HW6-MATH4323
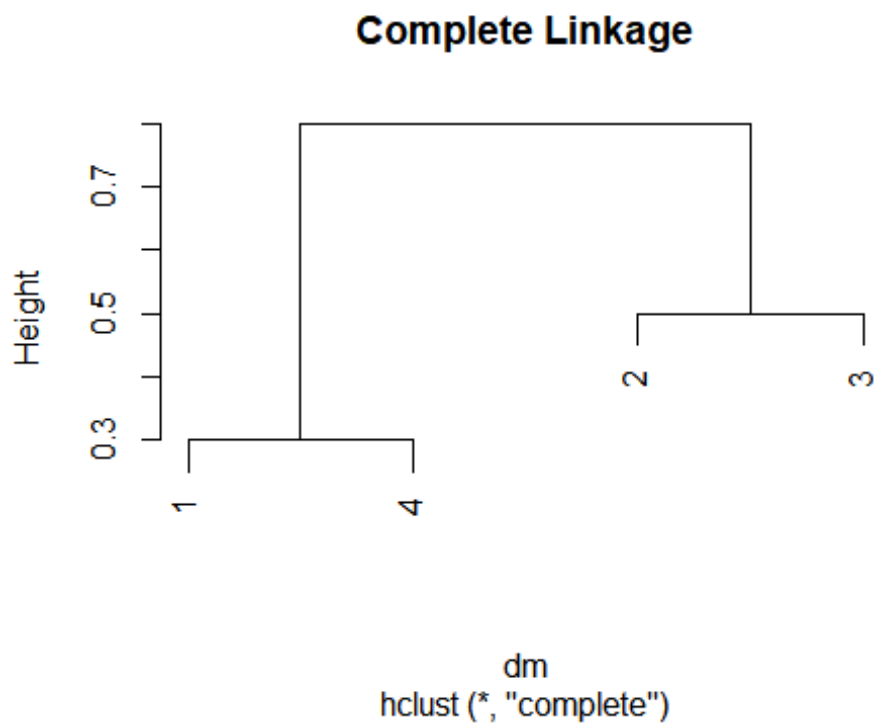
anthonycastillo ID:1670011

2022-11-09

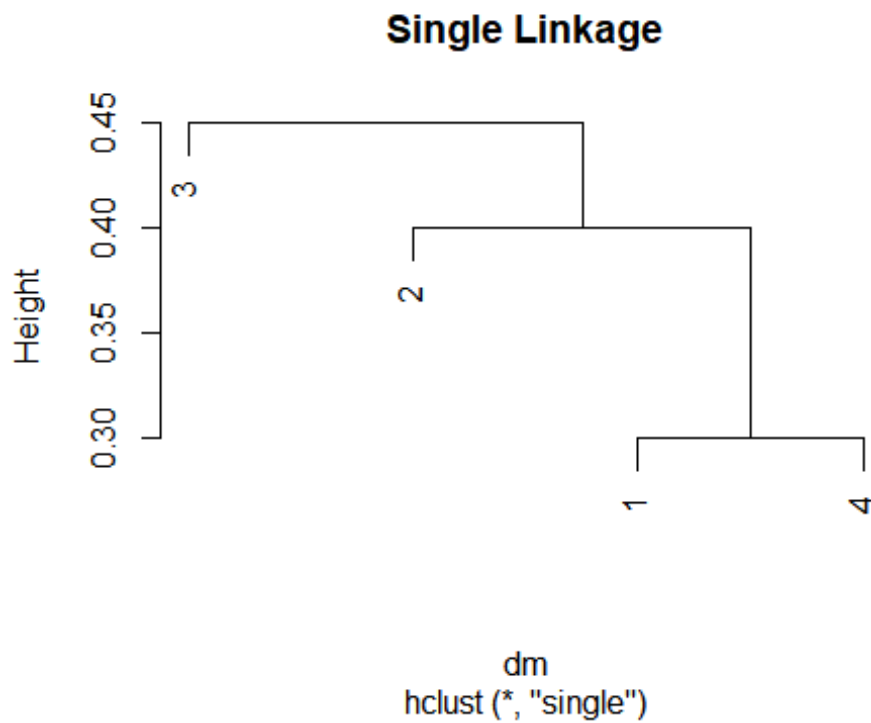Question 1 (a):

```
dm <- as.dist(matrix(c(0,0.4,0.75,0.3,
              0.4,0,0.5,0.8,
              0.75,0.5,0,0.45,
              0.3,0.8,0.45,0), nrow = 4))
hc.complete <- hclust(dm,
                      method = "complete")
plot(hc.complete,
     main = "Complete Linkage")
```


