3

Mapping the most and least

Mapping the most and least lets you compare places based on quantities so you can see which places meet your criteria, or understand the relationships between places.

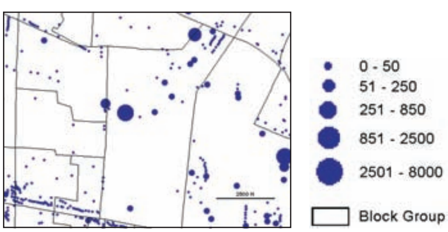
In this chapter:

- *Why map the most and least?*
- *What do you need to map?*
- *Understanding quantities*
- *Creating classes*
- *Making your map*
- *Looking for patterns*

People map where the most and least are to find places that meet their criteria and take action, or to see the relationships between places. To map the most and least, you map features based on a quantity associated with each. For example, a catalog company selling children's clothes would want to find ZIP Codes with many young families with relatively high income. Or, public health officials might map the numbers of physicians per 1,000 people in each census tract to see which areas are adequately served, and which are not.

Mapping features based on quantities adds an additional level of information beyond simply mapping the locations of features. For example, mapping the locations of businesses gives a sense of where workers are, information that might be useful for a transportation planner. But mapping the businesses based on the number of employees at each business gives a much better picture of where employees are.

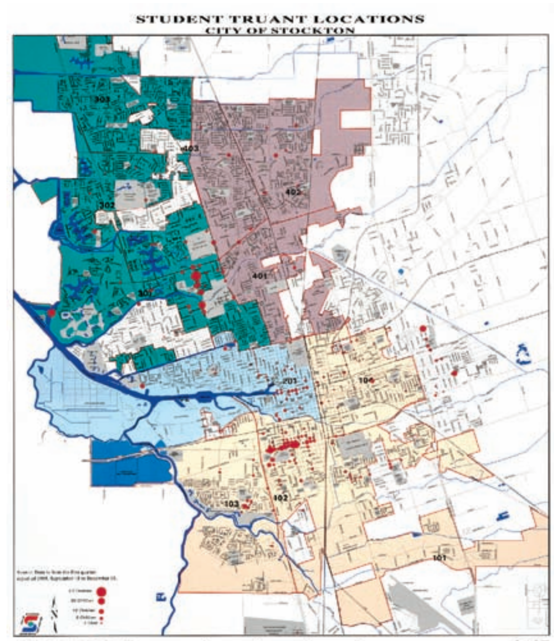
Locations of businesses



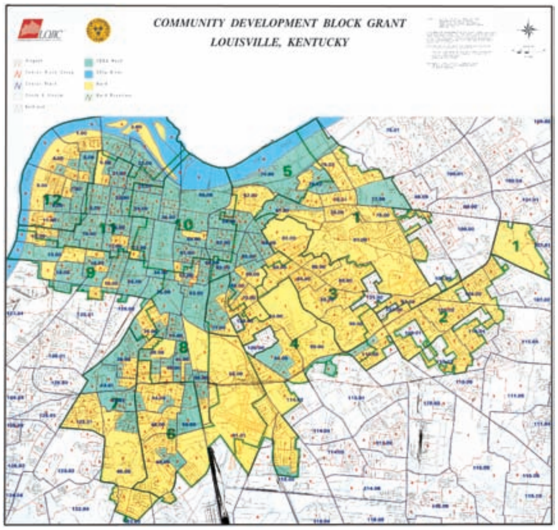
Businesses mapped by number of employees

MAP GALLERY

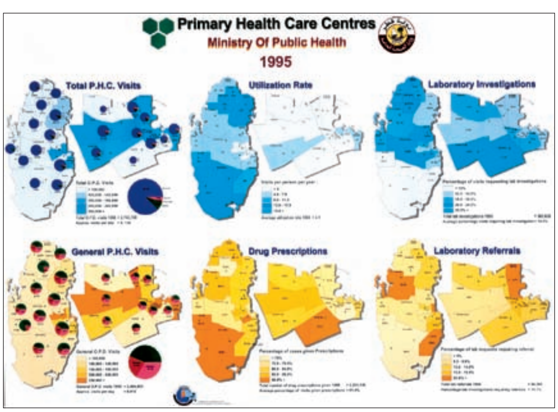
Police in Stockton, California, mapped locations where truant students were picked up. The size of the red dot indicates the number of children picked up at that location. The map lets school officials and police see where the highest truancy is, so they can focus on those areas.



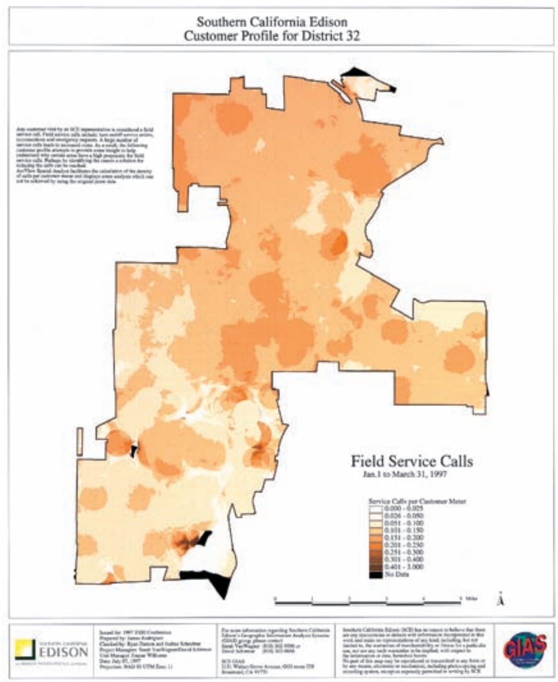
Analysts at the Department of Public Works in Louisville, Kentucky, created a map showing the location of census block groups qualifying for the Community Development Block Grant program. In the block groups shaded green, at least 51 percent of the people are of low and moderate income. The map also shows the city's ward boundaries, so aldermen can see what portions of their wards qualify for the grant program.



The Qatar Ministry of Public Health mapped public health statistics to compare the quality of health care across the country. Using records from state-run health care centers, they mapped the number of visits per person, the percentage of visits in which prescriptions were given, the percentage in which lab referrals were given, and other statistics. The maps were created by shading the areas covered by the health care centers, based on the statistics. Some differences among the centers can be seen. For instance, the centers in the south tend to prescribe drugs more often (indicated by the dark orange) and refer cases to labs less often (indicated by the light yellow). Officials use the maps to see how and where health care varies.



Analysts at Southern California Edison mapped the number of calls for service for a three-month period to see how different parts of the service area compared. They drew the location of each electric meter color-coded by the number of calls for service, from light red for few calls to dark red for many. The map shows that fewer calls were made in the southern portion of the area. There are several small areas where many calls were made, indicating a power outage or other problem affecting a number of people. The company uses the map to help find out why certain areas have more calls and others fewer.

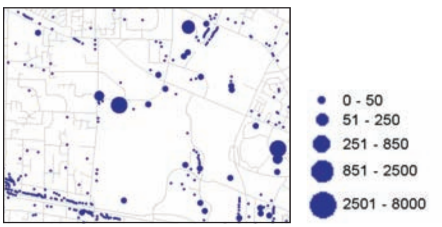


By mapping the patterns of features with similar values, you'll see where the most and least are. Knowing the type of features you're mapping, as well as the purpose of your map, will help you decide how to best present the quantities to see the patterns on your map.

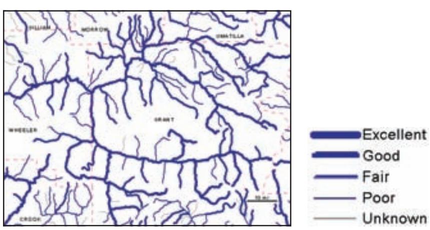
WHAT TYPE OF FEATURES ARE YOU MAPPING?

You can map quantities associated with discrete features, continuous phenomena, or data summarized by area.

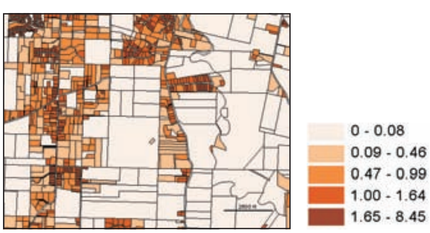
Discrete features can be individual locations, linear features, or areas. Locations and linear features are usually represented with graduated symbols, while areas are often shaded to represent quantities.



Locations—businesses by number of employees

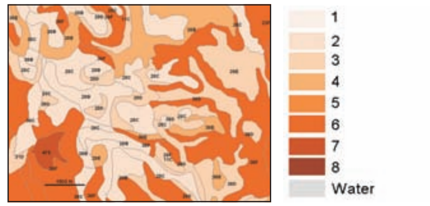


Lines—rivers ranked by fish habitat

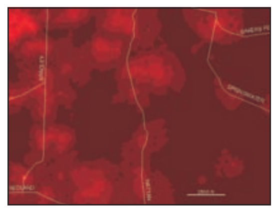


Areas—parcels by land value ($US per square foot)

Continuous phenomena can be defined areas or a surface of continuous values. Areas are displayed using graduated colors; surfaces are displayed using graduated colors, contours, or a 3-D perspective view.

