



# Tema 2: Análisis de datos univariantes

## 1. Representaciones y gráficos

- Tablas de frecuencias.
- Diagrama de barras, pictogramas, histograma, polígono de frecuencias, y diagrama de caja.
- Otros gráficos.

## 2. Resumen numérico

- Medidas de localización.
- Medidas de dispersión.
- Medidas de forma.

## Lecturas recomendadas:

Capítulo 2 del libro de Gonick y Smith (1999).

Capítulos 2 a 6 del libro de Peña y Romo (1997).

Capítulos 3 a 7 del libro de Portilla (2004).



## 2.1: Representaciones y gráficos

### Descripción de variables cualitativas

Frecuencia absoluta  
Frecuencia relativa  
Diagrama de barras  
Diagrama de sectores

### Descripción de variables cuantitativas

Frecuencia absoluta / frecuencia absoluta acumulada  
Frecuencia relativa / frecuencia relativa acumulada  
Histograma  
Polígono de frecuencias

### Lecturas recomendadas:

- Capítulos 2 y 3 del libro de Peña y Romo (1997)
- Capítulos 3 y 4 del libro de Portilla (2004)



## Descripción de variables cualitativas

MUESTRA: 70 estudiantes universitarios madrileños

VARIABLE: Partido político preferido

OBJETIVO: **Clasificación y representación de la información**

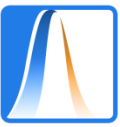
PP	IU	Otros	PP	PSOE	Otros	Otros
IU	PP	IU	PSOE	PSOE	UPD	IU
PP	PSOE	IU	PP	PSOE	Otros	PSOE
IU	IU	PSOE	IU	IU	PSOE	PSOE
PP	PSOE	PP	PP	PSOE	IU	UPD
PP	PSOE	UPD	PSOE	PP	Otros	IU
IU	PSOE	IU	PP	PSOE	IU	PSOE
IU	IU	PSOE	UPD	UPD	IU	PP
PSOE	IU	PSOE	IU	PP	PSOE	IU
PSOE	PSOE	UPD	UPD	PP	PP	PSOE





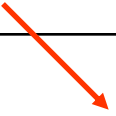
## La tabla de frecuencias

Frecuencia absoluta      Frecuencia relativa

Clase (i)	$n_i$	$f_i$
PSOE	23	0,33
PP	15	0,21 → = 15/70
IU	20	0,29
UPD	7	0,10
Otros	5	0,07
Total	70 = 23+15+20+7+5	1 = 0,33+0,21+ ...+0,07

