

Utilidad del análisis exploratorio de datos e importancia de la gráfica



Flor Acevedo De la Rivera
Javier Arredondo Carrasco
Matias Bolton Muñoz
Natalia Ortiz Valdés



Introducción

- Poco protagonismo en la enseñanza secundaria, se enseña como un contenido.
- Se enseña persiguiendo sólo objetivos de conocimientos: Estadística descriptiva.
- Rechazo de los estudiantes a la estadística.

- Crear actitudes positivas hacia la materia.
- La estadística como herramienta para conocer la realidad de forma objetiva.
- Desarrollar clases en torno a casos prácticos que atraigan la atención de los estudiantes.
- Problemas de interés.



Casos analizados



Caso de la panadería

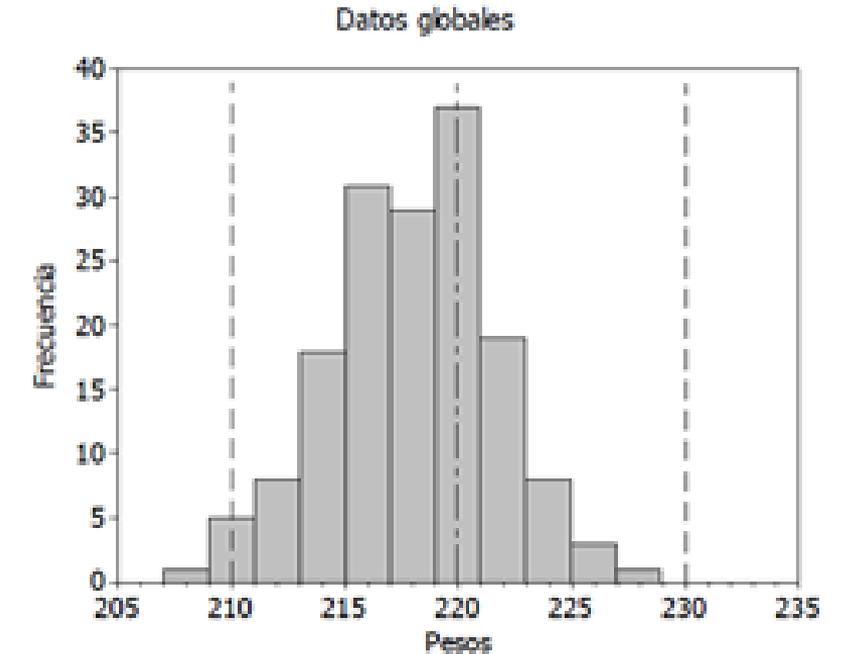


- Peso del pan producido es variable.
- El peso debe ser de 220 ± 10 g
- ¿Existe algún problema?
¿Qué está pasando?
¿Qué hay que hacer para resolver el problema, si es que existe?.

