

Documentation
(August 2013)

Class Separability Classification Tool

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

This project is partially funded by the U.S. Census Bureau contract Order # YA1323-12-SE-0387 (Contract Officer: Dr. Nancy Torrieri)

Project website: <http://gesg.gmu.edu/census>

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 provides a heuristic environment in which users can determine the trade-off between having a larger number of less separable classes or a smaller number of more separable classes.

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

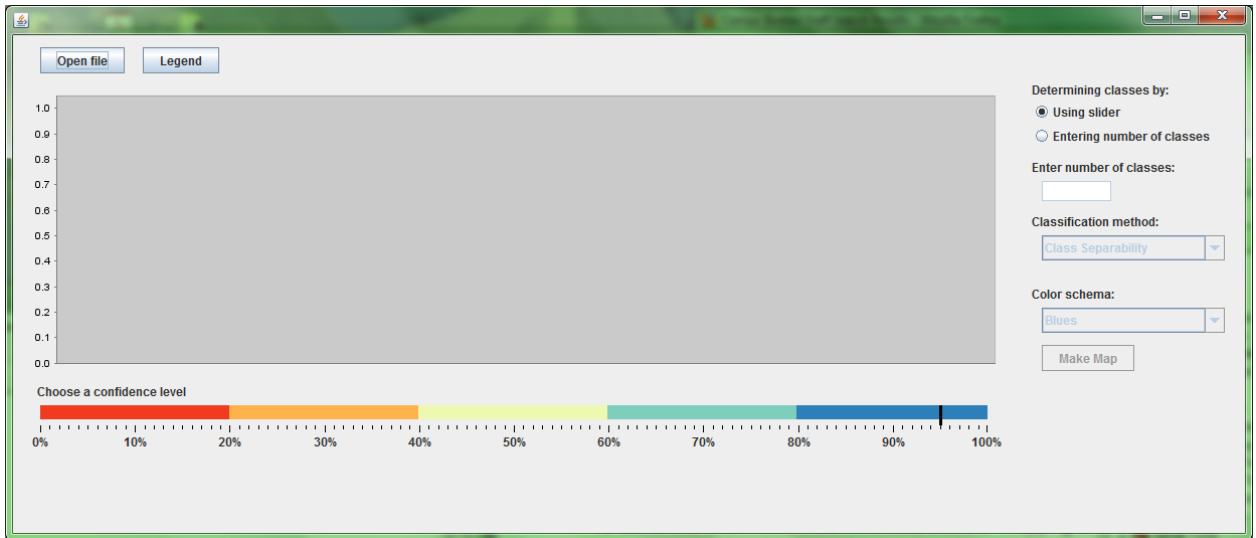
Sun, M., D. W. Wong, and B. J. Kronenfeld. A classification method for choropleth maps incorporating data reliability information. *The Professional Geographer* (to appear)

Another relevant paper is:

Sun, M., B. J. Kronenfeld, and D. W. Wong. 2013. [Cartographic techniques for communicating class separability: enhanced choropleth maps of median household income, Iowa](#). *Journal of Maps* 9(1): 43-49.

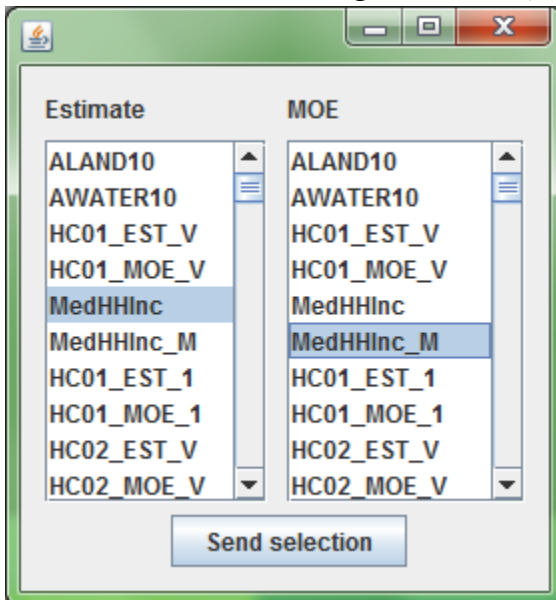
Structure of the Class Separability Classification Tool

To install the tool in ArcGIS, please refer to the installation guide available from the website <http://gesg.gmu.edu/census/download.htm>. Please note that this classification tool operates independently of the ACSMapping extension for ArcGIS we have developed earlier. Also, **Java is required for the classification tool** (install Java if your computer does not have Java). The following discussion assumes the user has downloaded and installed the classification tool in ArcGIS (ArcMap). To launch the classification tool in ArcGIS, click the “Class Separability” button on the tool bar. A new window with the interface shown below will appear outside of ArcGIS.



