

[0010] DEFAD: Representación y tabulación de datos

Motivación e introducción a gráficos

Elvira Ferre Jaén
elvira@um.es

Universidad de Murcia

Marzo 2018



Motivación

Consideremos algunos datos (cuatro pares de variables)

```
head( anscombe )
```

```
##      x1 x2 x3 x4   y1   y2   y3   y4
##  1 10 10 10  8 8.04 9.14  7.46 6.58
##  2  8  8  8  8 6.95 8.14  6.77 5.76
##  3 13 13 13  8 7.58 8.74 12.74 7.71
##  4  9  9  9  8 8.81 8.77  7.11 8.84
##  5 11 11 11  8 8.33 9.26  7.81 8.47
##  6 14 14 14  8 9.96 8.10  8.84 7.04
```



Motivación

¿Qué te gustaría calcular para cada variable?



Motivación

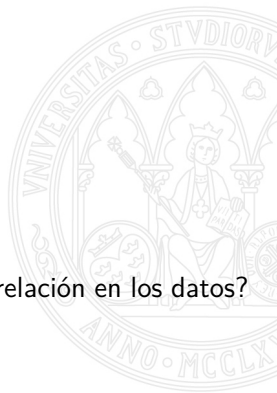
```
apply( anscombe, 2, summary )
```

```
##           x1    x2    x3 x4           y1           y2           y3           y4
## Min.      4.0   4.0   4.0  8   4.260000  3.100000   5.39   5.250000
## 1st Qu.   6.5   6.5   6.5  8   6.315000  6.695000   6.25   6.170000
## Median    9.0   9.0   9.0  8   7.580000  8.140000   7.11   7.040000
## Mean      9.0   9.0   9.0  9   7.500909  7.500909   7.50   7.500909
## 3rd Qu.  11.5  11.5  11.5  8   8.570000  8.950000   7.98   8.190000
## Max.     14.0  14.0  14.0 19  10.840000  9.260000  12.74  12.500000
```

Motivación

¿ Qué nos dicen los resultados anteriores ?

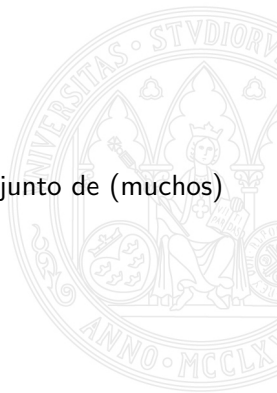
¿De verdad eres capaz de ver algún patrón, asociación, relación en los datos?



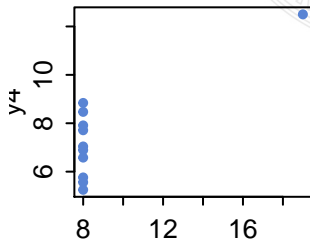
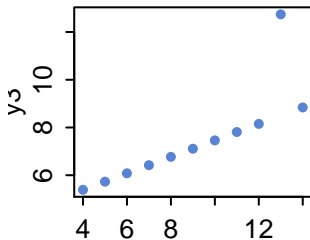
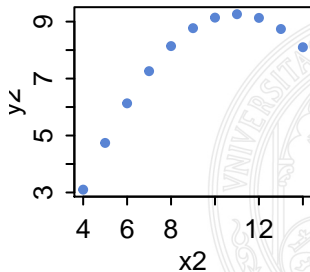
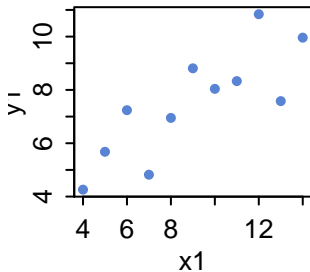
¿Por qué representar gráficamente?

Nuestros ojos no son buenos dándole sentido a un conjunto de (muchos) números.

Por ello necesitamos representarlos gráficamente.

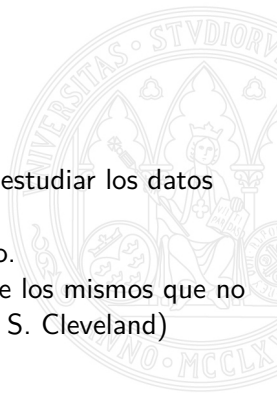


¿Por qué representar gráficamente?

