Analytics 512 Homework #9

Arif Ali

05/01/16

Exercise 10.7 #2

Please see the last page of the PDF

Exercise 10.7 #4

Part A

The maximal and minimal intercluster dissimilarities are unknown, so it can not be determined whether the fusion from complete linkage with be greater than the fusion from single or if the two are equal.

Part B

Linkage does not affect individual observations because there is no way to compute a maximal and minimal intercluster with only one possible match so the height should be the same

Exercise 10.7 #9

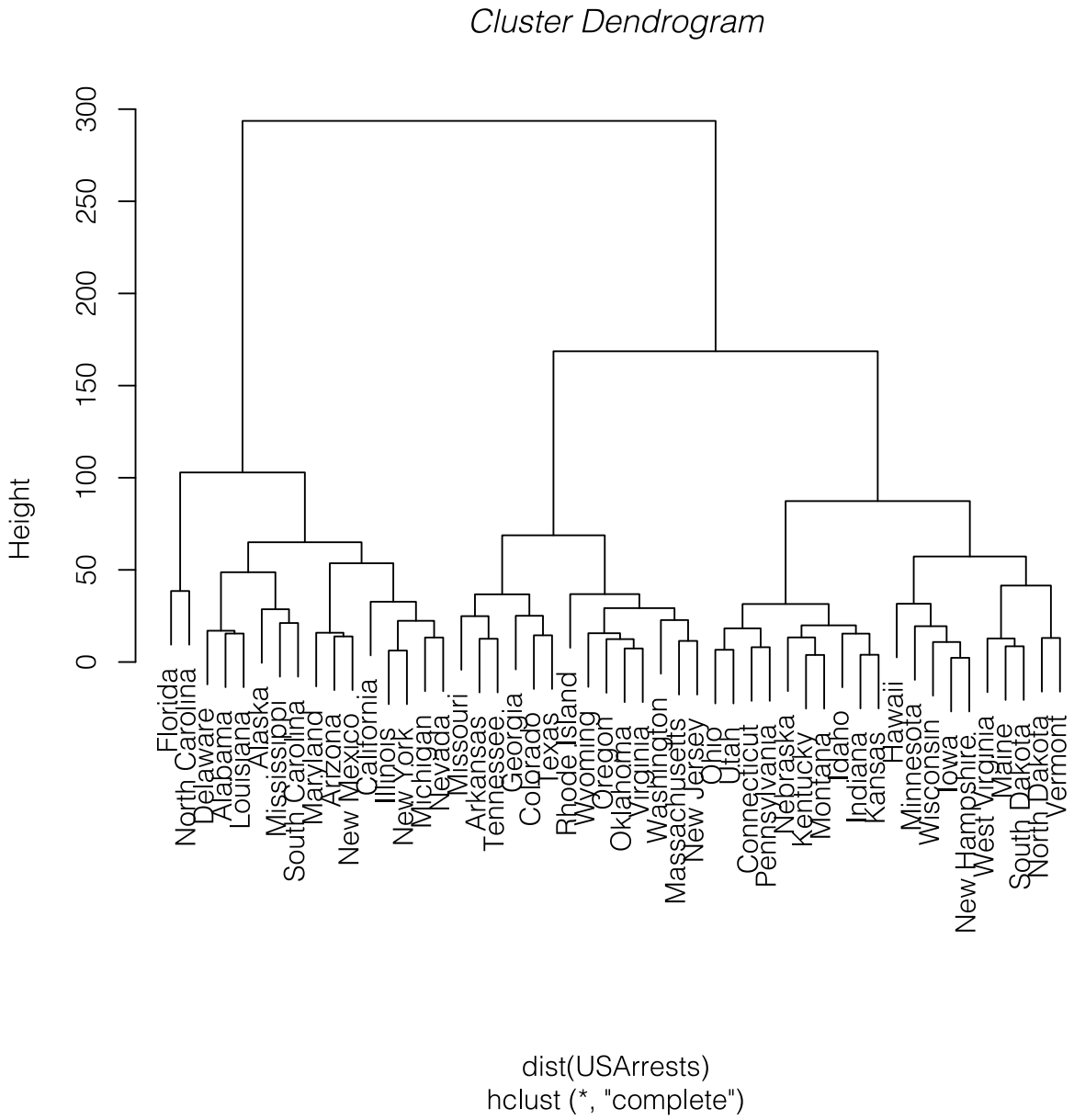
In [4]:
```
library(ISLR)
attach(USArrests)
```

```
The following objects are masked from USArrests (pos = 3):

    Assault, Murder, Rape, UrbanPop
```

Part A

In [7]:
```
part_a = hclust(dist(USArrests))
plot(part_a)
```



Part B

In [21]:
```
part_b = cutree(part_a,3)
states = names(part_b)
#Group 1
states[part_b==1]
#Group 2
states[part_b==2]
#Group 3
states[part_b==3]
```

Out[21]:
'Alabama' 'Alaska' 'Arizona' 'California' 'Delaware' 'Florida' 'Illinois' 'Louisiana' 'Maryland' 'Michigan' 'Mississippi' 'Nevada' 'New Mexico' 'New York' 'North Carolina' 'South Carolina'

