

Choosing a classification method

ArcView offers five classification methods for making a graduated color or graduated symbol map:

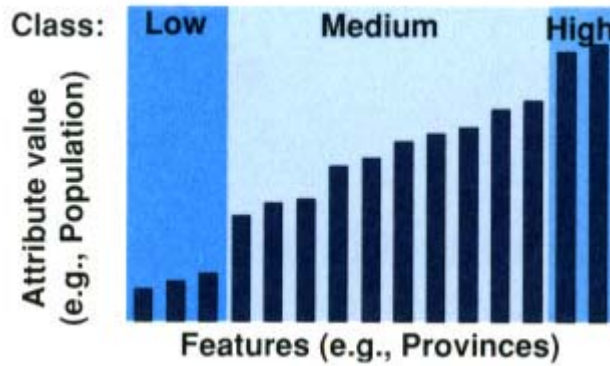
- Natural breaks
- Quantile
- Equal area (polygons only)
- Equal interval
- Standard deviation

You can also type your own class range values directly into the Legend Editor's Value field to specify your own classes.

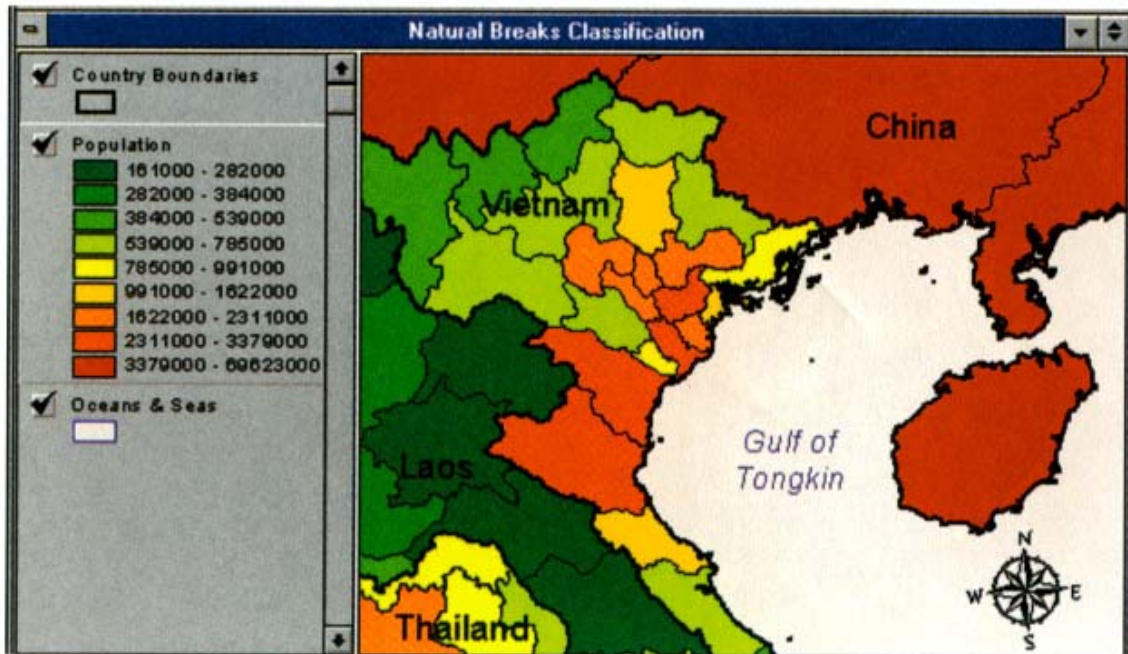
The classification method you choose depends on the nature of your data, and what you want to show about the data. The maps in the next section of this chapter illustrate how your choice of classification method, a chart is also shown illustrating how the same set of attribute values are divided into classes by that method.

