Frecuencias datos cuantitativos

Heiner Romero Leiva

02/24/2019

Estudio de frecuencias

```
edad = sample(10:45, size= 100, replace = TRUE)
table(edad)
```

```
## edad
## 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 33 35 36 37
##  4  1  2  4  3  6  3  3  5  3  3  1  1  5  5  1  3  2  4  2  1  5  3  1  4  4
## 38 39 40 41 43 44 45
##  3  1  5  3  2  3  4
```

```
# Tabla de frecuencias relativas
round(prop.table(table(edad)), 3) # 1 de cada 100 tienen 10, 4 de cada 100 tienen 11 (...)
```

```
## edad
##    10    11    12    13    14    15    16    17    18    19    20    21    22    23    24    25
## 0.04 0.01 0.02 0.04 0.03 0.06 0.03 0.03 0.05 0.03 0.03 0.01 0.01 0.05 0.05 0.01
##    26    27    28    29    30    31    33    35    36    37    38    39    40    41    43    44
## 0.03 0.02 0.04 0.02 0.01 0.05 0.03 0.01 0.04 0.04 0.03 0.01 0.05 0.03 0.02 0.03
##    45
## 0.04
```

```
# Tabla de frecuencias acumuladas
cumsum(table(edad)) # 37 personas tienen 24 annos o menos, 55 tenian 55 annos o menos
```

```
##  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29
##   4   5   7  11  14  20  23  26  31  34  37  38  39  44  49  50  53  55  59  61
##  30  31  33  35  36  37  38  39  40  41  43  44  45
##  62  67  70  71  75  79  82  83  88  91  93  96 100
```

```
# Tabla de frecuencias relativas
round(cumsum(prop.table(table(edad))),3) # 21% de 100 tenian 18 annos o menos, 96% de 100 tenian 43 ann
```

```
##    10    11    12    13    14    15    16    17    18    19    20    21    22    23    24    25
## 0.04 0.05 0.07 0.11 0.14 0.20 0.23 0.26 0.31 0.34 0.37 0.38 0.39 0.44 0.49 0.50
##    26    27    28    29    30    31    33    35    36    37    38    39    40    41    43    44
## 0.53 0.55 0.59 0.61 0.62 0.67 0.70 0.71 0.75 0.79 0.82 0.83 0.88 0.91 0.93 0.96
##    45
## 1.00
```

Experimentacion

Lanzamos 25 veces un dado de 6 caras y anotamos las puntuaciones obtenidas en cada tirada.

En este caso, $n = 25$ y, los distintos valores observados son:

$$X_1 = 1, X_2 = 2, X_3 = 3, X_4 = 4, X_5 = 5, X_6 = 6$$

Nos interesa ahora calcular las frecuencias de este experimento. Ademas, las organizaremos en un data frame para observarlas de formas mas clara y sencilla en una tabla.

```
set.seed(162017)
dados = sample(1:6, 25, replace = TRUE)
dados
```

```
##  [1] 1 1 5 5 5 5 1 6 5 4 1 3 1 3 2 2 1 1 1 4 2 1 6 3 1
```

```
set.seed(NULL)

# Tabla de frecuencias absolutas
table(dados)
```

```
## dados
##  1  2  3  4  5  6
## 10  3  3  2  5  2
```

```
# Tabla de frecuencias relativas
round(prop.table(table(dados)),3)
```

```
## dados
##    1    2    3    4    5    6
## 0.40 0.12 0.12 0.08 0.20 0.08
```

```
# Tabla de frecuencias absolutas acumuladas
cumsum(table(dados))
```

```
##  1  2  3  4  5  6
## 10 13 16 18 23 25
```

```
# Tabla de frecuencias relativas acumuladas
round(cumsum(prop.table(table(dados))),3)
```

```
##    1    2    3    4    5    6
## 0.40 0.52 0.64 0.72 0.92 1.00
```

```
# Creando dataframe para visualizar todas las frecuencias
dados.df = data.frame(Puntuacion = 1:6,
                      Fr.abs = as.vector(table(dados)),
                      Fr.rel = as.vector(round(prop.table(table(dados)),3)),
                      Fr.acu = as.vector(cumsum(table(dados))),
                      Fr.racu = as.vector(round(cumsum(prop.table(table(dados))),3)))
# Visualizando
dados.df
```

```
##   Puntuacion Fr.abs Fr.rel Fr.acu Fr.racu
## 1          1     10   0.40     10    0.40
## 2          2      3   0.12     13    0.52
## 3          3      3   0.12     16    0.64
## 4          4      2   0.08     18    0.72
## 5          5      5   0.20     23    0.92
## 6          6      2   0.08     25    1.00
```

Medidas de Tendencia Central

La media aritmetica o valor medio:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{k} n_j X_j}{n}$$

La mediana, que representa el valor central en la lista *ordenada* de observaciones.

Si

$$n$$

par, es el medio de los datos centrales. Si

$$n$$

impar, el dato central.

La moda es el valor (o valores) de maxima frecuencia (absoluta o relativa, el resultado sera el mismo).

Ejemplo:

```
sort(edad)
```

```
##   [1] 10 10 10 10 11 12 12 13 13 13 13 14 14 14 15 15 15 15 15 15 16 16 16 17 17
##  [26] 17 18 18 18 18 18 19 19 19 20 20 20 21 22 23 23 23 23 23 24 24 24 24 24 25
##  [51] 26 26 26 27 27 28 28 28 28 29 29 30 31 31 31 31 31 33 33 33 35 36 36 36 36
##  [76] 37 37 37 37 38 38 38 39 40 40 40 40 40 41 41 41 43 43 44 44 44 45 45 45 45
```

```
table(edad)
```

```
## edad
## 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 33 35 36 37
##  4  1  2  4  3  6  3  3  5  3  3  1  1  5  5  1  3  2  4  2  1  5  3  1  4  4
## 38 39 40 41 43 44 45
##  3  1  5  3  2  3  4
```

```
dados.df # tambien se pueden sacar en la tabla que construi.
```

```
##   Puntuacion Fr.abs Fr.rel Fr.acu Fr.racu
## 1          1     10   0.40     10    0.40
## 2          2      3   0.12     13    0.52
## 3          3      3   0.12     16    0.64
## 4          4      2   0.08     18    0.72
## 5          5      5   0.20     23    0.92
## 6          6      2   0.08     25    1.00
```

```
# Para el caso de la moda es 3.
# Para la mediana es el numero 3 porque es el que supera mas del 50 de frecuencia relativa acumulada.

mean(edad) # media
```

```
## [1] 26.55
```

```
mean(dados)
```

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

```
median(edad) # la mediana
```

```
## [1] 25.5
```

```
median(dados)
```

```
## [1] 2
```

```r
as.numeric(names(which(table(edad) == max(table(edad))))) # la moda
```

```
## [1] 15
```

```r
as.numeric(names(which(table(dados)==max(table(dados))))) # la moda
```

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

Tipos de Medias

```r
x = c(32, 45, 67, 43, 28, 17, 48, 95)
n = length(x)
```

Media aritmetica

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

.

```r
sum(x)/n # es lo mismo que hacerlo con la formula mean
```

```
## [1] 46.875
```

```r
mean(x)
```

```
## [1] 46.875
```

Media aritmetica ponderada

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i \cdot x_i}{\sum_{i=1}^{n} w_i}$$

Se ponderan los elementos por cada valor especifico para poder sacar un promedio final. Ejemplo: promedio de un curso de universidad.

```r
w = c(1,2,2,3,3,2,2,1)
sum(w*x)/sum(w)
```

```
## [1] 43.375
```

Media Geometrica

Es util cuando el conjunto de numeros que son interpretados necesitan o quedan explicados en forma de producto. Por ejemplo: V * T (velocidad * tiempo).

$$\bar{x}_G = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

```r
prod(x)^(1/n)
```

```
## [1] 41.62073
```

Media Armonica

Es muy util en conjuntos de numeros que se definen en relacion con alguna unidad, por ejemplo la distancia por unidad de tiempo.

$$\bar{x}_A = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

```
n/sum(1/x)
```

```
## [1] 36.77301
```

Minimo y maximo

```
min(x)
```

```
## [1] 17
```

```
max(x)
```

```
## [1] 95
```

Medidas de posicion:

```
set.seed(260798)
dado = sample(1:4, 50, replace = TRUE)
set.seed(NULL)
length(dado)
```

```
## [1] 50
```

```
dado = sort(dado) # los ordenamos de menor a mayor
dado
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 4 4
## [39] 4 4 4 4 4 4 4 4 4 4 4 4
```

```
dado.df = data.frame(Puntuacion = 1:4,
                     Fr.abs = as.vector(table(dado)),
                     Fr.rel = as.vector(round(prop.table(table(dado)),2)),
                     Fr.acu = as.vector(cumsum(table(dado))),
                     Fr.racu = as.vector(round(cumsum(prop.table(table(dado))),2)))
dado.df
```

```
##   Puntuacion Fr.abs Fr.rel Fr.acu Fr.racu
## 1          1     16   0.32     16    0.32
## 2          2     15   0.30     31    0.62
## 3          3      5   0.10     36    0.72
## 4          4     14   0.28     50    1.00
```

```
dado[15] # el cuantil 0.3 es 1.
```

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

Los cuartiles, son los tres numeros que dividen por cuartos a la poblacion total, es decir, el primer cuartil, segundo cuartil (mediana), y el tercer cuartil.