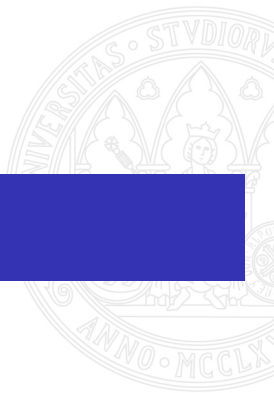


Visualización de los datos

- La visualización gráfica es una forma muy útil de estudiar los datos antes de analizarlos.
- Utilizar solo métodos numéricos es muy restrictivo.
- Visualizar los datos proporciona una percepción de los mismos que no se puede alcanzar desde ningún otro enfoque (W. S. Cleveland)



Gráficos estadísticos

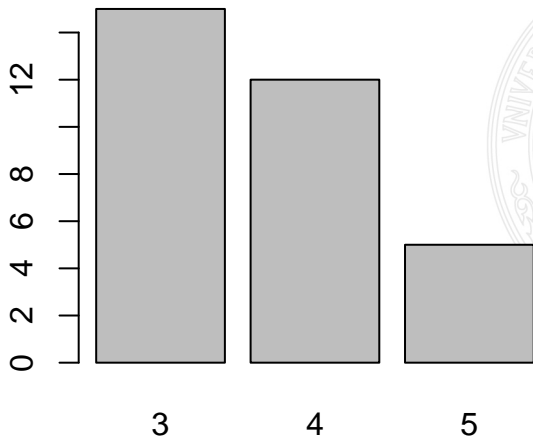


Gráficos estadísticos

- Gráficos para explorar
- Gráficos para comunicar

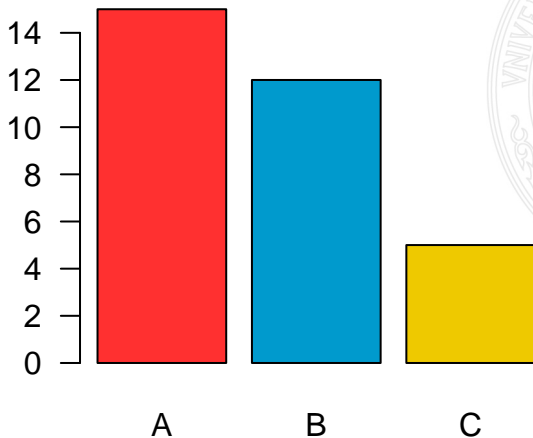


Gráficos para explorar

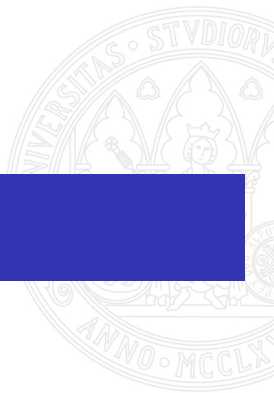


Gráficos para comunicar

Gráfico de barras



Consideraciones



Número y tipo de variables



¿Cuántas variables tenemos?

Variables en un conjunto de datos:

- 1 - datos univariantes
- 2 - datos bivariantes
- datos multivariantes



¿De qué tipo son las variables?

- Cuantitativa -vs- Cualitativa
- Continua -vs- Discreta

Dependiendo del tipo y el número de variables realizaremos un análisis u otro.



Datos univariantes

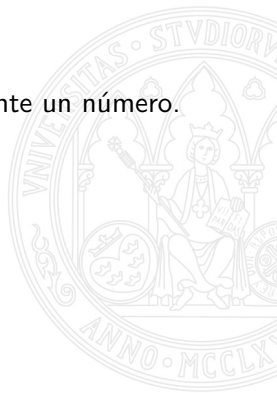


Datos univariantes

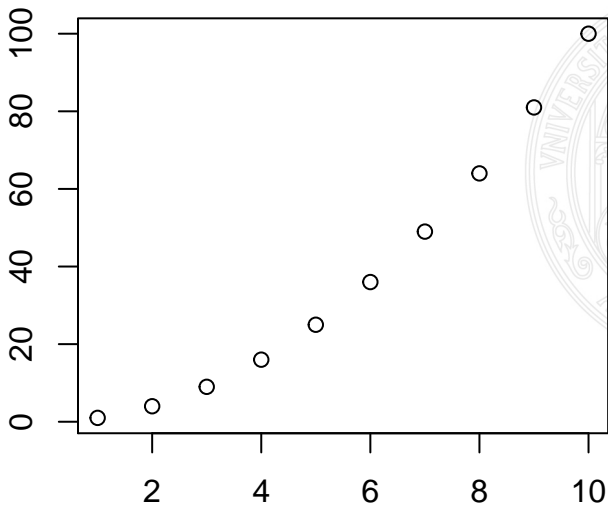
Variables **cuantitativas**: aquella que se expresa mediante un número.