Natural breaks

Natural Breaks is ArcView’s default classification method. This method identifies breakpoints by looking for groupings and patterns inherent in the data. ArcView uses a rather complex statistical formula (Jenk’s optimization) that minimizes the variation within each class. The example in the chart shows how this works: a set of features are ordered (left to right) from the smallest to the largest by population value. The features are divided into classes whose boundaries are set where there are relatively big jumps in the values.

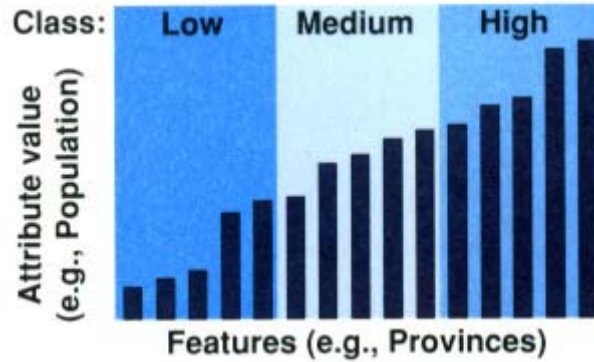


This map shows the population of provinces in part of Southeast Asia. The nine population classes were determined using natural breaks. On this map, the extreme values are obvious: the provinces in China have the highest population, while those in Laos have the lowest. Vietnam has provinces that range in population between the two extremes. This map presents population as you would expect it to look: it could be considered the most realistic view of the data.



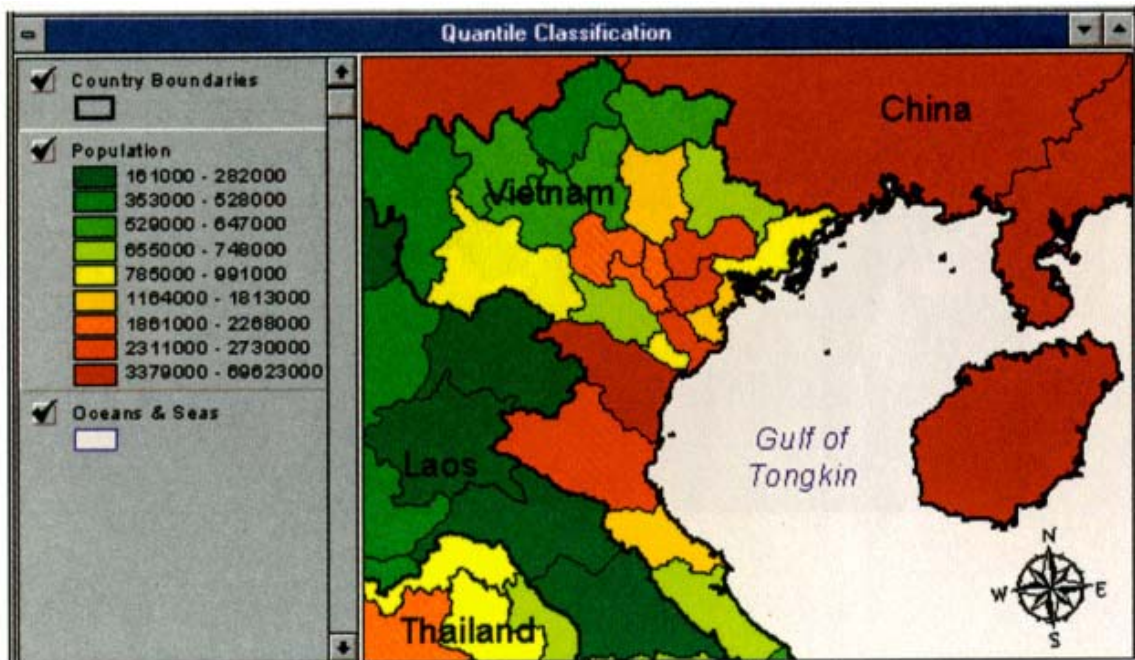
Quantile

In the quantile classification method, each class is assigned the same number of features. In the chart shown below, the lowest five provinces are put in the first class, the next five in the second class, and so on.



