

Expo: Estadística en el Entorno

"Agrupamiento Jerárquico y K-Means: Explorando Datos de Árboles de Naranja."

Participantes:

Ana Deneb Espinosa Castillo, Paola Alejandra García Morales, María Luisa Córdoba Tlaxcalteco

Introducción

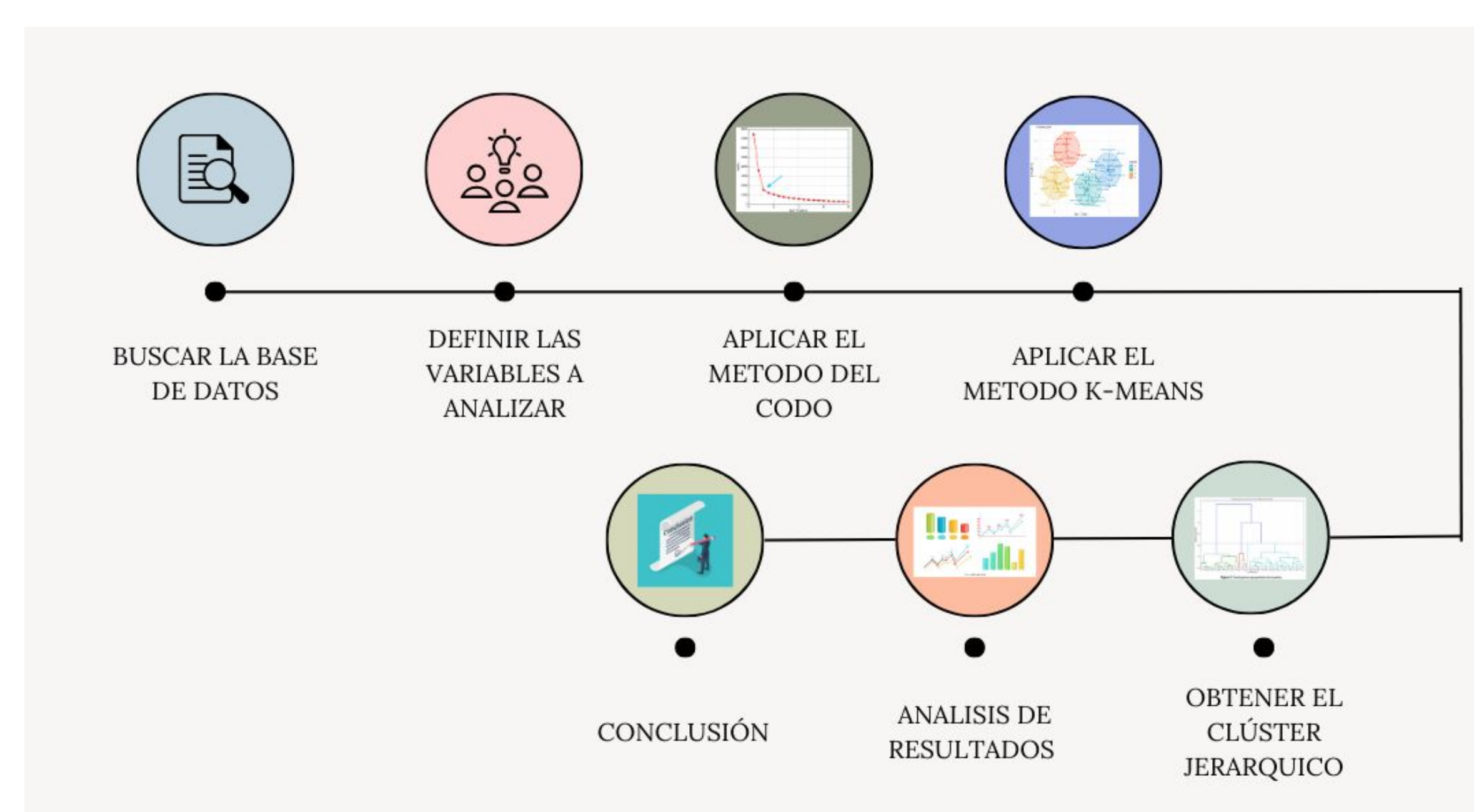
El análisis de clústeres se presenta como un conjunto de herramientas esenciales para la clasificación de individuos en grupos similares. En este trabajo, nos enfocamos en aplicar las técnicas de agrupamiento jerárquico y k-means a la base de datos "Árboles de Naranja", compuesta por 36 registros y tres variables clave: tipo de árbol, año de registro y circunferencia. Nuestra meta es identificar patrones y relaciones dentro de este conjunto de datos, ofreciendo una visión más precisa de su estructura y variabilidad.

