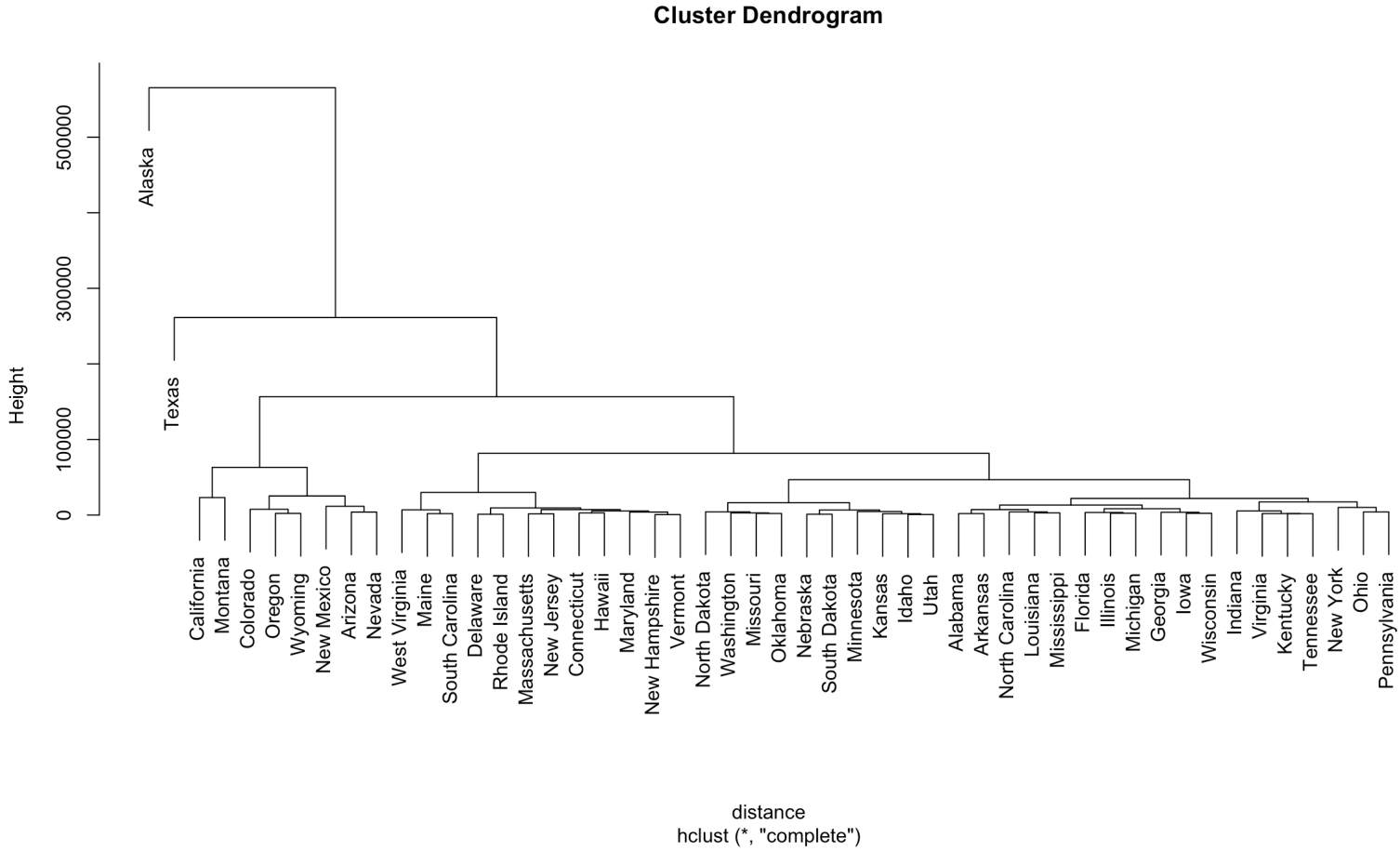
