

Mapping Spatial and Statistical Distributions in a Choropleth Map

By Naresh Kumar, Brown University

Choropleth maps, which are created by shading or coloring areal enumeration units with an intensity proportional to the data value associated with those units, are the most easily understood type of map. However, because mapping enumeration units vary in size, map readers cannot accurately guess the statistical distribution of a choropleth map. This necessitates the inclusion of statistical parameters and a frequency histogram in addition to a conventional legend.

Using a frequency histogram requires additional space at the expense of the map body. This goes against one of the key principles of map design—that the map body should be as large as possible given the constraints of the page margin and the need to adequately represent the other map components. This problem can be overcome by using frequency histogram legends that will not only show the statistical distribution of the data array but also provide a key to the map.

Designing choropleth maps with the integrated frequency histogram legends is a complicated process because GIS and mapping packages do not provide automated solutions for this procedure. However, the ability to programmatically modify GIS software can be harnessed to customize mapping applications so that they will generate the frequency legends. This article describes a customized solution for designing classed and unclassed frequency legends in choropleth maps implemented as an ARC Macro Language (AML) script for use with ArcGIS 8.2 (using an ArcInfo license).

Legends for Classed and Unclassed Maps

Frequency legends can be created for both classed and unclassed choropleth maps. In a classed map, the main task is to determine height and shading or color intensity of the bar for each class. Because unclassed maps lack classes, it is logical to use a frequency curve that varies the shading or color intensity from the lowest to the highest specified value.

In the application, a classed frequency legend includes a frequency histogram, a frequency curve, and shading or coloring. In the frequency histogram, the bar height of each class will vary in proportion to the number of observations as shown in the legends in the left side of Figure 1. A normal curve overlaid on a histogram can be useful in examining the symmetry of the distribution.

In a conventional unclassed map, a continuous ramp of shading or color intensity is used.

