

UNIVERSIDAD DE EL SALVADOR

FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE



Universidad de El Salvador

Hacia la libertad por la cultura

PRACTICA 6 y 7 DE R

DOCENTE:

LIC. JAIME ISAAC PEÑA

PRESENTADO POR:

CRISTIAN ALBERTO ZALDAÑA ALVARADO

Índice

1. PRÁCTICA 6: Análisis de datos categóricos.	3
1.1. ESCALAS DE MEDICIÓN	3
1.2. ANÁLISIS ESTADÍSTICO DE DATOS CATEGÓRICOS.	3
2. PRÁCTICA 7: Análisis estadístico de datos univariados discretos con R.	10
2.1. ANÁLISIS ESTADÍSTICO DE LOS DATOS.	10

1. PRÁCTICA 6: Análisis de datos categóricos.

1.1. ESCALAS DE MEDICIÓN

Como la estadística analiza los datos y éstos son producto de las mediciones, necesitamos estudiar las escalas de medición. Este tema es de suma importancia, pues el tipo de escala de medición utilizado para reunir los datos ayuda a determinar el tipo de análisis a utilizar en los datos. Existen cuatro clases de escalas que aparecen de manera común en las ciencias: nominal, ordinal, de intervalo y de razón. Ellas difieren en el número de atributos matemáticos que poseen.

Los tipos de datos univariados que vamos a analizar en esta práctica son:

Categóricos. Tienen la característica de que todos los miembros de una categoría se consideran iguales en lo que se refiere a ese tipo. Este tipo de datos se subdivide en nominales y ordinales.

- **Nominales.** Los valores que pueden asumir sirven para clasificarlos pero no para ordenarlos. En caso de usarse números, sólo se adoptan como nombres o identificaciones.
- **Ordinales.** Los valores que puede asumir este tipo de datos son categorías que conllevan un juicio de valor que exige comparar a los diferentes elementos de la muestra con respecto a este tipo con el objeto de establecer un orden. Es decir, que los datos se organizan a través de las relaciones de igualdad, mayor o menor.

1.2. ANÁLISIS ESTADÍSTICO DE DATOS CATEGÓRICOS.

Ejemplo: Se realiza un estudio para conocer las preferencias sobre el tipo de gaseosa que se consume: CC-Coca Cola, "PC-Pepsi Cola, "SC-Salva Cola, para ello se toma una muestra aleatoria de 20 personas.

1. Crear un vector con el tipo de gaseosa y otro con la muestra generada aleatoriamente:

```
> Tipo <- c("CC", "PC", "SC"); Tipo  
[1] "CC" "PC" "SC"  
  
> Consumo <- sample(Tipo, 20, replace=TRUE)  
> Consumo  
  
[1] "SC" "CC" "SC" "SC" "PC" "CC" "PC" "CC" "SC" "PC" "PC" "PC" "PC" "PC" "CC"  
[16] "PC" "SC" "PC" "SC" "CC"
```

Genera una muestra de tamaño 20 obtenida de los elementos del vector Tipo y los elementos se seleccionan con reemplazamiento

Suponiendo que se quiere editar o agregar datos

```
> data.entry(Consumo)
```

2. Guarde el vector en un archivo de datos

Guardar los datos en su directorio de trabajo

```
> write(Consumo, "Consumo.txt")
```

3. Eliminar los objetos que existen en el espacio de trabajo (Workspace)

```
> #ls()  
> #rm(list=ls(all=TRUE))  
> #ls()
```

4. Leer o recuperar el vector de datos o archivo de texto

```
> Consumo <- scan("Consumo.txt", what = character(), na.strings = "NA",
+ flush=FALSE);Consumo

[1] "SC" "CC" "SC" "SC" "PC" "CC" "PC" "CC" "SC" "PC" "PC" "PC" "PC" "PC" "CC"
[16] "PC" "SC" "PC" "SC" "CC"

> ls()

[1] "Consumo" "Tipo"
```

Si el vector contiene caracteres se ocupa: `what = character()` `na.strings = "NA"`, le indica a R que los valores faltantes son identificados con "NA"

