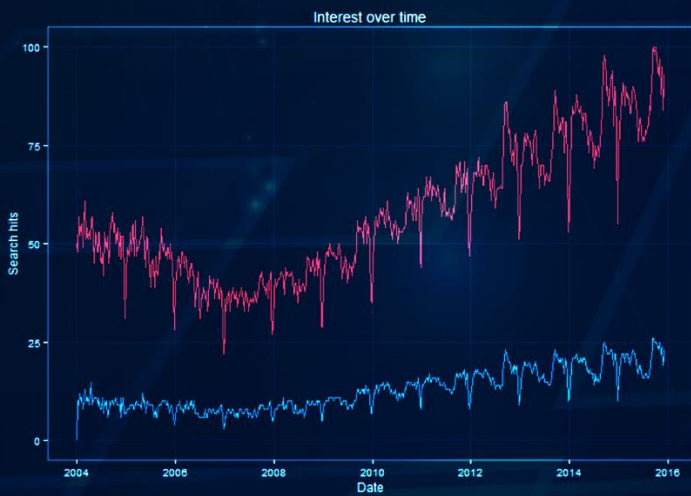


# Econometría Aplicada

con 



```
R Console (32-bit)
Archivo Editar Misc. Ejecutar Ventanas Ayuda

> x <- c(1,2,3,4,5,6)
> y <- x^2
> print(y)
[1] 1 4 9 16 25 36
> mean(y)
[1] 15.16667
> var(y)
[1] 178.9444
> lm_1 <- lm(y ~ x)
> print(lm_1)

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept) -9.3333
x 7.0000

> summary(lm_1)

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept) -9.3333
x 7.0000

Residuals:
1 2 3 4 5 6
3.3333 -0.6667 -2.6667 -2.6667 -0.6667 3.3333

Coefficients:
(Intercept) Estimate Std. Error t value Pr(>|t|)
1 -9.3333 2.8441 -3.282 0.030453 *
x 7.0000 0.7303 9.585 0.000662 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.055 on 4 degrees of freedom
Multiple R-squared: 0.9583, Adjusted R-squared: 0.9478
F-statistic: 91.87 on 1 and 4 DF, p-value: 0.000662

> |
```



## **SESIÓN 5: Análisis gráfico en R**

## Contenido

Introducción .....	4
Gráficos univariados .....	5
Gráfico de sectores .....	5
Crear un gráfico de sectores con datos agregados.....	5
Crear un gráfico de sectores a partir de una variable categórica .....	6
Gráfico de barras .....	6
Gráfico de barras simple.....	7
Gráfico de barras con más categorías.....	7
Diagrama de cajas .....	9
Histograma .....	10
Gráficos multivariados.....	13
Gráficos de dispersión .....	13
Matriz de dispersión.....	14
Bibliografía .....	16
Recursos informáticos .....	16

## Introducción

Para mejorar el entendimiento de los datos estadísticos se usa la presentación de estos por medio de gráficos, ya sean variables numéricas o variables continuas. Este proceso tiene gran importancia ya que dota al lector de una mayor facilidad para el entendimiento de los diversos resultados numéricos a los que se ha llegado

En la siguiente sesión se explicarán los diversos tipos de gráficos existentes, como hacerlo haciendo uso del lenguaje de R como herramienta principal.

## Gráficos univariados

Son aquellos que solo usan a una variable para mostrar sus resultados, entre los principales se encuentra, los histogramas, el gráfico de barras, el diagrama de cajas, entre otros.

### Gráfico de sectores

---

Los gráficos de sectores son aquellos que muestran un círculo dividido en sectores de tamaño proporcional a las frecuencias de los valores para una variable que es categórica.

Este tipo de gráfico es una excelente herramienta cuando se trabaja con datos porcentuales tomados de individuos u objetos, por ejemplo, en un análisis sobre el lugar de procedencia de cierto grupo de personas.

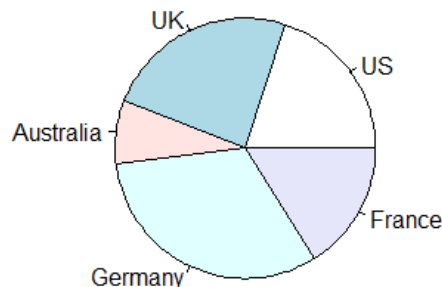
En R el gráfico de sectores se obtiene con la función **pie()**.

