HW5

Runye Hu A13410385

Yijie Fan A13485989

Xinyi He A13561164

Yifan Wu A14060535
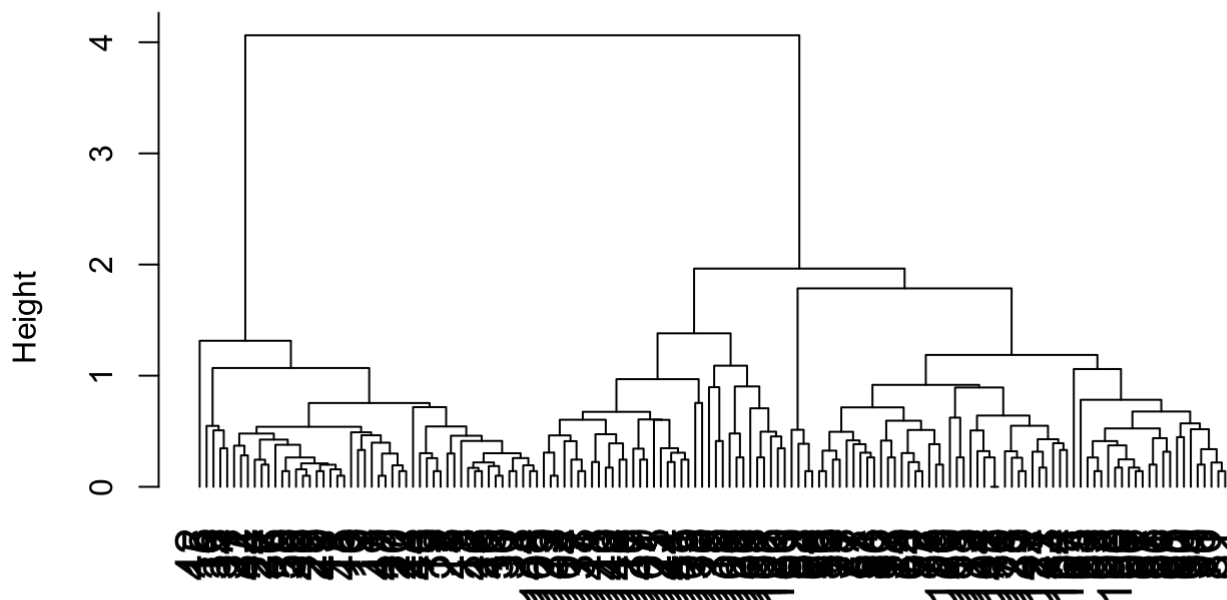
Yiwen Cai A13530685

```
############## Load Data #############
library(datasets)
data(iris)
```

Q1

```
dist<- dist(iris[1:4], method = "euclidean")
hc.average <- hclust (dist, method ="average")
clusterCut = cutree(hc.average, 3)

plot(hc.average , main=" Dendrogram: Average Linkage ",
xlab="", sub ="", cex =1.5 , hang = -1)
```