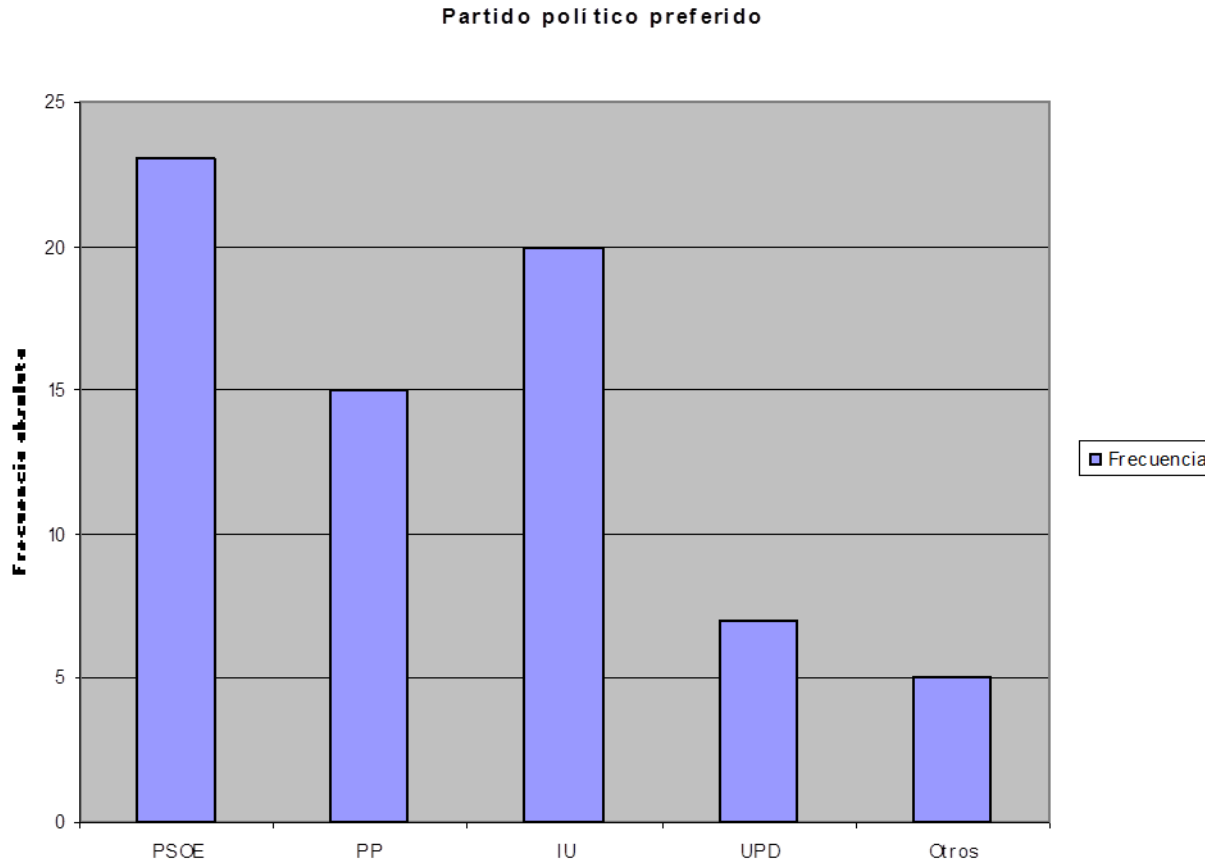


### Formato general de la tabla

Clase (i)	$n_i$	$f_i$
1	$n_1$	$f_1$  = $n_1/N$
2	$n_2$	$f_2$
3	$n_3$	$f_3$
⋮	⋮	⋮
k	$n_k$	$f_k$
Total	$N$  = $n_1 + n_2 + \dots + n_k$	$1$  = $f_1 + f_2 + \dots + f_k$



## El diagrama de barras





## Engañando con diagramas de barras

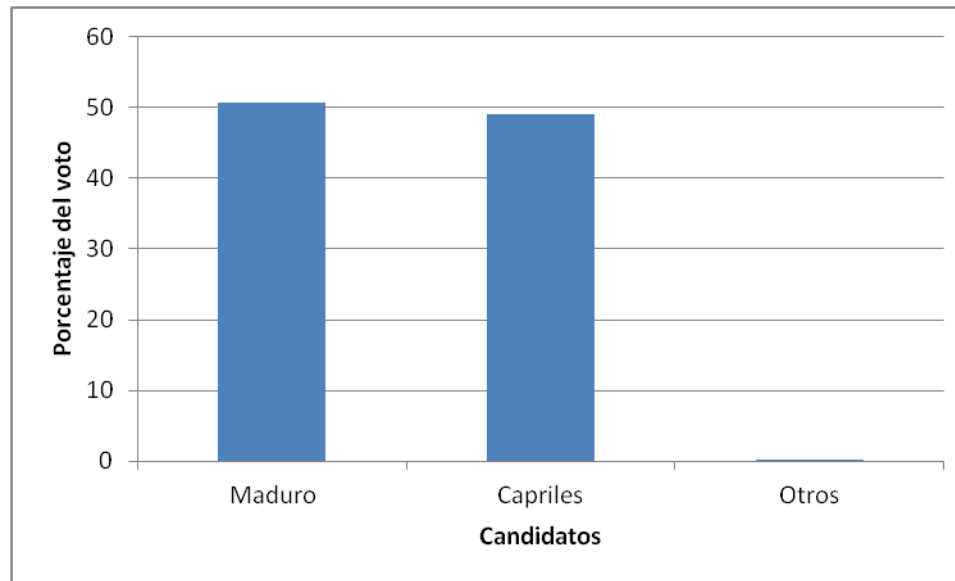
El siguiente gráfico apareció en la portada de la web de la televisión estatal de Venezuela después de las elecciones presidenciales de 2013.