Objetivo

El propósito de este estudio es analizar y determinar la relación existente entre la edad de los árboles y sus circunferencias. A través de un enfoque analítico y estadístico, buscamos identificar patrones significativos que arrojen luz sobre la influencia de la edad en el crecimiento de la circunferencia de los árboles en nuestro conjunto de datos. Este análisis contribuirá a una comprensión más detallada de la dinámica de crecimiento de los árboles y puede proporcionar información valiosa para la gestión y cuidado de estas especies.



Metodología



Resultados

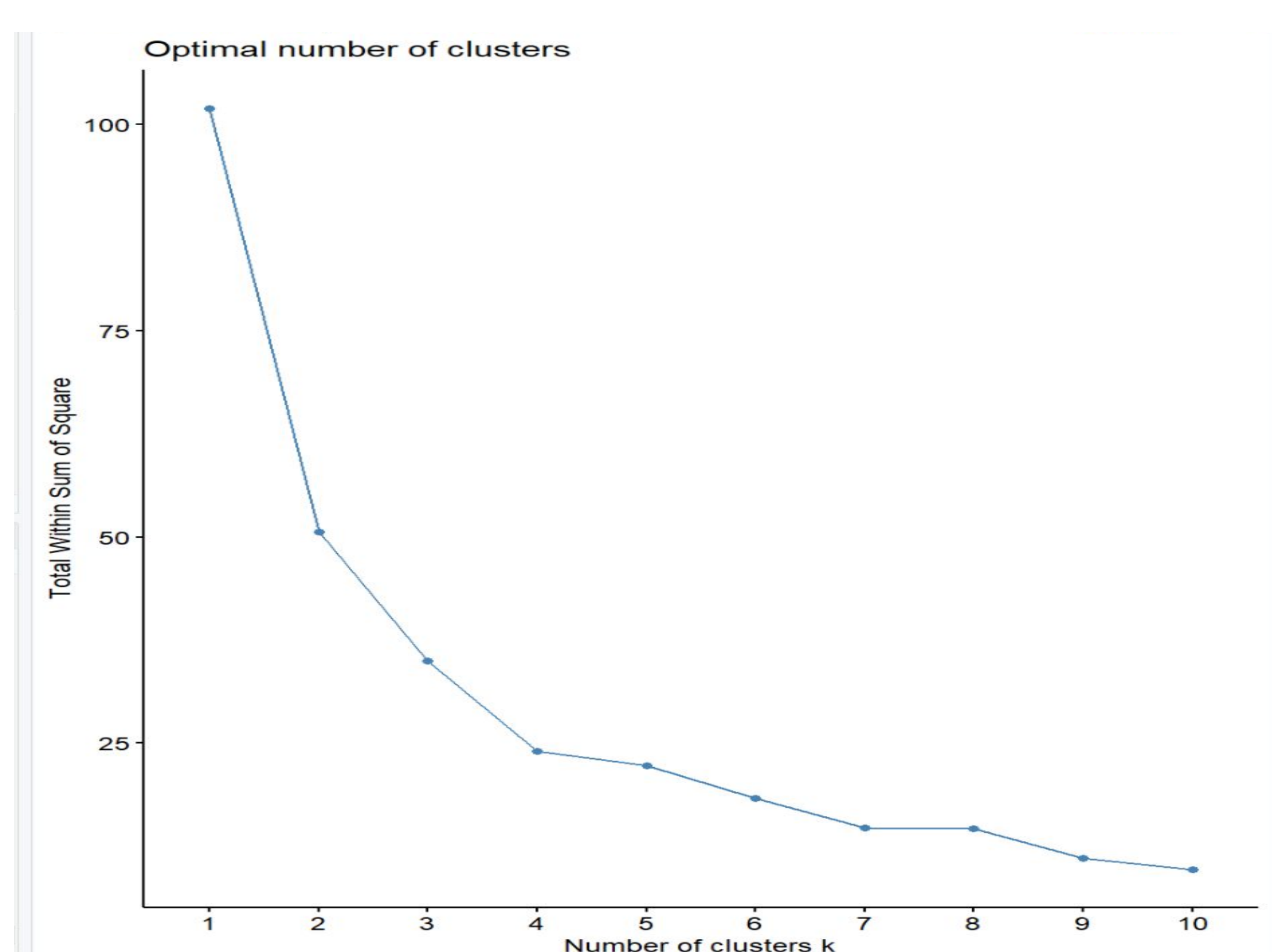


Gráfico 1. Método del codo.