#### Crear un gráfico de sectores con datos agregados

El uso de este comando se basa en un vector con cada uno de los porcentajes o los valores de cada categoría:

```
slices = c(10, 12, 4, 16, 8) # Valores  
lbls = c("US", "UK", "Australia", "Germany", "France") # Etiquetas  
pie(slices, labels=lbls)
```

El código anterior crea un vector de valores y para cada uno de estos se crea un etiqueta, luego se usa la orden **pie()**, que indica que el vector de valores es **slices** y las etiquetas se tomarán del objeto **lbls**.



### Crear un gráfico de sectores a partir de una variable categórica

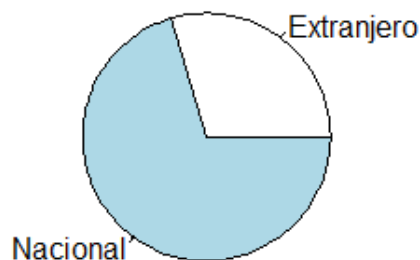
Cuando se tiene una variable categoría con diferentes valores y que no está a nivel agregado, se debe crear en primer lugar una variable que sea la recodificación de la variable categórica.

```
auto$extran [auto$foreign == 1] = "Extranjero"  
auto$extran [auto$foreign == 0] = "Nacional"
```

Luego se especifica con el comando **pie()** la variable, que este caso será por medio de una tabla, usando el comando **table()**.

```
pie(table(auto$extran))
```

La sintaxis indica que se quiere un gráfico de sectores haciendo uso de una tabla para variable recodificada que se ha creado, que tiene por nombre **extran**.



### Gráfico de barras

Son gráficos que muestran la frecuencia para cada categoría de valores en una variable, son una modificación de los gráficos de sectores. Estos gráficos son muy flexibles para ser adaptados en trabajos con variables categóricas.

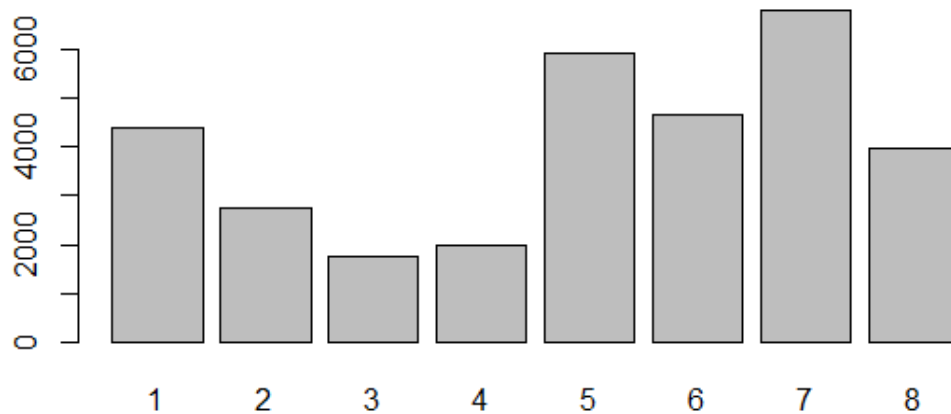
Para representar un gráfico de barras en R se hace uso de la función **barplot()**.

### Gráfico de barras simple

Al momento de crear un gráfico para una sola categoría para mejorar el aspecto se deberá hacer una recodificación que contenga a cada uno de los valores de la variable categórica.

```
barplot(table(sumaria_2015$dominio))
```

Esta orden indica que con la base de datos **sumaria\_2015**, se creará un gráfico de barras usando a la variable **dominio** a modo de tabla.



Para hacer un gráfico horizontal se debe especificar en las opciones **horiz=TRUE**, esto hará que el gráfico se horizontal y no vertical como se muestra por defecto.

