

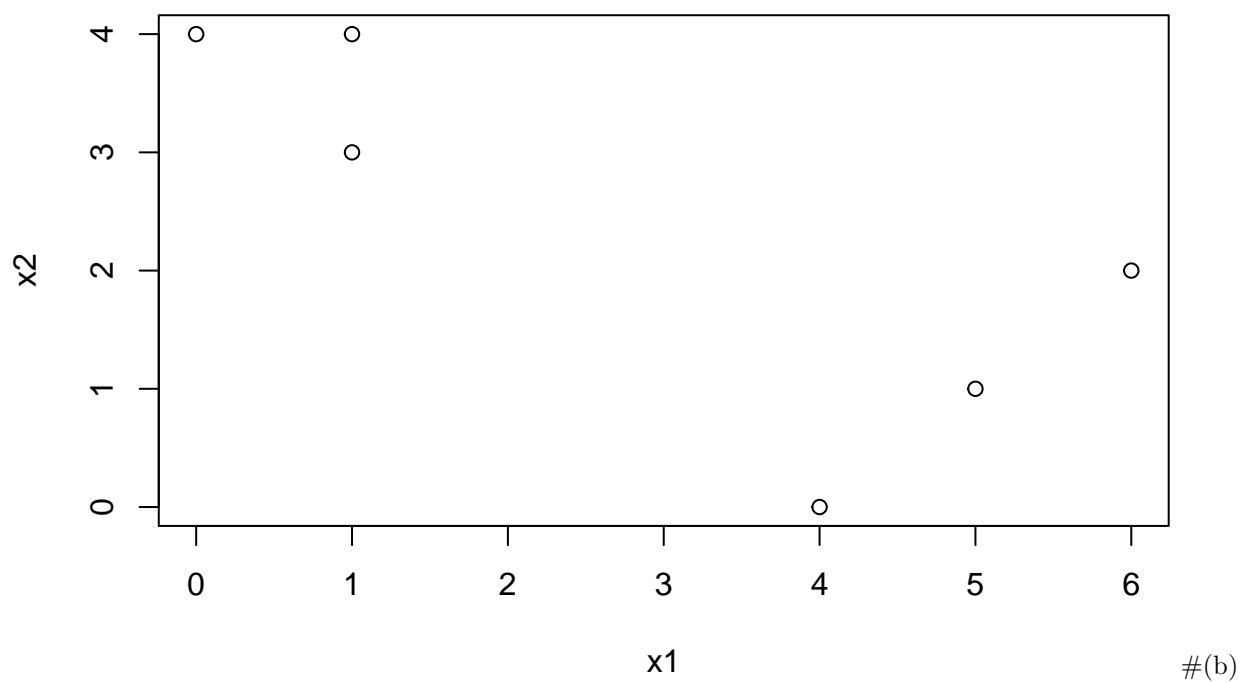
# ISLR-HW8

Chaoqun Yin

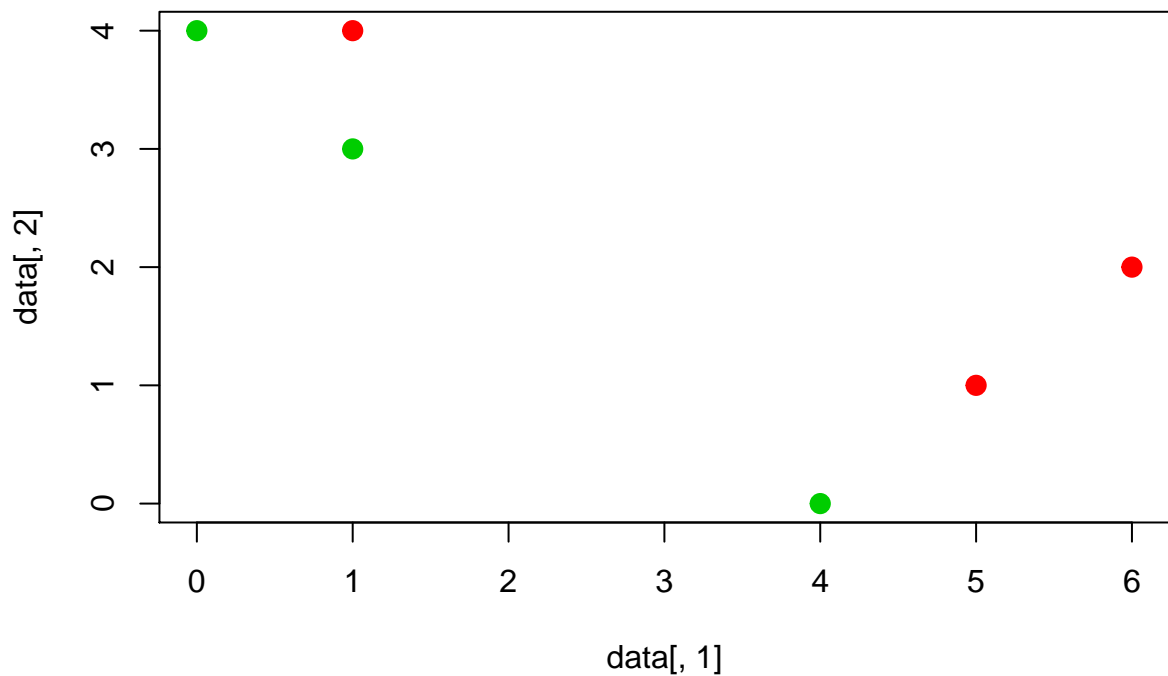
4/30/2019

## 10.3(a)

```
x1<-c(1, 1, 0, 5, 6, 4)
x2<-c(4, 3, 4, 1, 2, 0)
data<-data.frame(t(rbind(x1,x2)))
