**Term Project Phase II: Finding Natural Breaks** (Due 11:00 AM Tuesday November 22, 2016)


In phase I, you calculated the natural breaks for n = 2 (two groups). In phase II, you will calculate the natural breaks for n = 5 and n = 7 using the dataset  data30.py that includes 30 DSC ratio values:

Download data30.py here http://www.cs.umb.edu/~ding/classes/187/homework/data30.py

Thanks Daniel Werminghausen who helped us to randomly select 30 values out of the original 5362 data points.

Work on this with each of your team members, and remember the following guidelines for submitting your work:

Make sure to submit
1.      The main code file
2.      Any modules/txt files that you used as input
3.      Any instructions on how to run the code so that the proper result can be viewed

Basically, whatever is needed to see and run the results, you must include!

**One submission per team from the Team Lead. Please submit the source code of your program and results of your program.**

The example below explains the required output in Phase II. Output should have the following headings GreaterThanEqual, LessThan, PercentOfObservations, CountOfObservations.

To clarify, "% of observations in range" is calculated as (#number of  items in a natural break group / total #number of items). "Count of observations in range" is calculated as # number of items in a natural break group.


| DSC Range | | Observations | |
|---|---|---|---|
| >= | < | % of observations in range | Count of observations in range |
| 0.3303 | 1.0849 | 27% | 107 |
| 1.0849 | 1.4691 | 17% | 69 |
| 1.2583 | 1.4691 | 19% | 78 |
| 1.4691 | 1.8286 | 16% | 66 |
| 1.8286 | 9.7251 | 20% | 29 |
| 9.7251 | 45.112 | 16% | 5 |

| 45.112 | 60.7416 | 1% | 2 |
|--------|---------|-----|---|

Base Points: 100 points
Bonus Points: 100 Points
200 Points: If a team can complete the assignment on time
120 Points: If a team can finish the core calculations of the assignment on time