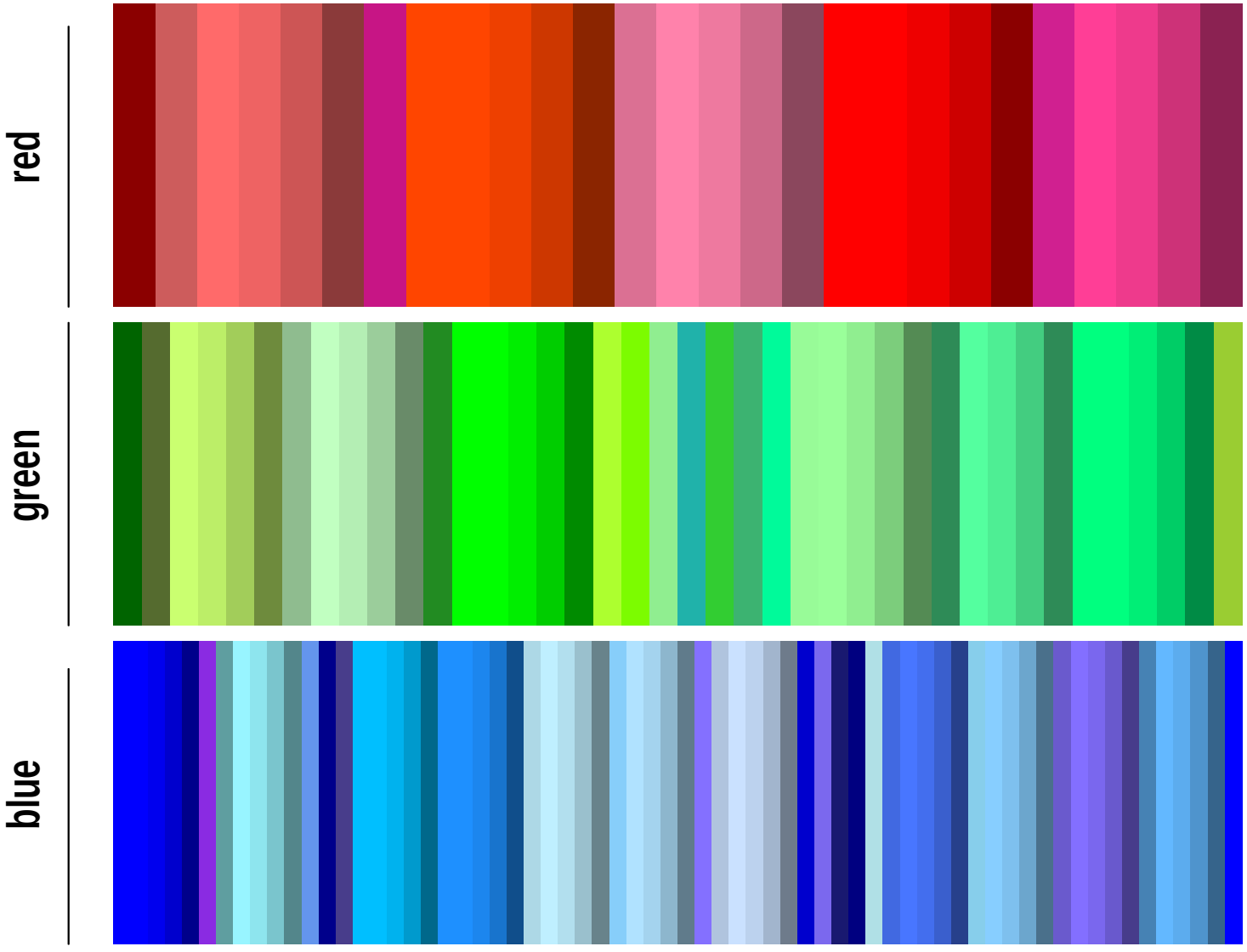
```
barplot(rep(8, 8), col = palette(), axes = FALSE, main = "pallete()", cex.main = 2,
        space = 0, border = NA )
barplot(rep(8, 8), col = rainbow(8), axes = FALSE, main = "rainbow()", cex.main = 2,
        space = 0, border = NA )
barplot(rep(8, 8), col = colors()[sample(1:length(colors()), 8)],
        axes = FALSE, main = "colors()", cex.main = 2, space = 0, border = NA )
```



