

Tarea 1

Equipo 15

3/9/2021

1. Lea el archivo

Cargamos el archivo con su dirección relativa. Es decir, primero nos situamos en la carpeta en donde tenemos el .csv con `>setwd("absolute_dir_path")`

```
data <- read.csv('BASE.csv')
head(data)
```

```
## Tipo_de_Inmueble Operacion Provincia Superficie Precio_Venta Fecha.Venta
## 1 Casa Alquiler Barcelona 153.51 1,747,005.43 25/10/04
## 2 Casa Venta Girona 149.16 1,849,498.82 27/09/04
## 3 Casa Alquiler Tarragona 171.28 2,107,069.30 12/02/05
## 4 Casa Alquiler Girona 151.42 1,546,162.61 23/05/05
## 5 Casa Venta Lleida 161.95 1,801,051.65 02/05/05
## 6 Casa Venta Tarragona 179.01 1,591,157.40 16/09/05
## Vendedor Satisfaccion
## 1 Carmen 2.Regular
## 2 Carmen 2.Regular
## 3 Carmen 2.Regular
## 4 Carmen 2.Regular
## 5 Carmen 2.Regular
## 6 Carmen 2.Regular
```

2. Para cada una de las variables categóricas `Tipo_de_inmueble`, `Operacion`, `Provincia`, realice una tabla que cuente cuánto se repite cada categoría (ver función `table`) y lleve a cabo una gráfica de barras de conteo (ver función `barplot`).

Con la función `table(...)` realizamos una tabla para el conteo de las categorías.

```
table(data$Tipo_de_Inmueble)
```

```
##
## Casa Local
## 373 382
```

```
table(data$Operacion)
```

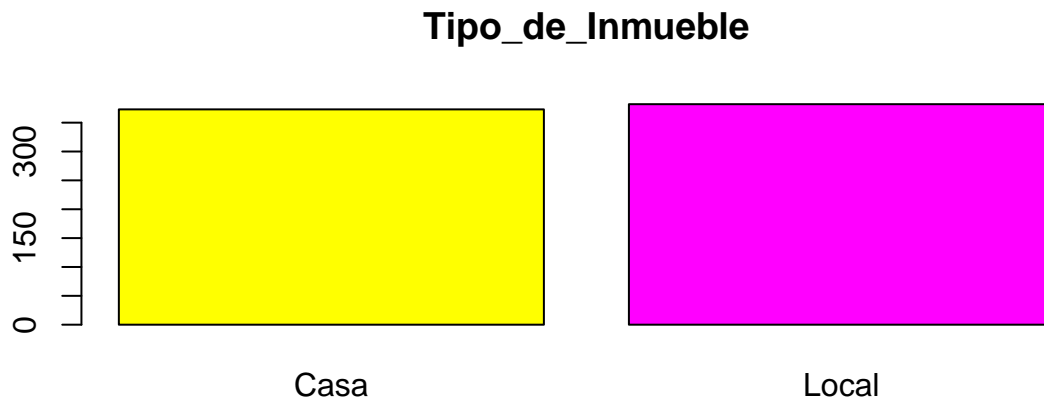
```
##
## Alquiler Venta
## 366 389
```

```
table(data$Provincia)
```

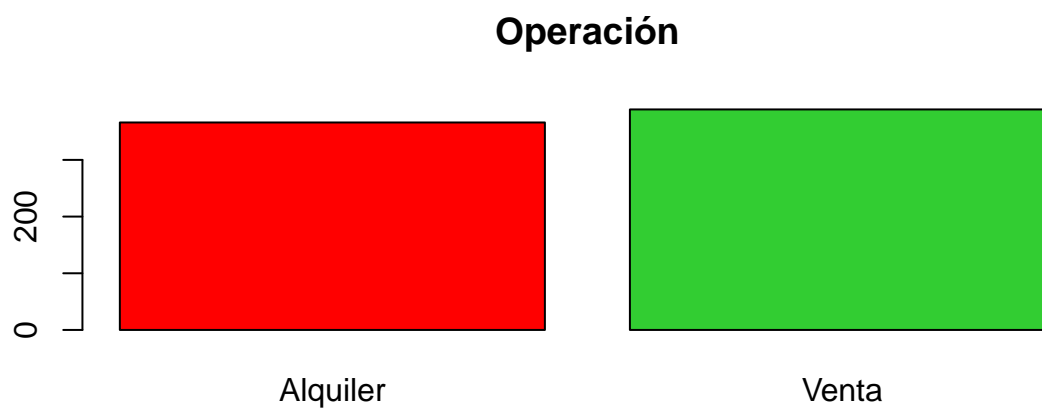
```
##  
## Barcelona      Girona      Lleida Tarragona  
##          204          210          158          183
```

Con la función `barplot(...)` generamos las gráficas de cada tabla mostrada.

