harinris_Homework1_ex

Harin Rishabh

2024-02-11

8 a)

Answer - Using the sdev output of the prcomp() function

```
set.seed(1)
scaled_data = scale(USArrests)
pr.out = prcomp(scaled_data)
pr.var = pr.out$sdev^2
pve = pr.var/sum(pr.var)
pve
```

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

8 b)

Answer - The PVE is same as the one calculated above using in-built functions.

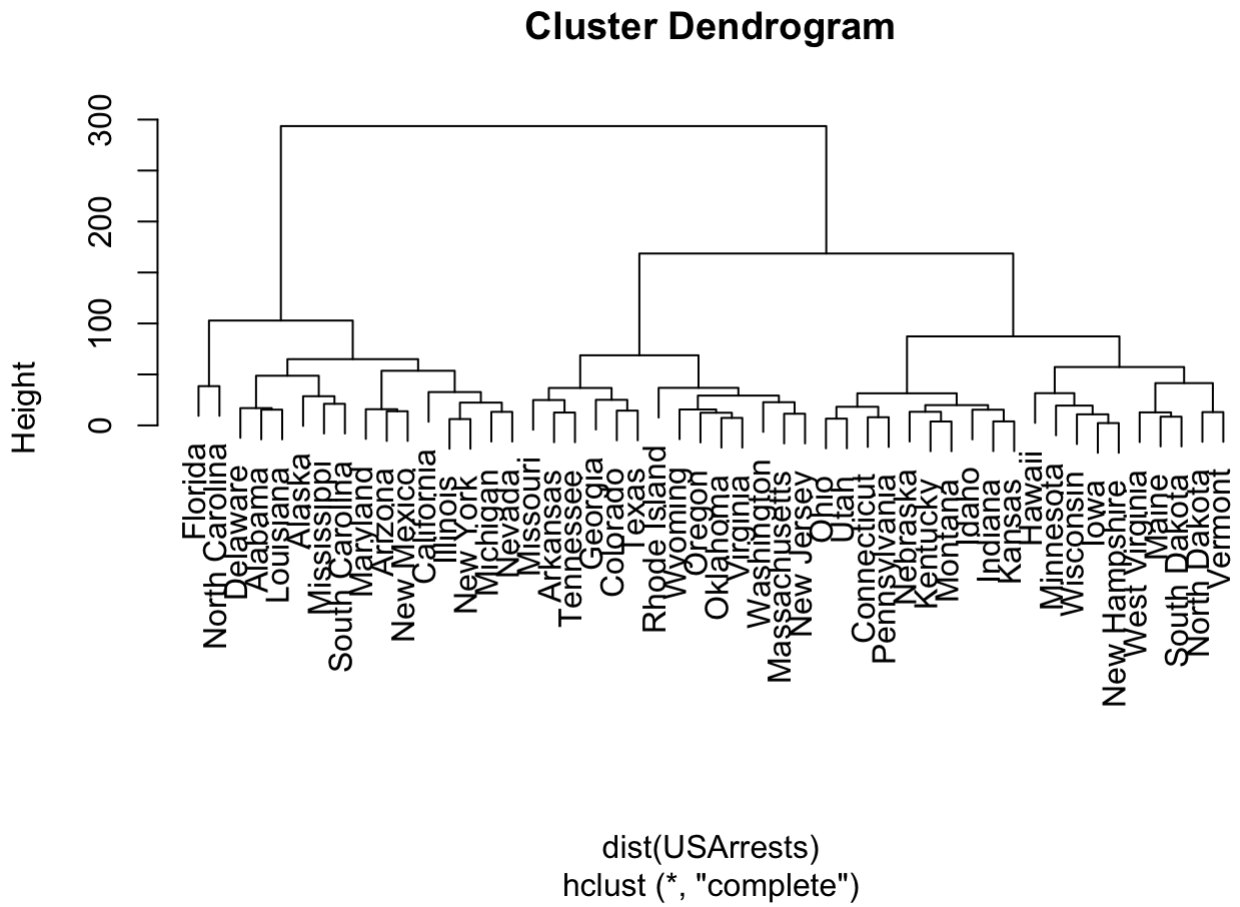
```
pcloadings = pr.out$rotation

num = apply((as.matrix(scaled_data) %*% pcloadings)^2, 2, sum)
denom = sum(apply(as.matrix(scaled_data)^2, 2, sum))

pve2 = num/denom
pve2
```

```
##        PC1        PC2        PC3        PC4
## 0.62006039 0.24744129 0.08914080 0.04335752
```

