

Expo: Estadística en el Entorno

Un Viaje de Análisis Clúster en el Desarrollo Humano

Participantes:

Jeovanni Vallejo Acalco, Jaime Alejandro Moreno Muñoz, María Luisa Córdoba Tlaxcalteco

Introducción

El Análisis de Clúster, o Análisis de Conglomerados, es una técnica de aprendizaje automático no supervisado que busca agrupar datos similares en clústeres distintos. Este trabajo se analiza la distancia entre cada objeto y su centroide, utilizando dos métodos principales: el agrupamiento jerárquico y el método k-means.

En el agrupamiento jerárquico, los datos se organizan en una jerarquía de clústeres, donde la proximidad entre los objetos determina la estructura. Por otro lado, el método k-means clasifica los datos en k clústeres, minimizando la distancia entre cada punto y el centroide del clúster al que pertenece.

Base de datos

La base de datos empleada se compone de cuatro grupos de desarrollo, en los cuales se categoriza el nivel de estudios a lo largo de los años comprendidos entre 2010 y 2021. Estas categorías incluyen niveles de clasificación como "Muy Alta", "Alta", "Medio" y "Bajo". El análisis abarca un conjunto total de 2340 datos.

Grupos.de..Desarrollo.Humano	X2010	X2011	X2012	X2013	X2014	X2015	X2016	X2017	X2018	X2019	X2020	X2021
1	0	2.746427	2.477821	2.475142	2.935539	2.832901	2.720964	2.672566	3.043233	3.056570	3.056570	3.056570
2	0	2.488360	3.667240	3.489860	4.297670	2.658110	2.607560	2.570640	2.999600	2.917450	2.522410	2.522410
3	1	3.727167	3.731171	3.702155	3.515670	3.491246	3.537645	3.549497	3.565062	3.565062	3.565062	3.565062
4	2	40.970000	40.970000	36.868090	36.868090	36.868090	36.868090	39.488210	39.488210	39.488210	39.488210	39.488210
5	0	7.627000	9.174500	9.174500	8.094880	9.276420	8.007100	8.693300	7.684310	8.180130	5.875610	5.875610

Fig. 1. Base de datos. Desarrollo humano.

