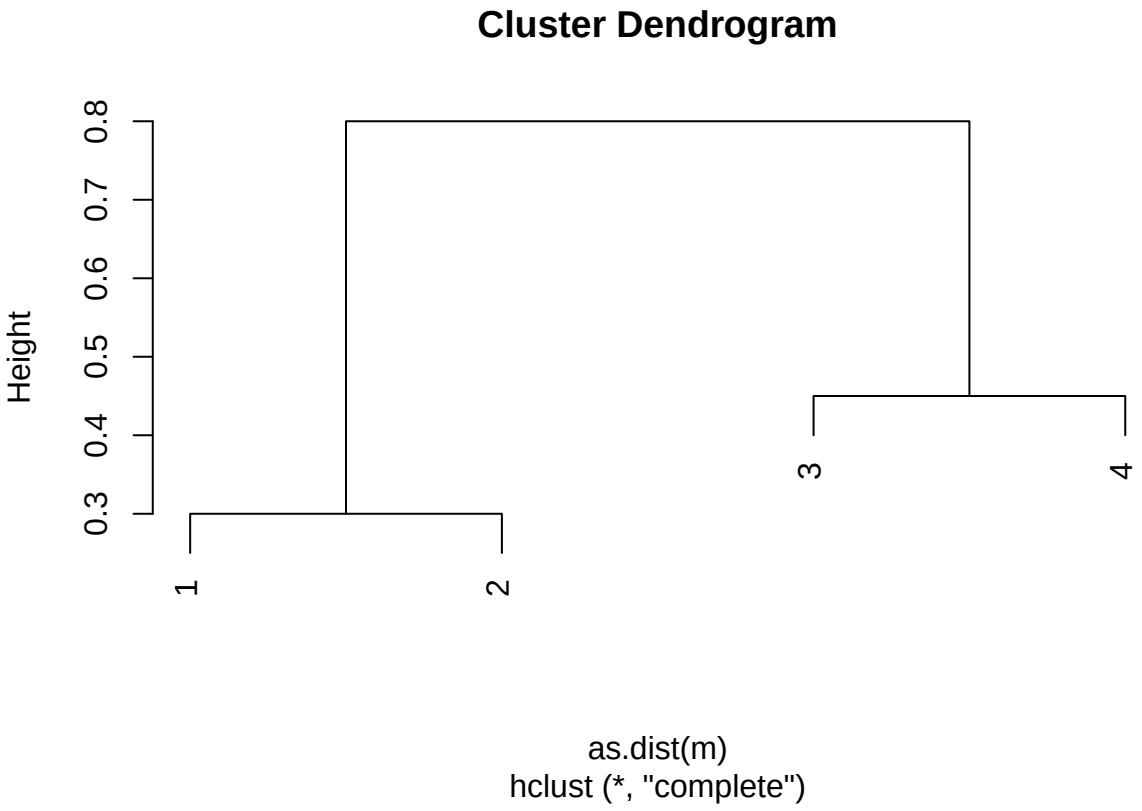
Homework Chapter 8

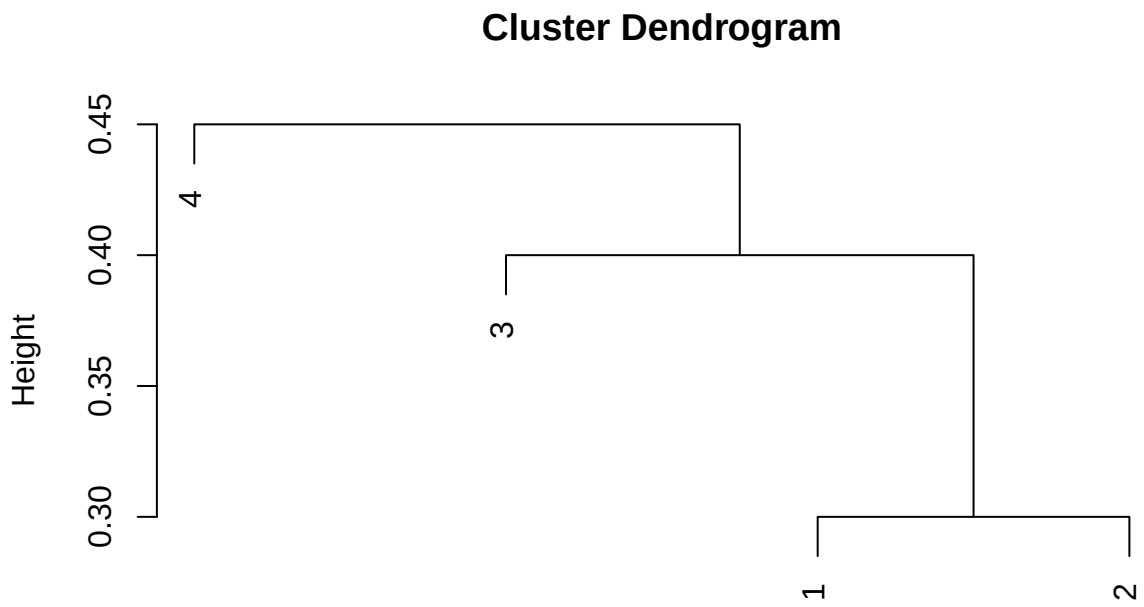
Jacob Thielemier

19 April 2024