Colores en R: paquete grDevices

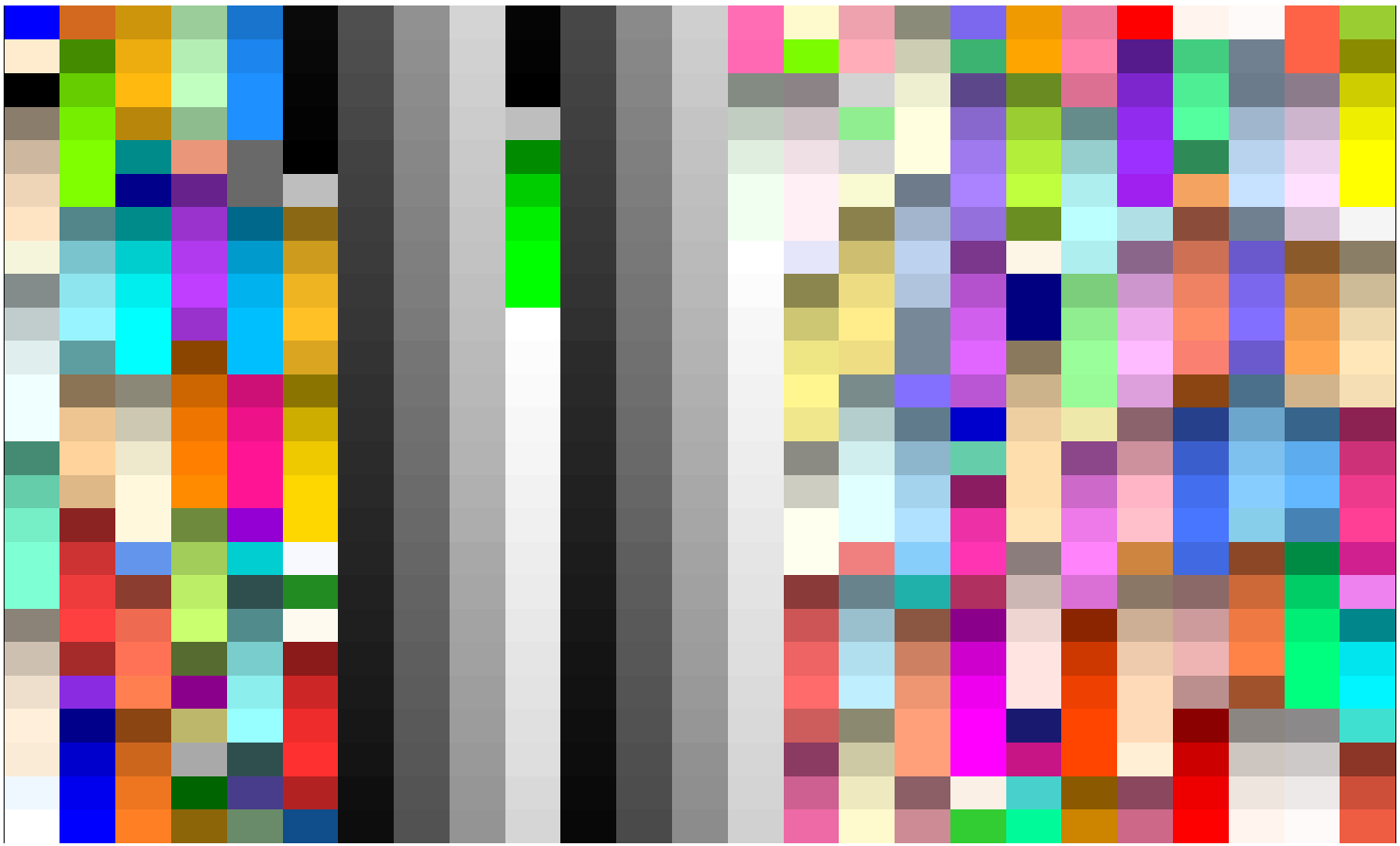
```
par(mfrow=c(3,1), mar=c(0.7, 3, 0, 0))

sec <- sum(grepl("red", colors()))
barplot(rep(sec, sec), col = colors()[which((grepl("red", colors())) == TRUE)],
    border = NA, space = 0, axes = FALSE)
axis(2, labels = FALSE, lwd.ticks = 0)
mtext("red", side = 2, line = 1, cex = 1.5, font = 2, las = 0)

sec <- sum(grepl("green", colors()))
barplot(rep(sec, sec), col = colors()[which((grepl("green", colors())) == TRUE)],
    border = NA, space = 0, axes = FALSE)
axis(2, labels = FALSE, lwd.ticks = 0)
mtext("green", side = 2, line = 1, cex = 1.5, font = 2, las = 0)

sec <- sum(grepl("blue", colors()))
barplot(rep(sec, sec), col = colors()[which((grepl("blue", colors())) == TRUE)],
    border = NA, space = 0, axes = FALSE)
axis(2, labels = FALSE, lwd.ticks = 0)
mtext("blue", side = 2, line = 1, cex = 1.5, font = 2, las = 0)
```

