



# Análisis exploratorio de datos

Wilmer Gonzalez

6213 - Minería de datos  
Facultad de ciencias  
Universidad Central de Venezuela

25 de junio de 2024

# Contenido

- 1 Introducción al análisis exploratorio
- 2 Algunas técnicas
- 3 Caso práctico en código

*"Notoriamente no hay clasificación del universo que no sea arbitraria y conjetural."*<sup>a</sup>

---

<sup>a</sup>El idioma analítico de John Wilkins - Jorge Luis Borges

# Introducción

# Inicios del análisis exploratorio

## Algunos hitos e ideas necesarias

- ▶ Inferir vs Deducir.
- ▶ *Confirmatory data analysis* y *Exploratory data analysis*.
- ▶ (Tukey et al., 1977); un grupo de investigación de estadística consolida en un mismo marco de trabajo varias técnicas de exploración de datos, argumentando al mismo tiempo que se hacía poco énfasis en el uso de datos para desarrollar teorías y otros aspectos de investigación.
- ▶ 80s - 90s: Las capacidades de cómputo comienzan a facilitar el uso de datos como punto de partida del proceso de investigación.

# Algunos rasgos (Jebb, Parrigon y Woo, 2017)

- ▶ *Openness - Skepticism.*
- ▶ Uso de técnicas de suavizado.
- ▶ Cálculos sencillos.
- ▶ Inspección visual.
- ▶ Entre otros...

**$EDA \approx$  Interacción con un conjunto de datos con el fin de comprender sus estructuras y dinámicas.**

Algunas técnicas

# EDA por inspección visual

Tipos de datos	Técnica
Categoricos	Descriptores estadísticos <sup>1</sup>
Univariable continua	Gráfico de línea, histograma
Conjunto bivariado continuo	2d Gráfico de puntos
Lista 2D	Mapa de calor
Conjunto bivariado continuo	2d Gráfico de puntos

<sup>1</sup>pe: promedio, varianza, desviación estándar, kurtosis, *skewness*, ...

# EDA por objetivos

Objetivo	Técnica
Análisis de distribución	Histograma
Valores anómalos	<i>box and whisker plots</i>
Cuantificar relación entre dos variables	Covarianza/Correlación
Visualización de datos en N-dimensiones ( $N \geq 4$ )	PCA*



Caso práctico en código

# Referencias



Jebb, Andrew T., Scott Parrigon y Sang Eun Woo (2017). “Exploratory data analysis as a foundation of inductive research”. [En.](#)



Tukey, John Wilder et al. (1977). *Exploratory data analysis*.

# ¡Gracias!

[github.com/ucvia/dm](https://github.com/ucvia/dm)