

Math189 HW5

Group Members:

Yaqi Chen (PID: A15742547; Section: A03)

Yuetong Lyu (PID: A13779993; Section: A04)

Siyi He (PID: A13400569; Section: A03)

Zhenyuan Xu (PID: A92067995; Section: A02)

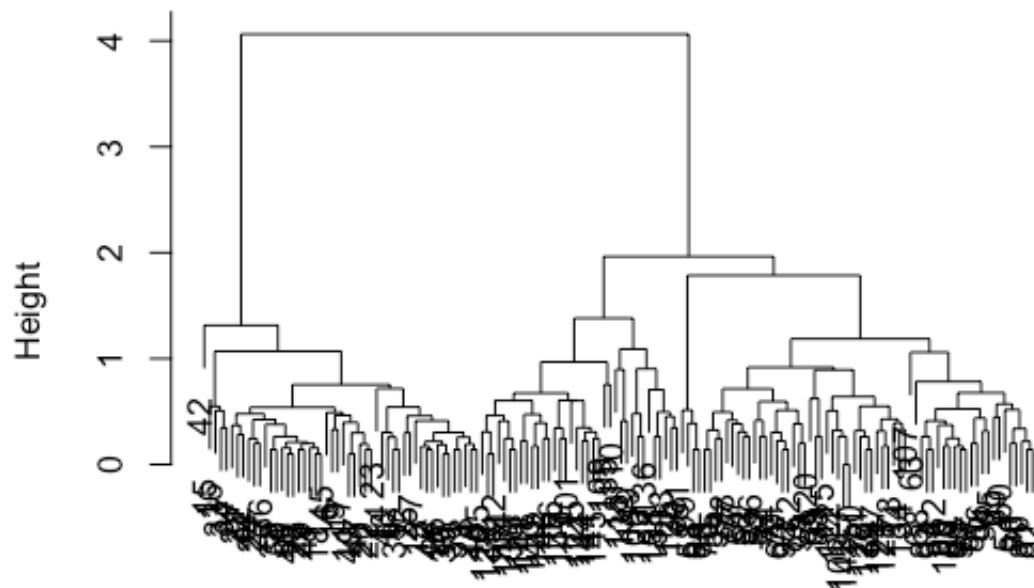
Jiawei Chao (PID: A13818001; Section: A02)

