

**UNIVERSIDAD DE EL SALVADOR  
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE  
DEPARTAMENTO DE MATEMÁTICA**



**Licenciatura en Estadística**

**Control Estadístico del Paquete R**

**”UNIDAD DOS”**

**Alumna:  
Erika Beatríz Guillén Pineda**

**Fecha de elaboración  
Santa Ana - 27 de noviembre de 2015**

# 1. ANÁLISIS ESTADÍSTICO DE LOS DATOS

- Ejemplo:
- 1) Activar el directorio de trabajo

```
getwd()

## [1] "C:/Users/User/Documents/TODAS_PRACTICAS"

setwd("C:/Users/User/Documents/TODAS_PRACTICAS")
```

- 2) Crear un nuevo Script y llamarle "Script07-DatosDiscretos"
- 3) Crear el vector de datos

```
Hijos <- c(2, 1, 2, 1, 4, 2, 3, 0, 2, 3, 3, 2, 1, 0, 2, 4, 1, 2,
          1, 3, 4, 1, 2, 3, 1, 5, 2, 3, 1, 2)
data.entry(Hijos)
Hijos

## [1] 2 1 2 1 4 2 3 0 2 3 3 2 1 0 2 4 1 2 1 3 4 1 2 3 1 5 2 3 1 2

length(Hijos)

## [1] 30
```

- 4) Guardar el vector de datos en un archivo de texto

```
write(Hijos, "Hijos.txt")
```

- 5) Limpiar el área de trabajo (Workspace)

```
ls()

## [1] "Hijos"

rm(list=ls(all=TRUE));
ls()

## character(0)
```

- 6) Leer o recuperar el vector de datos o archivo de texto

```
X <- scan("Hijos.txt", what = integer(0), na.strings = "NA", flush=FALSE)
ls()
```

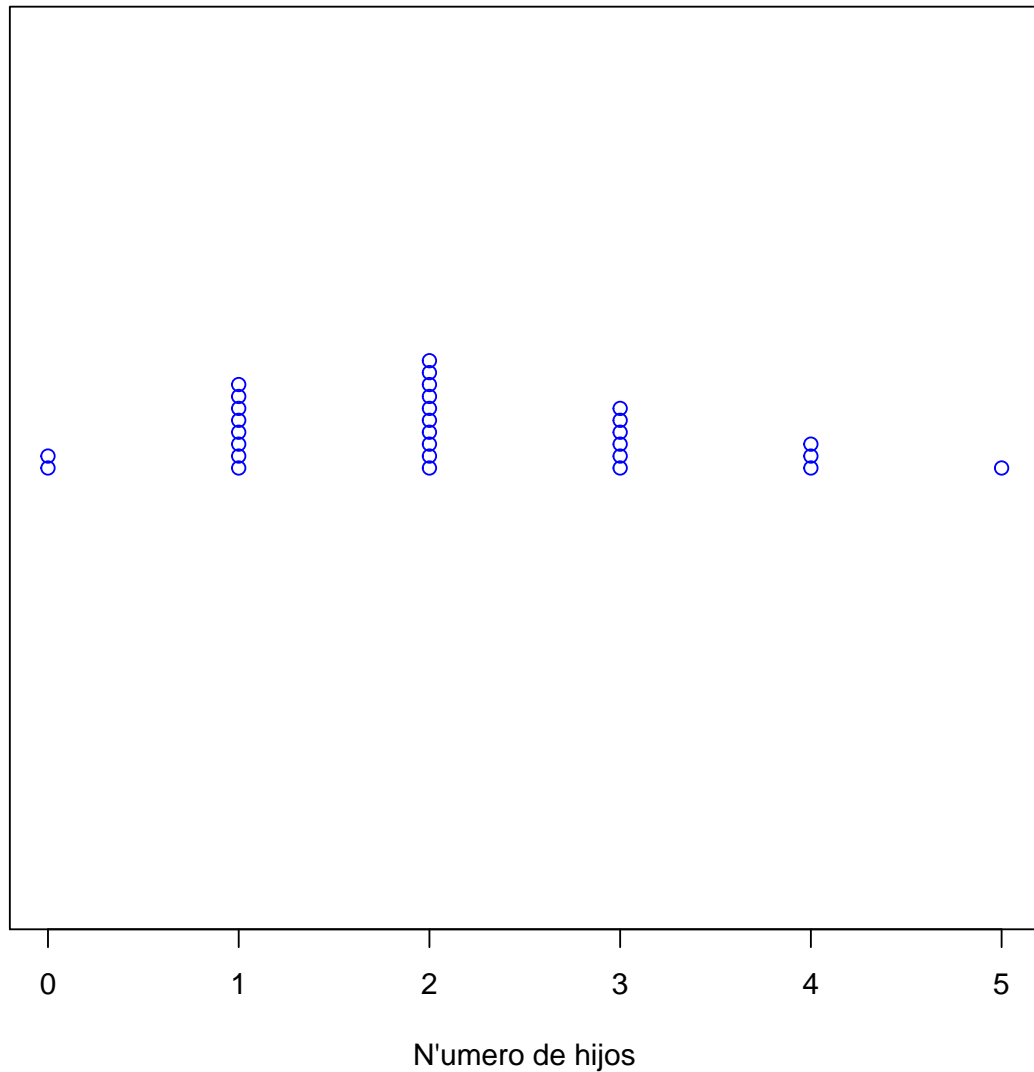
```
## [1] "X"
```

```
# Si el vector contiene caracteres se usa: what = character()
# Si el vector contiene reales se ocupa: what = double(0)
```