The logical steps of using the tool are as follows:

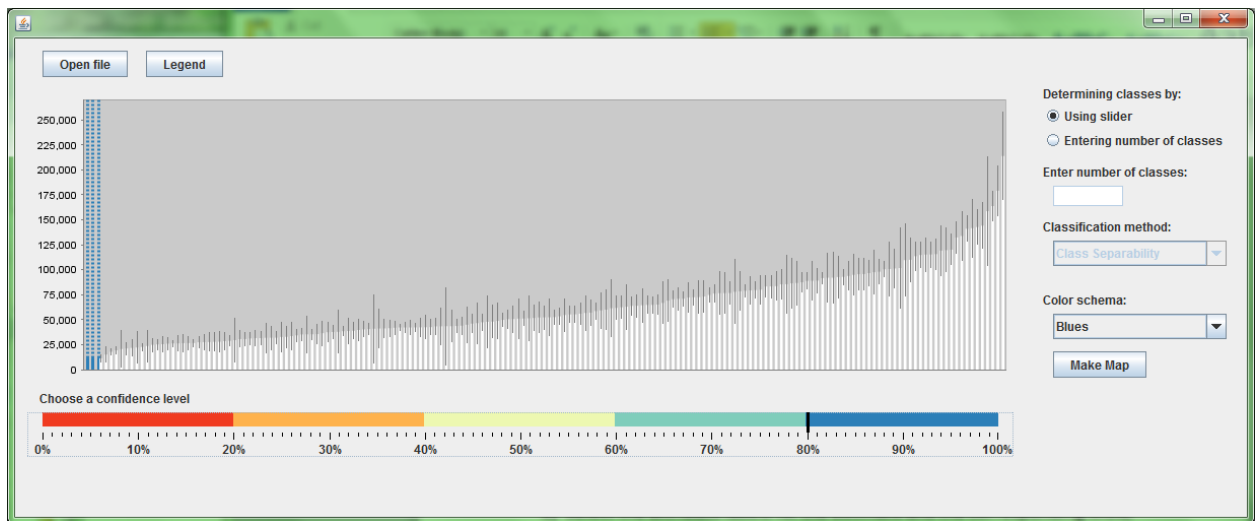
- 1) Click "Open file" to open a shapefile for mapping
- 2) After selecting the shapefile, a window will appear asking for the attribute labels for the estimate and margin of error (MOE).



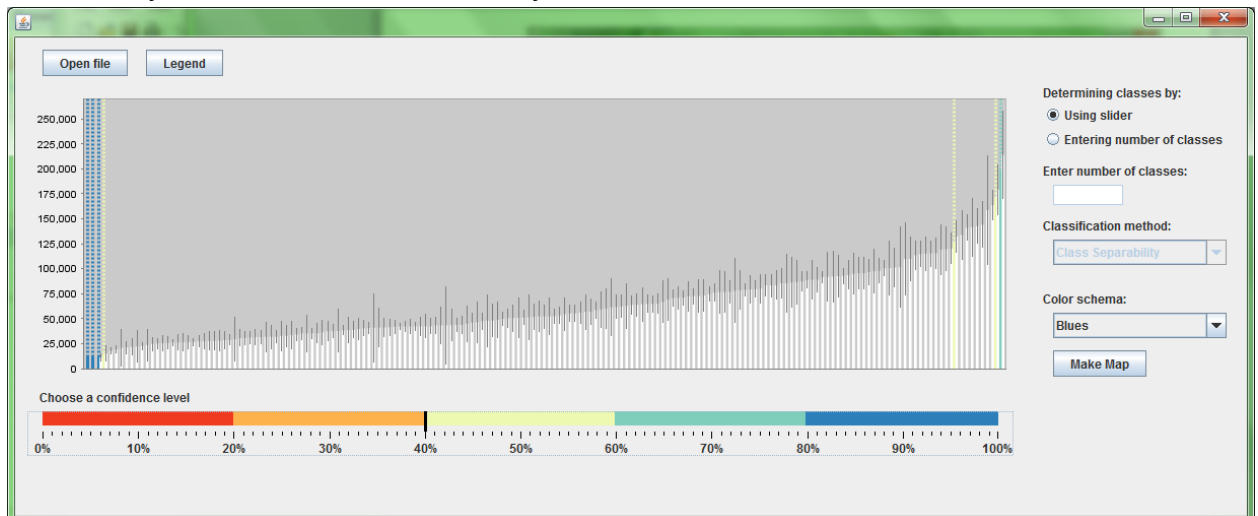
- 3) Select the attribute labels for the estimate and the corresponding MOE, and click "Send selection" (nothing will happen on the interface)
- 4) Use the mouse cursor to drag the black marker on the percent slider bar at the bottom to the desired confidence level (CL), say 80% to activate the bar plot. The 80% CL means that you are willing to accept the class breaks such that estimates in different classes are only different 80% 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 80%, 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.