Discusión

Al aplicar el método del codo a nuestro conjunto de datos, se sugiere la presencia de cuatro grupos distintos. En consecuencia, hemos decidido implementar el algoritmo k-means con k=4 para llevar a cabo la agrupación de los árboles de naranja.

La elección se fundamenta en la identificación de un punto de inflexión en la curva del método del codo, lo cual indica que cuatro clústeres son adecuados para capturar la variabilidad presente en la circunferencia de los árboles.

Este enfoque nos permitirá explorar de manera más específica las diferencias en el crecimiento de la circunferencia y contribuirá a una clasificación más precisa de los árboles en nuestro conjunto de datos.

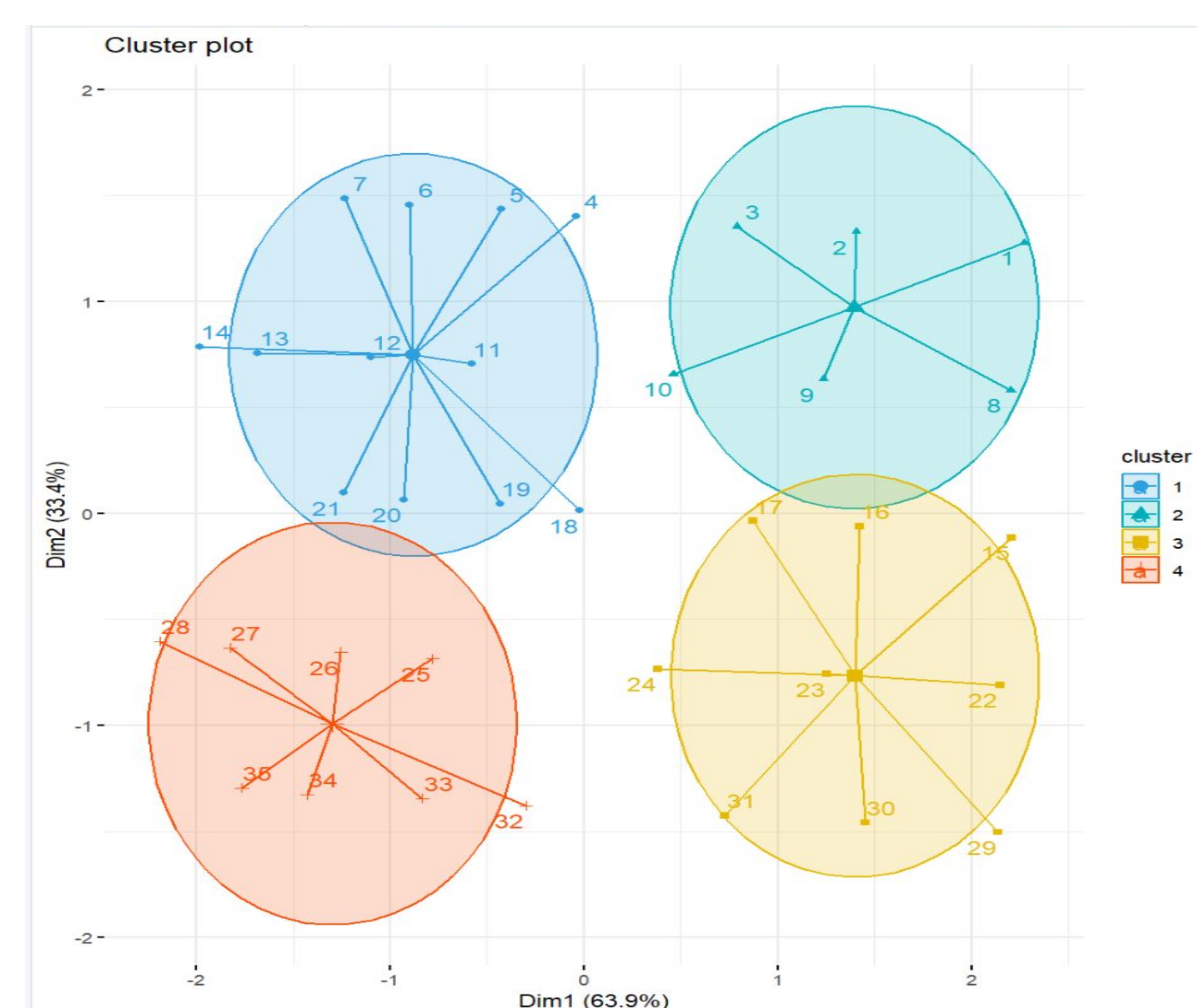


Gráfico 2. Representación de K- Means (k =4)