- 7) Elaborar el gráfico de puntos y diagrama de tallo-hojas (stem-and-leaf)

```
# Gráfico de puntos
stripchart(X, method="stack", vertical=FALSE, col="blue", pch=1, main="Gráfico de\npuntos", xlab="Número de hijos")
```

### Gráfico de puntos



```
# Observación: method puede ser:
# "overplot" (los puntos coincidentes son superpuestos)
# "jitter" (los puntos se ven como alejados o inquietos)
# "stack" (los puntos coincidentes son apilados, uno tras otro)
```

- 8) Crear la tabla de frecuencias completa

```
# Frecuencias individuales
fab <- table(X); fab # Frecuencias absolutas

## X
## 0  1  2  3  4  5
## 2  8 10  6  3  1
```

```

fre <- fab/length(X); fre # Frecuencias relativas

## X
##      0      1      2      3      4      5
## 0.06666667 0.26666667 0.33333333 0.20000000 0.10000000 0.03333333

Fac <- cumsum(fab); Fac # Frecuencias acumuladas

## 0 1 2 3 4 5
## 2 10 20 26 29 30

Far <- Fac/length(X); Far # Frecuencias acumuladas relativas

##      0      1      2      3      4      5
## 0.06666667 0.33333333 0.66666667 0.86666667 0.96666667 1.00000000

# Tabla de frecuencias completa
options(digits=2)
tabla <- data.frame(fab=fab, fre=fre, Fac=Fac, Far=Far)
names(tabla) <- c("X", "fab", "free.X", "fre", "Fac", "Far")
tabla

##   X fab free.X   fre Fac   Far
## 0 0   2     0 0.067   2 0.067
## 1 1   8     1 0.267  10 0.333
## 2 2  10     2 0.333  20 0.667
## 3 3   6     3 0.200  26 0.867
## 4 4   3     4 0.100  29 0.967
## 5 5   1     5 0.033  30 1.000

tfre <- data.frame(X=tabla$X, fab=tabla$fab, fre=tabla$fre,
                   Fac=tabla$Fac, Far=tabla$Far)
tfre

##   X fab   fre Fac   Far
## 1 0   2 0.067   2 0.067
## 2 1   8 0.267  10 0.333
## 3 2  10 0.333  20 0.667
## 4 3   6 0.200  26 0.867
## 5 4   3 0.100  29 0.967
## 6 5   1 0.033  30 1.000

# Note que el cuadro resultante no tiene la presentaci\on deseada para
# presentarla en un informe. Sin embargo, si estamos utilizando LATEX
# podemos utilizar la siguiente instrucc\on xtable(tfre) y con esto nos
# genera el c\odigo correspondiente para incorporarlo en nuestro archivo.