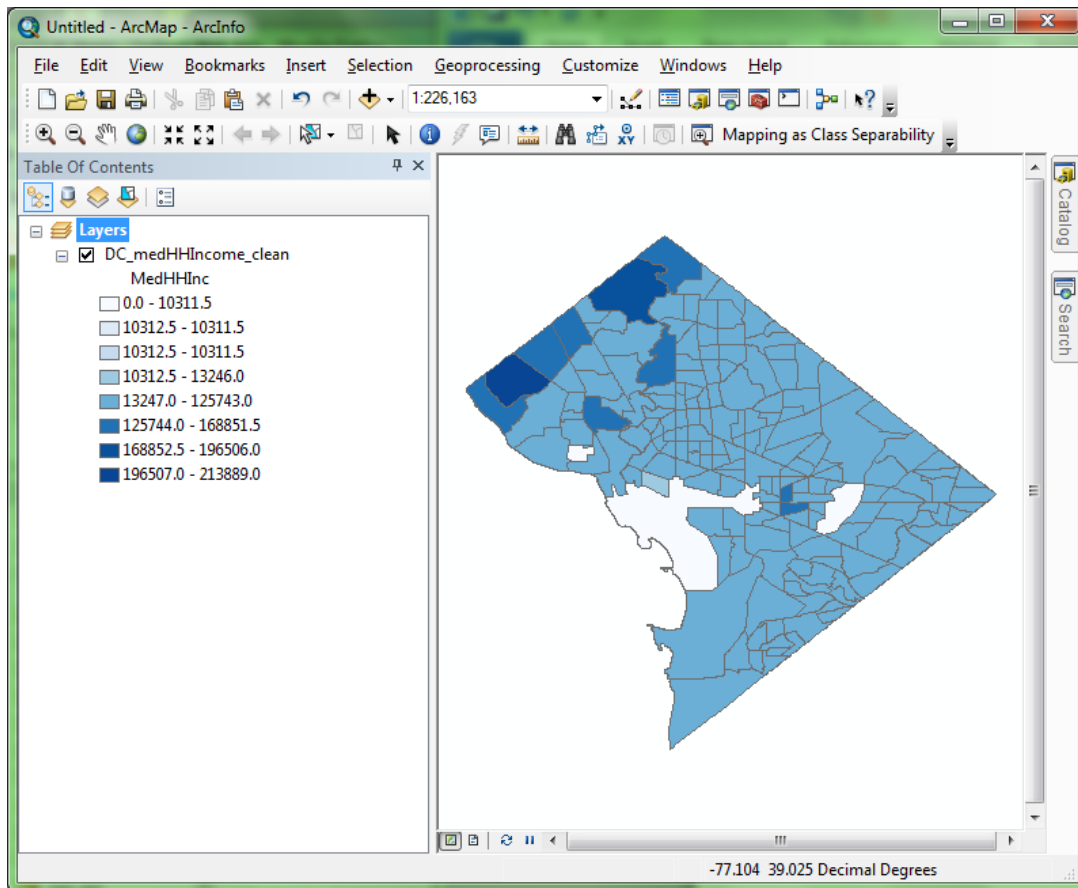
Also note that the bar plot includes two elements: a bar with its height representing the estimates, and an error line for each bar showing the corresponding margin of error. This bar plot also displays estimates in ascending order.



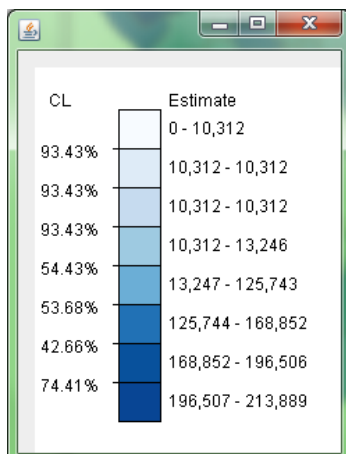
- 5) Users can drag the marker to a higher CL, likely resulting in fewer classes, or to a lower CL to have more classes. If the user can accept a CL of 40%, eight (8) classes would be formed in our example, but note that additional class breaks have different colors, reflecting their corresponding CLs that estimates are statistically different as determined by these class breaks.