### Gráfico de barras con más categorías

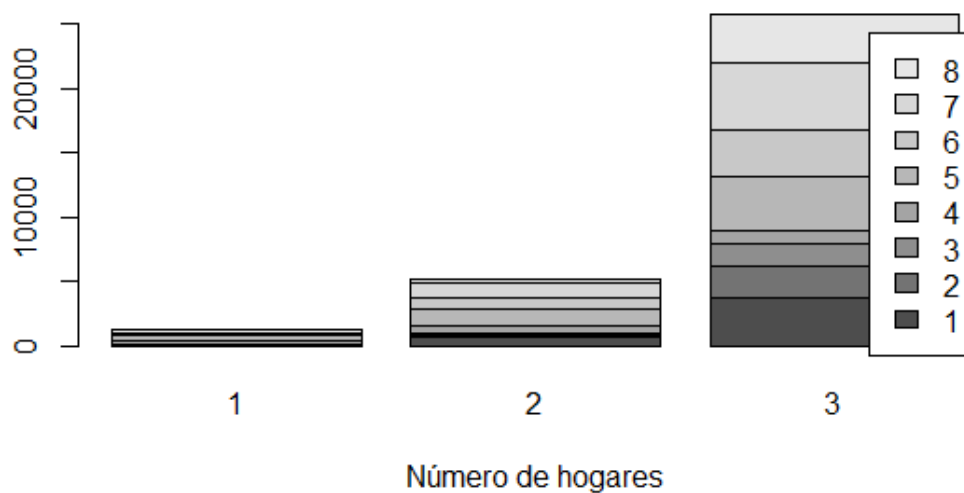
Este tipo de gráfico contendrá las frecuencias, pero separadas por variables. Se deberá hacer en base a una tabla de doble entrada entre 2 variables categóricas.

```
counts = table(sumaria_2015$dominio, sumaria_2015$pobreza)
barplot(counts, main="Condición de pobreza de hogares por dominio",
        xlab="Número de hogares", legend = rownames(counts))
```

Esta sentencia indica que se creará una tabla de doble entrada con variables categóricas, dominio y pobreza.

En el comando del gráfico de barras especifica que se usará dicha tabla creada y también que se usará un título en el gráfico, así como también etiqueta para eje x y una legenda con los nombres de la categoría.

### Condición de pobreza de hogares por dominio

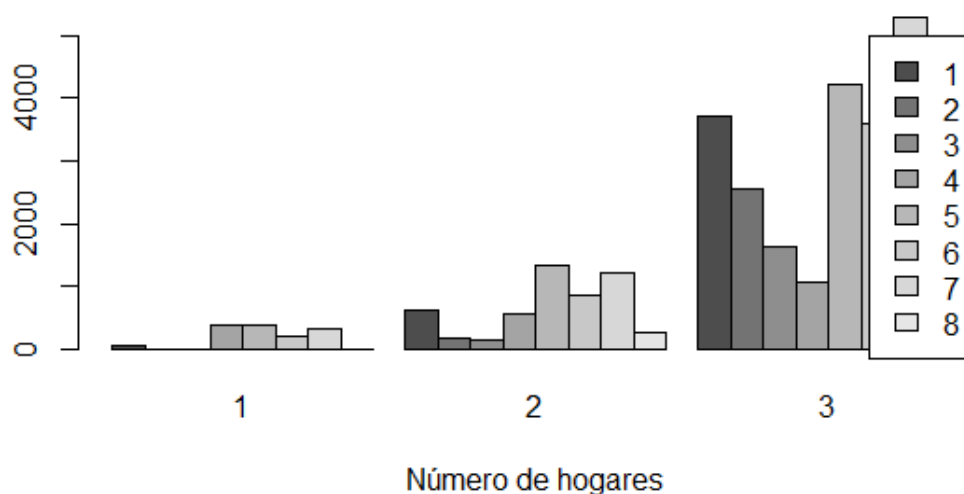


Para hacer gráficos agrupados, es decir con las barras separadas se debe hacer uso de la opción **beside=TRUE**.

```

counts = table(sumaria_2015$dominio, sumaria_2015$pobreza)
barplot(counts, main="Condición de pobreza de hogares por dominio",
        xlab="Número de hogares", legend = rownames(counts),
        beside=TRUE)
    
```

