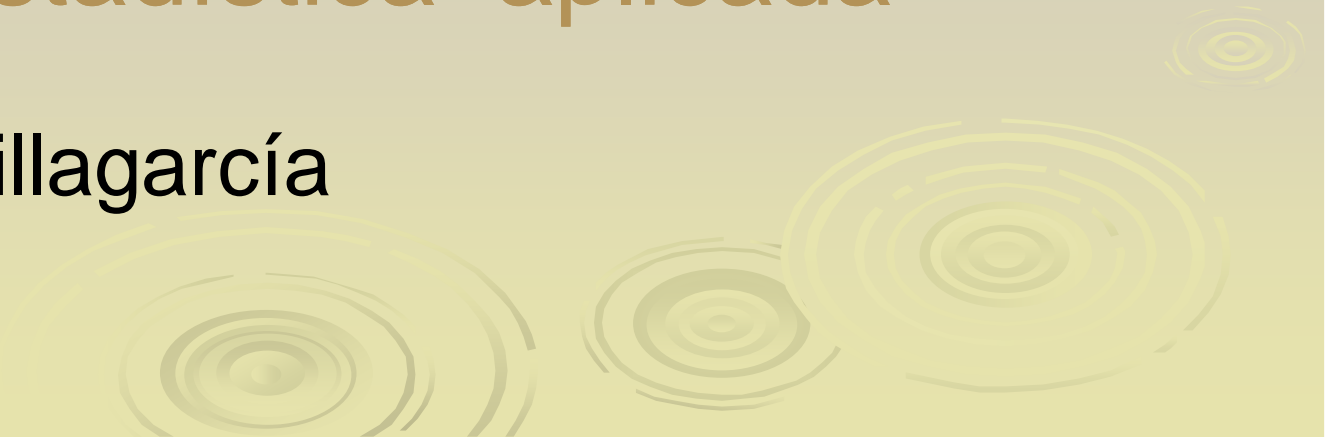


# Máster Sector Farmacéutico

## Estadística aplicada

Teresa Villagarcía



¿Para qué sirve la estadística?

**PARA SUFRIR MUCHO**



# Estudiaremos

- Formas eficaces de obtener información de los datos: medir.
- Muchos análisis gráficos
- Usaremos ordenadores: Statgraphics
- Software disponible en el mercado:  
**Statgraphics**, S, R, SPSS, SAS....

## Índice de la primera parte

- Tipos de datos
- Análisis gráficos
- Medidas resumen numéricas
- Análisis de varias variables cuantitativas.  
Correlación.




## Tipos de datos:

- Datos cualitativos
- Datos cuantitativos
  - Datos transversales
  - Datos temporales



## Datos Cualitativos: (No son números)

- Sexo de una persona
  - Nacionalidad
  - Estado civil
  - Hoteles: céntrico o extrarradio
  - ¿Datos de todo tipo?.....
- 


## Datos Cuantitativos

- Altura de una persona
- Peso
- Ingresos
- Hotel: Número de habitaciones.
- ¿Datos de todo tipo?.....

## Datos transversales:

Se obtienen de muchos individuos en el mismo instante de tiempo o **en tiempos equivalentes.**

**Típico de encuestas**

- Número de trabajadores en 45 hoteles
  - Altura de 200 personas
  - Ingresos de 3000 familias
  - ¿Datos vuestros?.....
- 



# Datos temporales:

Evolución de **una** variable en el tiempo

- Evolución de la inflación en España desde 1980
- Evolución de las ventas de la empresa desde hace 5 años
- Evolución de la calidad percibida por nuestros clientes.
- ¿Vuestros?



# Vamos a estudiar cómo se representan gráficamente

## ➤ Datos cuantitativos

- Transversales: Histograma, Box plot.
- Temporales: Gráfico de la serie

## ➤ Datos cualitativos

- Transversales: Tarta, Barras
- Temporales: Gráfico de la serie

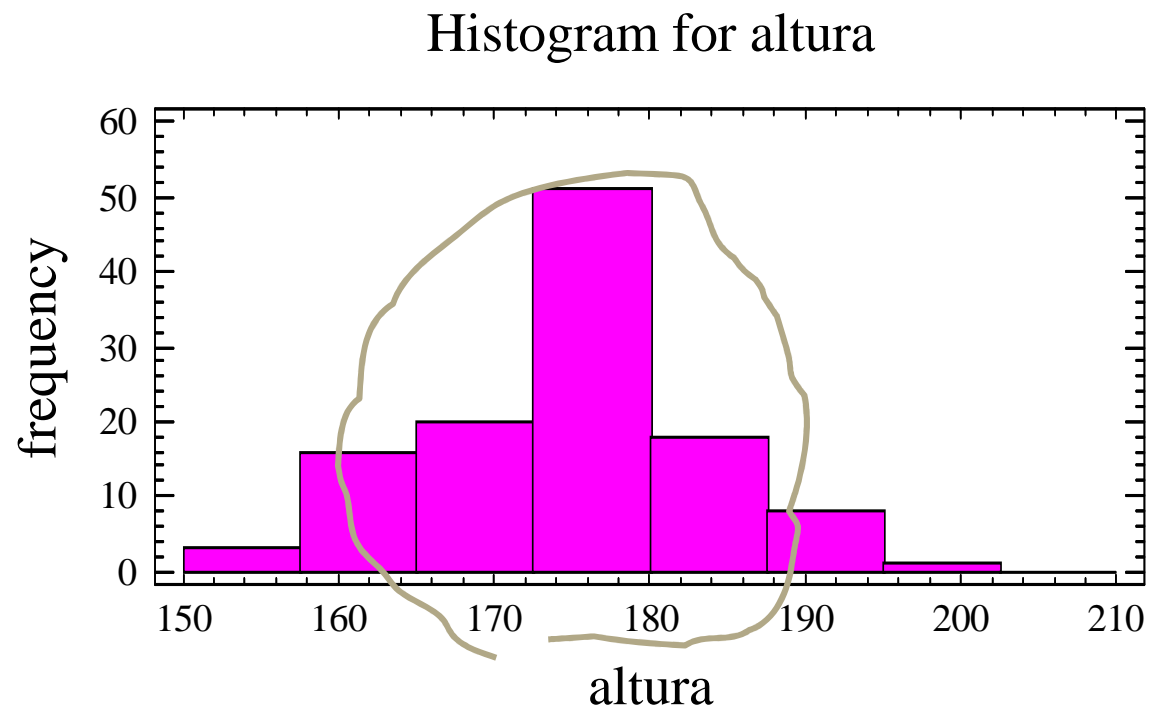


## Datos cuantitativos transversales: Alturas de 117 alumnos míos:

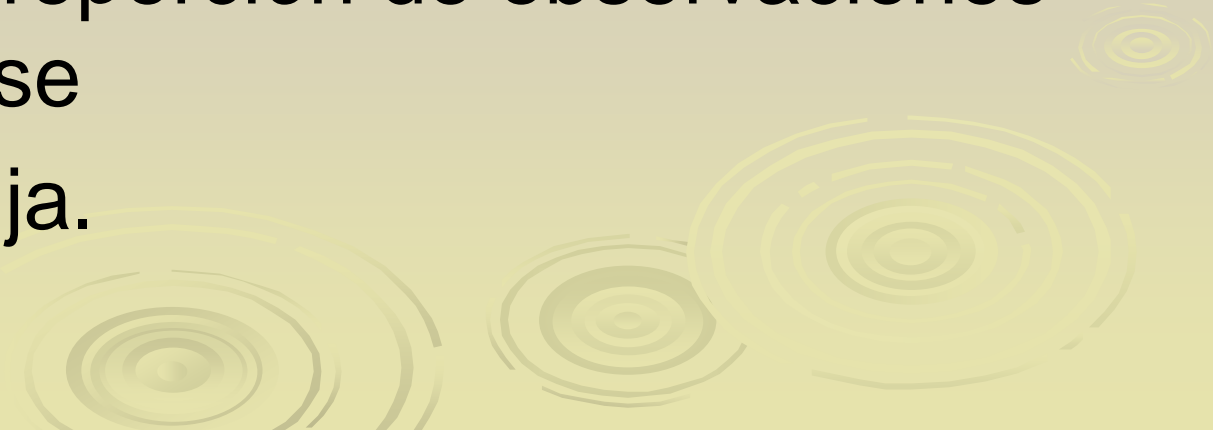
180 178 192 180 162 183 168 160 182 172 162 175  
163 182 179 174 182 178 159 157 155 171 161 175  
189 180 182 178 181 157 170 172 178 181 183 179  
171 151 163 170 174 190 178 163 170 180 189  
181 151 167 167 173 172 175 175 165 180 173 165  
163 169 162 169 178 163 184 172 169 176 164 178  
187 181 199 190 169 179 184 187 175 176 179 161  
178 178 169 179 175 177 169 175 178 177 184 180  
175 175 184 156 173 192 186 180 169 171 172 180  
193 182 185 177 170 173 192 166 173

# Hacemos un histograma

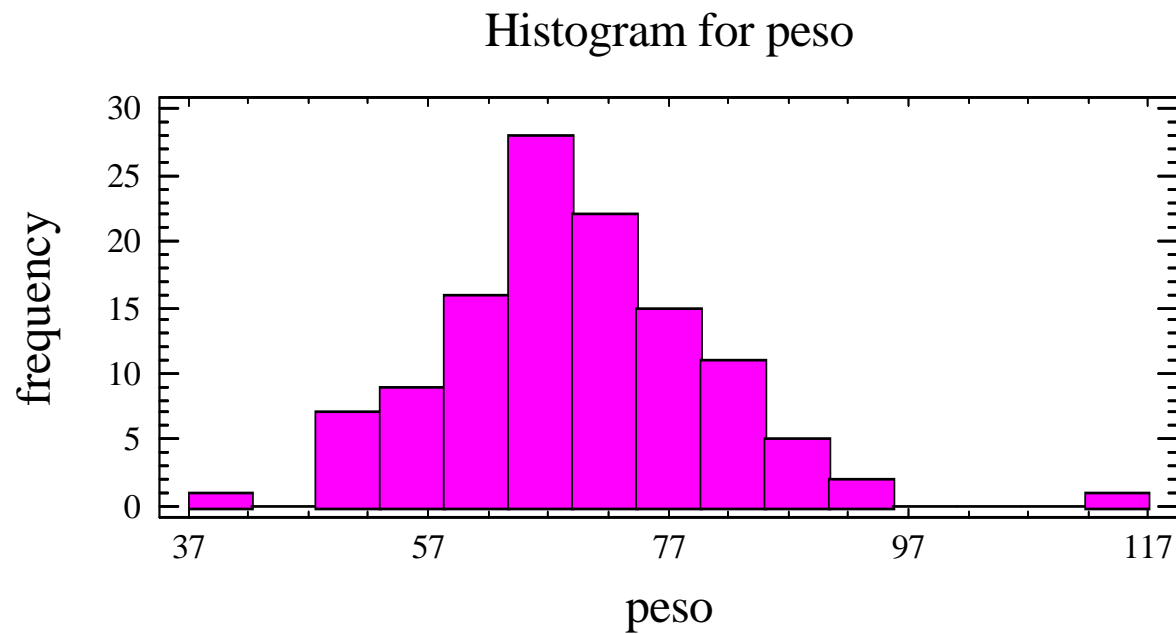
(Bueeno lo hace el ordenador que para eso está)



## ¿Cómo ha hecho el ordenador el histograma?

1. Calcula el valor mínimo y máximo observado
  2. Divide ese rango en una serie de clases
  3. Cuenta cuantas personas hay en cada una de esas clases
  4. Calcula la proporción de observaciones en cada clase
  5. Y.....lo dibuja.
- 

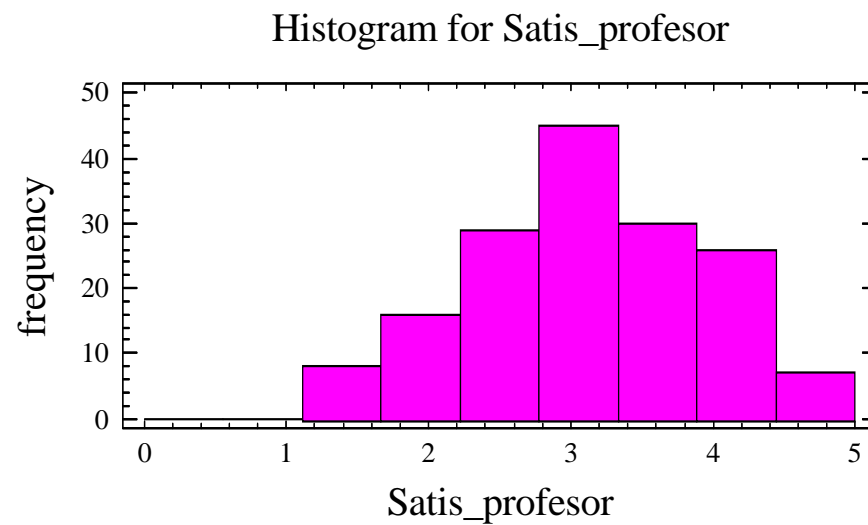
# Aprender a ver histogramas:



# Aprender a ver histogramas:

Satisfacción con los profesores (de 1 a 5).