Los deciles son los cuantiles Q_p con p un multiplo de 0.1

Los percentiles son los cuantiles Q_p con p un multiplo de 0.01, es decir, 1%, 2%, 3% etc.

```
quantile(dado)
```

```
##   0%  25%  50%  75% 100%
##    1    1    2    4    4
```

```r
set.seed(0)
dados2 = sample(1:6, 15, replace = TRUE)
dados2
```

```
##  [1] 6 1 4 1 2 5 3 6 2 3 3 1 5 5 2
```

```r
set.seed(NULL) # invalidamos la semilla para volver a la aleatoridad
quantile(dados2, 0.25) # primer cuartil, el 25% de los datos son 2 o menor que dos.
```

```
## 25%
##   2
```

```r
quantile(dados2, 0.8) # decil 8 # el 80% de los datos son 6 o menores que 6
```

```
## 80%
##   5
```

Medidas de dispersion

Evaluan los dispersos que estan los datos. Algunas de las mas importantes son:

- El rango o recorrido, que es la diferencia entre el maximo y el minimo de las observaciones.
- El rango intercuartil, que es la diferencia entre el tercer y el primer cuartil, $Q_0.75 - Q_0.25$.
- La **varianza**, a la que denotaremos por s^2, es la media aritmetica de las diferencias al cuadrado entre los datos x_i y la media aritmetica de las observaciones $\bar{x}$
- La desviacion tipica es la raiz cuadrada positiva de la varianza, $s = \sqrt(s^2)$.
- La varianza muestral es la correcion de la covarianza. La denotamois por s^2 y se corresponde con $s^2 = \frac{n}{n-1} * s^2$
- La desviacion tipica muestral, que es la raiz cuadrada positiva de la varianza muestral, $s = \sqrt(s^2)$

La varianza siempre es positiva o cero (nula).

Varianza y desviacion tipica

Notese que tanto la varianza como la desviacion tipica dan una informacion equivalente. Entonces, es comprensible preguntarse porque se definen ambas medidas sin con una basta. Pues bien, las unidades de la varianza (metros, litros, annos...), ya sea muestral o no, estan al cuadrado, mientras que las de la desviacion tipica no.

```r
dados2
```

```
##  [1] 6 1 4 1 2 5 3 6 2 3 3 1 5 5 2
```

```r
diff(range(dados2)) # como es 5, se debe saber que entre el primer numero y el segundo hay 5 unidades d
```

```
## [1] 5
```

```r
min(dados2) # minimo
```

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

```r
max(dados2) # maximo
```

```
## [1] 6
```

```r
IQR(dados2) # rango intercuartilico, esto quiere decir que entre el primer cuartil y el tercer cuartil,
```

```
## [1] 3
```

```r
var(dados2) # varianza muestral
```

```
## [1] 3.209524
```

```r
sd(dados2) # desviacion tipica muestral
```

```
## [1] 1.791514
```

```r
n = length(dados2)
var(dados2)*(n-1)/n # varianza verdadera
```

```
## [1] 2.995556
```

```r
sd(dados2) * sqrt((n-1)/n) # desviacion tipica verdadera
```

```
## [1] 1.730767
```

Datos cuantitativos por factor

```r
summary(dados2) # resumen de medidas de posicion
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   3.000   3.267   5.000   6.000
```

Funciones utilies para explorar datos cuantitativos.

Summary function:

Ejemplo:

```r
cangrejos = read.table("/Users/heinerleivagmail.com/Documents/GitHub/r-basic/data/datacrab.txt", header

cangrejos = cangrejos[-1]
summary(cangrejos)
```

```
##      color           spine           width          satell           weight
##  Min.   :2.000   Min.   :1.000   Min.   :21.0   Min.   : 0.000   Min.   :1200
##  1st Qu.:3.000   1st Qu.:2.000   1st Qu.:24.9   1st Qu.: 0.000   1st Qu.:2000
##  Median :3.000   Median :3.000   Median :26.1   Median : 2.000   Median :2350
##  Mean   :3.439   Mean   :2.486   Mean   :26.3   Mean   : 2.919   Mean   :2437
##  3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:27.7   3rd Qu.: 5.000   3rd Qu.:2850
##  Max.   :5.000   Max.   :3.000   Max.   :33.5   Max.   :15.000   Max.   :5200
```

```r
# Comparando diferentes colores de cangrejos

summary(subset(cangrejos, color == 3, c("weight", "width"))) # para cangrejos de 3 colores
```

```
##      weight         width
##  Min.   :1300   Min.   :22.5
##  1st Qu.:2100   1st Qu.:25.1
##  Median :2500   Median :26.5
##  Mean   :2538   Mean   :26.7
##  3rd Qu.:3000   3rd Qu.:28.2
##  Max.   :5200   Max.   :33.5
```

```r
summary(subset(cangrejos, color == 5, c("weight", "width"))) # para cangrejos de 5 colores
```

```
##      weight         width
##  Min.   :1300   Min.   :21.00
##  1st Qu.:1900   1st Qu.:23.90
##  Median :2125   Median :25.50
##  Mean   :2174   Mean   :25.28
```

```
##  3rd Qu.:2400   3rd Qu.:26.57
##  Max.   :3225   Max.   :29.30
```

```
# Los cangrejos con 5 colores pesan ligeramente menos y tienen menos ancgura que los que tienen 3 color
```

By Function:

```
by(iris[,c(1,3)], iris$Species, FUN = summary) # resumen estadistico por especies
```

```
## iris$Species: setosa
##   Sepal.Length    Petal.Length
##  Min.   :4.300   Min.   :1.000
##  1st Qu.:4.800   1st Qu.:1.400
##  Median :5.000   Median :1.500
##  Mean   :5.006   Mean   :1.462
##  3rd Qu.:5.200   3rd Qu.:1.575
##  Max.   :5.800   Max.   :1.900
## ------------------------------------------------------------
## iris$Species: versicolor
##   Sepal.Length    Petal.Length
##  Min.   :4.900   Min.   :3.00
##  1st Qu.:5.600   1st Qu.:4.00
##  Median :5.900   Median :4.35
##  Mean   :5.936   Mean   :4.26
##  3rd Qu.:6.300   3rd Qu.:4.60
##  Max.   :7.000   Max.   :5.10
## ------------------------------------------------------------
## iris$Species: virginica
##   Sepal.Length    Petal.Length
##  Min.   :4.900   Min.   :4.500
##  1st Qu.:6.225   1st Qu.:5.100
##  Median :6.500   Median :5.550
##  Mean   :6.588   Mean   :5.552
##  3rd Qu.:6.900   3rd Qu.:5.875
##  Max.   :7.900   Max.   :6.900
```

```
by(iris[,c(1,3)], iris$Species, FUN = max) # vemos los tamannos de las especies
```

```
## iris$Species: setosa
## [1] 5.8
## ------------------------------------------------------------
## iris$Species: versicolor
## [1] 7
## ------------------------------------------------------------
## iris$Species: virginica
## [1] 7.9
```

```
by(iris[,c(1,3)], iris$Species, FUN = min) # vemos los tamannos de las especies y podemos concluir que
```

```
## iris$Species: setosa
## [1] 1
## ------------------------------------------------------------
## iris$Species: versicolor
## [1] 3
## ------------------------------------------------------------
## iris$Species: virginica
```

```
## [1] 4.5
```