### Condición de pobreza de hogares por dominio



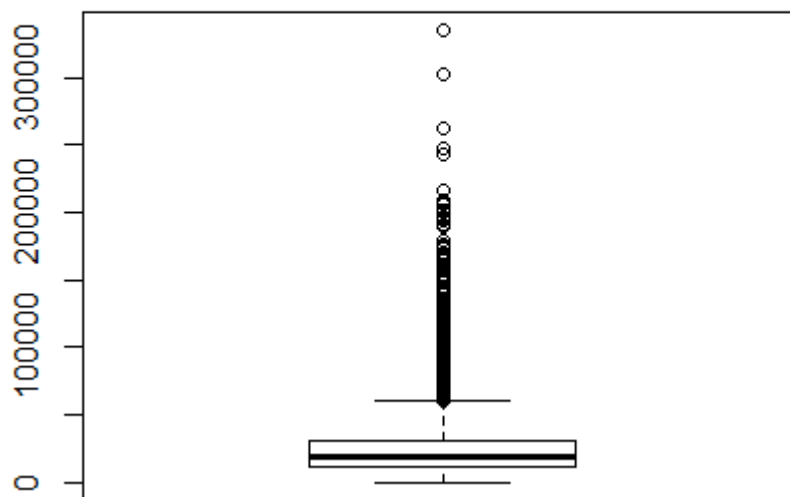


## Diagrama de cajas

El diagrama de cajas es una representación visual que describe características importantes para una variable numérica. En su representación se describen los valores de medidas de posición como los 3 cuartiles, el mínimo y el máximo, sobre un rectángulo alineado generalmente de forma vertical.

Para crear diagramas de caja en R se debe hacer uso de función **boxplot()**.

```
boxplot(sumaria_2015$gashog2d)
```

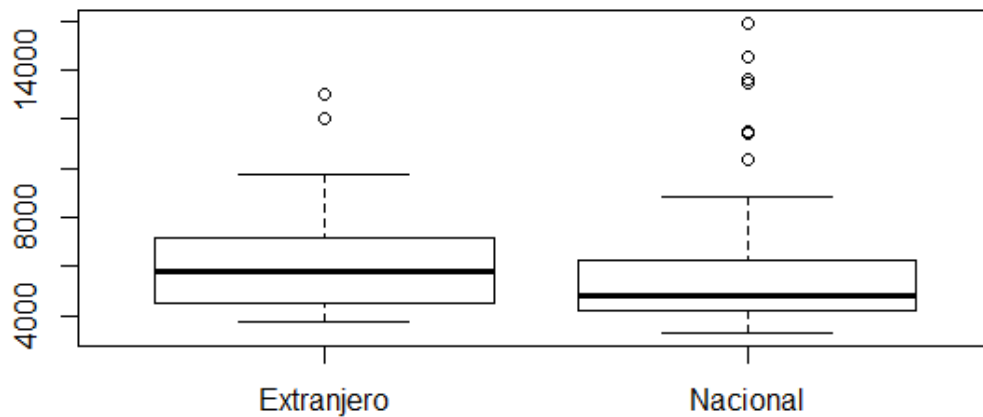


En los gráficos de cajas se pueden añadir categorías, para eso se deberá indicar en el código que variable será la categórica.

```
boxplot(auto$price ~ auto$extran, main="Diagrama de cajas para los  
precios\nde acuerdo a la procedencia del vehículo")
```

Esto indicará que se creará un gráfico de cajas para la variable **price**, pero que este será ordenado de acuerdo a la procedencia del vehículo, representado por la variable **extran**.

**Diagrama de cajas para los precios  
de acuerdo a la procedencia del vehículo**



## Histograma

El histograma para un conjunto de datos es un gráfico de barras que representa la frecuencia para cada intervalo de valores en un variable numérica.

Para realizar un histograma se debe dividir a la variable en rangos o intervalos que sean de preferencia de igual longitud. Para determinar el número de barras en un histograma se suele usar la regla de Sturges, que viene dada por la siguiente expresión:

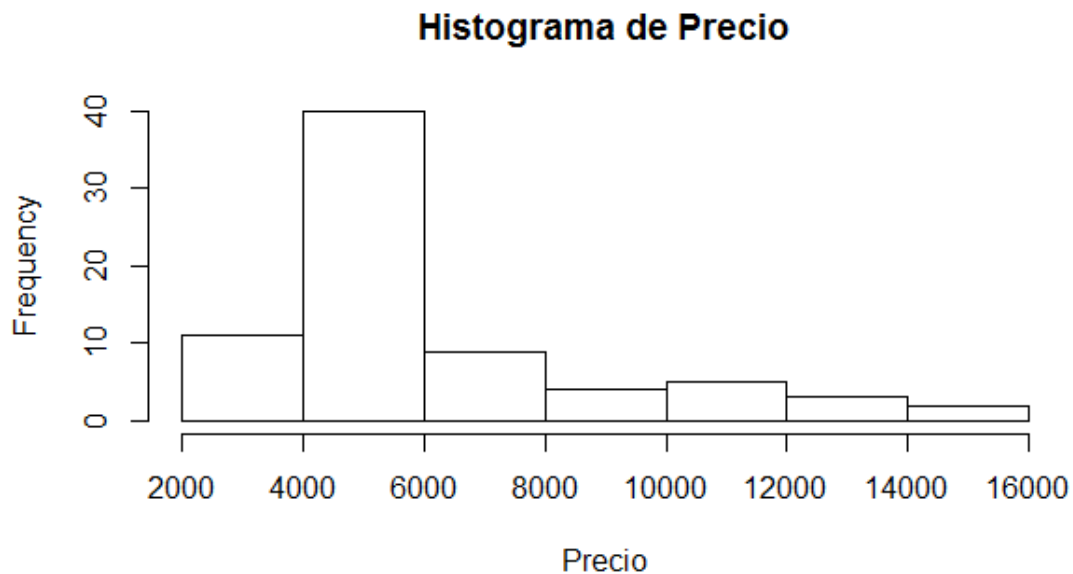
$$k = 1 + \log_2 n$$

Donde  $k$  es el número de barras y  $n$  es el tamaño de la muestra.

Para realizar histogramas en R se usa la función **hist()**, donde se debe especificar la variable a tratarse, y donde también se pueden especificar etiquetas de ejes, número de cortes, entre otros.

```
hist(auto$price , xlab="Precio", main="Histograma de Precio")
```

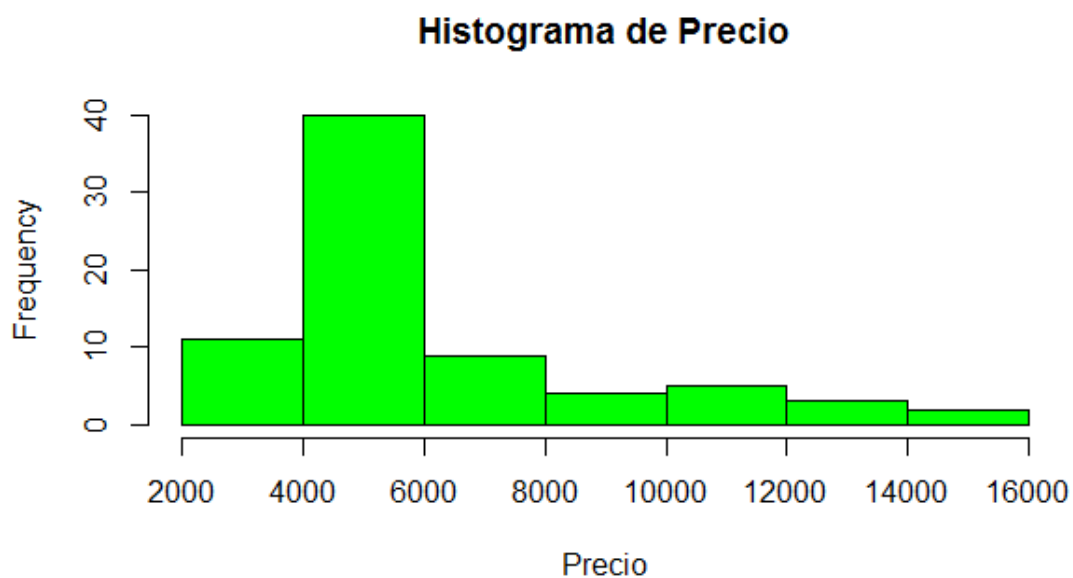
La sentencia anterior indica que se quiere crear un histograma con la **price** en la base de datos **auto**. Y se verá de la siguiente manera:



Para colorear el gráfico se pueden especificar la opción `col`, indicando el color en inglés:

```
hist(auto$price , col = "green", xlab="Precio", main="Histograma de Precio")
```

