# Chapter 2. Mapping GIS Data

### Objectives

- Using properties of symbols to differentiate features or rasters in maps
- Distinguishing among nominal, categorical, ordinal, and ratio/interval data
- Creating maps from attributes using different map types
- Selecting appropriate classification methods when displaying numeric attributes
- Displaying thematic and image rasters
- Understanding the relationship between a layer and its source data set

## Mastering the Concepts

### GIS Concepts

Geographic Information Systems are used for many purposes, but creating maps is one of the most common. The practice of cartography, or mapmaking, has a long history. Although a computer makes the process easier and offers the ability to explore and edit designs, the cartographer must still understand the basic principles behind portraying spatial data and the aesthetic challenges in designing an effective and attractive way to communicate ideas.

#### *Choosing symbols for maps*

Cartographers may choose from many strategies to symbolize features in a map. A layer may simply be portrayed using one symbol, or different features can be assigned different symbols depending on the value of an attribute field. Point data are shown with marker symbols, line features with linear symbols, and polygon features with shaded area symbols (Fig. 2.1).

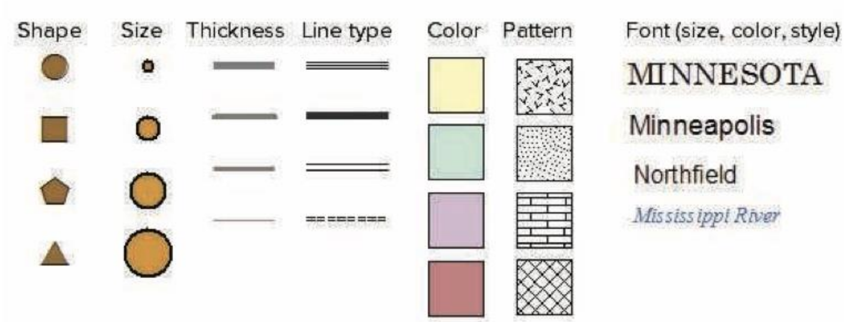


Fig. 2.1. Variations in symbols used to differentiate objects in a map

Cartographers have many ways to signify differences between features in a map: shape, size, thickness, line type, color, pattern, and font (Fig. 2.1). Traditionally these variations are used to show either changes in category (the type of thing) or changes in quantity. Shape, line type, pattern, and font are typically used to show changes in category, such as different types of wells (points), different classes of roads (lines), or different soil units (polygons). Size and thickness are generally used to indicate increases in quantity, such as the population of a city (points) or the discharge of a river (lines) (Fig. 2.1). Text symbol variations usually indicate categories (towns vs. rivers), although font size can indicate qualitative differences in value, such as town size, as long as not too many different sizes are used.

Color may be used to indicate either category or quantity. Colors may be designated using one of several common methods. The **CMYK** model is often used for printing, and it specifies mixtures of inks used in printers or plates (cyan, magenta, yellow, black).

The **RGB** model is based on mixing different proportions of red, green, and blue light on a scale from 0 to 255. If the value of the red light is high (close to 255) and the other two colors have low values near zero, a bright red color will result. If red and green are both high and blue is low, a mixture of red and green will create yellow. (Mixing light is a different process than mixing paint, and different colors will result.) Figure 2.2 shows two possible mixtures and the resulting colors. The RGB model can specify $256^3$ different colors (over 16 million).

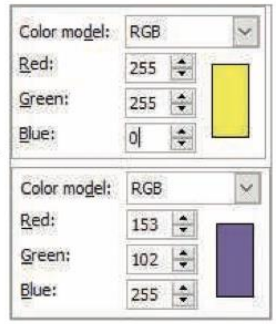


Fig. 2.2. Computers store and identify colors as mixtures of red, green, and blue light, each measured on a scale from 0 to 255.
Source: Esri

The **HSV** method is instructive for discussing the use of color in portraying features on a map. HSV stands for **hue**, **saturation**, **and value** (Fig. 2.3). **Hue** refers to the shade of color (such as red, blue, or yellow) and is established by the wavelength of the light observed. Typically hue is portrayed as a color wheel so that the color values range in degrees from 0 to 360. **Saturation** corresponds to the intensity of the color and is measured as a percentage. Imagine mixing a can of paint, starting with a white base and adding a single pigment—a small amount of pigment yields a low saturation, but a large amount of pigment results in high saturation. **Value** refers to how light or dark the color is, somewhat like putting pigment into a can of base paint that varies from black through gray to white. Value is also measured as a percentage. Any color can be defined using a combination of the three properties.

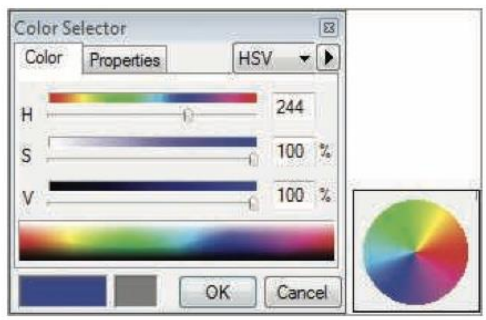


Fig. 2.3. Hue, saturation, and value method of defining color
Source: Esri

A fourth parameter, named **alpha**, may be used with either the RGB or HSV model. It refers to the opacity of a color, or the degree to which it is transparent or see-through, like tinted glass. A zero value indicates that the color is fully transparent, and the highest value, which may be represented either as DN = 255 or 100%, means it is not transparent at all.

