



# Expo: Estadística en el Entorno

## DATASET: IRIS

### Participantes:

Usbaldo Hernandez Landa, Danna Paola Hernandez Ortega y Maria Luisa Cordoba Tlaxcalteco

### Introducción

#### Escribir que es la tecnica

Iris es un conjunto de datos que consta de 150 observaciones de tres especies: setosa, versicolor y virginica. Para cada observación se registran cuatro características: longitud de sépalo, ancho del sépalo, longitud del pétalo y ancho del pétalo, todas medidas en centímetros.



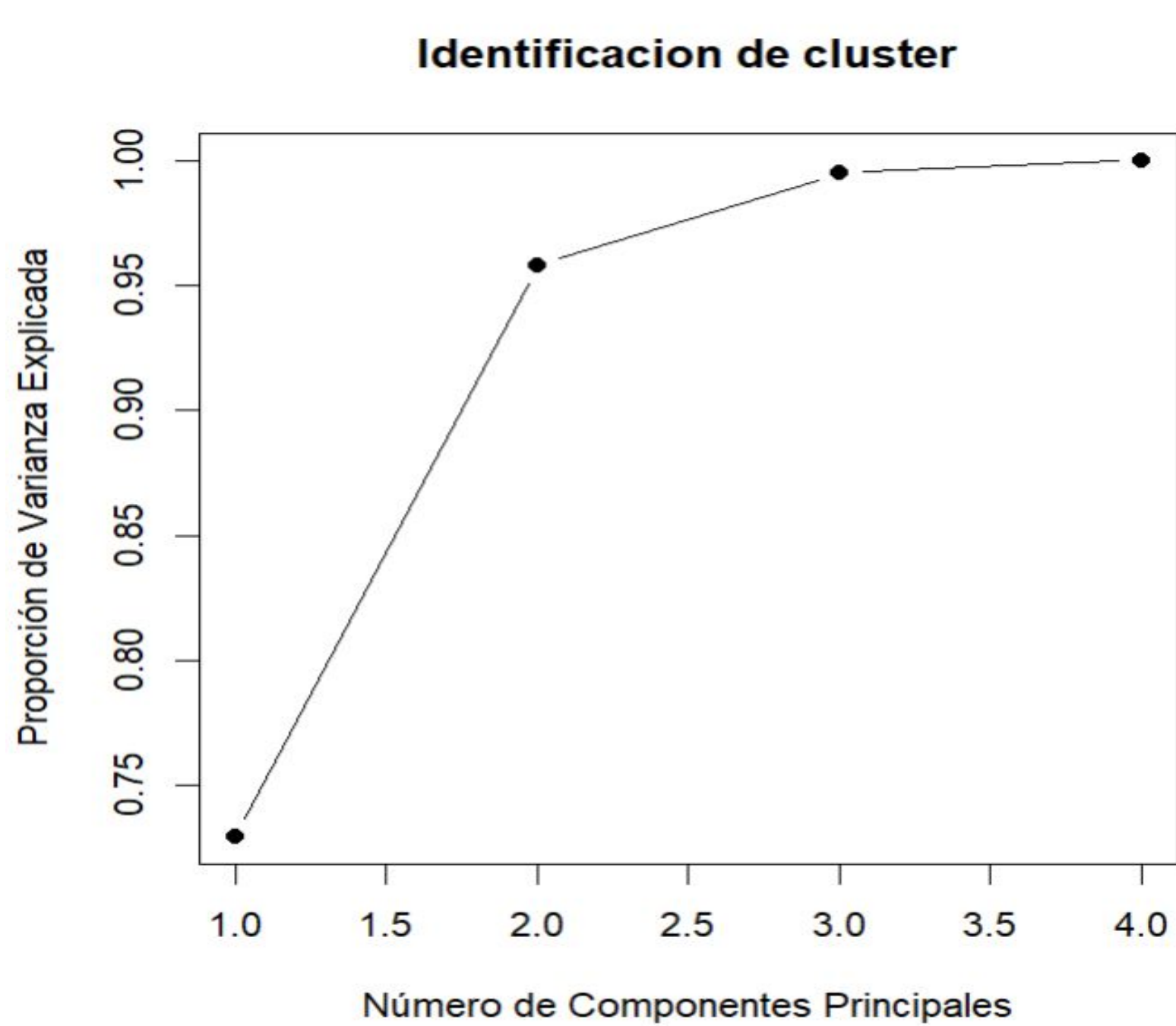
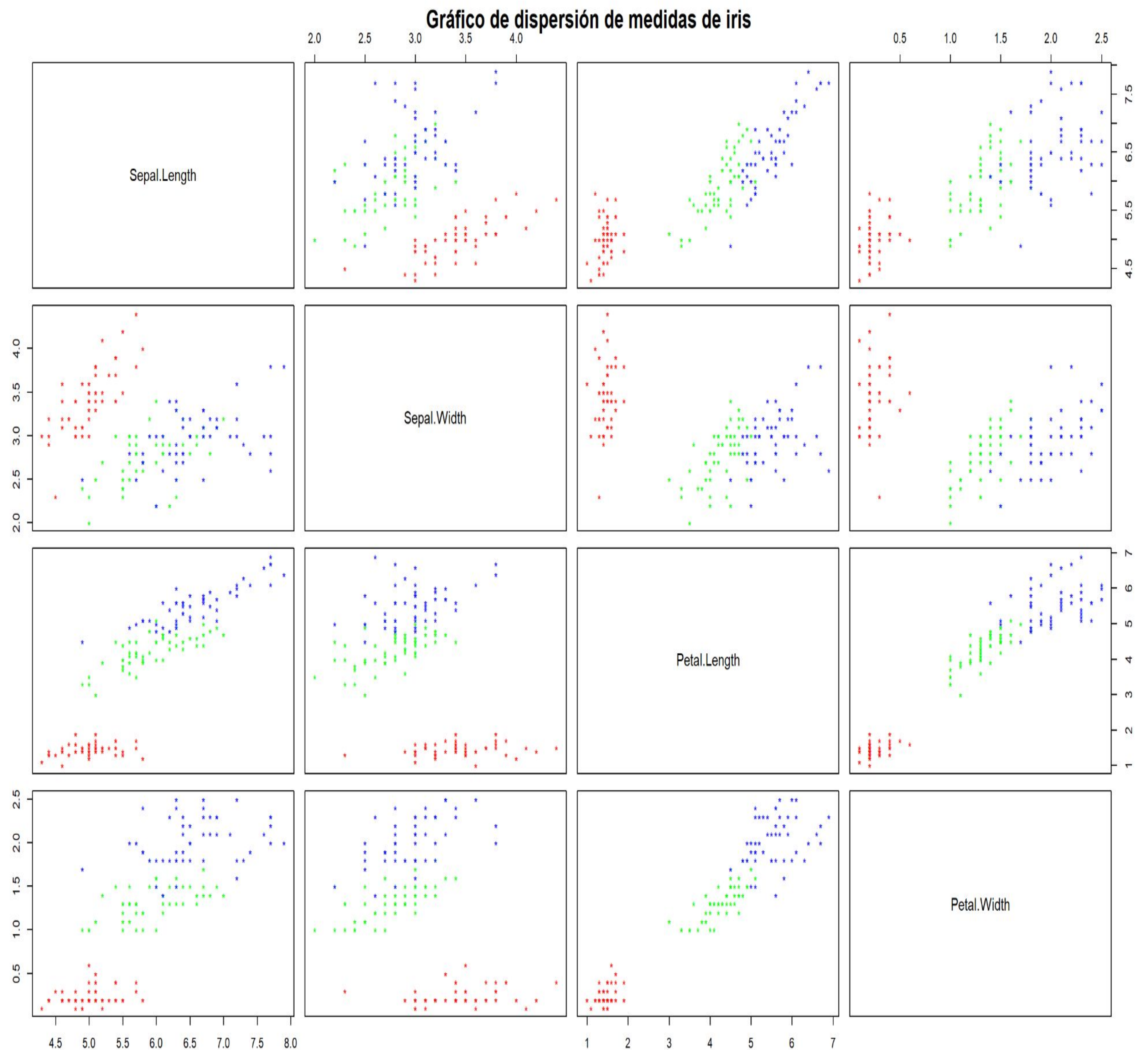
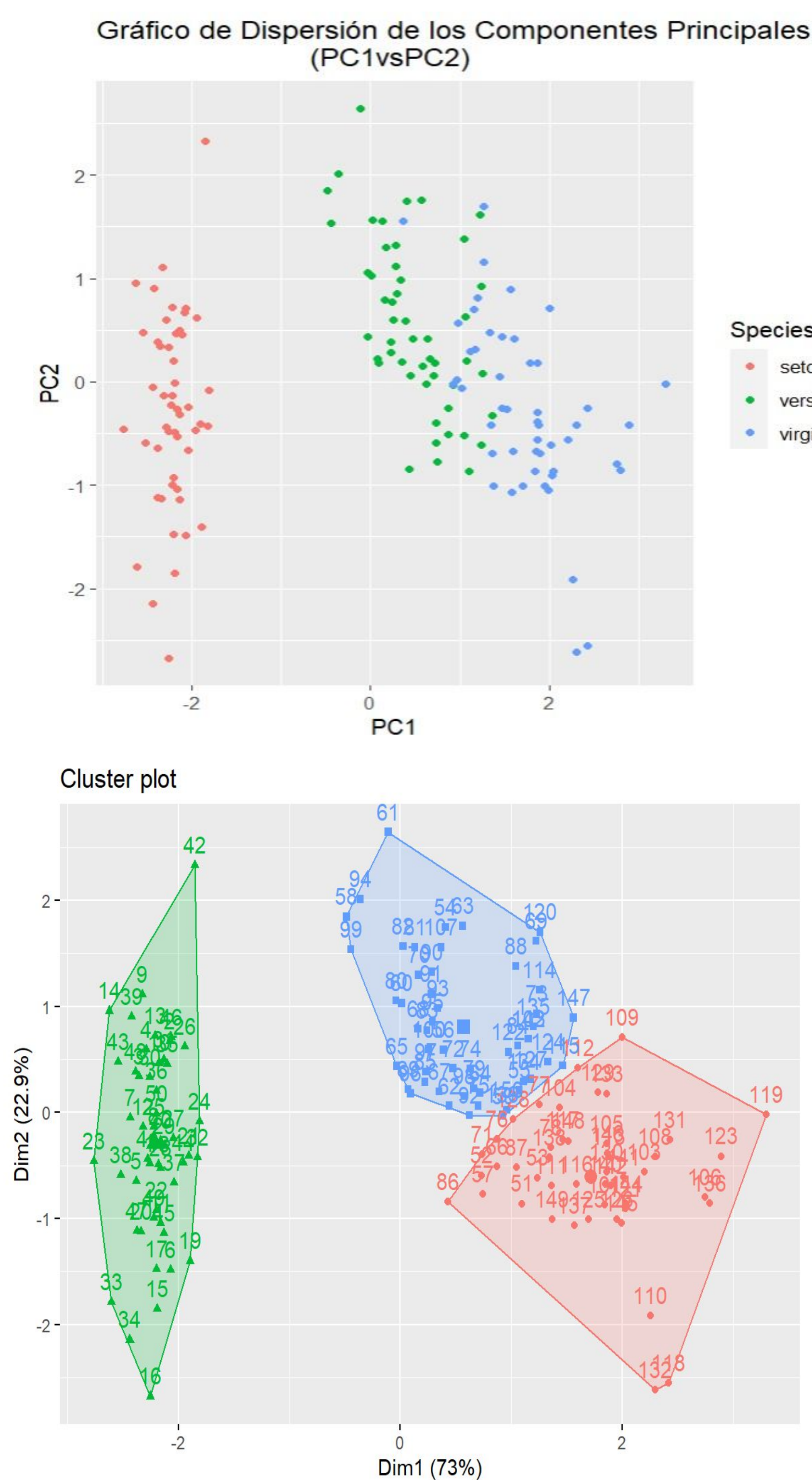
### Metodología