```

- 9) Calcular los estadísticos descriptivos de la variable

```
# Estadísticos de tendencia central de los datos
media <- mean(X, na.rm = FALSE);
media

## [1] 2.1

# na.rm = FALSE, le indica a R que los datos faltantes son omitidos
# en el cálculo de la media.
for(i in 1:length(X)) if (fab[i] == max(fab)) break()
moda <- names(fab[i]);
moda # R no tiene incorporada una función para la moda

## [1] "2"

mediana <- median(X);
mediana

## [1] 2

# Estadísticos de dispersión o variabilidad de los datos
range(X) # Devuelve el valor mínimo y máximo del conjunto de datos.

## [1] 0 5

cuasivar <- var(X);
cuasivar

## [1] 1.5

s <- sd(X);
s

## [1] 1.2

# Devuelve la cuasivarianza y la cuasivarianza muestral
quantile(X,c(0.25, 0.5, 0.75))

## 25% 50% 75%
##    1    2    3

# Cálculo de Q1, Q2, Q3
quantile(X, 0.6)
```

```
## 60%
## 2

# En general se pueden encontrar cualquier percentil
# Conocer un resumen de los datos
resumen <- summary(X);
resumen

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0      1.0      2.0      2.1      3.0      5.0

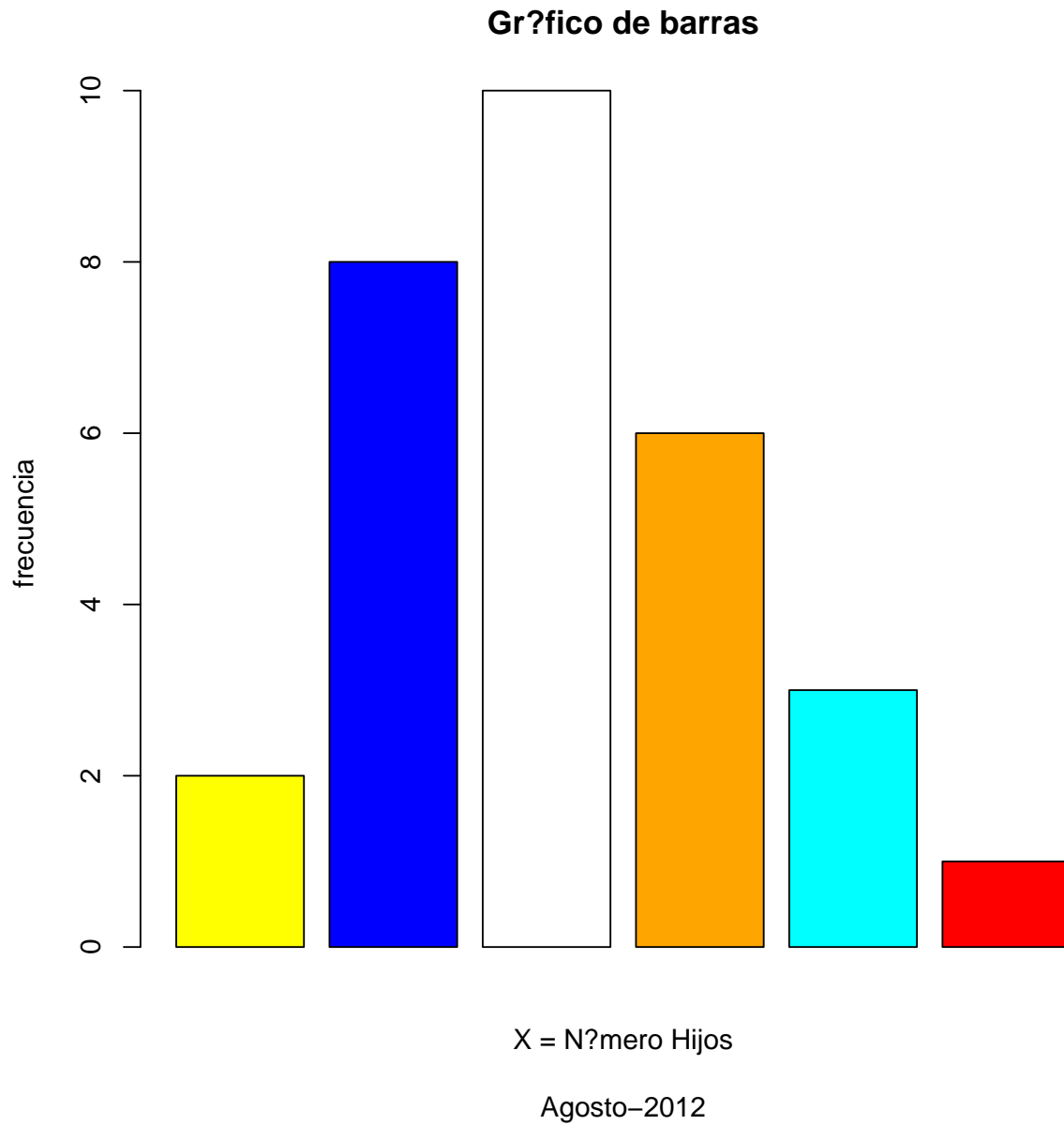
# Min, Q1, Median, Mean, Q3, Max
fivenum(X)

## [1] 0 1 2 3 5

# min, cuartil menor, mediana, cuartil mayor, max
```