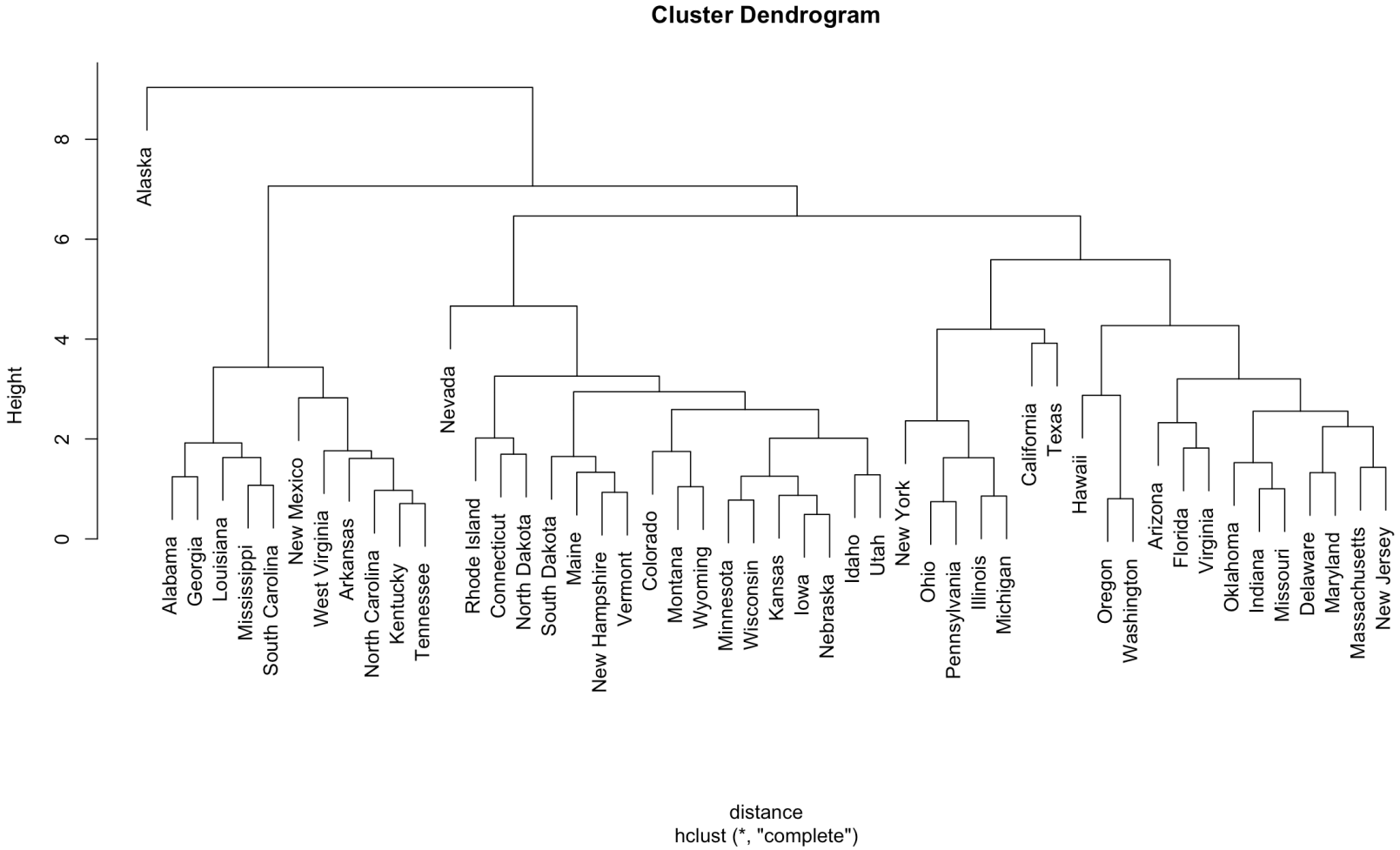
AGGLOMERATIVE HIERARCHICAL CLUSTERING