The set of colors chosen for a map layer should follow certain guidelines. Figure 2.4 shows three different sets of five colors. A set based on variations in hue should be used to depict changes in category, such as different soil units (Fig. 2.4a). For categories, the saturation and value of the colors should be similar. Quantities are generally indicated using differences in saturation or value, with light or unsaturated colors indicating lower quantities

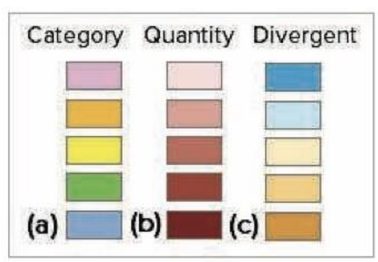


Fig. 2.4. Color ramp variations

and darker or more intense colors indicating higher quantities (Fig. 2.4b). Sometimes a **divergent color set** is helpful in showing variation around a significant middle value (Fig. 2.4c), such as in a climate change map. Areas with no significant temperature change are shown in the middle neutral color, colder temperatures are shown in blue, and warmer temperatures are shown in orange.

The importance of black-and-white symbol schemes cannot be neglected. In commercial printing or copying, color still costs more than black and white. Publishing figures in professional journals or reports may also cost significantly more in color. One should consider, too, how most viewers will see the map. Figures in a master's thesis may look wonderful in color, but some of the people reading it may receive a black-and-white version from inter-library loan. There are many reasons why one might wish to design a map in black and white at the outset. In a black-and-white map, only four or five different shades of gray are typically discernible, so symbol selection relies heavily on variations in shape, size, thickness, line type, or pattern rather than value.

Combinations of these factors may also be used when specifying symbols, such as using points that change in both shape and color, polygons that change color and pattern, or text that has differences in size and font and style. However, combinations should be used sparingly, as the more complex and varied the symbols become, the more difficult it may be to interpret them. In Figure 2.5, both volcanoes and highways are shown using combinations of color, shape, and size (thickness). The combinations serve to accentuate differences within the layer and could be used to emphasize one type of feature, such as using a larger, saturated pink symbol to call attention to the stratovolcanoes. However, this strategy would backfire if multiple point layers were being displayed because it would become difficult to interpret which features belonged to which point layer.


Fig. 2.5. Examples of categorical data representation for points, lines, and polygons

Color choice is complicated by the fact that about 10% of men and 1% of women are color-blind. Different types of color blindness are possible, although red-green color blindness is the most common. If layers are being symbolized for data exploration by one person, then the user may employ any set of colors that aids in the interpretation of the data. However, if the maps are intended for use by many people, it is better to avoid red-green color sets and rely on saturation, value, pattern, and shape variations instead. If divergent colors or multiple hues are needed, combinations such as brown-purple or orange-blue can be interpreted by most color-blind viewers.

Cartographers also need to be sensitive to conventions and connotations associated with different colors or symbols. A **convention** refers to the use of a particular color or symbol in a commonly understood way; for example, using blue to display water, blue and red to contrast cold with heat, or a blue cross to indicate a hospital. Conventional symbols help readers to interpret the map. **Connotation** is an emotional or psychological impact associated with a particular symbol, such as the color red indicating danger, or red, white, and blue evoking feelings of patriotism. However, connotations are often culturally specific. Red may indicate danger in the United States, but it is associated with joy in Thailand and with national pride in China. Making maps for an international audience demands particular care and knowledge in the choice of colors and symbols. Context can matter as well; even in the United States, a red-green map at Christmas might engender more joy than fear.

### *Types of data and types of maps*

Geographers, statisticians, and others characterize measurements or attributes as belonging to one of a set of data types: **nominal, categorical**, **ordinal**, **interval**, or **ratio**. The type of data affects how it should be stored in a database and what types of analyses or statistics are appropriate. In mapping, the data type also influences the kind of map used to display the feature attributes.

**Nominal data** name or identify objects, such as the names of states. Nearly every feature will have a different name. Nominal data are often text values, but they don't have to be: a parcel number or a tax identification number also serves the purpose of uniquely identifying an object. Nominal data are usually portrayed on a map by labels. Each layer typically has one text symbol, but different text symbols may be used to accentuate different layers, such as using all uppercase for state names or italic blue text for rivers.

**Categorical data** separate features into distinct groups or classes. Figure 2.5 shows examples of categorical data for point features (volcano type), linear features (road class), and polygon features (land cover). Other examples include soil types, ethnic groups, and geological rock formations. Categorical data are often stored as text, but it is also possible to represent the categories with numeric codes, such as Commercial Services = 20 and Industrial = 40. Categorical data are represented by a **unique values map**, which gives each category a different symbol based on shape, line type, color, or pattern (Fig. 2.5) as shown in the geological rock units map in Figure 2.6.

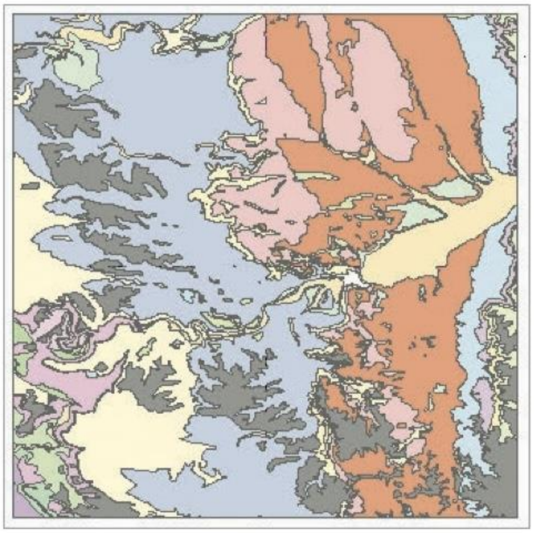

Fig. 2.6. A unique values map based on geological rock types
Source: South Dakota Geological Survey

**Ordinal data** have categories that are ranked based on some quantitative measure, although the measure may not be linear. For example, urban settlements might be classified as villages, towns, or cities. Students are assigned grades of A, B, C, D, or F. Soils are designated as A, B, C, or D depending on their infiltration. Ordinal data must be represented by unique values maps if the values are text, but value or saturation should be varied instead of hue or pattern so that the quantitative increase can be easily interpreted. Numeric ordinal data may be represented either with unique values maps or graduated color maps.