5. Crear la tabla de distribución de frecuencias y proporciones

```
> freq <- table(Consumo)
> freq

Consumo
CC PC SC
 5  9  6

> prop <- table(Consumo)/length(Consumo)
> prop

Consumo
  CC   PC   SC
0.25 0.45 0.30
```

Note que la salida por defecto no es para nada atractiva en comparación con el resto de paquetes estadísticos

En cambio, si estamos usando LATEX y queremos incorporar estos cuadros o cualquier otro podemos utilizar el comando `xtable(table(Consumo))` (NOTE QUE EL ARGUMENTO DEBE SER UN CUADRO), y con esto automáticamente se nos genera el código en LATEX y luego incorporarlo a nuestro informe, lo mejor de todo es que salida resultante es mucho más presentable.

6. Conocer un resumen de los datos

```
> summary(Consumo)

      Length      Class      Mode 
      20 character character
```

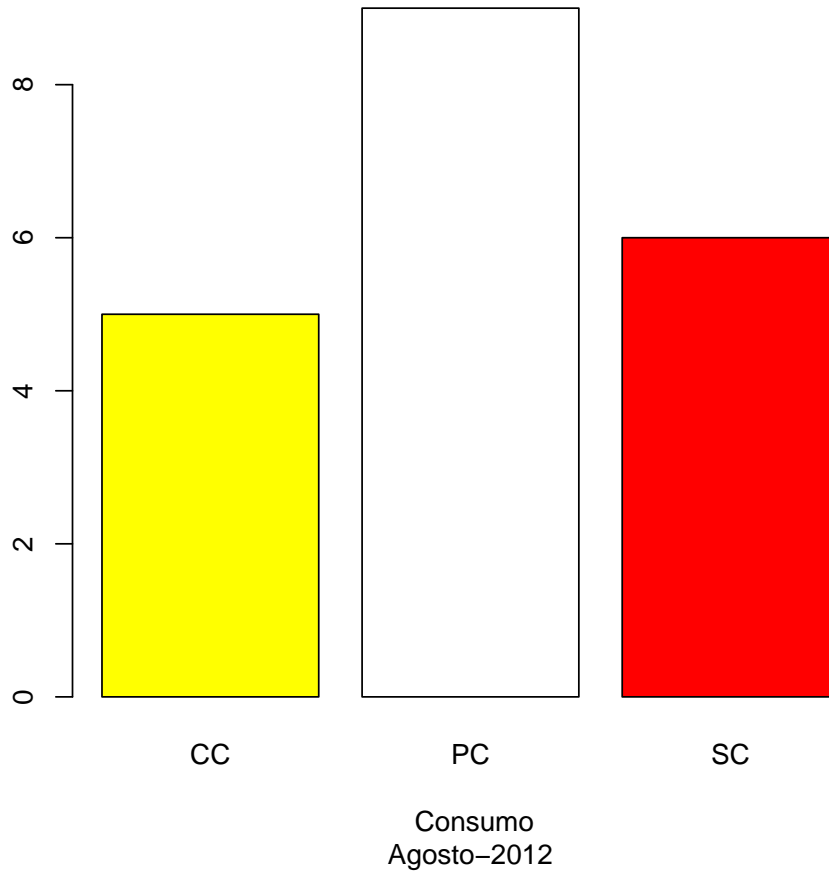
Note que por tratarse de variables cualitativas únicamente muestra el número de elementos, y el tipo de datos.

7. Realizar un gráfico de barras

Para las frecuencias absolutas

```
> barplot(freq, main="Gráfico de barras", xlab="Consumo", col=c("yellow", "white", "red"),
+ sub="Agosto-2012")
```