9 a)

Answer -

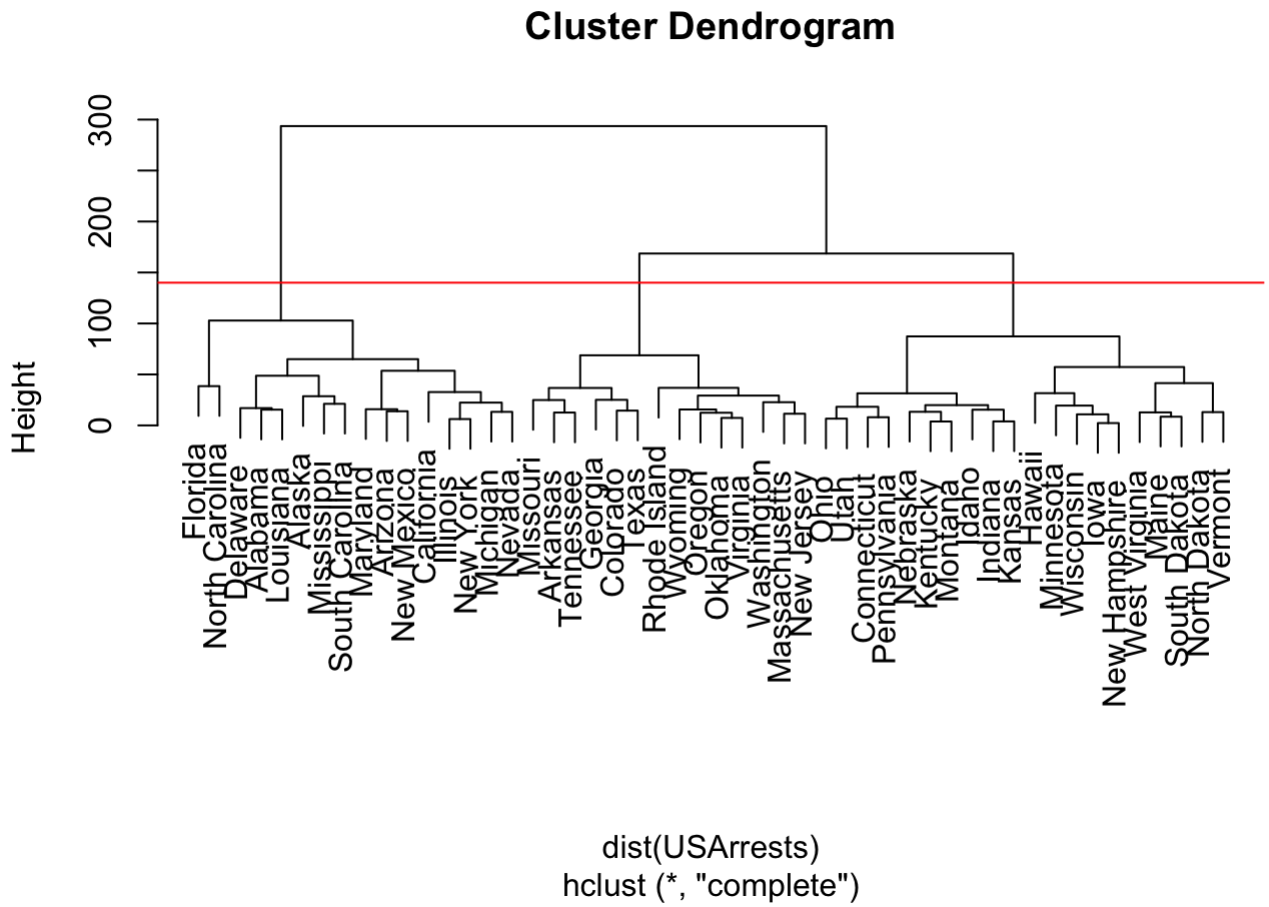
```
plot(hclust(dist(USArrests), method="complete"))
```