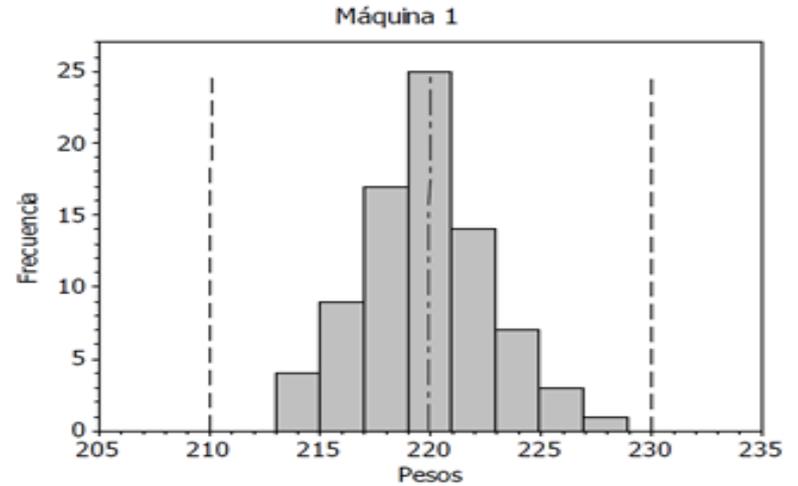
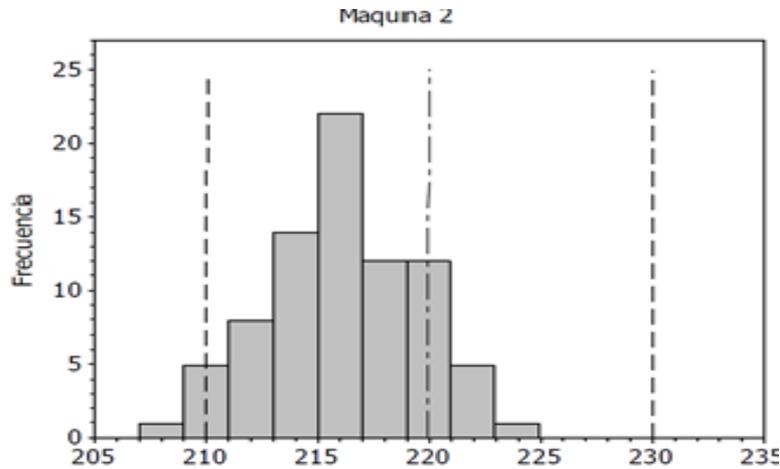
Día	Operario	Máquina 1				Máquina 2			
1	A	220.3	215.5	219.1	219.2	220.3	208.0	214.4	219.2
2	B	215.8	222.0	218.9	213.6	216.9	213.4	217.7	217.7
3	B	220.4	218.7	218.6	219.6	222.9	219.7	209.4	221.6
4	B	221.5	227.0	219.5	222.5	223.1	215.3	220.4	215.6
5	A	215.7	225.3	223.0	218.0	216.0	210.9	221.4	210.9
6	A	222.7	215.1	219.6	217.3	212.1	213.0	218.0	216.5
7	A	216.0	218.8	217.9	213.0	216.9	216.0	213.5	219.2
8	B	219.4	218.3	216.7	224.1	216.2	218.4	216.6	214.9
9	B	219.8	222.6	219.1	217.7	216.2	212.2	216.9	214.9
10	A	220.2	219.5	222.4	219.9	222.9	214.3	219.1	216.7
11	B	218.0	223.9	219.6	221.9	214.9	212.6	219.4	213.3
12	B	219.3	219.6	218.8	219.9	219.0	216.7	216.4	213.5
13	B	220.0	214.1	224.3	217.4	218.0	219.5	219.5	222.3
14	A	223.9	220.6	219.5	219.6	211.8	218.2	218.3	217.4
15	A	218.1	218.8	218.4	217.9	214.6	215.7	218.0	216.4
16	B	216.9	221.6	220.6	222.6	215.6	220.4	217.3	216.2
17	B	217.9	225.7	222.2	216.1	212.5	214.6	209.7	211.3
18	A	224.2	216.2	219.9	220.4	215.8	219.9	216.5	211.9
19	A	214.1	219.7	222.4	224.5	213.7	209.7	216.9	213.1
20	A	221.1	225.0	222.7	222.2	212.5	217.5	217.4	215.7



La solución aparece con las gráficas



La solución aparece con las gráficas



¿Qué podemos aprender de este caso?

- Los datos en una tabla pueden llegar a ocultar información a simple vista.
- Representar los datos facilita su interpretación.

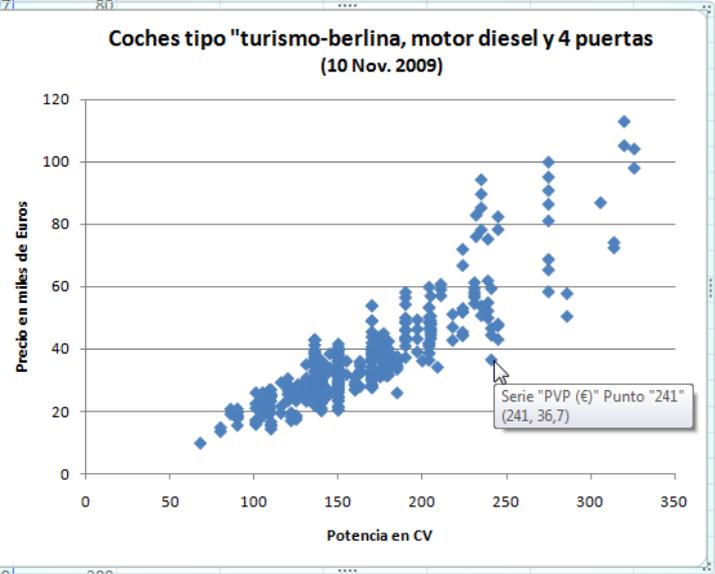
¿Qué autos resultan baratos y Cuáles resultan caros respecto a su potencia?