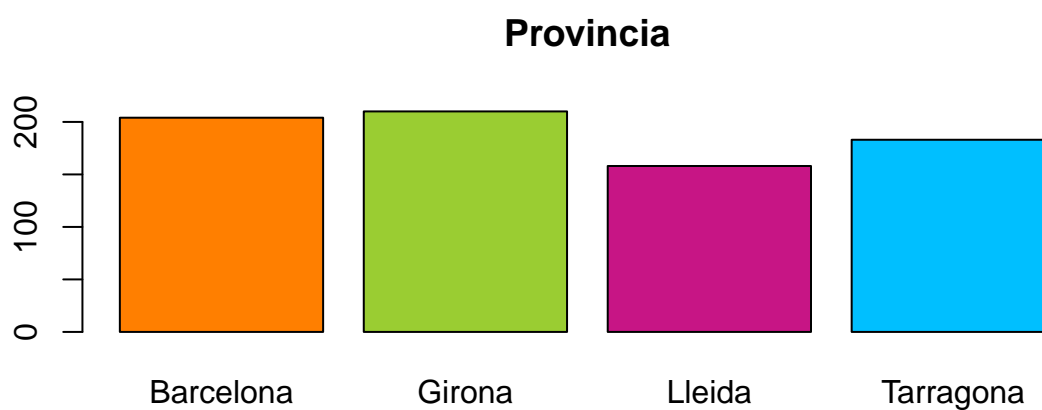
```
barplot(  
  table(data$Tipo_de_Inmueble),  
  main="Tipo_de_Inmueble",  
  col=c("yellow", "magenta"))
```



```
barplot(  
  table(data$Operacion),  
  main="Operación",  
  col=c("red1", "limegreen"))
```



```
barplot(
  table(data$Provincia),
  main="Provincia",
  col=c("darkorange1", "yellowgreen", "mediumvioletred", "deepskyblue1"))
```



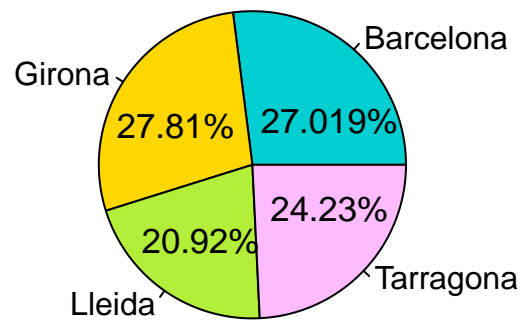
3. Investigue cómo hacer una gráfica de pie o pastel en R para describir la variable Provincia y grafíquela en R

Grafiquemos la variable Provincia en una grafica de pastel.

```
pie(table(data$Provincia), col=c("darkturquoise", "gold1", "olivedrab2", "plum1"))

text(0.40, 0.23, "27.019%", cex=1.1)
text(-0.40, 0.20, "27.81%", cex=1.1)
```

```
text(0.40,-0.23,"24.23%",cex=1.1)
text(-0.27,-0.40,"20.92%",cex=1.1)
```



4. Para las variables numéricas Superficie y Precio de Venta calcule las medidas de tendencia central Media, Mediana y los percentiles 10%, 25%, 50%, 75%, 90%.

Calculamos la mediana, la media y los percentiles para Superficie.

```
median(data$Superficie)
```

```
## [1] 139.3
```

```
mean(data$Superficie)
```

```
## [1] 134.03
```

```
quantile(data$Superficie,0.10)
```

```
## 10%
## 84.932
```

```
quantile(data$Superficie,0.25)
```

```
## 25%
## 98.42
```

```
quantile(data$Superficie,0.50)
```

```
##      50%  
## 139.3
```

```
quantile(data$Superficie,0.75)
```

```
##      75%  
## 169.245
```

```
quantile(data$Superficie,0.90)
```

```
##      90%  
## 181.494
```

Ahora, calculamos las Medidas de Tendencia Central para Precio de Venta.

