

Prueba Final de Visualización de Datos

Duración: 1h 45 minutos.

Desarrolle las respuestas.

1) Introducción

El “cuarteto de Anscombe” consiste en unas tablas con valores en X e Y que tienen igual media, desviación estándar y correlación. ¿La visualización complementa de alguna forma la descripción de los datos? Explique.

2) Modelo de Datos, Variables de la Imagen

- Describa qué son los tipos Nominal, Ordinal y Cuantitativo (Intervalo y Ratio) y mencione un ejemplo de cada tipo de variable.
- Según el “ranking de efectividad” la posición, longitud, ángulo, etc. son códigos visuales utilizados para representar variables Nominales, Ordinales y Cuantitativas. Describa otros códigos visuales que recuerde. ¿Todos los códigos sirven por igual para todos los tipos de variables? Desarrolle.

3) Análisis Exploratorio de Datos

- ¿Conviene eliminar los *outliers* en las visualizaciones?
- ¿Los *outliers* sólo surgen del análisis individual de cada variable o pueden surgir del análisis de dos o más variables? Desarrolle la respuesta.
- Defina y compare los pros y contras del *Box-plot* y del *Violin-plot*, para analizar la distribución de una variable cuantitativa.

4) Códigos Visuales

- Discuta sobre la conveniencia de incluir o no al 0 en los ejes de un gráfico.
- Defina qué son y cuándo usar las escalas de colores secuencial y divergente.
- Describa las reglas de Similitud, Proximidad, Clausura y Conectividad que se desarrollan en la teoría de la Gestalt.

5) Desarrolle

Al principio del curso se dijo que la visualización ayuda a “responder preguntas” y a “apoyar el razonamiento”, pero luego se dijo que “correlación no es inferencia”, que las interpolaciones/extrapolaciones pueden “inventar resultados” y que en definitiva los gráficos pueden inducir al error. En su opinión, ¿cuál es el papel verdadero de la visualización de datos en la ciencia de datos?