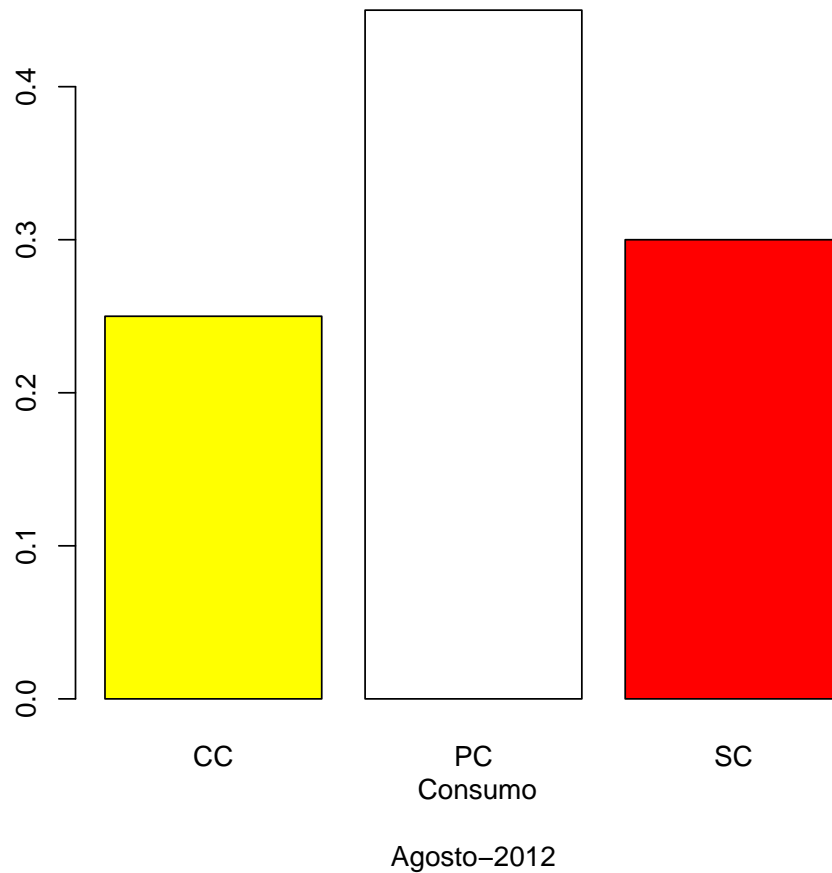
Gráfico de barras



Para las frecuencias relativas

```
> barplot(prop, main="Gráfico de barras", xlab=" Consumo\n", col=c("yellow", "white",  
+ "red"), sub="Agosto-2012")  
>
```

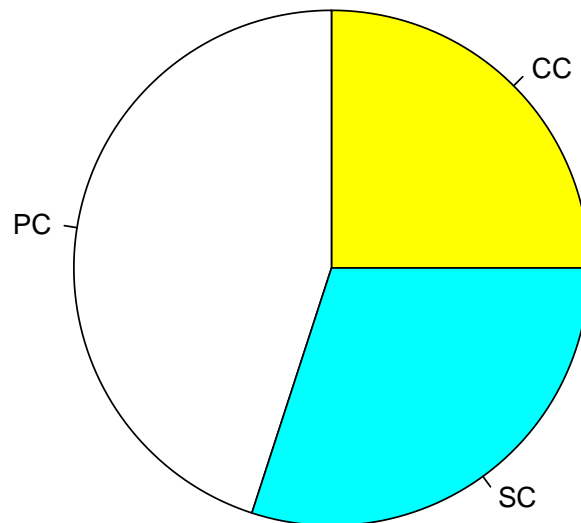
Gráfico de barras



8. Realizar un gráfico de pastel

```
> pie(frec, main="Gráfico de pastel", xlab="Tipo de Consumo", col=c("yellow", "white",  
+ "cyan"), sub="Agosto-2012")
```