Funcion aggregate

```
aggregate(cbind(Sepal.Length, Petal.Length)~Species, data = iris, FUN = summary)
```

```
##      Species Sepal.Length.Min. Sepal.Length.1st Qu. Sepal.Length.Median
## 1     setosa             4.300                4.800               5.000
## 2 versicolor             4.900                5.600               5.900
## 3  virginica             4.900                6.225               6.500
##   Sepal.Length.Mean Sepal.Length.3rd Qu. Sepal.Length.Max. Petal.Length.Min.
## 1             5.006                5.200             5.800             1.000
## 2             5.936                6.300             7.000             3.000
## 3             6.588                6.900             7.900             4.500
##   Petal.Length.1st Qu. Petal.Length.Median Petal.Length.Mean
## 1                1.400               1.500              1.462
## 2                4.000               4.350              4.260
## 3                5.100               5.550              5.552
##   Petal.Length.3rd Qu. Petal.Length.Max.
## 1                1.575             1.900
## 2                4.600             5.100
## 3                5.875             6.900
```

NA

```
x = c(1,2,34,NA)
var(x) # ojo con los NA no se pueden hacer operaciones
```

```
## [1] NA
```

```
var(x, na.rm = TRUE)
```

```
## [1] 352.3333
```

```
mean(x, na.rm = TRUE) # Se utiliza esta funcion para poder calcular o hacer operaciones
```

```
## [1] 12.33333
```

Boxplots, Diagramas de caja, o grafico de bigotes

El bigote que se esta dibujando puede llegar a ocupar como mucho 1.5 veces por debajo del primer cuartil el IQR o 1.5 veces por arriba del rango intercuartilico, el bigote alzara hasta el minimo o maximo de esos valores. Si el bigote llega a tener mas de 1.5 veces del IQR, esos valores quedaran siendo outliers o valores atipicos.

Explicacion: los extremos, los valores b_{inf}, b_{sup}, son los bigotes (whiskers) del grafico. Si m y M son el minimo y el maximo de la variable cuantitativa, entonces los extremos se calculan del siguiente modo:

$b_{inf} = \max\{m, Q_{0.25} - 1.5(Q_{0.75} - Q_{0.25})\}$

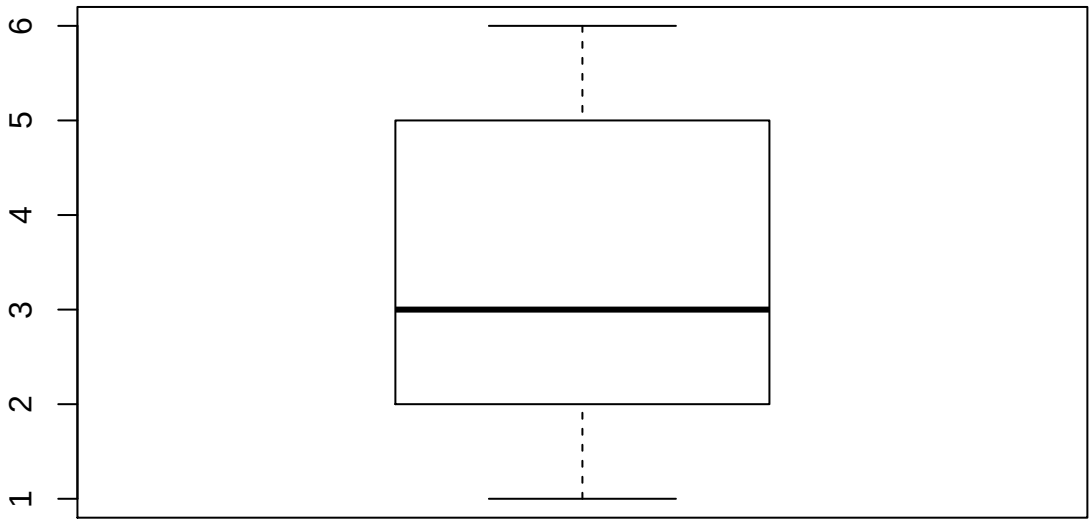
$b_{sup} = \max\{m, Q_{0.75} + 1.5(Q_{0.75} - Q_{0.25})\}$

Valores atipicos o outliers: que son los que estan mas alla de los bigores. Se marcan como puntos aislados.

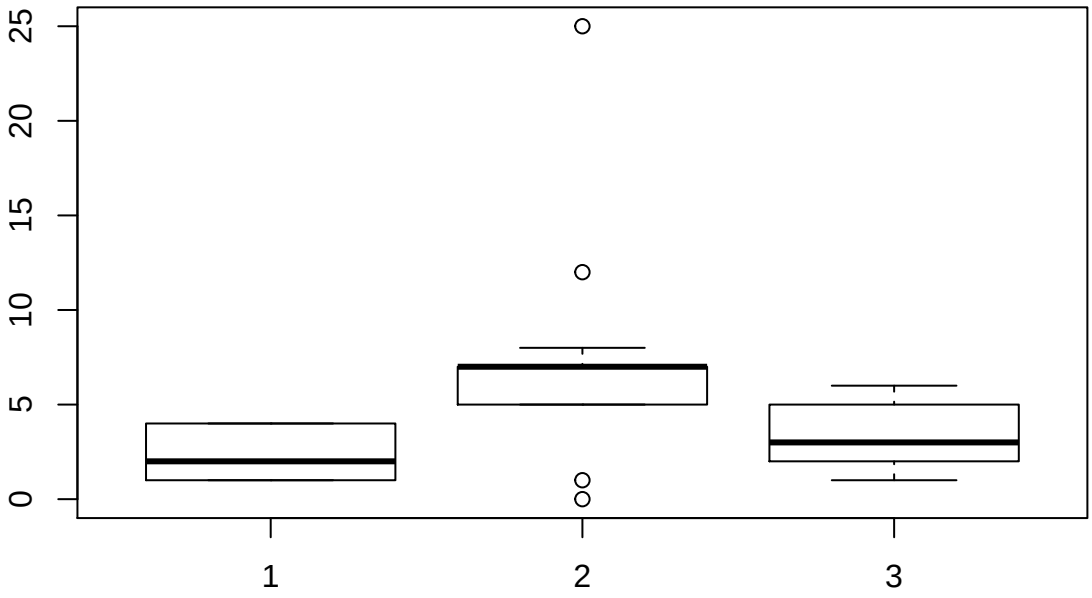
Por su definicion, concluimos que los bigotes marcan el minimo y el maximo de la variable cuantitativa, a no ser qiue haya datos muy alejados de la caja intercuantilica.

En tal caso, el bigote inferior marca el valor 1.5 veces el rango intercuantilico por debajo de $Q_{0,25}$, mientras que el superior marca el valor de 1.5 veces el rango intercuantilico por encima de $Q_{0.75}$.

```r
boxplot(dados2, main = "Un diagrama de caja")
```




```r
dados = c(1,6,7,5,7,8,5,7,0,7,5,7,12,25)
boxplot(dado, dados, dados2) # con tres cajas para comparar variables.
```




```r
iris # son 5 variables, pero la ultima es categorica
```