Parece que arrasó Nicolás Maduro ...



... ¡si no miras los porcentajes!

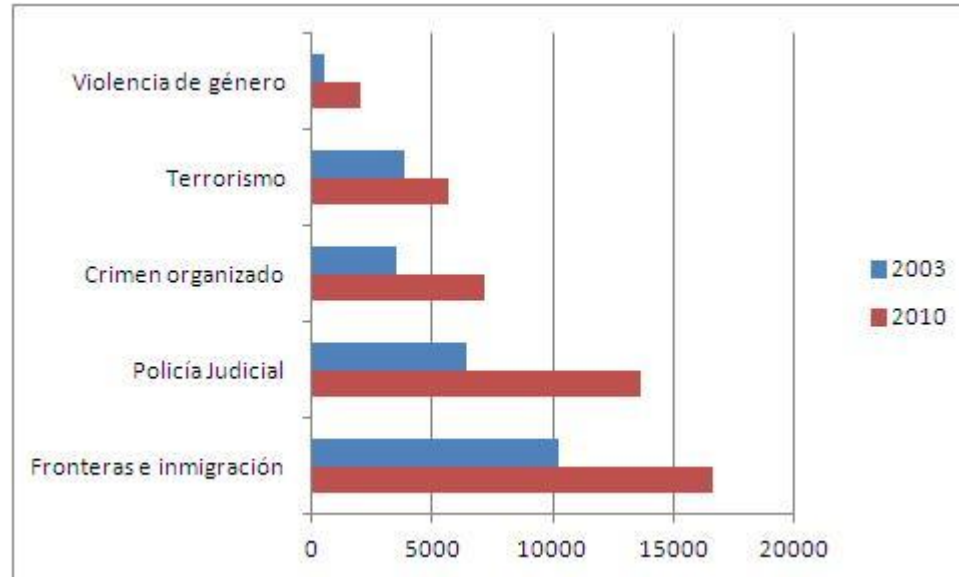


En el gráfico anterior, se ha cortado el eje vertical para dar una impresión engañosa. Ahora no se ve tanta diferencia entre los candidatos.





