#7

In [1]:
```
library(ISLR)
```

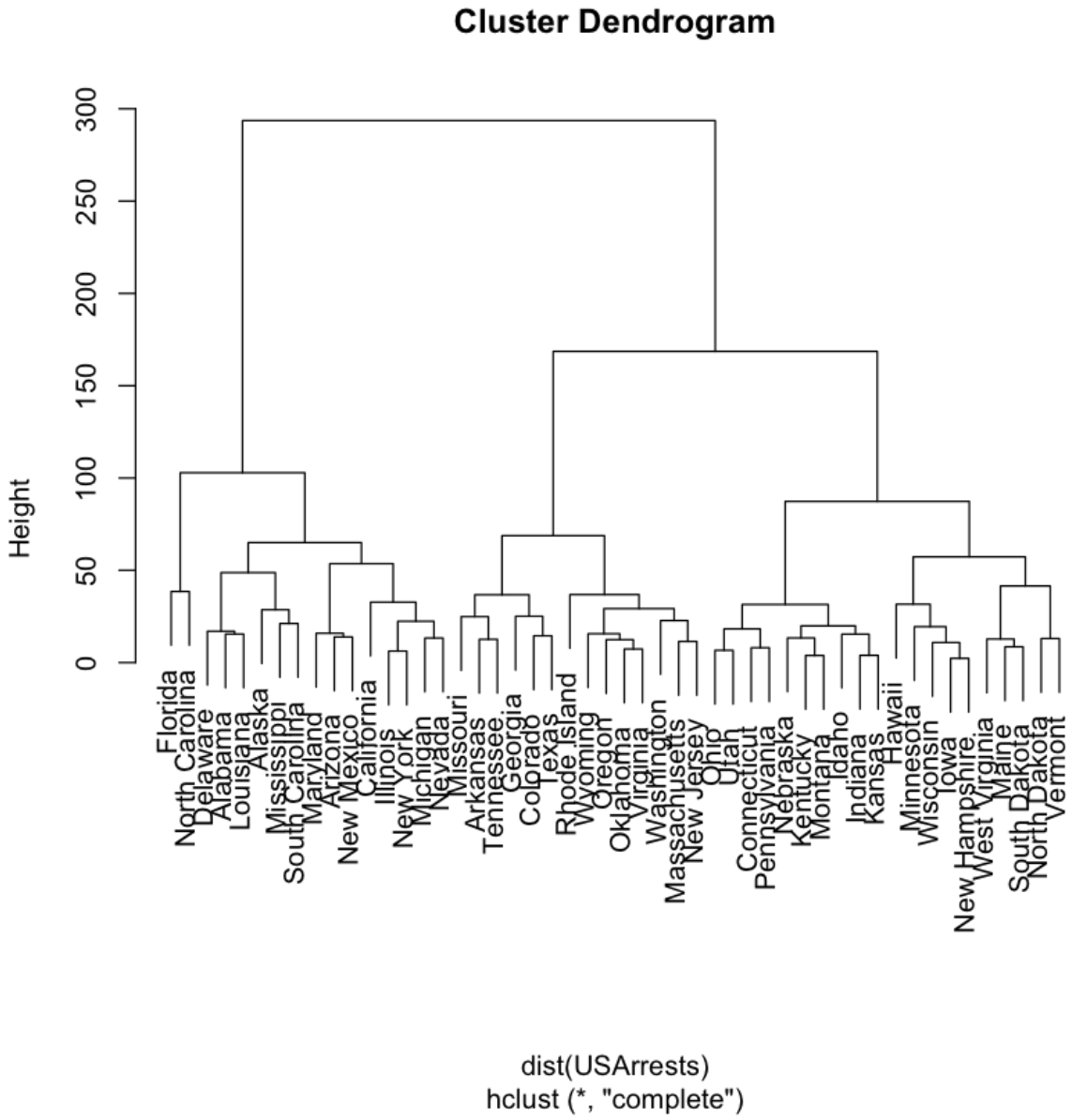
In [2]:
```
set.seed(1)
dsc = scale(USArrests)
a = dist(dsc)^2
b = as.dist(1 - cor(t(dsc)))
summary(b/a)
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
0.000086 0.069135 0.133943 0.234193 0.262589 4.887686 
```

