



## II Expo: Estadística en el Entorno

# Niveles de SO<sub>2</sub>: Un Enfoque de K-Means en el Análisis Ambiental y Económico

Participantes:

MARIA FERNANDA ALATRISTE LEAL, LIZETE LANDA PEREA, JOSE DE JESÚS VELASCO SÁNCHEZ, MARIA LUISA CORDOBA TLAXCALTECO.

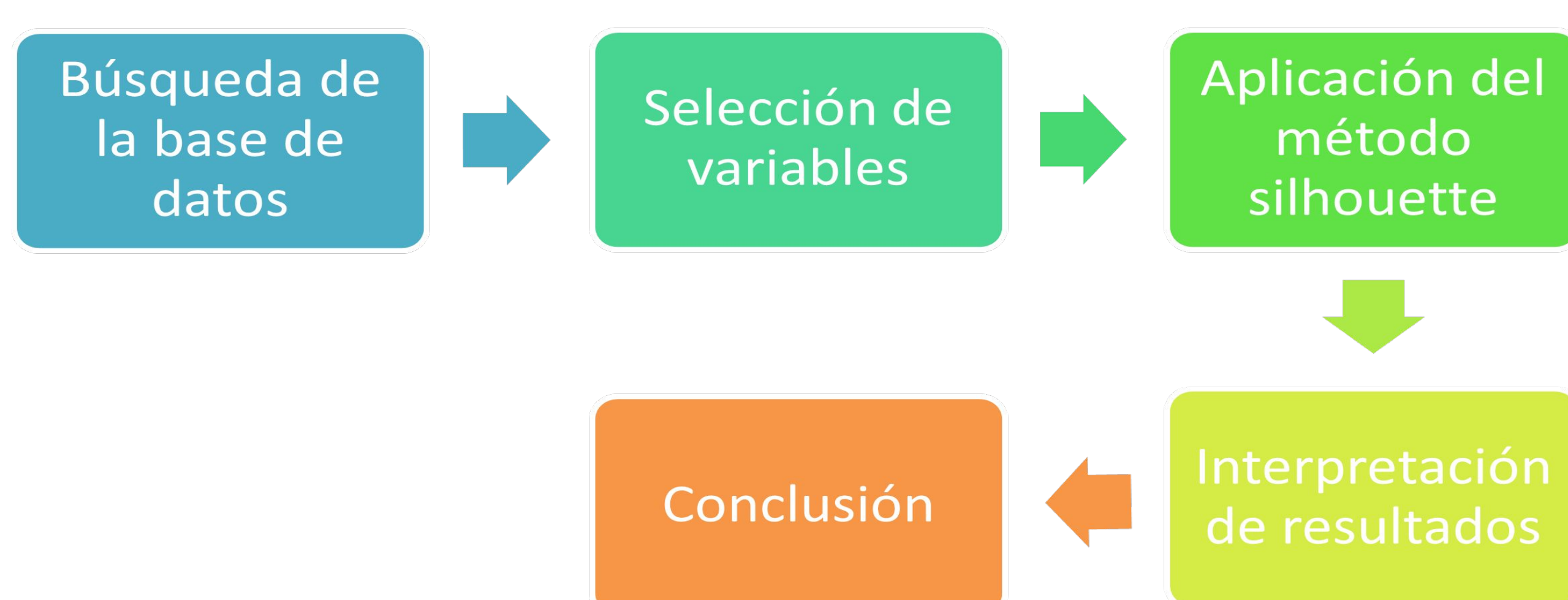
### Introducción

En este estudio, se aplicó la técnica de agrupamiento K-means a un conjunto de datos que abarca diversas variables relacionadas con diferentes ciudades. Estas variables incluyen la concentración de dióxido de azufre (SO<sub>2</sub>), la temperatura promedio, el número de empresas, la población, la velocidad del viento, la precipitación promedio y el número de días. El propósito fue explorar patrones intrínsecos y las estructuras fundamentales en los datos mediante la formación de agrupamientos similares entre ciudades. Además, se utilizó el análisis silhouette para evaluar la coherencia de los grupos identificados por el algoritmo K-means. Este enfoque proporciona una visión integral y estructurada de la variabilidad en las características de las ciudades, facilitando la interpretación de las similitudes y diferencias entre ellas.