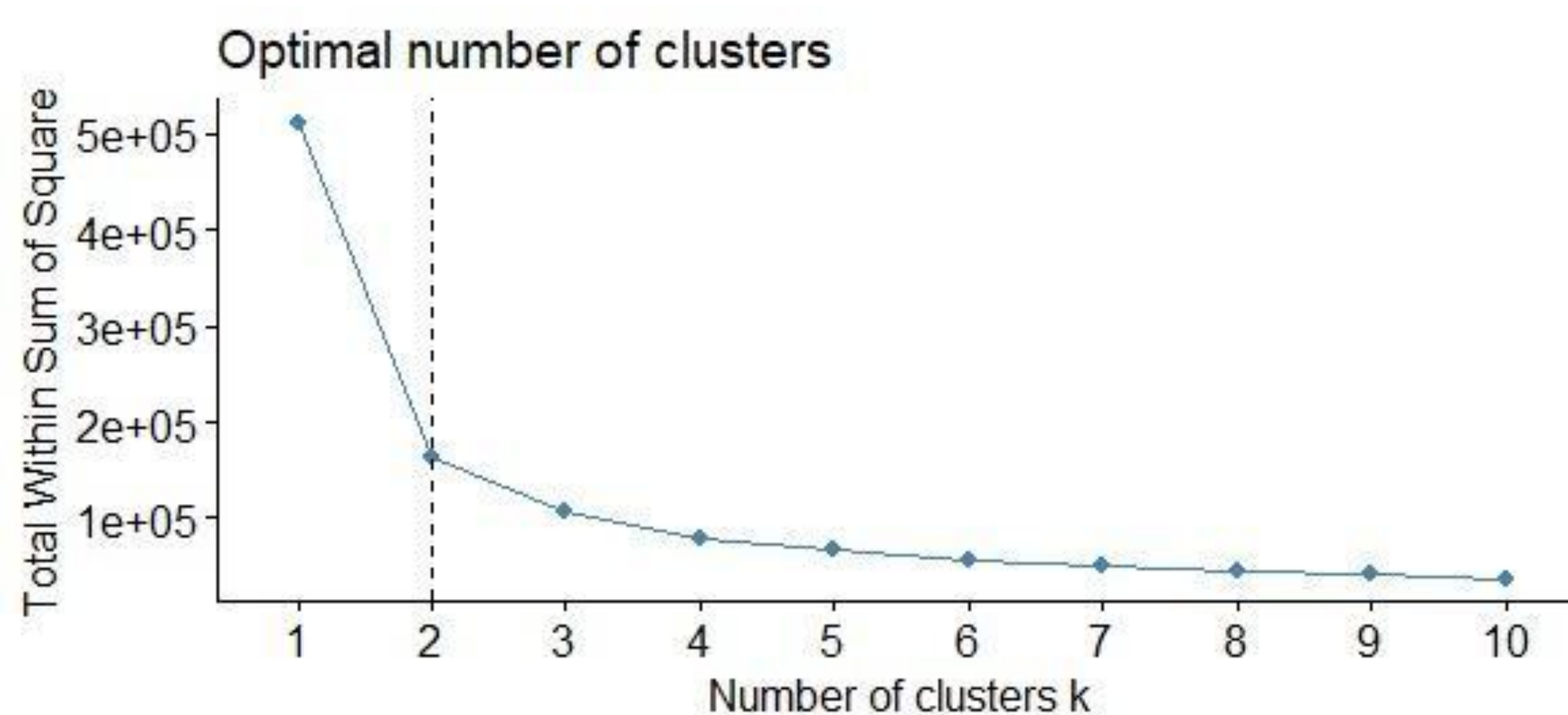


Fig. 2. Representación del número de clúster

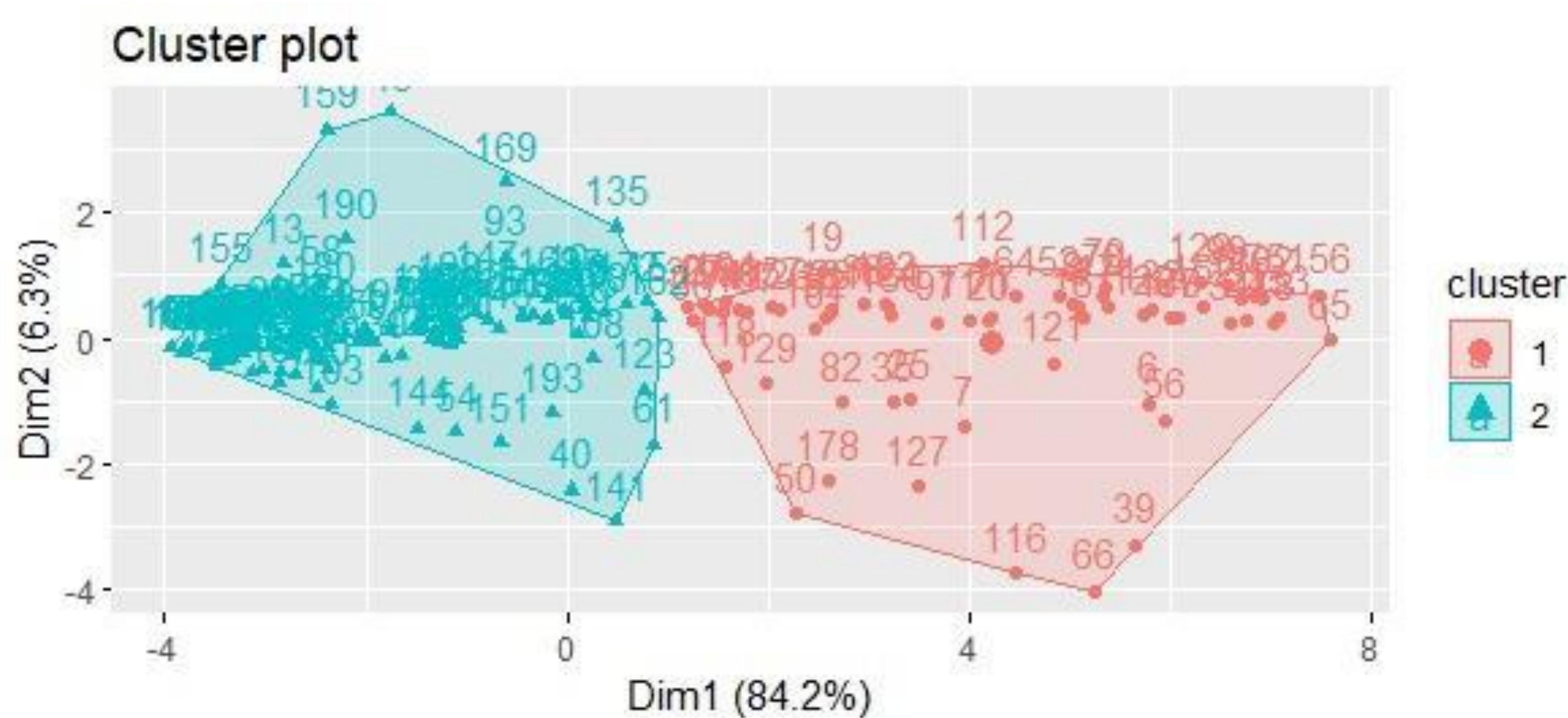


Fig. 3. Visualización del análisis k-means.