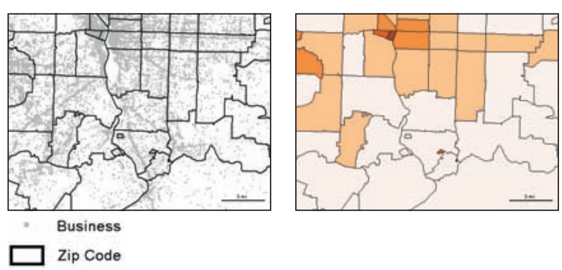


Soils ranked by suitability for growing crops. Soils with a rank of 1 are the most suitable.

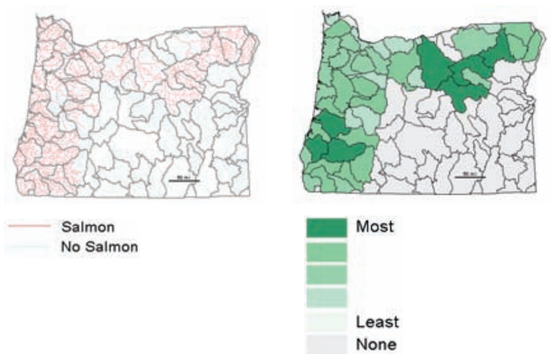


A surface of land value ($US per square foot) created by interpolating from tax parcel centroids. The lighter areas have a higher value.

Data summarized by area is usually displayed by shading each area based on its value, or using charts to show the amount of each category in each area. You can summarize individual locations, linear features, or areas. Chapter 5, 'Finding what's inside,' discusses summarizing by area.



The number of businesses in each ZIP Code is summarized to create a map showing ZIP Codes shaded by total number of businesses per square mile. The darker shades indicate a higher concentration of businesses.



Streams summarized by watershed show which watersheds have the longest salmon runs.

Poverty rate by block group (left) and by census tract

ARE YOU EXPLORING THE DATA OR PRESENTING A MAP?

Keeping the purpose of your map and the intended audience in mind will help you decide how to present the information on your map.

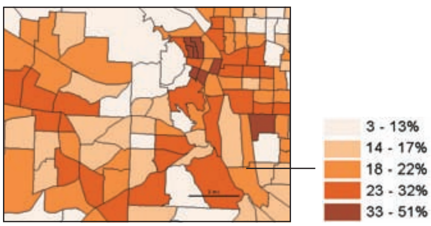
You might be exploring the data to see what patterns and relationships you can find. For example, you might look at the distribution of median income in an area to see the range of values, and where high- and low-income families cluster. When you're exploring, you'll display the data in more detail and try various ways of displaying it.

In another case, you might want to present a map showing specific patterns that answer a particular question. For example, you may want to show which areas have at least 35 percent of families living below poverty level, and which areas have 50 percent or more, so city officials can see areas potentially qualifying for economic development grants. When creating a map for presentation, you'll generalize the data to reveal the patterns.

In many cases, you'll start by exploring the data to see what patterns emerge, and what questions arise, and later create a generalized map to reveal specific patterns.

To map the most and least you assign symbols to features based on an attribute that contains a quantity. Quantities can be counts or amounts, ratios, or ranks. Knowing the type of quantities you're mapping will help you decide the best way to present the data.

Tract	Population	18-29 Years	%18-29
003603	1606	243	15
0074	2765	516	19
003702	2443	407	17
003803	4132	751	18
0076	3176	668	21

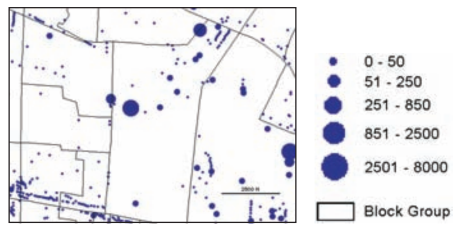


Percentage of 18- to 29-year-olds in each census tract

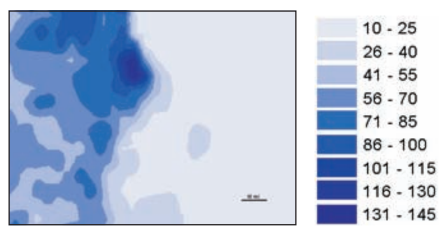
COUNTS AND AMOUNTS

Counts and amounts show you total numbers. A count is the actual number of features on the map. An amount is the total of a value associated with each feature. Using a count or an amount lets you see the value of each feature as well as its magnitude compared to other features.

You can map counts and amounts for discrete features, for example, the number of employees at each business; or for continuous phenomena, for example, the annual precipitation at any location.



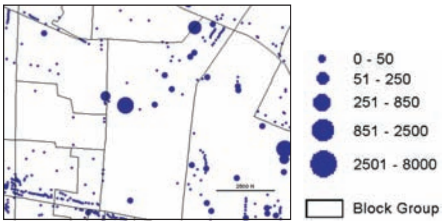
Businesses mapped by number of employees



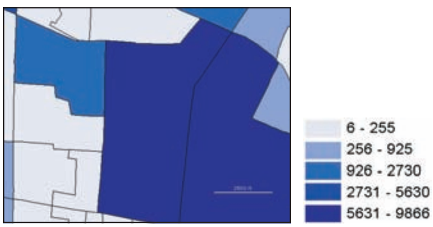
Average annual precipitation (inches)

If you're summarizing by area, using counts or amounts can skew the patterns if the areas vary in size. You should use ratios to accurately represent the distribution of features. However, you can map counts or amounts for data summarized by area if you want to focus on the quantity in each area, rather than see the patterns of where the most and least are.

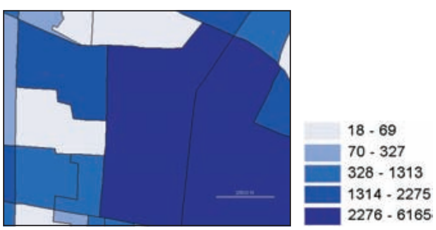
For example, you might map individual businesses by number of employees, but you wouldn't map block groups by number of employees (based on locations of businesses), because block groups vary in size. Larger block groups might have more workers, but more spread out. You'd need to map workers per square mile to see the distribution.



Businesses mapped by number of employees



Mapping the number of employees per block group shows the total in each.



Mapping the number of employees per square mile shows the distribution.

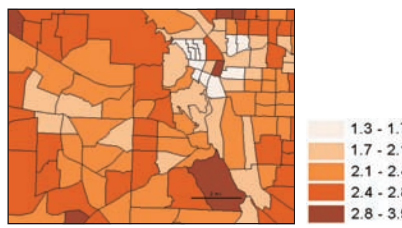
RATIOS

Ratios show you the relationship between two quantities, and are created by dividing one quantity by another, for each feature. Using ratios evens out differences between large and small areas, or areas with many features and those with few, so the map more accurately shows the distribution of features. Because of this, ratios are particularly useful when summarizing by area.

The most common ratios are averages, proportions, and densities.

Averages are good for comparing places that have few features with those that have many. To create an average, you divide quantities that use different measures. For example, dividing the number of people in each tract by the number of households gives you the average number of people per household.

```
People per HH = Population / Households
```

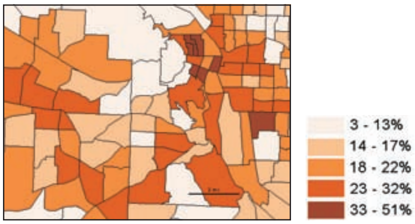


Tract	Population	Households	People per HH
003603	1606	643	2.5
0074	2765	1104	2.5
003702	2443	894	2.7
003803	4132	1591	2.6
0076	3176	1256	2.5

Number of people per household, by census tract

Proportions show you what part of a whole each quantity represents. To calculate a proportion, you divide quantities that use the same measure. For example, dividing the number of 18- to 30-year-olds in each tract by the total population of each tract gives you the proportion of people aged 18 to 30 in each tract.

```
% 18-29 = 18-29 Years / Population
```



Tract	Population	18-29 Years	%18-29
003603	1606	243	15
0074	2765	516	19
003702	2443	407	17
003803	4132	751	18
0076	3176	668	21

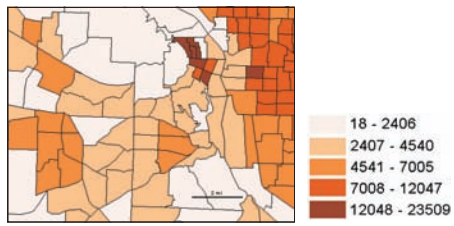
Percentage of population aged 18 to 29, by census tract

Proportions are often presented as percentages (the proportion multiplied by 100) as a kind of shorthand. People are used to thinking and talking in terms of percentages rather than ratios—for example, "22 percent," rather than "twenty-two one-hundredths."