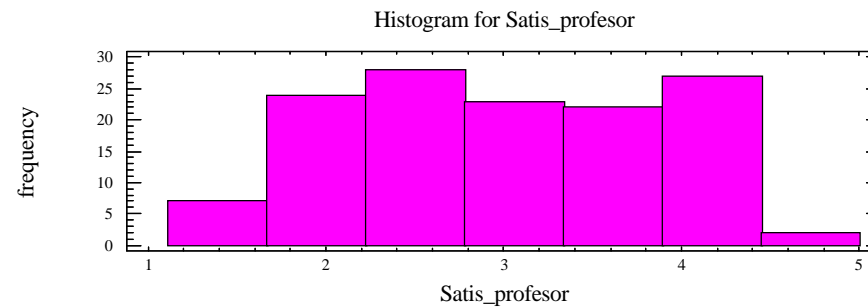
Encuesta en Ingeniería técnica mecánica



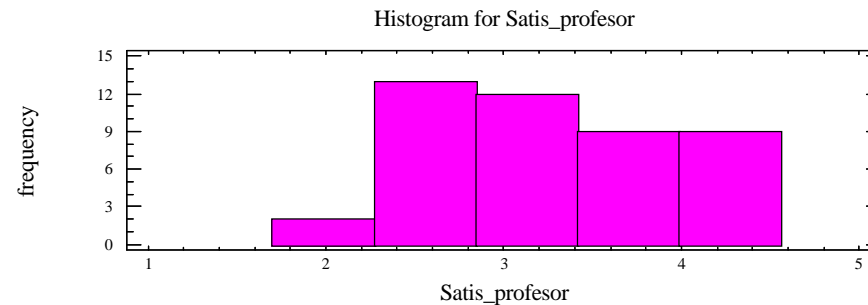
# Aprender a ver histogramas:

Satisfacción con los profesores.  
Encuesta en Ingeniería técnica mecánica

**Primer  
Curso**



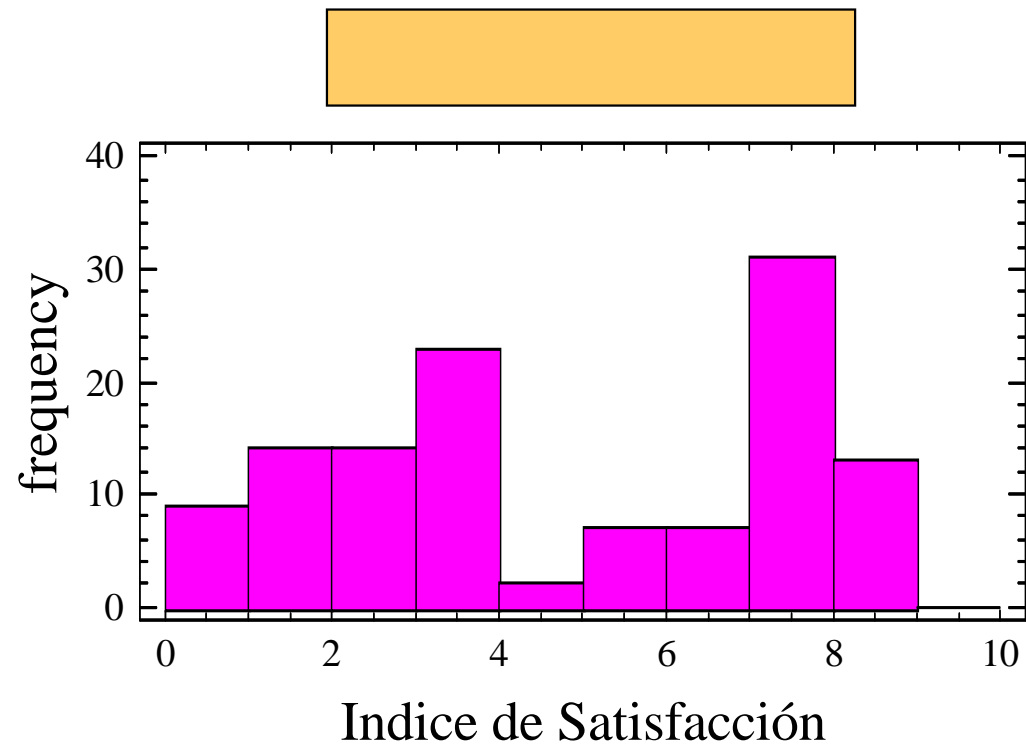
**Tercer  
Curso**





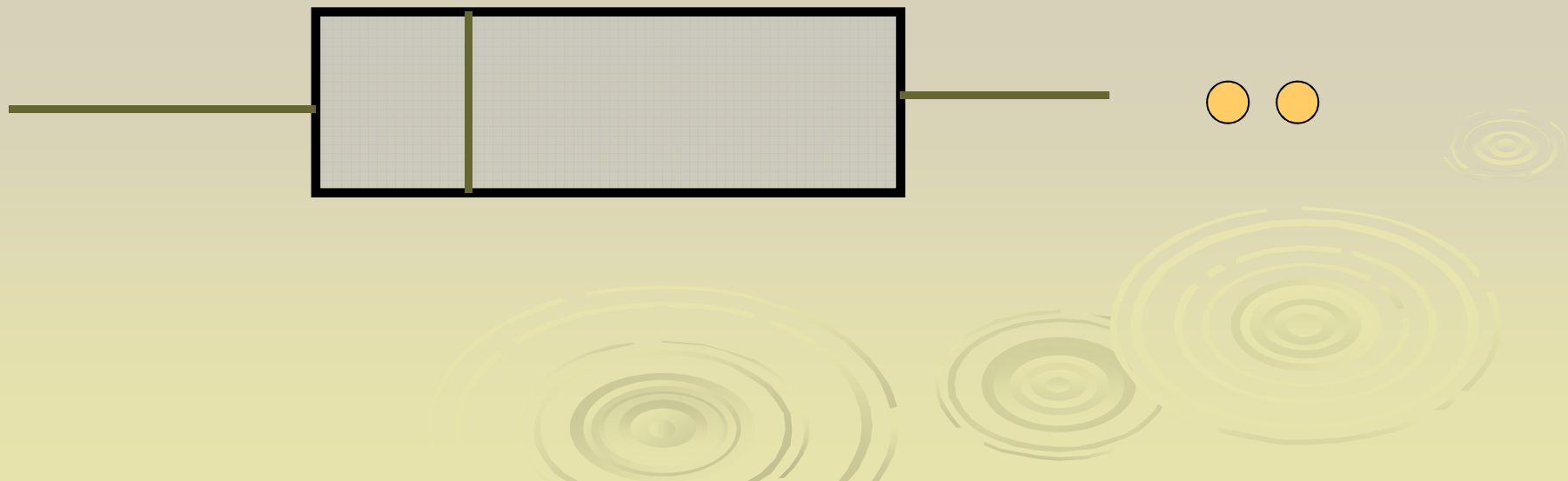
# Aprender a ver histogramas:

## Distribuciones bimodales

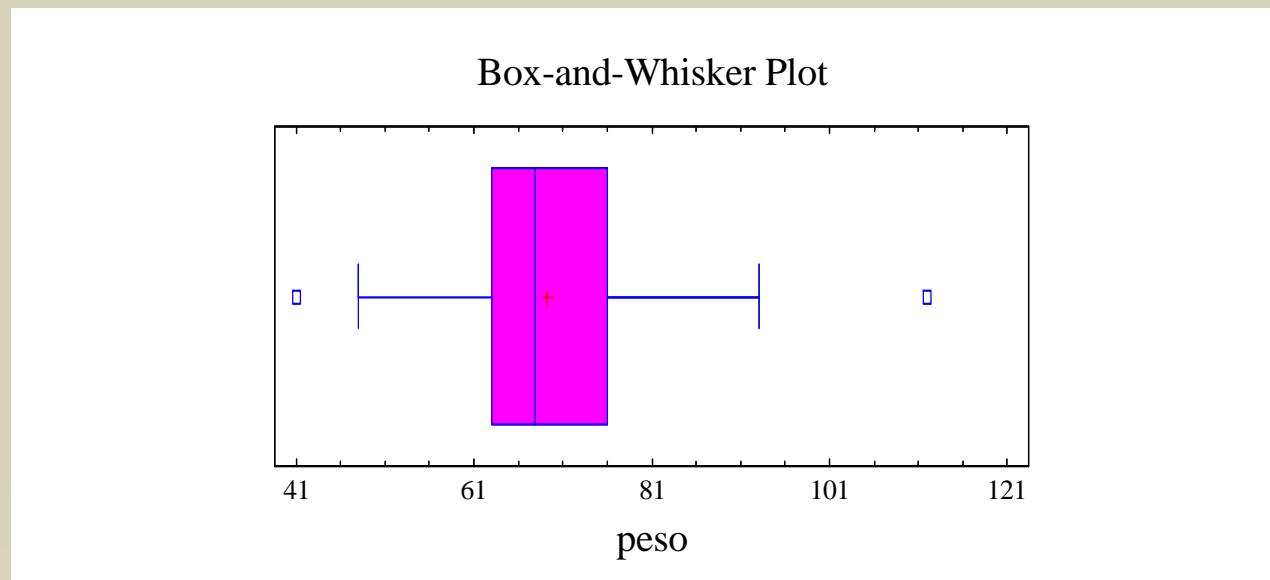


## Un gráfico fantástico: El diagrama de caja: Box-plot

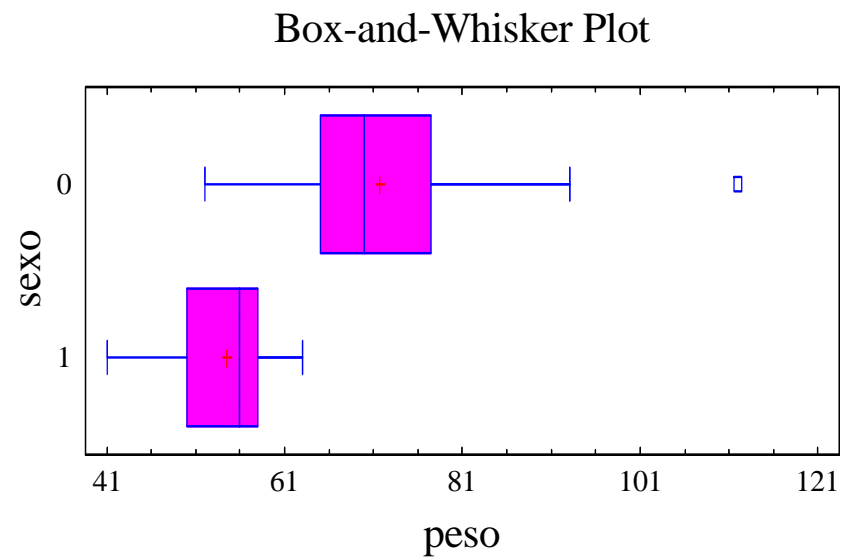
1. Se construye una caja que contiene el 50% central de los datos
2. Se dibuja la mediana
3. Se dibujan dos líneas hasta los puntos de corte (que calcula el ordenador)
4. Se dibujan los puntos que quedan fuera: Puntos atípicos