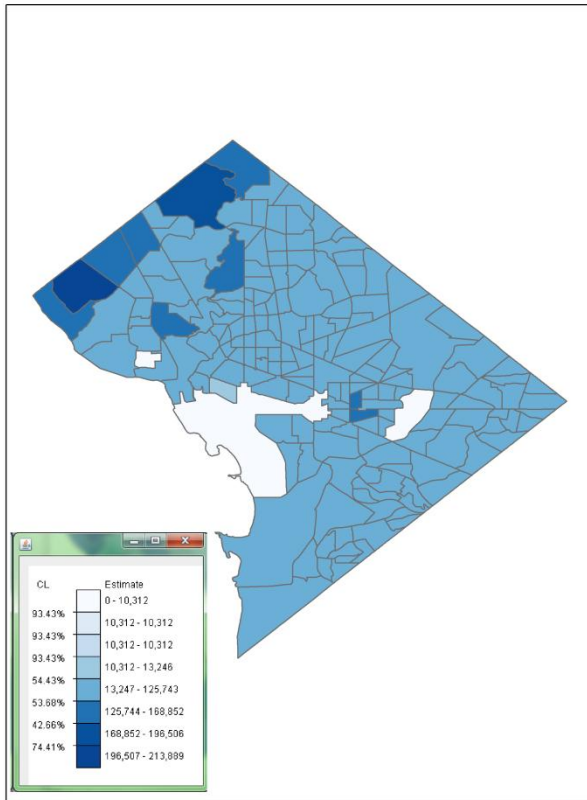
- 6) When users are satisfied with the classification, click the “Make Map” button to add a map layer in the Data View in ArcGIS, using the selected class breaks.



- 7) Users can click the Legend button to generate a legend designed to include the CLs of all class breaks, indicating the extent that estimates in different classes are different.