- Actualmente existe un interés en el mundo de la tecnología
- Resulta fácil obtener datos sobre estos temas a través de internet.
- Buscando coches nuevos tipo “turismo-berlina” con motor diesel y 4 puertas se obtienen los datos de 449 coches cuyo análisis con una hoja de cálculo permite responder a la pregunta planteada.

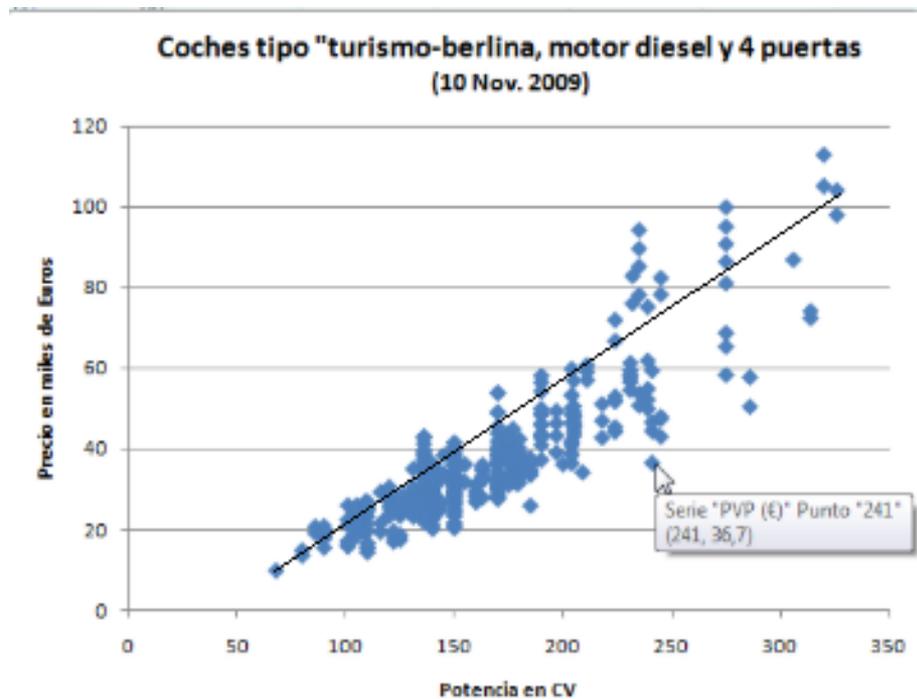


Nuevamente la elección correcta del gráfico es la solución

	A	B	C	D	E	F	G	H	I	J
1	Marca - Modelo	CV	PVP (€)							
2	DACIA LOGAN 1.5 dCi Ambiance 70	68	10	50						
3	SEAT CORDOBA 1.4 TDI 80cv Reference 4P	80	13,742	60						
4	HYUNDAI ACCENT 1.5 VGT 4P 4P	110	14,55	70						
5	SEAT CORDOBA 1.4 TDI 80cv Sportrider 4P	80	14,857	80						
6	HYUNDAI ACCENT 1.5 VGT AA 4P 4P	110	15							
7	FIAT LINEA 1.3 16v Active 90cv Multijet 4P	90	15							
8	KIA RIO 1.5 CRDi Emotion 4p 4P	110	15,9							
9	SEAT CORDOBA 1.9 TDI 100cv Reference Aire 4P	101	15,9							
10	SEAT CORDOBA 1.9 TDI 100cv Sportrider 4P	101	16,2							
11	HYUNDAI ACCENT 1.5 VGT FULL 4P 4P	110	16							
12	SEAT CORDOBA 1.9 TDI 100cv Sport 4P	101	17							
13	CHEVROLET NUBIRA 2.0 TCDi SX 16v (MY09 stock) 4P	122	17							
14	SEAT CORDOBA 1.9 TDI 100cv Stylance 4P	101	17,3							
15	CHEVROLET CRUZE 2.0 VCDi 16v LS 4P	125	17,3							
16	FIAT LINEA 1.3 16v Dynamic 90cv Multijet 4P	90	17,3							
17	TOYOTA COROLLA 1.4 D-4D Live 4P	90	18							
18	CHEVROLET CRUZE 2.0 VCDi 16v LS+ A/A 4P	125	18							
19	CHEVROLET NUBIRA 2.0 TCDi CDX 16v (MY09 stock) 4P	122	18							
20	FIAT LINEA 1.6 16v Dynamic 105cv Multijet 4P	105	18							
21	FORD FOCUS 1.6 TDCi Trend 90cv 4p 4P	90	18,5							
22	CHEVROLET CRUZE 2.0 VCDi 16v LS+ Clima 4P	125	18,5							
23	FIAT LINEA 1.3 16v Emotion 90cv Multijet 4P	90	18,5							
24	RENAULT MEGANE Sedan Emotion-2008 1.5 dCi 85cv Eco2 4P	86	18,5							
25	FORD FOCUS 1.6 TDCi Trend 109cv 4p 4P	109	19,3							
26	MAZDA 3 1.6 CRTD Pulse 109cv 4p 4P	109	19,3							
27	FIAT LINEA 1.6 16v Emotion 105cv Multijet 4P	105	19,59	300						
28	FORD FOCUS 1.8 TDCi Trend 4p 4P	116	19,595	310						
29	CHEVROLET NUBIRA 2.0 TCDi CDX 16v (MY09 stock) 4P	122	19,75	320						