Gráfico de pastel

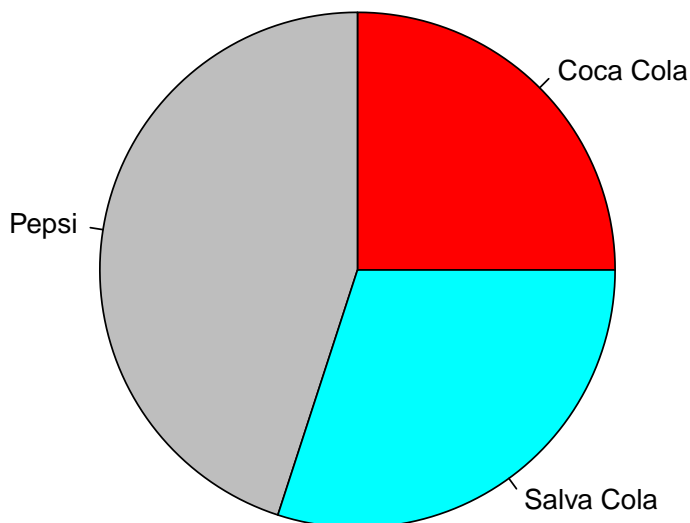


Tipo de Consumo
Agosto-2012

Se puede especificar nombres para las categorías y el color de los sectores

```
> names(frec) = c("Coca Cola", "Pepsi", "Salva Cola")  
> pie(frec, main="Gráfico de pastel", xlab=" Consumo", radius=0.8, col=c("red", "gray",  
+ "cyan"), sub="Agosto-2012")
```

Gráfico de pastel



Consumo Agosto-2012

Los colores se asignan dependiendo del orden en que han sido especificados por names()
Note con la instrucción radius se especifica el tamaño de la figura, mientras más cerca de uno (uno de menos uno) se encuentre más grande será (el ángulo cambia).

9. Colocar valores numéricos en los sectores del gráfico

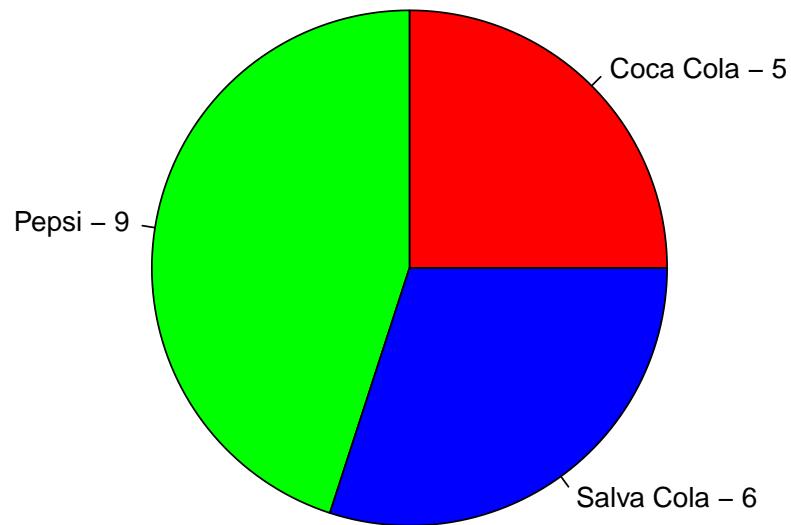
```
> n <- length(frec)
> hoja <- data.frame(frec); hoja

  Var1 Freq
1 Coca Cola    5
2  Pepsi     9
3 Salva Cola    6

> etiq <- c(paste(hoja$Var1, "-", hoja$Freq)); etiq
[1] "Coca Cola - 5" "Pepsi - 9"      "Salva Cola - 6"

> pie(frec, main="Gráfico de pastel", labels=etiq, col=rainbow(n), border=TRUE)
```


Gráfico de pastel



2. PRÁCTICA 7: Análisis estadístico de datos univariados discretos con R.

2.1. ANÁLISIS ESTADÍSTICO DE LOS DATOS.

1. 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)
> 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
```

2. Guardar el vector de datos en un archivo de texto.

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

3. 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] "Consumo" "etiq"      "frec"      "Hijos"     "hoja"      "n"          "prop"
[8] "Tipo"     "X"
```

