

# Expo: Estadística en el Entorno

## Análisis de Clúster en USArrests

### Participantes:

Cervantes Hernández Zulema Yamileth, Miranda Gabriel Karla Arely y Córdoba Tlaxcalteco María Luisa.

### Introducción

La base de datos para este estudio tiene cuatro grupos categorizados según características específicas. USArrests ofrece detalles sobre la criminalidad en los Estados Unidos en 1973, con estadísticas desglosadas por cada 100,000 residentes, abarcando agresiones, violaciones y asesinatos en los 50 estados. El estudio utiliza técnicas avanzadas de análisis de clúster, como el agrupamiento jerárquico y K-means.

### Objetivo

Explorar y analizar la cantidad de agrupamientos de los Estados en función de sus similitudes en indicadores criminales.



### Metodología

- Selección de la Base de Datos:** Se optó por utilizar una base de datos con una dimensión de 50 x 4, que contiene información relevante sobre crímenes en los Estados Unidos.
- Análisis de Clúster:** El análisis se realizó en el software .
- Métodos:**

#### K-means:

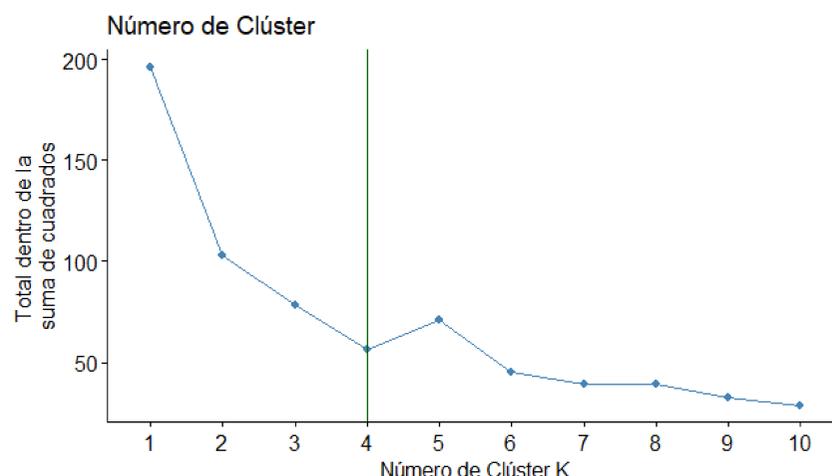
- Técnica para agrupar los Estados con respecto a sus indicadores criminales.

#### Jerárquico:

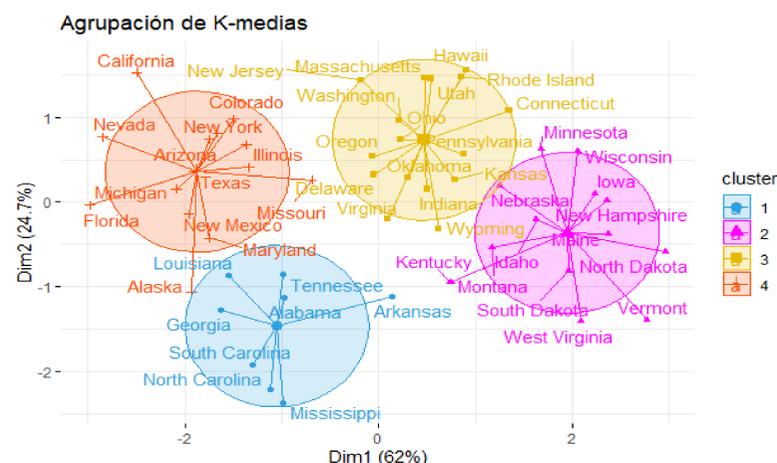
- Se aplicó el método jerárquico utilizando la distancia euclidiana como métrica para evaluar las similitudes entre los Estados.

Tamaño de los clústers	
Clúster 1	8
Clúster 2	13
Clúster 3	16
Clúster 4	13

### Resultados

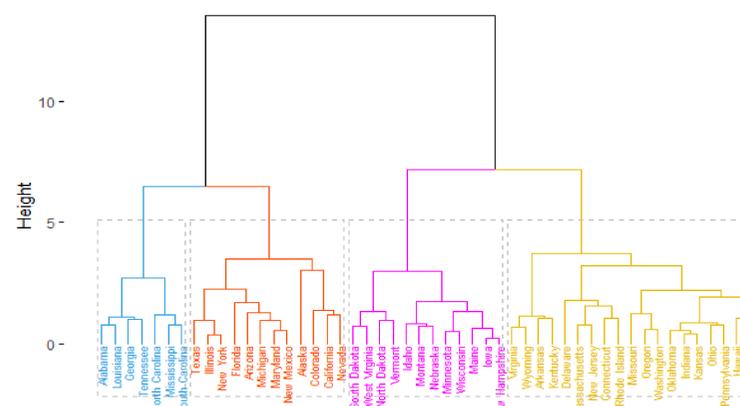


El codo indica el número ideal de clúster.  
K-means = 4



Los Estados se agrupan de acuerdo a sus similitudes, formando cuatro clústers.

#### Cluster Jerárquico



### Discusión

Basándonos en los resultados derivados del análisis de clúster realizado en RStudio, se concluye que se formaron cuatro grupos claramente definidos, cada uno caracterizado por similitudes en los arrestos por crímenes. Un aspecto significativo de este análisis es la notabilidad del clúster tres, que se destaca como el grupo más extenso, sugiriendo que la mayoría de los estados comparten similitudes notables en relación con las variables consideradas en el análisis.

### Referencias

USArrests. (2021). *Tasas de delitos violentos según estadísticas de EE. UU.* Kaggle. <https://www.kaggle.com/datasets/halimedogan/usarrests>  
Peña, D. (2002). *Análisis de Datos Multivariantes*. ResearchGate. [https://www.researchgate.net/publication/40944325\\_Analisis\\_de\\_Datos\\_Multivariantes](https://www.researchgate.net/publication/40944325_Analisis_de_Datos_Multivariantes)