Question 1



Part (a)



Part (b)

Part (c)

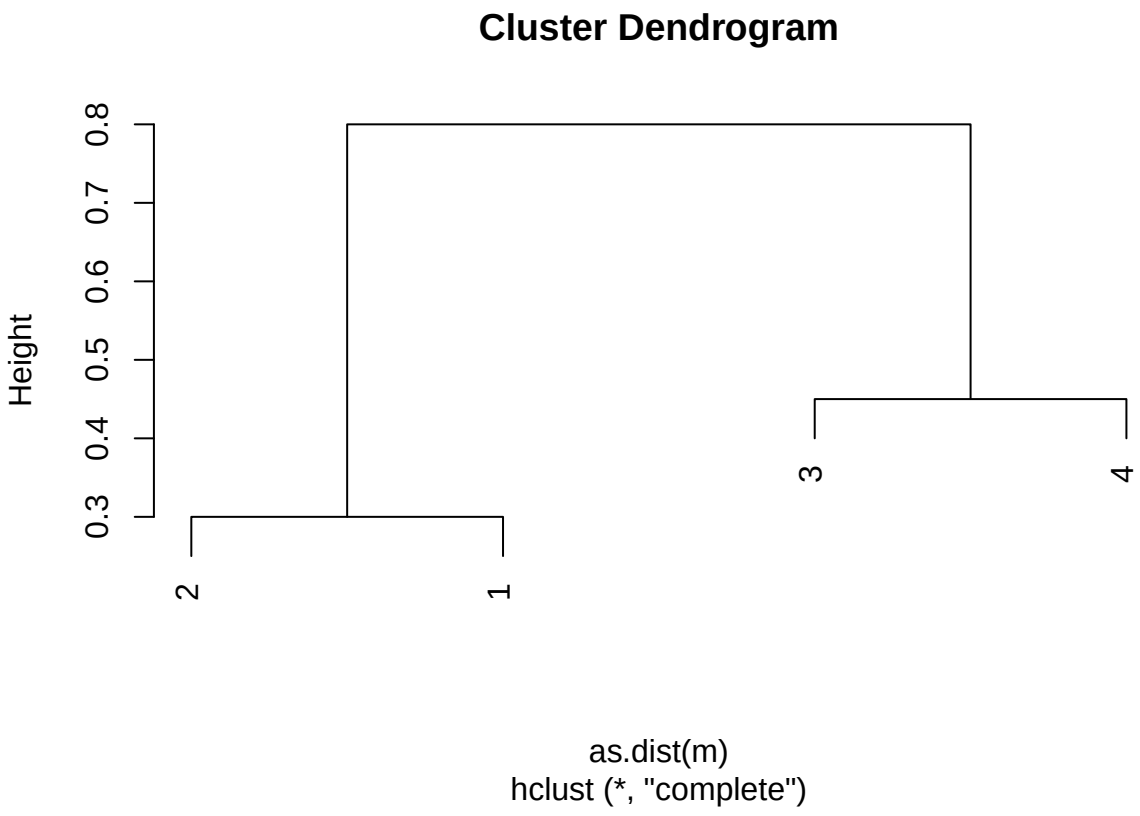
```
1 2

1 1 0 2 1 0 3 0 1 4 0 1
```

Part (d)