Tract	Population	18-29 Years	18-29 (Ratio)	18-29 (%)
031501	1853	241	0.13	13
003601	4439	799	0.18	18
003902	3072	461	0.15	15
003602	6578	1118	0.17	17

Densities show you where features are concentrated. To calculate density, you divide a value by the area of the feature to get a value per unit of area. For example, by dividing the population of a county by its land area in square miles, you get a value for people per square mile. Density is good for showing distribution when the size of the areas you're summarizing by varies greatly. For example, census tracts have roughly the same number of people, so some are small (where there are many people close together), and some are much larger (where the people are spread out). Density is the subject of chapter 4, 'Mapping density.'

```
Population per Square Mile =
Population / Square Miles
```



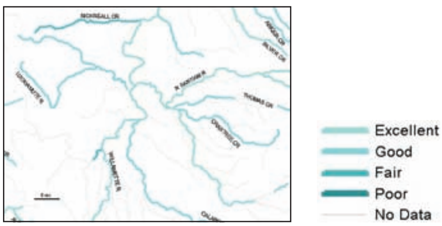
Tract	Population	Square Miles	People per Sq Mi
003603	1606	0.35	4589
0074	2765	0.58	4767
003702	2443	0.37	6603
003803	4132	0.48	8608
0076	3176	0.53	5992

Number of people per square mile, by census tract

You create ratios by adding a new field to the layer's data table, and calculating the new values by dividing the two fields containing the counts or amounts. Some GIS software, such as ArcInfo® and ArcView® GIS, lets you create ratios on the fly by doing the calculation when you create the map—you just tell it which fields you want to divide. Be careful not to create ratios from other ratios, or else you'll create values that don't really mean anything. For example, dividing the percentage of 18- to 30-year-olds in each tract by the area of the tract, to get a percentage per square mile, is meaningless.

RANKS

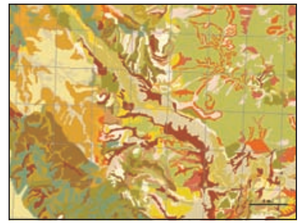
Ranks put features in order, from high to low. They show relative values rather than measured values. Ranks are useful when direct measures are difficult, or if the quantity represents a combination of factors. For example, it's hard to quantify the scenic value of a stream. You may be able to state, however, that the section that passes through a mountain gorge has a higher scenic value than the section passing near a dairy farm.

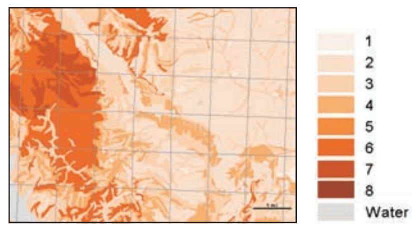


Streams ranked by recreation value

To indicate ranks, you can use text (for example, high, moderate, or low) or numbers (for example, 1 through 10). Since the ranks are relative, you only know where a feature falls in the order—you don't know how much higher or lower a value is than another value. For example, you may know a feature with a rank of "3" is higher than one ranked "2" and lower than a "4," but you don't know how much higher or lower.

You often assign ranks based on another feature attribute, usually a type or category, or a combination of attributes. For example, you'd assign all soils of a certain type the same suitability for growing a particular crop.



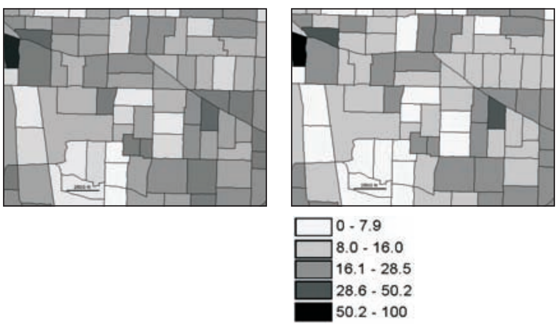


Each soil type (shown on the left) is assigned a rank based on its capability for growing crops. The highest capability is "1," while "8" is lowest.

Once you've determined what type of quantities you have, you need to decide how to represent them on the map; either by assigning each individual value its own symbol, or grouping the values into classes.

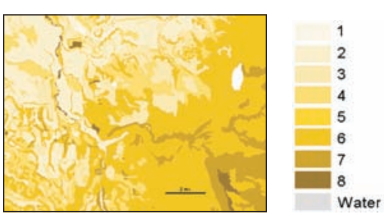
Mapping quantities involves a trade-off between presenting the data values accurately, and generalizing the values to see patterns on the map.

Usually, counts, amounts, and ratios are grouped into classes, since each feature potentially has a different value. This is especially true if the range of values is large. If each value were mapped using a unique symbol, your map would accurately reflect the data, but finding features with similar values would be difficult unless you were mapping just a few values. Using classes is especially valuable when the map will be used for public discussion, as it lets map readers compare areas quickly.



These maps show the poverty rate (percentage of people living below the poverty level) by block group. On the left, each block group is drawn using a unique gray shade based on its data value. Lighter block groups have lower values. Grouping the values into classes (right map) makes the patterns easier to see.

Ranks lend themselves to being mapped as individual values, since the values are not continuous—there is a fixed number of values, and several features are likely to have the same value.



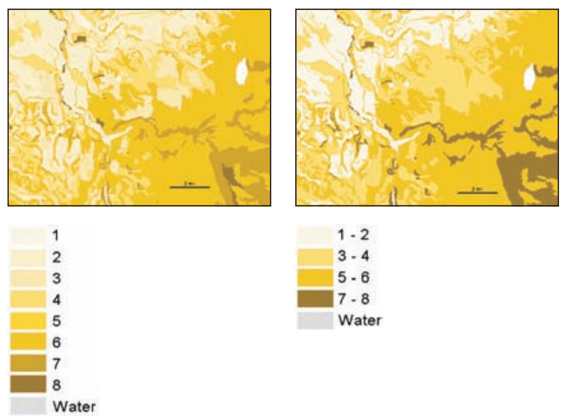
Here, soils are ranked by suitability for growing crops. Each rank is drawn with a different shade.

MAPPING INDIVIDUAL VALUES

By mapping individual values, you present an accurate picture of the data, since you don't group features together. However, this approach may require more effort on the part of map readers to understand the information being mapped, especially if the map contains many values.

Mapping individual values lets you search for patterns in the raw data. You may want to do this if you're unfamiliar with the data or area being mapped, or are looking for subtle patterns in the data. You might also do this to help decide how to group the values into classes.

If you're mapping ranks, assign one symbol to each rank. If you have more than eight or nine ranks, though, you may want to group them into classes, since too many different symbols on a map can make it difficult for map readers to distinguish the different ranks. You can do this by simply assigning the same symbol to adjacent ranks.



Both maps show soils ranked by suitability for growing crops. Combining the original eight ranks into four makes the patterns more distinct.

You can also map ratios, counts, or amounts using individual values if you have no more than 11 or 12 unique values, or fewer than 20 features.

USING CLASSES

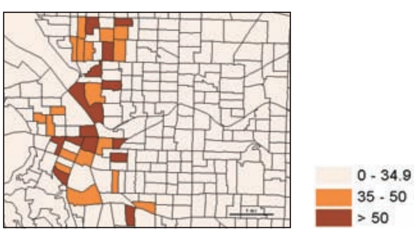
Classes group features with similar values, by assigning them the same symbol. This lets you see features with similar values. How you define the class ranges will determine which features fall into each class, and thus what the map will look like. By changing the classes, you can create very different maps. Generally, make sure features with similar values are in the same class, and make the difference in values between classes as great as possible.

You can create classes manually, or use a standard classification scheme.

Creating classes manually

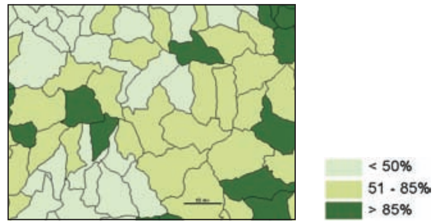
You should create classes manually if you're looking for features that meet specific criteria, or comparing features to a specific, meaningful value. You specify the upper and lower limit for each class and assign the symbols.

The classes might be based on a regulation that specifies a value above or below which some action occurs. If, for example, you want to find out where to designate an Urban Empowerment Zone, you'd map tracts with at least 35 percent of residents living below the poverty level. One of the class breaks for the map would be 35 percent. Another logical break would be 50 percent, to show which tracts have at least half the residents living below the poverty level.



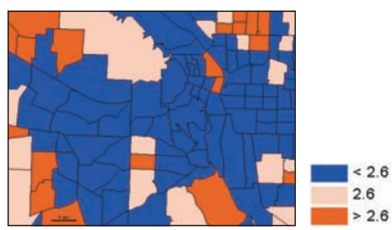
Percentage of residents living below the poverty level, by block group

The classes might also be based on standards or research in a particular discipline or industry. For example, conservation biologists creating a wildlife corridor would exclude watersheds with less than 50 percent forest cover, and try to include watersheds with more than 85 percent forest cover. In this case, the classes would be "Less than 50%", "51 – 85%", and "Greater than 85%".



