Multivariate Hw7

106070020

2021年6月15日

```
library(poppr)
library(pegas)
library(ape)
library(adegenet)
library(ade4)
library(smacof)
library(ggplot2)
library(ggrepel)
library(dendextend)
```

```
Mydata <- read.table("C:/Users/eva/Desktop/作業 上課資料(清大)/大四下/多變量/hw7/Master_Pinus_data_genotype.txt",
                      header = TRUE, check.names = FALSE)
dim(Mydata) # 550 individuals x 3082 SNPs
```

```
## [1]  550 3086
```

```
ind <- as.character(Mydata$tree_id) # individual labels
population <- as.character(Mydata$state) # population labels
locus <- Mydata[, -c(1, 2, 3, 4, 105:ncol(Mydata))]
Mydata1 <- df2genind(locus, ploidy = 2, ind.names = ind, pop = population, sep="")
distgenDISS <- diss.dist(Mydata1, percent = FALSE, mat = F)
```

Q1

```
a<-mds(delta = distgenDISS , ndim = 2 , type = "ratio")
ggplot() + geom_point(data = as.data.frame(a$conf) , mapping = aes(x = D1, y = D2, color = population), alpha = 0.5) +
labs(title = "MDS configuration ")
```
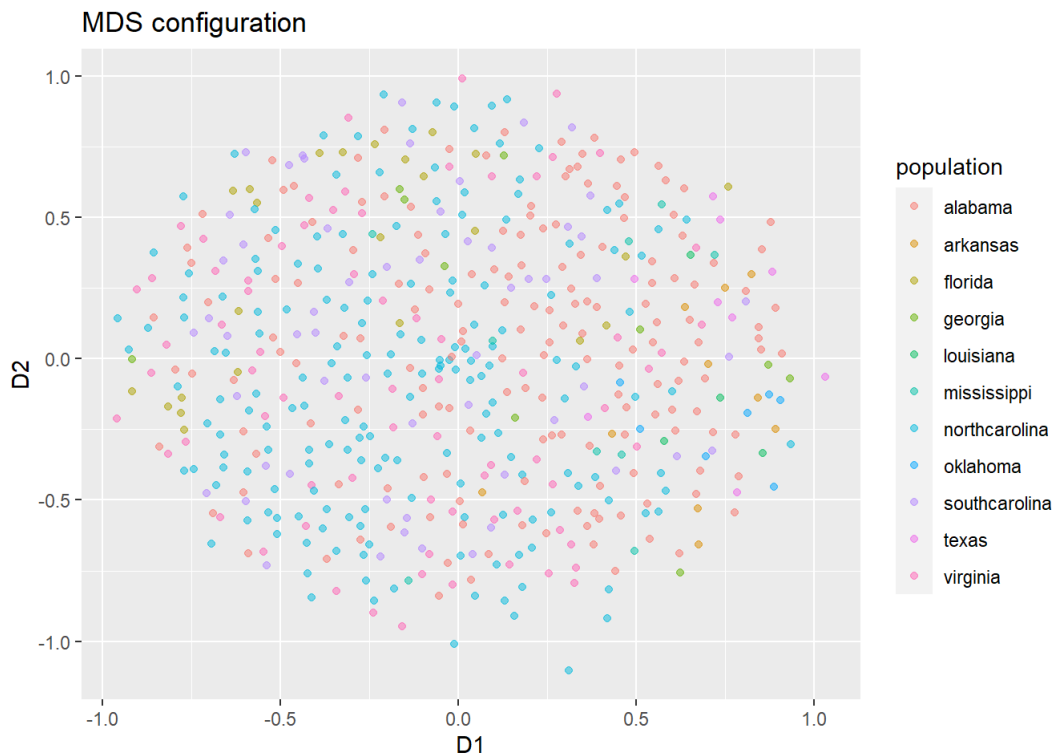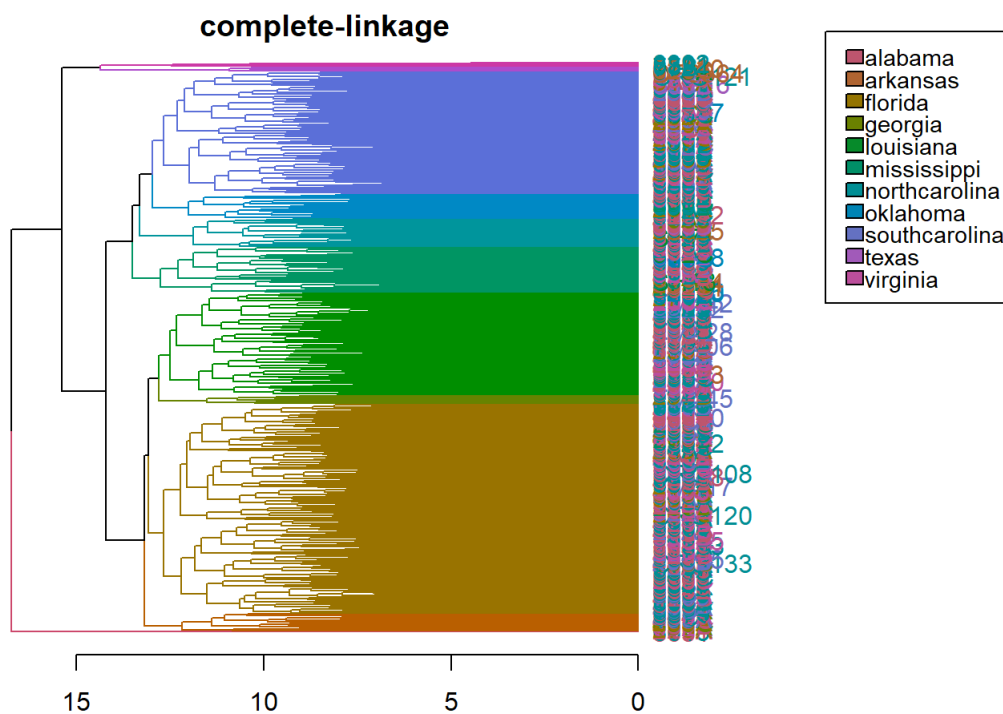