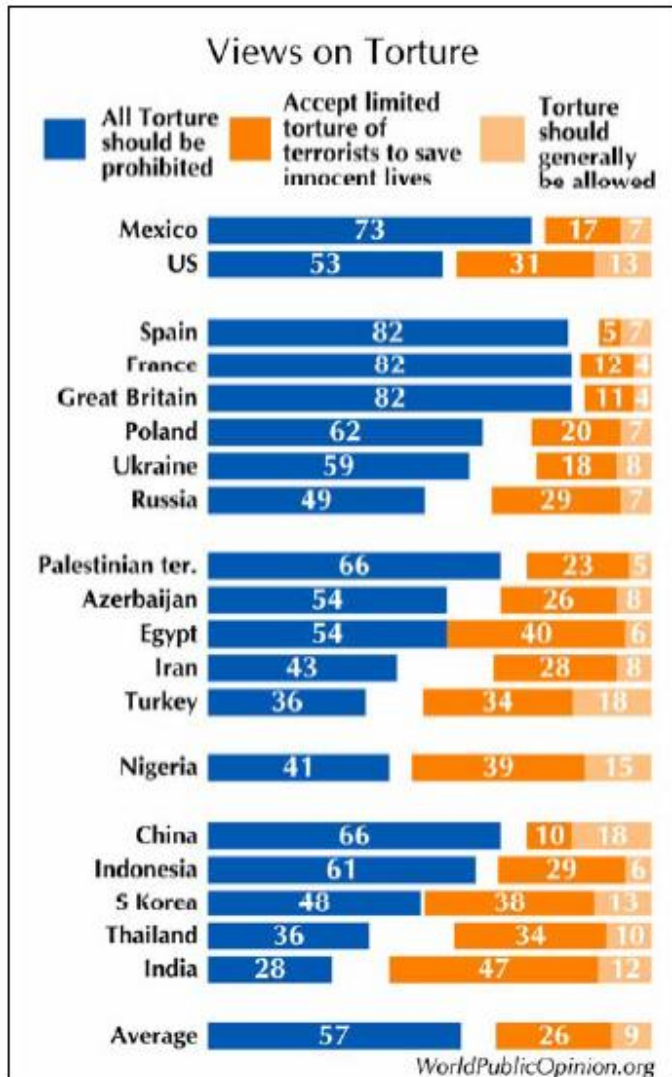
## Diagramas de barras comparativas



La tabla muestra los agentes de policía y Guardia Civil disponibles en varios sectores en los años 2003 y 2012. Se ve el crecimiento en todos los sectores.



## Diagramas de barras apiladas



El diagrama muestra las opiniones sobre la tortura en varios países en el año 2008.

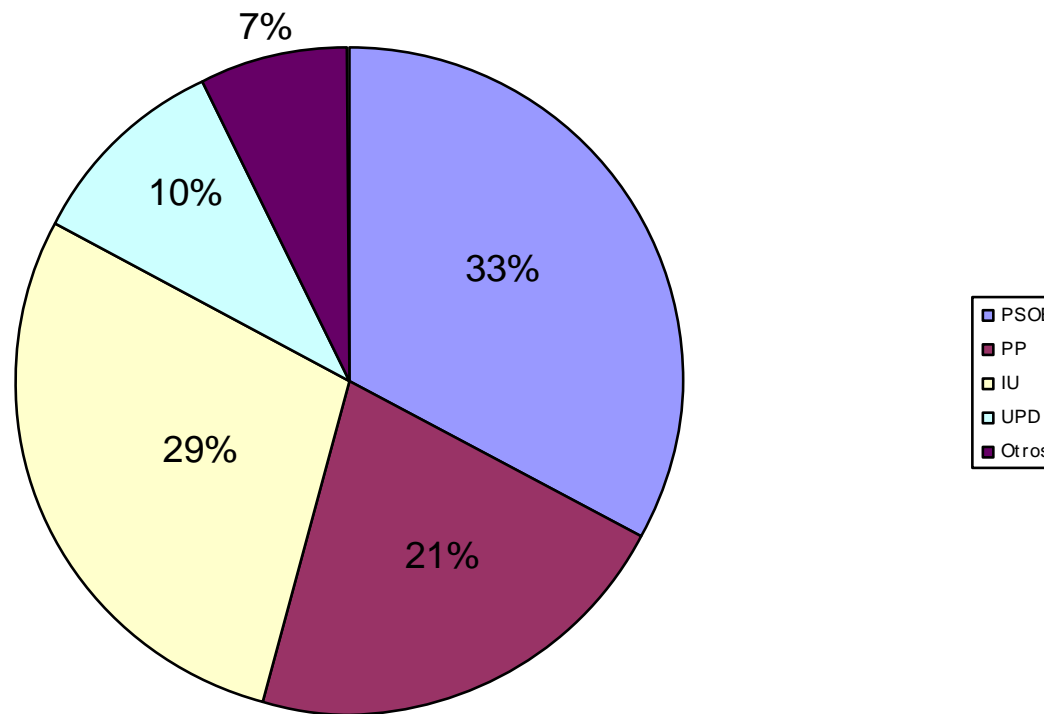
¿Qué información falta?

¿Se puede hacer con frecuencias absolutas?



## El diagrama de sectores

Partido político preferido

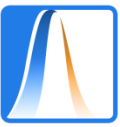




## Mintiendo con el diagrama de sectores



¡Sin comentarios!