Percentage of forest cover, by watershed

You might also create classes using a value based on a larger set of features. For example, if you're mapping census tracts for a county by average number of people per household, you might make one of the class breaks the value of the national average for this statistic (2.6 in the United States). Map readers would quickly see how tracts in the county compare to the national average.



Average number of people per household, by census tract. The national average in the United States is 2.6.

In any case, you should always state explicitly on the map what the classes represent.

Using standard classification schemes

Use a standard classification scheme if you want to group similar values to look for patterns in the data. You can choose from a number of schemes for grouping data values into classes; these are based on how the data values are distributed. You specify the classification scheme and the number of classes, and the GIS calculates the upper and lower limit for each class. The four most common schemes are natural breaks, quantile, equal interval, and standard deviation.

You'll figure out the best scheme for creating the class breaks by looking at the distribution of data values. Then, you can decide on the number of classes.

A good way of seeing how data values are distributed is to plot them on a chart. In this example, the charts and maps all use the same data: median household income, in U.S. dollars, by census block group. The chart on the facing page shows the distribution of data values for the area shown on the map. Median income is plotted along the horizontal axis, and the number of block groups with each value is shown on the vertical axis. The height of the bar indicates the number of block groups with that value. The shaded areas (corresponding to the shaded classes on the map) show the range for each scheme; the values of the class breaks are indicated on the horizontal axis. The width of each area shows how many block groups fall into each class.

Natural breaks (Jenk's)

Classes are based on natural groupings of data values. On the chart, class breaks are set where there is a jump in values, indicated by a large step between bars, so block groups having similar values are placed in the same class. The resulting map emphasizes the differences between the highest income block groups, in the lower left, and the next highest, in the center.

Quantile

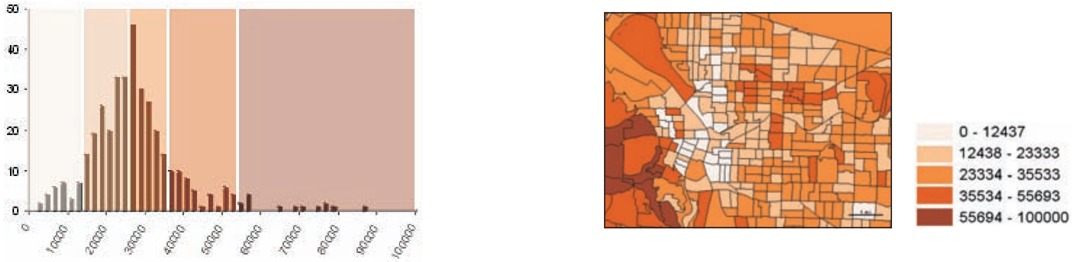
Each class contains an equal number of features. On the chart, the shaded areas show which block groups are in the same class, and indicate the class breaks where they cross the horizontal axis. On this map, block groups with similar values are forced into adjacent classes, and the block groups at the high end (with values ranging from $32,000 to $100,000) are lumped into one class.

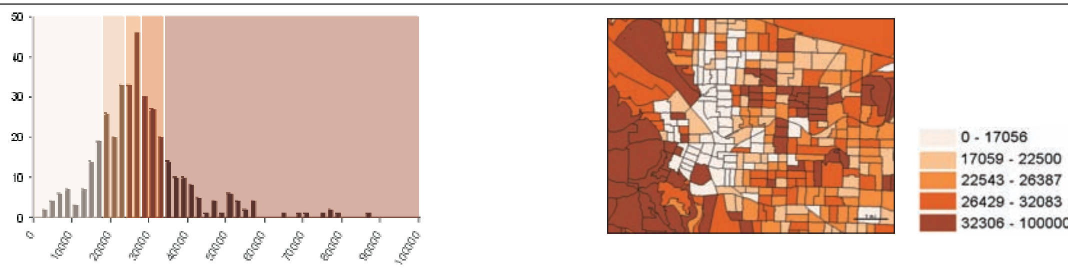
Equal Interval

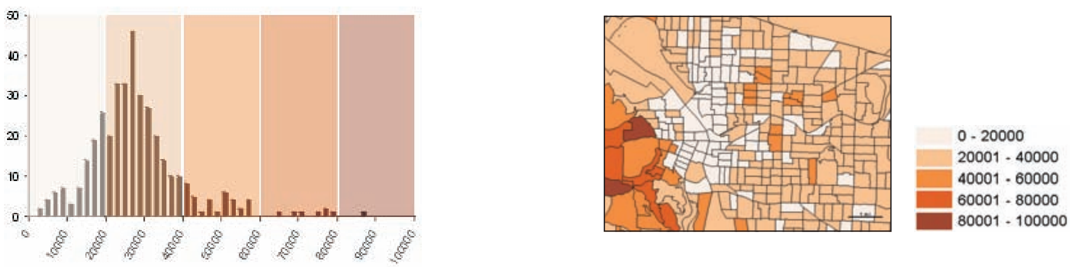
The difference between the high and low values is the same for every class—in this case, $20,000. On this map, almost all the block groups are contained in the two lowest classes. The map emphasizes the location of the few block groups with the very highest median income.

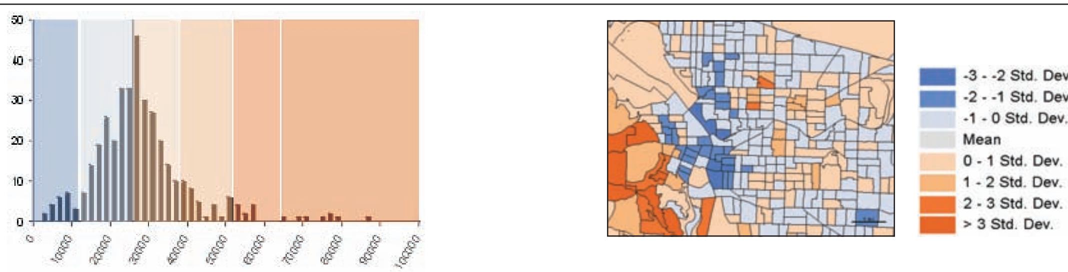
Standard Deviation

Features are placed in classes based on how much their values vary from the mean. The GIS calculates the mean (in this case, about $26,000) and the standard deviation (about $12,900). It successively adds or subtracts the standard deviation to or from the mean to set the class breaks. The map shows how many standard deviations each block group is from.


0 - 12437
12438 - 23333
23334 - 35533
35534 - 55693
55694 - 100000


0 - 17056
17059 - 22500
22543 - 26387
26429 - 32083
32306 - 100000


0 - 20000
20001 - 40000
40001 - 60000
60001 - 80000
80001 - 100000


-3 - -2 Std. Dev.
-2 - -1 Std. Dev.
-1 - 0 Std. Dev.
Mean
0 - 1 Std. Dev.
1 - 2 Std. Dev.
2 - 3 Std. Dev.
> 3 Std. Dev.

Comparing classification schemes

Natural breaks

Natural breaks finds groupings and patterns inherent in your data, so values within a class are likely to be similar, and values between classes, different. Data values that cluster are placed into a single class. Class breaks are defined where there is a gap between clusters of values.

How it works

The GIS automatically determines the high and low value for each class, using a mathematical procedure to test different class breaks. It picks the class breaks that best group similar values and maximizes the differences between classes.

What it's good for

Mapping data values that are not evenly distributed, since it places clustered values in the same class.

Disadvantages

- Since the class ranges are specific to the individual dataset, it's difficult to compare the map to other maps.
- Choosing the optimum number of classes is difficult, especially if the data is evenly distributed.

Quantile

Each class has an equal number of features in it.

How it works

The GIS orders the features, based on the attribute value—from low to high—and sums the number of features as it goes. It divides the total by the number of classes you've specified to get the number of features in each class. It then assigns the first features in the order to the lowest class until that class is filled, then moves on to the next class, fills it up, and so on.

What it's good for

- Comparing areas that are roughly the same size.
- Mapping data in which the values are evenly distributed.
- Emphasizing the relative position of a feature among other features. For example, you can show which counties in a state are in the top 20 percent, in terms of median income (those in the highest of five categories).

Disadvantages

- Features with close values may end up in different classes, especially if values cluster. This may exaggerate the differences between features. Conversely, a few widely ranging adjacent values may end up in the same class, minimizing the differences between these features.
- If the areas vary greatly in size, a quantile classification can skew the patterns on the map.

Equal interval

Each class has an equal range of values—that is, the difference between the high and low value is the same for each class.

How it works

The GIS subtracts the lowest value in the data set from the highest. It then divides that number by the number of classes you specified. It adds that number to the lowest data value to get the maximum value for the first class. It then adds to each maximum value to set the breaks for the rest of the classes.

What it's good for

- Presenting information to a nontechnical audience. Equal intervals are easier to interpret since the range for each class is equal. This is especially true if the data values are familiar to the reader, such as percentages.
- Mapping continuous data, such as precipitation and temperature.

Disadvantages

- If the data values are clustered rather than evenly distributed, there may be many features in one or two classes and some classes with no features.