Colores en R: paquete `grDevices`



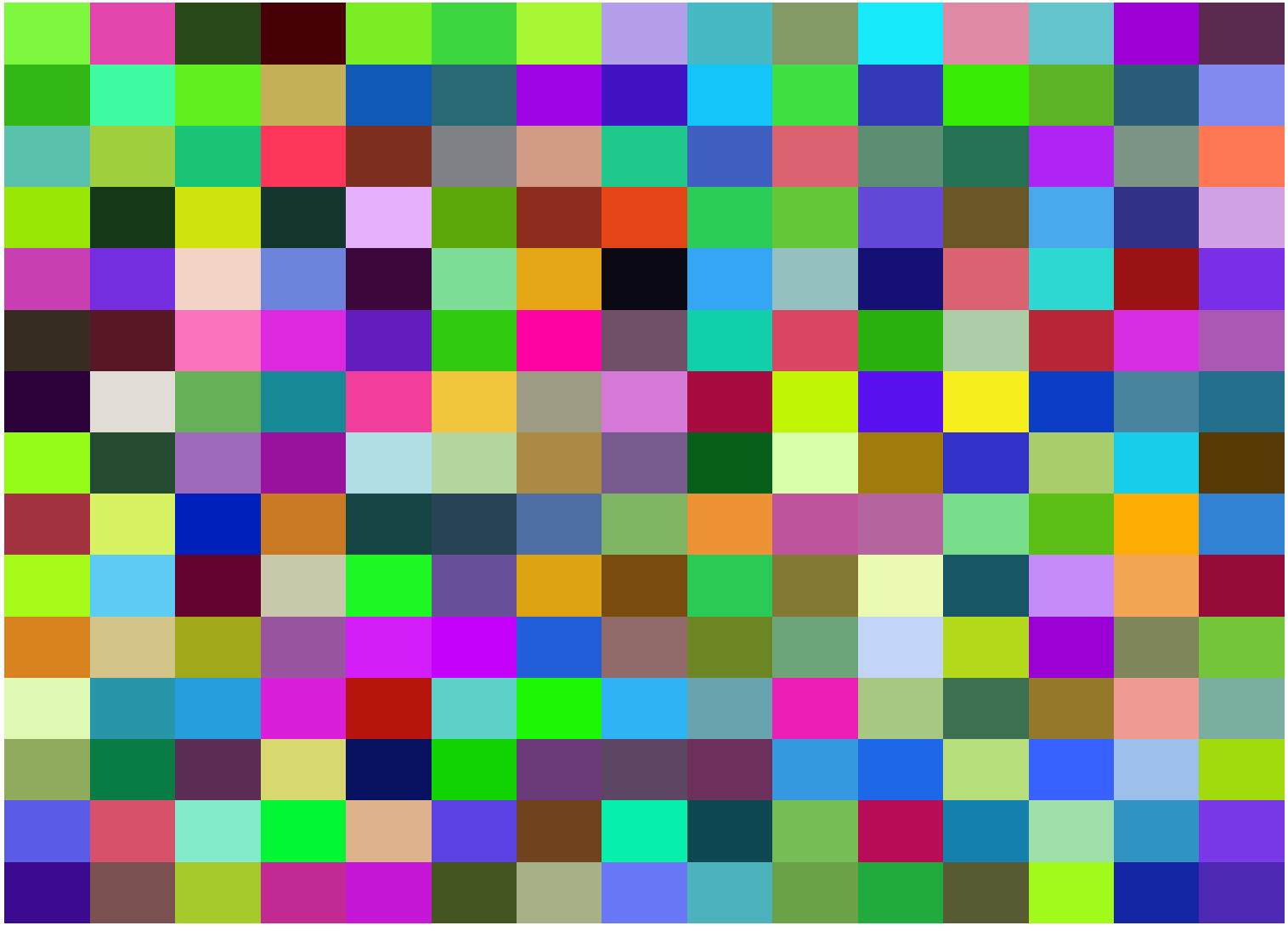
colors()

```
par(mar = rep(0,4))
image(matrix(1:625, 25, 25, byrow = TRUE), main = "", col = colors())
```



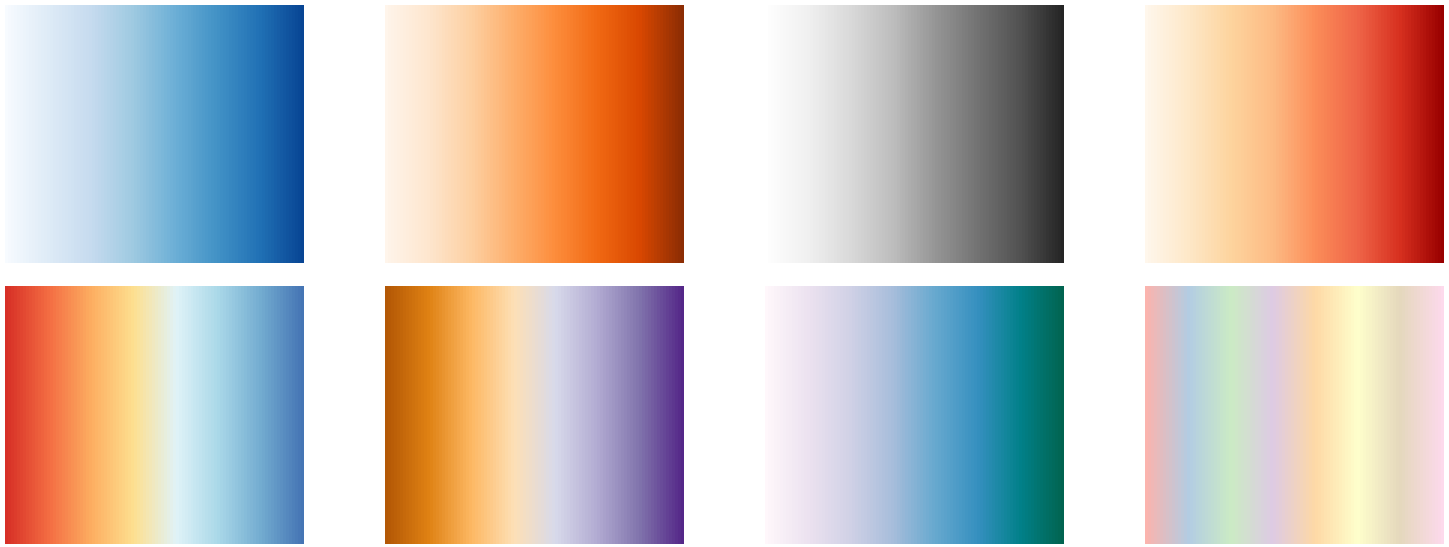
rgb()

```
animation::saveGIF({
    par(mfrow=c(1, 1),mar=rep(0, 4))

    sec_rgb <- character()
    for(i in 1:100){
            for(j in 1:225){
                    s <- seq(0, 1, 0.02)
                    sec_rgb[j] <- rgb(sample(s, 1), sample(s, 1),
                                      sample(s, 1), sample(s, 1))
            }
            image(matrix(1:225, 15, 15, byrow = TRUE), main = "", col = sec_rgb)
    }
}, interval = 0.8)
```