Quantitative data represent phenomena that fall along a regularly spaced measurement scale, such as distance or rainfall. Equal changes in interval involve equal changes in the quantity being measured. For example, the energy needed to heat a thimble of water from 16 to 17°F is the same as that needed to heat it from 96 to 97°F. **Ratio data** have a meaningful zero point that indicates the absence of the thing being measured. Precipitation is an example of ratio data; zero precipitation corresponds to a total lack of rain, and two inches of rain is twice as much as one inch. Ratio data support all four arithmetic operations of addition, subtraction, multiplication, and division. **Interval data** have a regular scale but are not related to a meaningful zero point. Temperature data measured in the familiar Celsius or Fahrenheit scale are interval data because a temperature of

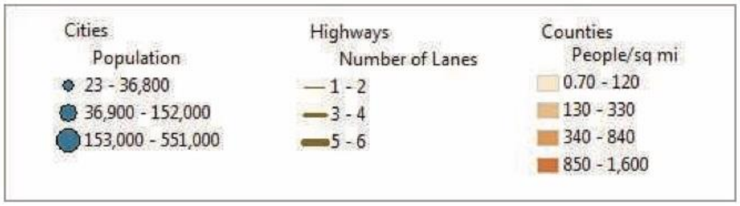


Fig. 2.7. Examples of numeric data representations for points, lines, and polygons

zero does not correspond to a complete lack of temperature (heat energy). Any scale that can have negative values, such as elevation, is an interval scale. Interval scales support only addition and subtraction. Today's high temperature of 80°F might be hotter than yesterday's high of 40°F, but it cannot be said to be twice as hot, except as a figure of speech.

Quantitative numeric data take on values along a continuous scale of possibilities; every state, for example, has its own population value. In order to symbolize numeric data, the values must be partitioned into groups with specific ranges.These ranges are called classes, and the maps are called classified maps. In Figure 2.7, the cities and roads have three classes of values, and population density has four classes. Each class is symbolized using variations in symbol size, thickness, or color saturation or value.

Point or line data are usually displayed by varying symbol size or thickness and are portrayed using a **graduated symbol map**. However, point or line data can also be portrayed using a **proportional symbol map** in which the numeric value is used to proportionally determine the size of the symbol. Instead of a few size classes, the map has a continuous range of symbol sizes. This style is often referred to as an unclassed map. Figure 2.8 compares graduated symbol and proportional symbol maps for the populations of state capitals.

Numeric data classes for polygons are represented using color-shaded symbols (Fig. 2.9); the resulting maps are called **graduated color maps** or **choropleth maps**. The maps are usually symbolized using changes in value or saturation (with a monochromatic color ramp) so that the increase in quantity is clear, as shown in the county population map in Figure 2.9a. Although rainbow color ramps are popular in software packages, using them for graduated color maps is usually a mistake because it is difficult for the mind to interpret hues as representing larger or smaller quantities (Fig. 2.9b). Generally the eye cannot distinguish between more than seven or eight levels of the same color, so no more than that number of classes should be used.

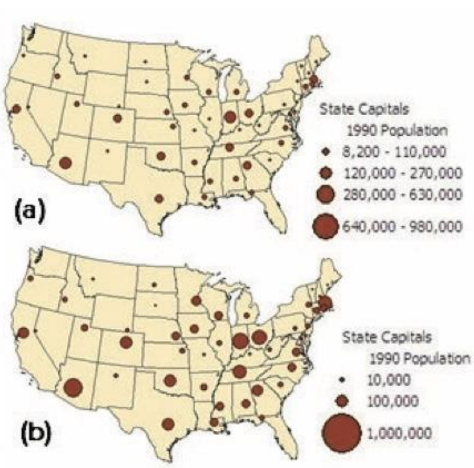


Fig. 2.8. Comparison of (a) a graduated symbol map and (b) a proportional symbol map showing the populations of state capitals
Source: Esri

### *The modifiable areal unit problem*

Data are often measured over one set of areal units, such as a census block, but then combined, or **aggregated,** into a set of larger areal units, such as block groups, tracts, and counties. When the measurement and aggregation units are arbitrary shapes and sizes instead of a regularly spaced grid, the size and shape of units will influence the recorded values. For some measurements, such as the number of farms or lightning strikes, units with larger areas tend to have higher values just because they are larger. In Figure 2.10a, the number of farms is greater in the bigger states like Texas and California. Certain variables, such as the number of home vacancies, are strongly linked to population, and maps made from them reflect population rather than giving insight into underlying patterns. In Figure 2.10b, the largest and most populous states have the most vacancies.

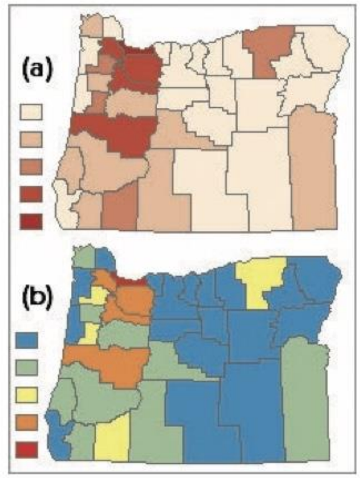


Fig. 2.9. A choropleth map showing population by varying (a) saturations and/or value, and (b) hue
Source: Esri

Another issue that can arise is that larger polygons receive greater prominence in the map simply because of their size. In Figure 2.10, the large western states attract more attention than the eastern states. Both of these issues are examples of a phenomenon known as the **modifiable areal unit problem**, or **MAUP**, which occurs when measurements are being aggregated over arbitrarily defined areas. Analyzing spatial or statistical patterns becomes more difficult because the patterns are obscured by the aggregation scheme. However, methods to reduce the effects of MAUP on maps are available.