Standard deviation

Each class is defined by its distance from the mean value of all the features.

How it works

The GIS first finds the mean value by adding all the data values and dividing by the number of features. It then calculates the standard deviation by subtracting the mean from each value and squaring it (to make sure it's positive), summing these numbers, and then dividing by the number of features. It then takes the square root to get the final number. The formula looks like this:

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

where s is the standard deviation, x is the value of a feature, $\bar{x}$ is the mean, and n is the number of features.

You can think of this as the average amount the data values vary from the mean. The GIS creates class breaks above and below the mean based on the number of standard deviations you specify, such as 1/2 or 1 standard deviation.

What it's good for

- Seeing which features are above or below an average value.
- Displaying data that has many values around the mean, and few further from the mean (a bell curve, or normal, distribution).

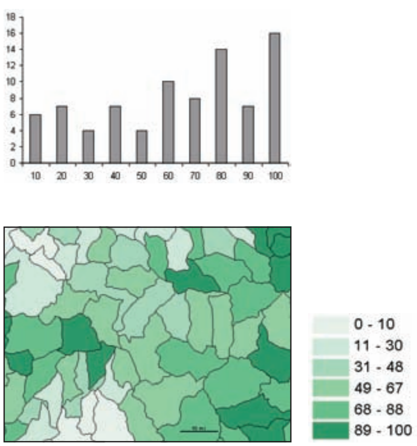
Disadvantages

- The map doesn't show the actual values of the features, only how far their value is from the mean.
- Very high or low values (outliers) can skew the mean so that most features will fall in the same class.

Choosing a classification scheme

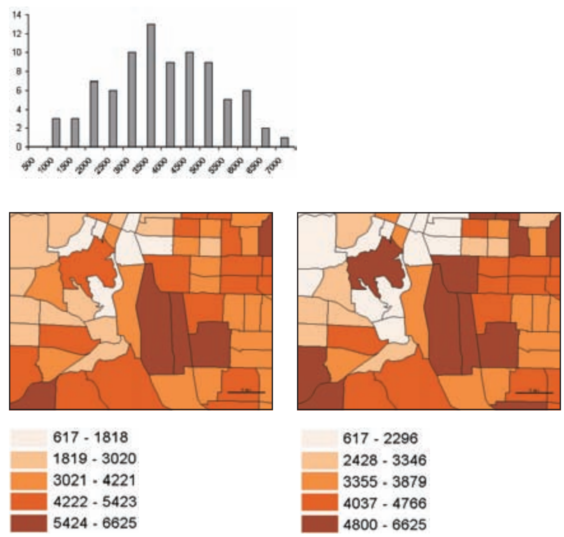
To decide which scheme to use, you need to know how the data values are distributed across their range. Create a bar chart and set the horizontal axis to be the attribute values. The vertical axis should represent the number of features having a particular value. Most spreadsheets can create charts, as can statistical programs and GIS programs such as ArcInfo and ArcView GIS. Follow these guidelines when choosing a classification scheme.

- If your data is unevenly distributed (many features have the same or similar values, and there are gaps between groups of values), use natural breaks.



Watersheds by percentage forested. The chart shows the gaps between groups of values (the shorter bars).

- If your data is evenly distributed and you want to emphasize the difference between features, use equal interval or standard deviation.
- If your data is evenly distributed and you want to emphasize the relative difference between features, use quantile.

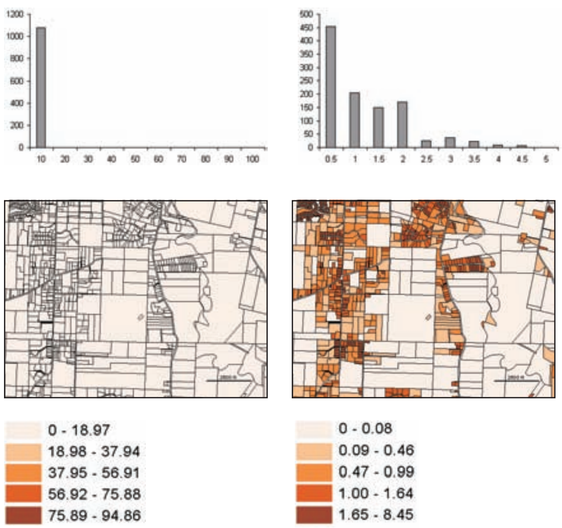


Population by census tract. The chart shows that the data values are fairly evenly distributed, with no large gaps between values. In this case, although the equal interval and quantile schemes produce similar results, equal interval places fewer features in the highest and lowest classes, emphasizing these extremes.

The GIS lets you change the class ranges, the number of classes, and the symbols you use to represent the features fairly quickly, so you can try several approaches to see which one communicates the information best. This is especially useful if you're exploring the data and searching for patterns.

Dealing with outliers

When you graph the data, you may find that you have a few extremely high or low values. These outliers can skew your class ranges, and hence the patterns on the map. This is particularly true when using the equal interval or standard deviation schemes, since all features, except the outlier, may be forced into a single class. Using natural breaks can isolate outliers in the highest or lowest class, but can still compress values into the remaining classes.



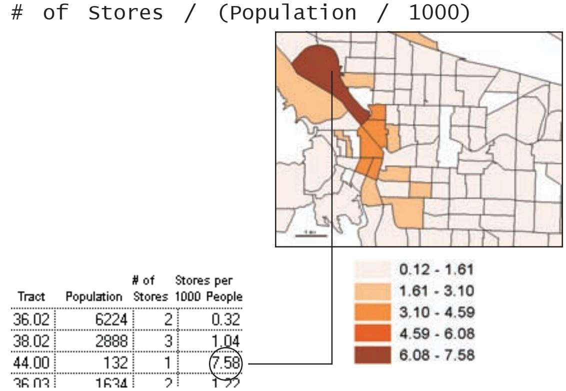
Land value per square foot, in U.S. dollars. In this example, an error in the database resulted in one extremely high data value, $94.86 per square foot (left chart and map). Using an equal interval classification, all features are in the lowest class, the outlier is in the highest, and the remaining classes have no features. With the error corrected, the bar chart and map show a more even distribution of values.

You should look at outliers closely. They may be the result of an error in the database, or they may be anomalies based on a small data sample. Or, they may be completely valid. If they're not outright errors that can be corrected in the database, you can handle them in several different ways, depending on how much they vary from the rest of the values, and how they affect the patterns on the map:

- Put each outlier in its own class. You might do this if the outliers are widely spread out.
- Group them together into a class. You might do this if the outliers cluster.
- Group them with the next closest class up or down, if they're not too far from other values in the class.
- Draw them using a special symbol, if you believe they're not really valid and should not be considered part of the pattern. For example, you might shade them in gray and label them as "Insufficient Data" in the legend.

Outliers can occur as the result of a ratio calculation. While each of the original values might be valid on their own, when divided, the resulting value might be misleading. This could be caused by the way features are stored in the database, or the relationship between the values you're mapping. Suppose, for example, you're mapping the number of grocery stores in each census tract, per 1,000 people. If there's a store in a tract where few people live (perhaps an area with many office buildings, but few houses), calculating stores per 1,000 people for this tract will result in a very high value, skewing the classes and obscuring the patterns in the map.

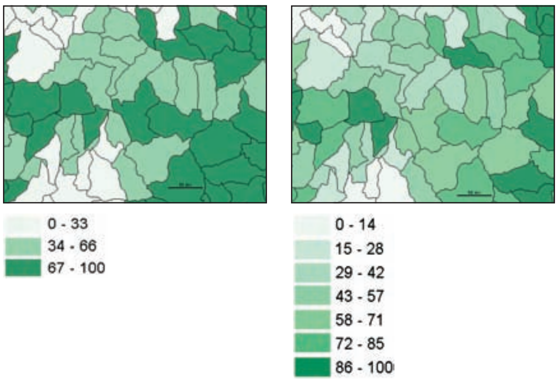
```
Stores per 1000 people =
# of Stores / (Population / 1000)
```



Number of grocery stores per 1,000 people, by census tract. The small population in tract 44.00 creates an outlier when the number of stores per 1,000 people is calculated.

Deciding how many classes

Once you've decided on an appropriate classification scheme, you need to decide how many classes to create. Based on this number, and the classification scheme, the GIS calculates the class ranges and breaks. If you've chosen an appropriate classification scheme, changing the number of classes shouldn't change the appearance of the data very much—only make the patterns more or less distinct.



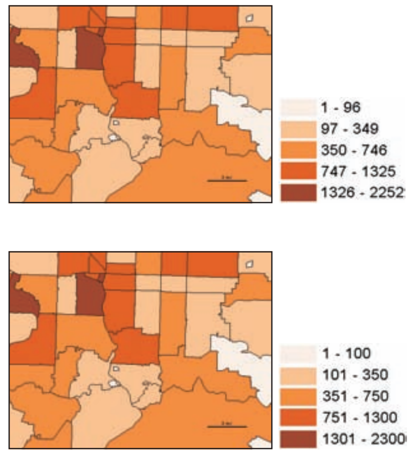
Watersheds by percentage of forest cover. Using four classes makes the patterns distinct; using more classes reveals subtleties in the patterns.