Question 1 (b):

```
hc.single <- hclust(dm,
                    method = "single")
plot(hc.single,
     main = "Single Linkage")
```

Single Linkage
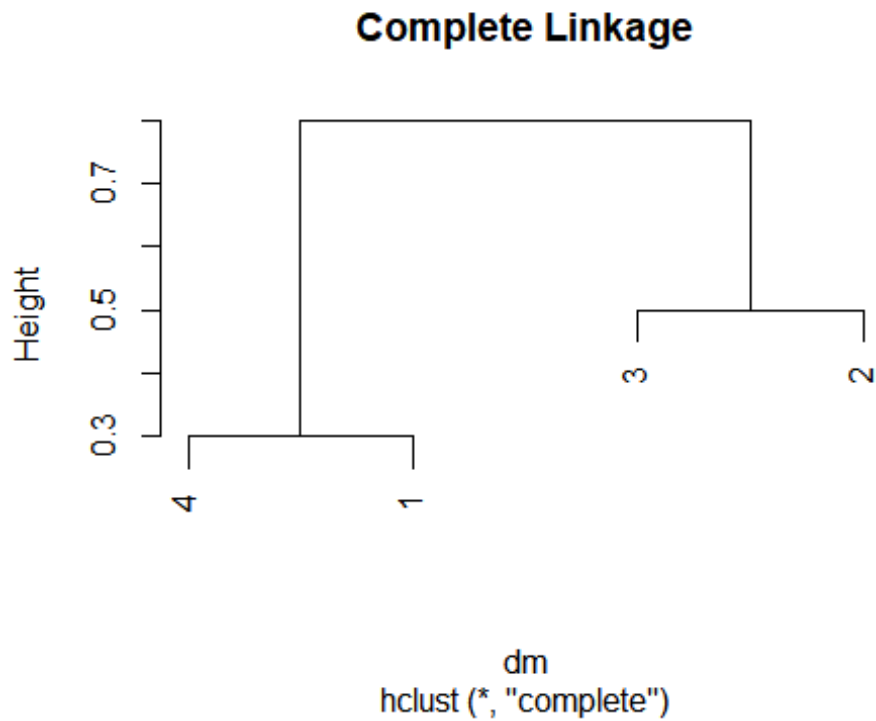

dm
hclust (*, "single")

Question 1 (c):

```
hc.compcut <- cutree(hc.complete, k=2)
hc.compcut
```

```
## [1] 1 2 2 1
```

Question 1 (d):

```
hc.singlecut <- cutree(hc.single, k=2)
hc.singlecut
```
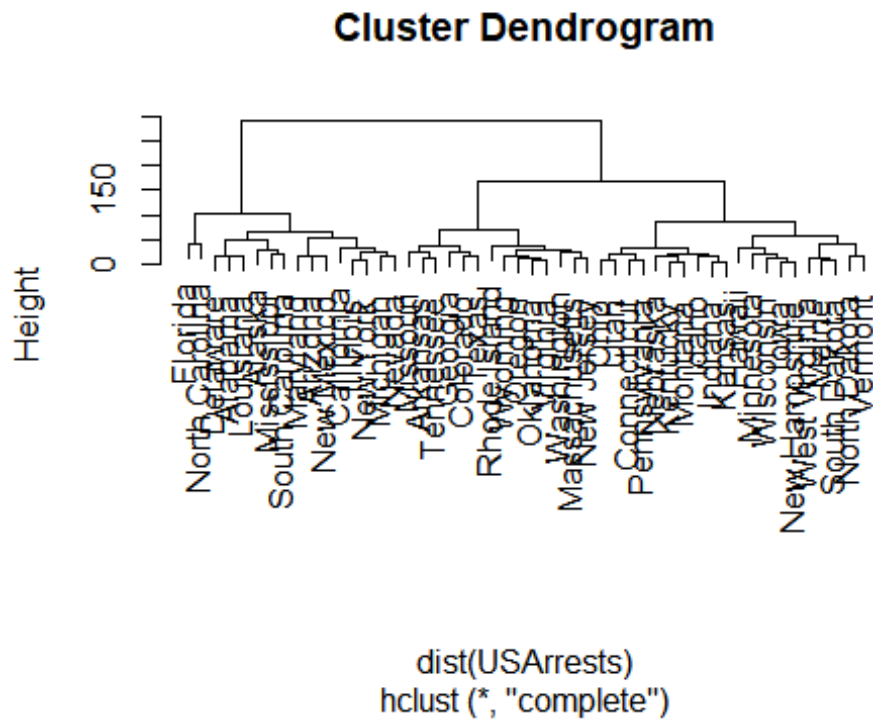
```
## [1] 1 1 2 1
```

