

Question_No_10

Utsab Bhattarai

2024-05-31

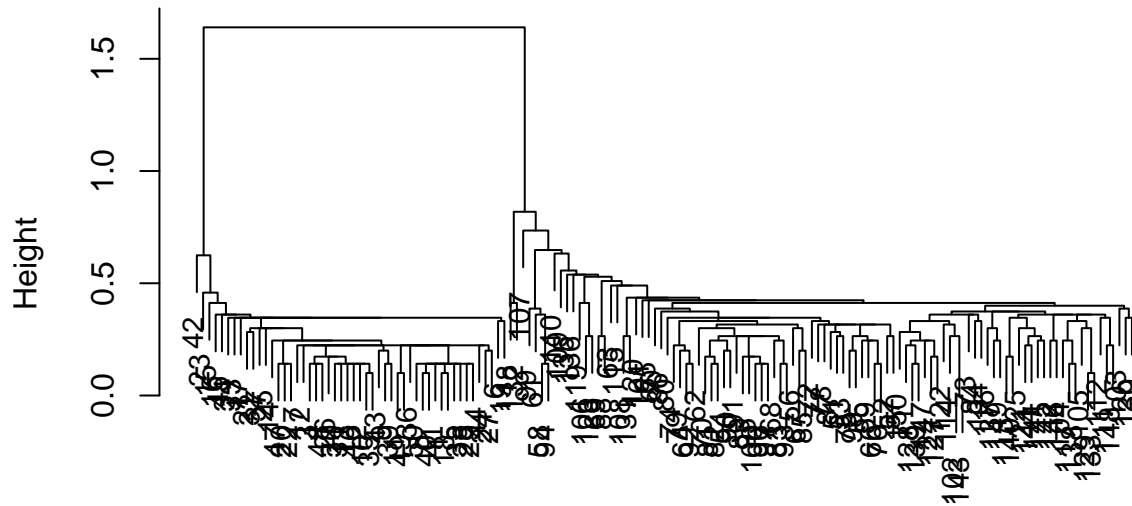
Use the first four variables of the “iris” data and do as follows with R script to knit PDF output:

```
data("iris")
```

a. Fit a hierarchical clustering model using single linkage and get the dendrogram for this model.

```
iris_data <- iris[, 1:4]
dist_iris <- dist(iris_data)
hc_single <- hclust(dist_iris,
                    method = "single")
plot(hc_single,
     main = "Dendrogram for Single Linkage",
     sub = "",
     xlab = "",
     cex = 0.9)
```

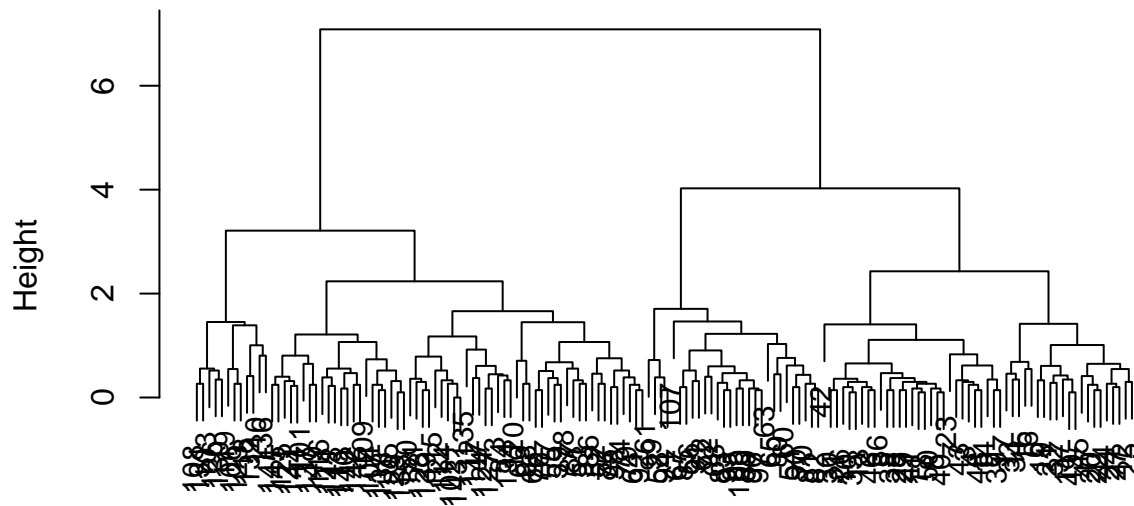
Dendrogram for Single Linkage



b. Fit a hierarchical clustering model using complete linkage and get the dendrogram for this model.

```
hc_complete <- hclust(dist_iris, method = "complete")
plot(hc_complete,
     main = "Dendrogram for Complete Linkage",
     sub = "",
     xlab = "",
     cex = 0.9)
```

