CSC 400, Data Analysis #3, Spring 2024   Name _____

Instructions: More information on expectations for the report can be found in the general directions for the data analysis sets. In this document, the specifics for the individual assignment will be discussed. Students are responsible for both the requirements in the general directions, and for the specific directions discussed below.

Topic 1: Clustering

The USArrests dataset is built in the {datasets} package in R. It should be pre-installed. Just open the data. More information on it can be found here:
https://www.rdocumentation.org/packages/datasets/versions/3.6.2/topics/USArrests

	Murder	Assault	UrbanPop	Rape
Alabama	13.2	236	58	21.2
Alaska	10.0	263	48	44.5
Arizona	8.1	294	80	31.0

```
data(USArrests)
View(USArrests)
```

You should apply three different clustering algorithms and experiment with different numbers of clusters. Compare your results. Describe which algorithms you used and why. Create appropriate graphs to illustrate your results. Describe any rescaling that was required (or compare to un-rescaled results to see if the results changed significantly). Consider alternative distance metrics.

When you have created your clusters for each algorithm, use the examples on this page https://www.geeksforgeeks.org/how-to-create-state-and-county-maps-easily-in-r/# to create a map of the results. You'll need to create a new column for the number of the cluster each state belongs to. The examples on this page are using the USArrests dataset. You may need to experiment a bit with the various settings to get a good-looking, non-distorted map. After creating your map and color-coding by cluster, do there appear to be any geographic relationships to the clustering or are the clusters geographically random?

You may include your code in an appendix or separate file, but the report of approximately 10 pages should focus on the analysis. It should look professionally formatted. Raw code and raw output is frowned upon.