One approach is to **normalize** the data, dividing each value by a specified variable. If one is concerned that the size of the aggregation area is influencing the magnitude of values, one can divide by the feature area as shown in Figure 2.11a, where the number of farms is divided by state area (compare to Fig. 2.10a). If the values are being influenced by population, then one can divide by population or other suitable variable, as in Figure 2.11b, where the vacancies are divided by the number of housing units, giving a far more interesting and informative map than Figure 2.10b. Another normalization method, less commonly used, shows percentages of the total quantity falling in each feature; Figure 2.11c shows the percentage of Congress controlled by each state, accomplished by showing the number of districts in each state as a percentage of the total number of districts.

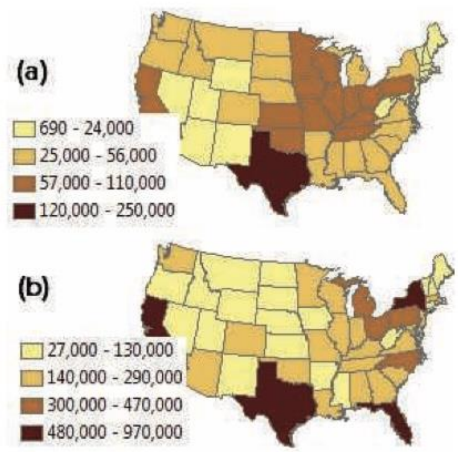


Fig. 2.10. (a) Number of farms; (b) number of vacant housing units
Source: Esri

However, not all attributes need normalization. In a precipitation map, each value already represents inches of rain falling at any location. The median rent for a county does not depend directly on the area of the county. These attributes would not be normalized. Before mapping any attribute, it is important to stop and consider how the aggregation unit might affect the map and determine whether the values should be normalized, and by what.

The visual MAUP issue, which occurs when large polygons dominate the map, can be addressed by changing the map type. A graduated symbol map applied to polygons places a symbol at the center of each polygon, minimizing the visual imbalance, as shown in a graduated symbol map of the density of farms (Fig. 2.12a; compare to Fig. 2.11a). A **dot density map** uses randomly placed dots to show the magnitude of a value in the attribute table (Fig. 2.12b). Each dot represents a certain number; in this map, 2000 farms. Note that the locations of the dots, which are randomly placed, do not necessarily represent the distribution of farms in the state.

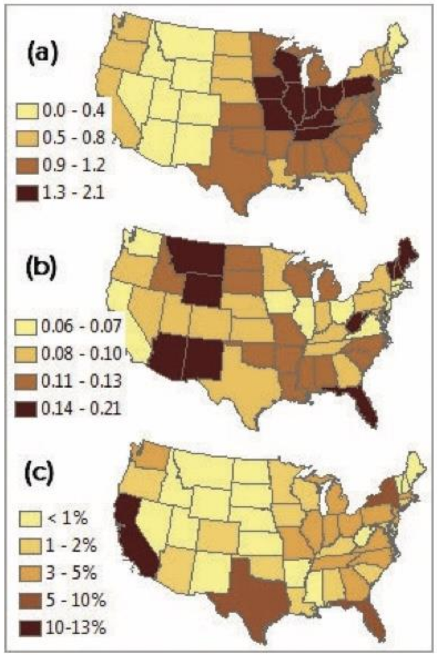


Fig. 2.11. (a) Farms normalized by state area; (b) vacancies normalized by housing units; (c) percentage of Congress controlled by each state
Source: Esri

#### Chart maps

A **chart map** expands the number of attributes that can be displayed on a map by replacing a single symbol with a chart representing several attributes. The chart could be a pie chart, bar chart, or stacked bar chart. Figure 2.13 shows a pie chart map with the proportions of Caucasians, African Americans, and Hispanics in each state. The pie sizes can be all the same (Fig. 2.13a) or proportional to the sum of the three categories, thus showing the relative number of people in the state as well (Fig. 2.13b). Notice how the pie colors are chosen to facilitate recognition of the classes: cream for Caucasians, dark brown for African Americans, and reddish for Hispanics. Such details help make maps easier for the reader to interpret, and they lessen the need for the reader to look back and forth between the legend and the map.

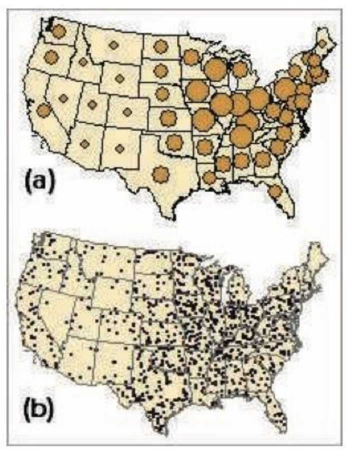


Fig. 2.12. Number of farms shown using (a) an area-normalized graduated symbol map or (b) a dot density map
Source: Esri

### *Displaying rasters*

Recall from Chapter 1 that the raster model is a cell-based data model in which an array of cells or pixels store numeric values relating to some feature or quantity on the earth's surface. Rasters may be broadly grouped into three main types, each of which is displayed differently: thematic rasters, image rasters, and indexed rasters.

#### Thematic rasters

A **thematic raster** represents features or quantities, such as roads, geology, elevation, or vegetation density (Fig. 2.14). The raster model is designed to store **continuous** values that may occur anywhere on the earth's surface, such as elevation. However, they may also store **discrete** objects, such as roads or land use polygons. Thus we can speak of a raster as being a continuous or a discrete raster.

Values in a raster are always stored as numbers and may be characterized as categorical, ordinal, or interval/ratio. (Nominal data are not stored in rasters—it would be inefficient.) The methods used to display thematic rasters are similar to those used for displaying vector data and also depend on the data type.