En este caso, notamos que los datos que tenemos para esta tabla son de tipo *character*, lo que nos imposibilita el cálculo. Por lo que debemos convertir a tipo *numeric*.

Primero, substraemos las comas de los valores y remplazamos los valores en la tabla:

```
data$Precio_Venta <- sub(",", "",  
                          sub(",", "",  
                              data$Precio_Venta,  
                                fixed=TRUE),  
                          fixed=TRUE)
```

Luego, convertimos:

```
data$Precio_Venta <- as.numeric(data$Precio_Venta)
```

Finalmente, calculamos las medidas (nota: usamos la función *options(...)* para poder imprimir los números decimales de nuestros resultados. Extrañamente, para este caso, los decimales no aparecían):

```
options(digits=12)  
mean(data$Precio_Venta)
```

```
## [1] 1328398.27837
```

```
median(data$Precio_Venta)
```

```
## [1] 1291349.49
```

```
quantile(data$Precio_Venta, 0.10)
```

```
##      10%  
## 802654.108
```

```
quantile(data$Precio_Venta, 0.25)
```

```
##          25%  
## 968112.82
```

```
quantile(data$Precio_Venta, 0.50)
```

```
##          50%  
## 1291349.49
```

```
quantile(data$Precio_Venta, 0.75)
```

```
##          75%  
## 1651658.355
```

```
quantile(data$Precio_Venta, 0.90)
```

```
##          90%  
## 1914446.996
```

5. Para las variables numéricas Superficie y Precio de Venta calcule las medidas de dispersión Varianza muestral, Desviación estándar, RIC (Rango intercuartílico).

Calculamos las medidas de Varianza, Desviación Estándar y el Rango Intercuartílico de la variable Superficie:

```
var(data$Superficie)
```

```
## [1] 1504.04459854
```

```
sd(data$Superficie)
```

```
## [1] 38.7820138536
```

```
quantile(data$Superficie, 0.75) - quantile(data$Superficie, 0.25)
```

```
##          75%  
## 70.825
```

Ahora, calculamos las medidas para la variable Precio de Venta:

```
var(data$Precio_Venta)
```

```
## [1] 187985327553
```

```
sd(data$Precio_Venta)
```

```
## [1] 433572.747705
```

```
quantile(data$Precio_Venta, 0.75) - quantile(data$Precio_Venta, 0.25)
```

```
##          75%  
## 683545.535
```

6. Para las variables numéricas Superficie y Precio de Venta investigue cómo calcular el coeficiente de asimetría y mencione si los datos son simétricos o asimétricos (si son asimétricos son sesgados a la izquierda o a la derecha).

Como vimos en clase, el Coeficiente de Asimetría (skewness) nos habla sobre la asimetría de los datos. Para calcular dicho valor, hacemos uso de la librería *moments*. Así, primero la instalamos y la cargamos a memoria:

```
install.packages("moments")
```

```
## Installing package into '/home/asus-julius/R/x86_64-pc-linux-gnu-library/3.6'  
## (as 'lib' is unspecified)
```

```
library(moments)
```

Calculando los coeficientes para la variable Superficie y para la variable Precio de Venta, respectivamente, obtenemos:

```
skewness(data$Superficie)
```

```
## [1] -0.0269056094677
```

```
skewness(data$Precio_Venta)
```

```
## [1] 0.330297378809
```

Notamos que el coeficiente de la columna Superficie es **negativo** siendo así que dicha variable está **sesgada a la derecha**. Mientras que el coeficiente de la columna Precio de Venta es **positivo**, diciendo que esta variable está **sesgada a la izquierda**.

Naturalmente podemos afirmar que los datos son **asimétricos**.

7. Para las variables numéricas Superficie y Precio de Venta investigue cómo calcular el coeficiente de curtosis y mencione si los datos son mesocráticos, platocúrticos o leptocúrticos.

Usando la misma librería que en el ejercicio anterior, calculamos el coeficiente de curtosis (kurtosis) de ambas variables, obteniendo:

```
kurtosis(data$Superficie)
```

```
## [1] 1.54706260544
```

```
kurtosis(data$Precio_Venta)
```

```
## [1] 2.30170763428
```

Con la información anterior, podemos decir que los datos de la variable de Superficie son **platicurtica** (porque el **coeficiente es negativo**), mientras que los datos correspondientes a la columna de Precio de Venta son **leptocurtica** (porque el **coeficiente es positivo**).