9 b)

Answer -

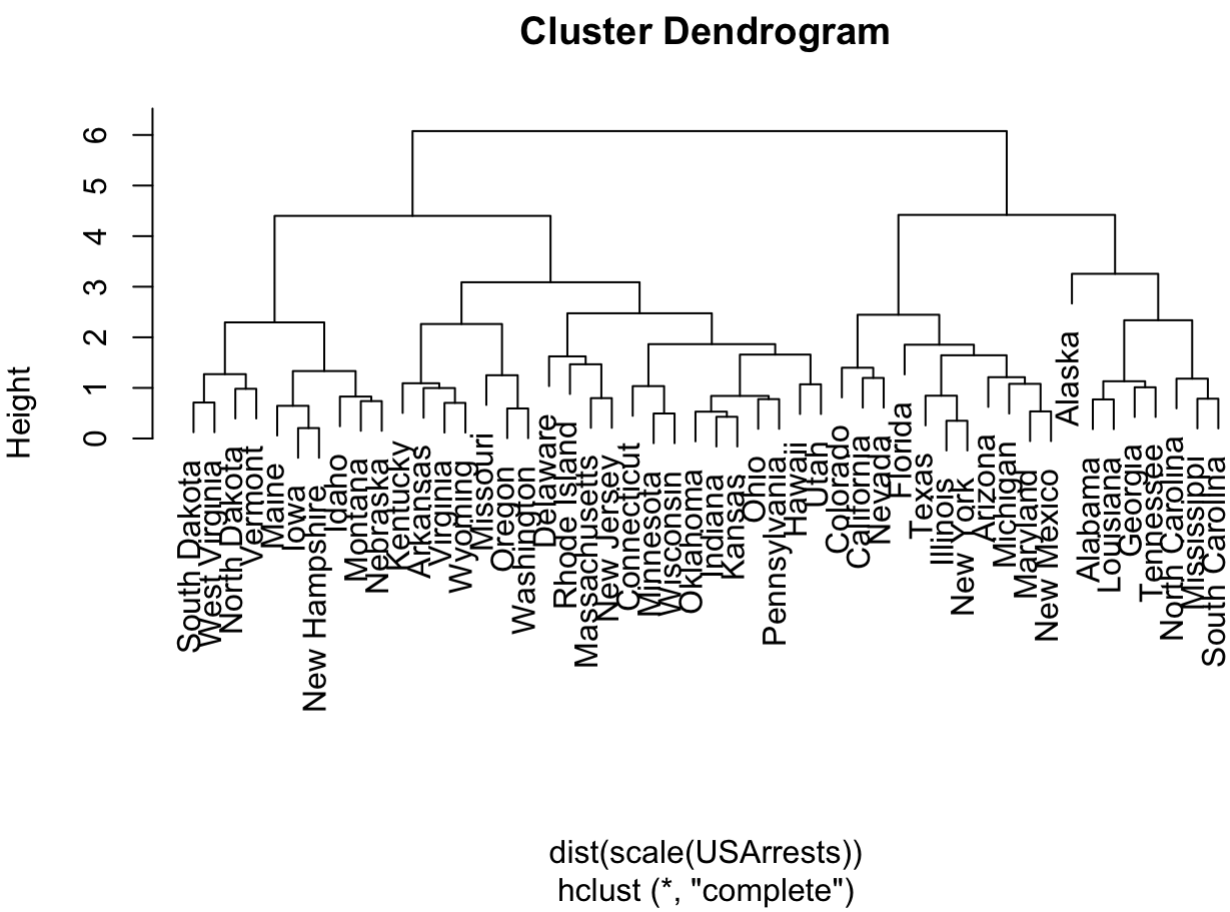
```
plot(hclust(dist(USArrests), method="complete"))
abline(h = 140, col = "red")
```



9 c)

Answer -

```
plot(hclust(dist(scale(USArrests)), method="complete"))
```



9 d)

Answer - The dendogram looks quite different after scaling the data. The height of the tree is reduced. The clustering of states is different but the breadth of the tree seems unchanged.