- Use hierarchical clustering to cluster the data on all attributes and produce a dendrogram

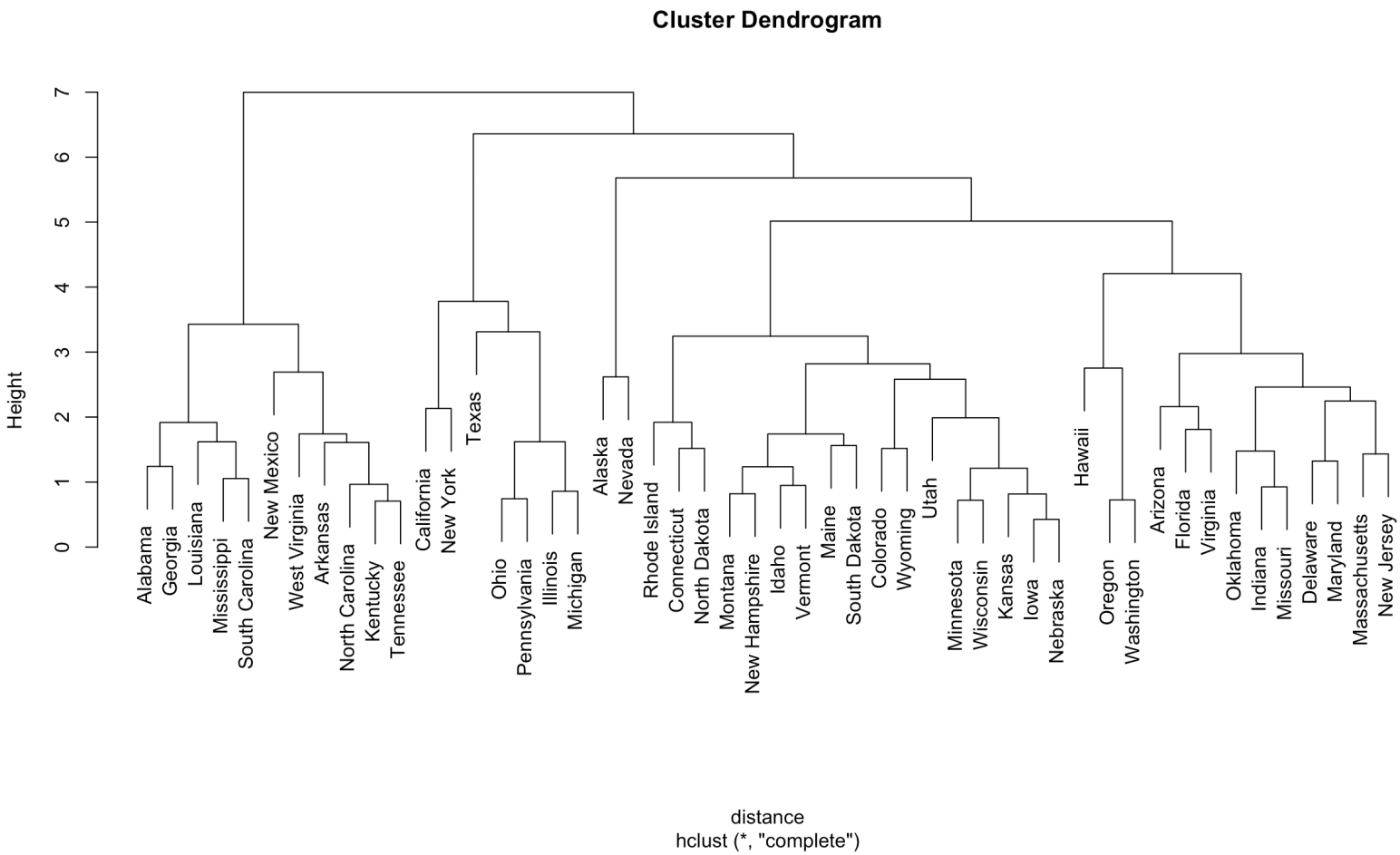


- Repeat the previous item with a normalized dataset and note any differences



More accurate but Nevada is out of place. Hawaii is also out of place.

- Remove "Area" from the attributes and re-cluster (and note any differences)



Area is important because this dendrogram does not make any sense. For example, Alaska isn't close the Nevada by any mean.

- Cluster only on the Frost attribute and observe the results

This one is interesting. All the warm state is grouped together. The state with low temperature are also grouped together.