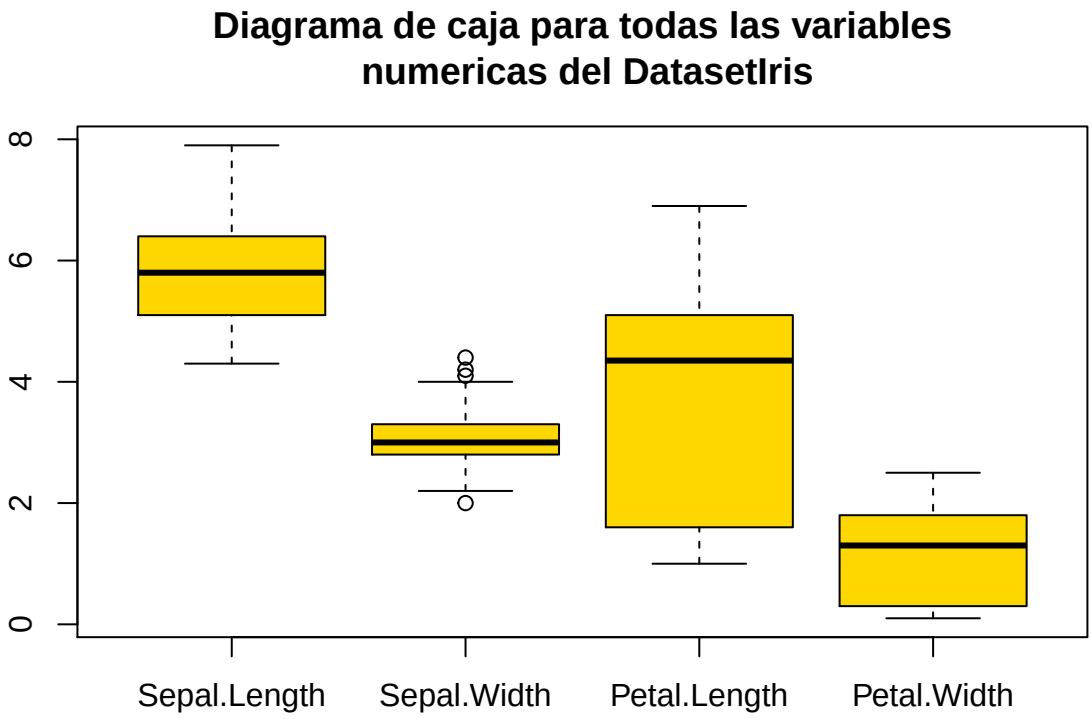
```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 1            5.1         3.5          1.4         0.2     setosa
## 2            4.9         3.0          1.4         0.2     setosa
## 3            4.7         3.2          1.3         0.2     setosa
## 4            4.6         3.1          1.5         0.2     setosa
## 5            5.0         3.6          1.4         0.2     setosa
## 6            5.4         3.9          1.7         0.4     setosa
## 7            4.6         3.4          1.4         0.3     setosa
## 8            5.0         3.4          1.5         0.2     setosa
```

```
## 9           4.4         2.9          1.4         0.2     setosa
## 10          4.9         3.1          1.5         0.1     setosa
## 11          5.4         3.7          1.5         0.2     setosa
## 12          4.8         3.4          1.6         0.2     setosa
## 13          4.8         3.0          1.4         0.1     setosa
## 14          4.3         3.0          1.1         0.1     setosa
## 15          5.8         4.0          1.2         0.2     setosa
## 16          5.7         4.4          1.5         0.4     setosa
## 17          5.4         3.9          1.3         0.4     setosa
## 18          5.1         3.5          1.4         0.3     setosa
## 19          5.7         3.8          1.7         0.3     setosa
## 20          5.1         3.8          1.5         0.3     setosa
## 21          5.4         3.4          1.7         0.2     setosa
## 22          5.1         3.7          1.5         0.4     setosa
## 23          4.6         3.6          1.0         0.2     setosa
## 24          5.1         3.3          1.7         0.5     setosa
## 25          4.8         3.4          1.9         0.2     setosa
## 26          5.0         3.0          1.6         0.2     setosa
## 27          5.0         3.4          1.6         0.4     setosa
## 28          5.2         3.5          1.5         0.2     setosa
## 29          5.2         3.4          1.4         0.2     setosa
## 30          4.7         3.2          1.6         0.2     setosa
## 31          4.8         3.1          1.6         0.2     setosa
## 32          5.4         3.4          1.5         0.4     setosa
## 33          5.2         4.1          1.5         0.1     setosa
## 34          5.5         4.2          1.4         0.2     setosa
## 35          4.9         3.1          1.5         0.2     setosa
## 36          5.0         3.2          1.2         0.2     setosa
## 37          5.5         3.5          1.3         0.2     setosa
## 38          4.9         3.6          1.4         0.1     setosa
## 39          4.4         3.0          1.3         0.2     setosa
## 40          5.1         3.4          1.5         0.2     setosa
## 41          5.0         3.5          1.3         0.3     setosa
## 42          4.5         2.3          1.3         0.3     setosa
## 43          4.4         3.2          1.3         0.2     setosa
## 44          5.0         3.5          1.6         0.6     setosa
## 45          5.1         3.8          1.9         0.4     setosa
## 46          4.8         3.0          1.4         0.3     setosa
## 47          5.1         3.8          1.6         0.2     setosa
## 48          4.6         3.2          1.4         0.2     setosa
## 49          5.3         3.7          1.5         0.2     setosa
## 50          5.0         3.3          1.4         0.2     setosa
## 51          7.0         3.2          4.7         1.4 versicolor
## 52          6.4         3.2          4.5         1.5 versicolor
## 53          6.9         3.1          4.9         1.5 versicolor
## 54          5.5         2.3          4.0         1.3 versicolor
## 55          6.5         2.8          4.6         1.5 versicolor
## 56          5.7         2.8          4.5         1.3 versicolor
## 57          6.3         3.3          4.7         1.6 versicolor
## 58          4.9         2.4          3.3         1.0 versicolor
## 59          6.6         2.9          4.6         1.3 versicolor
## 60          5.2         2.7          3.9         1.4 versicolor
## 61          5.0         2.0          3.5         1.0 versicolor
## 62          5.9         3.0          4.2         1.5 versicolor
```

```
## 63          6.0         2.2          4.0         1.0 versicolor
## 64          6.1         2.9          4.7         1.4 versicolor
## 65          5.6         2.9          3.6         1.3 versicolor
## 66          6.7         3.1          4.4         1.4 versicolor
## 67          5.6         3.0          4.5         1.5 versicolor
## 68          5.8         2.7          4.1         1.0 versicolor
## 69          6.2         2.2          4.5         1.5 versicolor
## 70          5.6         2.5          3.9         1.1 versicolor
## 71          5.9         3.2          4.8         1.8 versicolor
## 72          6.1         2.8          4.0         1.3 versicolor
## 73          6.3         2.5          4.9         1.5 versicolor
## 74          6.1         2.8          4.7         1.2 versicolor
## 75          6.4         2.9          4.3         1.3 versicolor
## 76          6.6         3.0          4.4         1.4 versicolor
## 77          6.8         2.8          4.8         1.4 versicolor
## 78          6.7         3.0          5.0         1.7 versicolor
## 79          6.0         2.9          4.5         1.5 versicolor
## 80          5.7         2.6          3.5         1.0 versicolor
## 81          5.5         2.4          3.8         1.1 versicolor
## 82          5.5         2.4          3.7         1.0 versicolor
## 83          5.8         2.7          3.9         1.2 versicolor
## 84          6.0         2.7          5.1         1.6 versicolor
## 85          5.4         3.0          4.5         1.5 versicolor
## 86          6.0         3.4          4.5         1.6 versicolor
## 87          6.7         3.1          4.7         1.5 versicolor
## 88          6.3         2.3          4.4         1.3 versicolor
## 89          5.6         3.0          4.1         1.3 versicolor
## 90          5.5         2.5          4.0         1.3 versicolor
## 91          5.5         2.6          4.4         1.2 versicolor
## 92          6.1         3.0          4.6         1.4 versicolor
## 93          5.8         2.6          4.0         1.2 versicolor
## 94          5.0         2.3          3.3         1.0 versicolor
## 95          5.6         2.7          4.2         1.3 versicolor
## 96          5.7         3.0          4.2         1.2 versicolor
## 97          5.7         2.9          4.2         1.3 versicolor
## 98          6.2         2.9          4.3         1.3 versicolor
## 99          5.1         2.5          3.0         1.1 versicolor
## 100         5.7         2.8          4.1         1.3 versicolor
## 101         6.3         3.3          6.0         2.5  virginica
## 102         5.8         2.7          5.1         1.9  virginica
## 103         7.1         3.0          5.9         2.1  virginica
## 104         6.3         2.9          5.6         1.8  virginica
## 105         6.5         3.0          5.8         2.2  virginica
## 106         7.6         3.0          6.6         2.1  virginica
## 107         4.9         2.5          4.5         1.7  virginica
## 108         7.3         2.9          6.3         1.8  virginica
## 109         6.7         2.5          5.8         1.8  virginica
## 110         7.2         3.6          6.1         2.5  virginica
## 111         6.5         3.2          5.1         2.0  virginica
## 112         6.4         2.7          5.3         1.9  virginica
## 113         6.8         3.0          5.5         2.1  virginica
## 114         5.7         2.5          5.0         2.0  virginica
## 115         5.8         2.8          5.1         2.4  virginica
## 116         6.4         3.2          5.3         2.3  virginica
```