### Datos

Los datos analizados incluyen información sobre diferentes ciudades, abordando variables como la concentración de dióxido de azufre (SO<sub>2</sub>), la temperatura promedio, el número de empresas, la población, la velocidad del viento, la precipitación promedio y el número de días. Estos indicadores ofrecen una visión amplia de la calidad del aire, el clima, la actividad económica y la demografía en cada ubicación. Este conjunto de datos puede analizarse para identificar patrones, relaciones y características distintivas entre las ciudades, proporcionando información valiosa sobre factores ambientales y socioeconómicos.

### Metodología



### Discusión

El análisis de agrupamiento K-Means ha permitido identificar patrones similares en las ciudades a través de diversas variables vinculadas con el medio ambiente y la demografía.

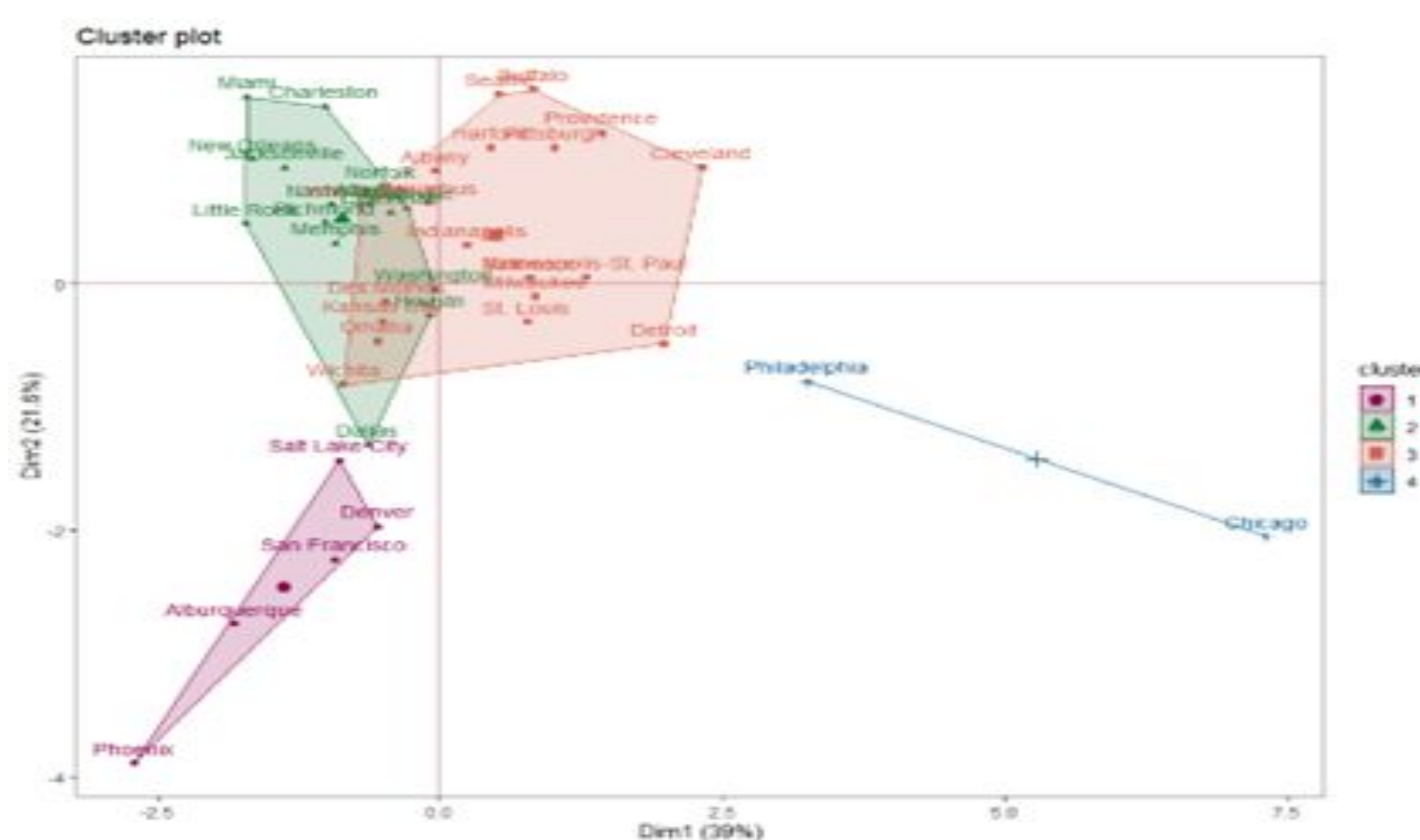
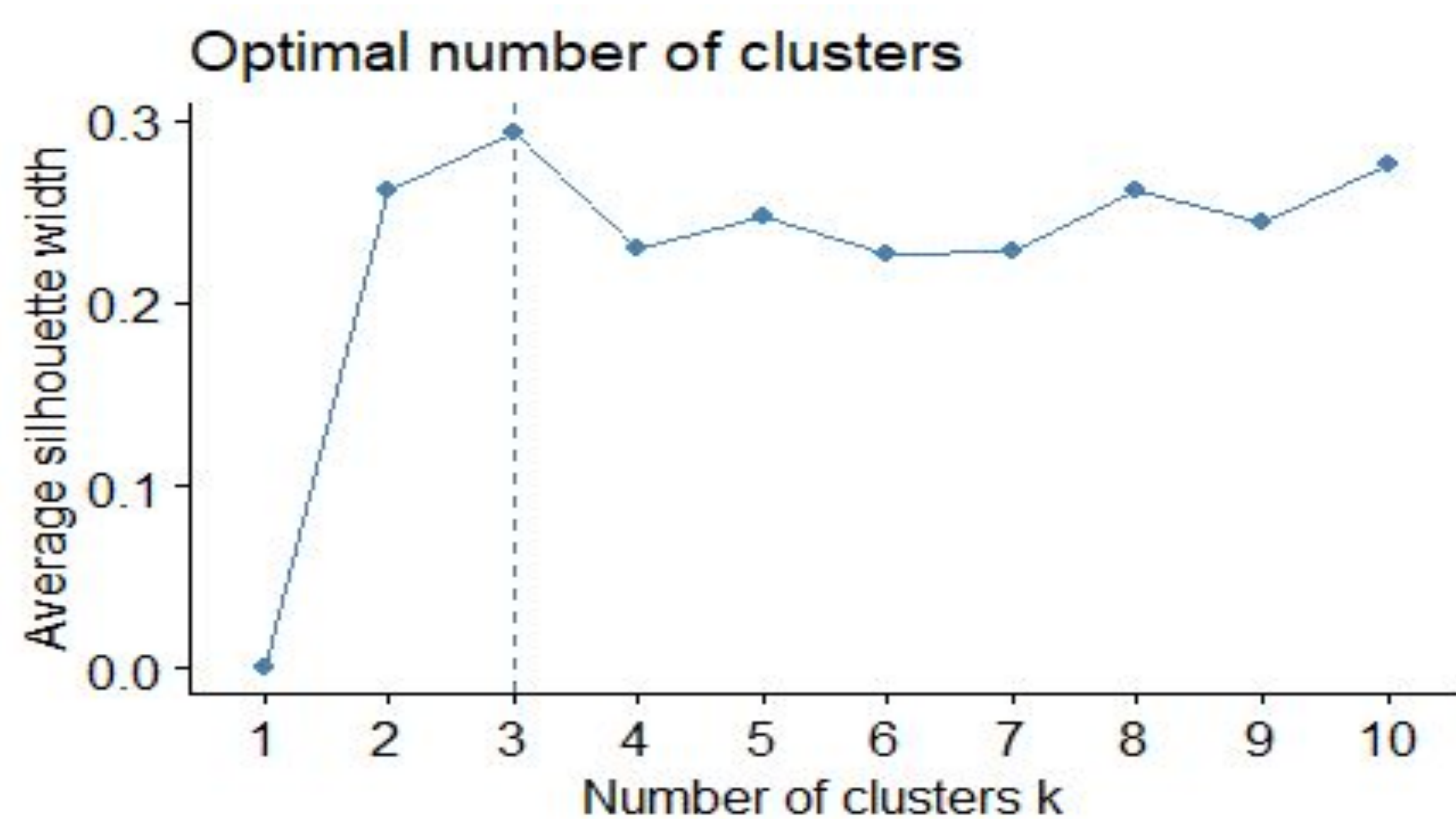
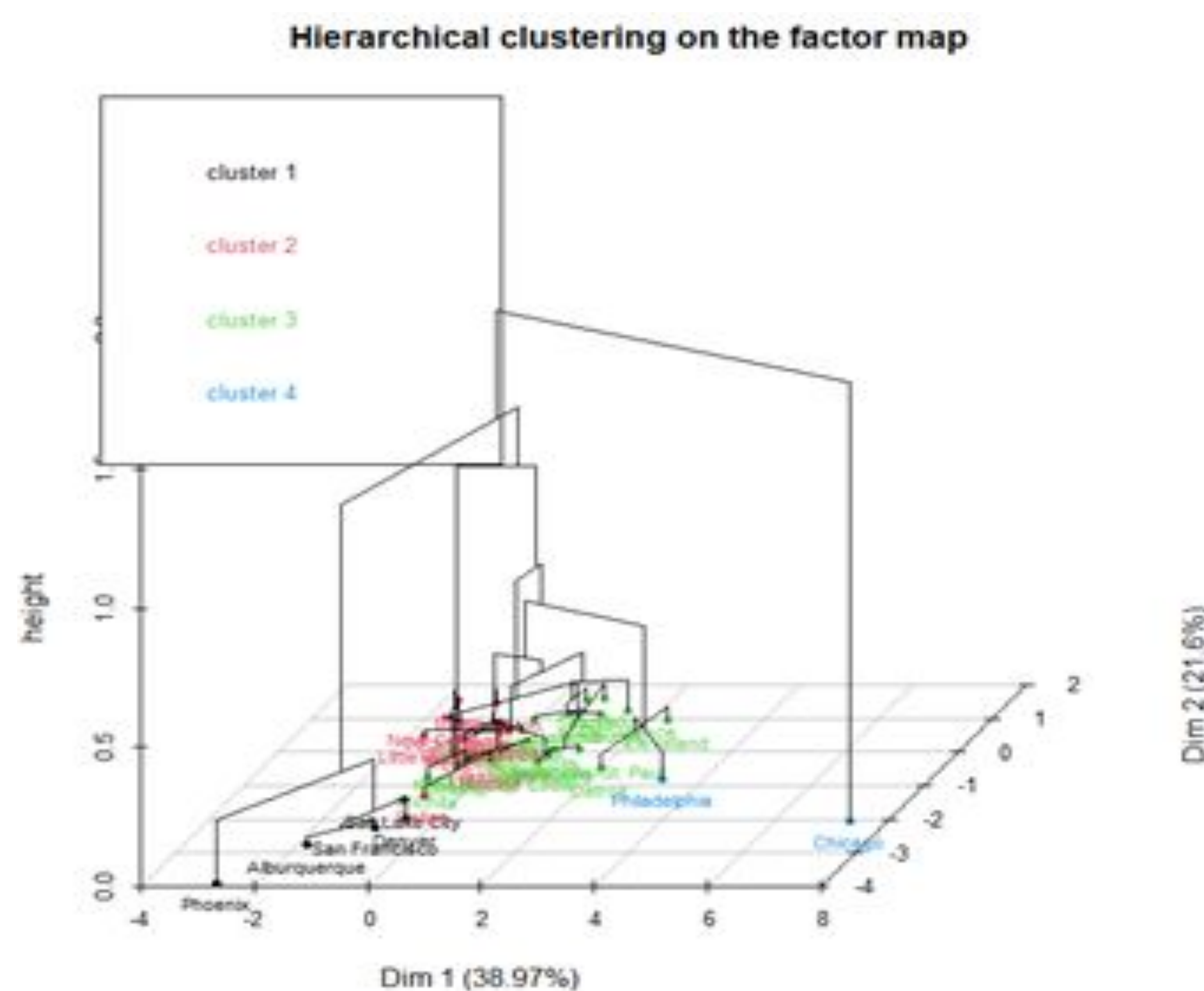
**Clúster 1:** Las ciudades en este clúster exhiben temperaturas más bajas y niveles de dióxido de azufre (SO<sub>2</sub>) por debajo del promedio en comparación con otras áreas urbanas.

**Clúster 2:** Las ciudades en este grupo comparten patrones climáticos caracterizados por altas temperaturas, precipitación significativa y niveles de SO<sub>2</sub> en el rango de 9 a 31.

**Clúster 4:** Se observa la concentración más alta de dióxido de azufre (SO<sub>2</sub>), así como los valores más elevados en cantidad de empresas y población.

### Resultados

Se observó el agrupamiento de las variables las cuales tenían similitud entre así, así como también se obtuvo el número de clústeres óptimos.



### Referencias

- Peña, D. (2002). Análisis multivariante de datos. Mc Graw Hill.
- Kassambara, A. (2018). Machine learning essentials: Practical guide in R. Sthda.
- R Dataset / Package Datasets / USARRests | R Datasets. (s. f.). Recuperado de <https://r-data.pmagunia.com/dataset/r-dataset-package-datasets-usarrests>