Individual Assignment 10

Zixiao Wu

2022-11-30

R Markdown

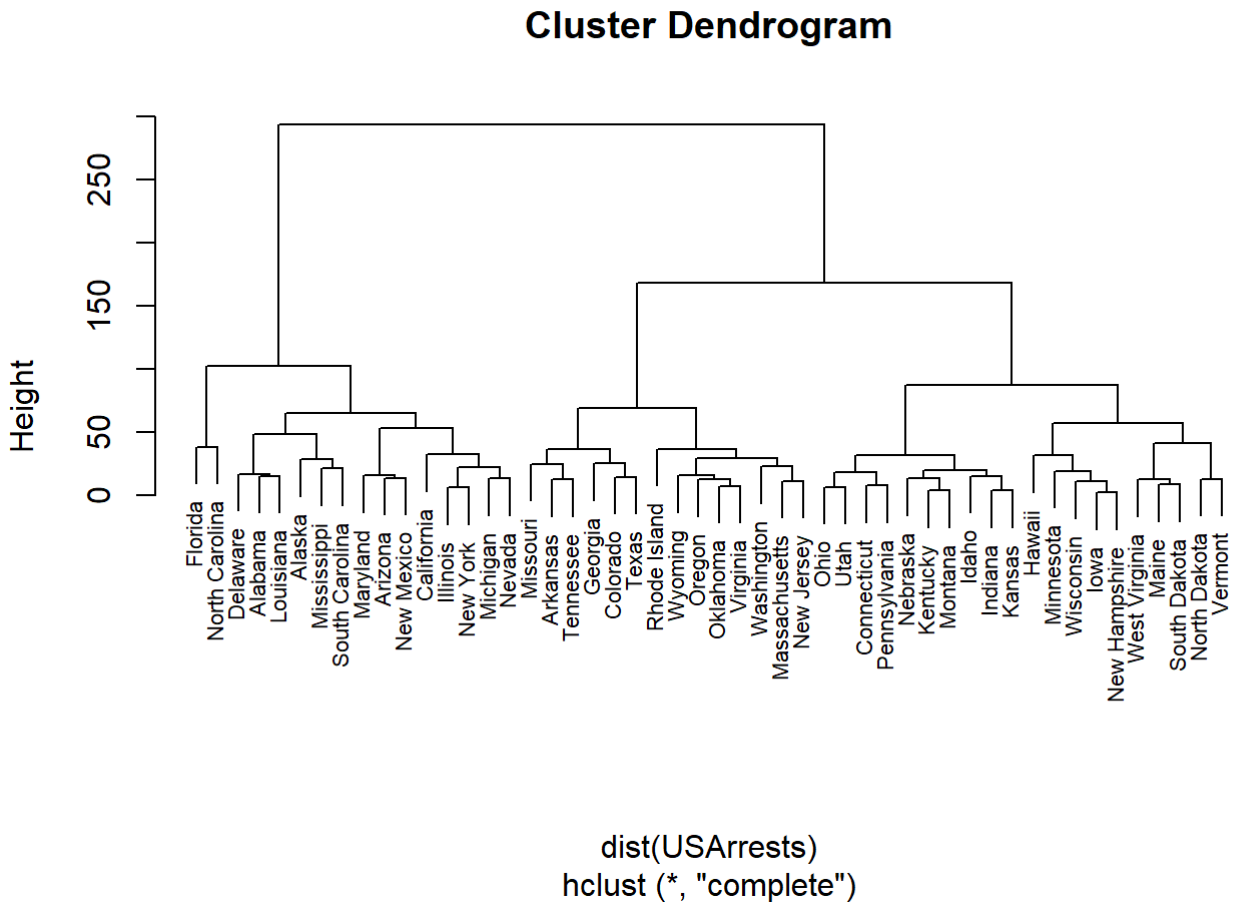
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

Exercise 10.7: Problem 9

a.

```
set.seed(114514)

hc = hclust(dist(USArrests), method = 'complete')
plot(hc, cex = 0.7)
```