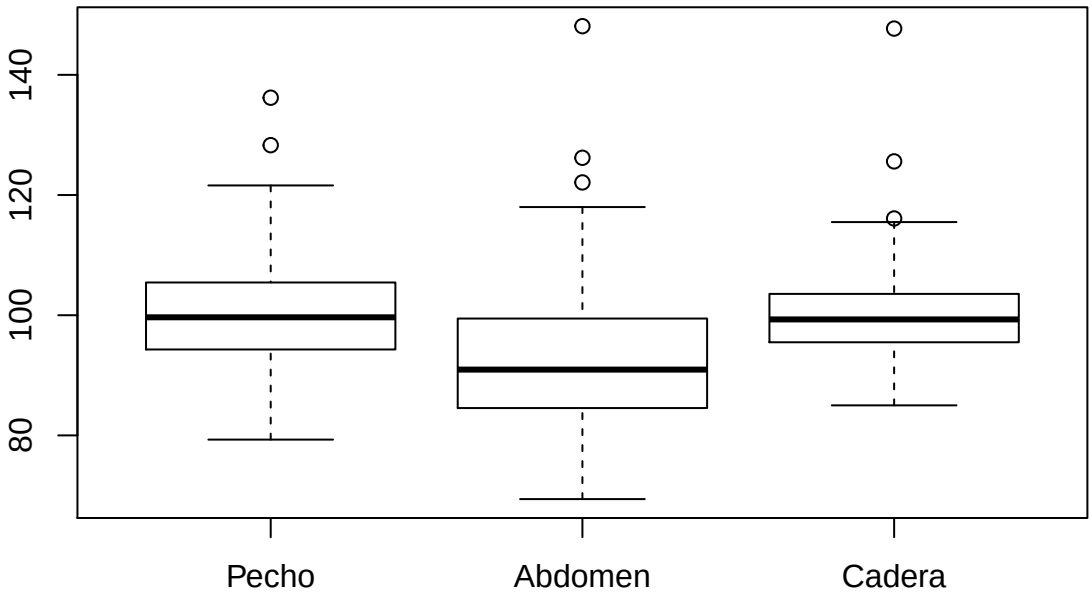
```
## 117          6.5         3.0          5.5         1.8  virginica
## 118          7.7         3.8          6.7         2.2  virginica
## 119          7.7         2.6          6.9         2.3  virginica
## 120          6.0         2.2          5.0         1.5  virginica
## 121          6.9         3.2          5.7         2.3  virginica
## 122          5.6         2.8          4.9         2.0  virginica
## 123          7.7         2.8          6.7         2.0  virginica
## 124          6.3         2.7          4.9         1.8  virginica
## 125          6.7         3.3          5.7         2.1  virginica
## 126          7.2         3.2          6.0         1.8  virginica
## 127          6.2         2.8          4.8         1.8  virginica
## 128          6.1         3.0          4.9         1.8  virginica
## 129          6.4         2.8          5.6         2.1  virginica
## 130          7.2         3.0          5.8         1.6  virginica
## 131          7.4         2.8          6.1         1.9  virginica
## 132          7.9         3.8          6.4         2.0  virginica
## 133          6.4         2.8          5.6         2.2  virginica
## 134          6.3         2.8          5.1         1.5  virginica
## 135          6.1         2.6          5.6         1.4  virginica
## 136          7.7         3.0          6.1         2.3  virginica
## 137          6.3         3.4          5.6         2.4  virginica
## 138          6.4         3.1          5.5         1.8  virginica
## 139          6.0         3.0          4.8         1.8  virginica
## 140          6.9         3.1          5.4         2.1  virginica
## 141          6.7         3.1          5.6         2.4  virginica
## 142          6.9         3.1          5.1         2.3  virginica
## 143          5.8         2.7          5.1         1.9  virginica
## 144          6.8         3.2          5.9         2.3  virginica
## 145          6.7         3.3          5.7         2.5  virginica
## 146          6.7         3.0          5.2         2.3  virginica
## 147          6.3         2.5          5.0         1.9  virginica
## 148          6.5         3.0          5.2         2.0  virginica
## 149          6.2         3.4          5.4         2.3  virginica
## 150          5.9         3.0          5.1         1.8  virginica
```

```r
nueva = iris[1:4]

boxplot(nueva, main =
          "Diagrama de caja para todas las variables\n numericas del DatasetIris",
        col = "gold") # para todas las variables numericas. Se puede ver que la Sepal.Width tiene outli
```



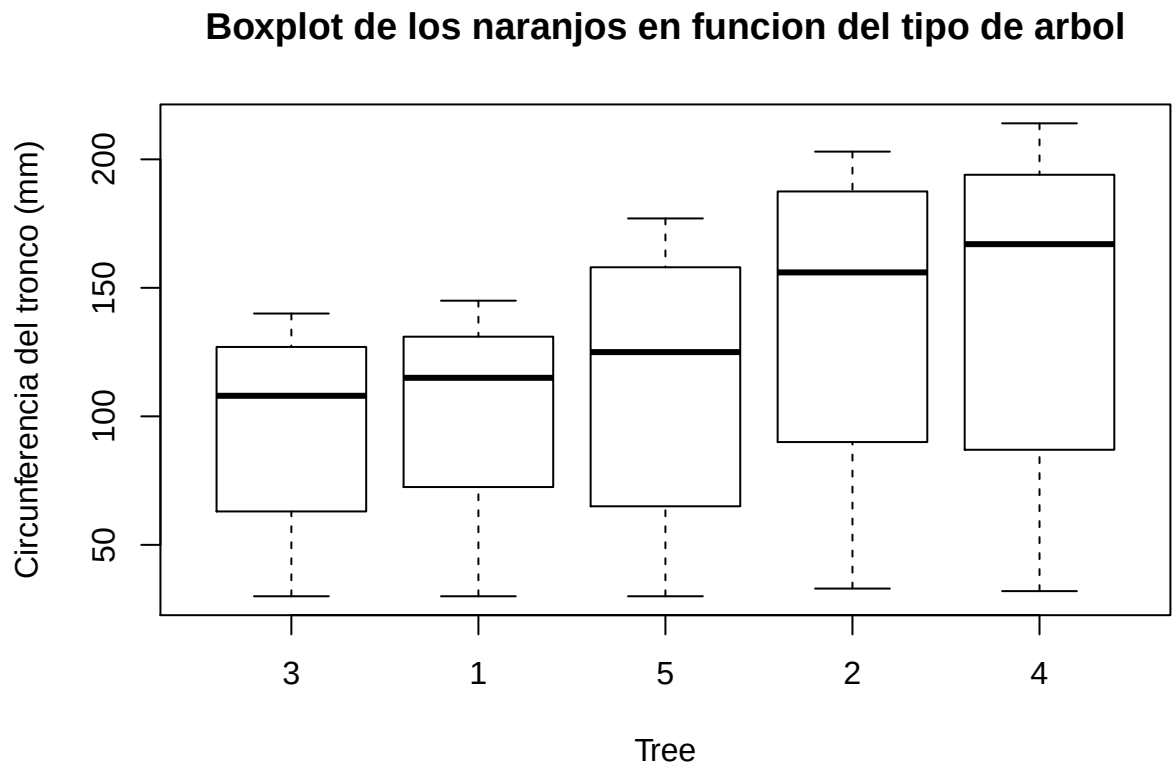
Ejemplo 2:

```
body = read.table("/Users/heinerleivagmail.com/Documents/GitHub/r-basic/data/bodyfat.txt", header = TRU

boxplot(body[,7:9], names = c("Pecho", "Abdomen", "Cadera"))
```