Dendrogram: Average Linkage

```
truth = as.numeric(iris$Species)
correct_rate = sum(truth == clusterCut)/length(truth)
```
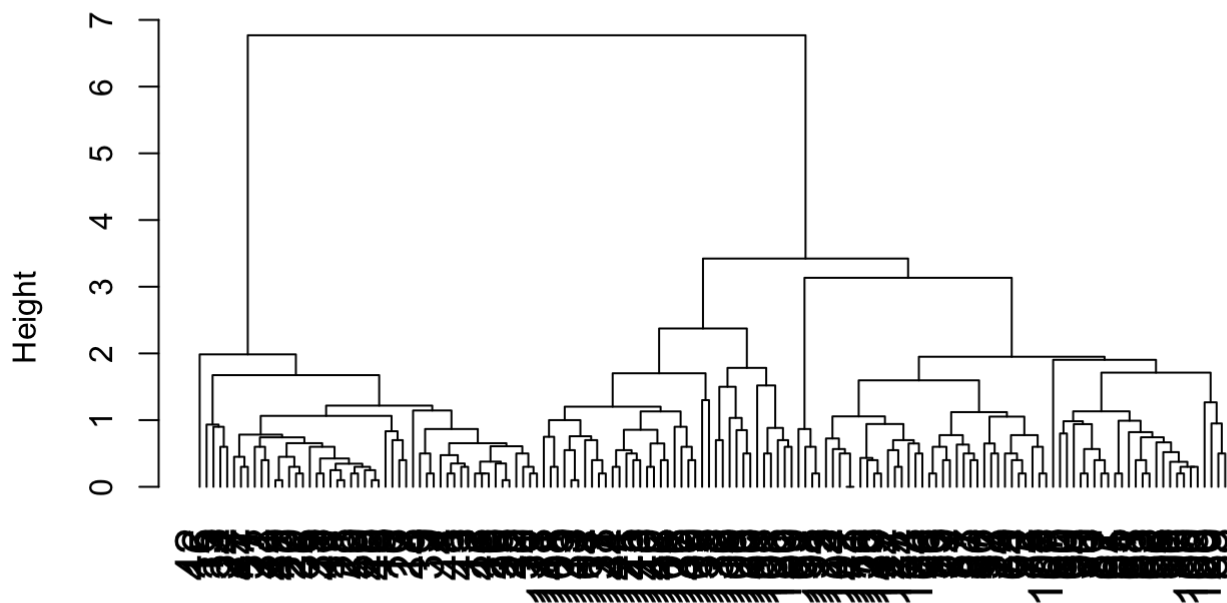
```
correct_rate
```

```
## [1] 0.9066667
```

Q2.1

```
dist<- dist(iris[1:4], method = "manhattan")
hc.average <- hclust (dist, method ="average")
clusterCut = cutree(hc.average, 3)

plot(hc.average , main=" Dendrogram: Average Linkage ",
xlab="", sub ="", cex =1.5 , hang = -1)
```

Dendrogram: Average Linkage


```
truth = as.numeric(iris$Species)

correct_rate = sum(truth == clusterCut)/length(truth)
```
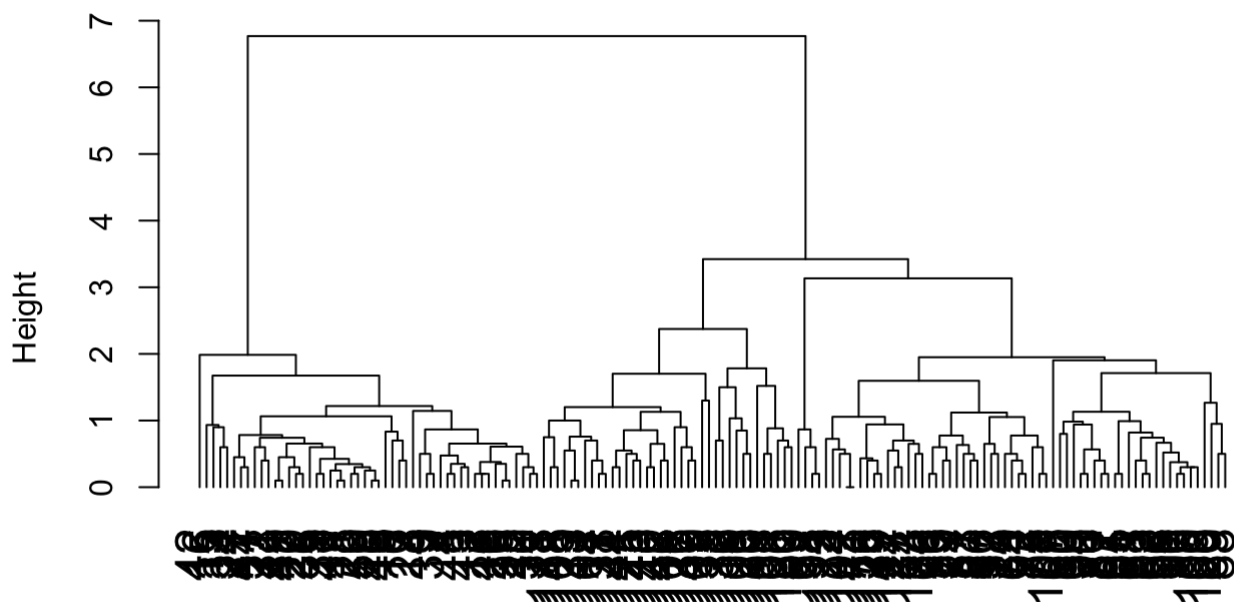
```
correct_rate
```

```
## [1] 0.9
```