#9(a)

In [3]:
```
library(ISLR)
set.seed(2)
```

In [4]:
```
hc.complete = hclust(dist(USArrests), method="complete")
plot(hc.complete)
```



9(b)

In [5]:
```
cutree(hc.complete,3)
```

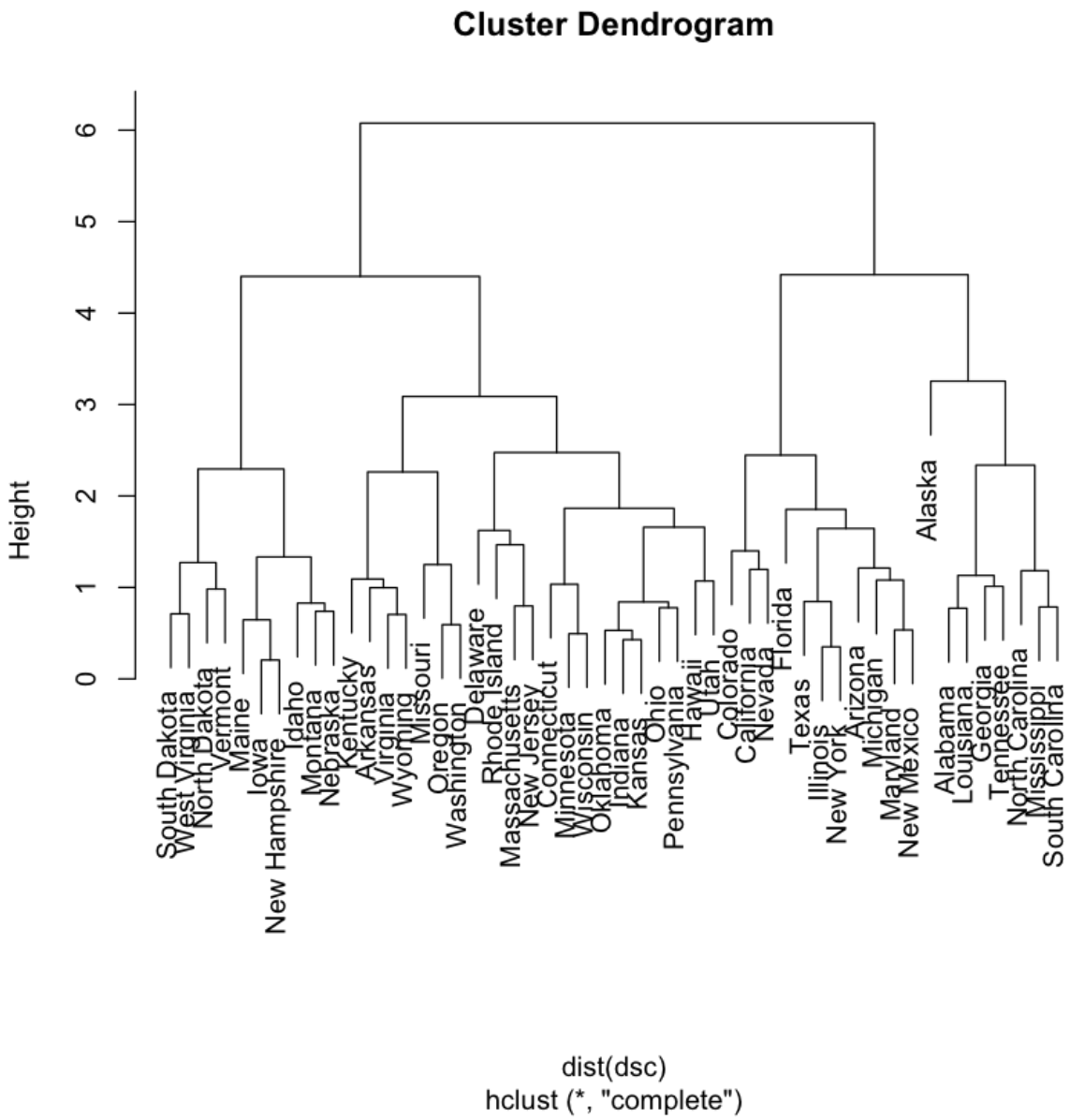
Alabama	1
Alaska	1
Arizona	1
Arkansas	2
California	1