plot(x1,x2)
```

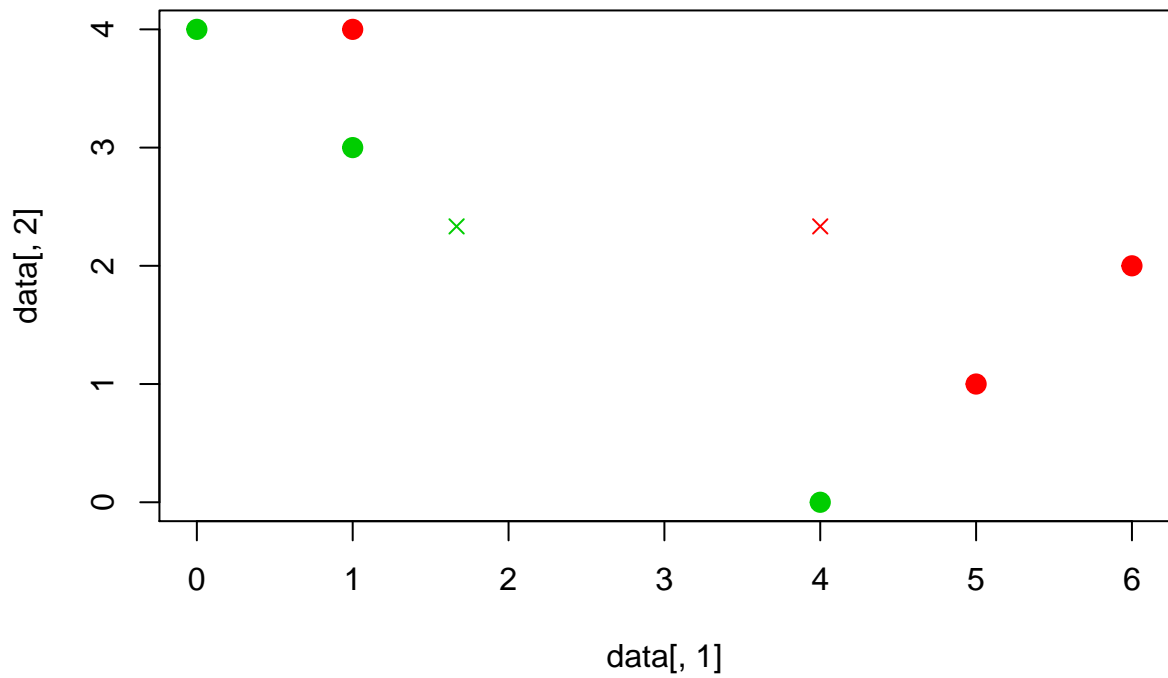


```
set.seed(5)
lab <- sample(2, nrow(data), replace = T)
plot(data[, 1], data[, 2], col = (lab + 1), pch = 20, cex = 2)
```