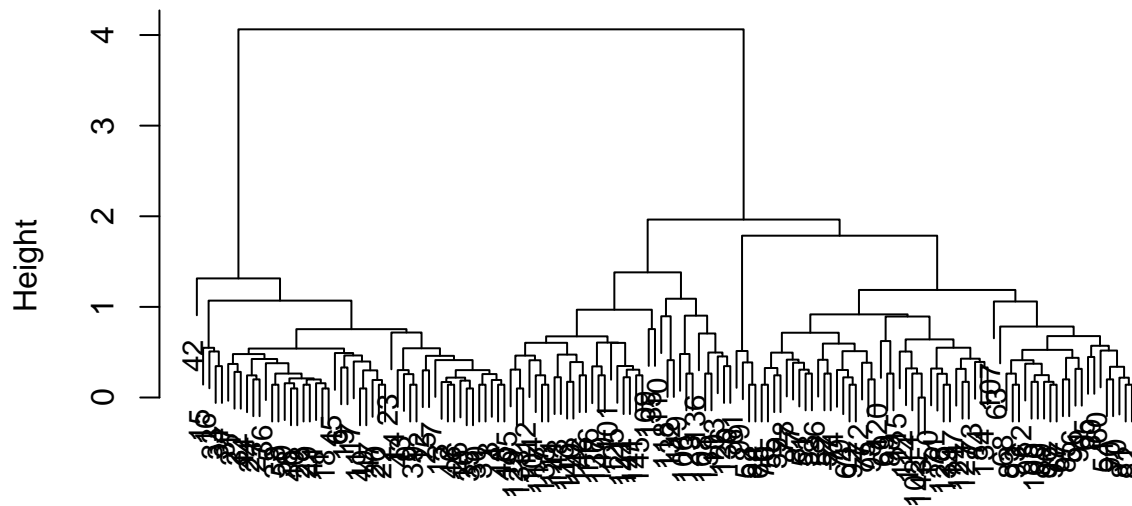
Dendrogram for Complete Linkage



c. Fit a hierarchical clustering model using average linkage and get the dendrogram for this model.

```
hc_average <- hclust(dist_iris,  
                     method = "average")  
plot(hc_average,  
     main = "Dendrogram for Average Linkage",  
     sub = "",  
     xlab = "",  
     cex = 0.9)
```

Dendrogram for Average Linkage



d. Find the best hierarchical clustering model, model for this data and locate the number of clusters for it.

```
library(cluster)
```

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

```
sil_width_single <- silhouette(cutree(hc_single,
                                     k = 3),
                              dist_iris)
sil_width_complete <- silhouette(cutree(hc_complete,
                                       k = 3),
                                dist_iris)
sil_width_average <- silhouette(cutree(hc_average,
                                       k = 3),
                                dist_iris)
# Calculate average silhouette width for each model
avg_sil_single <- mean(sil_width_single[,
                                     3])
avg_sil_complete <- mean(sil_width_complete[,
                                           3])
avg_sil_average <- mean(sil_width_average[,
```

```

3])
cat("Average Silhouette Width for Single Linkage:", avg_sil_single, "\n")

## Average Silhouette Width for Single Linkage: 0.5121108

cat("Average Silhouette Width for Complete Linkage:", avg_sil_complete, "\n")

## Average Silhouette Width for Complete Linkage: 0.5135953

cat("Average Silhouette Width for Average Linkage:", avg_sil_average, "\n")

## Average Silhouette Width for Average Linkage: 0.5541609

best_model <- which.max(c(avg_sil_single,
                        avg_sil_complete,
                        avg_sil_average))
model_names <- c("Single Linkage",
                "Complete Linkage",
                "Average Linkage")
cat("The best hierarchical clustering model is:", model_names[best_model], "\n")

## The best hierarchical clustering model is: Average Linkage

best_hc <- switch(best_model,
                 hc_single,
                 hc_complete,
                 hc_average)

```