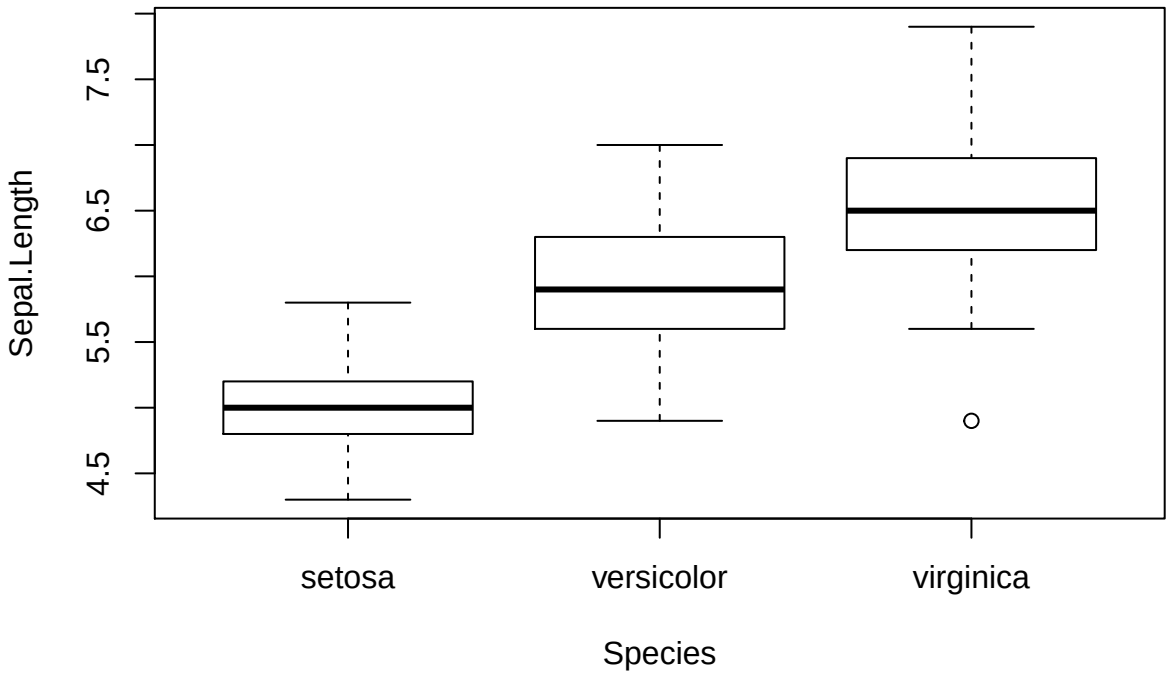
Ejemplo 3:

```
boxplot(circumference~Tree, data = Orange, ylab = "Circunferencia del tronco (mm)", main = "Boxplot de
```



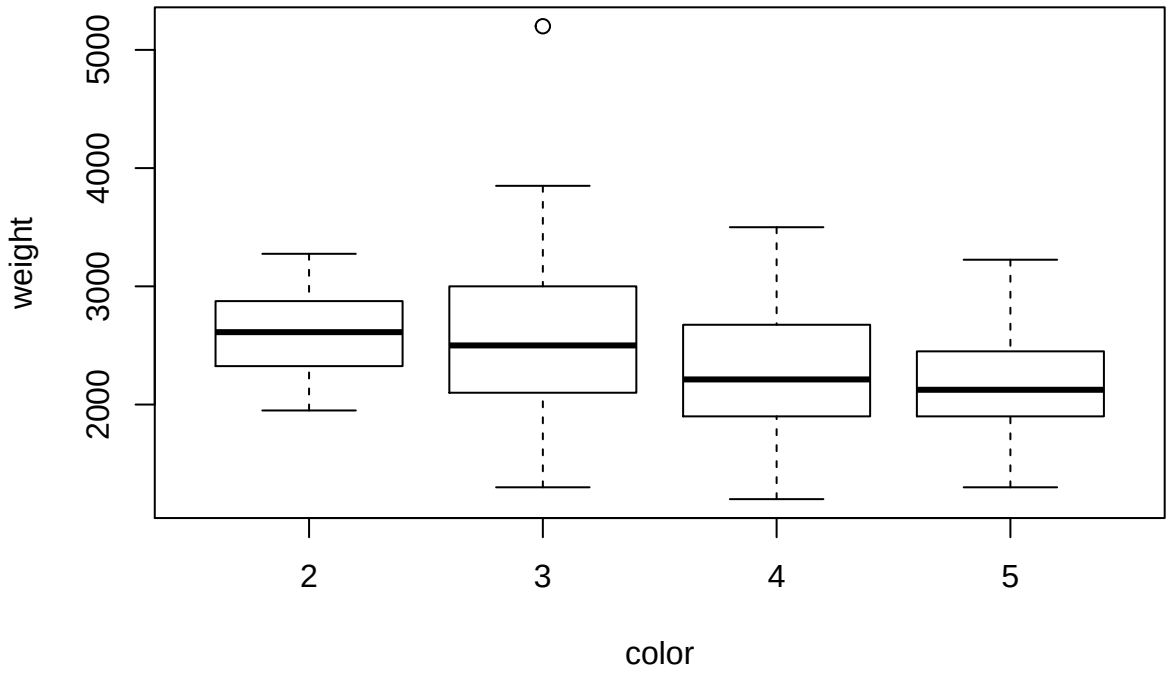
Ejemplo 4:

```
boxplot(Sepal.Length~Species, data = iris) # utilizo el Sepal Length como un factor
```



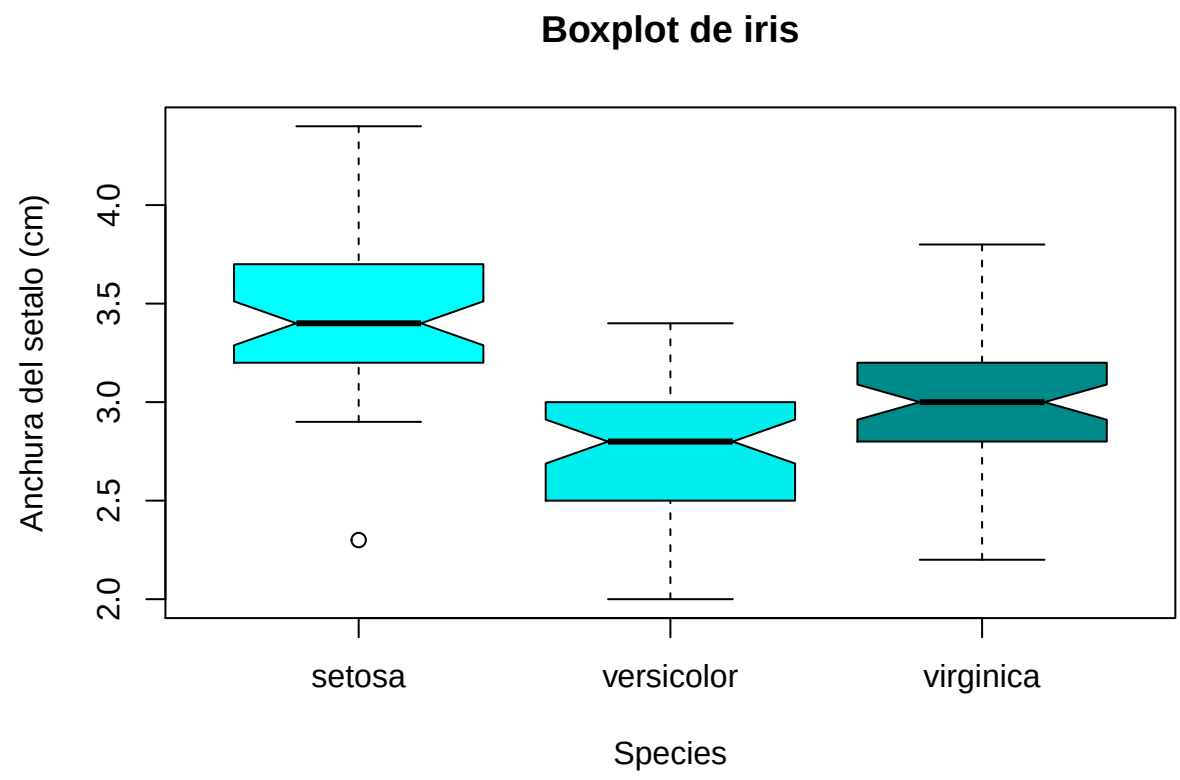
Ejemplo 5:

```
boxplot(weight~color, data = cangrejos)
```



Ejemplo 6:

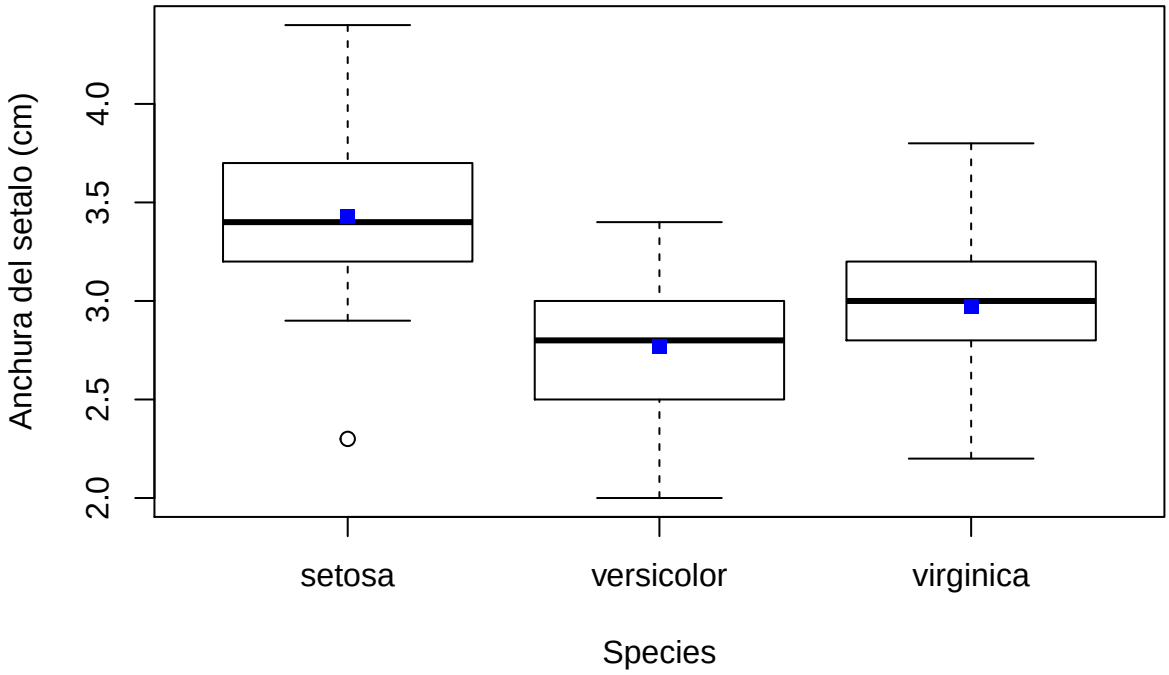
```
boxplot(Sepal.Width~Species, data = iris, ylab = "Anchura del setalo (cm)", notch = TRUE, col = c("cyan
```



```r
# En este caso no se solapan las muescas.
```

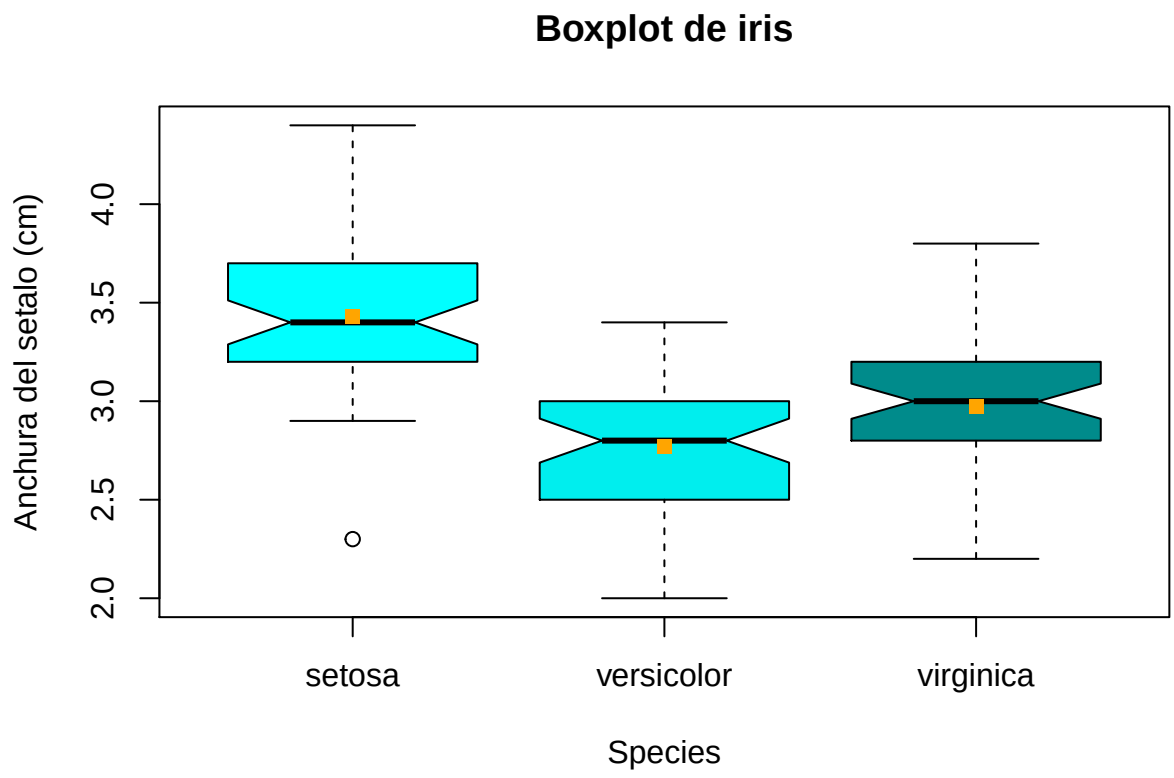
Ejemplo 7:

```r
# Agregando un distintivo a la media

boxplot(Sepal.Width~Species, data = iris, ylab = "Anchura del setalo (cm)")
medias = aggregate(Sepal.Width~Species, data = iris, FUN = mean)
points(medias, col = "blue", pch = 15)
```



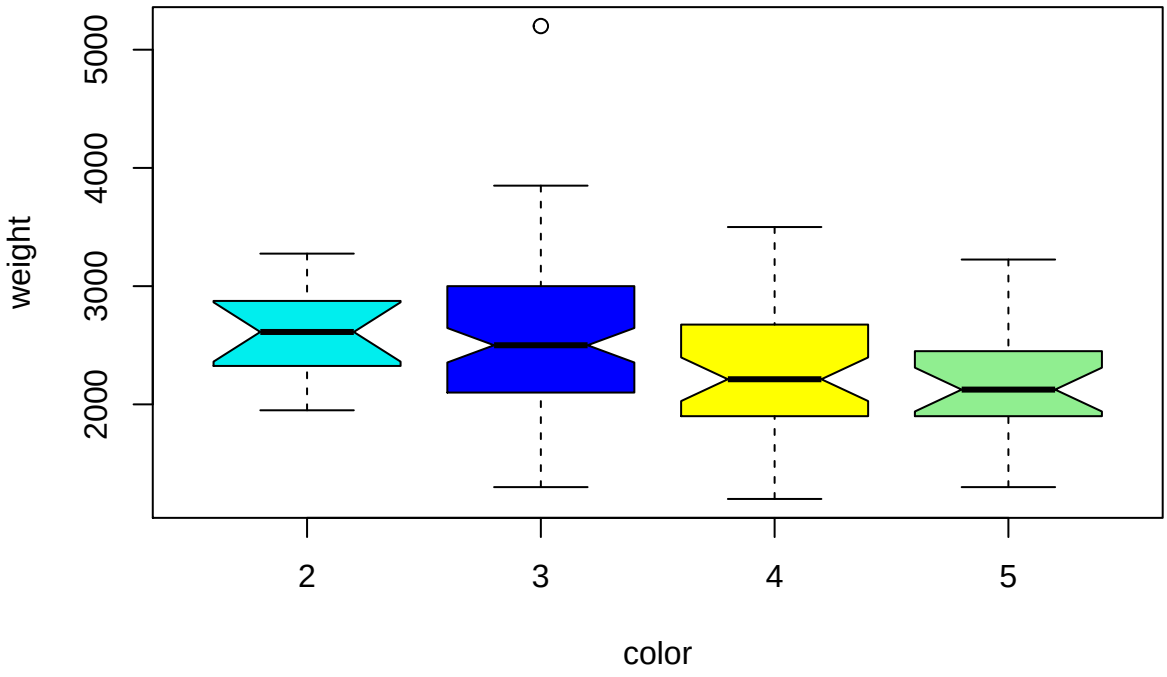
Ejemplo 8:

```r
boxplot(Sepal.Width~Species, data = iris, ylab = "Anchura del setalo (cm)", notch = TRUE, col = c("cyan
medias = aggregate(Sepal.Width~Species, data = iris, FUN = mean)
points(medias, col = "orange", pch = 15)
```