Resultado

El análisis de clúster ha revelado dos agrupamientos finales que representan los grupos de desarrollo. Estos dos grupos se distinguen por niveles contrastantes, clasificados como "Alta" y "Baja".

Esta diferenciación subraya las similitudes intrínsecas dentro de cada agrupamiento y proporciona una visión clara de las disparidades en los niveles de desarrollo identificados en el estudio

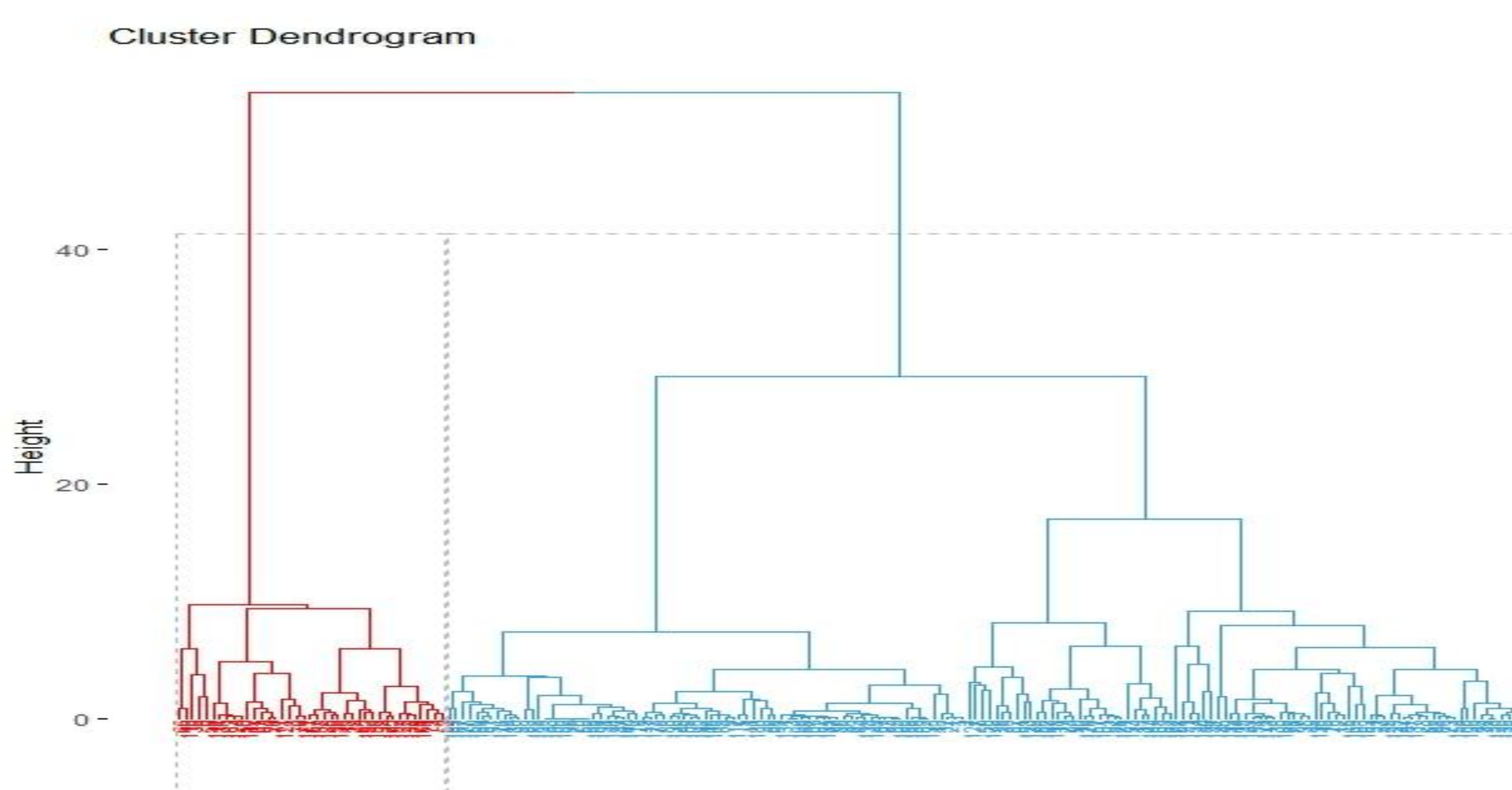


Fig. 4. Visualización del análisis de agrupamiento jerárquico.

Referencias

1. Srivastava, M. S. (2002). Methods of multivariate statistics (Vol. 419). John Wiley & Sons.
2. Johnson, R. A., Wichern, D. W. (2002). Applied multivariate statistical analysis. Upper Saddle River, NJ: Prentice Hall.