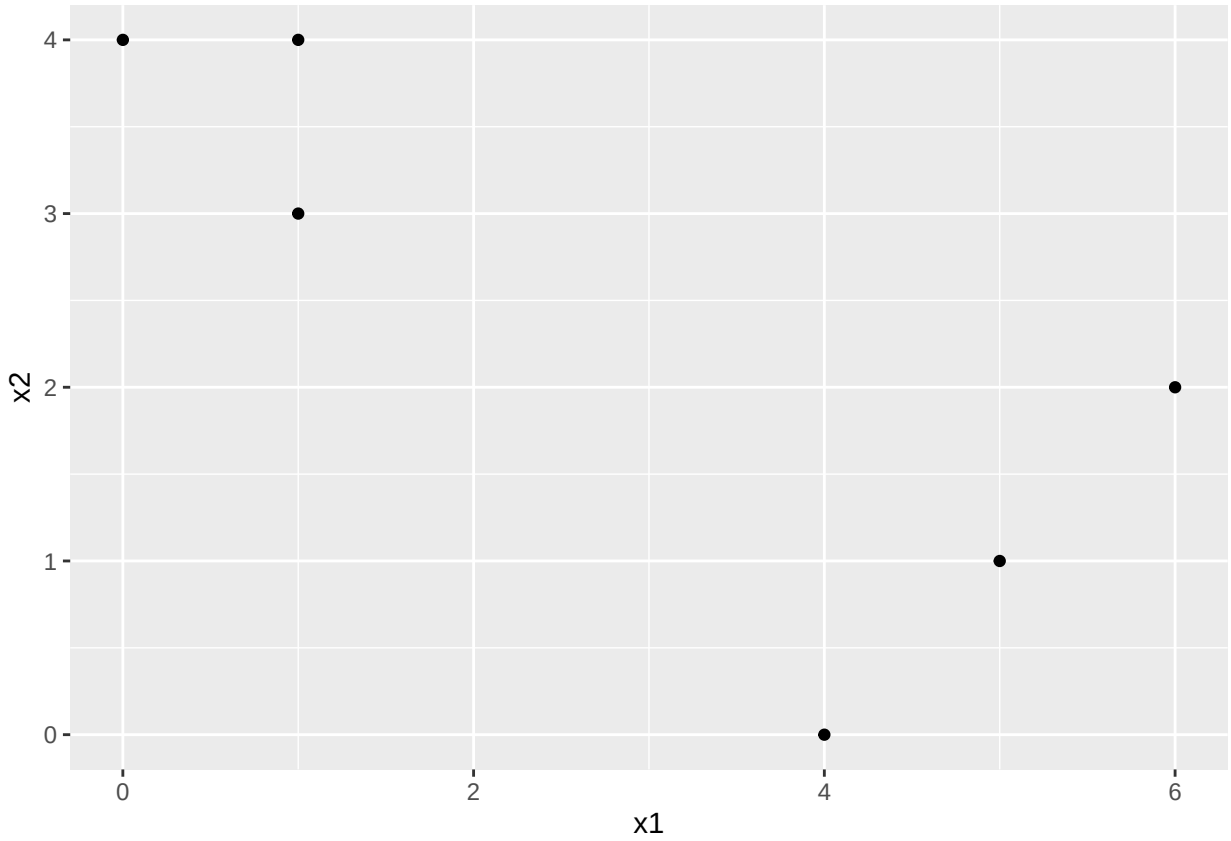
```
1 2

1 1 0 2 1 0 3 1 0 4 0 1
```


Cluster Dendrogram
Height
0.3
0.4
0.5
0.6
0.7
0.8
2
1
3
4
as.dist(m)
hclust (*, "complete")

Part (e)

Question 2



Part (a)