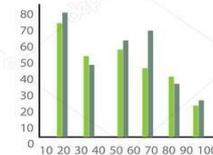
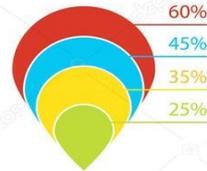
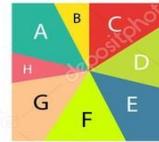
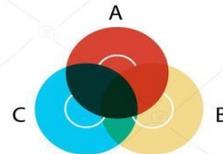
Nuevamente la elección correcta del gráfico es la solución



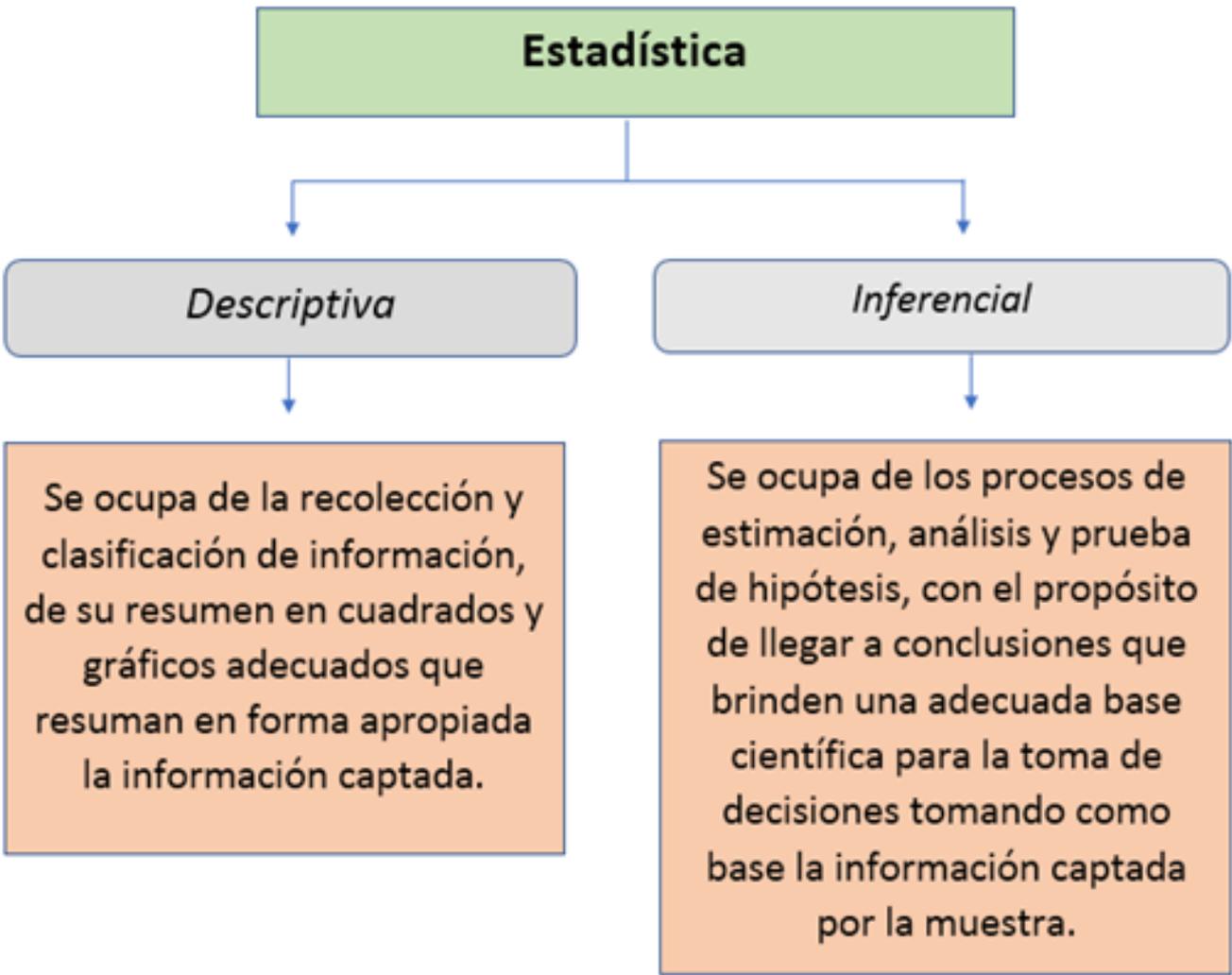
¿Qué podemos aprender de este caso?