Question 1 (e):

```
plot(hc.complete,
     main = "Complete Linkage",
     label = c(4,3,2,1))
```

Complete Linkage

dm
hclust (*, "complete")

Question 2 (a):

```r
x <- matrix(c(-0.6,-1,
              0.05,-1.1,
              -1.5,-0.4,
              -1.4,-1.55,
              1.3,-0.3,
              -0.9,-1.2,
              1.4,0,
              0.65,-0.25,
              -0.1,0.9),
            ncol=2,
            byrow=T)
lol <- dist(x)
ligma <- as.matrix(lol)
for(i in 1:9){
  num <- (min(ligma[i,5],ligma[i,7]))
  print(round(num,2))
}

## [1] 2.02
## [1] 1.48
## [1] 2.8
## [1] 2.98
## [1] 0
## [1] 2.38
```

```
## [1] 0
## [1] 0.65
## [1] 1.75
```

Question 2 (b):

```r
lol <- dist(x)
ligma <- as.matrix(lol)
for(i in 1:9){
  num <- (mean(ligma[i,5],ligma[i,7]))
  print(round(num,2))
}

## [1] 2.02
## [1] 1.48
## [1] 2.8
## [1] 2.98
## [1] 0
## [1] 2.38
## [1] 0.32
## [1] 0.65
## [1] 1.84
```

Question 3 (a):
The complete linkage dendogram will most of the time be higher on the tree since it uses maximum intercluster dissimilarity but there can be cases where the maximum intercluster dissimilarity and minimum intercluster dissimilarity could be the same distance and complete linkage and single linkage would fuse at the same height
Question 3 (b):
The fusion will occur at the same height since it is two single elements they always fuse at the same height no matter the type of linkage

Question 4 (a):

```r
us.hclust <- hclust(dist(USArrests),
                    method = "complete")
plot(us.hclust)
```

Cluster Dendrogram

dist(USArrests)
hclust (*, "complete")

Question 4 (b):

```
us.cutree <- cutree(us.hclust, k=3)
table(us.cutree)

## us.cutree
##  1  2  3
## 16 14 20

us.cutree

##        Alabama         Alaska        Arizona        Arkansas      California
##              1              1              1               2               1
##       Colorado    Connecticut       Delaware         Florida         Georgia
##              2              3              1               1               2
##         Hawaii          Idaho       Illinois         Indiana            Iowa
##              3              3              1               3               3
##         Kansas       Kentucky      Louisiana           Maine        Maryland
##              3              3              1               3               1
##  Massachusetts       Michigan      Minnesota     Mississippi        Missouri
##              2              1              3               1               2
##        Montana       Nebraska         Nevada   New Hampshire      New Jersey
##              3              3              1               3               2
##     New Mexico       New York North Carolina    North Dakota            Ohio
##              1              1              1               3               3
##       Oklahoma         Oregon   Pennsylvania    Rhode Island  South Carolina
```

```
##              2              2              3              2              1
##    South Dakota      Tennessee          Texas           Utah        Vermont
##              3              2              2              3              3
##        Virginia     Washington  West Virginia      Wisconsin        Wyoming
##              2              2              3              3              2
```

Question 4 (c):

```
us.scale <-scale(USArrests)
us.hclust <- hclust(dist(us.scale),
                    method = "complete")
plot(us.hclust)
```

