

Q9.R

Manoj

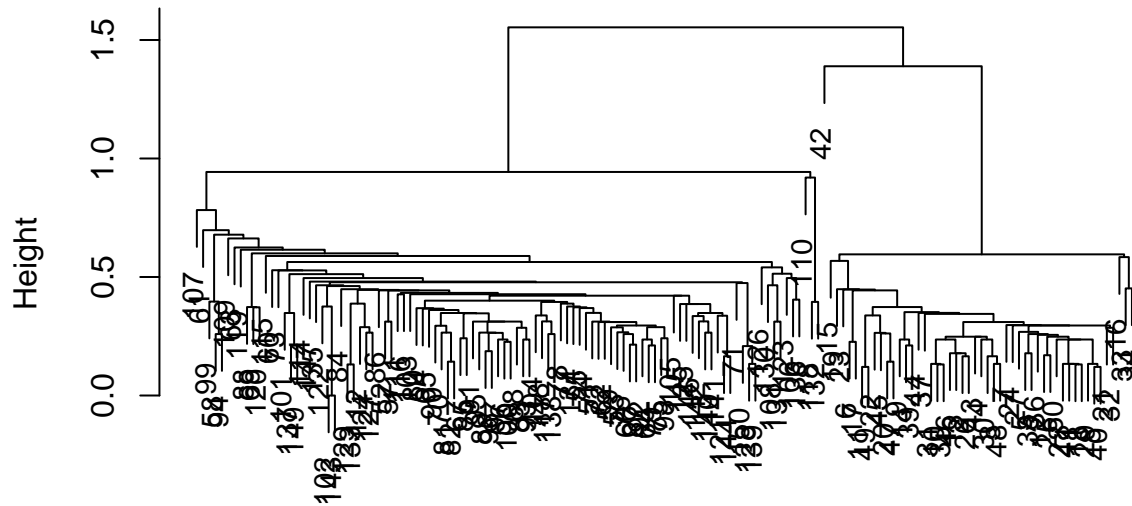
2024-05-31

```
# Load necessary libraries  
library(datasets)  
library(cluster)  
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
# Load the iris dataset  
data(iris)  
  
# Extract the first four variables  
iris_data <- iris[, 1:4]  
  
# Standardize the data  
iris_data <- scale(iris_data)  
  
# a) Fit hierarchical clustering model using single linkage  
hc_single <- hclust(dist(iris_data), method = "single")  
  
# Plot the dendrogram for single linkage  
plot(hc_single, main = "Dendrogram (Single Linkage)", xlab = "", sub = "", cex = 0.9)
```

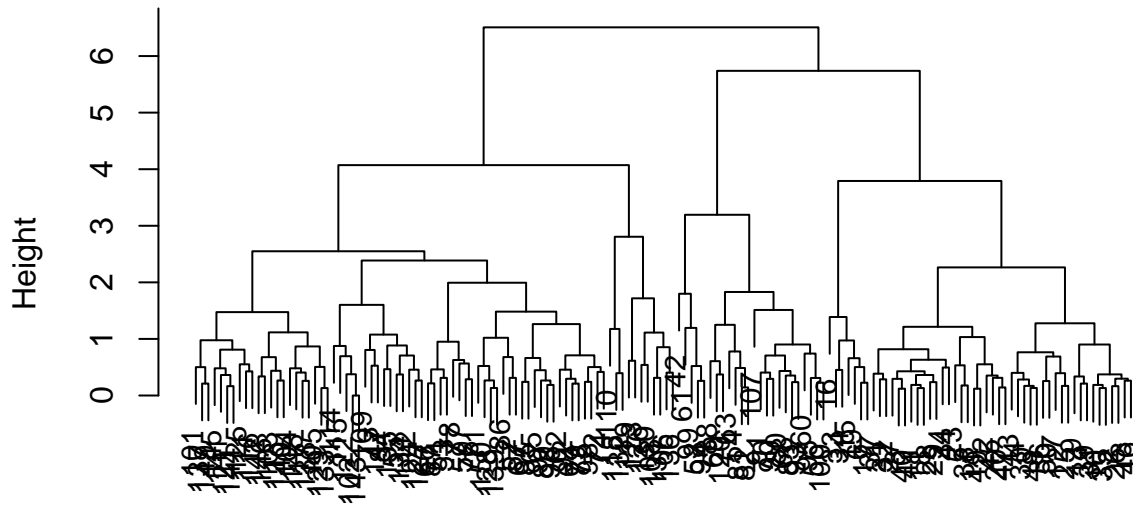
Dendrogram (Single Linkage)



```
# b) Fit hierarchical clustering model using complete linkage
hc_complete <- hclust(dist(iris_data), method = "complete")

# Plot the dendrogram for complete linkage
plot(hc_complete, main = "Dendrogram (Complete Linkage)", xlab = "", sub = "", cex = 0.9)
```