- Utilizar hojas de cálculo para construir gráficas.
- Formas de analizar gráficos de datos.

Similitudes de ambos casos.



Estadística



```
graph TD; A[Estadística] --> B[Descriptiva]; A --> C[Inferencial]; B --> D[Se ocupa de la recolección y clasificación de información, de su resumen en cuadros y gráficos adecuados que resuman en forma apropiada la información captada.]; C --> E[Se ocupa de los procesos de estimación, análisis y prueba de hipótesis, con el propósito de llegar a conclusiones que brinden una adecuada base científica para la toma de decisiones tomando como base la información captada por la muestra.];
```

Descriptiva

Se ocupa de la recolección y clasificación de información, de su resumen en cuadros y gráficos adecuados que resuman en forma apropiada la información captada.

Inferencial

Se ocupa de los procesos de estimación, análisis y prueba de hipótesis, con el propósito de llegar a conclusiones que brinden una adecuada base científica para la toma de decisiones tomando como base la información captada por la muestra.



Análisis exploratorio de datos o Estadística descriptiva

Examinamos los datos en forma descriptiva con el fin de:

- ★ Organizar la información.
- ★ Sintetizar la información.
- ★ Ver sus características más relevantes.
- ★ Presentar la información.

Factores necesarios para un buen análisis estadístico:

- Diseño del experimento o investigación.
- Calidad de datos.
- Representaciones acordes al objetivo.





El análisis con este enfoque descriptivo se puede realizar mediante:

- Tablas de frecuencia.
- Gráficos descriptivos: histogramas, sectores, diagramas de dispersión, etc.
- Cálculo de estimadores puntuales de parámetros: porcentajes, tasas, media, desviación estándar, mediana, percentiles...etc.

Tipos de Gráficos

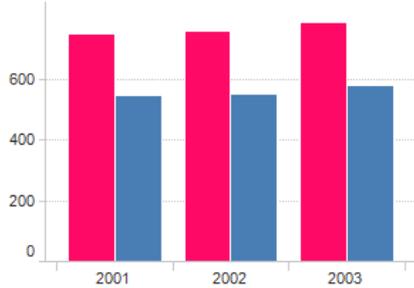
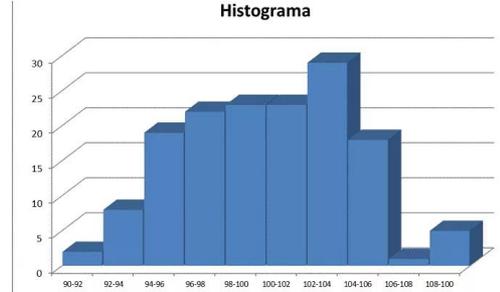