Problem 1

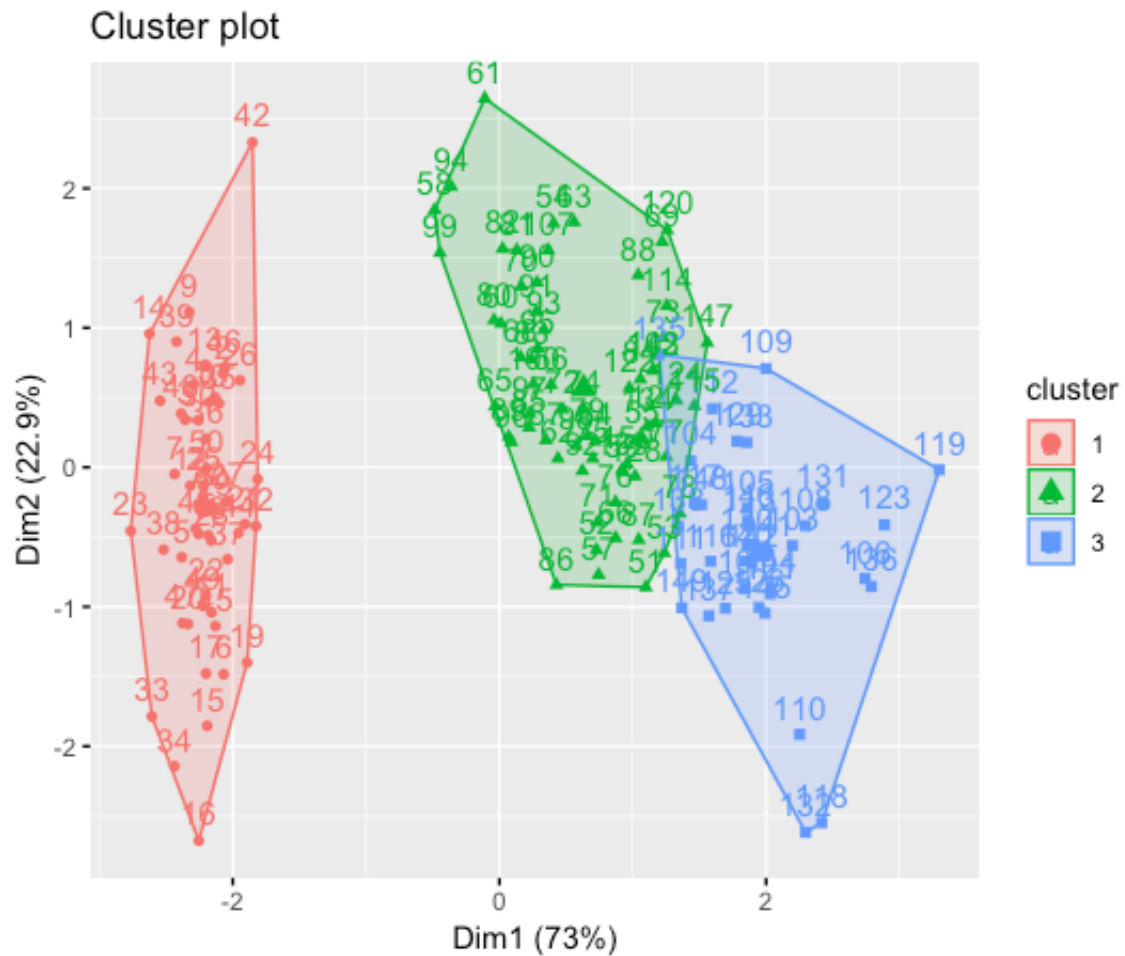
```
install.packages("factoextra")
library(factoextra)
data(iris)
data<-iris[c(1,2,3,4)]
data<-na.omit(data)
d<-dist(data,method="euclidean")
hc<- hclust(d,method="average")
plot(hc,main=" Dendrogram: Average Linkage ")
sub_grp<-cutree(hc,k=3)
fviz_cluster(list(data = data, cluster = sub_grp))
species <- c(rep(1,50),rep(2,50),rep(3,50))
count <- 0
for (i in 1:150){
  if (sub_grp[i] == species[i]){
    count <- count + 1
  }
}
rate <- count/150
print(paste("Correct clustering rate =",rate))
```

"Correct clustering rate = 0.9066666666666667"

Dendrogram: Average Linkage



d
hclust (*, "average")