Se mostrará el siguiente gráfico:



Para añadir la curva de distribución normal, se debe crear objetos que contengan a la media y a la desviación estándar de la variable, donde

el histograma en lugar de mostrar las frecuencias deberá mostrar las probabilidades:

```
m = mean(auto$price)
std = sqrt (var(auto$price))
```

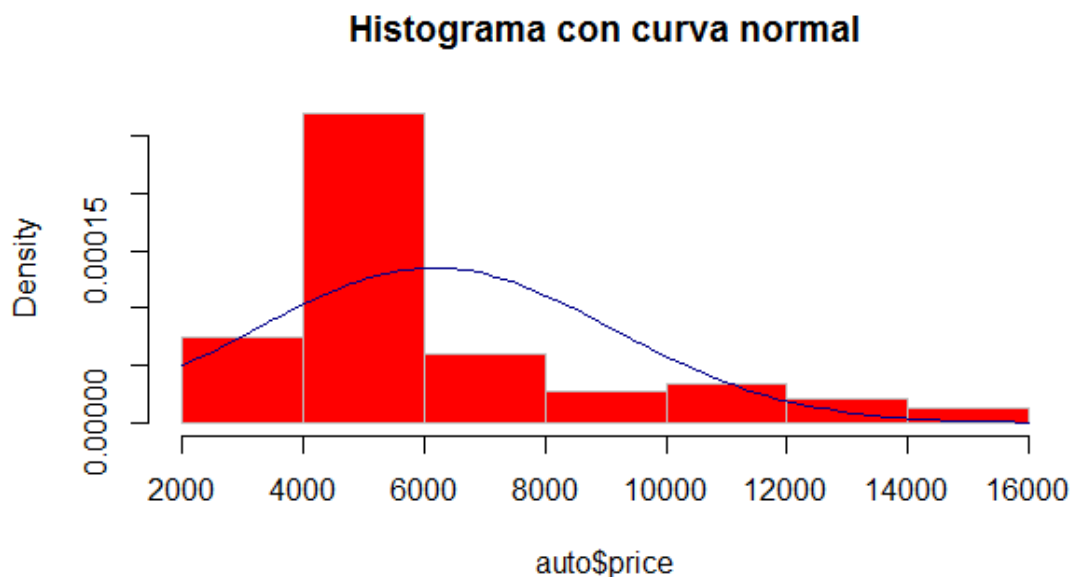
Una vez creados los objetos se hace el histograma haciendo uso de la función **prob**, para mostrar las probabilidades en lugar de las frecuencias:

```
hist(auto$price, probability = TRUE, col = "red",
      main = "Histograma con curva normal", border = "grey")
```

Luego de esto se debe crear una curva con distribución normal especificando la media y la desviación estándar que se generó con la variable de análisis, la función a usarse será **curve()** y **dnorm()** para la distribución normal.

```
curve(dnorm(x, mean = m, sd=std) , col = "darkblue",
      add = TRUE)
```

Una vez que se corra toda la sintaxis se mostrará el siguiente gráfico, en el que se verá la distribución de frecuencias con la curva normal superpuesta en el histograma.



## Gráficos multivariables

En los gráficos multivariables se va a representar la relación existente entre dos o más variables.

### Gráficos de dispersión

Este tipo de gráficos se usa para trazar puntos de datos para dos variables, en un eje horizontal y en otro vertical, en donde se trata de demostrar la relación existente entre ambas variables.