Q2.2

```
dist<- dist(iris[1:4], method = "euclidean")
hc.single <- hclust (dist, method ="single")
clusterCut = cutree(hc.single, 3)

plot(hc.average , main=" Dendrogram: Single Linkage ",
xlab="", sub ="", cex =1.5 , hang = -1)
```

Dendrogram: Single Linkage


```
truth = as.numeric(iris$Species)
correct_rate = sum(truth == clusterCut)/length(truth)
```
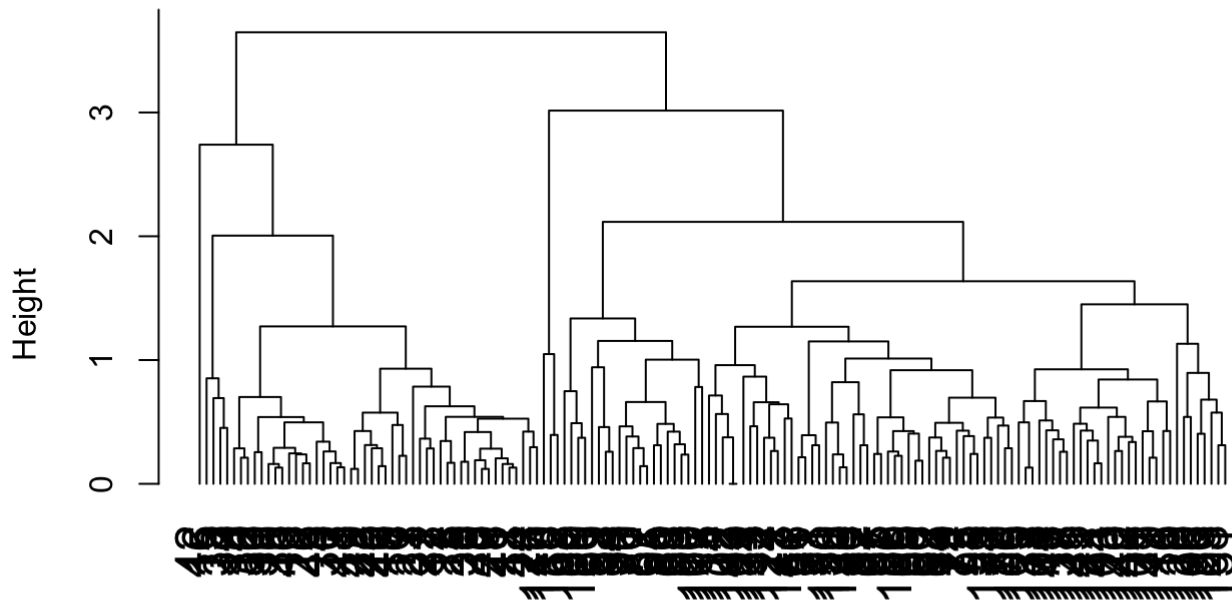
```
correct_rate
```

```
## [1] 0.68
```

Q2.3

```
data = scale(iris[1:4])
dist<- dist(data, method = "euclidean")
hc.average <- hclust (dist, method ="average")
clusterCut = cutree(hc.average, 3)

plot(hc.average , main=" Dendrogram: Average Linkage ",
xlab="", sub ="", cex =1.5 , hang = -1)
```

Dendrogram: Average Linkage


```
truth = as.numeric(iris$Species)
correct_rate = sum(truth == clusterCut)/length(truth)
```

```
correct_rate
```

```
## [1] 0.6866667
```