- Most map readers can distinguish up to seven colors on a map, so using more than seven classes will make it hard to find features with similar values. Four or five classes will usually reveal patterns in the data, without confusing the reader. Using fewer than three or four classes doesn't show much variation between features, and hence shows no clear patterns.
- If you're exploring the data to see what kinds of feature groupings and patterns emerge, you may want to use more classes at first. Each feature will be in a narrower range, with values closer to its actual value.

Making the classes easier to read

Once the GIS has defined the ranges, you may want to adjust them to make the classes easier to interpret quickly.

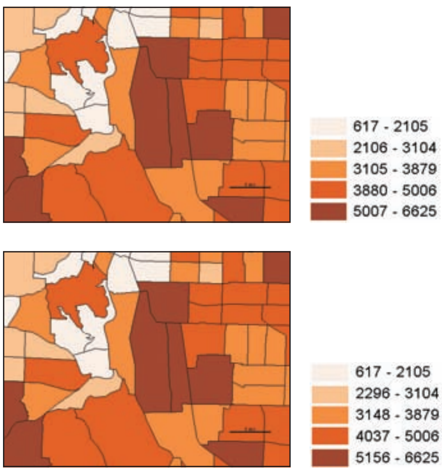
If you don't have to show the exact data values, rounding the minimum and maximum values for each class can make the legend easier to read, without changing the patterns in the map.



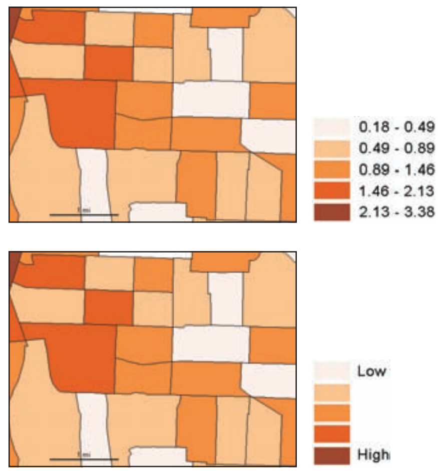
Number of businesses per ZIP Code. Rounding the values makes the numbers easier to read.

Some GIS software creates continuous class ranges by default, making the maximum value in each class the minimum of the next higher class. In fact, the lowest actual data value for the next higher class may be well above the low value shown in the legend for that class. Have the GIS define the classes. Then, change the lowest value for each class to match the lowest value of its features. While the patterns on the map won't change, the legend will better reflect the actual value ranges. This is especially true if you're using a natural breaks classification. However, you should avoid doing this if you're using an equal interval classification, since these ranges are continuous by definition.

You can rename classes in the legend to emulate ranked values, such as "very high," "high," "moderate," "low," or "none." This can make your map easier to understand quickly. You may want to do this if the relative values are more important than the actual values. This is often the case with ratios or large numbers. For example, when calculating the number of grocery stores per 1,000 people for each census tract, you may end up with decimal values that are meaningless on their own—they are only important relative to the other values. To make the map easier for readers to understand, you would change the numeric values to labels such as "high," "medium," and "low."



Population per census tract. Using noncontinuous ranges presents a more detailed picture of the data.



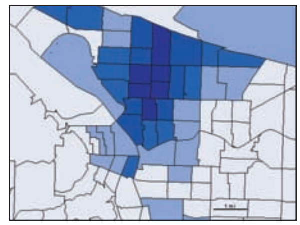
Once you've decided how to classify the data values, you'll want to create a map that presents the information to map readers as clearly as possible. Keep the map simple and present only the information necessary to show patterns in the data. Because GIS makes it easy to create maps and the database often has so much available information, there is a temptation to present more information on the map than the reader can readily comprehend.

The GIS gives you a number of options for creating maps to show quantities:

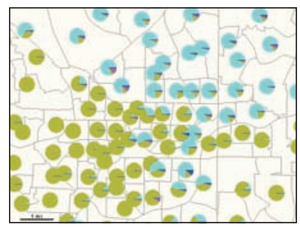
- Graduated symbols
- Graduated colors
- Charts
- Contours
- 3-D perspective views

Graduated symbols

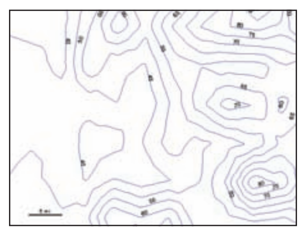
Graduated colors



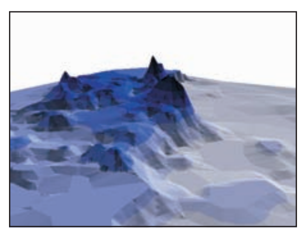
Charts



Contours



3-D perspective views



Which Features	Which Values	Advantages	Disadvantages
Locations Lines Areas	Counts/Amounts Ratios Ranks	Intuitive—people associate symbol size with magnitude	May be difficult to read if many features on map
Areas Continuous phenomena	Counts/Amounts Ratios Ranks	Makes it easy to read patterns and feature values	Colors not intuitively associated with magnitude
Locations Areas	Counts/Amounts Ratios	Shows categories as well as quantities	May present too much information, obscuring patterns
Continuous phenomena	Amounts Ratios	Easy to see rate of change across an area	May make it hard to read patterns and individual feature values
Continuous phenomena Locations Areas	Counts/Amounts Ratios	High visual impact	May make it hard to read values of individual features

CHOOSING A MAP TYPE

The option you choose depends on the type of features and data values you're mapping.

If you have discrete locations or lines, use

- Graduated symbols to show value ranges
- Charts to show both categories and quantities
- A 3-D view to show relative magnitude

If you have discrete areas, or data summarized by area, use

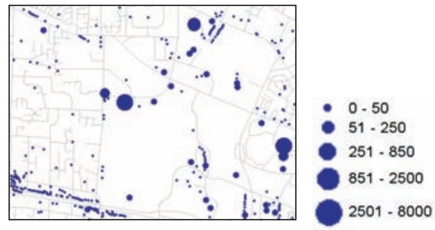
- Graduated colors to show value ranges
- Charts to show both categories and quantities
- A 3-D view to show relative magnitude

If you have spatially continuous phenomena, use

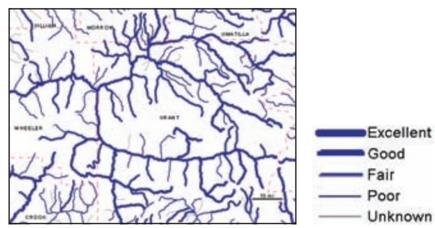
- Graduated colors to show value ranges
- Contours to show the rate of change
- A 3-D view to show relative magnitude

USING GRADUATED SYMBOLS

Use graduated symbols to map discrete locations or lines. Graduated point symbols are drawn at the locations of individual features to show the magnitude of the data value. Graduated line symbols are used to show the volumes or ranks for linear networks, such as roads, utility lines, or rivers.

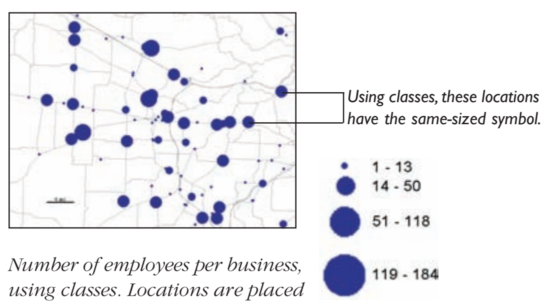


Locations—number of employees per business



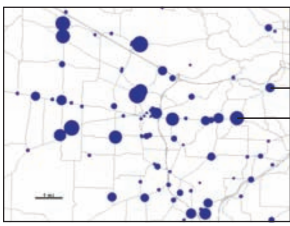
Lines—rivers ranked by fish habitat

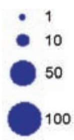
If you're using graduated symbols with classes, you specify the minimum and maximum sizes for the symbol, and the number of classes; the GIS figures out the size of the intermediate symbols.



Number of employees per business, using classes. Locations are placed in a class based on their values. All features in each class are drawn using the same-sized symbol.

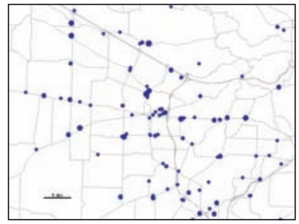
If you're using graduated symbols with individual values, you specify one data value and a corresponding symbol size, and the GIS scales all other symbols accordingly. The legend displays a subset of values and symbols to indicate the relative value of individual features. For example, when mapping the number of employees at each business, the legend displays four circles representing businesses with 1, 10, 50, and 100 employees. By comparing these with the circles on the map, you can tell approximately how many employees a particular business has.



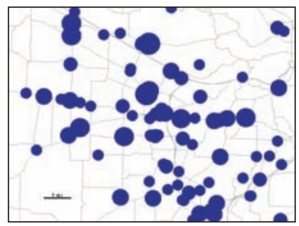


Number of employees per business, using individual values. Each circle is sized according to the value of its location.