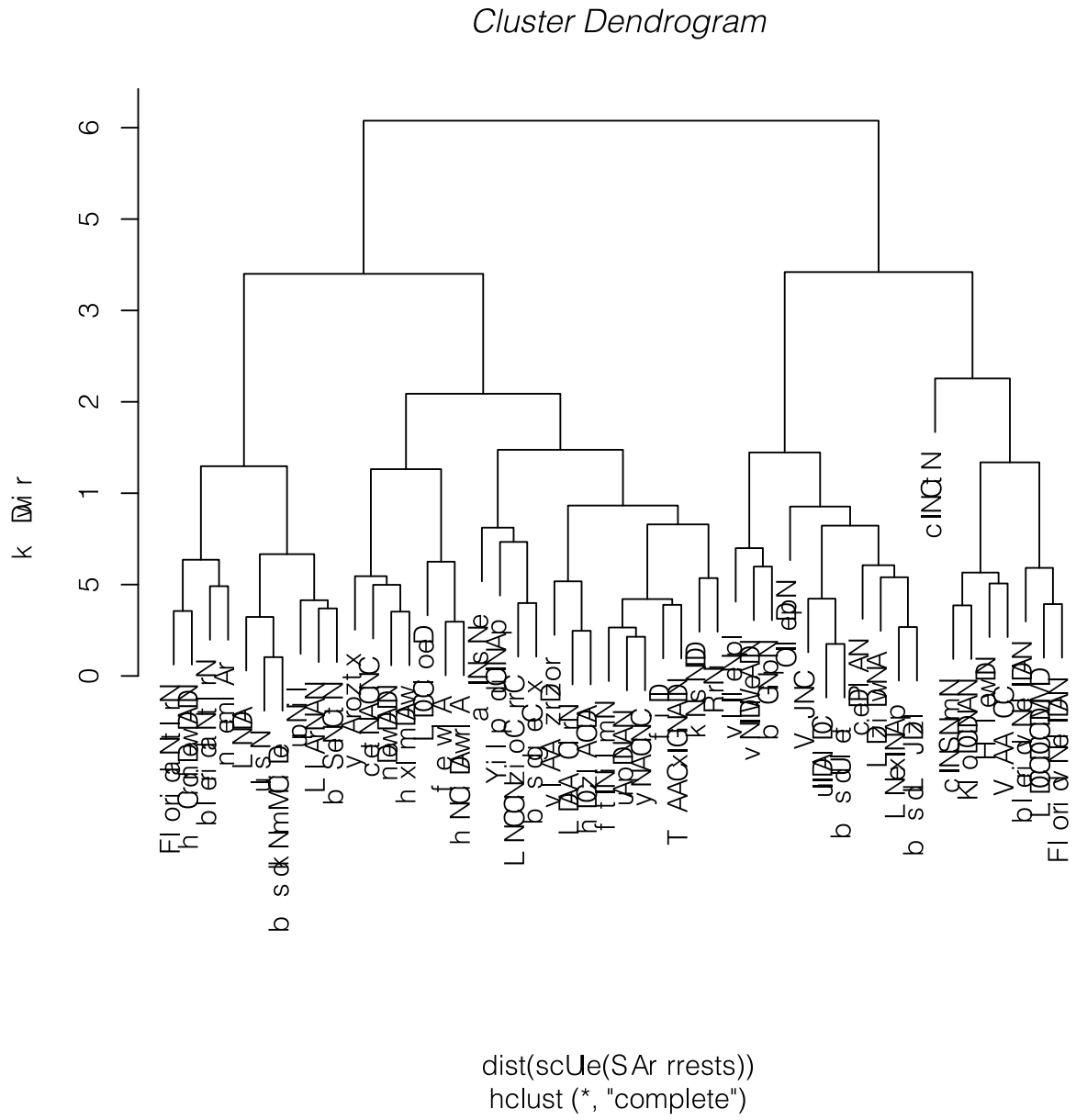
Out[21]:
'Arkansas' 'Colorado' 'Georgia' 'Massachusetts' 'Missouri' 'New Jersey' 'Oklahoma' 'Oregon' 'Rhode Island' 'Tennessee' 'Texas' 'Virginia' 'Washington' 'Wyoming'

Out[21]:
'Connecticut' 'Hawaii' 'Idaho' 'Indiana' 'Iowa' 'Kansas' 'Kentucky' 'Maine' 'Minnesota' 'Montana' 'Nebraska' 'New Hampshire' 'North Dakota' 'Ohio' 'Pennsylvania' 'South Dakota' 'Utah' 'Vermont' 'West Virginia' 'Wisconsin'

Part C

In [25]:
```
part_c = hclust(dist(scale(USArrests)))
```

In [26]:
```
plot(part_c)
```



Part D

In [27]:
```
part_d = cutree(part_c,3)
```

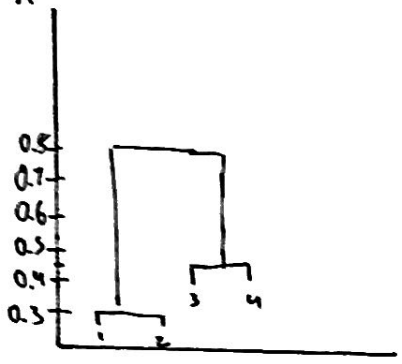
In [29]:
```
table(part_d == part_b)
```

Out[29]:
```
FALSE  TRUE
   22    28
```

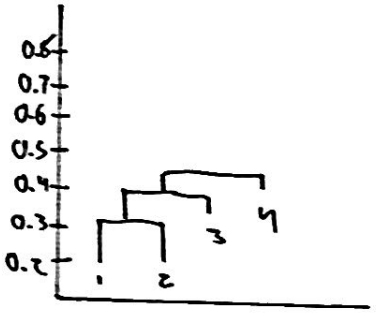
The one thing that initially stands out is Alaska. By scaling the data, Alaska seems to be significiantly further from any of the observations, which could indicate a very different rate of crime, which is something to be more concerned of as opposed to the whole numbers as crime rates are usually what most individuals are concerned about.

Exercise 2

Part A



Part B



Part C

Cluster 1: $\{1,2\}$

Cluster 2: $\{3,4\}$

Part D

Cluster 1: $\{1,2,3\}$

Cluster 2: $\{4\}$

Part E