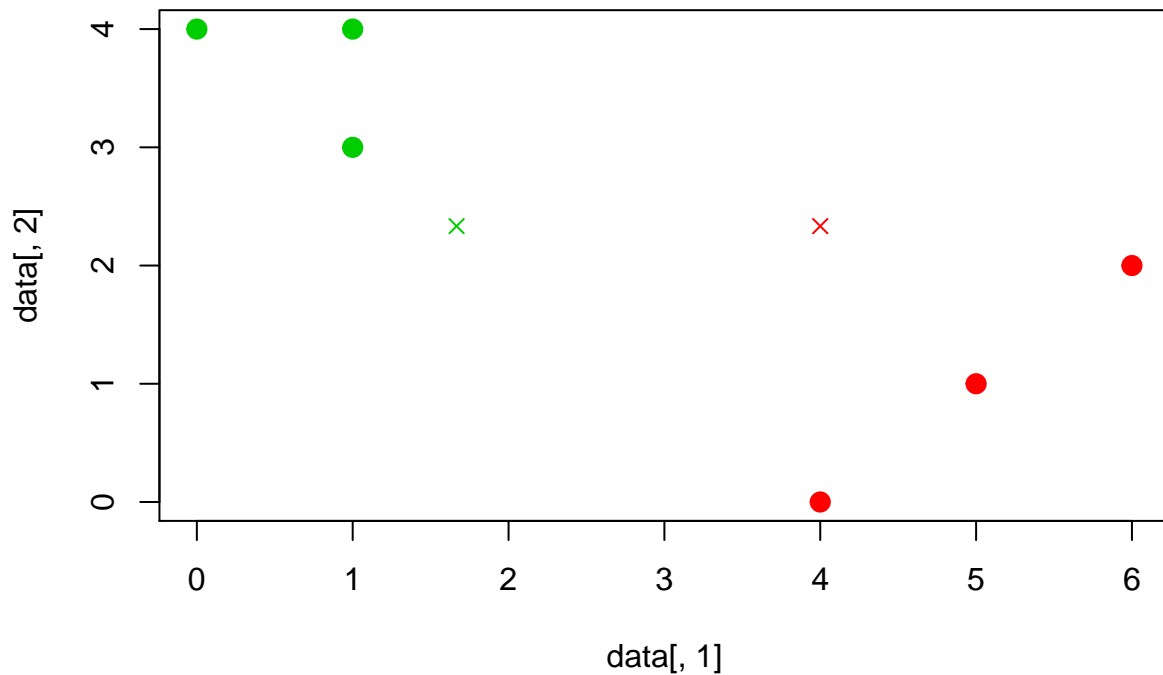
#(c)

```
centroid1 <- c(mean(data[lab == 1, 1]), mean(data[lab == 1, 2]))
centroid2 <- c(mean(data[lab == 2, 1]), mean(data[lab == 2, 2]))
plot(data[,1], data[,2], col=(lab + 1), pch = 20, cex = 2)
points(centroid1[1], centroid1[2], col = 2, pch = 4)
points(centroid2[1], centroid2[2], col = 3, pch = 4)
```



