Brady Field

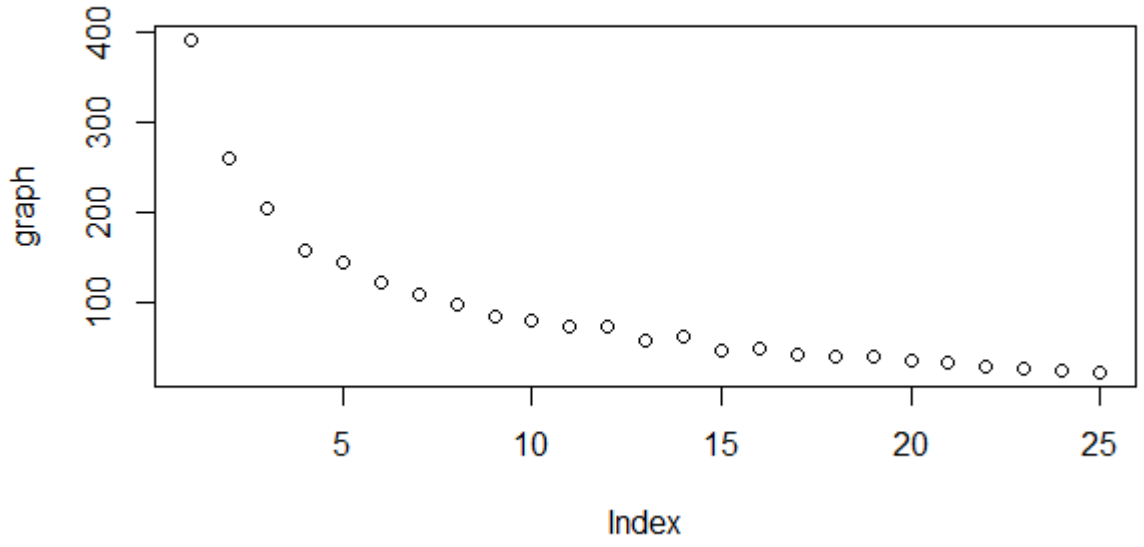
1. Make sure to use a normalized version of the dataset.

done

2. Using k-means, cluster the data into 3 clusters. Note the size of each cluster and the mean values. Do you have any insight into why they were divided this way?

They grouped the things with numbers that were fairly close together

3. Using a for loop, repeat the clustering process for k = 1 to 25, and plot the total within-cluster sum of squares error for each k-value.



4. Evaluate the plot from the previous item, and choose an appropriate k-value using the "elbow method" mentioned in your reading. Then re-cluster a single time using that k-value. Use this clustering for the remaining questions.

after 6 it dropped steadlily, but only a little bit at a time

5. List the states in each cluster.

1- Alabama, Arkansas, Georgia, Kentucky, Louisiana, Mississippi, New Mexico, North Carolina, South Carolina, Tennessee, West Virginia

2- California, New York, Texas

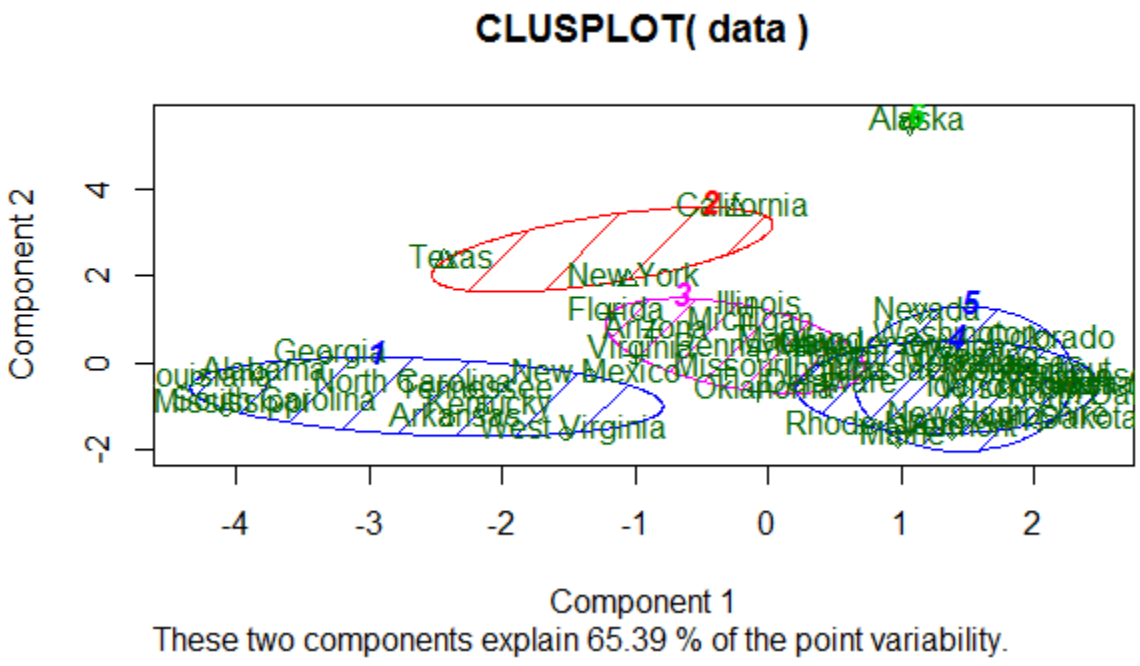
3- Arizona, Delaware, Florida, Illinois, Indiana, Maryland, Michigan, Missouri, New Jersey, Ohio, Oklahoma, Pennsylvania, Virginia

4- Connecticut, Hawaii, Iowa, Kansas, Massachusetts, Minnesota, Nebraska, North Dakota, Oregon, Rhode Island, Washington, Wisconsin

5- Colorado, Idaho, Maine, Montana, Nevada, New Hampshire, South Dakota, Utah, Vermont, Wyoming

6- Alaska

6. Use "clusplot" to plot a 2D representation of the clustering.



7. Analyze the centers of each of these clusters. Can you identify any insight into this clustering?

	Population	Income	Illiteracy	Life Exp	Murder	HS Grad	Frost
1	-0.2269956	-1.3014617	1.3915271	-1.1773136	1.0919809	-1.4157826	-0.7206500
2	2.8948232	0.4869237	0.6507713	0.1301655	1.0172810	0.1393257	-1.1310576

3 0.4824142 0.4116869 -0.2031820 -0.2241576 0.2018336 -0.0990848 -0.1686716

4 -0.3466061 0.5361532 -0.4976487 1.1842528 -1.0979159 0.5117826 0.1274187

5 -0.7430230 -0.1988706 -0.9187360 0.2431478 -0.5574910 0.8619539 1.1069462

6 -0.8693980 3.0582456 0.5413980 -1.1685098 1.0624293 1.6828035 0.9145676

Area

1 -0.23402899

2 0.99272004

3 -0.28893759

4 -0.24286635

5 0.04573941

6 5.80934967

It actually ended up graphing them by geography and population pretty well. Heck, it even reflects political alignment

to a degree in some cases.

This info could be used to infer some interesting things.

8. Please select the category you feel best describes your assignment:

D - Meets requirements

9. Provide a brief justification (1-2 sentences) for selecting that category.

Followed instructions and had fun