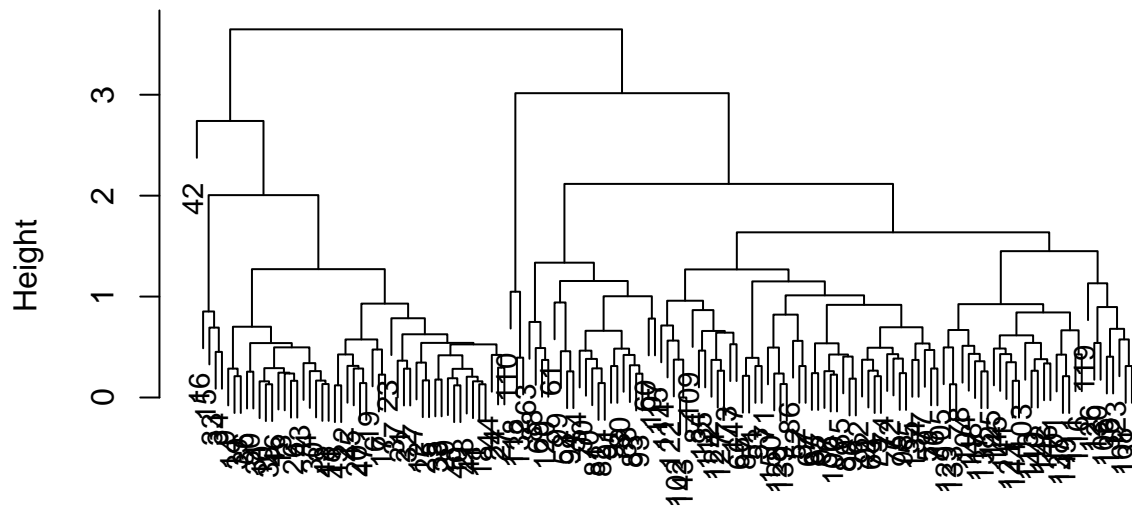
Dendrogram (Complete Linkage)



```
# c) Fit hierarchical clustering model using average linkage
hc_average <- hclust(dist(iris_data), method = "average")

# Plot the dendrogram for average linkage
plot(hc_average, main = "Dendrogram (Average Linkage)", xlab = "", sub = "", cex = 0.9)
```

Dendrogram (Average Linkage)



```
# d) Find the best hierarchical clustering model and locate the number of clusters
# Use the silhouette method to find the optimal number of clusters for each model
sil_width_single <- silhouette(cutree(hc_single, k = 3), dist(iris_data))
sil_width_complete <- silhouette(cutree(hc_complete, k = 3), dist(iris_data))
sil_width_average <- silhouette(cutree(hc_average, k = 3), dist(iris_data))

# Calculate average silhouette width for each method
avg_sil_width_single <- mean(sil_width_single[, 3])
avg_sil_width_complete <- mean(sil_width_complete[, 3])
avg_sil_width_average <- mean(sil_width_average[, 3])

# Determine the best model
best_model <- which.max(c(avg_sil_width_single, avg_sil_width_complete, avg_sil_width_average))

# Display the best model and the corresponding number of clusters
cat("Best hierarchical clustering model is: ")
```

```
## Best hierarchical clustering model is:
```

```
if (best_model == 1) {
  cat("Single Linkage\n")
} else if (best_model == 2) {
  cat("Complete Linkage\n")
} else {
```

```
cat("Average Linkage\n")
}
```

Single Linkage

```
# Determine the optimal number of clusters for the best model using silhouette method
optimal_clusters <- function(hc) {
  sil_widths <- sapply(2:10, function(k) {
    mean(silhouette(cutree(hc, k), dist(iris_data))[, 3])
  })
  which.max(sil_widths) + 1
}

optimal_clusters_single <- optimal_clusters(hc_single)
optimal_clusters_complete <- optimal_clusters(hc_complete)
optimal_clusters_average <- optimal_clusters(hc_average)

cat("Optimal number of clusters for the best model: ")
```

Optimal number of clusters for the best model:

```
if (best_model == 1) {
  cat(optimal_clusters_single, "\n")
} else if (best_model == 2) {
  cat(optimal_clusters_complete, "\n")
} else {
  cat(optimal_clusters_average, "\n")
}
```

2