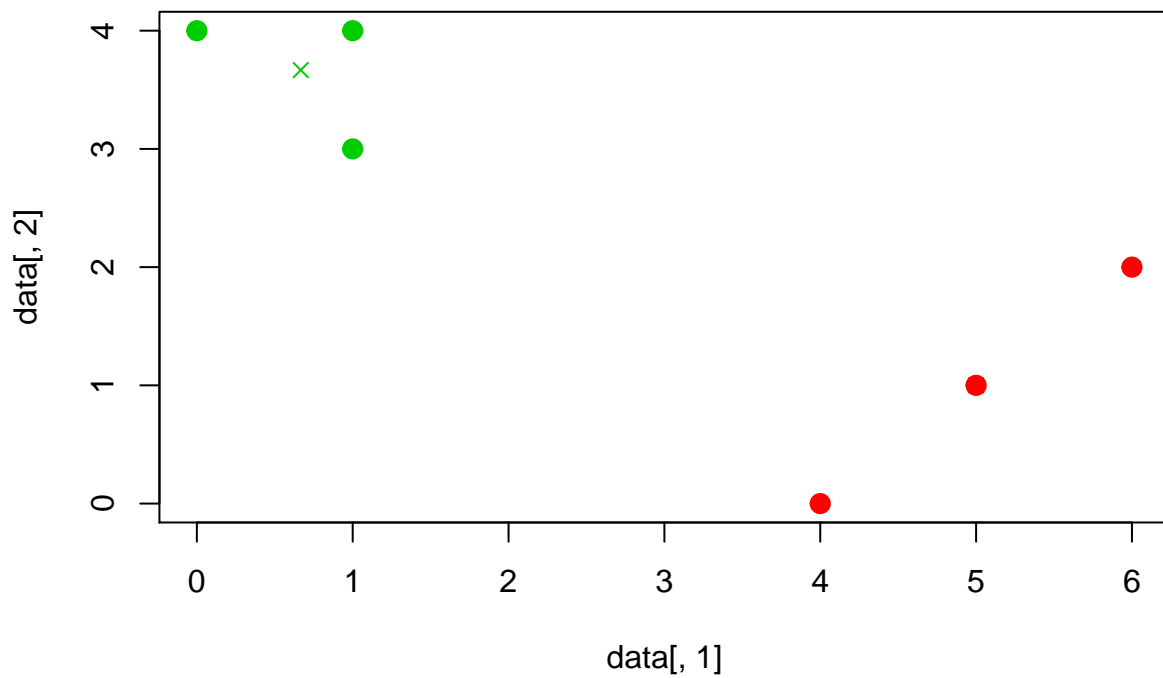
(d)

```
labels <- c(2,2,2, 1, 1, 1)
plot(data[, 1], data[, 2], col = (labels + 1), pch = 20, cex = 2)
points(centroid1[1], centroid1[2], col = 2, pch = 4)
points(centroid2[1], centroid2[2], col = 3, pch = 4)
```