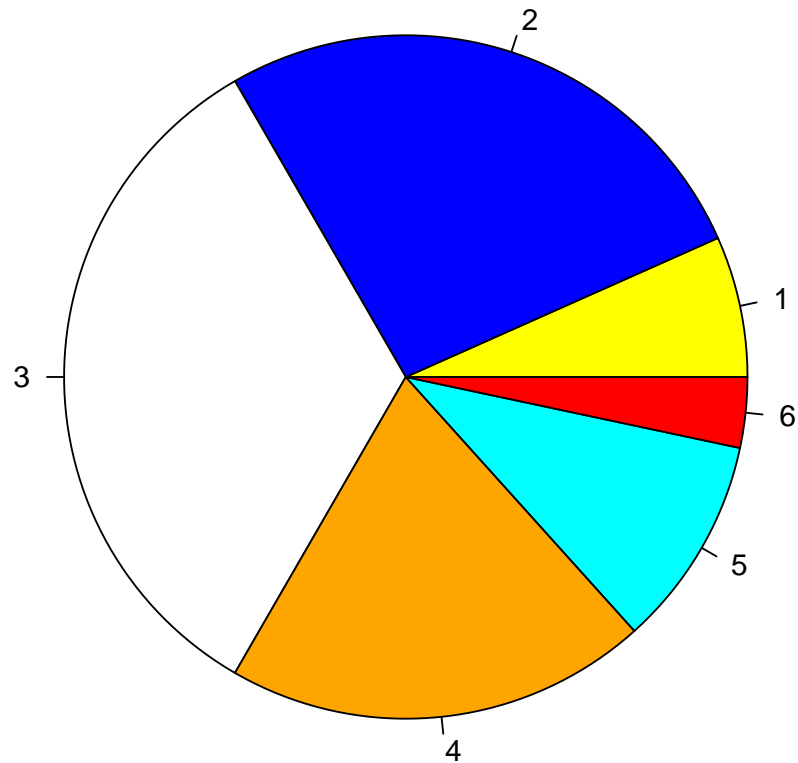
- 10) Elaborar los gráficos que se le pueden aplicar a la variable discreta

```
# Gráfico de barras (por ser pocos valores)
barplot(tfre[[2]], main="Gráfico de barras", xlab="X = Número Hijos\n",
        ylab="frecuencia", col=c("yellow", "blue", "white", "orange",
                                "cyan", "red"), sub="Agosto-2012")
```



```
# Gráfico de pastel (por ser pocos valores)
pie(tfre[[2]], main="Gráfico de pastel", xlab="Número Hijos \n",
    col=c("yellow", "blue", "white", "orange", "cyan", "red"),
    sub="Agosto-2012")
```