# Interpretar Box-plots

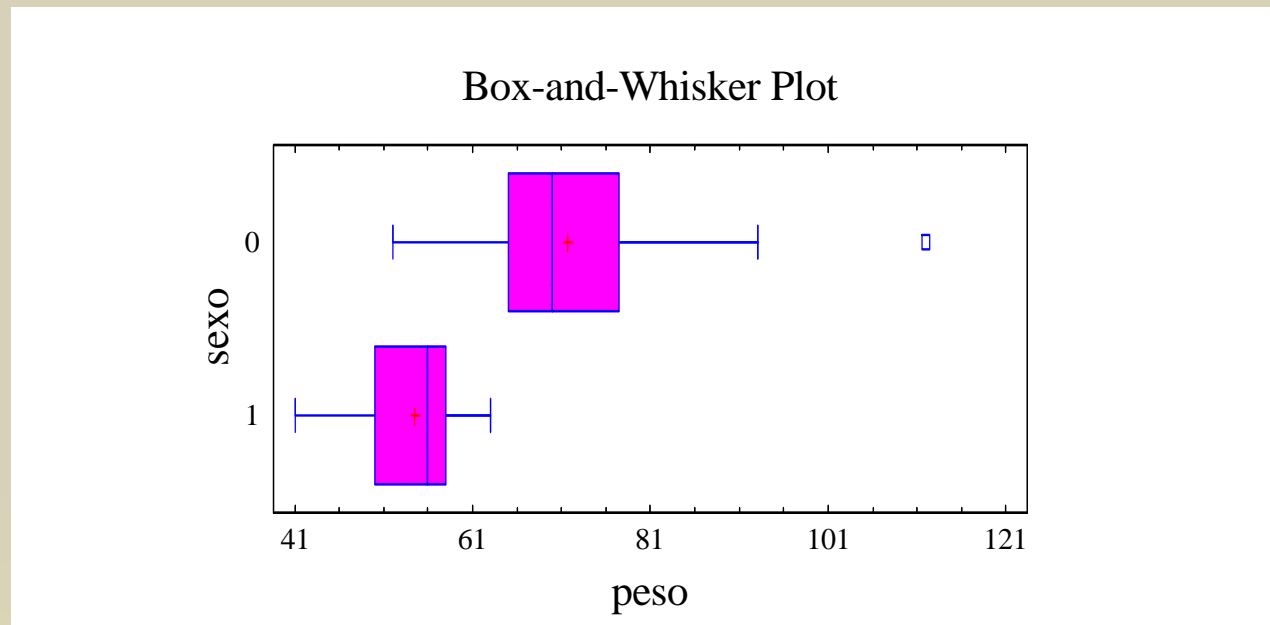


# Interpretar Box-plots por sexo



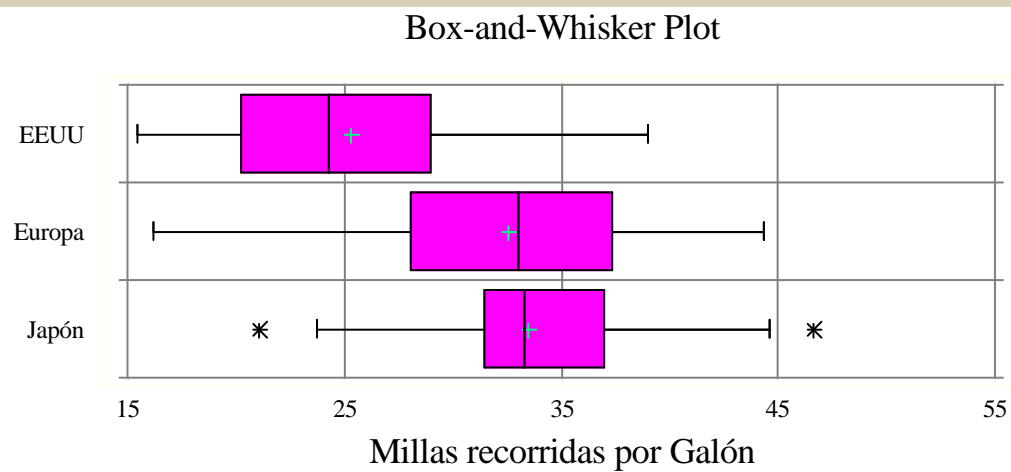


# Interpretar Box-plots por sexo



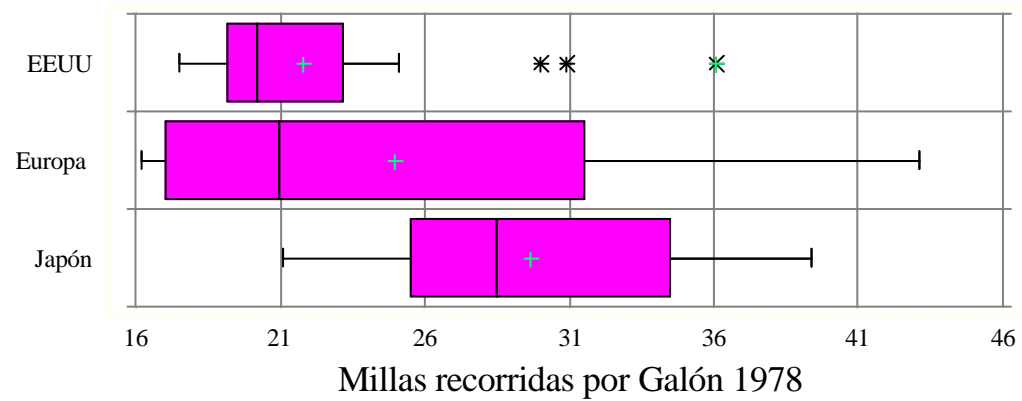
**ESTAMOS analizando una variable cuantitativa  
según otra cualitativa**