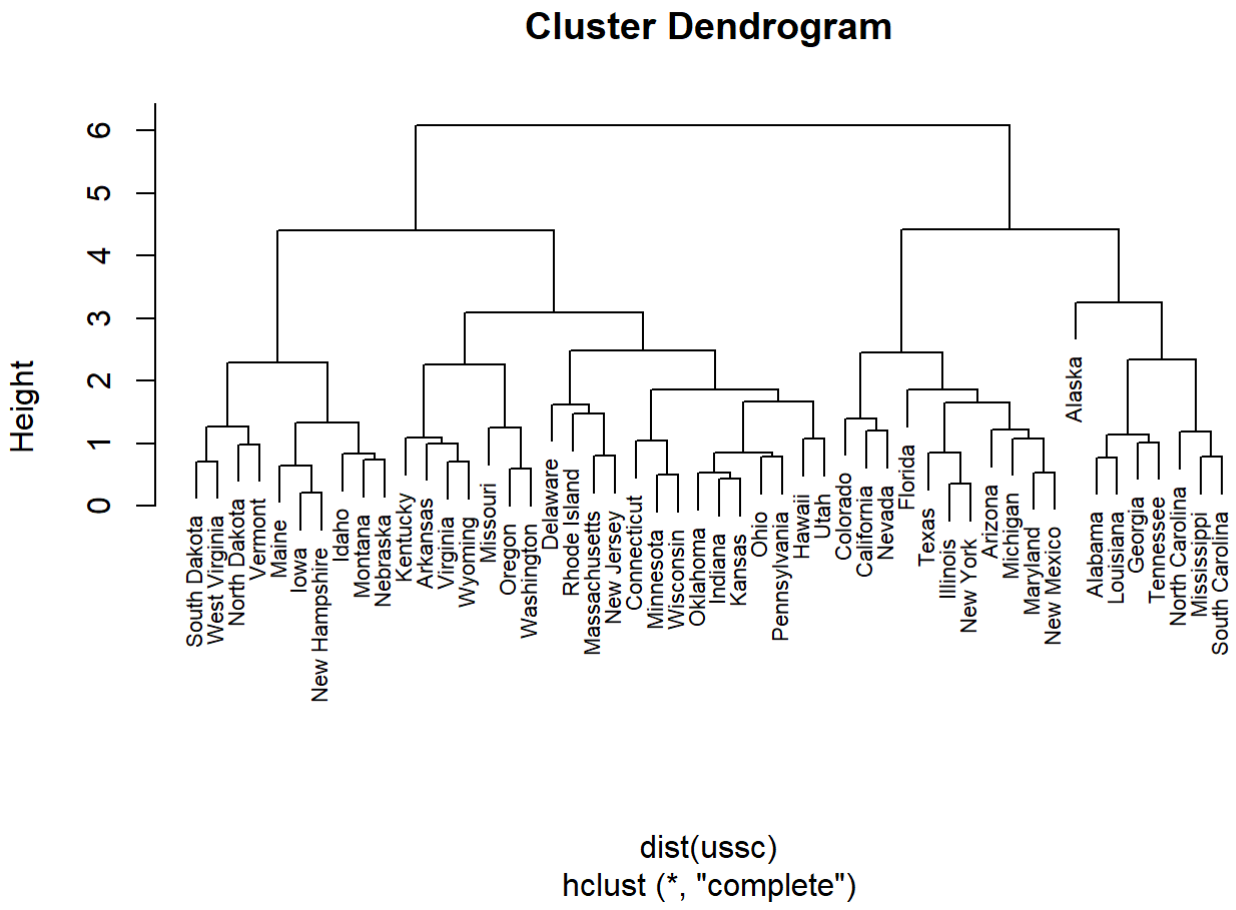
b.

```
hccut = cutree(hc, k=3)
hccut
```

```
##        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.

```
ussc = scale(USArrests)
hcsc = hclust(dist(ussc), method = 'complete')
plot(hcsc, cex = 0.7)
```



d. We can find that original model has a height of 10, and the scaled model is 6, means that the clusters are different. In general, variables with different measurement units should be scaled before computing the inter-observation dissimilarities. From this model we can also scale the variable to make every variable has a equal measurement.