# Load necessary libraries

library(ggplot2)

library(cluster)

# Read the dataset

mall\_data <- read.csv("Mall\_Customers.csv")

# Select relevant columns for clustering

mall\_features <- mall\_data[, c("age", "annual.income", "spending.score")]

# Standardize the features

mall\_features\_scaled <- scale(mall\_features)

# Determine optimal number of clusters using the Elbow method

wss <- sapply(1:10, function(k) sum(kmeans(mall\_features\_scaled, k)$withinss))

plot(1:10, wss, type="b", xlab="Number of clusters", ylab="Within cluster sum of squares")

# From the plot, select the optimal number of clusters (the "elbow point")

# Perform K-means clustering

k <- 5 # Change this number to the optimal number of clusters

kmeans\_result <- kmeans(mall\_features\_scaled, centers = k, nstart = 25)

# Add cluster assignments to original dataset

mall\_data$cluster <- as.factor(kmeans\_result$cluster)

# Visualize the clusters

ggplot(mall\_data, aes(x = annual.income, y = spending.score, color = cluster)) +

geom\_point() +

labs(title = "K-means Clustering of Mall Customers",

x = "Annual Income",

y = "Spending Score",

color = "Cluster") +

theme\_minimal()