**Lab3**

**Cord**

# Load required libraries

library(stats)

library(cluster)

# Function to simulate student features

simulate\_student\_features <- function(n = 100) {

# Set the random seed

set.seed(260923)

# Generate unique student IDs

student\_ids <- seq(1, n)

# Simulate student engagement

student\_engagement <- rnorm(n, mean = 50, sd = 10)

# Simulate student performance

student\_performance <- rnorm(n, mean = 60, sd = 15)

# Combine the data into a data frame

student\_features <- data.frame(

student\_id = student\_ids,

student\_engagement = student\_engagement,

student\_performance = student\_performance

)

# Return the data frame

return(student\_features)

}

# Simulate student features

student\_data <- simulate\_student\_features(n = 100)

# Perform PCA for dimensionality reduction

pca\_result <- prcomp(student\_data[, -1], scale. = TRUE)

# Get PCA results

pca\_data <- as.data.frame(pca\_result$x)

# Cluster using KMeans

kmeans\_clusters <- kmeans(pca\_data, centers = 3)

# Cluster using Hierarchical clustering

hierarchical\_clusters <- hclust(dist(pca\_data))

# Plot the PCA results with KMeans clusters

plot(pca\_data[, 1:2], col = kmeans\_clusters$cluster, pch = 19,

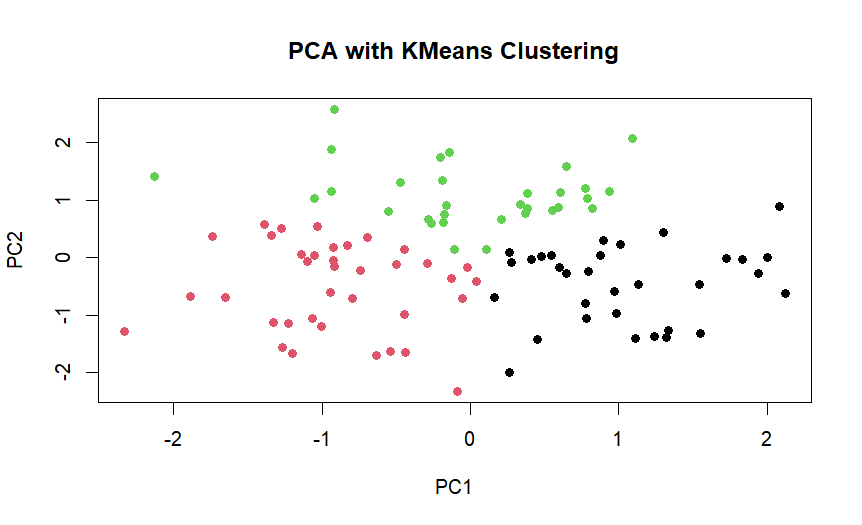
main = "PCA with KMeans Clustering", xlab = "PC1", ylab = "PC2")

# Plot the dendrogram from Hierarchical clustering

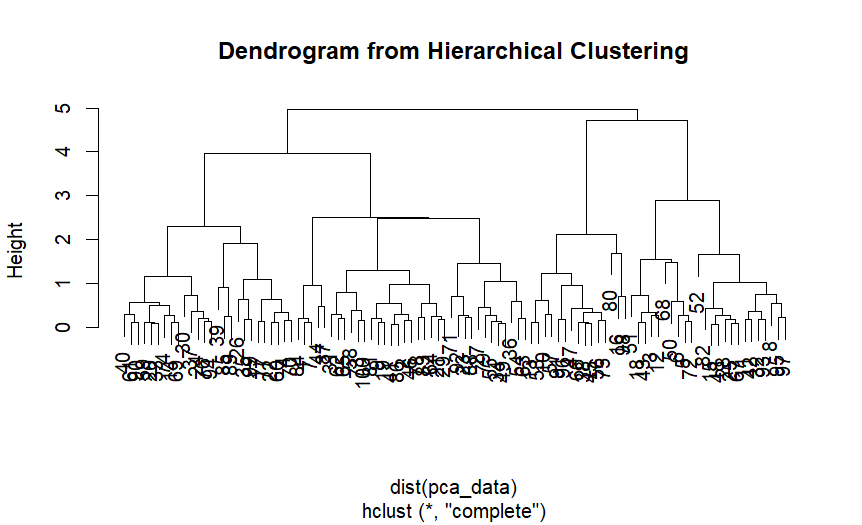
plot(hierarchical\_clusters, main = "Dendrogram from Hierarchical Clustering")

1. **Approach to Dimensionality Reduction and Clustering**

**The study used Principal Component Analysis (PCA) to decrease dimensionality, KMeans clustering to identify distinct clusters, Hierarchical clustering to explore alternative clustering structures within the data, and simulated student engagement and performance data for a total of 100 students.**

****

**2.Results of the Analysis**

**Using KMeans and hierarchical clustering, the study noticed three clusters: moderate, high, and low. While students in Cluster 3 could need substantial support, students in Cluster 2 might benefit from focused therapy. However, Cluster 1 students can benefit from focused therapy because they are academically successful and highly motivated.**

****3.Implications for Learning Analytics****

**Personalized assistance strategies can be developed using learning analytics for students specific clusters, such as Cluster 3. Additionally, predictive models can be used to think learning outcomes and address possible problems. To maximize the impact on student success, assets like technology support, guidance, and tutoring can be provided based on these clusters.**

****4. Scholarly Reference****

****Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.****

**This reference provides insights into the use of data mining techniques, including clustering, for educational data analysis and decision-making.**