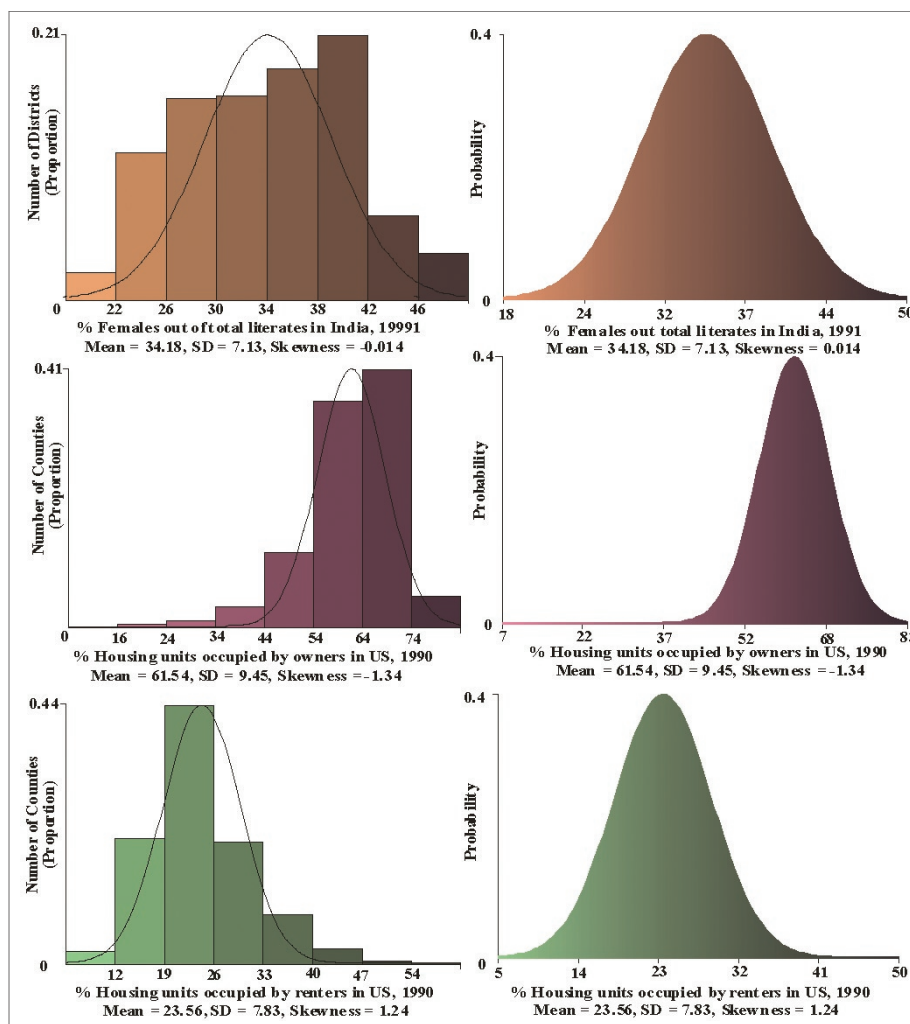


Figure 1: Classed and unclassed frequency legends for different statistical distributions. Secondary sources, 1990 United States Census and Indian Census data, were used for the frequency legends.

The lowest and highest intensities are assigned to the lowest and highest values respectively. A distinctive feature of an unclassed map is that each unique value has unique shading or color intensity. This characteristic requires as many unique bars as unique values in the data. If there are many observations, the width for each x value will be so narrow that any shading pattern will be almost invisible. Consequently, a frequency curve is used for an unclassed frequency legend with the variations in shading and color intensity along the x -axis showing the trend of the statistical distribution. Examples of frequency curve use for unclassed frequency legends are illustrated on the right side of Figure 1.

Methods Used

The frequency legend maps were developed using the ArcPlot module in ArcInfo Workstation 8.2. To design choropleth maps, the user needs an error-free polygon coverage with proper topology, a mapping variable in the polygon attribute table (PAT), and a variable with a unique identification number (UIDN) for each administrative (i.e., enumeration) unit.

In addition, a projected coverage should be used so a map scale can be drawn. The inclusion of the UIDN in the coverage is also necessary so that it can be used to filter out two or more values of the same administrative/enumeration unit, especially for those coverages with islands. For example, an ArcInfo coverage of the

Form

Program for Frequency Legend Choropleth Map

Select a Polygon Coverage
(Right Click to Browse the Coverage)

Select a variable (numeric) to be mapped
(Right Click to Browse the Variable)

Check this box to avoid duplicate frequency, e.g. frequency for each country is 1, but a country with 30 islands will have a frequency of 31 ☒

Check this box if you want to include 0 (zero) in mapping ☐

Classed or Classless Map: ☒ Classed ☐ Classless

Select a Hue Value in Between 1 to 360 (0 for Grey)
Yellow = 60, Green = 120, Cyan = 180
Blue = 240, Magenta = 300, Red = 360

Figure 2: The basic input form for the frequency legend map design application

frequency legends for choropleth maps. The algorithm begins with input from a form in

state of Rhode Island will generate a frequency of more than one because the state coverage contains several polygons (i.e., the coverage includes islands) although, in a state-level map for the United States, it should generate a frequency count of one.

A user-friendly form was designed in ArcPlot that is linked to the main algorithm. This algorithm, written in an AML, produces classed and unclassed

which the user selects an ArcInfo coverage and a variable from the PAT of the selected coverage. The user also enters input parameters, such as legend type and color hue, as shown in Figure 2. Next, the user specifies a page layout and map scale parameters. The user is prompted with the statistical parameters that include arithmetic mean, standard deviation (SD), and minimum and maximum values. These parameters help the user select class limits (in a classed map) and change the mean, SD, minimum, and maximum values for the frequency curve as needed. A map with a frequency legend incorporating the selected options is produced as illustrated in Figures 3 and 4.