A discrete raster representing categorical or ordinal data is displayed using a unique values map. Just like the unique

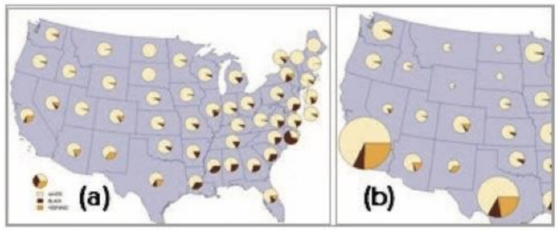


Fig. 2.13. (a) A chart map can represent several attributes, such as the proportion of ethnic groups. (b) The pie size can represent the total population.
Source: Esri

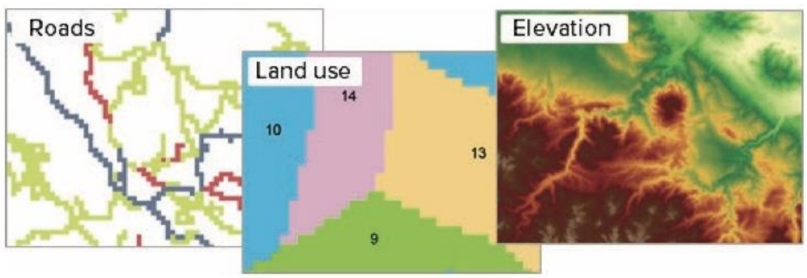


Fig. 2.14. Thematic rasters. Discrete rasters store feature data, such as road types or land use polygons. Continuous rasters store values that exist everywhere, such as elevation.
(Roads): Source: Black Hills National Forest; (Land use and Elevation): Source: USGS

values map used for data sets, each value of the raster receives its own color, as in the geology map shown in Figure 2.15a. Color schemes for unique values maps have 32 possible values and work best with data that have relatively few categories. When representing ordinal data, a monochromatic color scheme is used to communicate a sense of increase.

A thematic raster containing interval or ratio data must be classified into ranges in one of two ways. The **classified** display method divides the values into a small number of bins, similar to

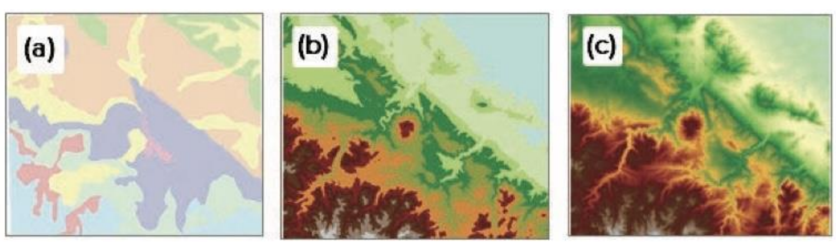


Fig. 2.15. Display methods for thematic rasters: (a) unique values geology; (b) classified elevation; (c) stretched elevation
(a): Source: South Dakota Geological Survey; (b-c): Source: USGS

a graduated color map. A color ramp is chosen to assign the colors to each bin. The elevation map in Figure 2.15b has 12 classes represented by 12 colors from a color ramp. The **stretched** display method scales the image values to a color ramp with 256 shades (Fig. 2.15c). The raster is first subjected to a **slice**, which rescales the elevation values (ranging from 800 to 1600 m) into 256 bins, then matches the bins to the 256 shades in the color ramp. A slice is essentially a classification that creates 256 classes.

The colorful ramps, like those in the elevation rasters in Figure 2.15, ignore the guideline to use saturation or value rather than hue to represent quantities, yet the eye has no trouble interpreting the elevation increase. Continuous rasters typically represent surfaces with an underlying structure, such that low values occur near low values and high ones near high. This structure imposes a visual order and allows the eye to interpret the greater number of colors. Such ramps are not suitable for features or all rasters, but they can be effective when appropriate.

The property that values close to each other tend to be more similar than values farther apart is present, to a greater or lesser extent, in many geographic data sets. It is so common, in fact, that this property is often called the first law of geography, or **Tobler's law**. Not only does it play a role in data visualization, as discussed previously, but it is also a fundamental factor in spatial analysis.

#### Image rasters

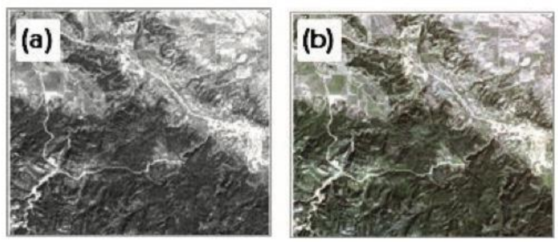


Fig. 2.16. Raster display methods for images: (a) stretched image; (b) RGB composite image.
Source: USGS

**Image** rasters include aerial photography and satellite data. The pixels represent degrees of brightness caused by light reflecting from materials on the surface. The brightness values are usually placed on a scale of 0 to 255 DN (digital numbers). If one brightness value is given to each pixel, the image is usually displayed with a grayscale ramp; a dark shadow would have a low brightness and a low DN, and a white cement road would have a high brightness and a high DN (Fig. 2.16a). Such images are displayed using the same **stretched** method that is applied to continuous thematic rasters. Images based on the 0 to 255 scale do not need to be sliced, but some images contain larger values and do require slicing. To display a color image, each pixel is given three DN values representing red, green, and blue light, thereby specifying a unique color combination, as discussed earlier. This display method is called an **RGB composite** (Fig. 2.16b).

#### Stretching image values

Image DN values are often normally distributed. Figure 2.17a shows a typical **histogram** of cell values, with the horizontal axis showing the range of values with the corresponding grayscale color ramp and the vertical axis showing the number of cells for each value (color). The image has few values near 0 or 255, with most of the pixels occurring in the middle gray range. The image is shown in Figure 2.17b, and it appears dim and featureless with little contrast.

The **stretched** display method can improve the display of normally distributed values by ignoring the tails of the distribution. One method assigns the lowest and highest image values to 0 and 255, respectively, and stretches the remaining colors along the ramp. This is called a minimum-maximum stretch. Even greater contrast can be developed using a standard-deviation stretch, which uses only the values within two standard deviations of the mean (Fig. 2.17c).