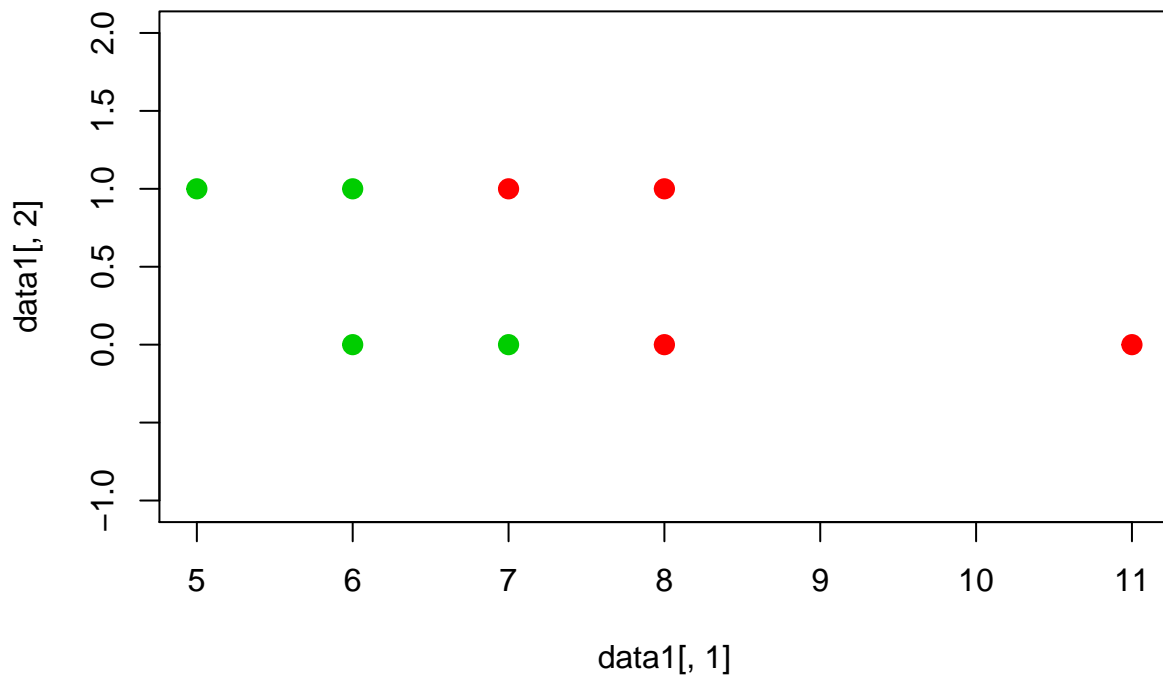
#(e)

```
centroid1 <- c(mean(data[labels == 1, 1]), mean(data[labels == 1, 2]))
centroid2 <- c(mean(data[labels == 2, 1]), mean(data[labels == 2, 2]))
plot(data[,1], data[,2], col=(labels + 1), pch = 20, cex = 2)
points(centroid1[1], centroid1[2], col = 2, pch = 4)
points(centroid2[1], centroid2[2], col = 3, pch = 4)
```



## 10.5

```
socks <- c(8, 11, 7, 6, 5, 6, 7, 8)
computers <- c(0, 0, 0, 0, 1, 1, 1, 1)
#situation one
data1 <- cbind(socks, computers)
labels <- c(1, 1, 2, 2, 2, 2, 1, 1)
plot(data1[, 1], data1[, 2], col=(labels + 1), pch = 20, cex = 2, asp = 1)
```

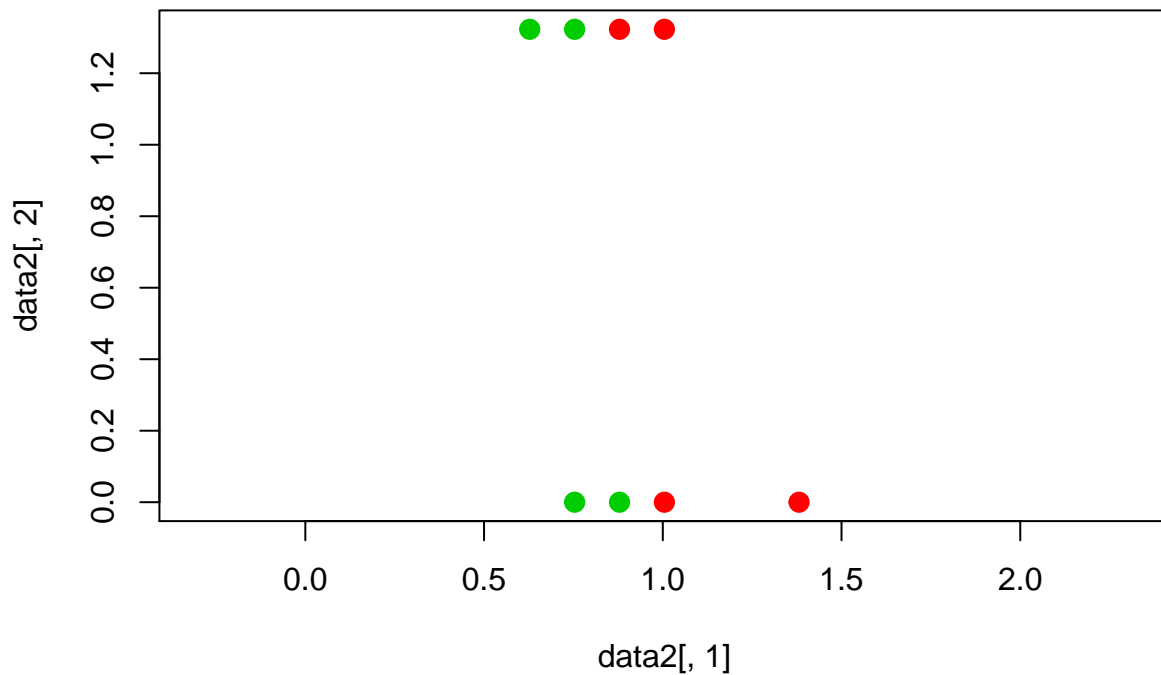


```
#situation two
```

```
data2 <- cbind(scale(socks, center = FALSE), scale(computers, center = FALSE))
```

```
labels <- c(1, 1, 2, 2, 2, 2, 1, 1)
```

```
plot(data2[, 1], data2[, 2], col=(labels + 1), pch = 20, cex = 2, asp = 1)
```