paletas

```
library(RColorBrewer)
paletas <- c("Blues", "Oranges", "Greys", "OrRd", "RdYlBu", "PuOr", "PuBuGn",
             "Pastel1")
par(mfrow = c(2, 4), mar = c(1, 1, 1, 1))
for(i in 1:length(paletas)){
    barplot(rep(10, 100),
            col = colorRampPalette(brewer.pal(n = 8, name = paletas[i]))(100),
            space = 0, border = NA, axes = F)
}
```



Paquete: paletteer


```
colores <- paletteer::paletteer_c(package = "scico", palette = "oslo", n = 50)

## Error in loadNamespace(name): there is no package called 'paletteer'

barplot(rep(10, length(colores)), col = colores,
    space = 0, border = NA, axes = F)

## Error in barplot(rep(10, length(colores)), col = colores, space = 0, border =
NA, : object 'colores' not found
```

Estructura de la presentación

par()

La función par() permite controlar todos los aspectos de bajo nivel de un gráfico. Si se ejecuta la función par() aparece la lista de todos los argumentos con los valores por defecto. Cada vez que se modifica algún parámetro no hay retorno hasta que vuelva a ser cambiado o al reiniciar sesión. Por eso es recomendable guardar los valores por defecto:

Ejemplo:

```
op <- par(no.readonly = TRUE)
par(op)
```

Parámetro: `par()`

Parámetro	Descripción
`adj`	Justificación del texto `adj = 0` → izquierda `adj = 0,5` → centrado `adj = 1` → derecha
`bty`	Tipo de caja del gráfico Tipos posible: "o", "l", "7", "c", "u", "]"
`cex`	Controla el tamaño de símbolos y textos
`cex.axis`	Tamaño de los valores de los ejes
`cex.lab`	Tamaño de las etiquetas de los ejes
`cex.main`	Tamaño del titulo
`cex.sub`	Tamaño del subtitulo
`col.axis`	Color de los valores de los ejes
`col.lab`	Color de las etiquetas de los ejes
`col.main`	Color del titulo
`col.sub`	Color del subtitulo
`col`	Color de símbolos

Parámetros gráficos: `par()`

Parámetro	Descripción
`font`	Tipo de letra
	`font = 1` → normal `font = 2` → cursiva `font = 3` → negrita `font = 4` negrita y cursiva
`las`	Controla la orientación de las etiquetas `las = 0` →paralelo a los ejes `las = 1` →horizontal `las = 2` →perpendicular a los ejes `las = 3` →vertical
`lty`	Tipo de linea. Numero entero del 1 al 5
`lwd`	Ancho de linea. Numero entero
`mfcol`	Vector de dos enteros. Divide ventana gráfica
`mfrow`	Vector de dos enteros. Divide ventana gráfica
`pch`	Tipo de símbolo. Numero entero del 1 al 25
`bg`	Color de fondo