Medimos:

- Cómo están distribuidos los valores
- Máximos, mínimos, rangos
- Medidas de centralización
- Medidas de dispersión
- Áreas de concentración
- Valores atípicos
- Patrones interesantes



Variables cuantitativas



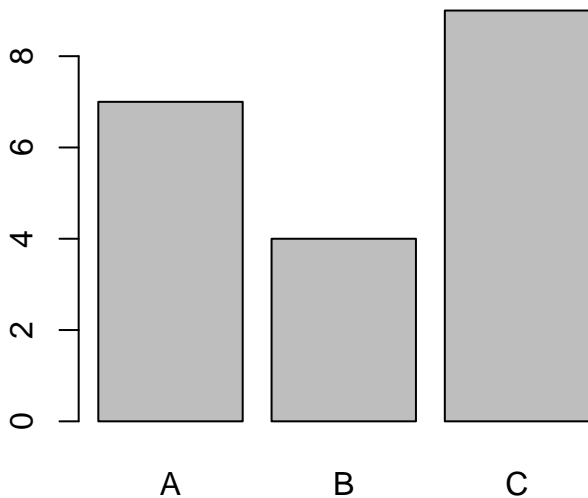
Datos univariantes

Variables **cualitativas**: expresan cualidades, atributos, categorías o características. Medimos:

- Totales y proporciones
- Valores comunes
- Valores más usuales
- Frecuencias de distribución



Variables cualitativas



Datos bivariantes



Datos bivariantes

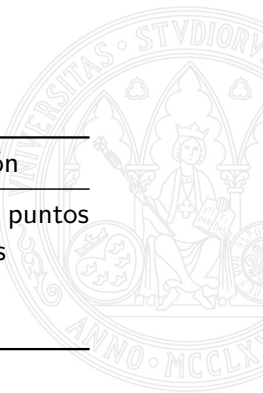
- Cuantitativos -vs- Cuantitativos
- Cualitativos -vs- Cuantitativos
- Cualitativos -vs- Cualitativos

Utilizamos la función `plot()` para ver los diferentes tipos de gráficos.



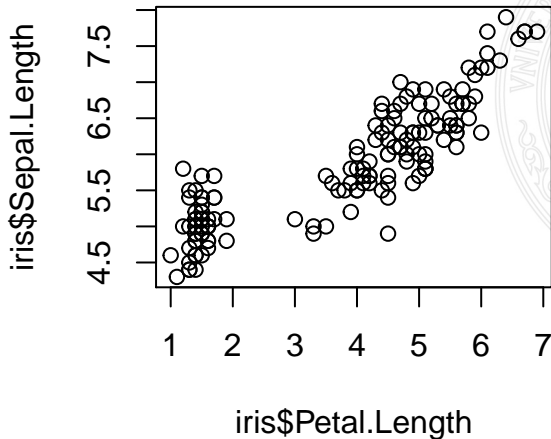
Función `plot()`

Función	Tipo de dato	Descripción
<code>plot()</code>	numérico, numérico	gráfico de puntos
<code>plot()</code>	numérico, factor	stripcharts
<code>plot()</code>	factor, numérico	boxplots
<code>plot()</code>	factor, factor	spineplot



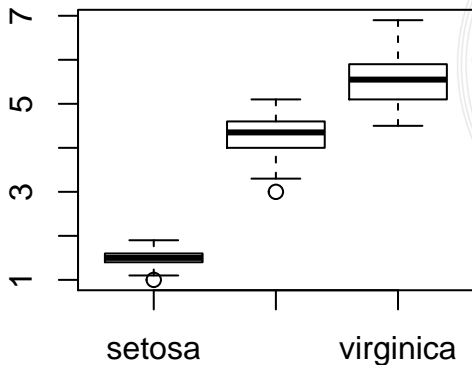
Variables numéricas

```
plot( iris$Petal.Length, iris$Sepal.Length )
```



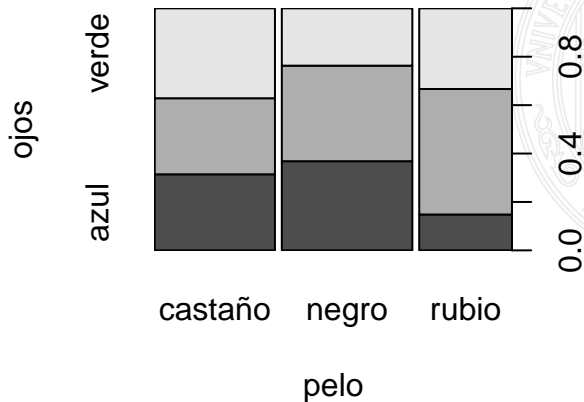
Variable factor y numérica

```
plot( iris$Species, iris$Petal.Length )
```



Variable factor y factor

```
plot( df$pelelo, df$ojos, xlab = "pelelo", ylab = "ojos" )
```



Datos multivariantes



Datos multivariantes

- Cuantitativos
- Cualitativos
- Mixtos

En general estamos interesados en las asociaciones (correlaciones, relaciones)



Datos multivariantes

Gráfico de dispersión 3D