Cluster Dendrogram

dist(us.scale)
hclust (*, "complete")
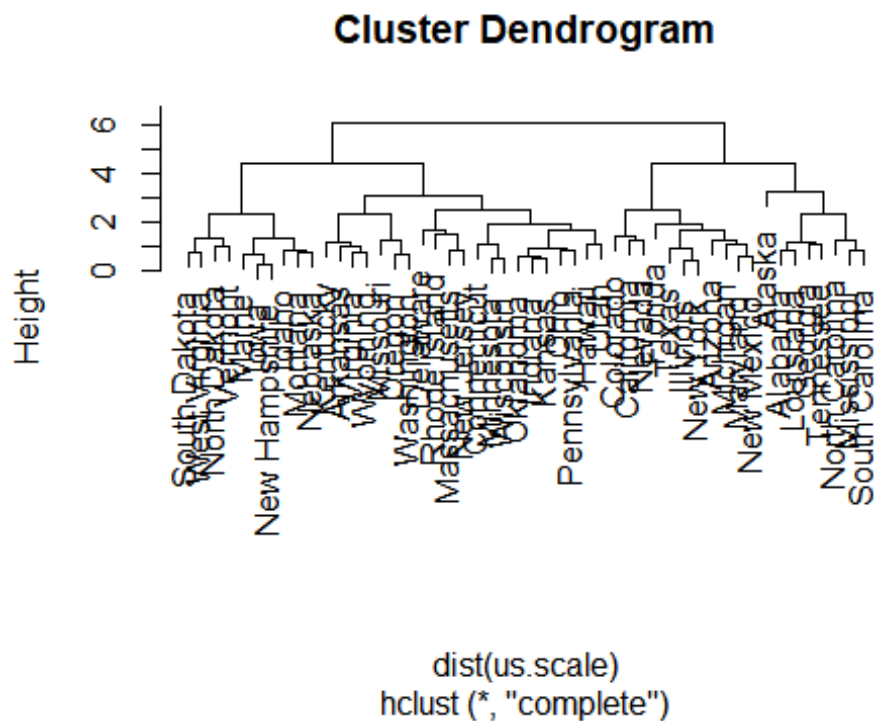
Question 4 (d):

```
library(factoextra)

## Loading required package: ggplot2

## Welcome! Want to learn more? See two factoextra-related books at
## https://goo.gl/ve3WBa

hc.out <- eclust(USArrests,
                 FUNcluster = "hclust",
                 hc_method = "complete")
```

```
## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use
"none" instead as
## of ggplot2 3.3.4.
## i The deprecated feature was likely used in the factoextra package.
##    Please report the issue at
<]8;;https://github.com/kassambara/factoextra/issueshttps://github.com/kassam
bara/factoextra/issues]8;;>.

fviz_nbclust(USArrests,hcut,
             hc_method = "complete")
```


Optimal number of clusters
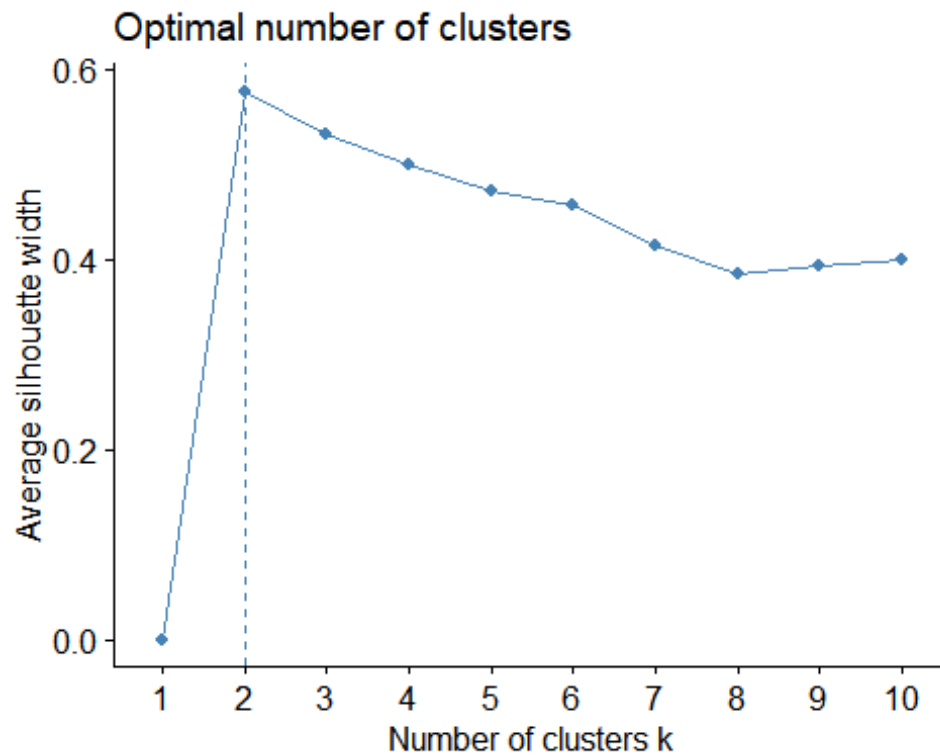
```
optimal.k <- cutree(us.hclust, k = 2)
```