Gráfico de barras



Histograma

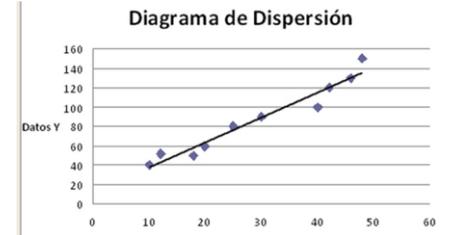


Gráfico de Dispersión



Gráfico circular

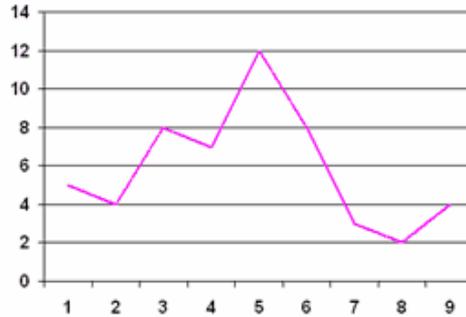


Gráfico lineal

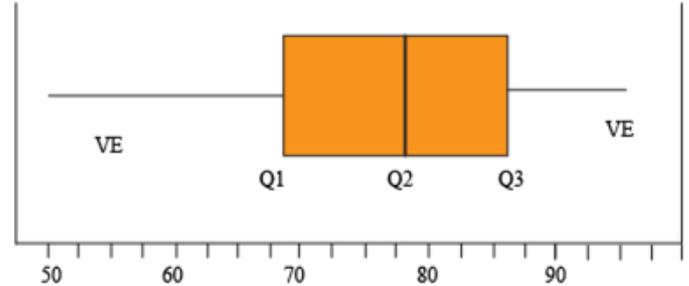
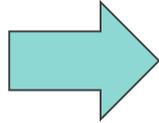


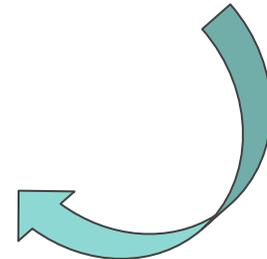
Gráfico de caja y bigote

Ejemplo de estadística descriptiva en el aula

- 12 alumnos dijeron: Los pitufos
- 16 alumnos dijeron: Thor
- 10 alumnos dijeron: Linterna verde
- 6 alumnos dijeron: Crepúsculo



Películas	Alumnos del curso
Los Pitufos	12
Thor	16
Linterna Verde	10
Crepúsculo	6





Cómo elegir el mejor gráfico

A la hora de elegir un gráfico que represente los datos, es necesario identificar el tipo de mensaje que queremos transmitir.

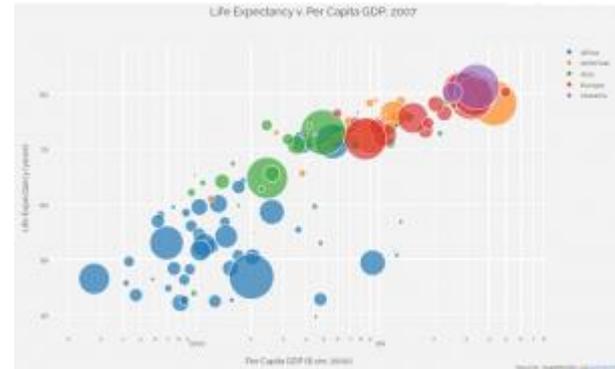
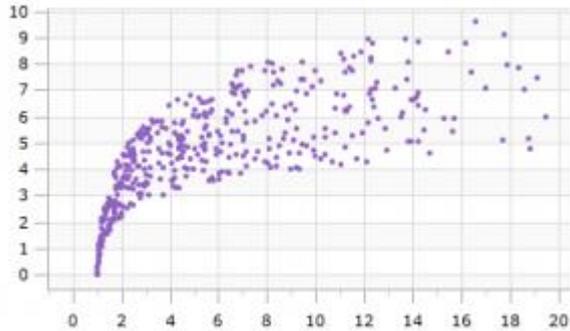
Existen cuatro tipos de mensajes que podemos transmitir con datos:

- Una relación.
- Una distribución.
- Una comparación.
- Una composición.



Relación

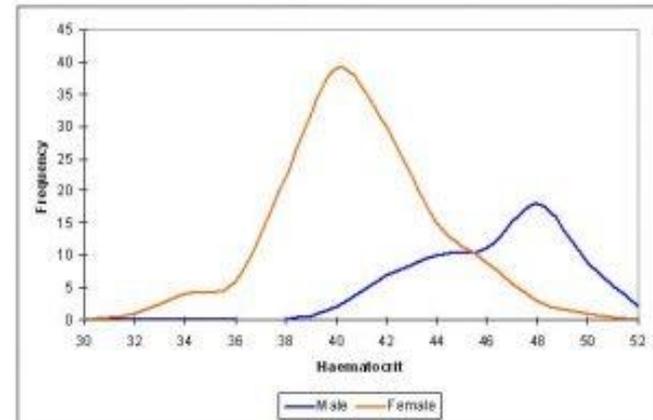
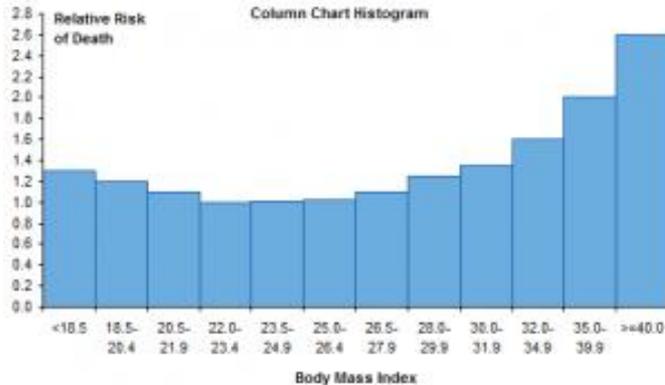
Muestra una conexión entre una o varias variables que se han analizado.