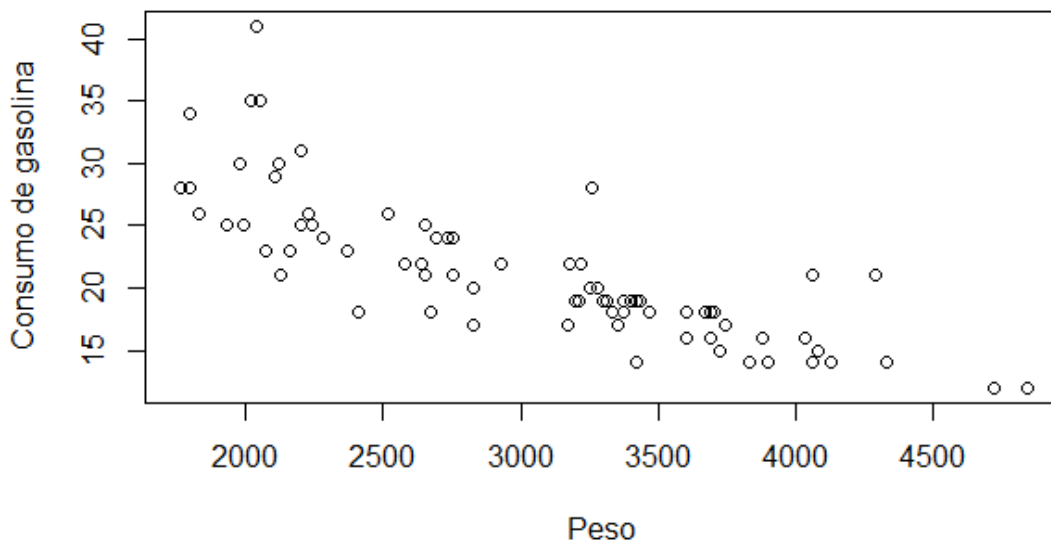
Para hacer gráficos de dispersión en R se hace uso de la función **plot()**.

```
plot(auto$weight, auto$mpg, xlab = "Peso",  
      ylab = "Consumo de gasolina",  
      main = "Dispersión entre el peso y el consumo de combustible")
```

La sentencia anterior indica que se quiere hacer un gráfico de dispersión con las variables **weight** (eje X) y **mpg** (eje Y), ambas de la base de datos **auto**, donde las etiquetas de los ejes estarán especificadas, así como también el título del gráfico.

El gráfico resultante se verá de la siguiente manera:

**Dispersión entre el peso y el consumo de combustible**

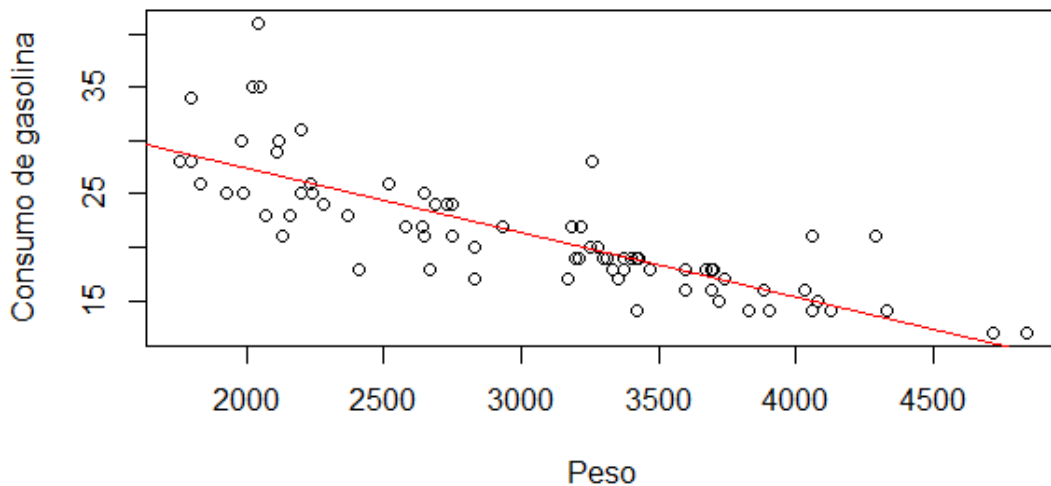


Para añadir la línea de regresión de las dos variables en dicho gráfico de dispersión se debe hacer uso de la función **abline()**.

```
abline(lm(mpg ~ weight) , col = "red")
```

Esta instrucción indicará que se generará una línea de regresión donde la variable dependiente es **mpg** y la independiente es **weight**. El gráfico se verá de la siguiente manera:

### Dispersión entre el peso y el consumo de combustible



### Matriz de dispersión

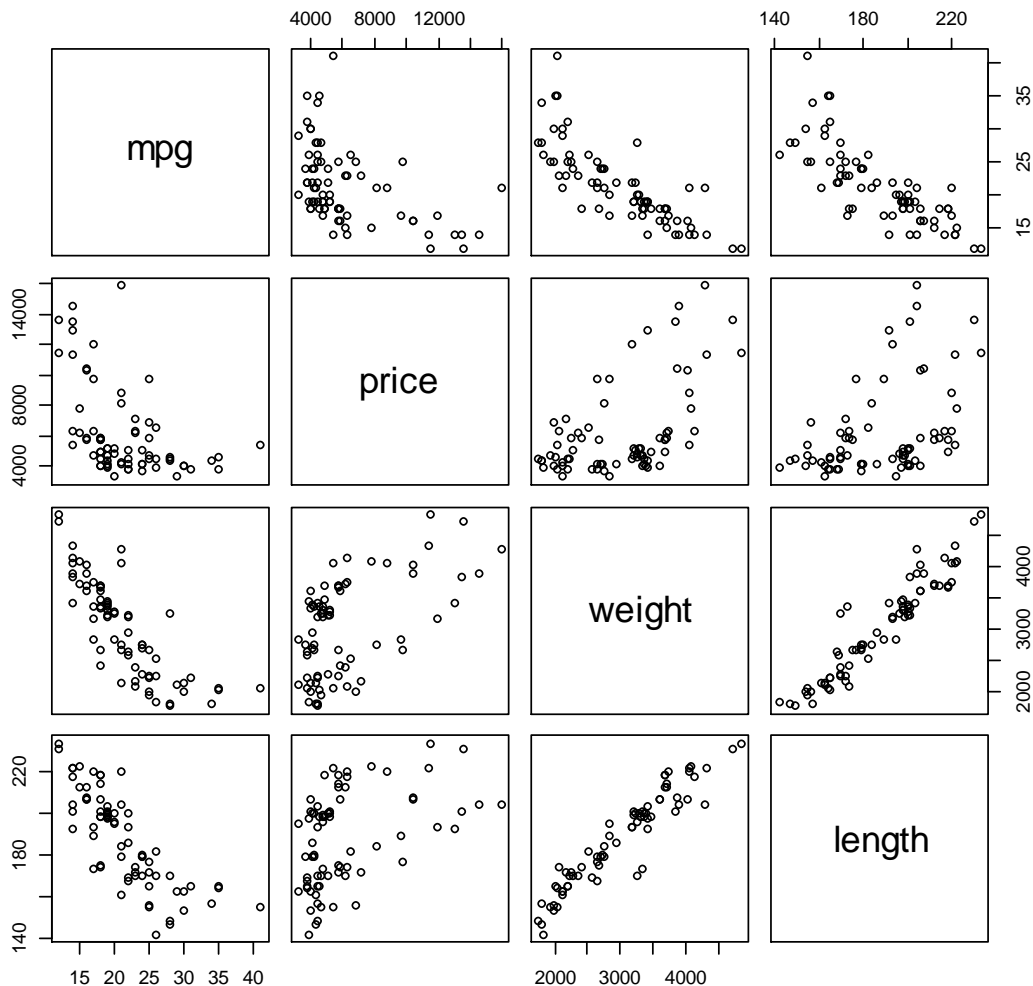
Las matrices de dispersión muestran dispersiones entre más de dos variables. En R la función para realizar matrices de dispersión es **pairs()**.

```
pairs(~mpg + price + weight + length, data = auto,  
      main = "Matriz de dispersiones simple")
```

La sintaxis anterior le indicará al programa que se quiere generar una matriz de dispersiones para las variables en cuestión (**mpg**, **price**, **weight** y **length**).

El gráfico en R se mostrará de la siguiente manera:

### Matriz de dispersiones simple



Aparte del comando **pairs()**, existen otros paquetes que ofrecen funciones para realizar matrices de dispersiones, tal es el caso del paquete **lattice** y la función **sploim()**, o del paquete **car** y la función **scatterplot.matrix()**.

## Bibliografía

Correa, Carlos & González Nelfi. (2002). *Gráficos estadísticos con R*. Universidad Nacional - Sede Medellín. Consultado el 25 de noviembre de 2016.

<ftp://cran.r-project.org/pub/R/doc/contrib/grafi3.pdf>

## Recursos informáticos

---

Quick R - Graphs:

<http://www.statmethods.net/graphs/>

R Bloggers - How to make a Histogram with basic R:

<https://www.r-bloggers.com/how-to-make-a-histogram-with-basic-r/>