Si el vector contiene caracteres se usa: `what = character()`

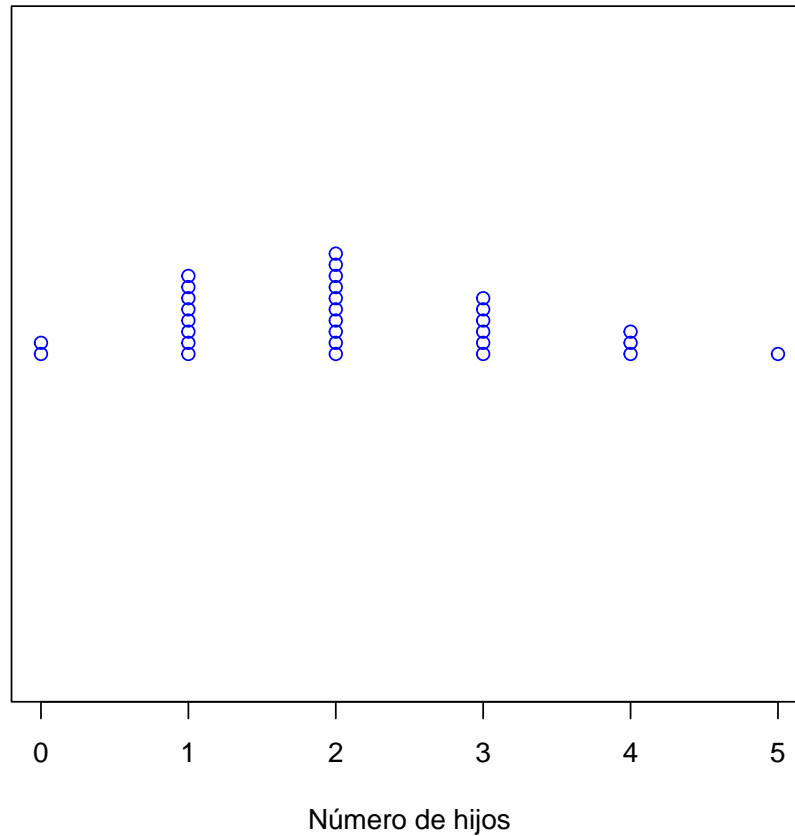
Si el vector contiene reales se ocupa: `what = double(0)`

4. 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\n
+ puntos", 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)

5. 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 presentación deseada para presentarla en un informe. Sin embargo, si estamos utilizando LATEX podemos utilizar la siguiente instrucción `xtable(tfre)` y con esto nos genera el código correspondiente para incorporarlo en nuestro archivo.

6. 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

Devuelve el valor mínimo y máximo del conjunto de datos.

```
> range(X)
```

```
[1] 0 5
```

Devuelve la cuasivarianza y la cuasivarianza muestral

```
> cuasivar <- var(X)
> cuasivar
```

```
[1] 1.5
```

```
> s <- sd(X)
> s
```

```
[1] 1.2
```

Cálculo de Q1, Q2, Q3

```
> quantile(X, c(0.25, 0.5, 0.75))
```

```
25% 50% 75%
  1   2   3
```

En general se pueden encontrar cualquier percentil

```
> quantile(X, 0.6)
```

```
60%
  2
```

Conocer un resumen de los datos

Min, Q1, Median, Mean, Q3, Max

```
> 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, cuartil menor, mediana, cuartil mayor, max

```
> fivenum(X)
```

```
[1] 0 1 2 3 5
```

7. 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")
```