The optimal k values seems to be k=2

Question 4 (e):

```
table(optimal.k)

## optimal.k
##  1  2
## 19 31

optimal.k
```

```
##      Alabama         Alaska        Arizona       Arkansas     California
##            1              1              1              2              1
##     Colorado    Connecticut       Delaware        Florida        Georgia
##            1              2              2              1              1
##       Hawaii          Idaho       Illinois        Indiana           Iowa
##            2              2              1              2              2
##       Kansas       Kentucky      Louisiana          Maine       Maryland
##            2              2              1              2              1
## Massachusetts       Michigan      Minnesota    Mississippi       Missouri
##            2              1              2              1              2
##      Montana       Nebraska         Nevada  New Hampshire     New Jersey
##            2              2              1              2              2
##   New Mexico       New York North Carolina   North Dakota           Ohio
##            1              1              1              2              2
##     Oklahoma         Oregon   Pennsylvania   Rhode Island South Carolina
##            2              2              2              2              1
## South Dakota      Tennessee          Texas           Utah        Vermont
##            2              1              1              2              2
##     Virginia     Washington  West Virginia      Wisconsin        Wyoming
##            2              2              2              2              2
```
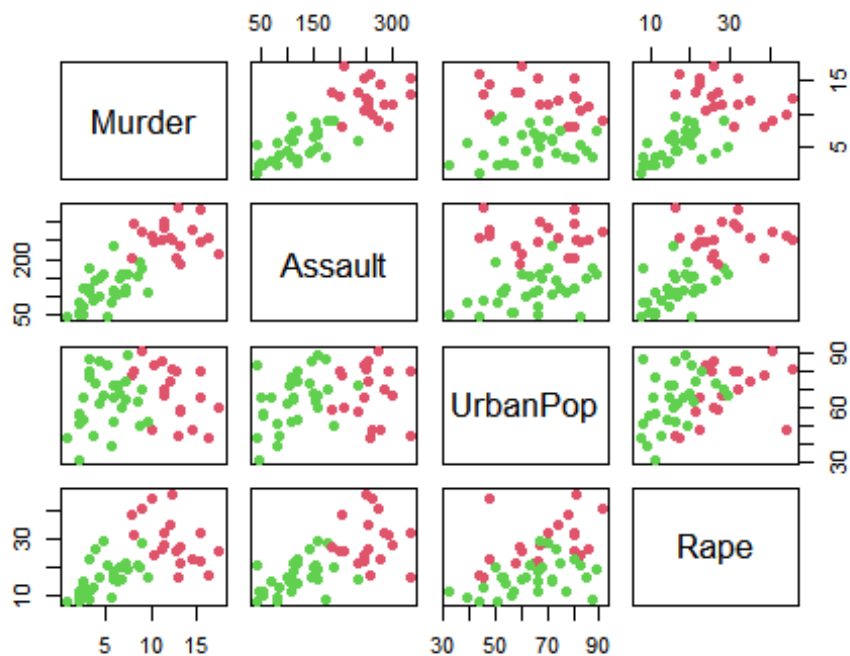
```
plot(us.hclust)
rect.hclust(us.hclust, k=2,border = 2:3)
```


Cluster Dendrogram

```
plot(USArrests, col = optimal.k+1,
     pch = 19, cex = 1)
```

Cluster 2 contains states with low urban pops and more high counts of rape,assault,and murder

Cluster 1 contains states with high urban population and low counts of rape,assault, and murder

Question 4 (f):
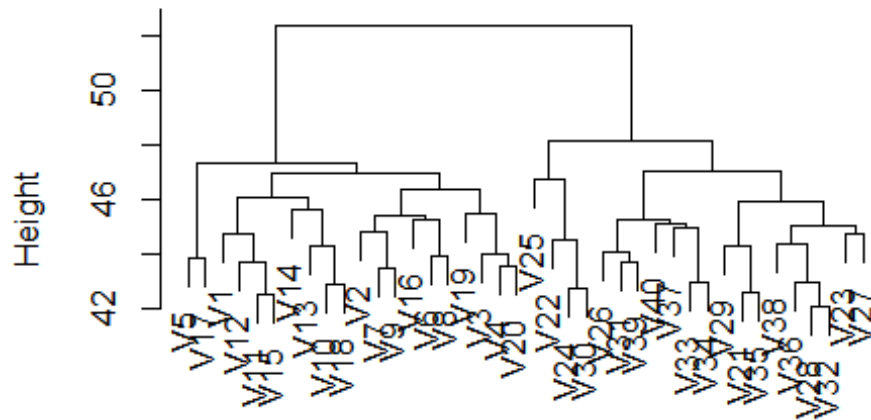The variables for the data has 2data types which are numeric and integer so yes scaling could help