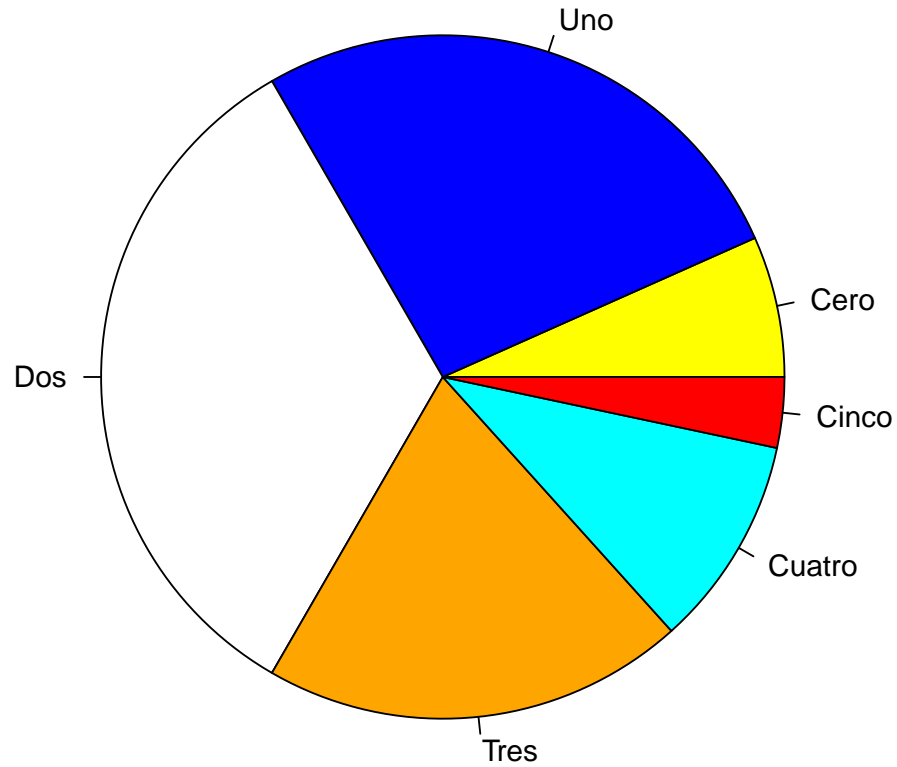


**Gráfico de pastel**

N'umero Hijos

Agosto-2012

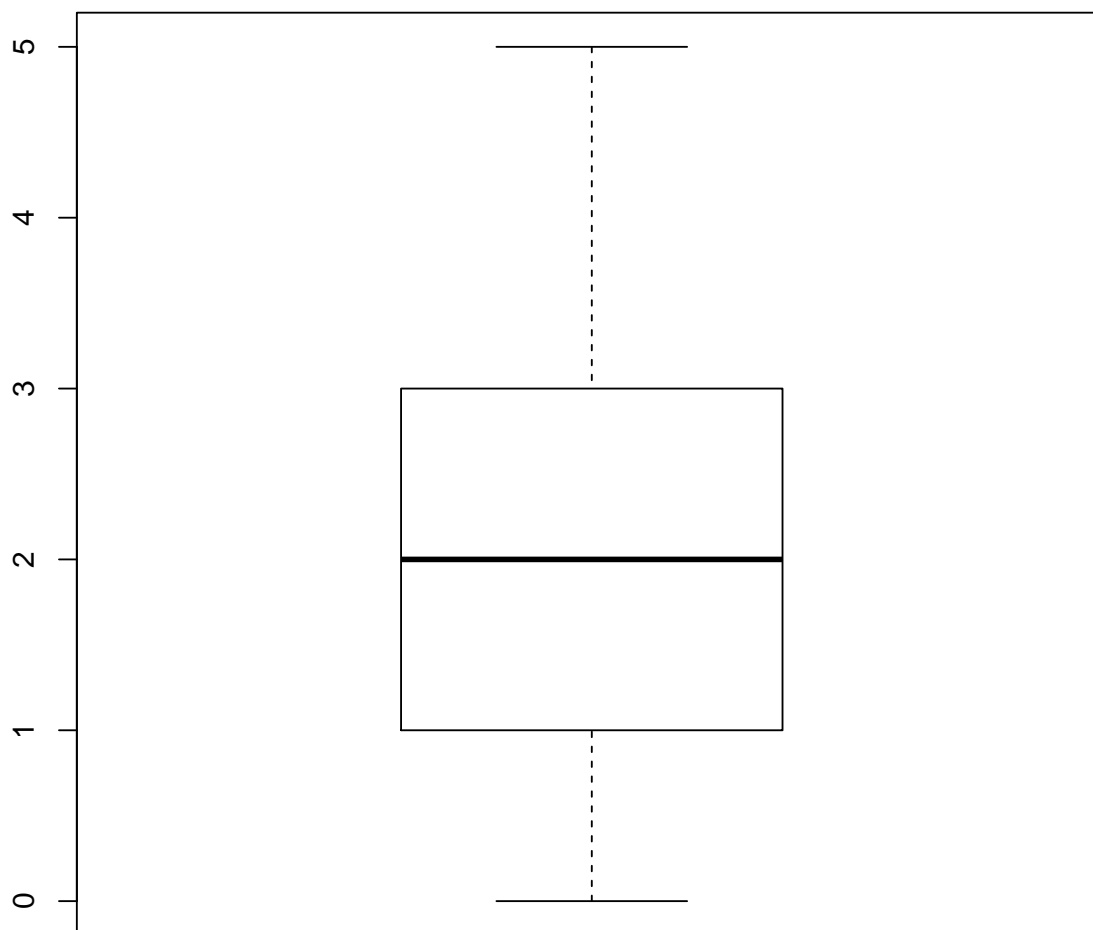
```
# Se puede especificar nombres para las categorías
names(fab) = c("Cero", "Uno", "Dos", "Tres", "Cuatro", "Cinco")
pie(fab, main="Gráfico de pastel", xlab="X = N'umero Hijos\n",
    col=c("yellow", "blue", "white", "orange", "cyan", "red"),
    sub="Agosto-2012")
```

