

**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. ESCALAS DE MEDICIÓN

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

**Categoricos:** 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.

## 1.1. 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) 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 Script06-DatosCategoricos
- 3) Crear un vector con el tipo de gaseosa y otro con la muestra generada aleatoriamente:

```
Tipo <- c("CC", "PC", "SC");
Tipo # crea un vector en las que contiene los tres tipos de refrescos

## [1] "CC" "PC" "SC"

Consumo <- sample(Tipo, 20, replace=TRUE);
Consumo

## [1] "CC" "PC" "PC" "SC" "PC" "CC" "PC" "CC" "CC" "PC" "SC" "SC" "PC" "PC"
## [15] "CC" "CC" "SC" "PC" "SC" "PC"

# 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)
```

- 4) Guarde el vector en un archivo de datos

```
# Guardar los datos en su directorio de trabajo
write(Consumo, "Consumo.txt")
```

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

```
ls()

## [1] "Consumo" "Tipo"

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

## character(0)
```

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

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

## [1] "CC" "PC" "PC" "SC" "PC" "CC" "PC" "CC" "CC" "PC" "SC" "SC" "PC" "PC"
## [15] "CC" "CC" "SC" "PC" "SC" "PC"

ls()

## [1] "Consumo"

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

7) Crear la tabla de distribución de frecuencias y proporciones

```
frec <- table(Consumo);
frec

## Consumo
## CC PC SC
## 6 9 5

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

## Consumo
## CC PC SC
## 0.30 0.45 0.25

# Note que la salida por defecto no es para nada atractiva en comparaci?n
# con el resto de paquetes estad?sticos. En cambio, si estamos usando
# LATEX y queremos incorporar estos cuadros o cualquier otro podemos utilizar
```

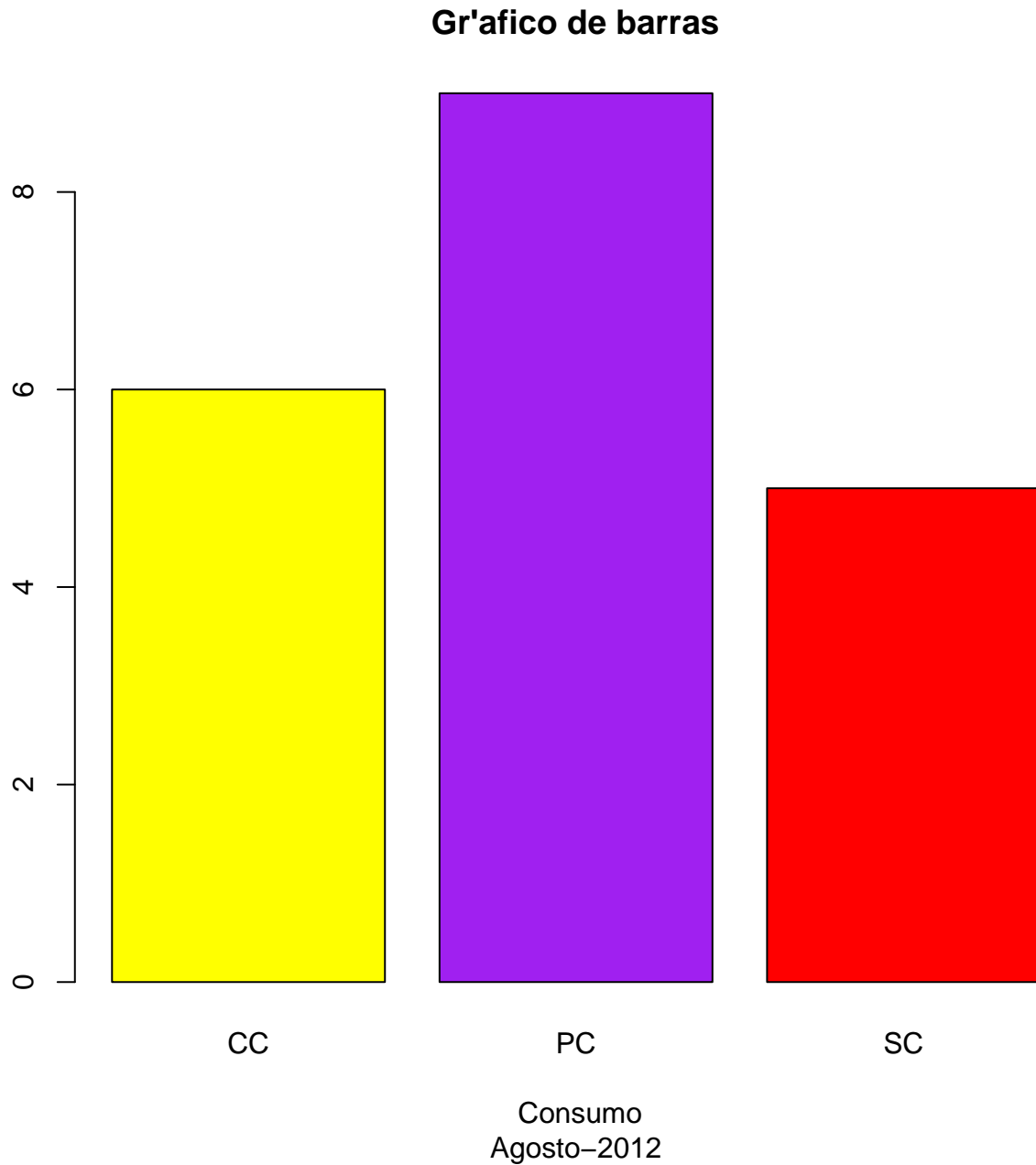
```
# el comando xtable(table(Consumo)) (NOTE QUE EL ARGUMENTO DEBE SER UN CUADRO),  
# y con esto automáticamente se nos genera el código en LATEX y luego  
# incorporarlo a nuestro informe, lo mejor de todo es que salida  
# resultante es mucho más presentable.
```