Q2(complele linkage)

```
distgenEUCL <- dist(Mydata1, method = "euclidean", diag = FALSE, upper = FALSE, p = 2)
hclust_comp<-hclust(distgenEUCL, method="complete")
dend  <- as.dendrogram(hclust_comp)
n_population <- length(unique(factor(population)))
cols_11 <- colorspace::rainbow_hcl(n_population, c = 70, l  = 50)
col_population <- cols_11[factor(population)]

# extra: showing the various clusters cuts
k234 <- cutree(dend, k = 2:4)

# color labels by car company:
labels_colors(dend) <- col_population[order.dendrogram(dend)]
# color branches based on cutting the tree into 4 clusters:
dend <- color_branches(dend, k = 11)

### plots
par(mar = c(4,1,1,12))
plot(dend, horiz = TRUE, main='complete-linkage')
legend(x = -5, y = 580, legend = levels(factor(population)), fill = cols_11, cex=0.9, bty='o',
       lty=0, y.intersp = 0.8, x.intersp =0.1, xpd=TRUE)
```


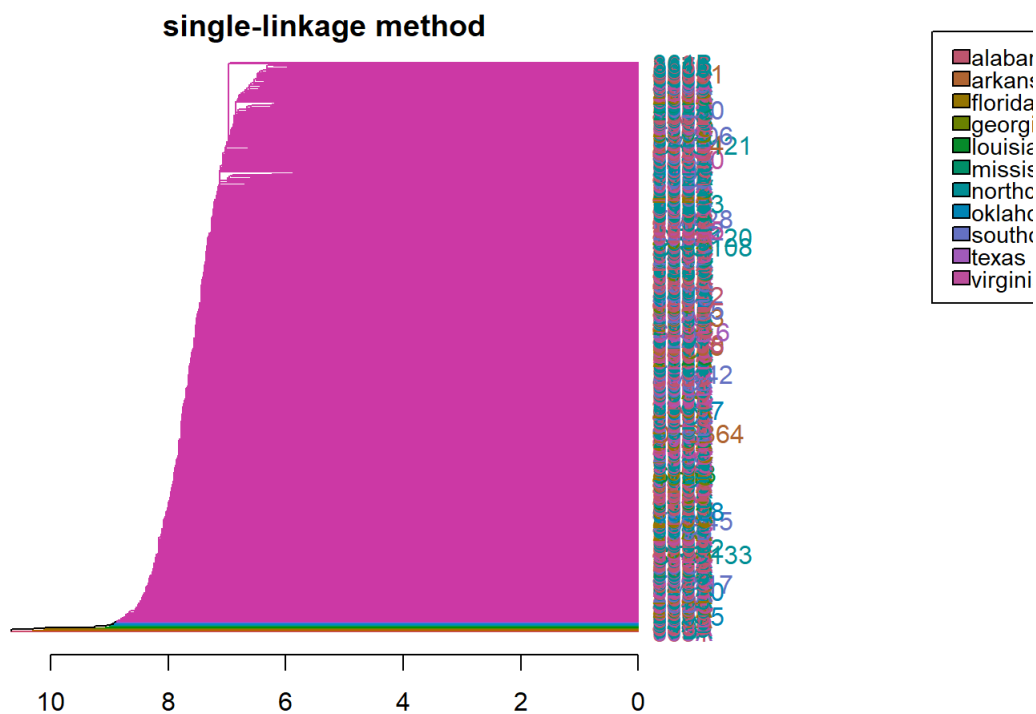**complete-linkage**

Q2(single linkage)

```
distgenEUCL <- dist(Mydata1, method = "euclidean", diag = FALSE, upper = FALSE, p = 2)
hclust_comp<-hclust(distgenEUCL, method="single")
dend  <- as.dendrogram(hclust_comp)
n_population <- length(unique(factor(population)))
cols_11 <- colorspace::rainbow_hcl(n_population, c = 70, l  = 50)
col_population <- cols_11[factor(population)]

# extra: showing the various clusters cuts
k234 <- cutree(dend, k = 2:4)

# color labels by car company:
labels_colors(dend) <- col_population[order.dendrogram(dend)]
# color branches based on cutting the tree into 4 clusters:
dend <- color_branches(dend, k = 11)

### plots
par(mar = c(4,1,1,12))
plot(dend, horiz = TRUE, main='single-linkage method')
legend(x = -5, y = 580, legend = levels(factor(population)), fill = cols_11, cex=0.9, bty='o',
       lty=0, y.intersp = 0.8, x.intersp =0.1, xpd=TRUE)
```


single-linkage method

Q2(ward's method linkage)

```
distgenEUCL <- dist(Mydata1, method = "euclidean", diag = FALSE, upper = FALSE, p = 2)
hclust_comp<-hclust(distgenEUCL, method="ward.D2")
dend  <- as.dendrogram(hclust_comp)
n_population <- length(unique(factor(population)))
cols_11 <- colorspace::rainbow_hcl(n_population, c = 70, l  = 50)
col_population <- cols_11[factor(population)]

# extra: showing the various clusters cuts
k234 <- cutree(dend, k = 2:4)

# color labels by car company:
labels_colors(dend) <- col_population[order.dendrogram(dend)]
# color branches based on cutting the tree into 4 clusters:
dend <- color_branches(dend, k = 11)

### plots
par(mar = c(4,1,1,12))
plot(dend, horiz = TRUE, main="Ward's Method")
legend(x = -5, y = 580, legend = levels(factor(population)), fill = cols_11, cex=0.9, bty='o',
       lty=0, y.intersp = 0.8, x.intersp =0.1, xpd=TRUE)
```