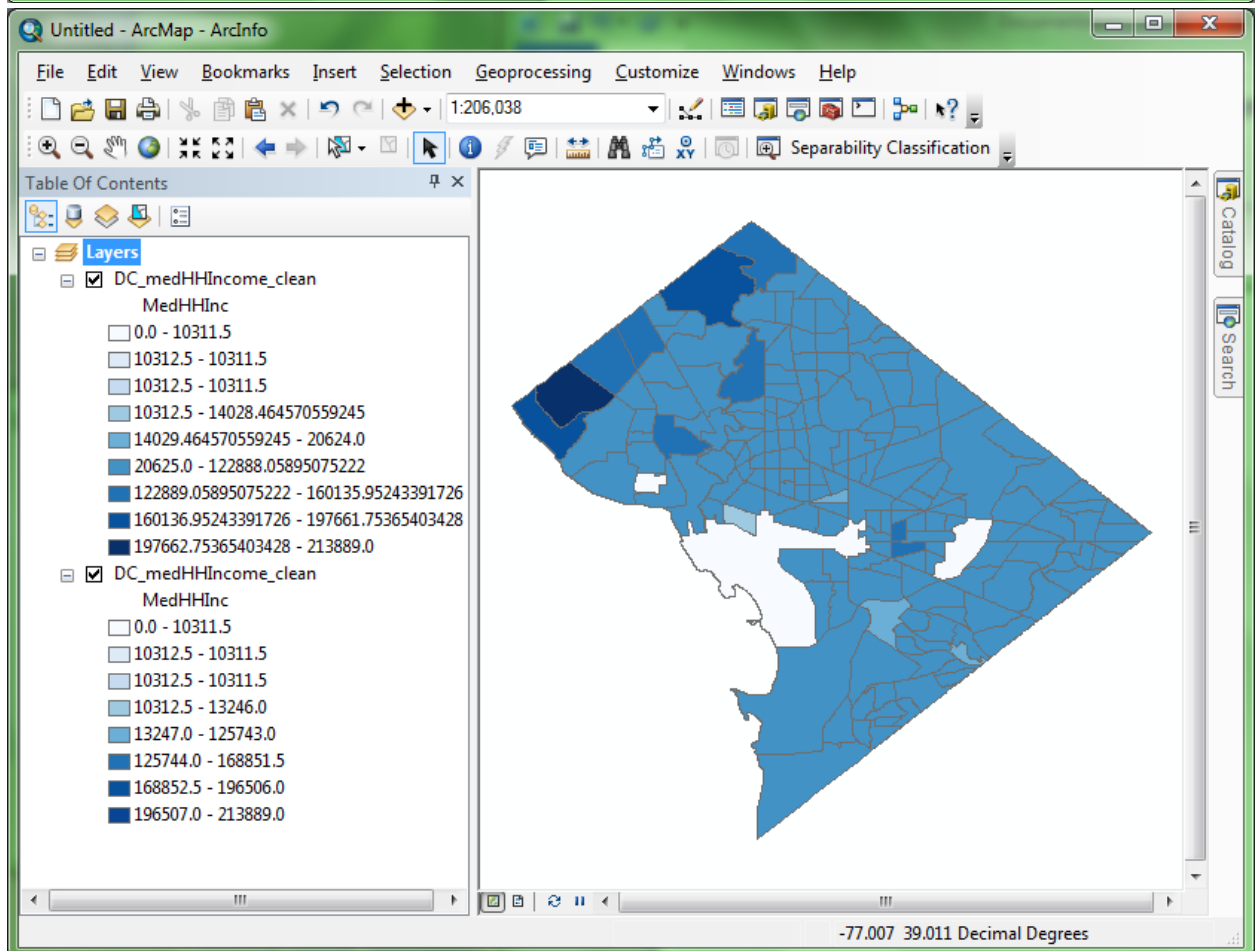
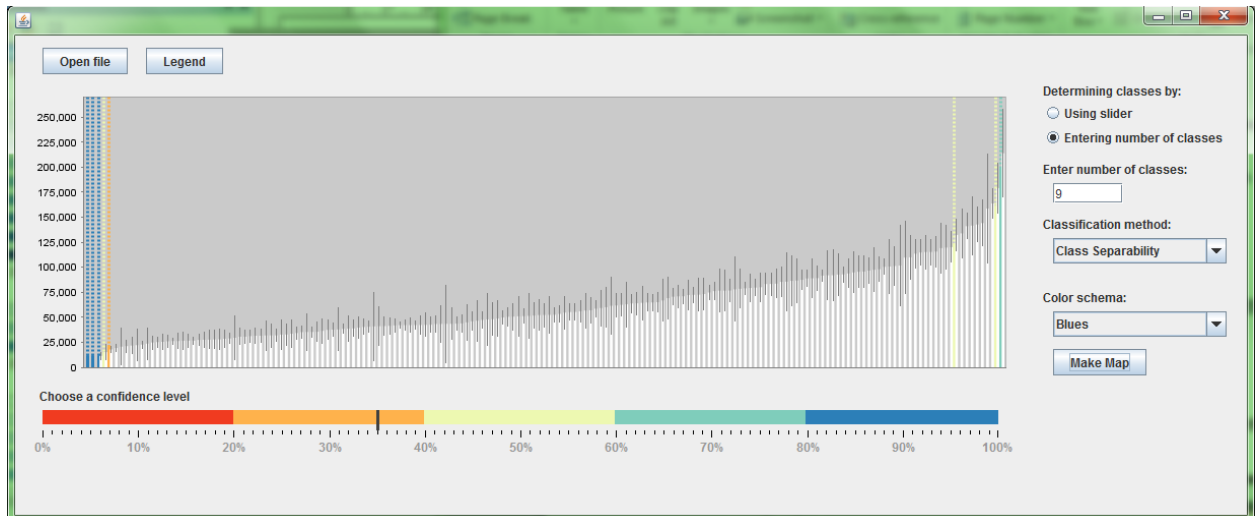
- 8) To use this legend in a map created from ArcGIS, a user may use the combined key strokes of “Function-Alt-Print Screen” to capture the legend window, and then paste it in the layout of ArcGIS instead of inserting the default map legend from ArcGIS.



We understand that this is not the most elegant way to incorporate the legend into the map, but we do have significant technical difficulties in exporting the legend to any graphic format after our tool (a program in Java) was packaged by ArcGIS to develop the Add-In. We will appreciate any solution to this technical challenge.

- 9) The classification tool has two other cartographic functions:
- Determining class breaks by the number of classes: Instead of dragging the marker on the slider bar to the acceptable confidence level to determine class breaks, users may enter the desired number of classes, and click the “Make Map” button again. Then the histogram will be

updated with the class breaks based upon the desired number of classes (the marker of the slider bar will be moved to the new CL level determined by class number), and a new map layer with the revised classification will be added to ArcMap.



- b. Using different color schema: we have adopted some color designs suggested by <http://colorbrewer2.org/> and users may choose different color schemes to render the map.