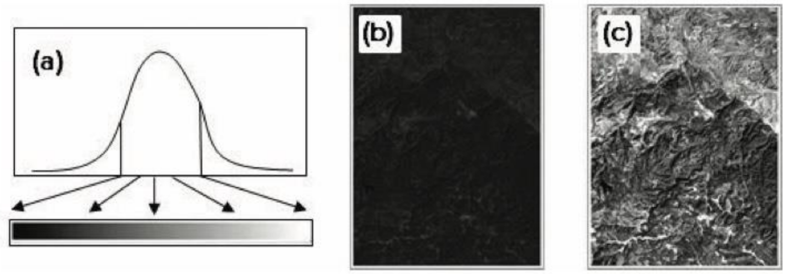


Fig. 2.17. The effects of stretching: (a) stretching an image histogram; (b) no stretch applied; (c) a standard-deviation stretch
(b–c): Source: USGS

Stretches can be applied to thematic continuous rasters as well as images and can be helpful when the image values are not evenly distributed.

#### Indexed color rasters and colormaps

The RGB color model is capable of storing millions of different colors and is best for fine rendering of images. However, it uses large amounts of memory and storage space. An alternate method for storing color identifies a restricted set of colors that appear in the image. It assigns an integer from 0 to 255 to each color and stores the RGB proportions needed to make it. This list of colors is called the **colormap** and is stored with the image. Each value in the raster is portrayed using its assigned color. Figure 2.18 shows part of a digital raster graphic (DRG). A DRG is a scanned US Geological Survey topographic map. The contours are represented by pixels with the value 4 and displayed using the assigned brown color. Every pixel value and defined color used in the map is shown in the accompanying colormap: green for vegetation, black for text, and so on.

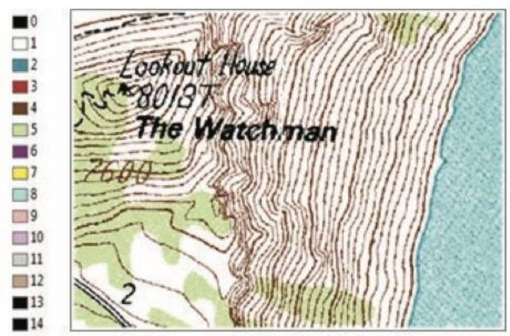


Fig. 2.18. An indexed-color raster stores a colormap that assigns a color to each value in the raster.
Source: USGS

Indexed color rasters are commonly produced when scanning large color maps. Storing them in RGB could use hundreds of megabytes or more. An indexed-color map typically uses less than a tenth of the space needed for an RGB map, and the eye cannot easily distinguish between the two.

### *Classifying numeric data*

Mapping numeric data, raster or vector, requires classifying a range of values into a small number of groups, each of which can be represented by a different color or symbol size. This process is called **classification**. There are many ways to classify data, and the choice of method affects the appearance of the map and the message that it portrays. Some common methods are compared in Figure 2.19 using the same data set of average farm size by state.

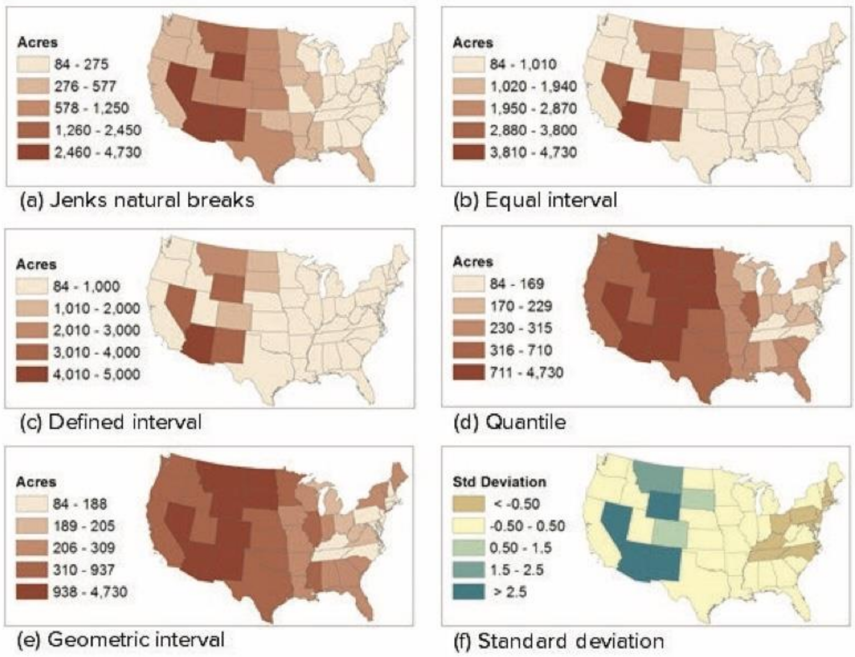


Fig. 2.19. Comparison of different classification methods using the same data set of average farm size in acres by state
Source: Esri

The **Jenks method** sets the class breaks at naturally occurring gaps between groups of data (Fig. 2. 19a). Each class interval can have its own width, and the number of features in each class will vary. The Jenks method works well on unevenly distributed data, such as populations of the capitals shown in Figure 2.8. There are many low-population and medium-population capitals and a few very high-population capitals. The Jenks method works well with almost any data set, making it a natural choice for the default classification scheme in ArcGIS Pro.

The **equal interval** classification divides the values into a specified number of classes of equal size (Fig. 2.19b). This method is useful for ratio data, such as income or precipitation, because it gives a sense of regularity to the observed increases. However, it is hard to predict how many features will end up in each class. Notice in the farm size example that nearly all of the states fall into the first class. Compare this map to Figure 2.19a.

A **defined interval** classification is similar to an equal interval one, except that the user specifies the size of the class interval, and the number of classes then depends on the range of values (Fig. 2.19c). This method will create rounded values in the classes that are easy to interpret. Defined interval maps are ideal when comparing classes composed of percentages, dollars, temperatures, and other values when specific break values are desired (100, 200, etc.). It does, however, suffer the same disadvantages as the equal interval classification.