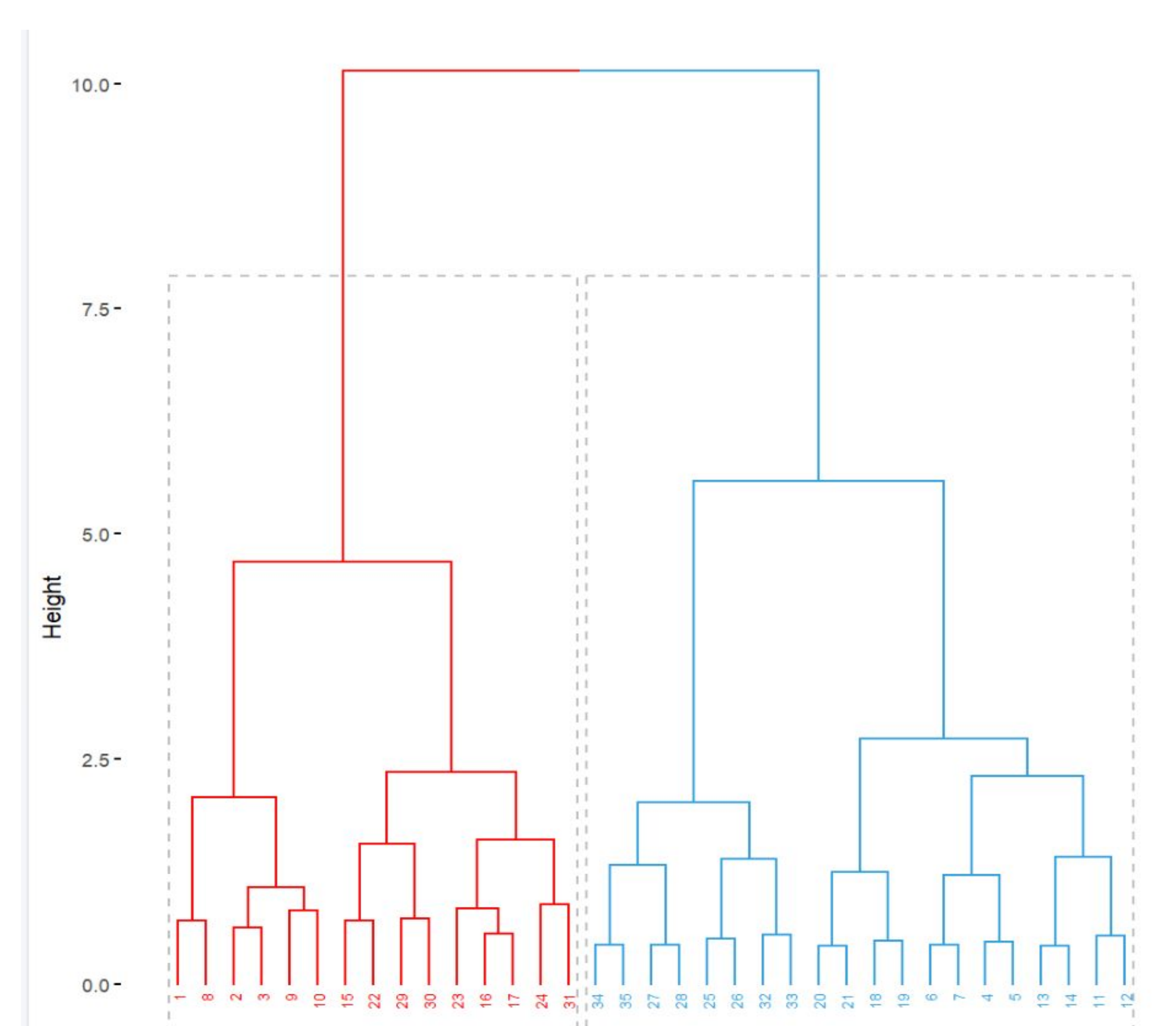


Gráfico 3. Dendrograma.

Referencias

Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. STHDA

Análisis de datos multivariantes. (2013)