## Consumo de los automóviles vendidos en Estados Unidos desde 1978 a 1981 por origen

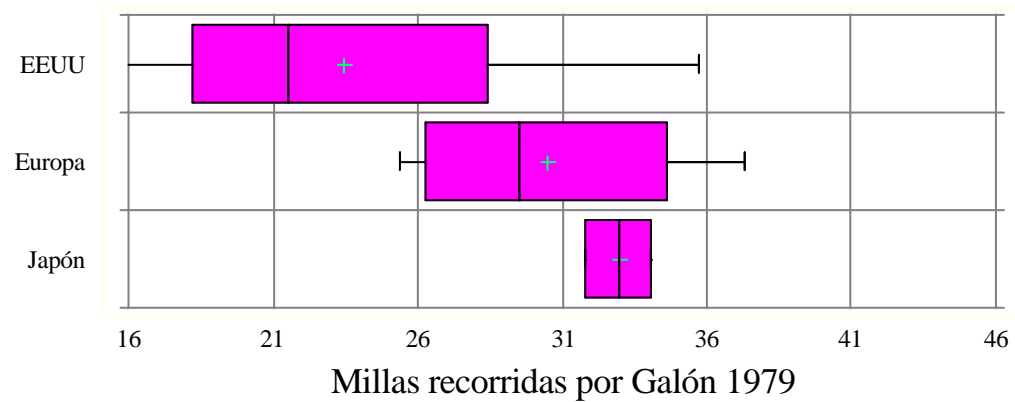


## Consumo de los automóviles vendidos en Estados Unidos por origen separando por años.

Box-and-Whisker Plot

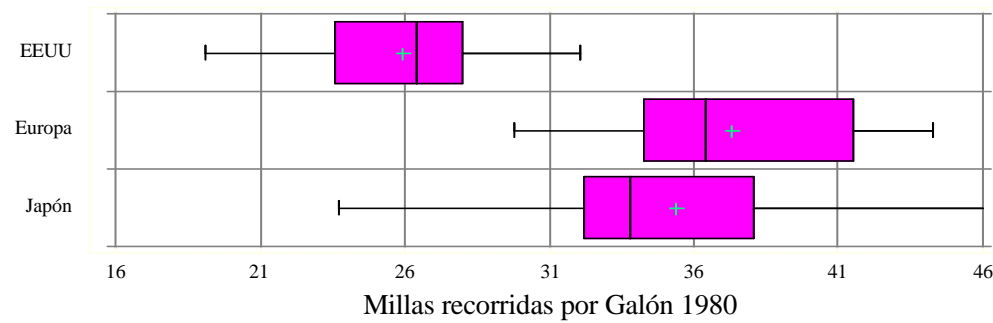


Box-and-Whisker Plot

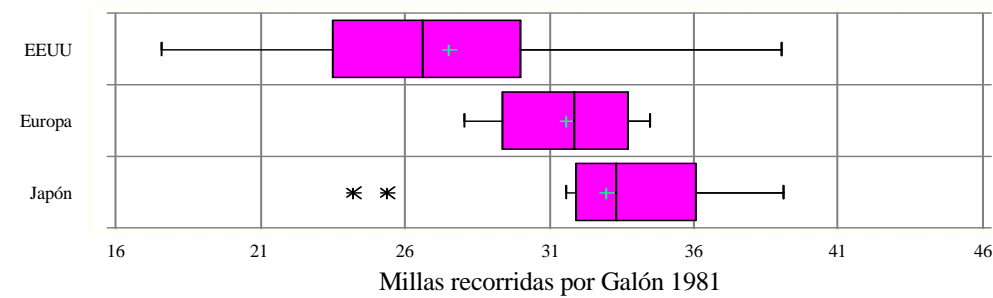




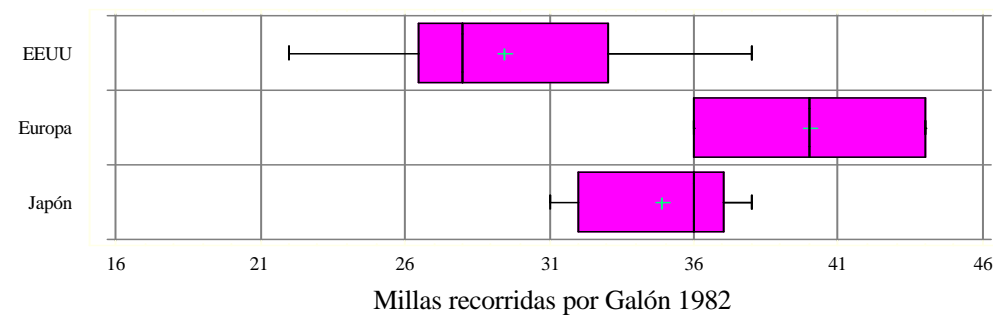
Box-and-Whisker Plot



Box-and-Whisker Plot

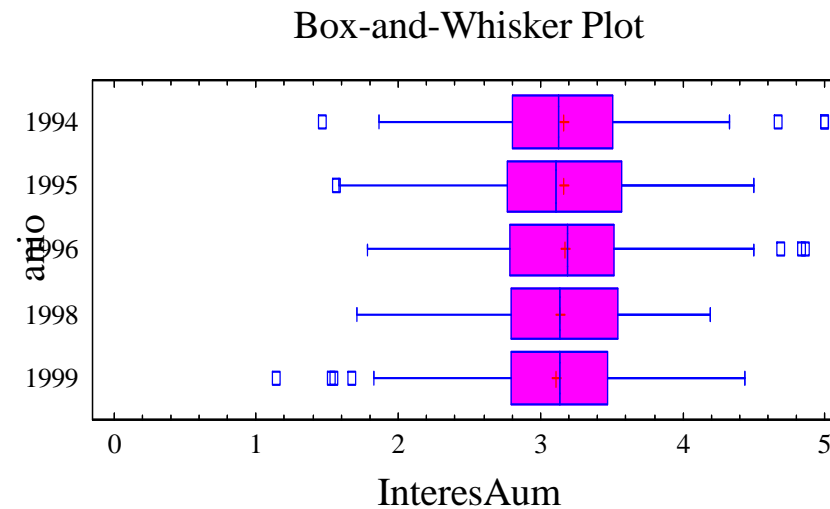


Box-and-Whisker Plot



# Interpretar Box-plots:

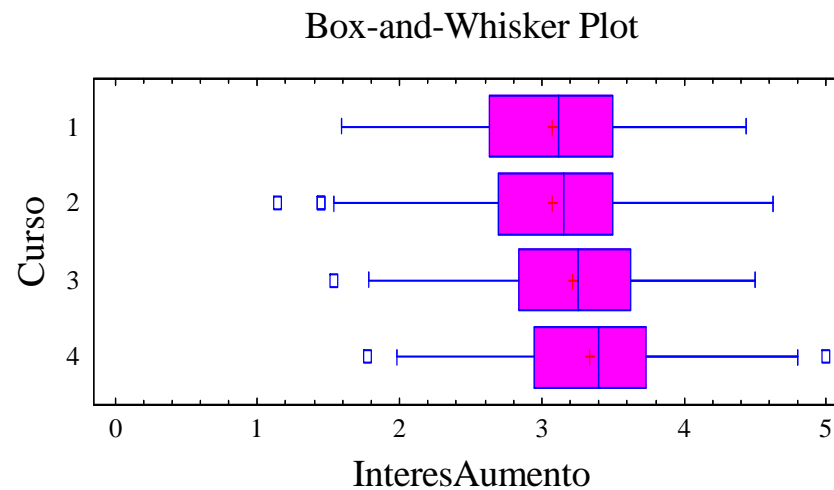
Aumenta el interés: Ing. Industrial



Hay que identificar esas asignaturas complicadas

# Interpretar Box-plots:

Aumenta el interés:Admón empresas



Hay que identificar esas asignaturas complicadas

# Vamos a estudiar cómo se representan gráficamente

## ➤ Datos cuantitativos

- Transversales: Histograma, Box plot.
- Temporales: Gráfico de la serie

## ➤ Datos cualitativos

- Transversales: Tarta, Barras
- Temporales: Gráfico de la serie



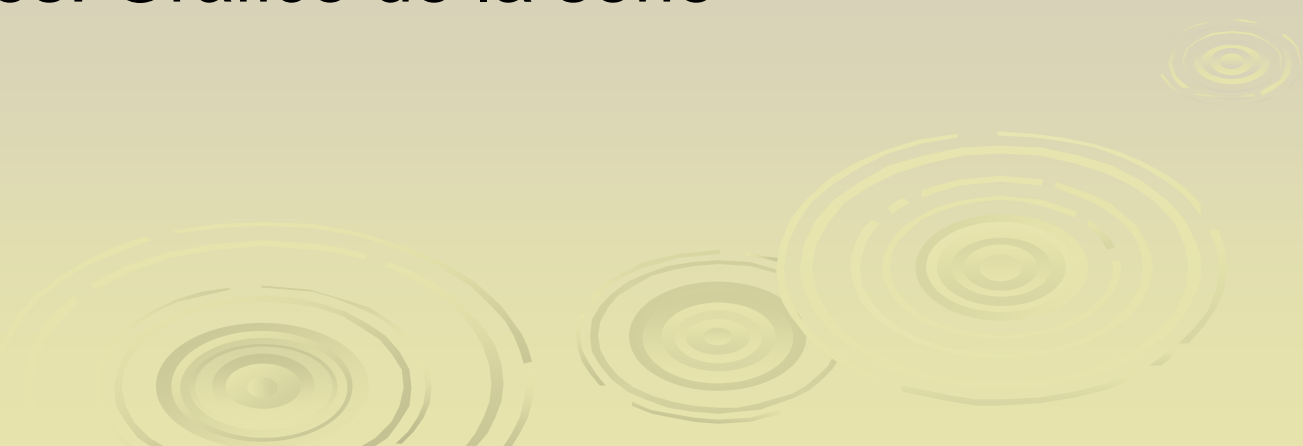
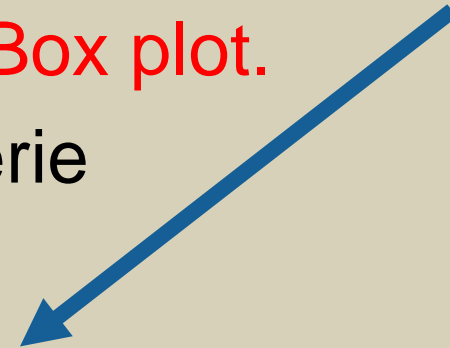
# Vamos a estudiar cómo se representan gráficamente

## ➤ Datos cuantitativos

- Transversales: Histograma, Box plot.
- Temporales: Gráfico de la serie

## ➤ Datos cualitativos

- Transversales: Tarta, Barras
- Temporales: Gráfico de la serie



# Datos cualitativos: Bar chart

## Encuesta de satisfacción de clientes:

- Muy satisfechos
- Bastante satisfechos
- Medianamente satisfechos
- Descontentos

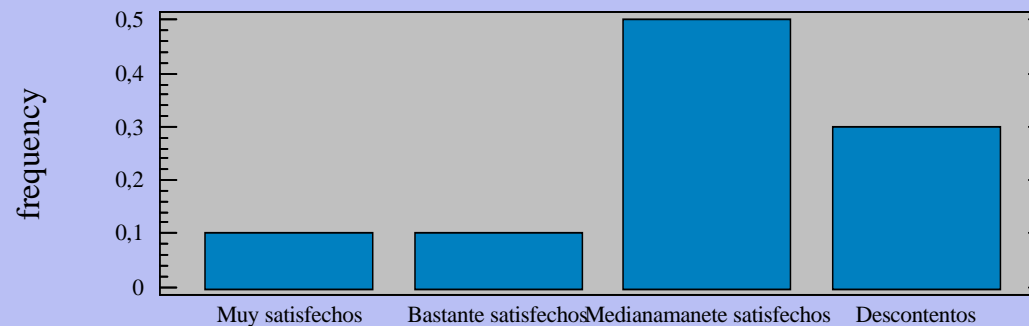


Diagrama de barras

¿Diferencia con el histograma?

# Datos cualitativos: Bar chart

## Nivel de estudios de 75 personas:

1 Analfabeto

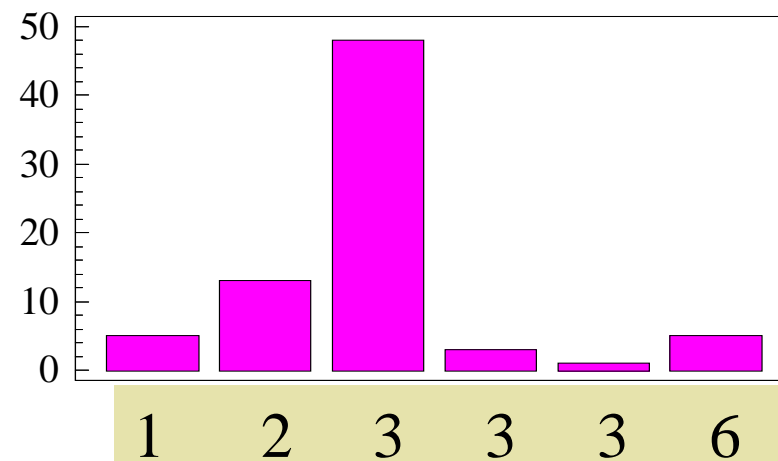
2 Sin estudios

3 Primarios

4 Secundarios

5 Medios

6 Estudios superiores



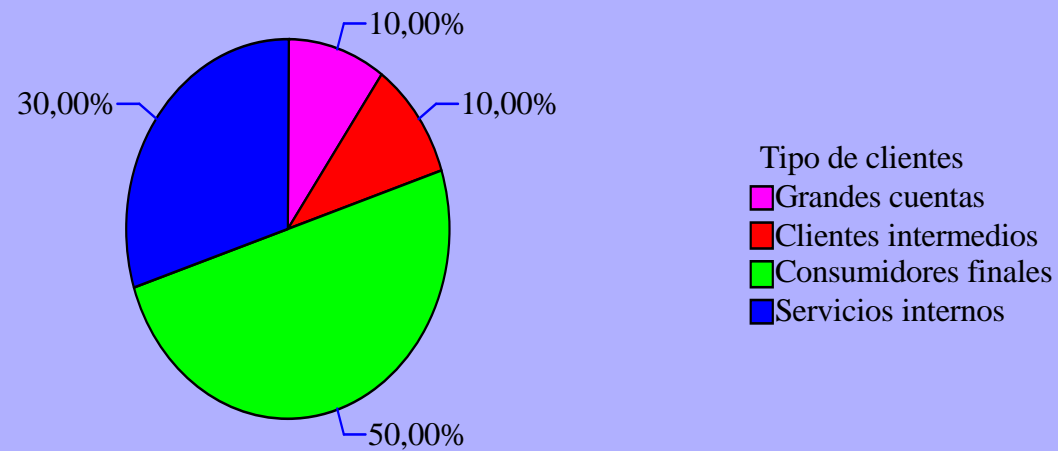
# Datos cualitativos transversales : Pie chart



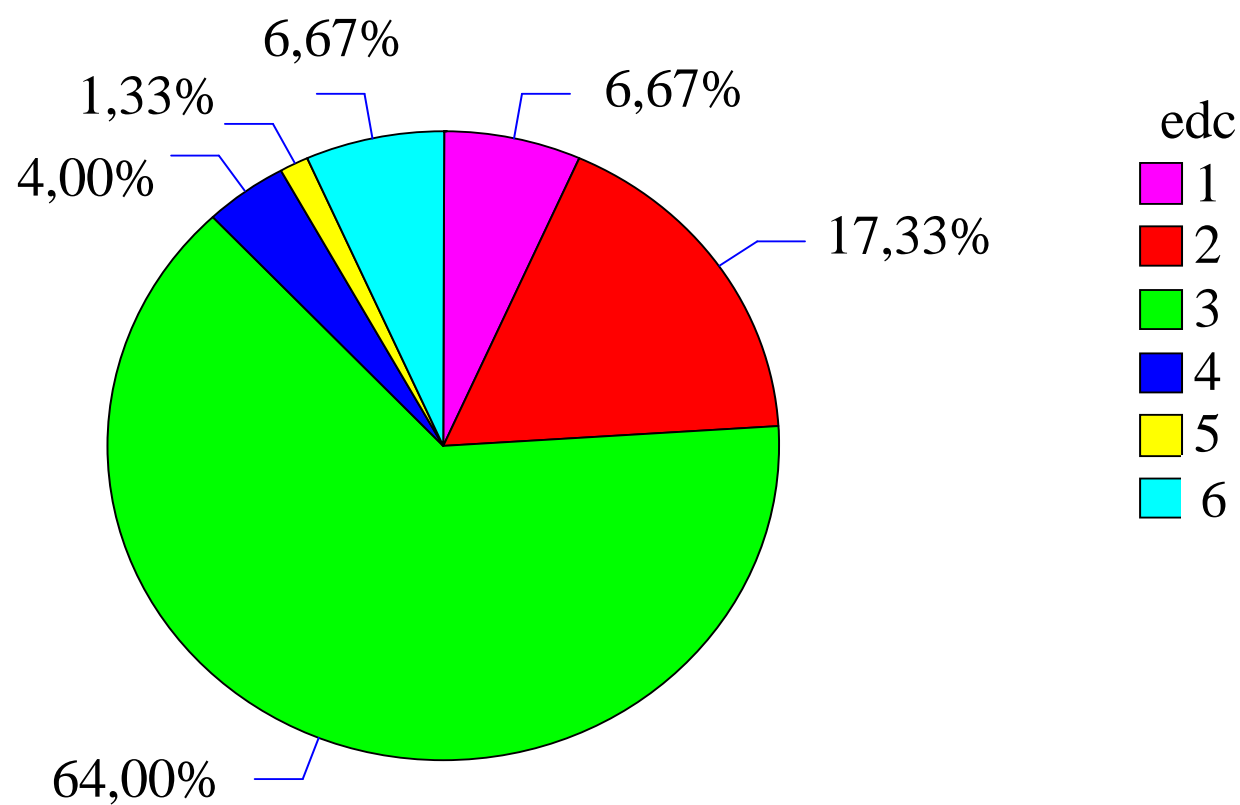


# Datos cualitativos transversales : Pie chart

Gráfico de proporciones



## Datos cualitativos transversales : Pie chart



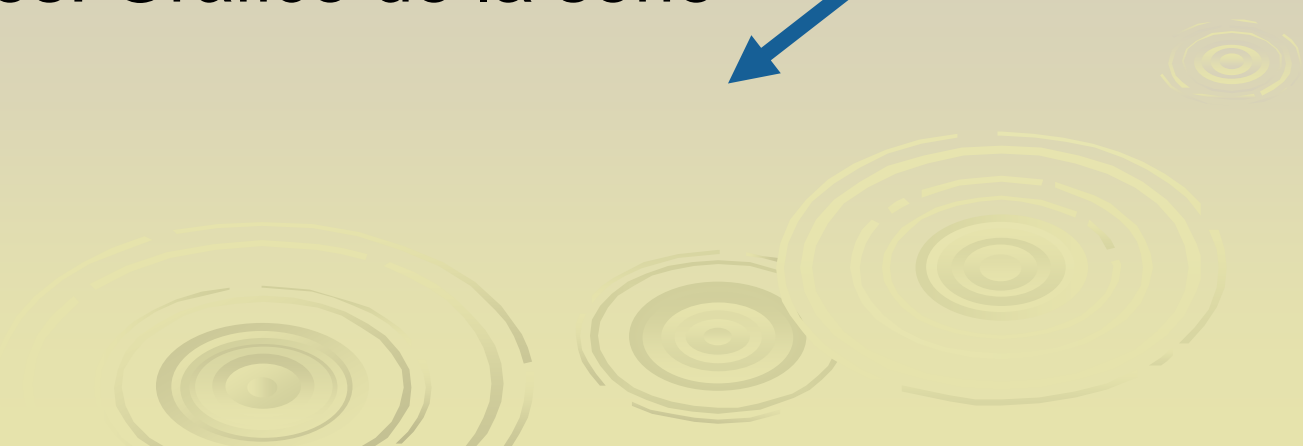
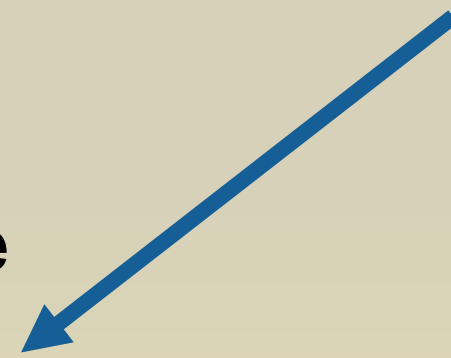
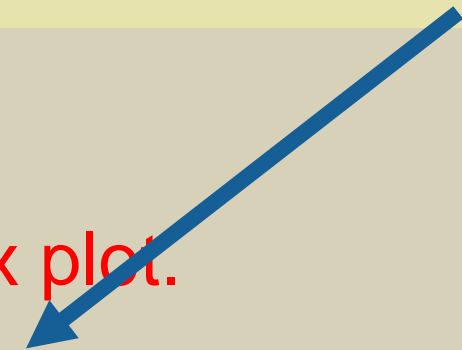
# Vamos a estudiar cómo se representan gráficamente