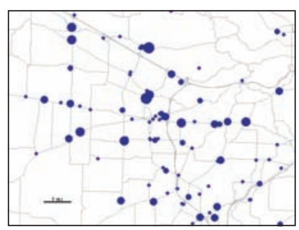
- Most commonly, circles are used for individual locations, although any symbol can be specified. Most readers can interpret relative magnitude more easily from circles than from any other symbol.
- All the symbols should be the same color. Make sure they're dark enough to be seen easily. If you're drawing them inside areas, use a light color to shade the areas.
- Make the difference between the largest and smallest symbols great enough to show the difference in data values.



Symbols are too close to the same size to show patterns.



Symbols are different enough to show patterns, but obscure individual locations.



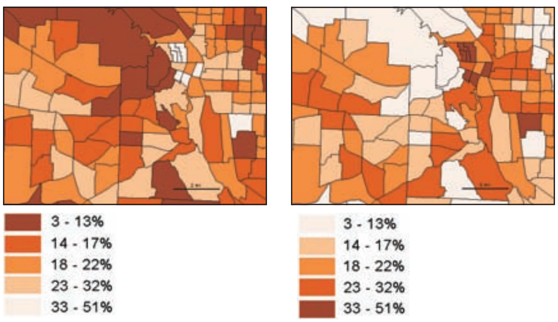
Symbols show patterns without obscuring feature locations.

- Make sure that the smallest symbols will be visible at the scale you're using, and that the largest symbols don't overlap so much that they obscure patterns on the map.

USING GRADUATED COLORS

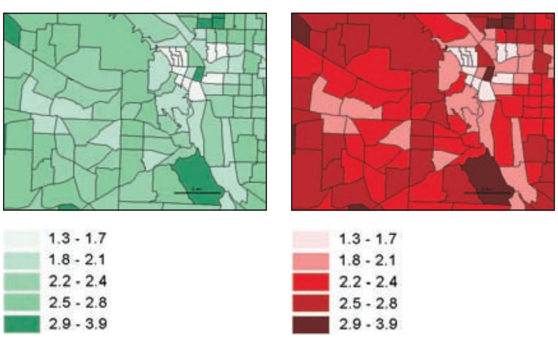
Use graduated colors to map discrete areas, data summarized by area, or continuous phenomena. Usually, you assign shades of one or two colors to the classes. You pick the colors representing the lowest and highest classes, then pick the intermediate colors or let the GIS pick them.

If you have fewer than five or six classes, use one color and vary the shade. Remember that most people can only distinguish up to seven colors. Most people also interpret darker colors to mean "more" or "greater," so assign the darkest shade to the highest class.



Percentage of population aged 18 to 29, by census tract. People generally associate darker colors with higher values, so the map on the left may initially be misleading.

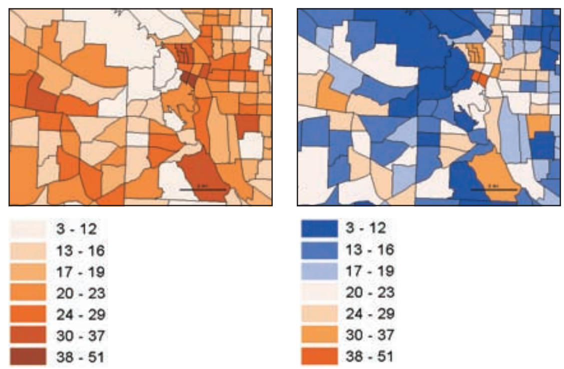
Different colors have different visual impacts. Reds and oranges attract the most attention; blue-green, the least.



Number of people per household, by census tract. The contrasts between the reds are greater, highlighting the high and low values.

It's easier to distinguish between shades of purples and blues than shades of other colors, so you might use these if you have more than four or five classes. Keep in mind that certain colors have special meanings for some people. For example, red is often used to indicate hot spots, such as areas with many crimes, or areas unsuitable for a particular use, such as those too steep to build on.

If you have more than seven or eight classes, you may want to use a combination of colors and shades, using two or even three colors (blue to orange, or blue to green to yellow) to help distinguish the classes. Warm colors (red, orange, or yellow) are a good choice for the classes representing higher values since they highlight these values; cool colors (green, blue, or purple) can be used for lower values.



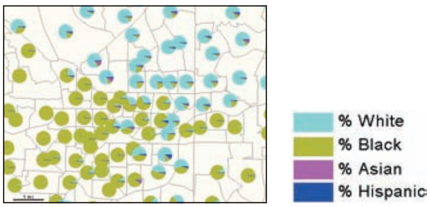
Percentage of population aged 18 to 29, by census tract. Using shades of two or more hues helps distinguish the classes.

Using two colors is also good for showing data with both positive and negative values, such as percentage above or below an average value: shades of one color (such as red) would show percentages above the average, while shades of another color (such as blue) would show percentages below, with a neutral color showing the average. This approach is particularly good for maps using classes based on standard deviation.

A narrow range of values spanning several adjacent classes may end up represented as quite different shades on the map, especially if your data is not evenly distributed. This is one reason it's important to try to make classes, and their symbols, reflect the actual distribution of data values.

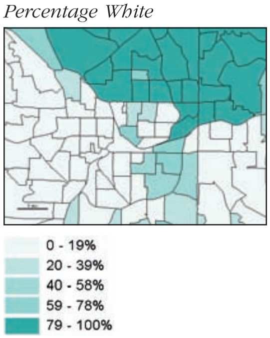
USING CHARTS

Use charts to map data summarized by area, or discrete locations or areas. With charts, you can show patterns of quantities and categories at the same time (see chapter 2, 'Mapping where things are'). That lets you show more information on one map, rather than showing each category on its own map. For example, if you're mapping population by county, you can use a pie chart to show the percentage of the population by ethnic group for each county.

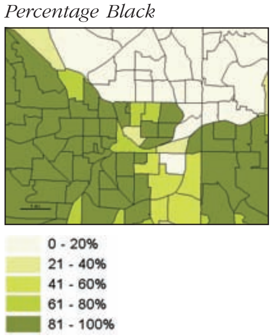


Ethnic population by census tract. The charts show a clear trend from upper right to lower left.

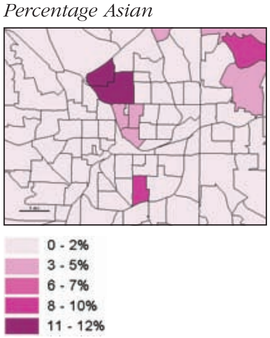
Percentage White



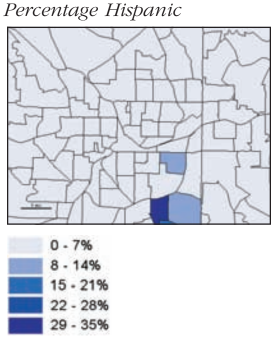
Percentage Black



Percentage Asian



Percentage Hispanic

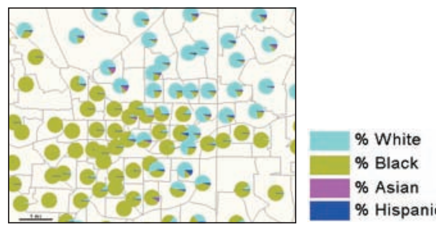


These four maps using graduated colors show more detailed patterns than the map using pie charts.

Charts are useful for a quick study of the patterns. However, some patterns may not be as readily apparent as when presented on separate maps. Maps with charts also require more effort to interpret, since you're looking at both quantities and categories together.

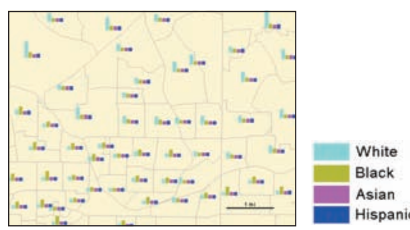
You can create pie charts or bar charts.

- Use pie charts if you want to show how much of the total amount each category takes up. You specify the categories and the attribute to use as the "total" value. The GIS calculates the percentage for each category and shades the chart accordingly.



Pie charts are good for showing percentages.

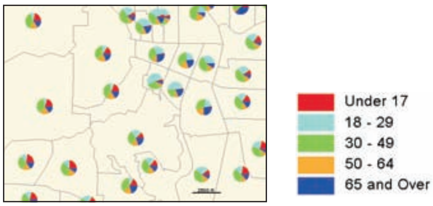
- Use bar charts to show relative amounts, rather than a proportion of a total. You specify a minimum and maximum bar height, and each category is drawn according to its value.



Bar charts are good for showing amounts.

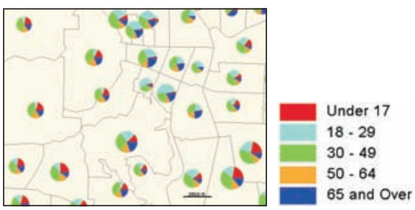
You can make pie charts the same size, or vary the size according to the total amount of the attribute. For example, you could draw the charts larger or smaller, based on the total population of each census tract.