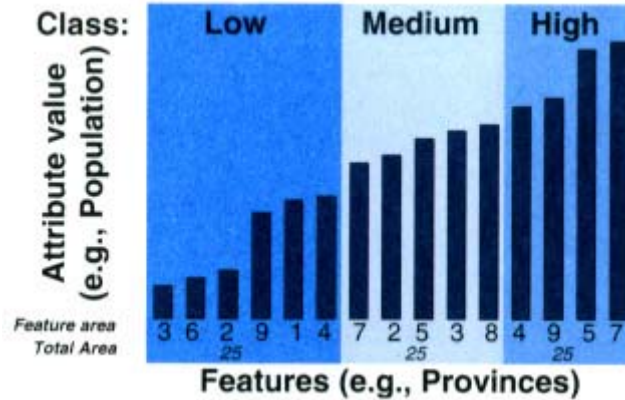
It doesn't matter that the provinces on either side of a class boundary have almost the same population. Quantile classes can therefore be misleading because low values are often included in the same class as high values. You can help overcome this distortion by increasing the number of classes. Quantile classification is best suited for data that is linearly distributed; in other words, data that does not have disproportionate numbers of features with similar values. It is especially useful when you want to emphasize the relative position of a feature among other features, for example, to show that a store is in the top one-third of all stores by sales.

This map shows the same data as the last map, but the population of the provinces is mapped using quantile classification. Differences among the provinces with an intermediate population, such as in Vietnam, can be more easily distinguished on this map than on the natural breaks map.



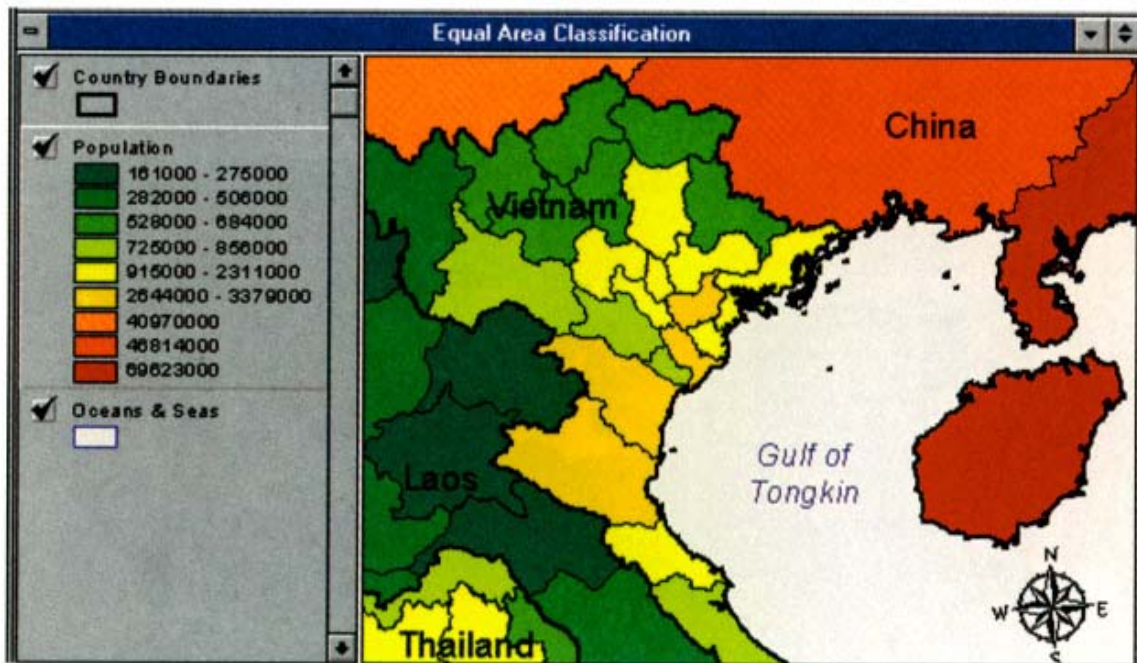
Equal area

The equal area method classifies polygon features by finding breakpoints in the attribute values so that the total area of the polygons in each class is approximately the same. ArcView determines the total area only from polygons that have valid data values for the attribute. Classes formed with the equal area method are typically similar to quantile classes.