## El pictograma



PSOE



PP



IU



UPD



OTROS

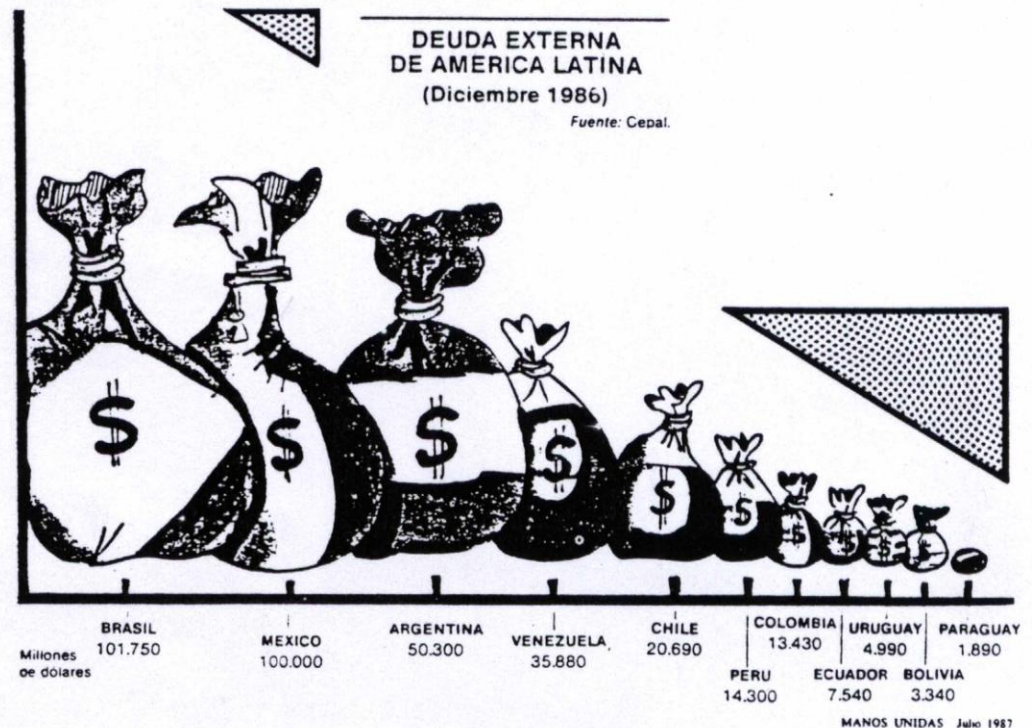
El **área** del gráfico es proporcional a la frecuencia.



## Ejemplo real de un pictograma

Según fuentes de CEPAL, la deuda externa de América latina en diciembre de 1986 viene dada por la tabla de la izquierda. A su derecha reproducimos el pictograma publicado en la revista *Manos Unidas* en julio de 1987.

Países	Millones de dólares
Brasil.....	101 750
México.....	100 000
Argentina.....	50 300
Venezuela.....	35 880
Chile.....	20 690
Perú.....	14 300
Colombia.....	13 430
Ecuador.....	7 540
Uruguay.....	4 990
Bolivia.....	3 340
Paraguay.....	1 890

