Question10

Chonnikarn Charoenpanich

#Installing Biobase

```
library(Biobase)
```

```
## Loading required package: BiocGenerics
```

```
## Loading required package: parallel
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, parApply, parCapply, parLapply,
##     parLapplyLB, parRapply, parSapply, parSapplyLB
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, basename, cbind, colnames,
##     dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##     grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,
##     order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##     rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
##     union, unique, unsplit, which.max, which.min
```

```
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.
```

Question 10

Load the Montgomery and Pickrell eSet:

```
con =url("http://bowtie-bio.sourceforge.net/recount/ExpressionSets/montpick_eset.RData")
load(file=con)
close(con)
mp = montpick.eset
pdata=pData(mp)
edata=as.data.frame(exprs(mp))
fdata = fData(mp)
```

Cluster the samples using k-means clustering after applying the `log2`log2 transform (be sure to add 1). Set a seed for reproducible results (use `set.seed(1235)`set.seed(1235)). If you choose two clusters, do you get the same two clusters as you get if you use the `cutree`cutree function to cluster the samples into two groups? Which cluster matches most closely to the study labels?
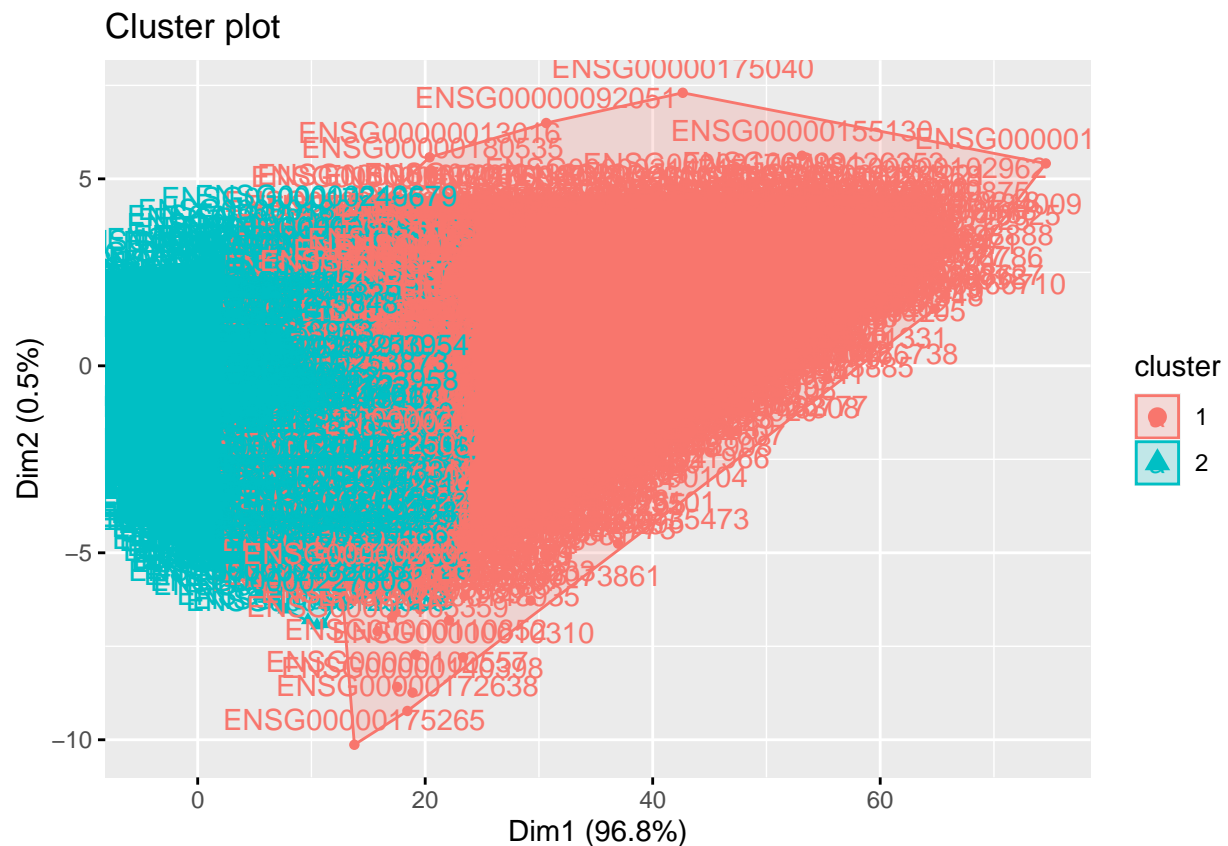
```
library(factoextra)
```

```
## Loading required package: ggplot2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```
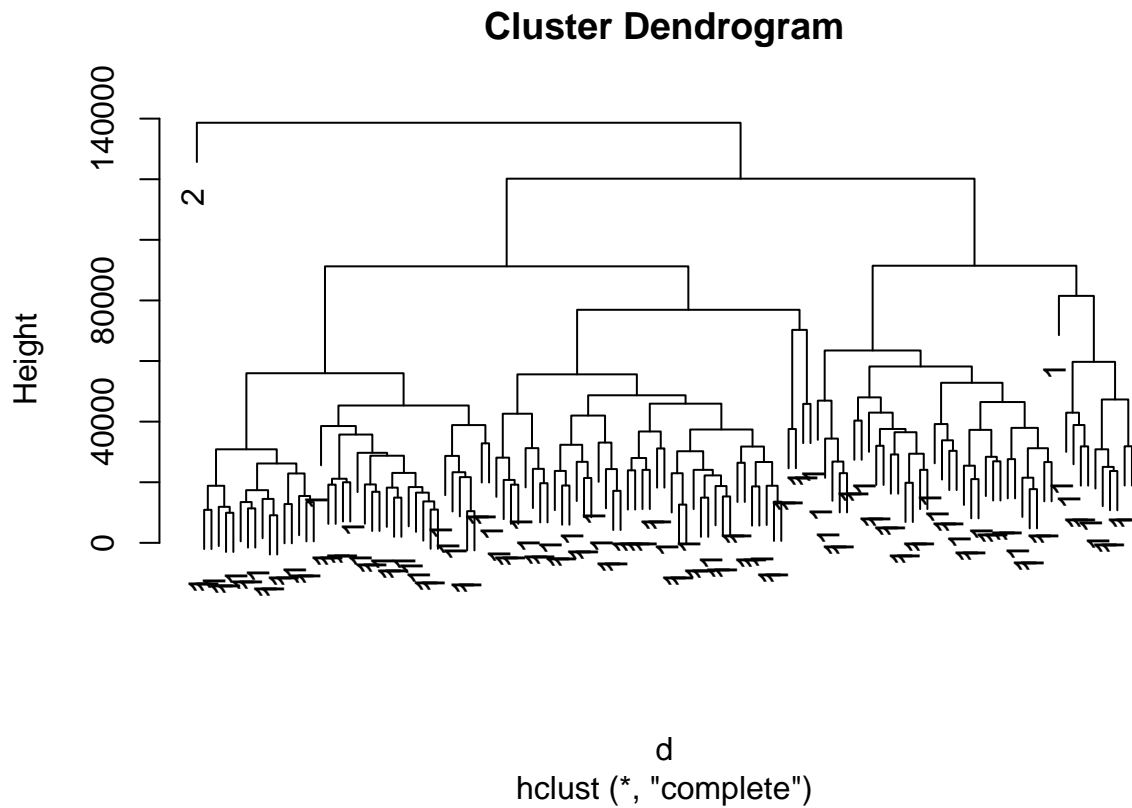
```
e = log2(edata + 1)
set.seed(1235)
k = kmeans(e, centers = 2)
factoextra::fviz_cluster(k, data = e)
```

```
d = dist(t(edata))
h = hclust(d)
tree = cutree(h, 2)
plot(h, tree)
```

Cluster Dendrogram

d
hclust (*, "complete")

They produce different answers, with hierarchical clustering giving a much more unbalanced clustering. The k-means clustering matches study better.