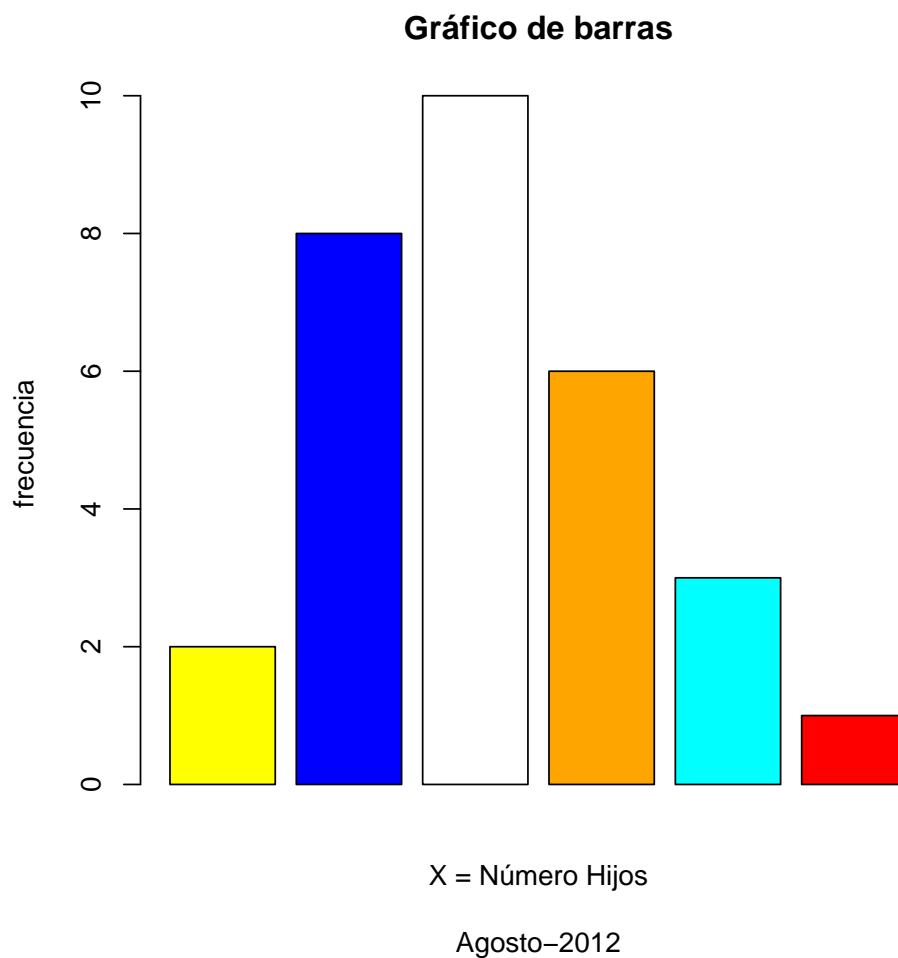
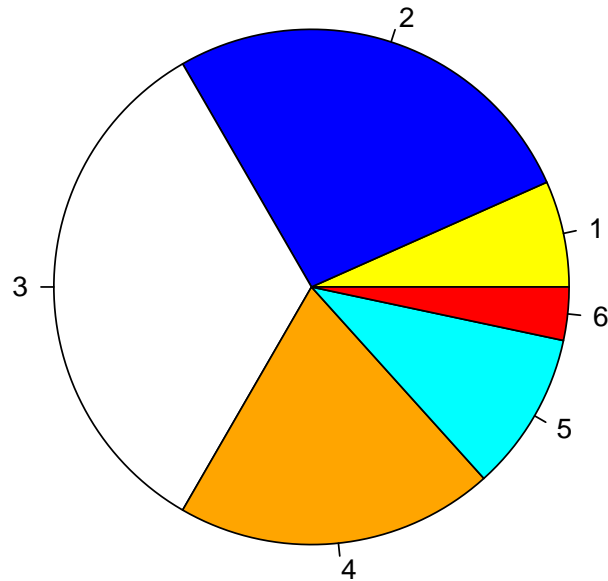


Gráfico de pastel (por ser pocos valores)

```
> 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



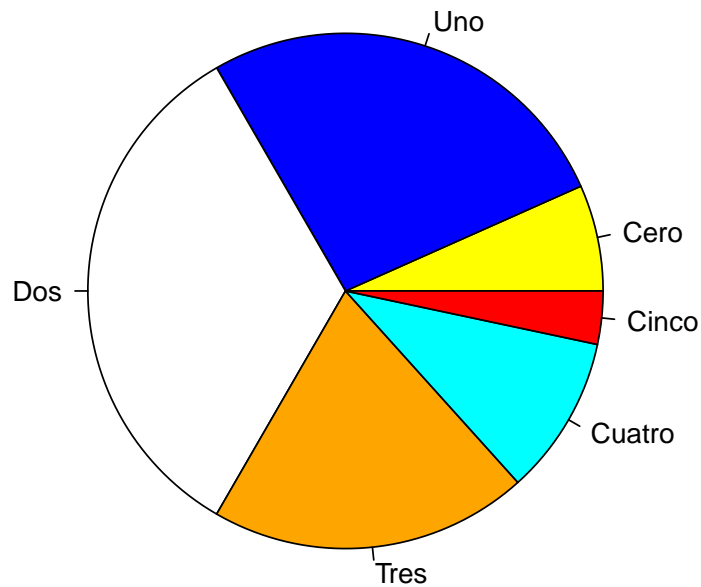
Número Hijos

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 = Número Hijos\n", col=c("yellow", "blue",  
+ "white", "orange", "cyan", "red"), sub="Agosto-2012")
```