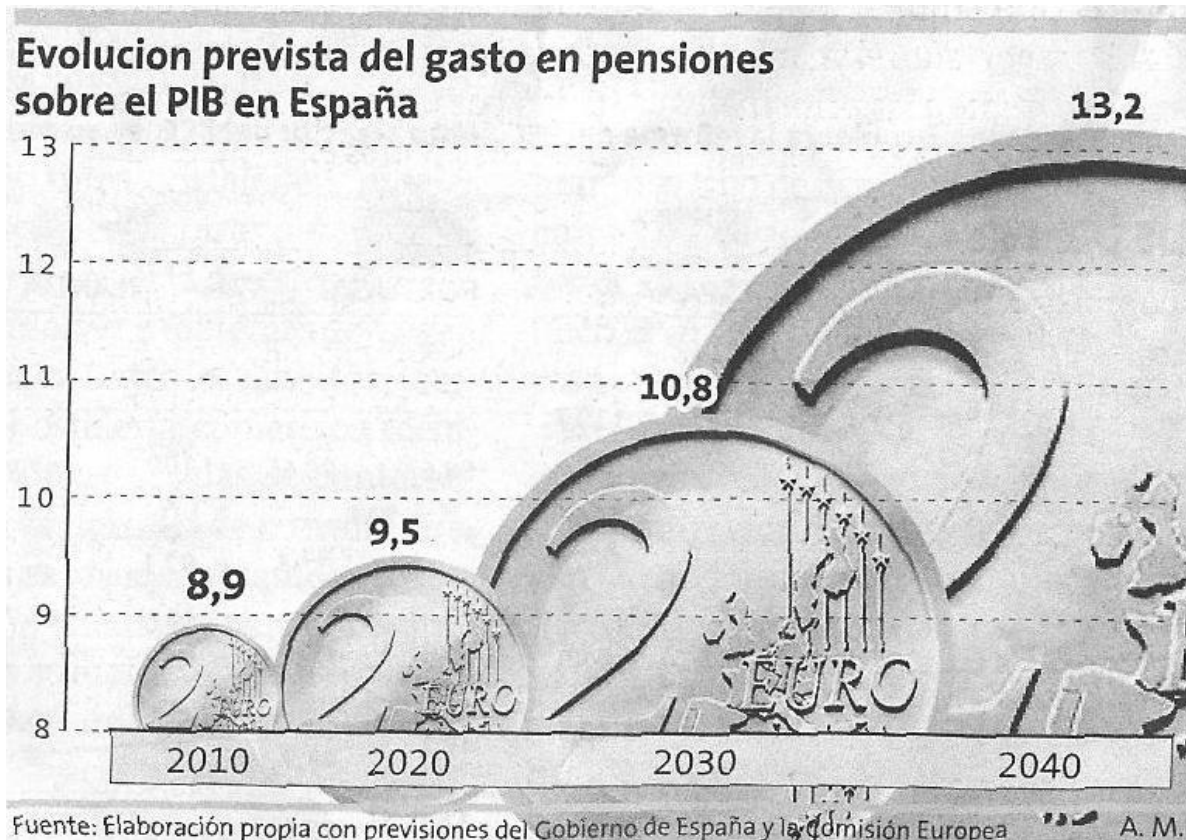






## Engañando con pictogramas

El siguiente gráfico fue publicado en La Voz de Galicia del 24 de octubre de 2010.



Haciendo que la altura es proporcional a la frecuencia da una impresión falsa.

¿Hay algo más que os parece mal?



## Ejercicio

<u>TIPO DE DELITO</u>	Nº
Hurto	5996
Agresión con lesiones	5059
Daños y perjuicios	3866
Narcóticos	3134
Robo	2289
Agresión	1641
Otro	1802

Delitos cometidos en Chicago entre 5/11/2009 y 5/12/2009

[Resumir estos datos](#)





## Ejercicio

Resultados de las notas finales de un curso de Estadística:

<u>CLASE</u>	Nº Alumnos
No presentado	17
Suspenso	16
Aprobado	29
Notable	24
Sobresaliente	2

¿Esta variable es nominal o ordinal?

¿Qué porcentaje de los alumnos no han aprobado?

¿Existe otro gráfico que puede ser útil?



## Ejercicio (Examen)

Los 40 estudiantes de un grupo de Estadísticas evalúan su profesor desde 1 (aburridísimo) a 5 (fantástico). La tabla muestra los resultados parciales.

Evaluación	Frecuencia absoluta	Frecuencia relativa
1		0,05
2		
3	5	
4	9	
5	19	
TOTAL		

Completar la tabla.