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

West Virginia	3
Wisconsin	3
Wyoming	2

9(c)

In [6]:
```
dsc = scale(USArrests)
hc.s.complete = hclust(dist(dsc), method="complete")
plot(hc.s.complete)
```



9(d)

In [7]:
```
1 cutree(hc.s.complete,3)
```

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

Pennsylvania	3
Rhode Island	3
South Carolina	1
South Dakota	3
Tennessee	1
Texas	2
Utah	3
Vermont	3
Virginia	3
Washington	3
West Virginia	3
Wisconsin	3
Wyoming	3

In [8]:
```
table(cutree(hc.s.complete,3))
```

```
 1  2  3
 8 11 31
```

In [9]:
```
table(cutree(hc.s.complete, 3), cutree(hc.complete, 3))
```

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

#10(a)

In [10]:
```
set.seed(5)
x = matrix(rnorm(20*3*50, mean=0, sd=0.001), ncol=50)
x[1:20, 2] = 1
x[21:40, 1] = 2
x[21:40, 2] = 2
x[41:60, 1] = 1
```

#10(b)

In [11]:
```
pca.out = prcomp(x)
summary(pca.out)


pca.out$x[,1:2]

plot(pca.out$x[,1:2], col=2:4, xlab="Z1", ylab="Z2", pch=19)
```

```
Importance of components:
                          PC1    PC2      PC3      PC4      PC5
PC6
Standard deviation     1.0085 0.5823 0.001834 0.001698 0.001675 0.001
633
Proportion of Variance 0.7499 0.2500 0.000000 0.000000 0.000000 0.000
000
Cumulative Proportion  0.7499 1.0000 0.999970 0.999970 0.999970 0.999
970
                            PC7      PC8      PC9     PC10     PC11
PC12
Standard deviation     0.001575 0.001548 0.001476 0.001439 0.001405 0
.001345
Proportion of Variance 0.000000 0.000000 0.000000 0.000000 0.000000 0
.000000
Cumulative Proportion  0.999980 0.999980 0.999980 0.999980 0.999980 0
.999980
                           PC13     PC14    PC15     PC16     PC17
PC18
Standard deviation     0.001311 0.001265 0.00124 0.001196 0.001193 0
```

10(c)

In [12]:
```
km.out = kmeans(x, 3, nstart=20)
table(km.out$cluster, c(rep(1,20), rep(2,20), rep(3,20)))
```

```
   1  2  3
1 20  0  0
2  0 20  0
3  0  0 20
```

10(d)

In [13]:
```
km.out = kmeans(x, 2, nstart=20)
km.out$cluster
```

```
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2
```

10(e)

In [14]:
```
km.out = kmeans(x, 4, nstart=20)
km.out$cluster
```

```
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 4 4 4 2 2 4 2 4 4 2 4 2 2 2
4 4 4 4 4 4
```

#10(g)

In [15]:
```
km.out = kmeans(scale(x), 3, nstart=20)
km.out$cluster

km.out = kmeans(pca.out$x[,1:2], 3, nstart=20)
```

```
2 3 3 1 2 3 3 3 3 3 3 1 3 1 3 3 1 3 2 1 3 1 1 3 2 2 1
2 2 3 2 2 2 2 2 1 2 2 2 2 1 2 2 1 3 3 1 3 1 1 3 1 1 1
2 3 3 3 2 2
```

In [ ]:
```

```