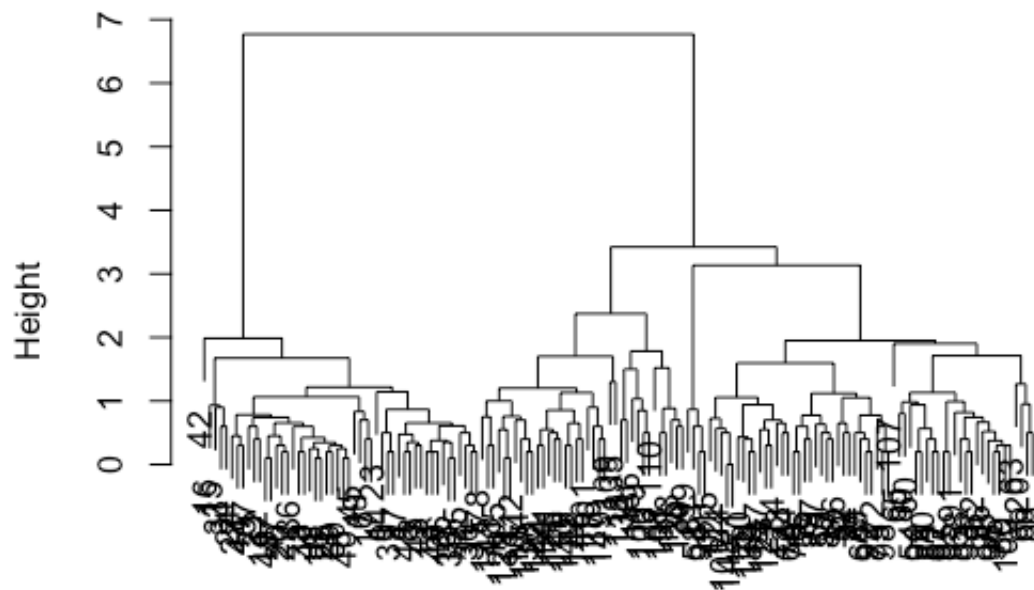


Problem 2A

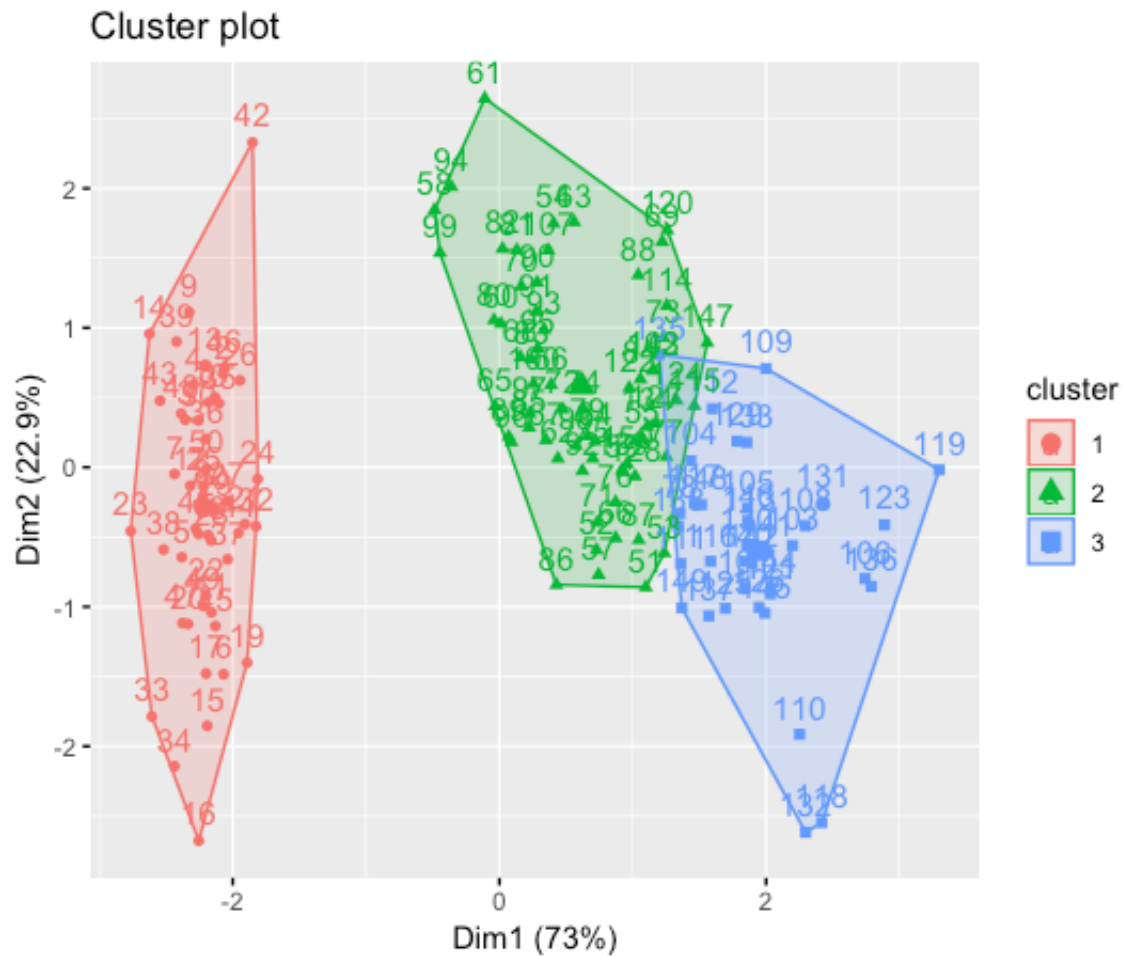
```
d2<-dist(data,method="manhattan")
hc2<- hclust(d2,method="average")
plot(hc2,main=" Dendrogram: Average Linkage ")
sub_grp2<-cutree(hc2,k=3)
fviz_cluster(list(data = data, cluster = sub_grp2))
count <- 0
for (i in 1:150){
  if (sub_grp2[i] == species[i]){
    count <- count + 1
  }
}
rate <- count/150
print(paste("Correct clustering rate =",rate))
```

"Correct clustering rate = 0.9"

Dendrogram: Average Linkage



d2
hclust (*, "average")

