

Tema 10

Visualización de los datos

Técnicas multivariantes

Dr. Antoni Ferragut



- La visualización de datos es parte del análisis descriptivo y aporta un mayor conocimiento de los mismos
- Sirve también para representar los resultados de manera simple
- Veremos los modos de representar los datos y problemas frecuentes
- Finalmente veremos instrucciones en python para realizar estas visualizaciones

- La visualización de los datos forma parte del análisis estadístico descriptivo
- La visualización adecuada proporciona más información que trabajar con los datos en crudo

- En un gráfico podemos ser capaces de condensar mucha información y de mostrar patrones, tendencias y hallazgos de una forma rápida y directa
- Podemos seguir las siguientes pautas:
 - Gráficos sencillos, intuitivos y comprensibles
 - Guías de estilo, coherencia en un mismo documento
 - Gráficos que aporten: más cognición, menos búsqueda de información en el texto

- Se deben presentar los datos limpios, y por tanto tratar antes con:
 - Valores anómalos
 - Valores contradictorios
 - Identificadores no únicos
 - Valores ausentes
- Pasos a realizar:
 - Análisis de información, para detectar estos problemas
 - Verificación de inconsistencias, para corregirlas
 - Almacenamiento de datos limpios (sin eliminar los brutos)

- Unidimensionales: código fuente, listas secuenciales... representables en una línea o eje
- Bidimensionales: dos variables interrelacionadas
- Tridimensionales: objetos con volumen; una variable frente a otras dos... la perspectiva es importante. Ok en pantalla, no tan bien en papel
- Multidimensionales: para la representación necesitaremos aplicar transformaciones, por ejemplo PCA
- Temporales: ordenados cronológicamente pueden ser superpuestos y representados como un dato bidimensional (con la variable tiempo)

- Gráfico de líneas: cada par de valores se une al siguiente mediante una línea. Usado en series temporales
- Gráfico de dispersión (scatterplot): cada par se representa con un punto
- Histograma: representa frecuencias de una variable discreta (o discretizada)
- Diagrama de caja (boxplot): representa estadísticos descriptivos de una distribución
- Diagrama de barras: representa la frecuencia de variables no comparables
- Diagrama de sectores: representa porcentajes de una variable
- La librería Matplotlib de Python contiene todos estos tipos de gráficos, es la más usada. También están Nineplot o Seaborn