10.8(a)

```
store <- prcomp(USArrests, scale = TRUE)
```

```
pr.var <- store$sdev^2
```

```
pve <- pr.var / sum(pr.var)
```

```
pve
```

```
## [1] 0.62006039 0.24744129 0.08914080 0.04335752
```

```
sum(pr.var)
```

```
## [1] 4
```

(b)

```
rot <- store$rotation
```

```
sumvar <- sum(apply(as.matrix(scale(USArrests))^2, 2, sum))
```

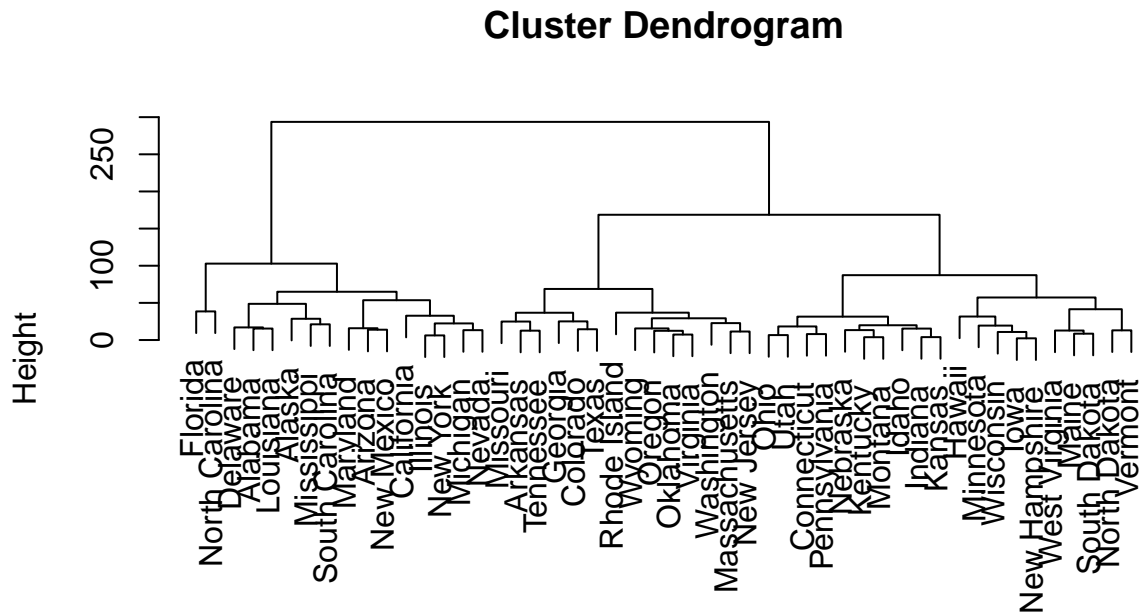
```
apply((as.matrix(scale(USArrests)) %*% rot)^2, 2, sum) / sumvar
```

```
##          PC1          PC2          PC3          PC4
```

```
## 0.62006039 0.24744129 0.08914080 0.04335752
```