A **quantile** classification puts about the same number of features in each class (Fig. 2.19d) and enables the display of groups, such as quartiles, commonly of interest in statistics. This method will create a balanced map with all classes equally well represented, but some of the features in the same class could have very different values, and features in different classes could have similar values. Quantile classifications are best applied to uniformly distributed data. Notice that a quantile classification highlights differences between the eastern states that are hidden in the Jenks, equal interval, and defined interval classifications.

A **geometric interval** classification (Fig. 2.19e) bases the class intervals on a geometric series in which each class is multiplied by a constant coefficient to produce the next higher class. A geometric interval classification is designed to work well with continuous data like precipitation and to provide about the same number of values in each class range. It works especially well with positively skewed data distributions.

A **standard deviation** classification apportions the values based on the statistics of the field. The user selects the class breaks as the number of standard deviations, and the data range determines the number of classes needed. This method excels at highlighting which values are typical and which are outliers, especially since a divergent color set is used to accentuate below- versus above-normal values. In Figure 2.19f, the yellow states are close to the mean farm size, several eastern states appear to have smaller-than-average farms, and some western states have much larger farms than average. This map is best applied to normally distributed data.

Finally, if none of the preceding options gives the desired map, the user can manually set class break points to any chosen values. This option works for assigning a logarithmic or exponential scale as well.

#### Choosing the classification method

The choice of classification method depends on the mapmaker's purposes and on the type of data. Jenks shows the "nearest neighbors" in a distribution, whereas a defined interval or equal interval map does a better job of portraying relative magnitude. Notice the difference between the maps in Figures 2.19a and 2.19b. The Jenks map gives the impression that many states have large farms because the class sizes increase slowly; an equal interval map shows that farms in most states actually average 1000 acres or less. Figure 2.19b makes this observation clear at first glance.

Some data possess the quality that the magnitude of the values has an intrinsic meaning. Percentage data, for example, have a physical and a psychological meaning—people have an intuitive understanding of the difference between 50% and 100%. Differences in median rent occur in dollar values to which people can attach meaning. When dealing with such data, it is wise to use a defined interval map to choose classes with logical break points, such as 20%, 5 inches of rain, or $200, rather than 12.6%, 1.47 inches of rain, or $187. The reader can more effectively interpret the classes.

The distribution of values has impact also. Jenks and geometric interval classifications are designed to work well with unevenly distributed data. Equal interval, defined interval, and quantile maps can be used with any data but show better results with evenly distributed data. The statistics behind standard deviation maps assume that the data are normally distributed.

## About ArcGIS

### *Layers*

When a data set is added from a folder to a map in ArcGIS Pro, we call it a **layer**. Although we tend to think of these layers as data sets, there is actually an important distinction. When a picture is inserted in a Word document or a slide show, a *copy* of the picture is stored inside the document, so it is always there if the document gets copied or e-mailed. Even if the original picture is deleted from the hard drive, the copy remains in the document. We are accustomed to this treatment of information, so it is difficult to realize that Pro does NOT treat map data this way. We will learn more about the reasons in Chapter 5.

When data are added to a map, no copy is made. The information remains in its original location, and the map layer simply points to it. The original data on which the layer is based are known as the **source data**. When a layer's properties are changed, such as giving it labels or displaying it with different colors, the source data set does not change. The layer finds the source data each time and manipulates them within the map to produce the desired effect. A layer is like a cooking recipe—it describes where to find the data file (the ingredients) and how to display it (the cooking instructions). The information in the layer, such as which symbols should be used and whether the features have labels, is called the **layer properties**. Layers are held in memory and stored when the project is saved. If the source data set is deleted, renamed, or moved, the map will be unable to display it.

When working in Pro, it is important to always remember that a layer and the geospatial data it references are different objects. The data set may be stored on a disk, or it may come through a portal from ArcGIS Online or another server. ***A layer points to the location of a geospatial data set and stores information about how to display and use it***. The same data set can be accessed by multiple layers and maps to present different views of the same information, whether within a single map, within the same project, or across many maps created by different users. Figure 2.20 shows a counties data set stored on a network drive; the same data set has been used by several students in their projects.

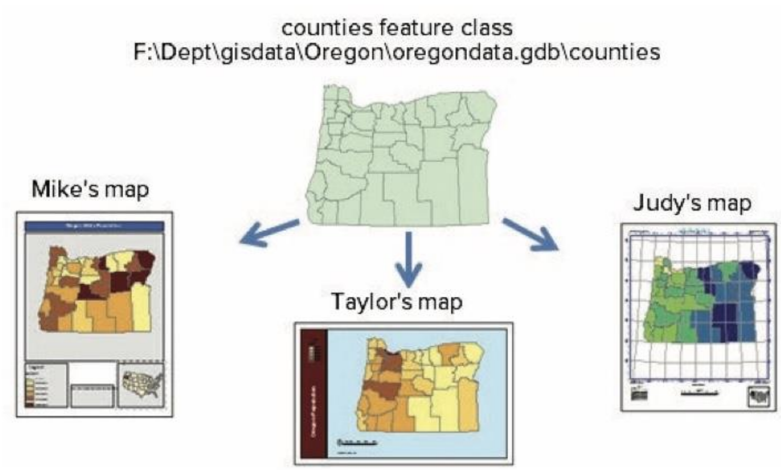


Fig. 2.20. The same feature class can be used in many maps, even by different people.
Source: Esri

A layer may be saved as a **layer file**. This file stores the location of the referenced data set and the properties of the layer. Saved layer files can be used to quickly add a layer with predefined symbols to a map document. Group layer files can organize multiple layers with predefined symbols into thematic maps or base maps. Geologic map symbols, for example, use many specific colors and patterns and can take hours to set up.