## Ejercicio (Examen)

**Pregunta 1.A)** En el siguiente recuadro se presentan los resultados (porcentajes) de la pregunta 28a del Barómetro de abril de 2010 publicado por el Centro de Investigaciones Sociológicas:

<b>CIS</b>	
<b>Estudio nº 2.834</b>	<b>Barómetro de abril</b>
<b>Abril 2010</b>	
<b>Pregunta 28a</b>	
¿Con qué frecuencia se conecta Ud. a Internet?	
Todos o casi todos los días	64.9
De tres a cinco días por semana	13.3
Uno o dos días por semana	11.6
Algunas veces al mes	4.8
Con menor frecuencia o de manera ocasional	2.2
N.C.	3.2
(N)	(1449)

¿Qué proporción aproximada de encuestados se conecta uno o más días por semana?

- (a) 11,6%
- (b) 89,8%
- (c) 21,8%
- (d) Ninguna de las anteriores.



## Ejercicio (Test)

En una encuesta sobre hábitos de salud se les pregunta a 30 alumnos de la universidad seleccionados al azar por el dinero que, en media, cada fin de semana gastan en consumo de alcohol obteniéndose los siguientes resultados:

¿Qué proporción de alumnos gastan en alcohol 20 o más euros?

- a) 20%
- b) 40%
- c) 80%
- d) Ninguna de las anteriores.

Gasto en alcohol (€)	$n_i$	$f_i$
[0-10)	9	
[10-20)	9	
[20-30)	6	
[30-40)	3	0.1
[40-50]	3	0.1
Total	30	1



## Ejercicio (Test)

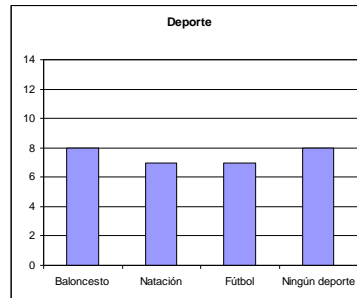
En una encuesta sobre hábitos de salud se les pregunta a 30 alumnos de la universidad, seleccionados al azar, sobre el deporte que practican habitualmente. Los resultados se muestran en la siguiente tabla:

Deporte	$n_i$	$f_i$
Baloncesto	12	0.4
Natación	3	0.1
Fútbol	9	0.3
Ningún deporte	6	0.2
<b>Total</b>	<b>30</b>	<b>1</b>

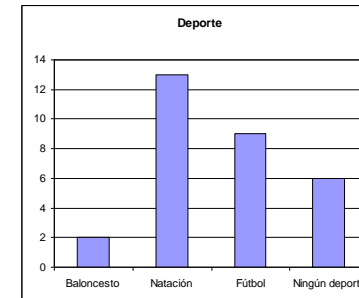
Señala cuál de los siguientes diagramas de barras correspondería a estos datos:



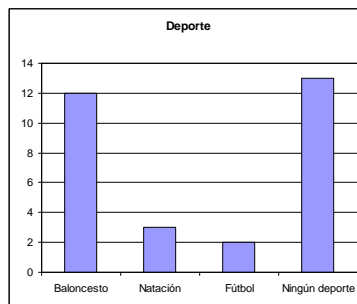
a)



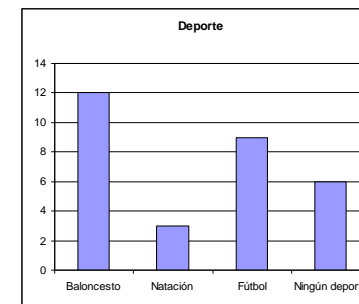
c)



b)



d)





## Ejercicio (Examen)

Según la Encuesta de Población Activa (EPA) en 2008 el número total de activos se sitúa en 22,8 millones de personas. Basándonos en el siguiente diagrama de sectores donde se recoge el reparto de la población activa por sector económico para 2008:

¿Cuántos millones de personas de la población activa dedican su actividad a los sectores económicos de Industria o Construcción?

- a) 28,0
- b) 3,648
- c) 6,384
- d) 2,736