**Análisis de Componentes Principales (PCA):** Método estadístico que reduce la dimensionalidad de datos, identificando y preservando las variables más informativas para revelar patrones y simplificar la interpretación.

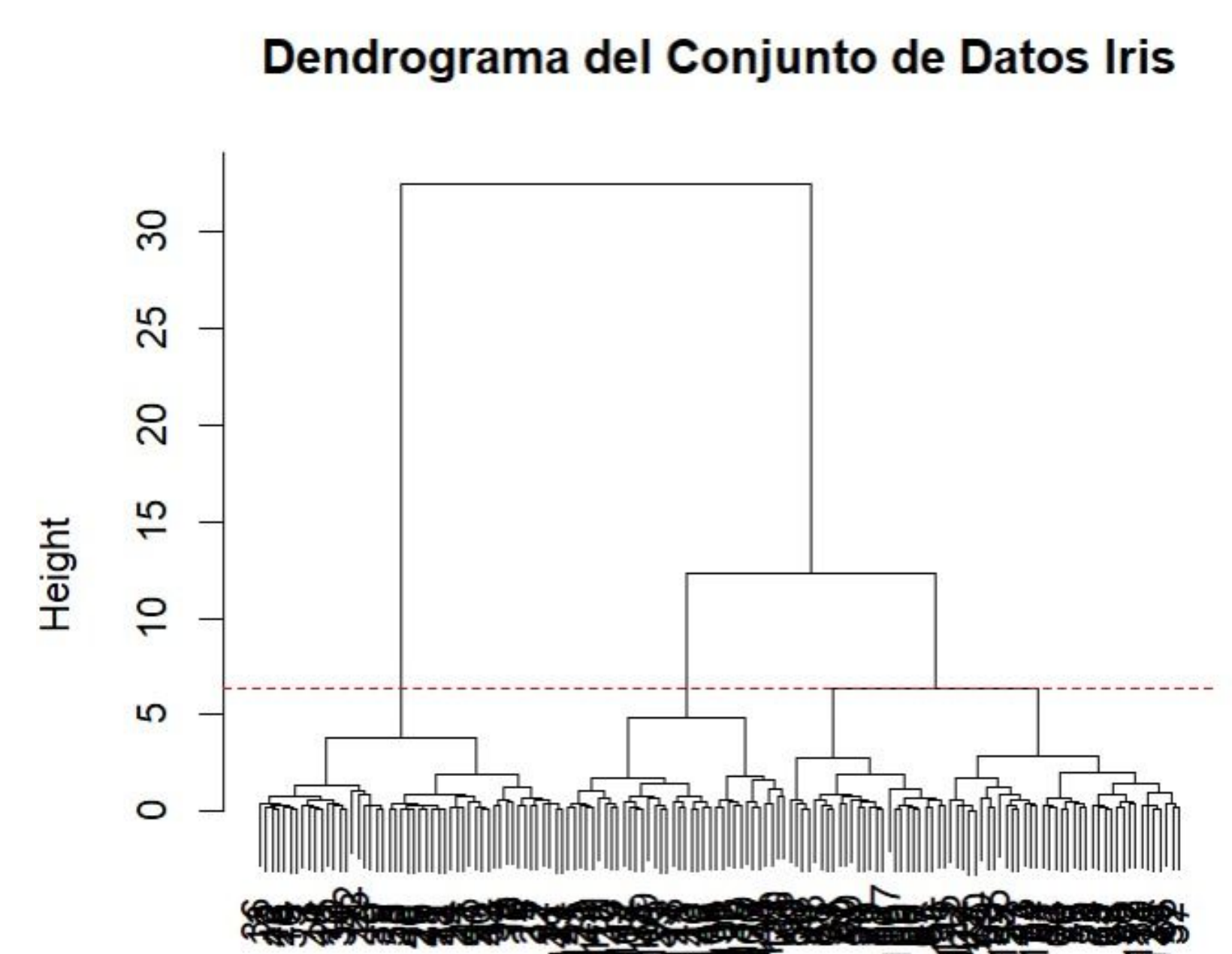
**Análisis Cluster:** Técnica estadística que agrupa datos similares en conjuntos llamados clusters, facilitando la identificación de patrones internos y la comprensión de la estructura.

**k-medias (k-means):** Algoritmo de agrupamiento que asigna puntos de datos a  $k$  clústeres, minimizando la varianza.

### Resultados



Actual	Predicted		
	setosa	versicolor	virginica
setosa	10	0	0
versicolor	0	10	0
virginica	0	3	7



### Discusión

Aplicando análisis de componentes principales en el conjunto de datos Iris, se identificó una estructura clara que permite reducir la dimensionalidad de los datos. Mediante cluster y k-medias, se identifican grupos coherentes, demostrando la utilidad de estas técnicas en la exploración de patrones y relaciones en las especies de flores estudiadas.

### Referencias

Peña, D. (2002). Análisis de Datos Multivariantes. Madrid: McGraw Hill

Fisher, R. A., Irwin, F. G., & Royal, H. W. (1936). The Iris Dataset. UCI Machine Learning Repository. <https://archive.ics.uci.edu/ml/datasets/iris>