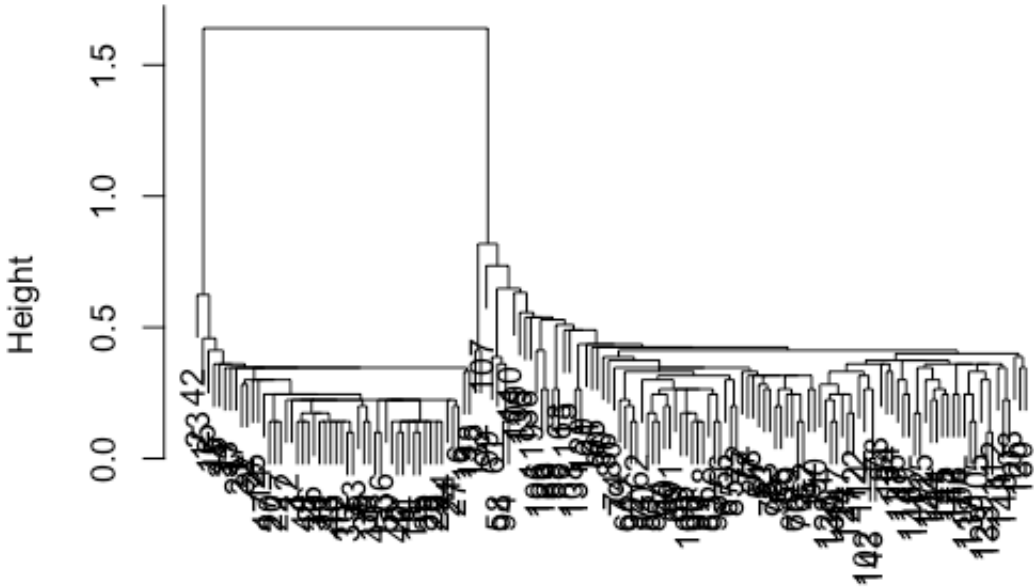


Problem 2B

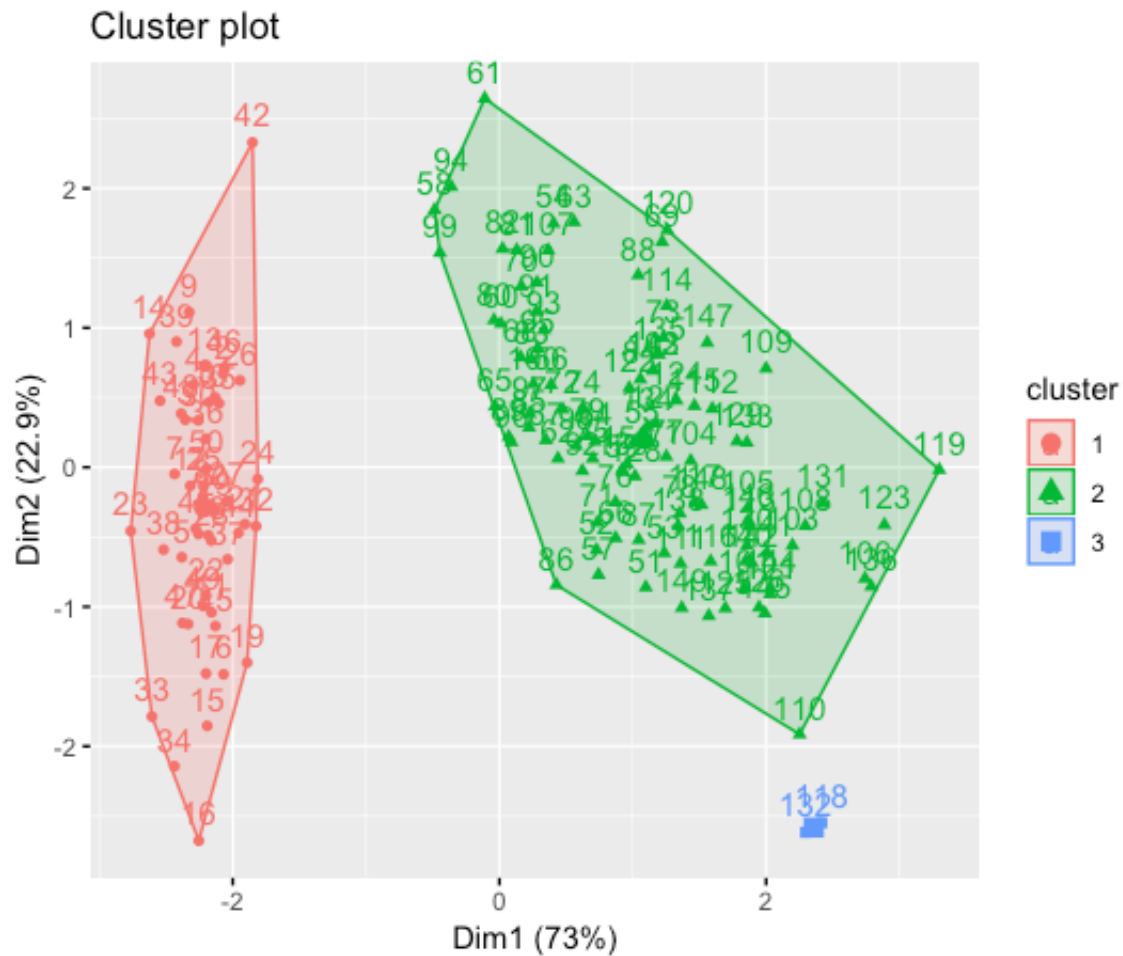
```
d3<-dist(data,method="euclidean")
hc3<- hclust(d3,method="single")
plot(hc3,main=" Dendrogram: Single Linkage ")
sub_grp3<-cutree(hc3,k=3)
fviz_cluster(list(data = data, cluster = sub_grp3))
count <- 0
for (i in 1:150){
  if (sub_grp3[i] == species[i]){
    count <- count + 1
  }
}
rate <- count/150
print(paste("Correct clustering rate =",rate))
```

