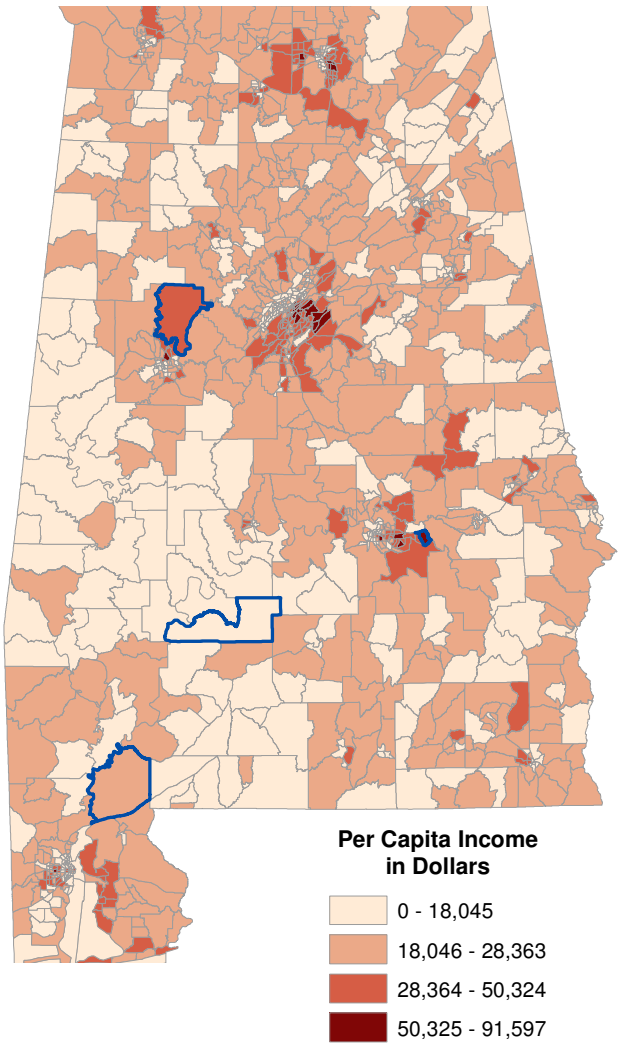
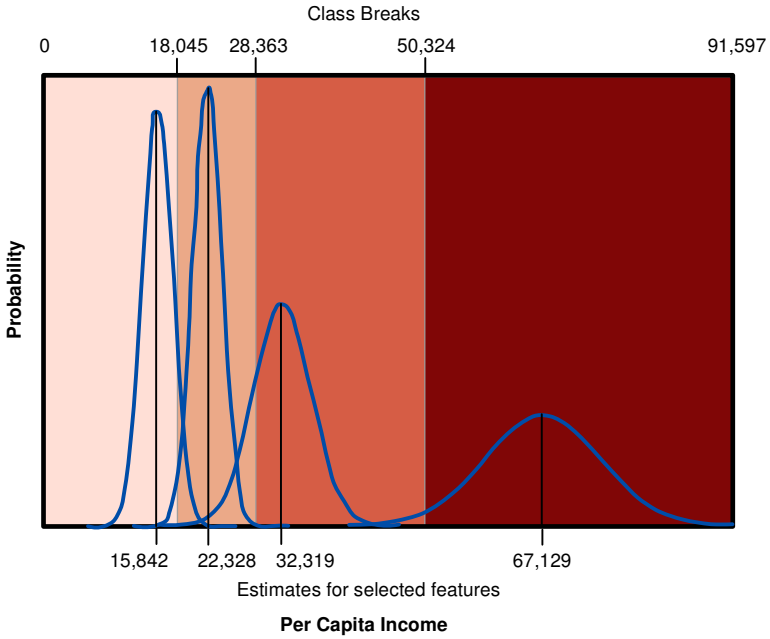
Classing Method Assistant Help

Summary -

Selecting a good classification method for data having a relatively large degree of uncertainty can be difficult for reasons described below. The Classing Method Assistant tool will assist the user in choosing a suitable classification scheme, based upon the data uncertainty, from among natural breaks, quantiles, equal intervals, and manual classification. The method used in this tool is discussed in Konstantin Krivoruchko's text *Spatial Statistical Data Analysis for GIS Users*, published by Esri in 2011.

The output results are presented in the Geoprocessing Results window. With regard to data uncertainty, the classification method with the highest **Total Probability** is likely a good method to choose for mapping. The user still must consider the data distribution and the purpose of the map, however. For some classification schemes, examples are often seen with equal intervals, a large proportion of the data estimates and their associated confidence intervals fall completely within a single class resulting in a high Total Probability.





Selected Features

	GEOID	Estimate	MOE_99	Lower	Upper
	01131	15,842	5,174	10,668	21,016
	01003	22,328	4,916	17,412	27,244
	01125	32,319	9,930	22,389	42,249
	01101	67,129	19,823	47,306	86,952

The figure above is adapted from Figure 2.11 in Krivoruchko's text. The map shows per capita income, in dollars, for census tracts in Alabama using a Natural Breaks classification scheme. We'll look more closely at the four selected features outlined on the map in dark blue. The rectangle to the right illustrates 1) the class intervals, in colors corresponding to the map, along with class break values, and 2) curves depicting the sample distribution for each of the four selected features, as well as the estimate (mean) for each feature. The Selected Features table shows the estimate, the margin of error (MOE) at the 99% confidence level, and the confidence interval, which is the range between the lower confidence interval boundary (estimate minus MOE) and the upper confidence interval boundary (estimate plus MOE).