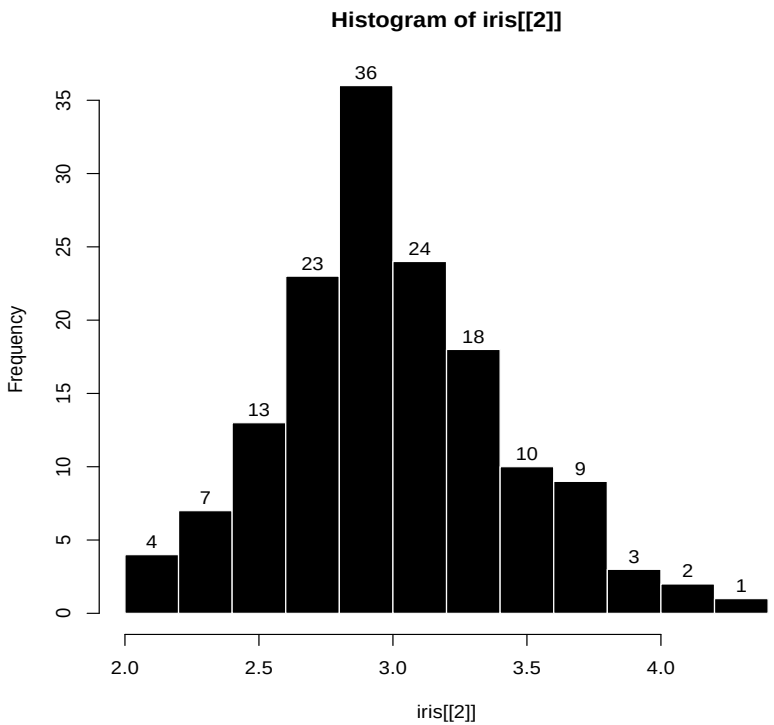
Estructura de la presentación

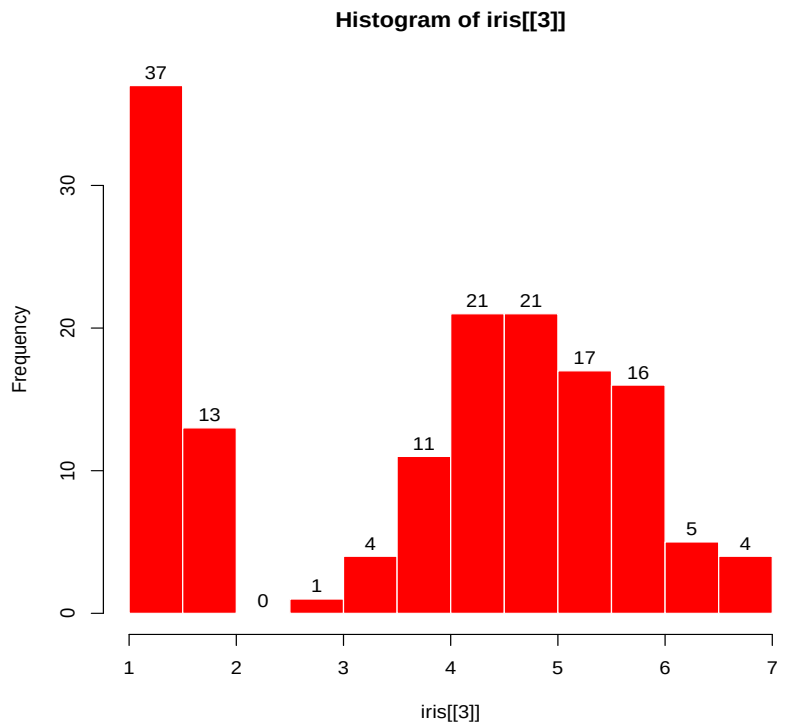
hist()

Un histograma es una representación gráfica de una variable continua que se visualiza como un diagrama de barras con intervalos constantes.

Ejemplo:

```
hist(iris[[2]], col = 1, border = "white", labels = TRUE)
hist(iris[[3]], col = 2, border = "white", labels = TRUE)
```





Todos los parámetros de construcción de un histograma se pueden modificar. Si desea ver los cálculos hechos por defecto puede guardar como un objeto un histograma y ver el contenido del mismo.