Gráfico de pastel



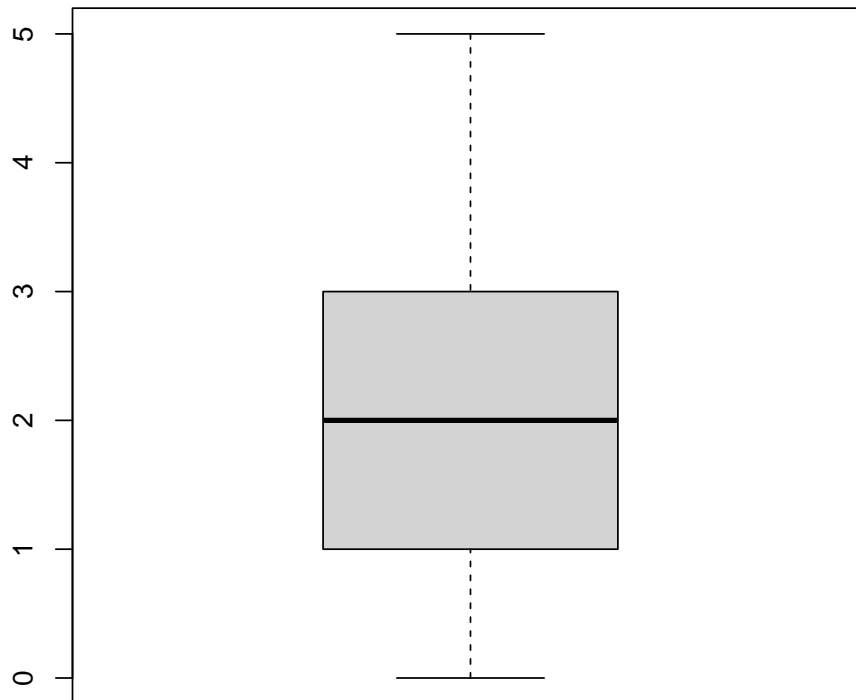
X = Número Hijos

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\n")
```

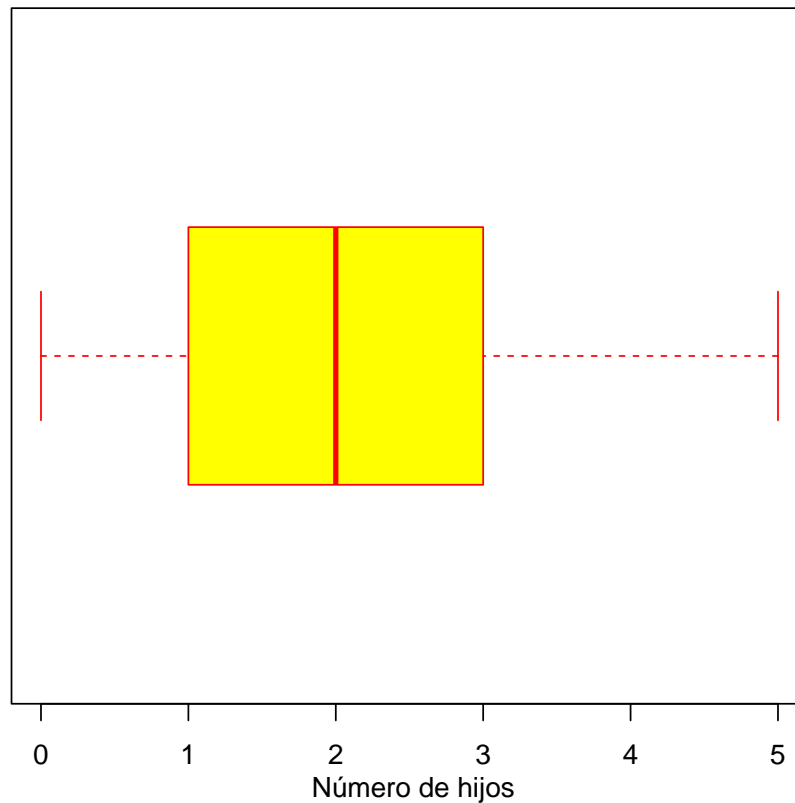

Gráfico de caja



Vertical

```
> boxplot(X, main="Gráfico de caja", xlab=" Número de hijos\n", 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]]`.