Although we do not know the exact per capita income value for each feature, the mean is the value with the highest probability of being the true value. The curves have different degrees of spread because the margin of error (MOE) for each is different; the greater the MOE, the larger the spread of the curve. For example, the estimate of $67,129 has a MOE of $19,823, meaning the actual per capita income for this tract has a 99% chance of being between $47,306 and $86,952.

Examining the curves and class breaks, we can see that although the mean falls within the assigned class interval, there is some probability that the true per capita income value for each of the features may fall in a different class than that to which it is assigned. In fact, a substantial portion of possible values for the distribution with a mean of $32,319 fall in the class immediately below its assigned class.

The tool works as follows. The user specifies the number of classes, optimally a small number such as three or four, and the tool determines class intervals for natural breaks, quantiles, and equal intervals methods. Optionally, the user can manually specify breaks. For each method, the tool calculates the probability that the true value of each feature falls within its assigned class by calculating the area under the curve between the upper and lower class limits. Then, for each method, the tool sums the probabilities of all features to arrive at a Total Probability. The classification method with the highest Total Probability is likely the best method to choose for mapping. **The output results are presented in the Geoprocessing Results window.**

To further improve class break selection, the user may interactively work with the Statistical Difference – Geo tool to determine if features in different classes are statistically different from one another.

Notes:

- If input estimates or margins of error are undefined (e.g. -999), they are excluded in all calculations.
- If some features must be excluded from classing, simply select the features you *do* wish to include and run the tool.
- The Classing Method Assistant algorithms differ from Esri's algorithms, so the class breaks may not be identical. To be consistent with Esri class breaks, the user may enter breaks manually for various classing methods using Manual class breaks entry.

Input feature layer
Input feature layer name.

Number of classes
The number classes into which your data will be divided.

Estimate Field
Input estimate field.

Margin of error field
Margin of error field associated with the estimate field.

Margin of error confidence level
The confidence level, i.e. the level of precision, of the margin of error (MOE) field. MOEs are provided for all ACS data at the 90% confidence level, therefore 90% is the default value. However, users may have calculated MOEs at either the 95% or the 99% confidence level. Input confidence levels of 95% and 99% are available on the dropdown menu.

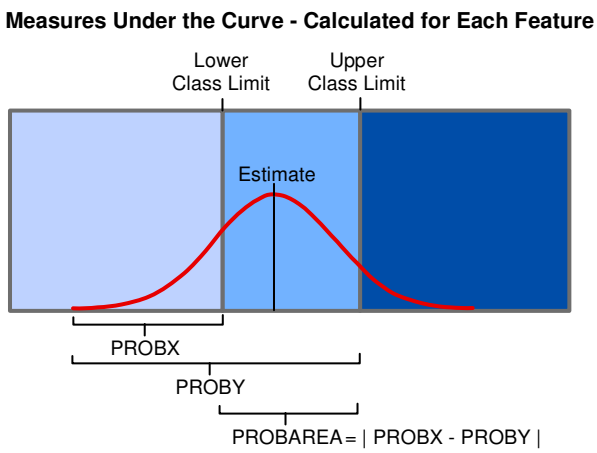
Manual class breaks entry (optional)
By default the tool calculates Total Probability for national breaks, quantiles, and equal intervals classification methods. This input box optionally enables you to add manual class breaks. To use this option:

- Note the Min. Value and Max. Value of the Min and Max Field Values box just above the Manual class breaks entry box. These are the lower and upper values of the features to be classed.
- In the Manual class breaks entry box, enter the upper value (i.e. limit) of your first data class, the click the + icon to add the value to your list of break values.
- Continue adding the upper value of the remaining data classes until you have specified the Max Field Value. The number of values you enter must equal the number of classes you specified.
- If you make an error, you can use the X icon to delete your entry.

Create feature class with additional probability fields

This is an optional toggle that enables you to output a new layer containing intermediate statistical calculations for each feature in the attribute table. The statistics returned include:

- SE – Standard error, which is MOE/1.645 for a 90% confidence level
- PROBX – the probability the estimate will be less than or equal to the lower class limit
- PROBY – the probability the estimate will be less than or equal to the upper class limit
- PROBAREA – the probability that the estimate falls within the upper and lower class limits for a specified classification scheme. In other words, the absolute value of PROBX-PROBY



Notes:

- PROBX, PROBY, and PROBAREA will be returned for each classification method (SE will be the same for all). Therefore, a prefix corresponding to the first letter of the method type will be added to the field name. For example, the PROBX for the equal intervals method will be E_PROBX. For a manual method, PROBX will be M_PROBX and so on.
- For estimates and/or MOEs with null or -999 values, the returned fields will contain a 0 value because they cannot be calculated. They are therefore not included in the total probability calculation.

Output feature class

The output layer location and name.