## ➤ Datos cuantitativos

- Transversales: Histograma, Box plot.
- Temporales: Gráfico de la serie

## ➤ Datos cualitativos

- Transversales: Tarta, Barras
- Temporales: Gráfico de la serie



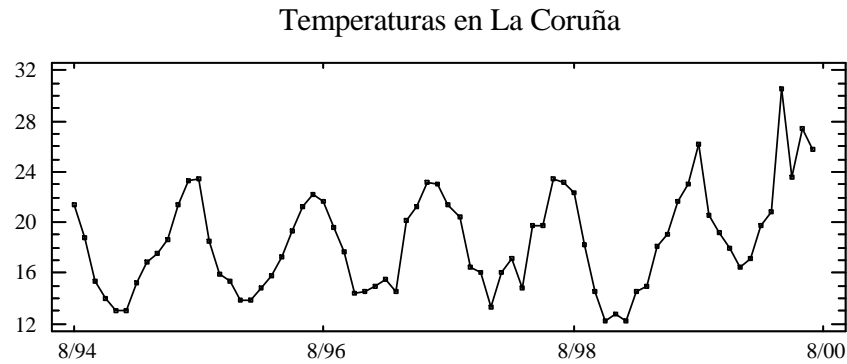
# Series temporales:

## ➤ Datos de evolución de variables en el tiempo:

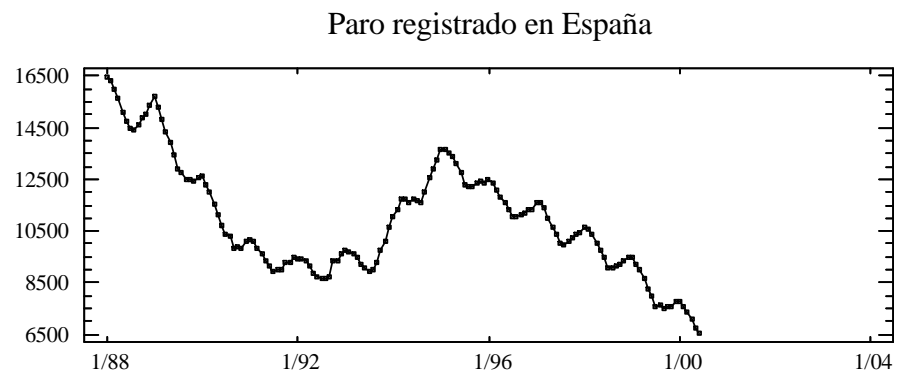
1. Periodicidad: Frecuencia de recogida de datos. Anual, mensual
2. Tendencia: Si aumenta o disminuye con el tiempo
3. Ciclo estacional Se observa un ciclo ligado al momento del año en que se ha recogido el dato

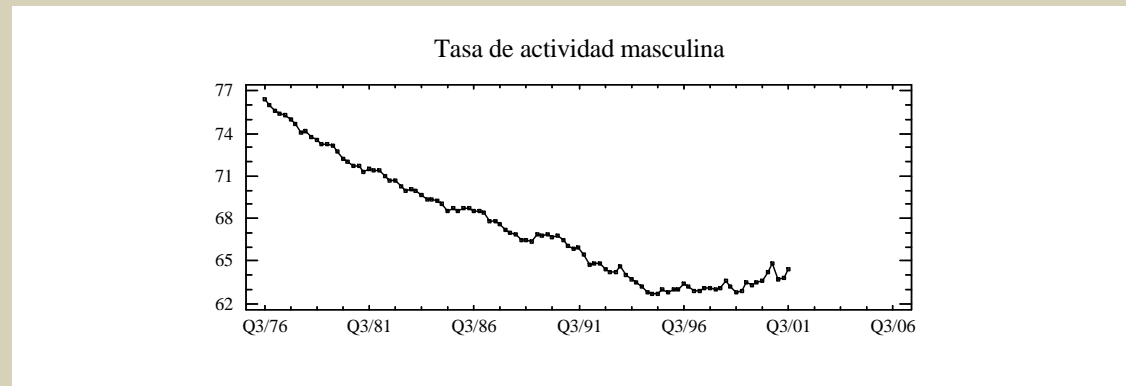


## Periodicidad mensual- Ciclo estacional



## Periodicidad mensual- Tendencia- Ciclo estacional





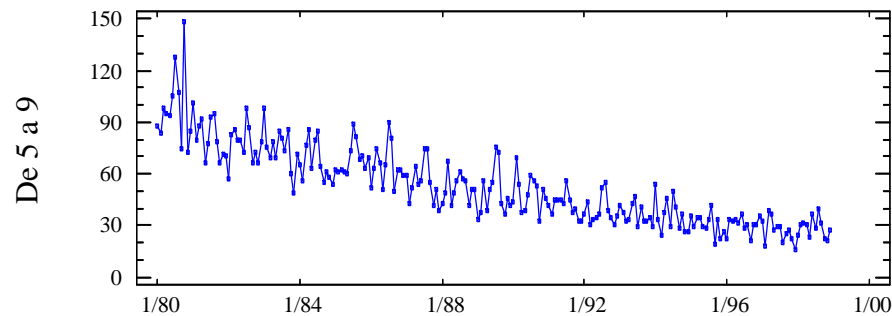
¿Tiene tendencia?

# Mortalidad Española por edades e infantil en la UE desde 1975

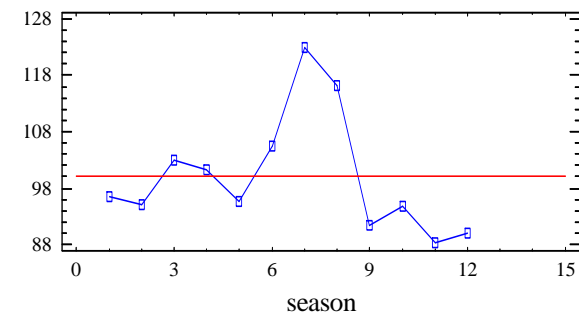
- Fuente de datos: INE. [www.ine.es](http://www.ine.es)  
INEBASE

# Mortalidad española. Número de personas Por edades desde Enero de 1980. Datos mensuales Jóvenes: de 5 a 9 y de 20 a 24 años

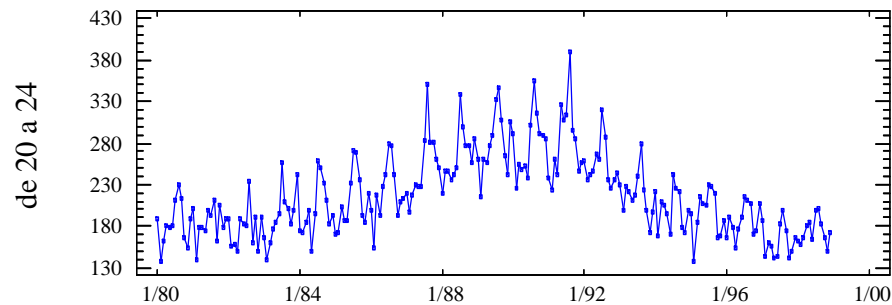
Mortalidad personas de 5 a 9 años



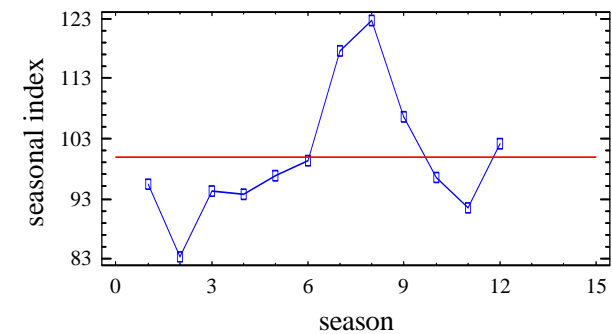
Seasonal Index Plot for De 5 a 9



Mortalidad personas de 20 a 24 años



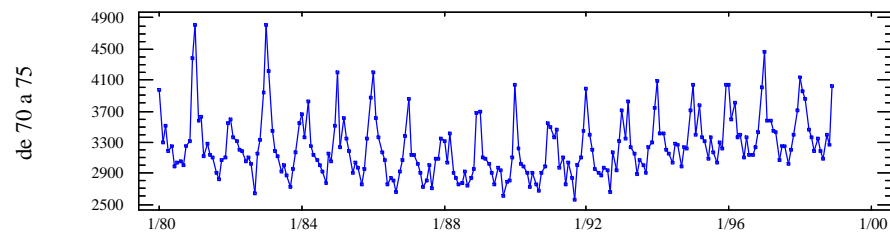
Seasonal Index Plot for de 20 a 24



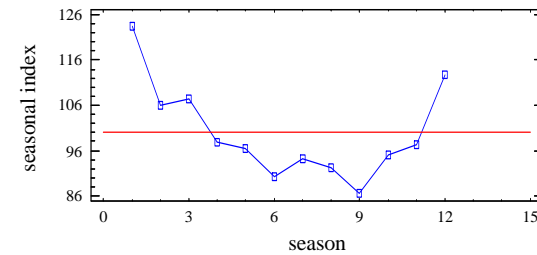


# Personas mayores: 70 a 75 y 80 a 85

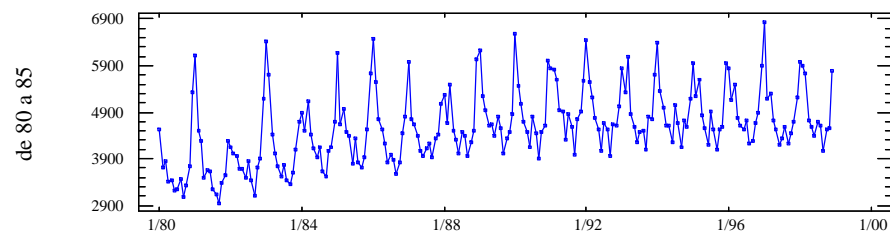
Mortalidad personas de 70 a 75 años



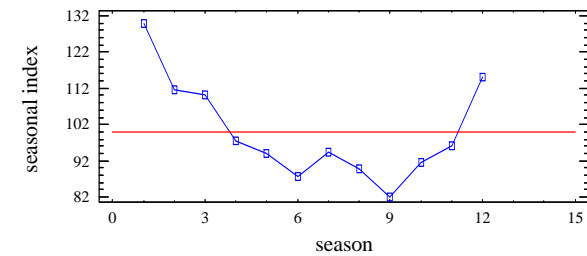
Seasonal Index Plot for de 70 a 75



Mortalidad personas de 80 a 85 años



Seasonal Index Plot for de 80 a 85



# Vamos a estudiar cómo se representan gráficamente

## ➤ Datos cuantitativos

- Transversales: Histograma, Box plot.
- Temporales: Gráfico de la serie

## ➤ Datos cualitativos

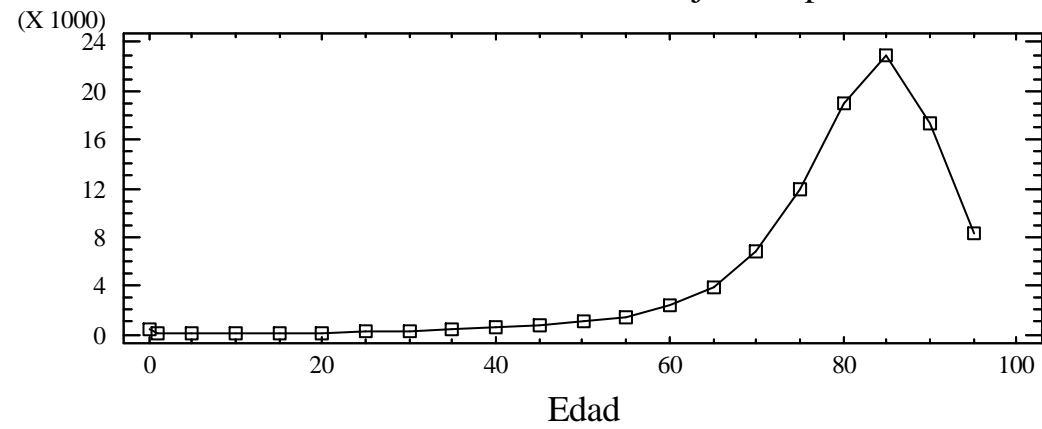
- Transversales: Tarta, Barras
- Temporales: Gráfico de la serie