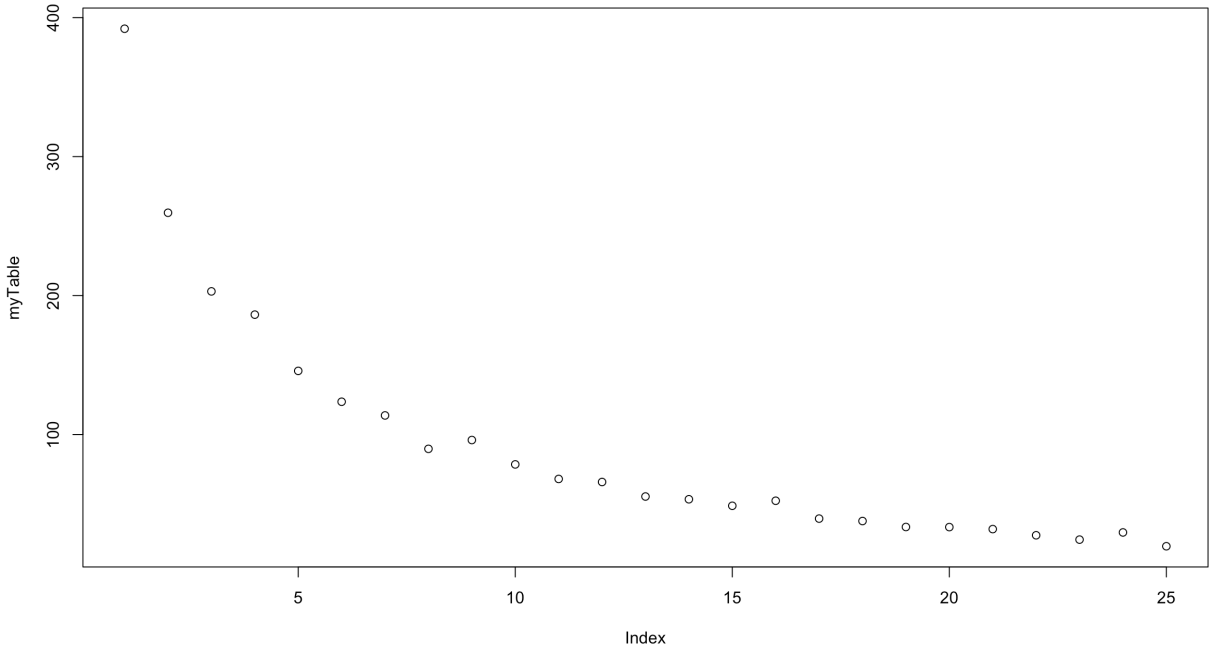
USING K-MEANS

- 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?
  [1] 111.66951 67.72742 23.62227

	Population	Income	Illiteracy	Life Exp	Murder	HS Grad	Frost	Area
1	0.9462026	0.7416690	0.005468667	-0.3242467	0.5676042	0.1558335	-0.1960979	0.4483198
2	-0.4873370	0.1329601	-0.641201154	0.7422562	-0.8552439	0.5515044	0.4528591	-0.1729366
3	-0.2269956	-1.3014617	1.391527063	-1.1773136	1.0919809	-1.4157826	-0.7206500	-0.2340290

  It is because of the starting point.

- 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.



- 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.

