Documentation
(February 2016)

Class Separability Classification Tool
&
Heuristic Classification Tool

By David Wong (dwong2@gmu.edu) and Min Sun (msun@gmu.edu)

This project is partially funded by
(1) the U.S. Census Bureau contract
Order # YA1323-12-SE-0387 (Contract Officer: Dr. Nancy Torrieri)
&
(2) the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award R01HD076020.

Background

An important feature of the ACS data is that each estimate is accompanied by a margin of error (MOE), indicating how reliable the estimate is. Previous project effort was to develop mapping tools in GIS (specifically ArcGIS) that take into account the MOEs of attributes so that when users want to compare attribute values implicitly or explicitly, the tools will render results based upon statistical comparisons of estimates with MOEs accounted for. Thus, maps using these tools can show if values are significantly different or not given a certain level of statistical significance. This is one possible approach to take into account the error information of the ACS estimates when mapping them.

Another approach is geared toward the more fundamental concept in choropleth maps. In making a choropleth map, values of the concerned variable are divided into classes and each class is associated with a color or a pattern. Then areal units are colored according the class memberships of their attribute values. The implication of such a mapping scheme is that areal units with different colors are expected to have different values, as these values are in different classes. In other words, the assumption is that values between classes are different. However, if values or estimates have error, they
may not be statistically different even if they are different numerically and are assigned to different classes (and colors).

Therefore, the objective here is to identify class breaks that can maximize the probabilities that values in different classes are statistically different. Unfortunately, such objective is limited by several constraints. Depending on the empirical distribution of the estimates, and the sizes and distributions of error, estimates may not be highly separable statistically, especially when the estimates have large errors. When estimates have large errors, they are still separable, but with low confidence levels (CLs) that they are different. Another constraint in determining highly separable classes is the number of classes. In general, estimates are more separable (with high CLs) when fewer classes are needed. For larger numbers of classes, estimates become less separable (low CLs). Therefore, our classification tool can determine the trade-off between having a larger number of less separable classes or a smaller number of more separable classes.

However, using the class separability method to determine class breaks, the resultant maps usually have very unbalanced classes, i.e., many observations fall into one or a few classes while other classes have very few observations. Such maps may not be useful to reflect spatial patterns, if any, in the data. Therefore, other classification criteria, such as evenness of distribution across classes, variation within classes may need to be considered, in the expenses of lowering separability between classes. To assist users to determine.

The detailed conceptual discussion of this classification method, the Class Separability Classification method, can be found in the following manuscript:


Another relevant paper is:

The paper that describes the heuristic classification method in detail is


**How to Use the Heuristic Classification Tool**

**Java (version 1.6 and up) is required for the classification tool.** The following discussion assumes the user has downloaded and have Java installed on computer. To launch the classification tool, double click the download file “MappingUncertaintyTool.jar”. A window with the interface shown below will appear.

![Image of the interface](image)

The logical steps of using the tool are as follows:

1) Click “Open file” to open a shapefile for mapping
2) After selecting the shapefile, a window will appear asking for the attribute labels for the estimate and margin of error (MOE).
3) Select the attribute labels for the estimate and the corresponding MOE, and click “Send selection” (nothing will happen on the interface)

4) To start mapping considering class separability, open the chart window by click “classification window”->”Chart”, a new sub-window will appear. Use the mouse cursor to drag the black marker on the percent slider bar at the bottom to the desired confidence level (CL), say 30% to activate the bar plot. The 30% CL means that you are willing to accept the class breaks such that estimates in different classes are only different 30% of the chance. Although some class breaks may have a CL higher than the chosen level, the chosen level is the lowest CL among all class breaks. In our example of choosing 30%, only four classes can be formed, shown by the four dotted lines on the left. Note that their colors correspond to the colors on the confidence level slider bar, reflecting the CLs of the corresponding class breaks having statistically different values or estimates.
Also note that the bar plot includes two elements: a bar with its height representing the estimates, and an error line for each bar showing the corresponding margin of error. This bar plot also displays estimates in ascending order.

5) Users can drag the marker to a higher CL, likely resulting in fewer classes, or to a lower CL to have more classes. If the user can accept a CL of 40%, six classes would be formed in our example, but note that additional class breaks have different colors, reflecting their corresponding CLs that estimates are statistically different as determined by these class breaks.

6) Open map window by click “classification window”->”Map”. You are able to see a map has been automatically created as long as the number of classes is not more than 9.

7) In order to support the heuristic mapping procedure of considering criteria such as in-class variability and evenness of observation being assigned to different classes, the
tool offers users star plot window (accessed by clicking “classification window” -> “star plot”) to overview and evaluate the trade-off relationship between those multiple criteria.

8) Meanwhile, the tool will allow users to add, remove or move class breaks to adjust the width of classes based on users’ cognition on the quality and other mapping criteria.

For more heuristic mapping procedure, please refer to the YouTube video linked to the website.