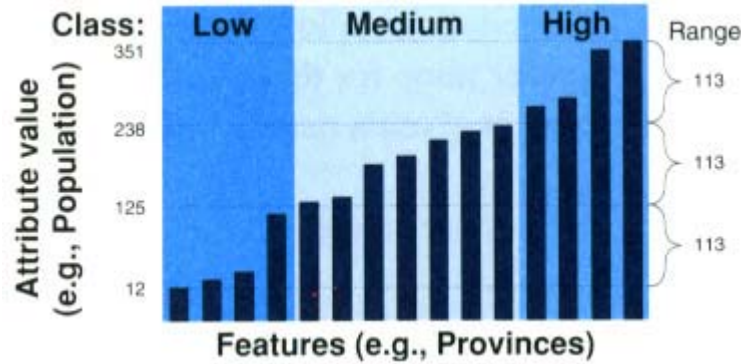
Equal area classification is similar to quantile classification except that each feature is given a weight in the classification equal to its area, rather than equal to 1.

When the population data is classified with the equal area method, the very largest provinces (in China) are in classes by themselves. The smaller provinces are put into the remaining classes. This tends to hide the variation in population between smaller provinces.



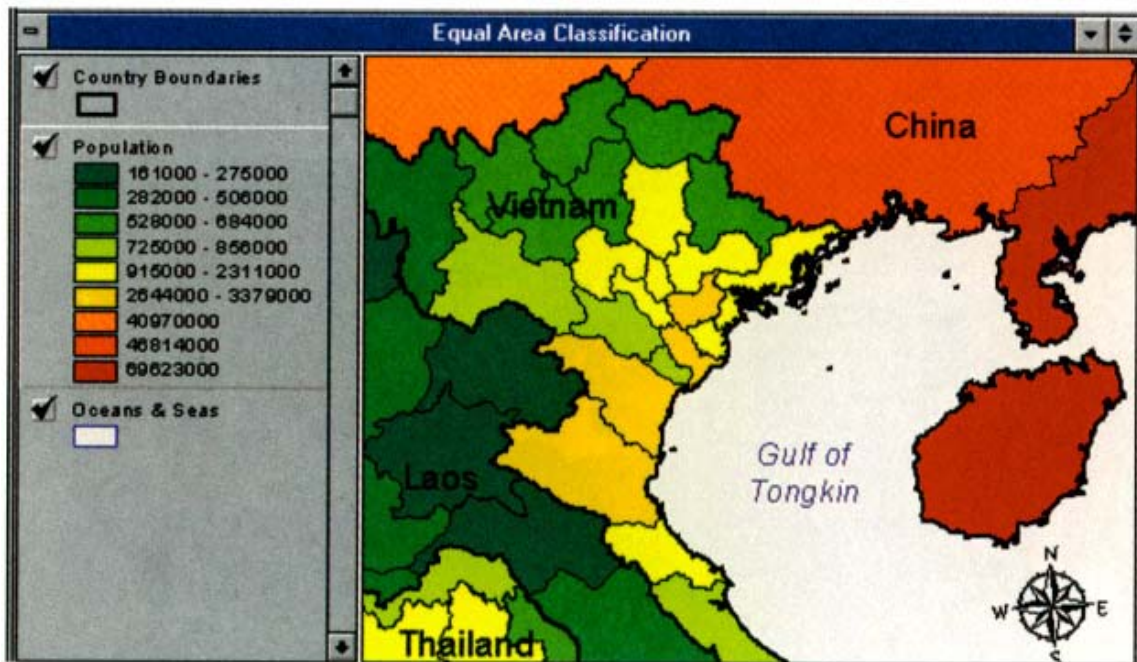
Equal interval

The equal interval method divides the range of attribute values into equal sized sub-ranges. For example, if the features in your theme have attributes values ranging from 12 to 351, the total range of the values is 339, so if you classify these features into three classes using the equal interval method, each class will represent a range of 113, and the class value ranges will therefore be 12-125, 126-238, and 239-351, as shown in the chart.



