

individual assignment 11

student ID:474084

(a)

```
library(ISLR)
attach(USArrests)

set.seed(1)
us.complete=hclust(dist(USArrests),method="complete")
```

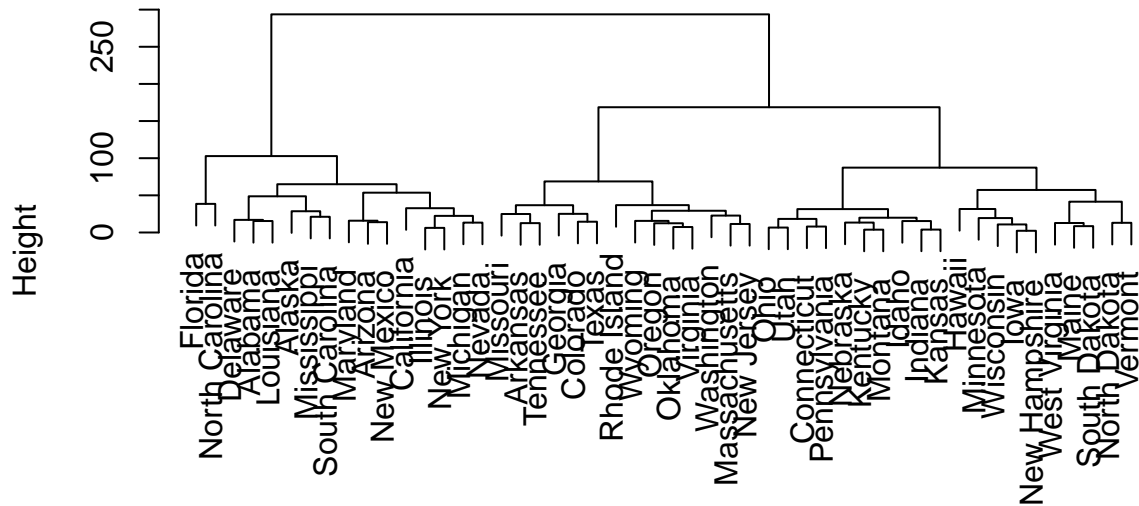
(b)

```
cutree(us.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

```
plot(us.complete)
```

Cluster Dendrogram



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

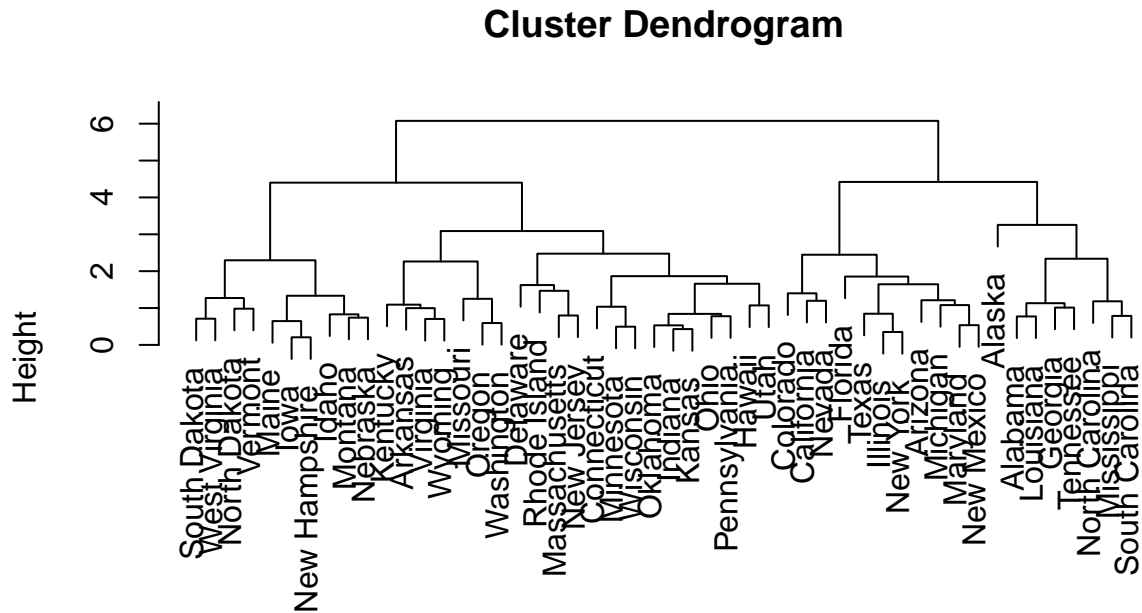
In this chunk, I cut the dendrogram at a height that results in three distinct clusters. The specific cluster of every state is shown in the plot.

(c)

```
sus=scale(USArrests)
sus.complete=hclust(dist(sus),method="complete")
cutree(sus.complete,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

```
plot(sus.complete)
```



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

In this chunk, I scale the variables to get the new clusters and plot it.

(d)

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

```
##
##  1  2  3
## 16 14 20
```

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

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

```
table(cutree(us.complete, 3), cutree(sus.complete, 3))
```

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

Scaling influences the maximum height the dendrogram can obtain. Before scaling, the biggest height is more than 250, and after scaling, it is 6. Although the tree looks similar to some extent, the clusters are affected by the scaling. As for my opinion, it should use scaling, because the unit of every variable is different, therefore using scaling is more accurate.