Ejemplo 9:

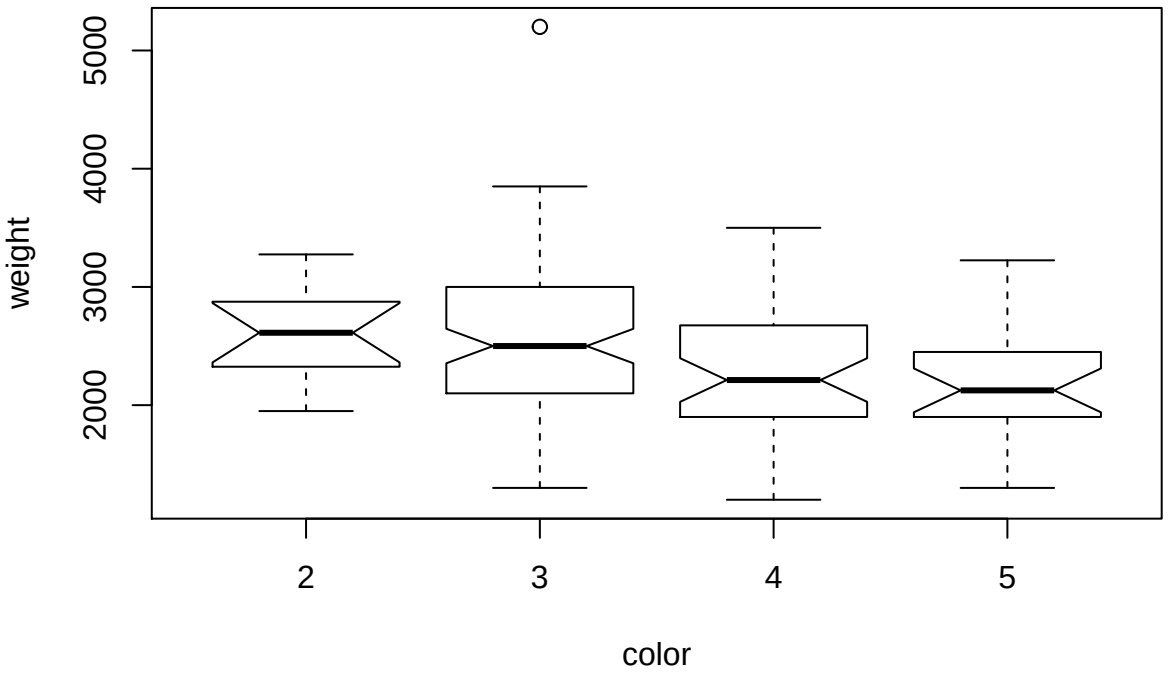
```
str(boxplot(weight~color, data = cangrejos, notch = TRUE, col = c("Cyan2", "Blue", "Yellow", "LightGreen
```




```
## List of 6
##  $ stats: 'integer' num [1:5, 1:4] 1950 2325 2612 2875 3275 ...
##  $ n    : num [1:4] 12 95 44 22
##  $ conf : num [1:2, 1:4] 2362 2863 2354 2646 2028 ...
```

```
##  $ out  : num 5200
##  $ group: num 2
##  $ names: chr [1:4] "2" "3" "4" "5"
```

```
boxplot(weight~color, data = cangrejos, notch = TRUE)$out # para consultar por partes lo que da el str,
```




```
## [1] 5200
```

Tarea:

Analisis de spray insecticida

Cargar los datos del data set `R` e inspeccionar con str.

```
data = InsectSprays
head(data)
```

```
##   count spray
## 1    10     A
## 2     7     A
## 3    20     A
## 4    14     A
## 5    14     A
## 6    12     A
```

```
str(data)
```

```
## 'data.frame':    72 obs. of  2 variables:
##  $ count: num  10 7 20 14 14 12 10 23 17 20 ...
##  $ spray: Factor w/ 6 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...
```

Realice un resumen estadistico de los datos:

```
by(data$count, data$spray, FUN = summary)
```

```
## data$spray: A
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##    7.00   11.50   14.00   14.50   17.75   23.00
## ------------------------------------------------------------
## data$spray: B
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    7.00   12.50   16.50   15.33   17.50   21.00
## ------------------------------------------------------------
## data$spray: C
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.000   1.500   2.083   3.000   7.000
## ------------------------------------------------------------
## data$spray: D
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.000   3.750   5.000   4.917   5.000  12.000
## ------------------------------------------------------------
## data$spray: E
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    2.75    3.00    3.50    5.00    6.00
## ------------------------------------------------------------
## data$spray: F
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.00   12.50   15.00   16.67   22.50   26.00
```

```
# Los botes que matan mas bichos son el A, B y F, pero los botes C, D y E, tienen menos recorrido, es d
```

Inspeccionar la desviacion:

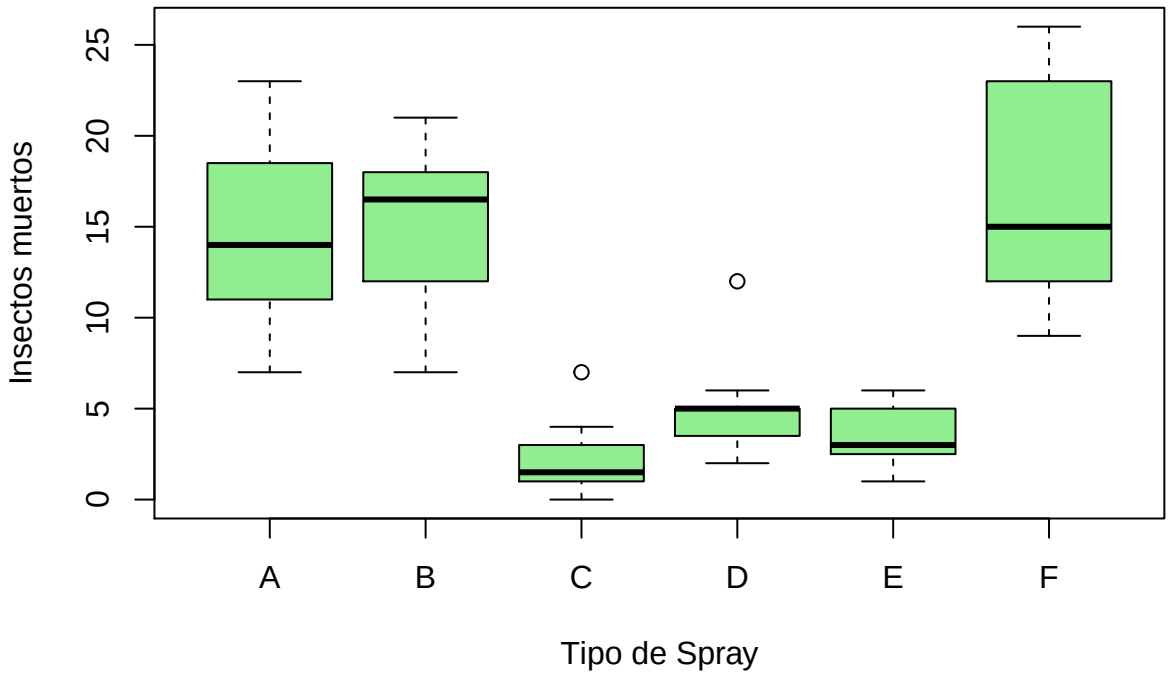
```
aggregate(count~spray, data = data, FUN = sd)
```

```
##   spray    count
## 1     A 4.719399
## 2     B 4.271115
## 3     C 1.975225
## 4     D 2.503028
## 5     E 1.732051
## 6     F 6.213378
```

```
# En efecto los botes C, D, y E son mas especificos y su boxplot sera mas pequenno.
```

Cree graficos de caja de los insecticidas por tipos

```
boxplot(count~spray, data = data, col = "lightgreen", xlab ="Tipo de Spray", ylab = "Insectos muertos")
```




```
# La caja tipo F tiene el mayor IQR.
```