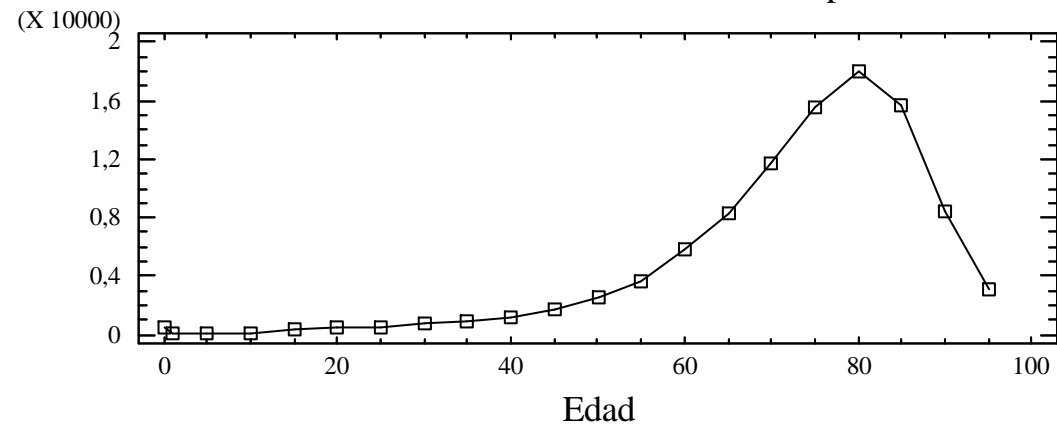
## Datos especiales

- Mortalidad por edades y sexo en España, Namibia, Afganistán y Francia
- Fuente OMS

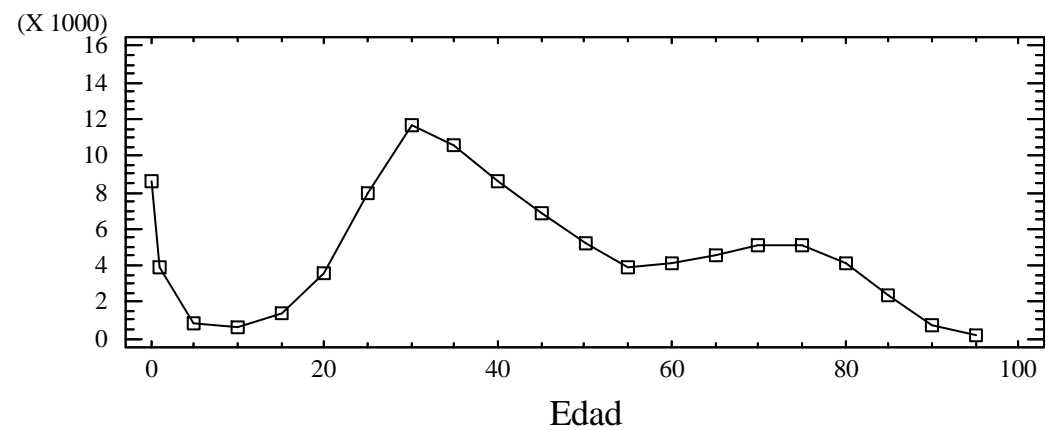
Distribución de muertes Mujeres España



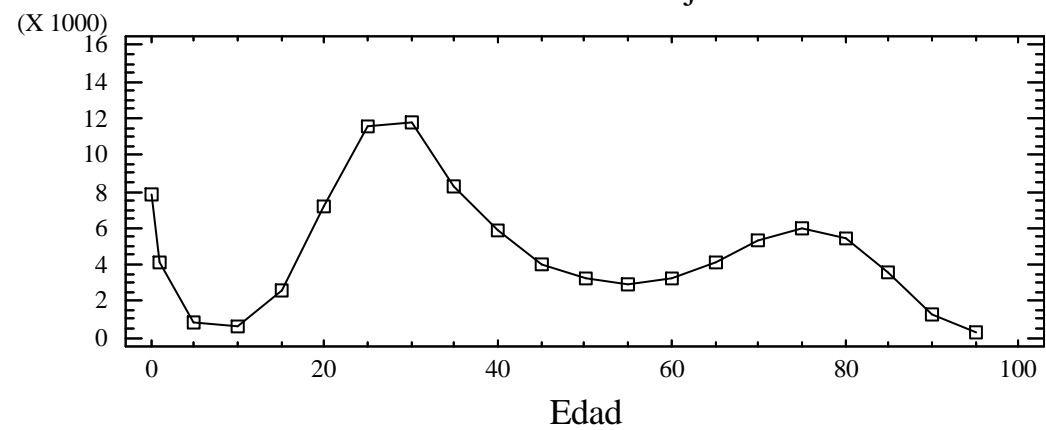
Distribución de muertes Hombres España

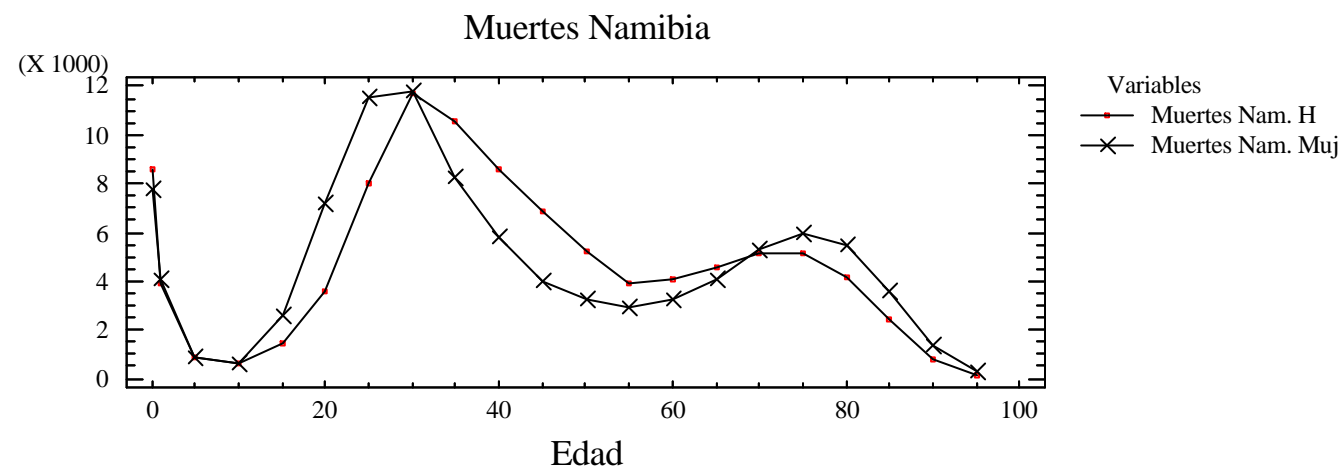
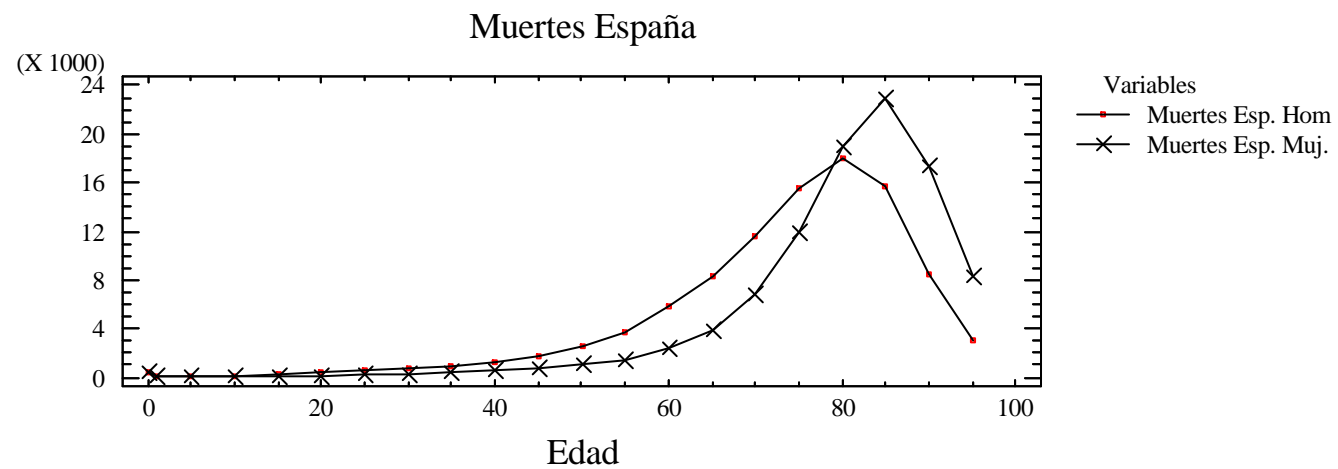


Distribución de muertes Hombres Namibia



Distribución de muertes Mujeres Namibia





## Medidas analíticas (números) para describir los datos

- Posición de la variable:
  - Media (No la explico por obvia)
  - Mediana (Observación del medio)
- Dispersión o amplitud de la variable
  - Desviación típica: Mide la amplitud de los datos



# Medidas analíticas (números) para describir los datos

Mediana (Observación del medio de los datos)



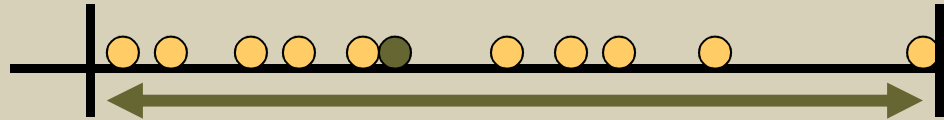
Representa estupendamente la ubicación de los datos.  
Es una medida con ventajas respecto a la media



# Medidas analíticas (números) para describir los datos

Dispersión: mide la amplitud de los datos

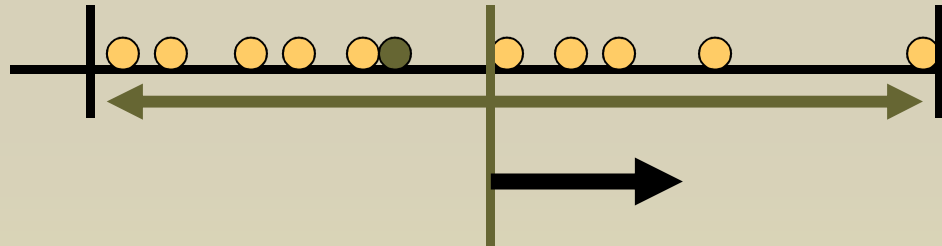
Rango mide la amplitud de los datos



# Medidas analíticas (números) para describir los datos

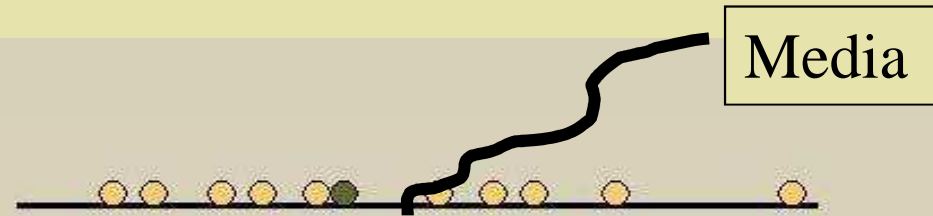
Dispersión: mide la amplitud de los datos