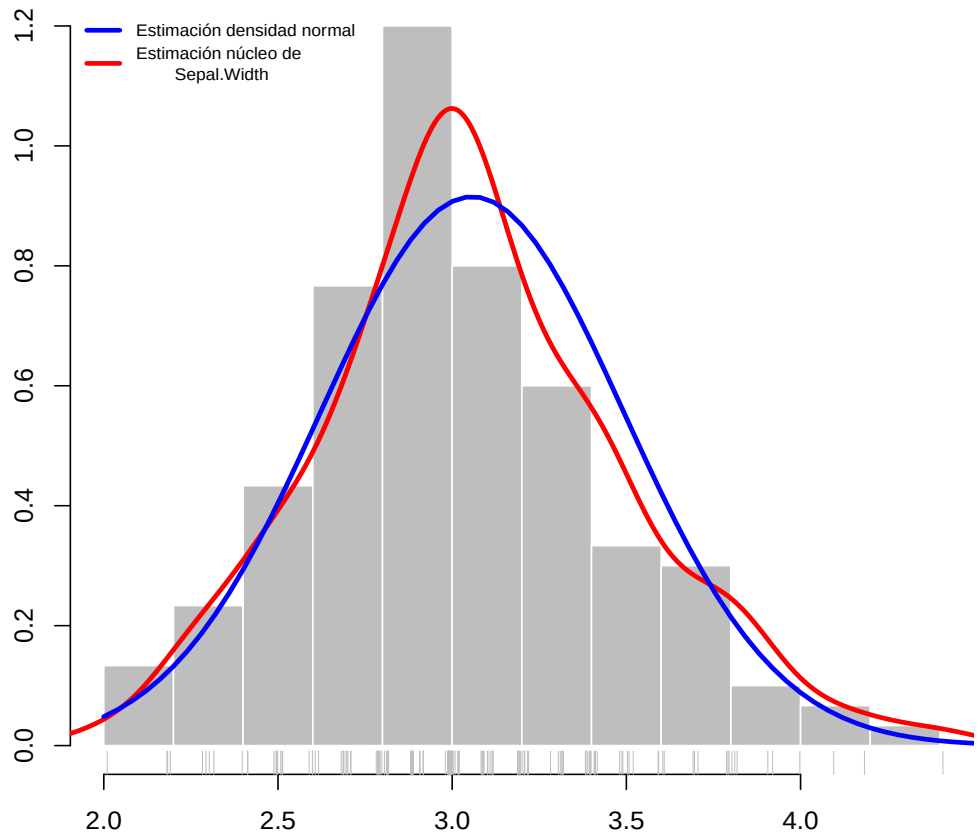
```
h <- hist(iris[[2]], plot = FALSE); h
```

```
## $breaks
##  [1] 2.0 2.2 2.4 2.6 2.8 3.0 3.2 3.4 3.6 3.8 4.0 4.2 4.4
##
## $counts
##  [1]  4  7 13 23 36 24 18 10  9  3  2  1
##
## $density
##  [1] 0.13333333 0.23333333 0.43333333 0.76666667 1.20000000 0.80000000
##  [7] 0.60000000 0.33333333 0.30000000 0.10000000 0.06666667 0.03333333
##
## $mids
##  [1] 2.1 2.3 2.5 2.7 2.9 3.1 3.3 3.5 3.7 3.9 4.1 4.3
##
## $xname
## [1] "iris[[2]]"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

hist()

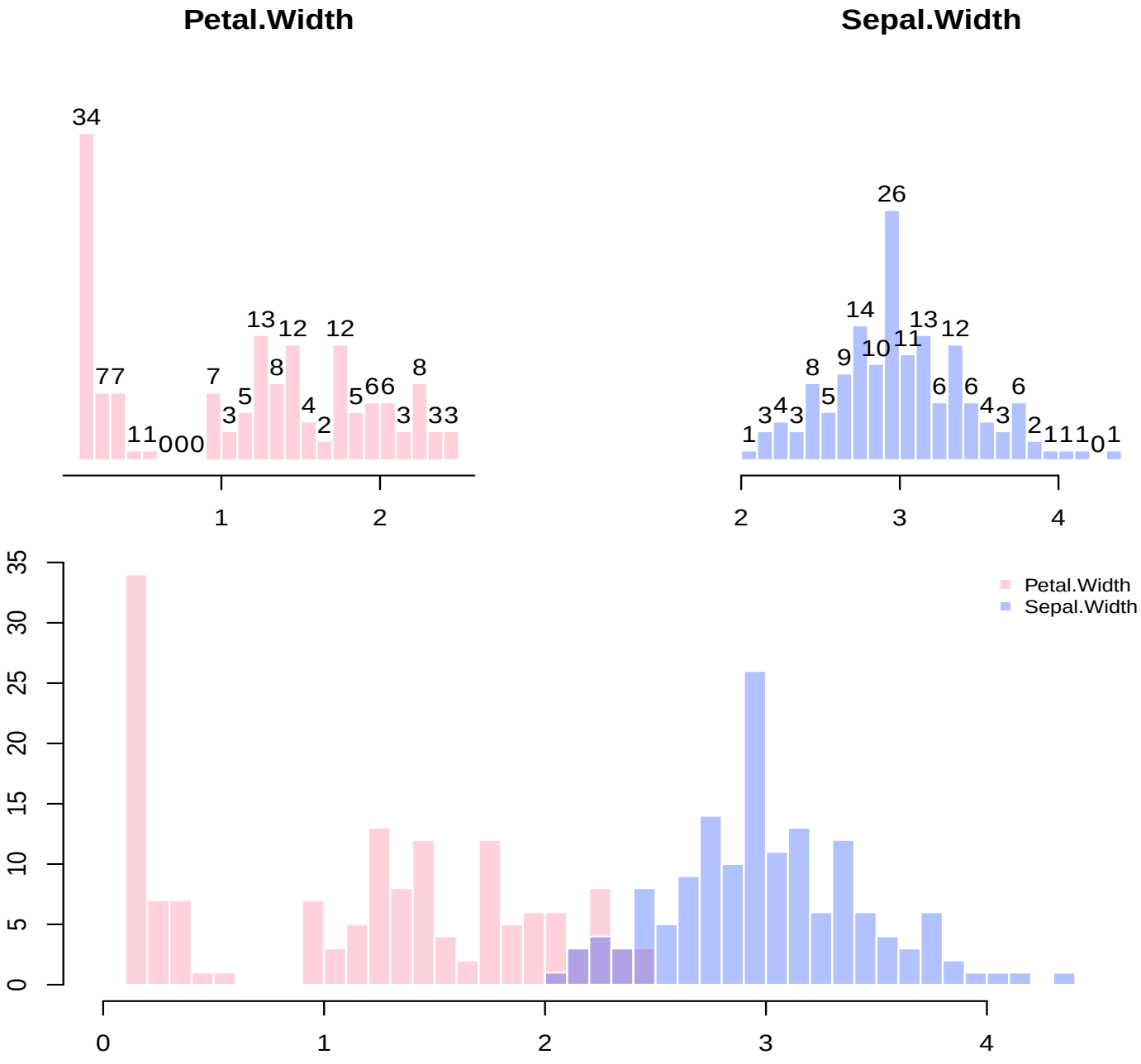
Parámetro	Descripción
`breaks`	Indica los valores de corte de los intervalos de la variable por defecto son todos iguales.
`counts`	Indica la frecuencia de los valores de la variable en cada intervalo.
`density`	Cuando el argumento `freq` es `FALSE` se calcula la densidad y el eje `y` cambia. Se calcula como: la frecuencia en el intervalo dividido el numero total datos por el ancho del intervalo. a la unidad.
`mids`	Punto medio de los intervalos.
`equidist`	Indica si la distancia entre los intervalos son iguales.