#### 8) Conocer un resumen de los datos

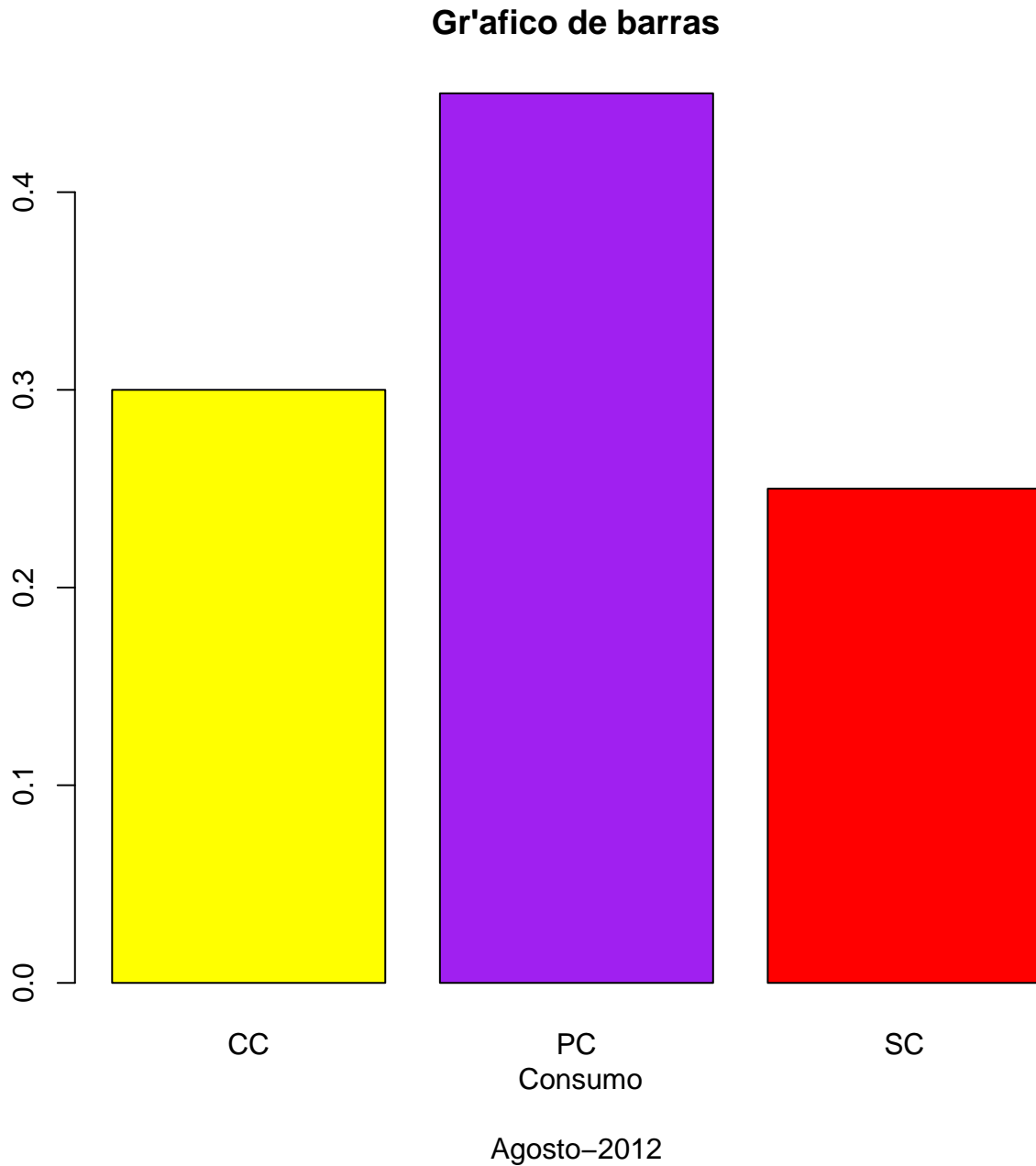
```
summary(Consumo)  
  
##      Length      Class      Mode  
##           20 character character  
  
# note que por tratarse de variables cualitativas únicamente  
# muestra el número de elementos, y el tipo de datos.
```

#### 9) Realizar un gráfico de barras

```
# Para las frecuencias absolutas  
barplot(frec, main="Gráfico de barras", xlab=" Consumo",  
        col=c("yellow", "purple", "red"), sub="Agosto-2012")
```