Distribución

Muestra una colección de datos para comprobar su comportamiento.



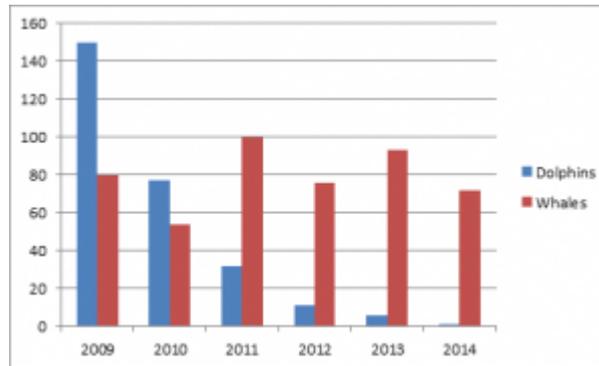
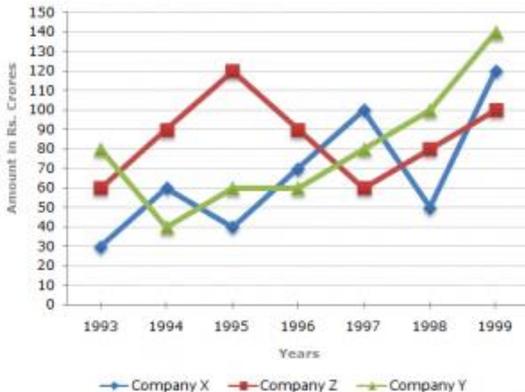


Comparación

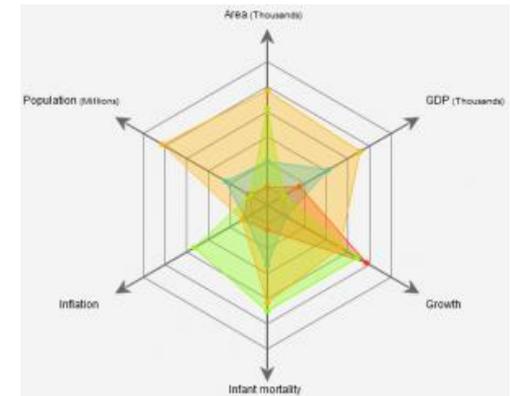
Muestra cómo una serie de variables reaccionan frente a una variable común.

Aquí debemos hacer hincapié si queremos realizar una comparación a lo largo del tiempo o en un momento puntual.

A lo largo del tiempo



En un momento definido



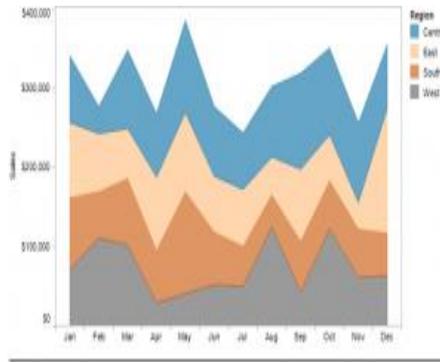
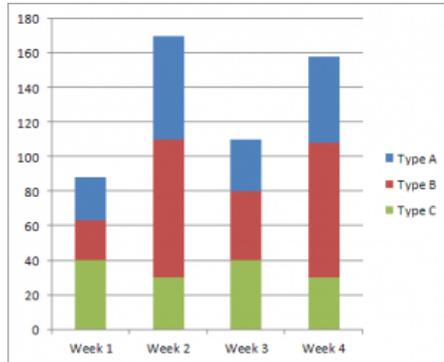


Composición

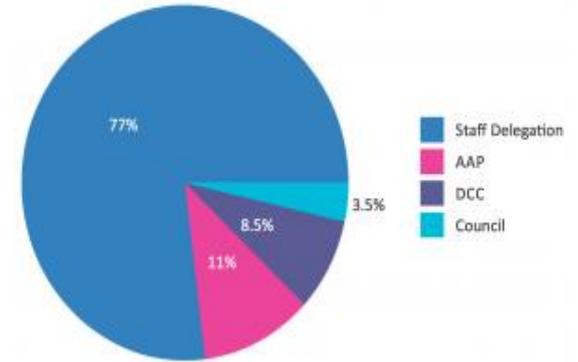
Recoge una muestra de datos y revela cómo es la composición de la muestra según variables definidas.