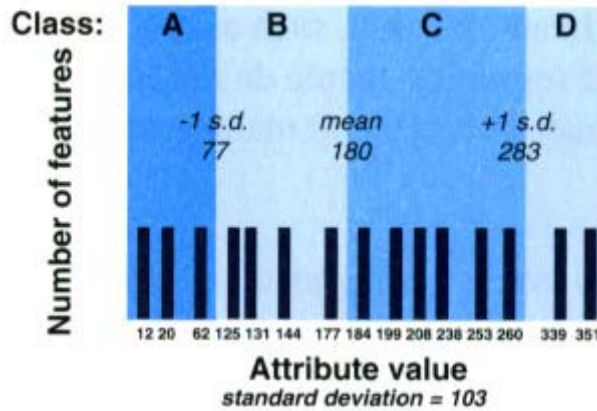
Equal interval classifications are useful when you want to emphasize the amount of an attribute value relative to other values, for example, to show that a store is part of the group of stores that made up the top one-third of all sales. Equal interval classification is ideal for data whose range is already familiar, such as percentages or temperatures. Population counts or other data for which people do not have strong conceptual association with a data range may be better communicated using another classification method.

In the map below, population is mapped using an equal interval classification. This map shows that there is a very great disparity between the least populous and the most populous provinces in this area. Because the Chinese provinces are so populous, all the provinces in Vietnam, Laos and Thailand fall into the first class (161,000 – 7,879,000). Clearly, the equal interval classification is not a good one to use if you want to reveal subtle differences between features with fairly similar values.

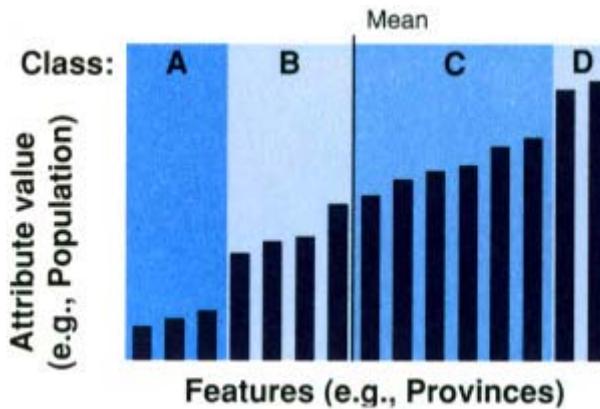


Standard Deviation

Standard deviation shows you the extent to which an attribute's values differ from the mean of all the values. When data is classified using the standard deviation method, ArcView finds the mean value and then places class breaks above and below the mean at intervals of either 1, 0.5, or 0.25 standard deviations, until all the data values are included in a class. ArcView will aggregate any values beyond three standard deviations from the mean into two classes: greater than three standard deviations above the mean ('>3 Std. Dev.') and less than three standard deviations below the mean ('< -3 Std. Dev.'). On a graduated color map, the default color ramp for this classification is *dichromatic* (e.g., blue to red) and the mean data value is given a neutral color (e.g., white).



The chart below shows the same set of attribute values presented on the vertical axis so you can compare this chart to those for the other classification methods.



In the map below, population is mapped using a standard deviation classification. All the provinces in Vietnam, Laos, and Thailand all fall slightly below the mean, while the provinces in China fall well above the mean. It is clear from the population classes listed in the view's Table of Contents that the population data is skewed by the highly populous Chinese provinces, because there is only one class below the mean but there are seven classes above the mean.