"Correct clustering rate = 0.68"

Dendrogram: Single Linkage



```
d3
hclust (*, "single")
```

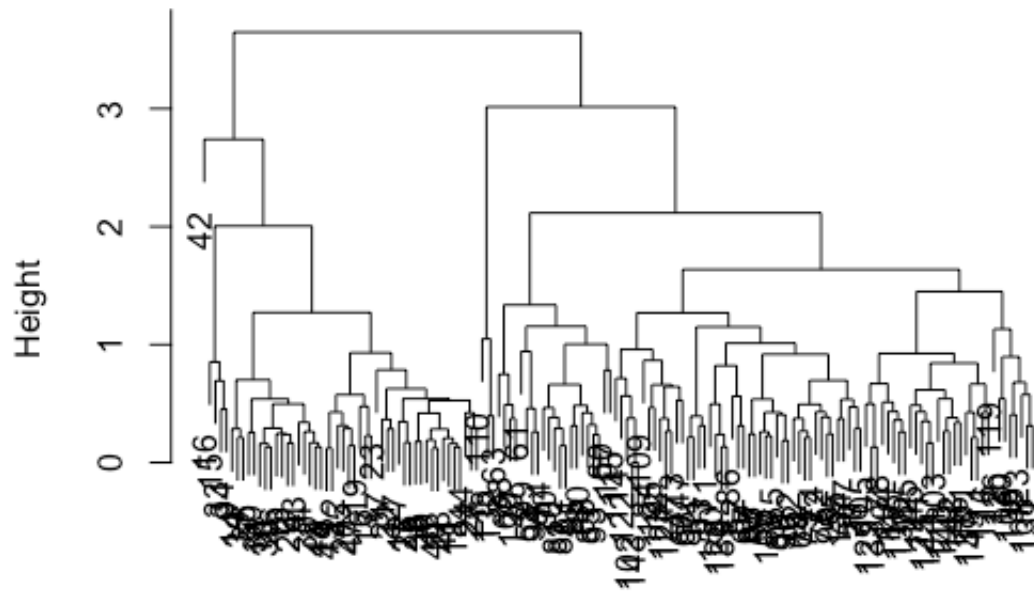


Problem 2C

```
data_standard <- scale(data, center = TRUE, scale = TRUE)
d4<-dist(data_standard,method="euclidean")
hc4<- hclust(d4,method="average")
plot(hc4,main=" Dendrogram: Average Linkage ")
sub_grp4<-cutree(hc4,k=3)
fviz_cluster(list(data = data_standard, cluster = sub_grp4))
count <- 0
for (i in 1:150){
  if (sub_grp4[i] == species[i]){
    count <- count + 1
  }
}
rate <- count/150
print(paste("Correct clustering rate =",rate))
```

"Correct clustering rate = 0.686666666666667"

Dendrogram: Average Linkage



d4
hclust (*, "average")

cluster

- 1
- 2
- 3