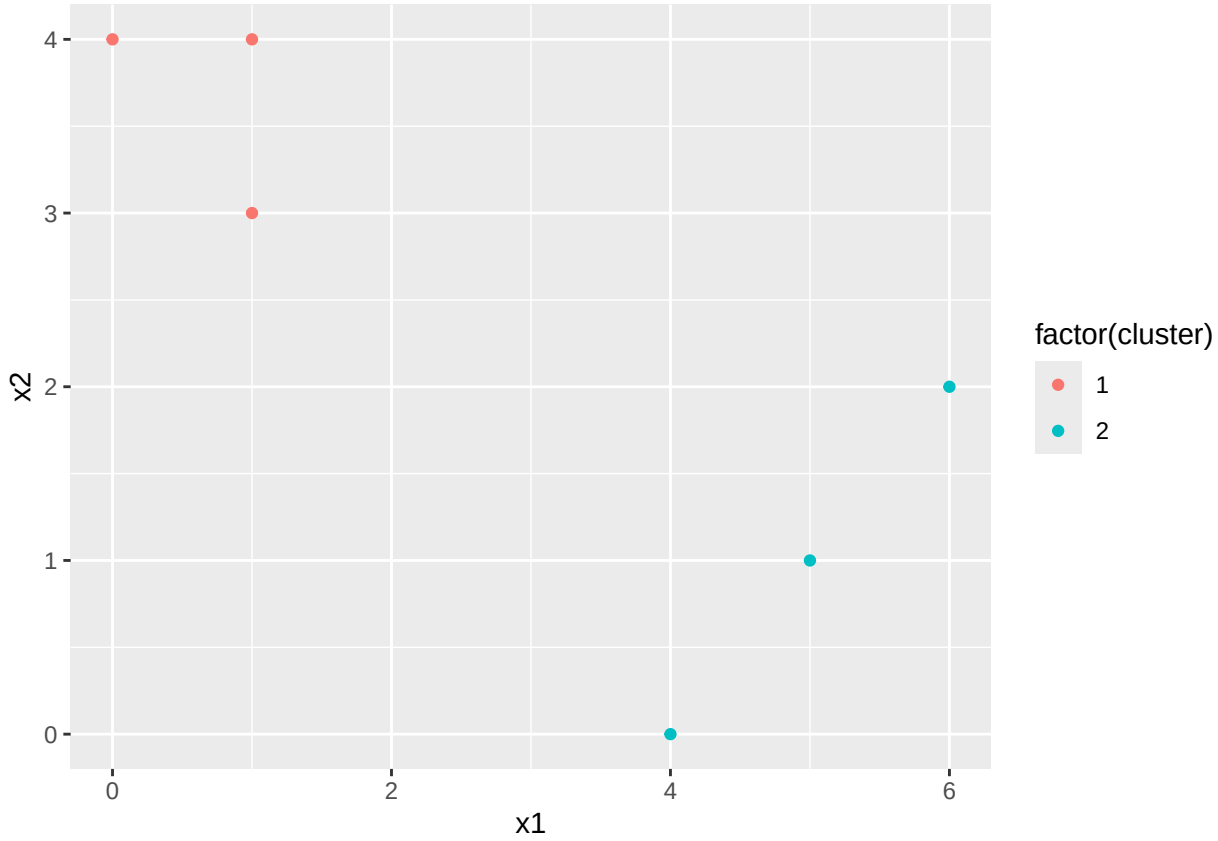
Part (b) [1] 1 1 1 1 2 2

Part (c) [,1] [,2] x1 1.75 5 x2 3.00 1

Part (d) [1] 1 1 1 2 2 2

Part (e)

- In this case, we get stable labels after the first iteration.



Part (g)

Question 3

Part (a and b)

Table 1: Cluster Assignment of States

Cluster	States
1	Alabama, Alaska, Arizona, California, Delaware, Florida, Illinois, Louisiana, Maryland, Michigan, Mississippi, Nevada, New Mexico, New York, North Carolina, South Carolina
2	Arkansas, Colorado, Georgia, Massachusetts, Missouri, New Jersey, Oklahoma, Oregon, Rhode Island, Tennessee, Texas, Virginia, Washington, Wyoming
3	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 and d)