Question 5 (a):

```
data <- read.csv("D:/fall22/math4323/Ch10Ex11.csv",header = F)
temp <- t(data)
temp <- data.frame(temp)
```

Question 5 (b):

```
#complete linkage
hc.completeData <- hclust(dist(temp),
                          method = "complete")
plot(hc.completeData,
     main = "Complete Linkage")
```
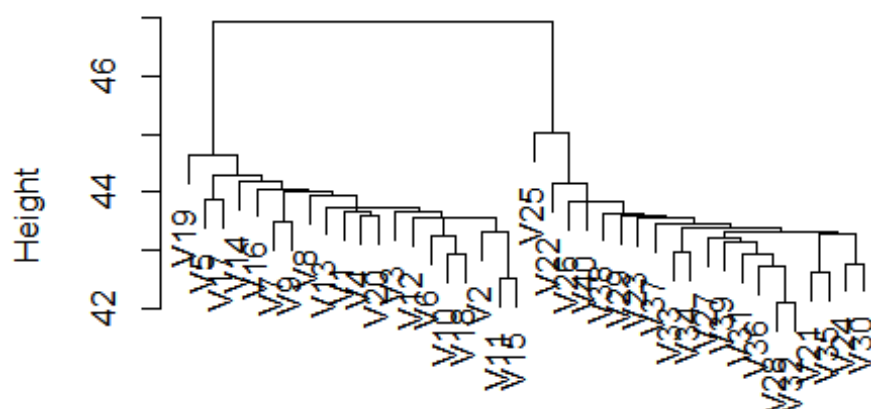
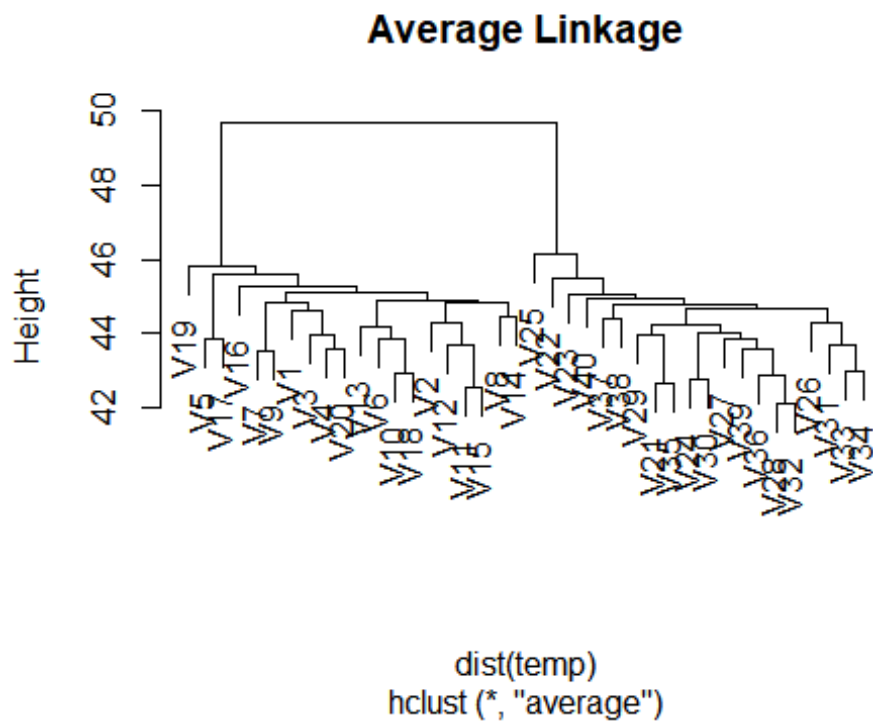## Complete Linkage

dist(temp)
hclust (*, "complete")

```r
#single linkage
hc.singleData <- hclust(dist(temp),
                        method = "single")
plot(hc.singleData,
     main = "Single Linkage")
```

Single Linkage

dist(temp)
hclust (*, "single")

```r
#average Linkage
hc.averageData <- hclust(dist(temp),
                         method = "average")
plot(hc.averageData,
     main = "Average Linkage")
```

Average Linkage

dist(temp)
hclust (*, "average")

Yes the genes separate into two main clusters for all linkages