

STATS 415 HW 8

Jiahao Cheng

April 2025

Problem 1

- (a) Figure 1 and Figure 2 show the hierarchical clustering with complete linkage and Euclidean distance.

```
> library(ISLR2)
> library(cluster)
> data("USArrests")
> hc.complete <- hclust(dist(USArrests), method = "complete")
> plot(hc.complete, main = "Complete")
```

Figure 1: Code

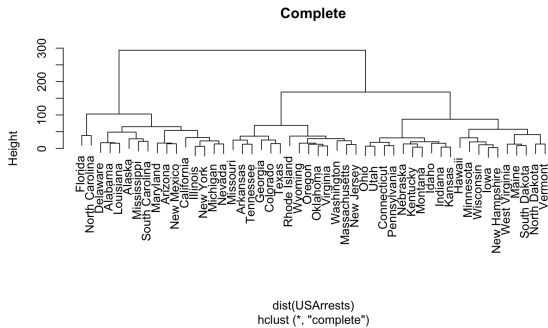


Figure 2: Plot

- (b) Figure 3 shows the result of 3 groups clustering. Each state belongs to a $1/2/3$ group.

```
> print(cutree(hc.complete, 3))
```

Alabama	Alaska	Arizona	Arkansas
1	1	1	2
California	Colorado	Connecticut	Delaware
1	2	3	1
Florida	Georgia	Hawaii	Idaho
1	2	3	3
Illinois	Indiana	Iowa	Kansas
1	3	3	3
Kentucky	Louisiana	Maine	Maryland
3	1	3	1
Massachusetts	Michigan	Minnesota	Mississippi
2	1	3	1
Missouri	Montana	Nebraska	Nevada
2	3	3	1
New Hampshire	New Jersey	New Mexico	New York
3	2	1	1
North Carolina	North Dakota	Ohio	Oklahoma
1	3	3	2
Oregon	Pennsylvania	Rhode Island	South Carolina
2	3	2	1
South Dakota	Tennessee	Texas	Utah
3	2	2	3
Vermont	Virginia	Washington	West Virginia
3	2	2	3
Wisconsin	Wyoming		
3	2		

Figure 3: 3 groups

- (c) Figure 4 and Figure 5 show the the hierarchical clustering with complete linkage and Euclidean distance after scaling .

```

> scaled_data <- scale(USArrests)
> hc.complete <- hclust(dist(scaled_data), method = "complete")
> plot(hc.complete, main = "Complete")
> print(cutree(hc.complete, 3))

```

Alabama	Alaska	Arizona	Arkansas
1	1	2	3
California	Colorado	Connecticut	Delaware
2	2	3	3
Florida	Georgia	Hawaii	Idaho
2	1	3	3
Illinois	Indiana	Iowa	Kansas
2	3	3	3
Kentucky	Louisiana	Maine	Maryland
3	1	3	2
Massachusetts	Michigan	Minnesota	Mississippi
3	2	3	1
Missouri	Montana	Nebraska	Nevada
3	3	3	2
New Hampshire	New Jersey	New Mexico	New York
3	3	2	2
North Carolina	North Dakota	Ohio	Oklahoma
1	3	3	3
Oregon	Pennsylvania	Rhode Island	South Carolina
3	3	3	1
South Dakota	Tennessee	Texas	Utah
3	1	2	3
Vermont	Virginia	Washington	West Virginia
3	3	3	3
Wisconsin	Wyoming		
3	3		

Figure 4: After scaling code

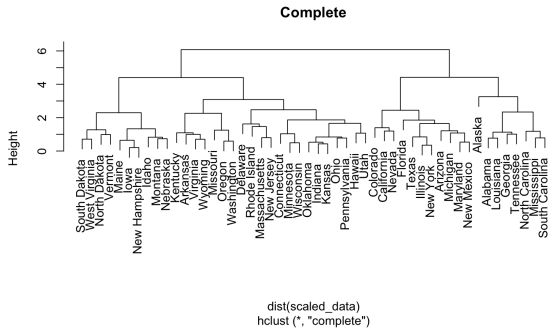


Figure 5: After scaling plot

- (d) Scaling the variable allows variables in different units to be computed under the same scale. Before scaling, some variables with large unit and variance tend to

influence more on distance computation than other variables.
The variables should be scaled before the inter-observation dissimilarities are computed. Because variables are in different units.