```r
hist(iris[[2]], col = "gray", border = "white", freq = FALSE,
        main = "", xlab = "", ylab = "")
lines(density(iris$Sepal.Width),col="red",lwd=3)
curve(dnorm(x,mean = mean(iris$Sepal.Width),sd = sd(iris$Sepal.Width)),from = 2,
        to = 6, add = TRUE, col = "blue", lwd = 3)
legend("topleft", col = c("blue","red"), lwd = 3, bty = "n", cex = 0.7,
        legend = c("Estimación densidad normal","Estimación núcleo de
        Sepal.Width")); rug(jitter(iris[[2]]), col = "gray")
```



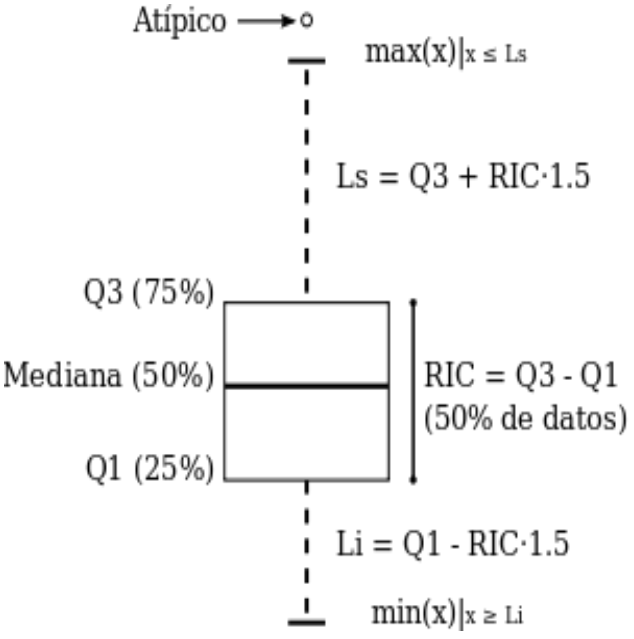
Ejemplo:

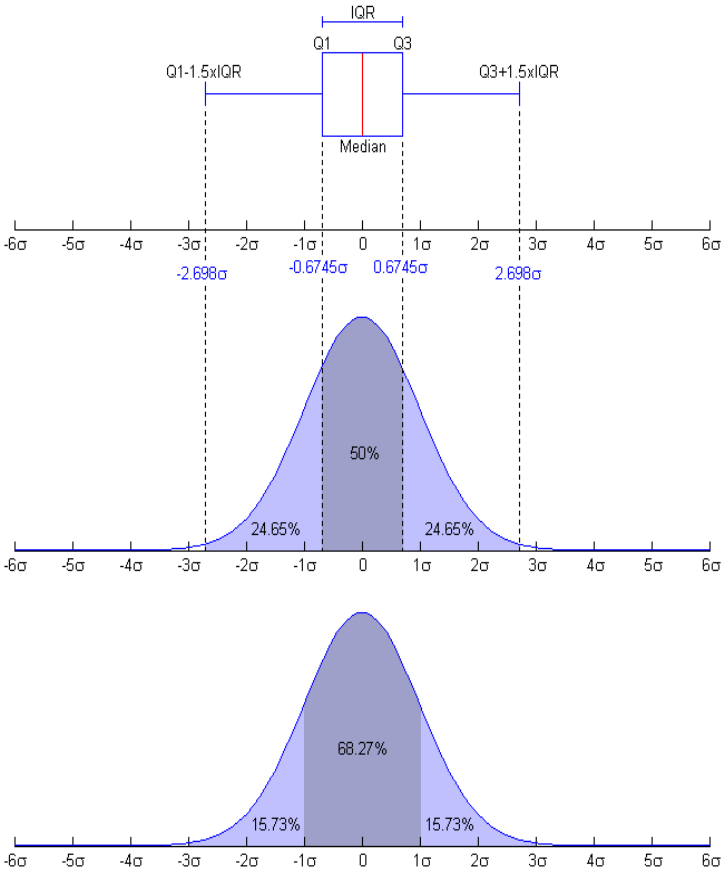
```
col1 <- rgb(1,0.1,0.3,0.2)
col2 <- rgb(0,0.2,  1,0.3)
par(mar=c(3,5,3,3))
layout(matrix(c(1,2,3,3), 2, 2, byrow = TRUE))