## 10.9(a)

```
set.seed(5)
complete <- hclust(dist(USArrests), method = "complete")
plot(complete)
```



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

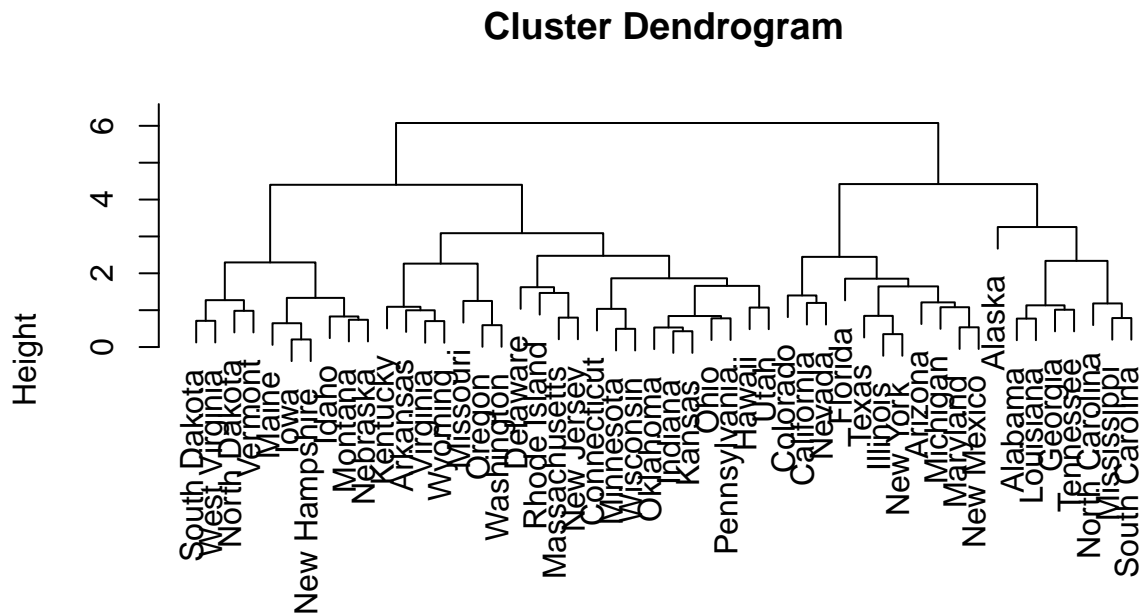
#(b)

```
cutree(complete, 3)
```

##	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

(c)

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



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

#(d)

```
cutree(complete.sd, 3)
```

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

```
table(cutree(complete, 3), cutree(complete.sd, 3))
```

```
##
##      1  2  3
##    1  6  9  1
##    2  2  2 10
##    3  0  0 20
```

## 10.10(a)

```
set.seed(5)
df<- matrix(rnorm(20 * 3 * 50, mean = 0, sd = 0.001), ncol = 50)
df[1:20, 1] <- 2
df[21:40, 1] <- 2
df[21:40, 2] <- 1
df[41:60, 2] <- 2
labels <- c(rep(1, 20), rep(2, 20), rep(3, 20))
```

## (b)

```
pr.out <- prcomp(df)
#plot(pr.out$df[, 1:2], col =(labels + 1), pch = 19)
```

## (c)

```
km1 <- kmeans(df, 3, nstart = 20)
table(labels, km1$cluster)
```

```
##
## labels  1  2  3
##      1  0 20  0
##      2  0  0 20
##      3 20  0  0
```

## (d)

```
km2<- kmeans(df, 2, nstart = 20)
table(labels, km2$cluster)
```

```
##
## labels  1  2
##      1  0 20
##      2  0 20
##      3 20  0
```



(e)

```
km3 <- kmeans(df, 3, nstart = 20)
table(labels, km3$cluster)
```

```
##
## labels  1  2  3
##        1  0  0 20
##        2 20  0  0
##        3  0 20  0
```