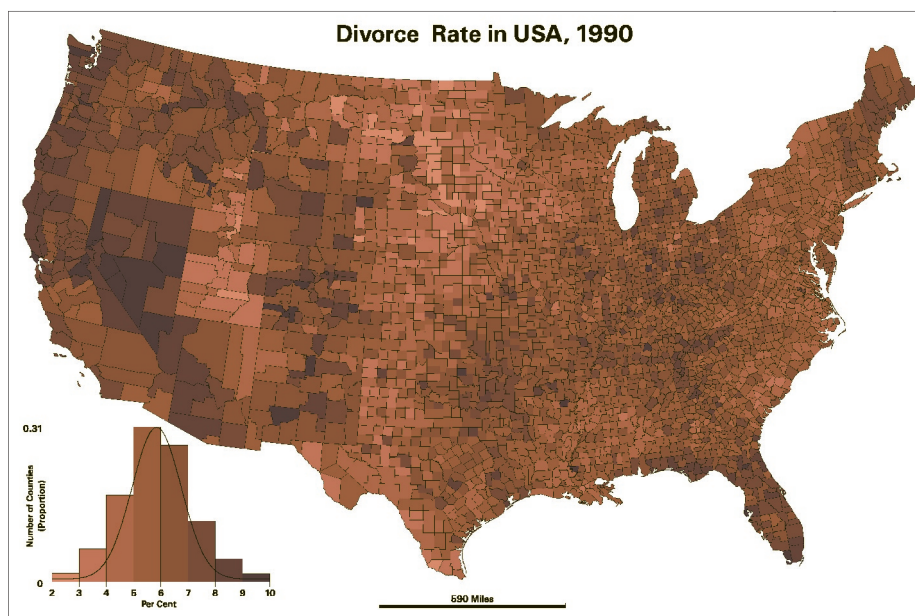


Figure 3: Classed frequency legend showing the divorce rate in the United States in 1990

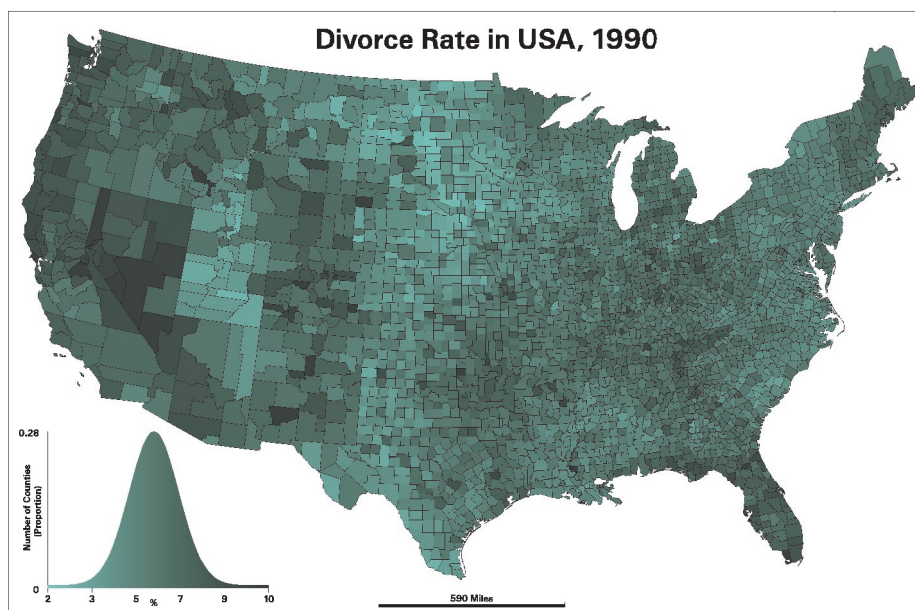


Figure 4: Unclassed frequency legend showing the divorce rate in the United States in 1990

Conclusion

The use of frequency legend maps allows mapmakers to depict both spatial and statistical characteristics of mapped phenomena. These statistical characteristics are unavailable from a conventional map. Although both classed and unclassed frequency legends can be created for any statistical distribution, these legends work best for a symmetrical data distribution.

As part of this project, a survey that tested how ordinary users responded to frequency legend maps vis-à-vis maps with a conventional legend was administered to 121 students. University undergraduates and graduates from several disciplines participated in the survey. Three important findings emerged from the analysis of this survey data. First, the majority of respondents were able to capture both spatial and statistical distribution from the maps with frequency legends—one of the main objectives of designing these maps. Second, frequency legends worked best for the statistically trained audience. Finally, users responded more effectively to frequency legends in a normal statistical distribution vis-à-vis in skewed statistical distribution.

The lack of automated solutions generating frequency legends in GIS and mapping packages restricts their use to experts. Consequently, options for creating these legends should be integrated into future versions of GIS and mapping software so that novice GIS users and cartographers can take advantage of this new dimension in choropleth mapping. The use of frequency legends with the main statistical parameters can help with the data exploration that is a necessary component of GIS.

For more information, contact
Naresh Kumar, Assistant Professor
Population Studies and Training Center
Box 1836
Brown University
Providence, Rhode Island 02912
E-mail: Naresh_Kumar@Brown.edu
Web: nk.gis.brown.edu