hist(iris[[4]], breaks = 20, col = col1, main = "Petal.Width",
        xlab = "",  ylab = "", labels = TRUE, axes = FALSE)
axis(1, seq(0, 3))

hist(iris[[2]], breaks = 20, col = col2, main = "Sepal.Width",
        xlab = "",  ylab = "", labels = TRUE, , axes = FALSE)
axis(1, seq(2, 4.5))

par(mar=c(5,5,0,3))
hist(iris[[4]], breaks = 20, xlim = c(0,4.5), col = col1,
        main = "", xlab = "",  ylab = "")

hist(iris[[2]], breaks = 20, xlim = c(0,4.5), col = col2, add = T,
        main = "", xlab = "",  ylab = "")

legend(4, 35, legend=c("Petal.Width","Sepal.Width"),
        col=c(col1, col2), pch=15, bty = "n", cex = 0.8)
```


Petal.Width
Sepal.Width
34
77
11
000
7
3
5
13
8
12
4
2
12
5
66
8
3
33
1
26
3
4
3
8
5
9
14
10
11
13
6
12
6
6
4
3
6
2
1
1
1
0
1
1
2
2
3
4
0
5
10
15
20
25
30
35
Petal.Width
Sepal.Width
0
1
2
3
4

Estructura de la presentación

boxplot()


Atípico
max(x)|x ≤ Ls
Ls = Q3 + RIC·1.5
Q3 (75%)
Mediana (50%)
Q1 (25%)
RIC = Q3 - Q1
(50% de datos)
Li = Q1 - RIC·1.5
min(x)|x ≥ Li

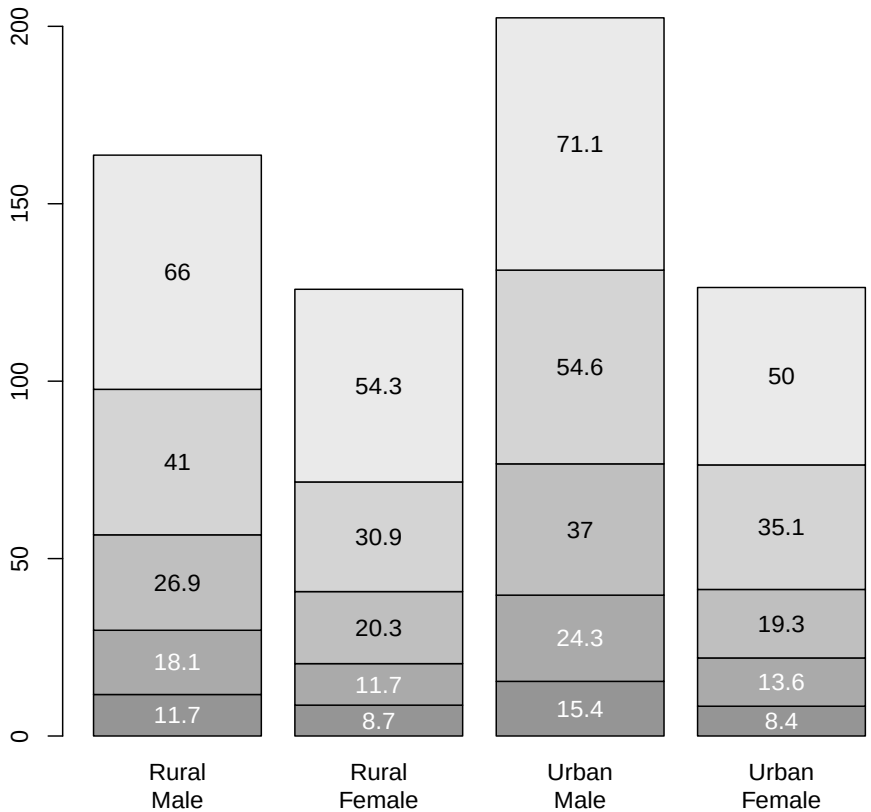

IQR
Q1
Q3
Q1-1.5xIQR
Q3+1.5xIQR
Median
-6σ -5σ -4σ -3σ -2σ -1σ 0 1σ 2σ 3σ 4σ 5σ 6σ
-2.698σ
-0.6745σ
0.6745σ
2.698σ
50%
24.65%
24.65%
68.27%
15.73%
15.73%

Estructura de la presentación

barplot()

a

```
midpts <- barplot(VADeaths, col=grey(0.5 + 1:5/12), names=rep("", 4))
mtext(sub(" ", "\n", colnames(VADeaths)),at=midpts, side=1, line=1.5, cex=1)
text(rep(midpts, each=5), apply(VADeaths, 2, cumsum) - VADeaths/2,
     VADeaths, col=rep(c("white", "black"), times=2:3, cex=0.8))
```



Fuente: https://www.stat.auckland.ac.nz