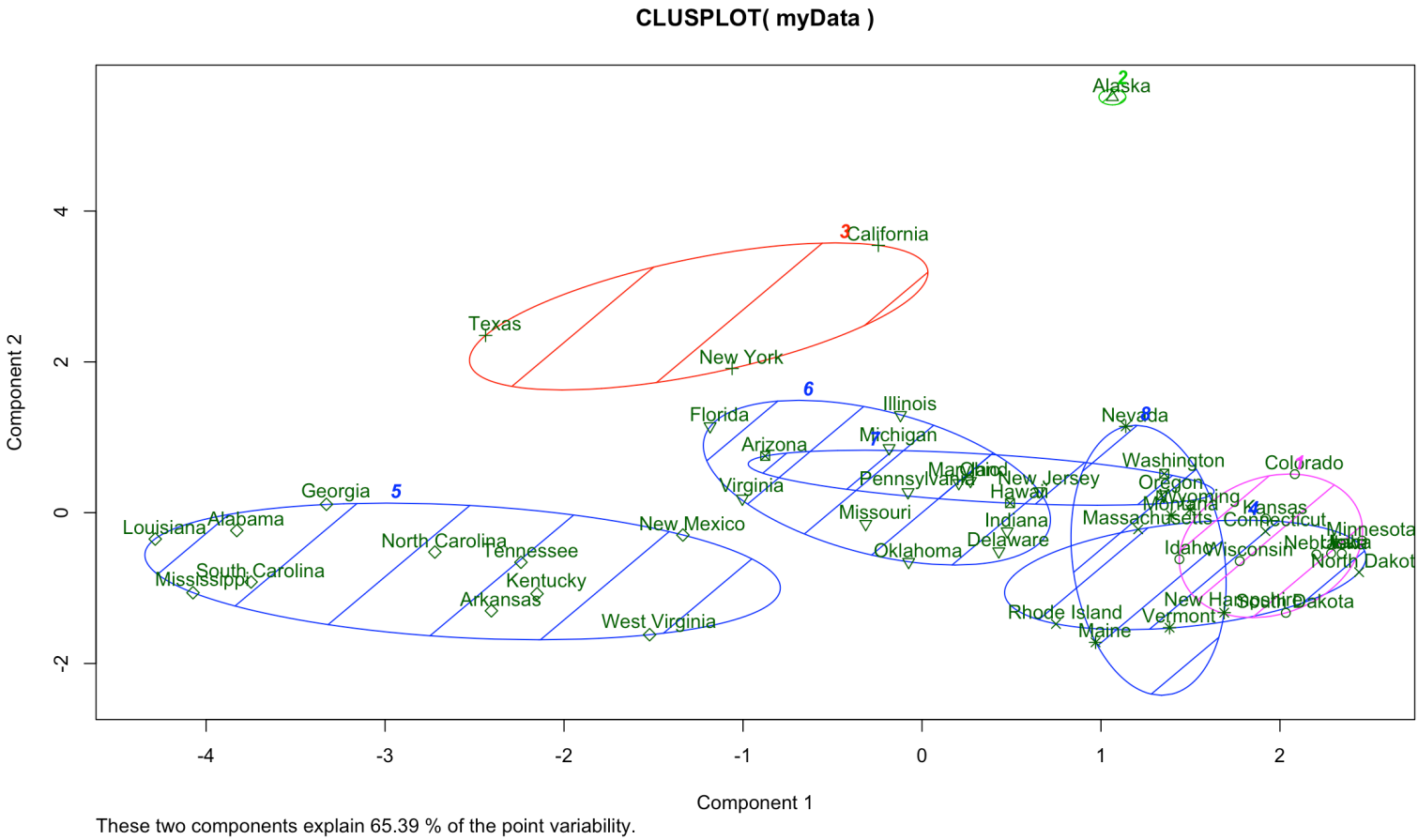
  Using elbow method, I think k = 8 is appropriate

- List the states in each cluster.

	myClusters.cluster.o.
Colorado	1
Idaho	1
Iowa	1
Kansas	1
Minnesota	1
Nebraska	1
South Dakota	1
Utah	1

Wisconsin	1
Alaska	2
California	3
New York	3
Texas	3
Connecticut	4
Massachusetts	4
North Dakota	4
Rhode Island	4
Alabama	5
Arkansas	5
Georgia	5
Kentucky	5
Louisiana	5
Mississippi	5
New Mexico	5
North Carolina	5
South Carolina	5
Tennessee	5
West Virginia	5
Delaware	6
Florida	6
Illinois	6
Indiana	6
Maryland	6
Michigan	6
Missouri	6
New Jersey	6
Ohio	6
Oklahoma	6
Pennsylvania	6
Virginia	6
Arizona	7
Hawaii	7
Oregon	7
Washington	7
Maine	8
Montana	8
Nevada	8
New Hampshire	8
Vermont	8
Wyoming	8

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



This plot showed Alaska as an outlier, which it is.

- Analyze the centers of each of these clusters. Can you identify any insight into this clustering?
  Some center are really close with each other,