**Gráfico de pastel**

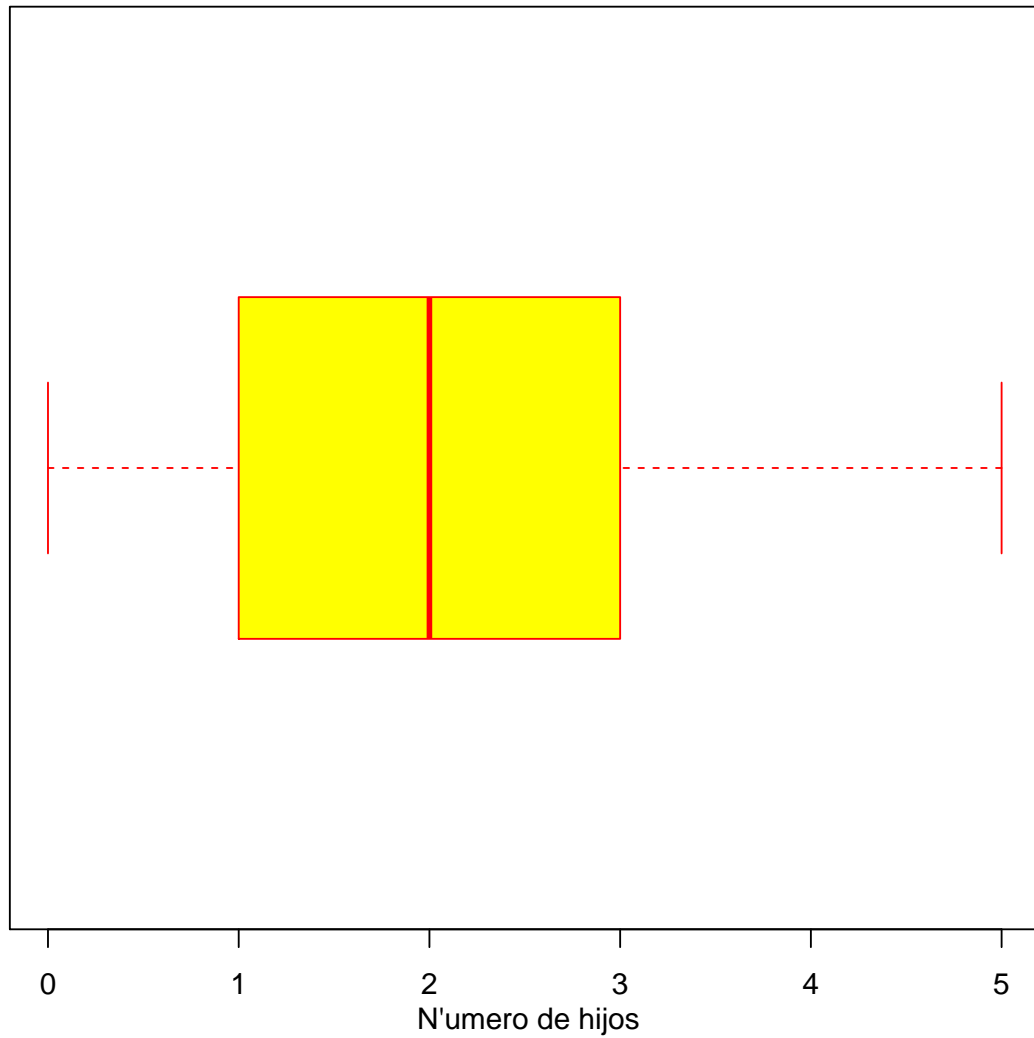
X = N'umero Hijos

Agosto-2012

```
# Gráfico de cajas (box-plot) es la representación gráfica de los cinco números  
# Horizontal  
boxplot(X, main="Gráfico de caja", ylab="Número de hijos")
```

**Gráfico de caja**

```
# Vertical  
boxplot(X, main="Gráfico de caja", xlab="Número de hijos",  
        plot=TRUE, border="red", col="yellow", horizontal=TRUE)
```

**Gráfico de caja**

```
# NOTE QUE TODOS LOS GRÁFICOS DE BARRAS Y DE PASTEL SON REALIZADOS  
# APARTIR DE UNA TABLA DE FRECUENCIA, LA CUAL SE INDICA EN tfre[[2]].  
# TAMBIÉN SE PUDO UTILIZAR tabla[[2]].
```