Al igual que en caso anterior, se hace la distinción entre sí se estudia a lo largo del tiempo o un momento preciso.

A lo largo del tiempo



En un momento definido





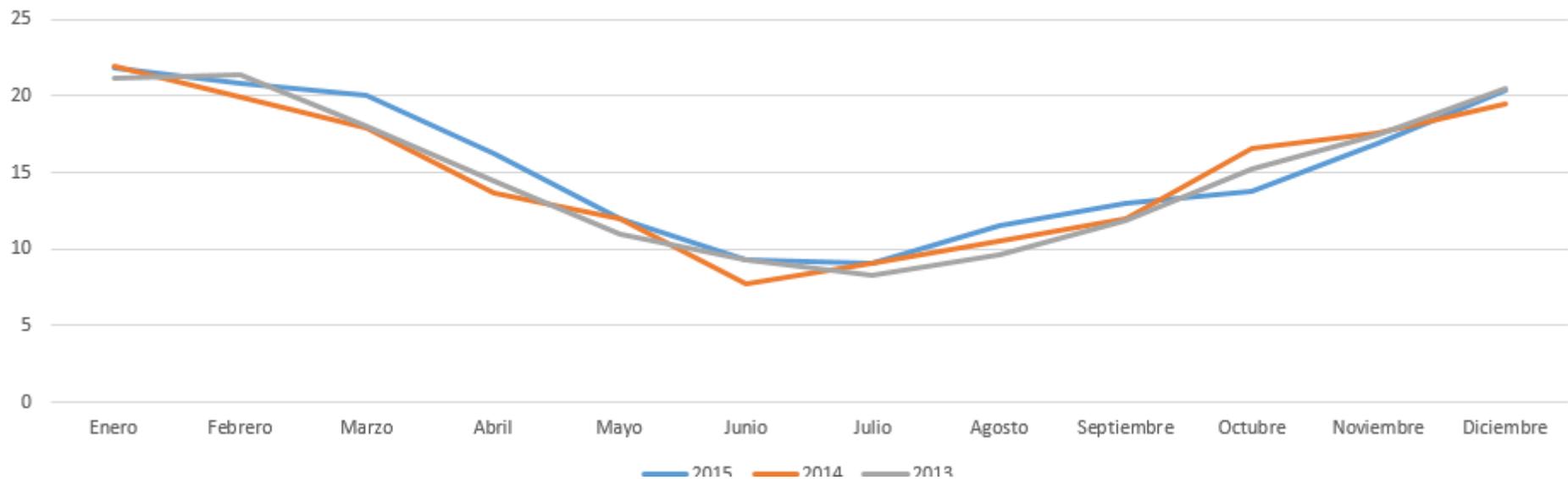
Actividad:

Uno grupo de meteorólogos necesita determinar cuál de los tres años entre 2015 y 2013 fue el más helado y cuál fue el más caluroso. A su disposición tienen las temperaturas medias de cada mes desde 2013 hasta 2015.

Con la información de la tabla que se le entregará, realice un gráfico en su cuaderno que resuelva el problema de los meteorólogos.

Se recomienda utilizar colores distintos.

Temperaturas medias durante los años 2015, 2014 y 2013





Conclusión

- Muchas veces los datos en bruto ocultan información y que al representarlas de manera adecuada, podemos concluir ideas de manera más rápida y simple.
- Permite a las personas no especializadas en el tema, puedan interpretar mejor la información.
- La estadística descriptiva por lo general, es más cercana a lo que se enseña en los colegios.
- La elección del gráfico adecuado es esencial para comunicar una idea.



Bibliografía

- Maria Tatay. (2016). Cómo elegir la mejor gráfica para mostrar los datos. 17 de Julio de 2019, de Be Prisma Sitio web: <https://beprisma.com/como-elegir-la-mejor-grafica-mostrar-datos/>