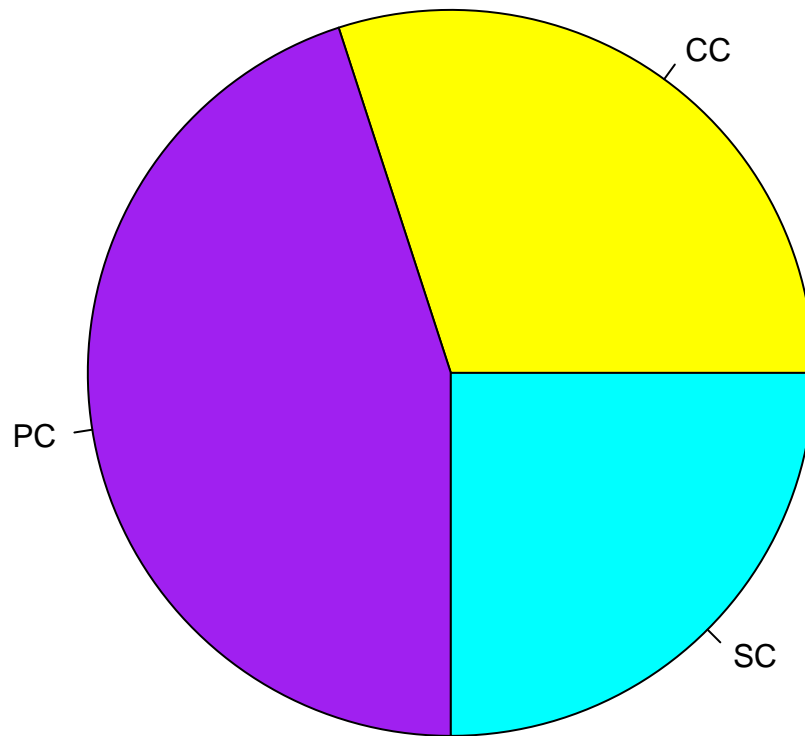


```
# Para las frecuencias relativas  
barplot(prop, main="Gráfico de barras", xlab=" Consumo\n",  
        col=c("yellow", "purple", "red"), sub="Agosto-2012")
```



10) Realizar un gráfico de pastel

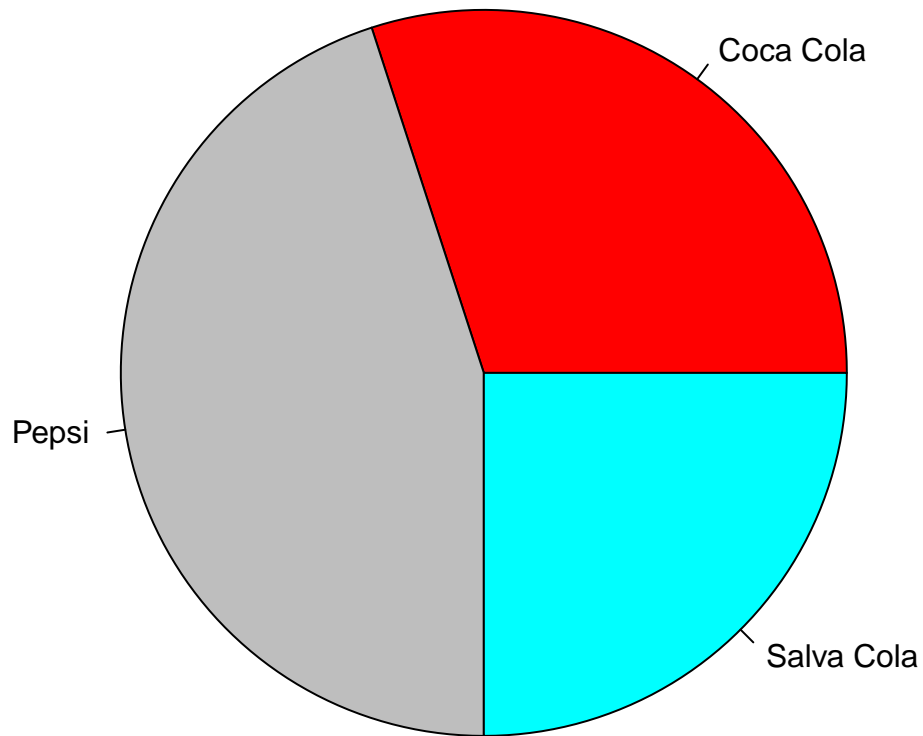
```
pie(frec, main="Gráfico de pastel", xlab="Tipo de Consumo",  
    col=c("yellow", "purple", "cyan"), sub="Agosto-2012")
```

**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).
```

11) Colocar valores numéricos en los sectores del gráfico

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

##          Var1 Freq
```



```
## 1 Coca Cola 6
## 2 Pepsi 9
## 3 Salva Cola 5

etiql <- c(paste(hoja$Var1, "-", hoja$Freq));
etiql

## [1] "Coca Cola - 6" "Pepsi - 9" "Salva Cola - 5"

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

**Gráfico de pastel**