In Figure 2.21, a user has combined different sizes, colors, and even shapes to produce a specific classification that shows the arsenic values for water wells, with x indicating values below the detection limit, blue circles to show values less than the Environmental Protection Agency (EPA) standard, and orange circles for values above the EPA standard. By saving the finished layer as a layer file, we can quickly transfer the classification scheme to new layers and preserve the consistency of the symbols.

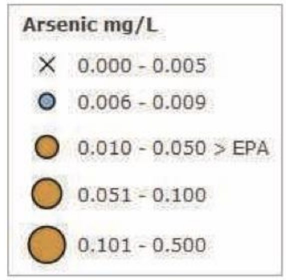


Fig. 2.21. Custom arsenic classification combining shape, color, and size

### *Styles*

ArcGIS provides an extensive set of tools for symbolizing features and rasters. Within the Symbology pane (see Fig. 2.22), the user can choose from a variety of predefined symbols and modify the properties of those symbols, such as the fill color, outline width, and outline color, when the available ones are not suitable. The software provides complete flexibility to modify individual symbol layers and create virtually any symbol desired.

Symbols are organized into groups for ease of use. A **style** contains a set of symbols with a related theme and typically includes different symbol types, including point markers, line styles, polygon shades and patterns, color sets, text styles, and even north arrow or scale bar symbols. Pro has two default styles named ArcGIS 2D and ArcGIS 3D. The ArcMap program included many additional styles customized for specific industries or purposes, with themes such as Geology 24K, Environmental, Civic, Crime Analysis, or Hazmat. Although these styles do not appear within Pro, they can be imported if needed.

Users with very specific symbol needs can create their own styles and fill them with symbols copied from other styles, modified from other styles, or newly created. A search function in the Symbology pane allows the user to search by keyword for matching symbols in any of the installed styles. Searching for `tree`, for example, yields various symbols from different styles.

## Summary

- Differentiating between features on a map requires variations in symbol shape, size, thickness, line type, color, pattern, or font.
- Attribute data have a data type designation: nominal, categorical, ordinal, interval, or ratio. The data type determines the kind of map and even the types of analysis that may be performed on that attribute.
- Nominal data name things or uniquely identify them and may be text or numbers. Each feature usually has its own value, and repeats are the exception.
- Categorical data group objects into smaller sets identified by a unique value. Numbers may be categorical when they are used as codes. Ordinal data consist of categories that are ranked in some way.
- Interval data are measured on a regular scale, and ratio data are measured on a regular scale with a meaningful zero point.
- Single symbol maps and labels are used to map nominal data. Unique values maps are used for categorical or ordinal data. Interval or ratio data are displayed using graduated color, graduated symbol, proportional symbol, dot density, or chart maps.
- Rasters may contain thematic or image data. Discrete thematic rasters can be displayed using a unique values method. Continuous thematic rasters may be classified or stretched. Image rasters use a stretched or RGB composite display.

- Continuous numeric data are classified before being mapped. Classification methods include Jenks natural breaks, equal interval, defined interval, quantile, geometric interval, standard deviation, and manual. The best classification method depends on the type of data and the data distribution.
- Data sets become layers when added to a map. Layers reference the original data and store information on how to display them.
- In ArcGIS, symbols are grouped into similar themes called styles. Tools are available to allow the user to create virtually any symbol needed for a map.

**Important Terms**

| | | | |
|---|---|---|---|
| aggregated | discrete | interval data | ratio data |
| alpha | divergent color set | Jenks method | RGB composite |
| categorical data | dot density map | layer | saturation |
| chart map | dynamic labels | layer file | slice |
| choropleth map | equal interval | layer properties | source data |
| classification | geometric interval | modifiable areal unit problem (MAUP) | standard deviation |
| classified | graduated color map | | stretched |
| CMYK | graduated symbol map | nominal data | style |
| colormap | histogram | normalize | thematic raster |
| connotation | HSV | ordinal data | Tobler's law |
| continuous | hue | proportional symbol map | unique values map |
| convention | image | quantile | value |
| defined interval | | | |

## Chapter Review Questions

1. For each of the following types of data, state whether it is nominal, categorical, ordinal, interval, or ratio. Explain your reasoning.

   arsenic concentration in mg/L
   vegetation type
   annual precipitation in inches
   customer zip codes
   social security numbers
   elevations of climate stations
   football team rankings
   number of students in universities
   letter grades given to students
   college student majors

2. For each of the following attributes, state whether a single symbol, graduated color, or unique values map would be most appropriate. Explain your reasoning.

   political party chosen by voting districts
   lung cancer rates by county
   flow rates of wells in gallons/minute
   river flow volumes (cfs)
   restaurant locations
   soil class, such as loam, clay, etc.

3. If mapping the following attributes for counties, indicate which ones would generally be normalized, and discuss what attribute field(s) should be used.

   average annual snowfall
   home vacancy rate
   Native American population
   median home price
   number of crimes reported
   number of car accidents

4. State whether a unique values, classified, stretched, or RGB composite display method should be used for each of the following rasters, and justify the choice.

| | |
|---|---|
| land cover classes | black-and-white aerial photo |
| 7-band satellite image | tree canopy percent |
| elevation | slope in degrees |

5. Explain the merits of unclassed maps as compared to classed maps.
6. Describe the difference between a geospatial data set and a layer.
7. Explain the difference between thematic rasters and image rasters.
8. For each of the following map types, find a map from the Internet that uses this type of symbolization: graduated color, graduated symbol (lines), graduated symbol (points), unique values. Turn in a screenshot of each map with a citation. For each one, explain how the differentiation between features is achieved, and critique the choice of symbols used.

**Expand your knowledge**

Expand your knowledge of ArcGIS Pro by exploring these sections of the Help.

Help > Maps > Author maps > Symbology > Symbolization

Help > Maps > Author maps > Text

Help > Projects > Project Items > Styles

Help > Data > Data types > Imagery and raster > Appearance and Symbology