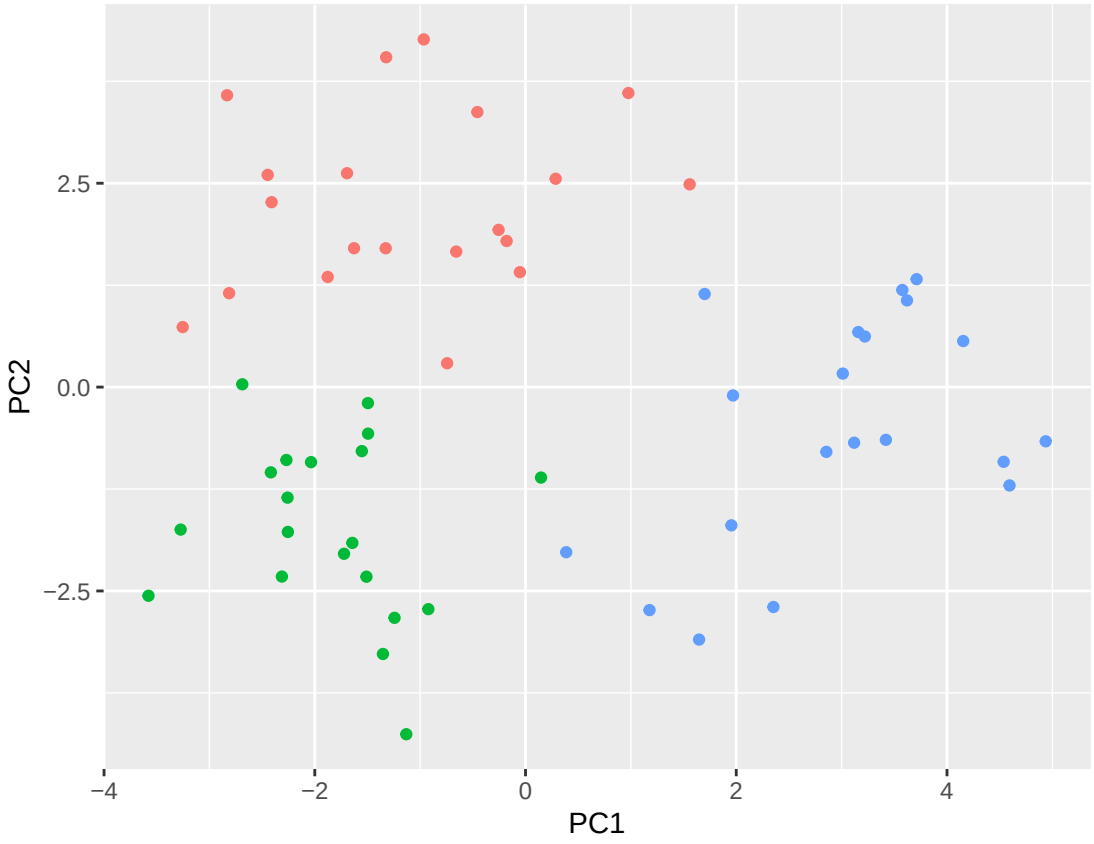
Table 2: Cluster Assignment of States

Cluster	States
1	Alabama, Alaska, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee
2	Arizona, California, Colorado, Florida, Illinois, Maryland, Michigan, Nevada, New Mexico, New York, Texas

Cluster	States
3	Arkansas, Connecticut, Delaware, Hawaii, Idaho, Indiana, Iowa, Kansas, Kentucky, Maine, Massachusetts, Minnesota, Missouri, Montana, Nebraska, New Hampshire, New Jersey, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Dakota, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming

- Scaling results in different clusters and the choice of whether to scale or not depends on the data in question. In this case, the variables are:
  - Murder numeric Murder arrests (per 100,000)
  - Assault numeric Assault arrests (per 100,000)
  - UrbanPop numeric Percent urban population
  - Rape numeric Rape arrests (per 100,000)
- These variables are not naturally on the same unit and the units involved are somewhat arbitrary (so for example, Murder could be measured per 1 million rather than per 100,000) so in this case I would argue the data should be scaled.

Question 4



Part (a and b)

Part (c) km A B C 1 1 20 1 2 0 0 19 3 19 0 0

K-means separates out the clusters nearly perfectly.

Part (d) km A B C 1 18 20 1 2 2 0 19

K-means effectively defines cluster 2 to be class B, but cluster 1 is a mix of classes A and B.

Part (e) km A B C 1 0 7 2 2 18 1 0 3 0 0 18 4 2 12 0

K-means effectively defines cluster 1 to be class B, cluster 2 to be class A but clusters 3 and 4 are split over class C.

Part (f) km A B C 1 0 20 2 2 20 0 0 3 0 0 18

K-means again separates out the clusters nearly perfectly.

Part (g) km A B C 1 1 20 1 2 19 0 0 3 0 0 19

K-means appears to perform less well on the scaled data in this case.