Rango mide la amplitud de los datos



Desviación típica mide la amplitud de los datos

# ¿Por qué la mediana es muy útil?



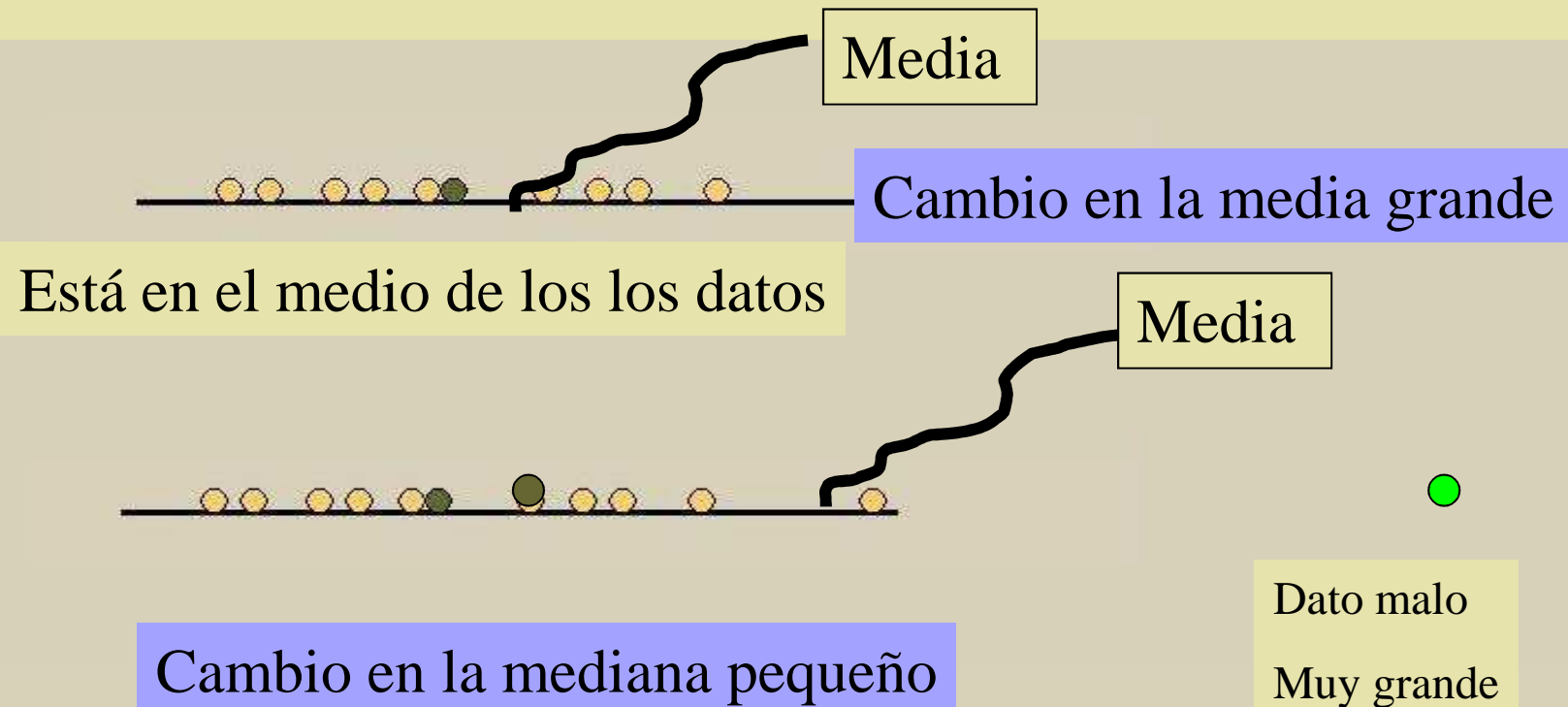
Está en el medio de los los datos



Dato malo

Muy grande

# ¿Por qué la mediana es muy útil?



## Medidas analíticas para describir los datos

- ***La media no siempre es una medida satisfactoria.***  
***Ejemplo***

**En un hotel se pregunta a los clientes por su valoración de:**

- Limpieza de la habitación
- Rapidez en los trámites de entrada
- Iluminación del BAR

Las valoraciones para dos hoteles han sido:

Variable – Atributo de Calidad	Valor medio obtenido: 6 HOTEL 1	Valor medio obtenido: 6 HOTEL 2
Limpieza Habitación	3	8
Rapidez trámites entrada	5	7
Iluminación BAR	10	3

¿A qué hotel irías?

## ¿Son todos los atributos de calidad igualmente importantes?

Si pensamos que la importancia es:

- Limpieza habitación: 50%
- Recepción: 40%
- Iluminación del Bar: 10%

➤ HOTEL 1:  $0.5 \times 3 + 0.4 \times 5 + 0.1 \times 10 = 4.5$

➤ HOTEL 2:  $0.5 \times 8 + 0.4 \times 7 + 0.1 \times 3 = 7.1$

# Relación entre dos variables

- Gráfico de dispersión
- Gráfico de dispersión múltiple
- Correlaciones



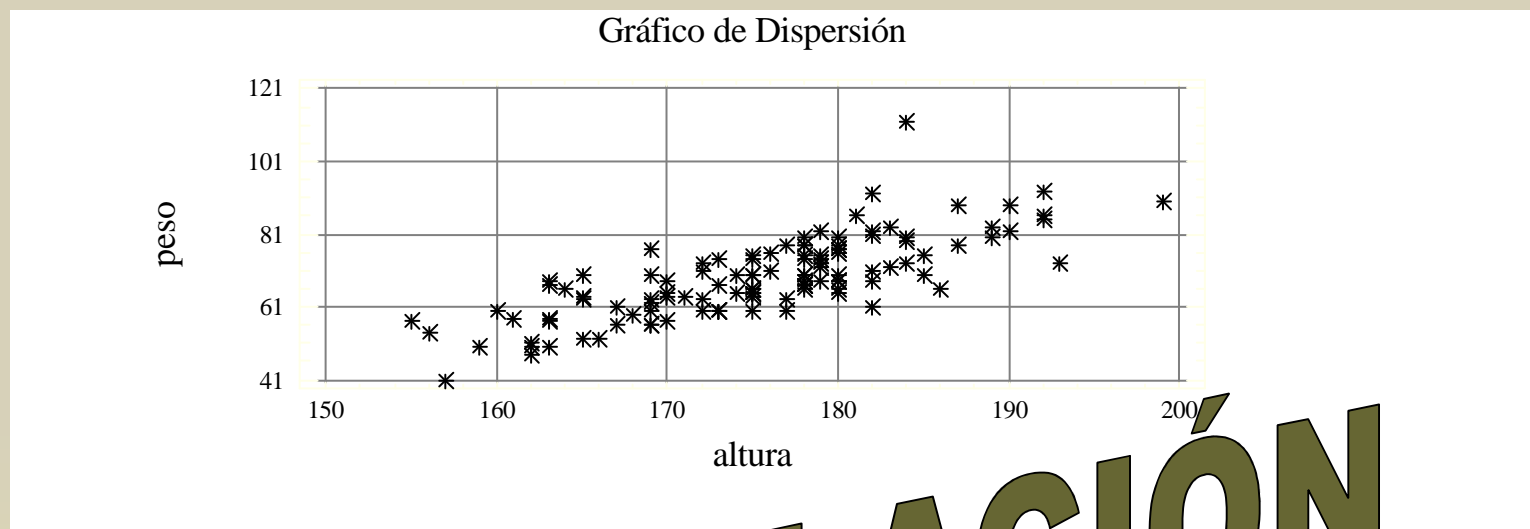


## Relación entre dos variables:

- Normalmente hay que analizar más de un aspecto de interés:
  - VARIAS VARIABLES.
- Gráfico de dispersión (*Scatterplot*) resulta muy útil.

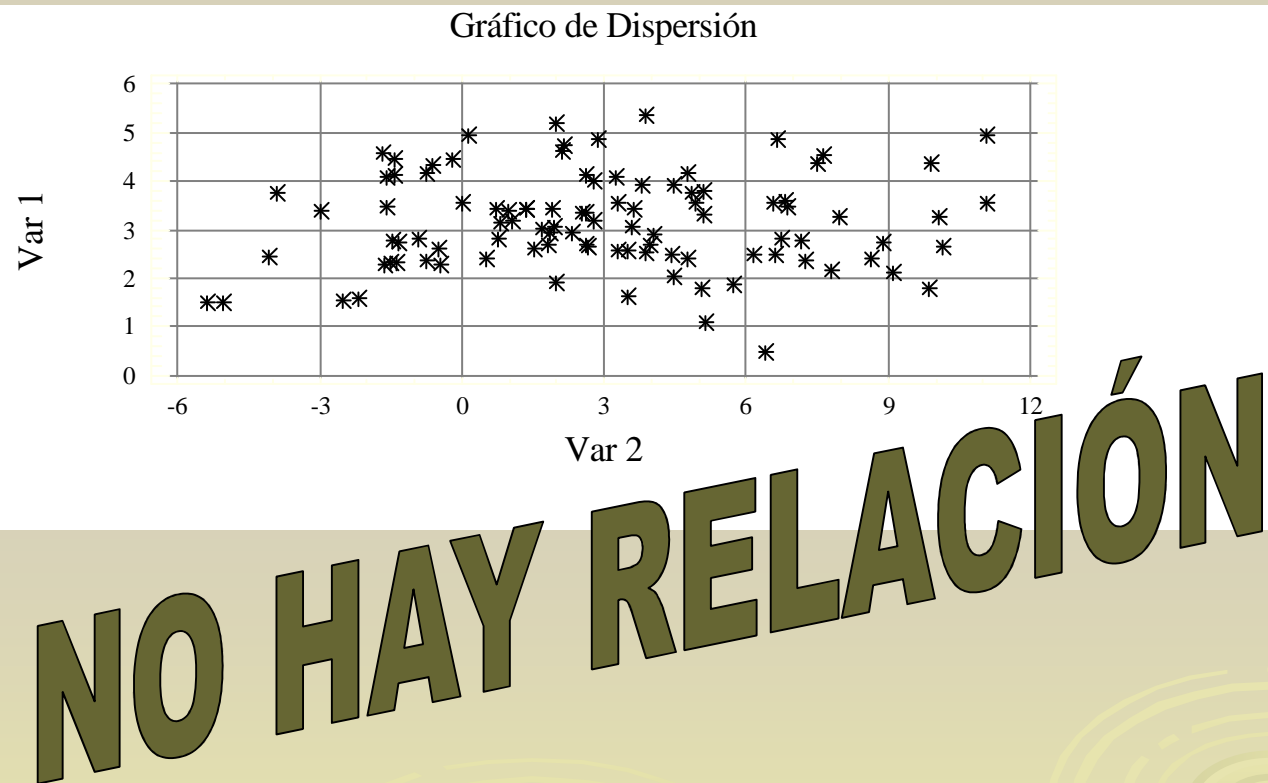


# Relación entre dos variables: PESO Y ALTURA

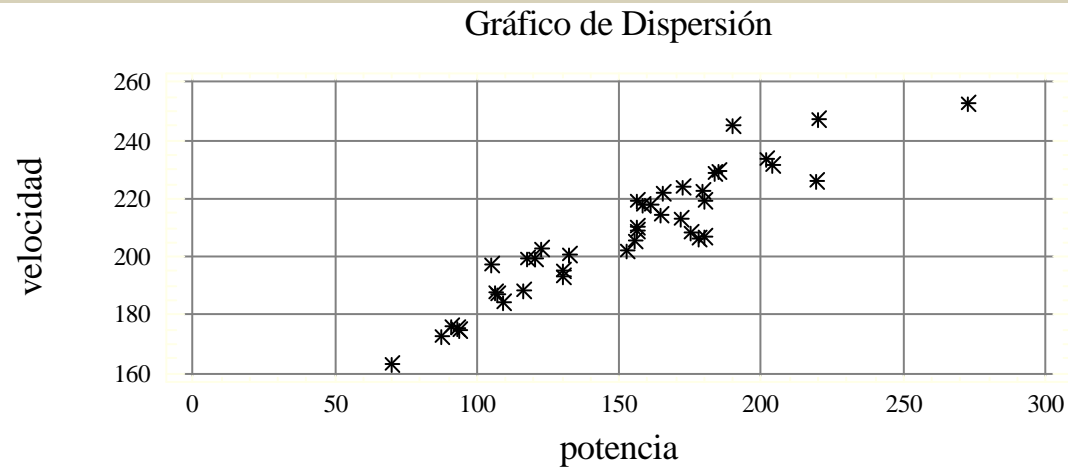


**HAY RELACIÓN**

No hay relación entre dos variables:



# ¿Hay relación entre estas variables?



# Para medir el grado de relación entre variables

- Utilizamos la correlación.
- Varía entre  $-1$  y  $+1$

## Interpretación de la correlación:

**-1**

**0**

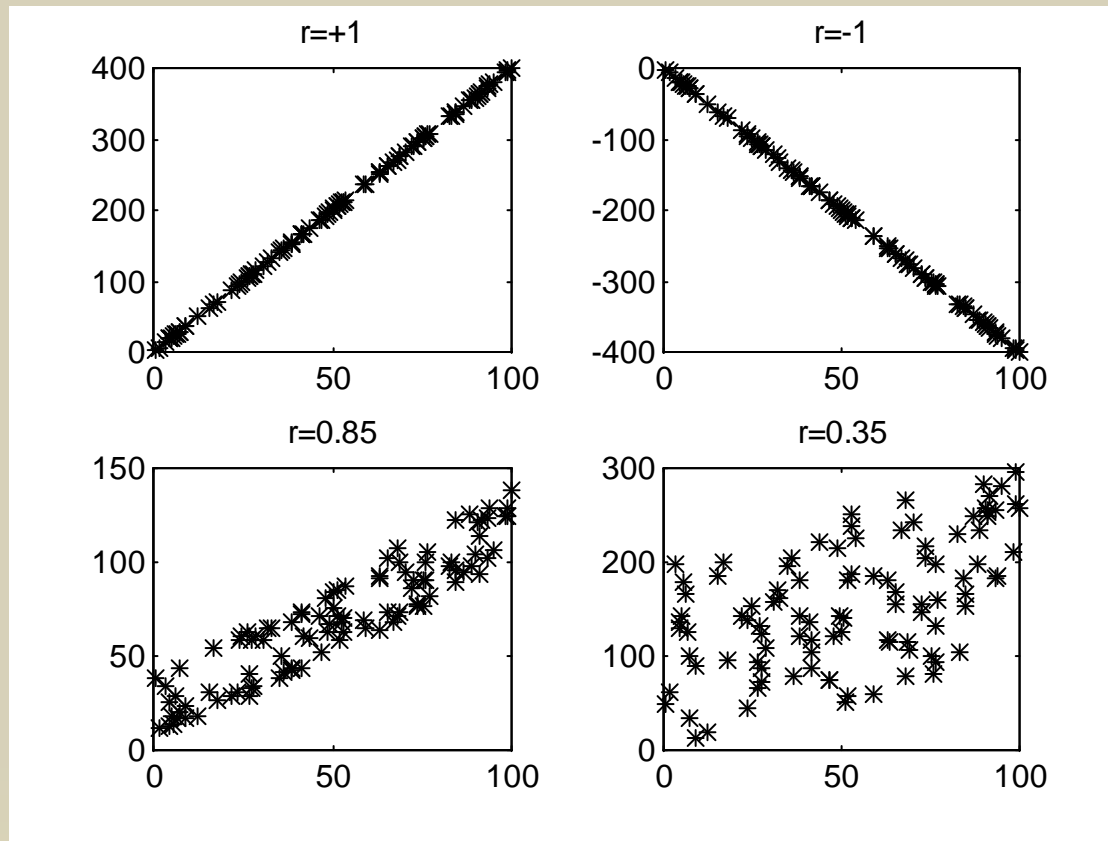
**+1**

**Mucha relación  
Decreciente**

**No hay relación**

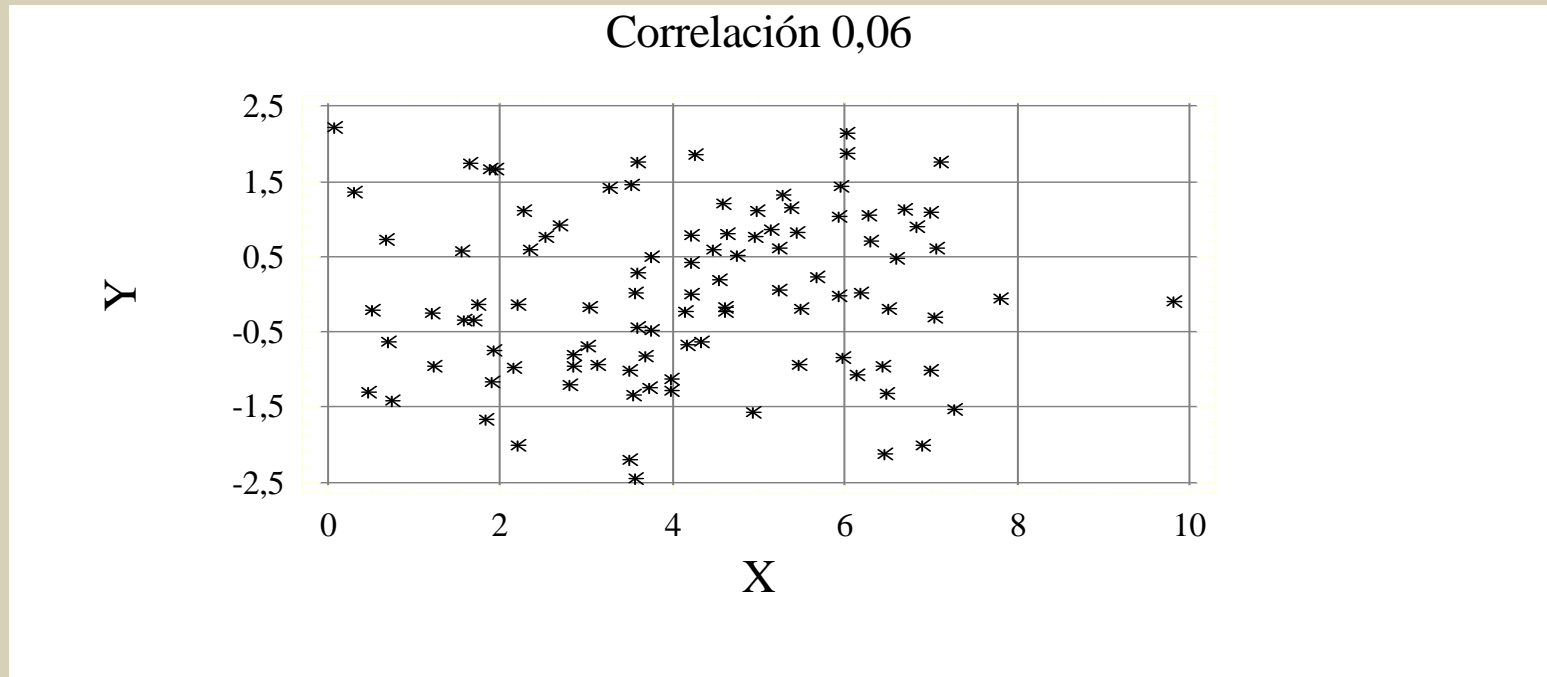
**Mucha relación  
Creciente**

# Interpretación de la correlación



- + Relación creciente: Si una variable aumenta, la otra también
- Relación decreciente: Si una variable aumenta, la otra disminuye

# Interpretación de la correlación



Si la correlación es muy pequeña indica falta de relación entre las variables.

## Dispersión múltiple

- Cuando tenemos muchas variables hacer los gráficos de dos en dos es muy latoso
- Los gráficos múltiples hacen de golpe todas los graficos:

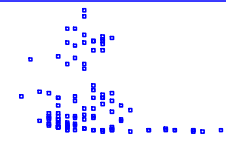
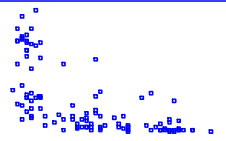
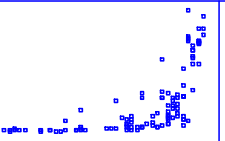
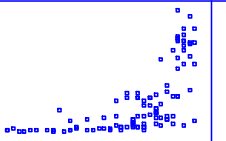
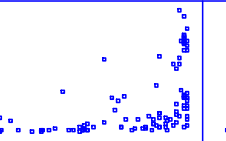
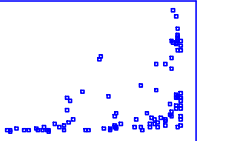
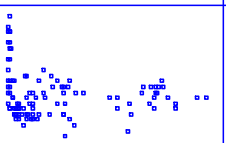
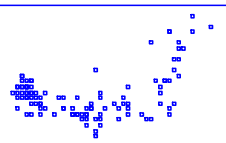
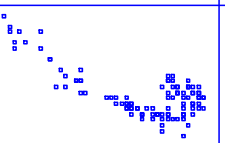
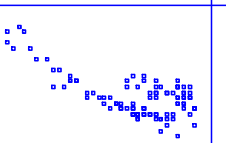
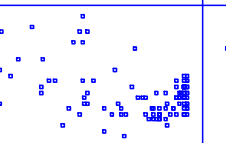
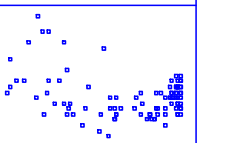
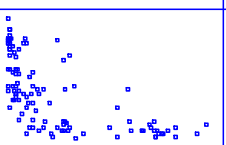
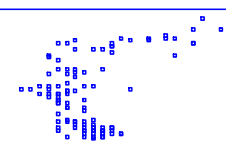
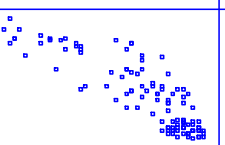
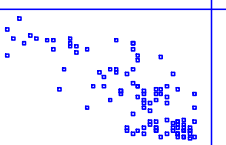
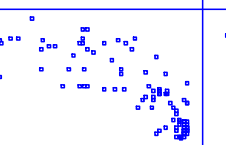
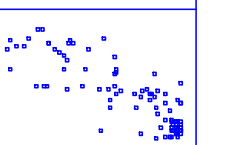
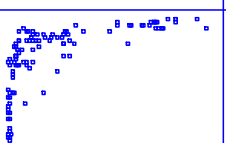
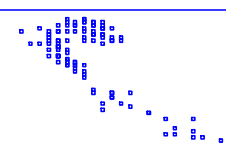
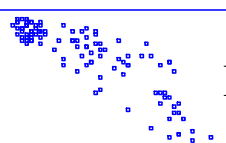
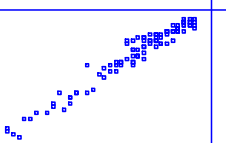
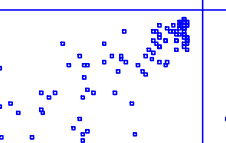
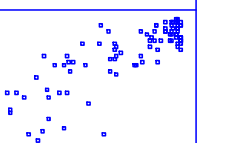
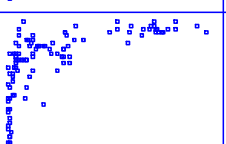
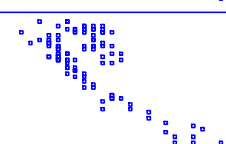
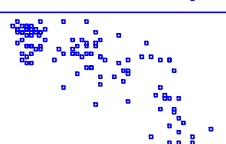
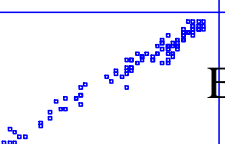
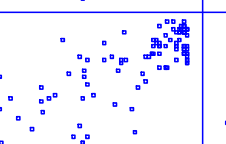
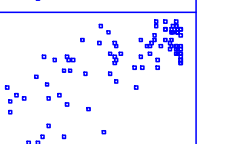

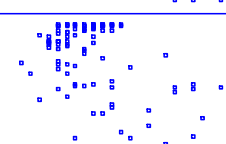
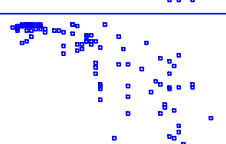
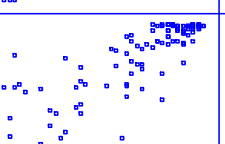
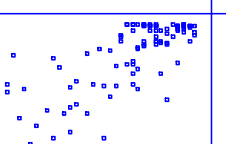
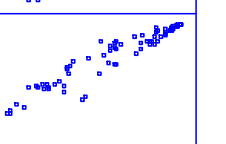

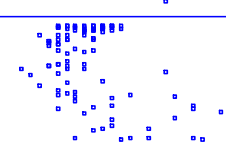
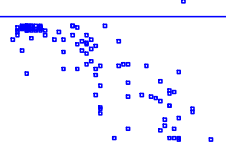
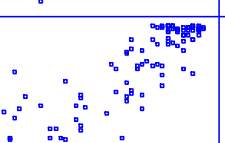
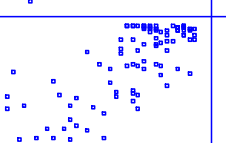
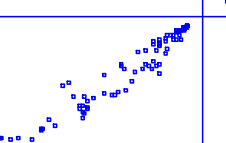




## Dispersión múltiple

- Saca todos los gráficos de dispersión entre un grupo de variables
- Para países del mundo en 1995



PIB CAP						
	TasaMort					
		TasaNata				
			Esp vida Fem			
				Esp vida Masc		
					Alfabmasc	
						Alfabfem

# Y podemos obtener las correlaciones por parejas

## Correlaciones

	Esp vida Fem	Hijpromedio	Inc Pobl anual	Alfabfem	TasaNata
Esp vida Fem		-0,8374 (107)	-0,5804 (109)	0,8187 (85)	-0,8620 (109)
Hijpromedio	-0,8374 (107)		0,8410 (107)	-0,8386 (85)	0,9749 (107)
Inc Pobl anual	-0,5804 (109)	0,8410 (107)		-0,6407 (85)	0,8618 (109)
Alfabfem	0,8187 (85)	-0,8386 (85)	-0,6407 (85)		-0,8349 (85)
TasaNata	-0,8620 (109)	0,9749 (107)	0,8618 (109)	-0,8349 (85)	