Keep the charts the same size if you want to focus on the amount of each category relative to the total.



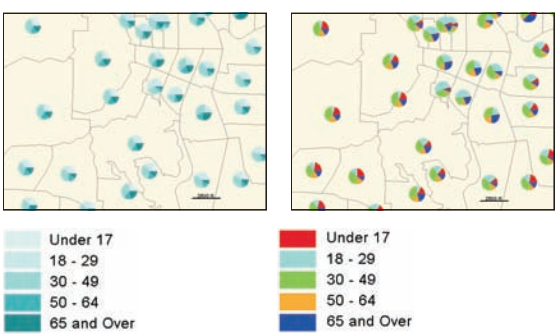
Population of census tracts by age category

Use graduated charts to show the relative size of each feature.



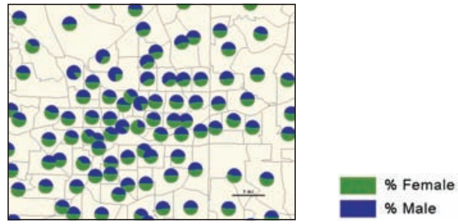
The size of each circle indicates the relative population of each tract.

Since the charts represent categories, not relative amounts, draw the bars or wedges using different colors, rather than shades of one color.



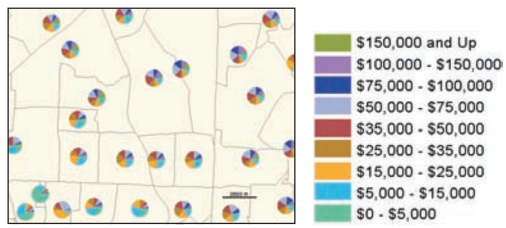
It's easier to distinguish the categories when using different colors, rather than shades of one color.

Charts are most effective when mapping no more than 30 features. Otherwise, the patterns on the map will be difficult to see.



Charts aren't effective for mapping many features; the patterns are hard to see.

Don't use more than five categories on a chart; if you want to show more categories, a series of shaded maps showing each category will work better.



Median income, by census tract. Using too many categories makes the map hard to interpret.

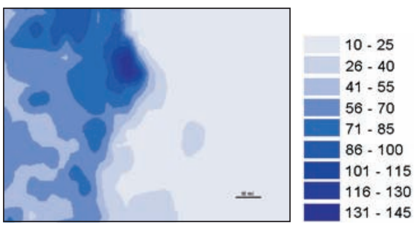
Make sure the charts are large enough to read at the scale you're using. If you're using graduated charts, make sure the smallest one is big enough to be readable, and the biggest ones don't obscure the area boundaries or overlap each other.

USING CONTOUR LINES

Use contour lines to show the rate of change in values across an area for spatially continuous phenomena. Where the lines are closer together, the change is more rapid. Elevation and barometric pressure are commonly mapped using contours.

Contour lines are drawn at an interval that you specify. For instance, a contour map of precipitation with a contour interval of 10 inches would have contour lines at 10, 20, 30, and so on. Each point on a line has the same value, while a point between the lines has a value between the values of the two lines on either side of it.

The interval determines the number of lines and the distance between them. When you use a smaller interval, you create a map with more lines.

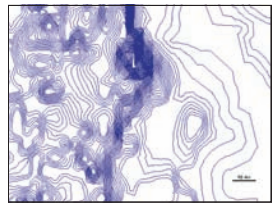


Average annual precipitation, in inches, as a surface

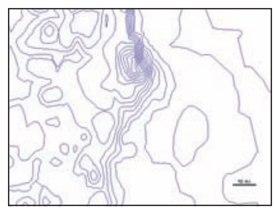
Average annual precipitation, in inches, as contour lines

- Choose an interval for the contour lines that is small enough to give some definition to the surface but not so small that the lines become too close together and the map difficult to read.

The contour interval is too large to show definition in the surface.

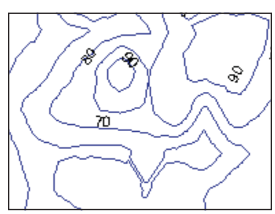


The contour interval is too small to read lines clearly.



The contour interval shows definition in the surface, while making individual lines easy to see.

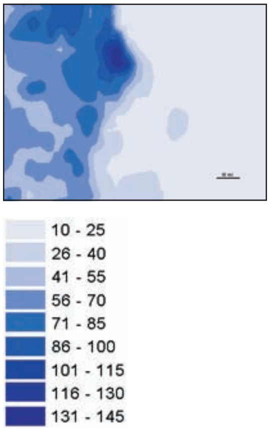
- Contour lines should be labeled with their value to make it easier to see the actual values, as well as the patterns.

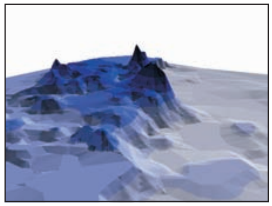


- Using a bold line for every fifth interval makes the values easier to read. For example, if the interval is 10 inches, you'd use bold lines for values of 50, 100, and 150 inches.

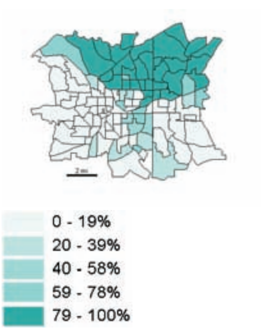
CREATING 3-D PERSPECTIVE VIEWS

Three-dimensional (3-D) perspective views are most often used with continuous phenomena to help people visualize the surface. You can also create 3-D views for areas or points—the height of the feature indicates the magnitude of the location or area.





Average annual precipitation, as a surface (left), and as a 3-D perspective view (right)



Percentage white population by census tract, using graduated colors (left) and displayed as a perspective view

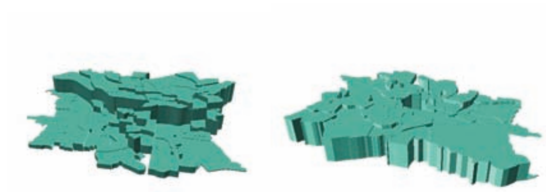
Although you can see where values are higher or lower when using 3-D views, it's difficult to determine the value at any particular location.

To create a 3-D view you specify three parameters that determine what your view will look like: the viewer's location, the vertical exaggeration (z-factor), and the location of the light source.

Viewer location

The viewer's location determines which features are visible in the view, since taller features may block features behind them. You set the location by rotating the view until you have the viewing position you want. Or, you specify the coordinates of the viewer and a target location on the view (usually the center of the features you're displaying), along with the viewer's angle above the surface.

Since you might rotate the view off a north–south axis to highlight patterns in the data, it's important to orient the viewer by providing a north arrow, or drawing recognizable features on the view (such as boundaries or major roads).



Rotated 180 degrees (right view), the 3-D perspective of census tracts doesn't show the steep drop-off in white population.

z-factor

When creating the view, you can specify a value, called a "z-factor," to increase the variation in the surface so that the differences are easier to see. This value is multiplied by each feature's data value. For example, using a factor of 2, a tract that has 40 percent white population will have a value of 80, and a tract that is 10 percent white will have a value of 20. The original difference between the values (30) is now 60, so the difference in heights in the view will also be greater. The goal is to use a z-factor large enough to show variation in the surface but not overly exaggerate the differences between the values.

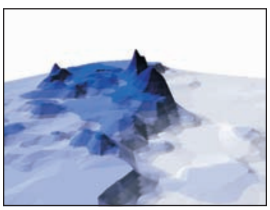
If the vertical exaggeration is too small (right), the view doesn't show enough variation.

Light source

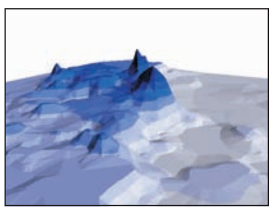
The location of the light source, when combined with the z-factor, determines how shadows will appear on the surface, and thus how distinct the features on the surface will be. You specify two values for the light source: the direction and the angle.

The direction the light is coming from is usually specified in degrees (from 0 to 359, with 0 being north). Unless the location of the shadows is important (for example, you're mapping terrain and want to see which areas are in shadow at a particular time), you can set the direction to highlight the features in the view. This may take some experimenting.

The angle is the height of the light source above the horizon, specified in degrees. The lower the angle, the longer the shadows.



Light source from the northwest

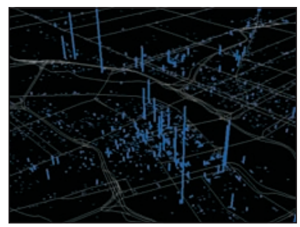


Light source from the southeast

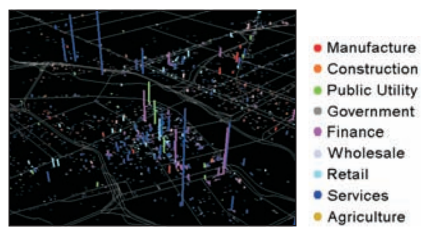
Displaying a perspective view

If you're mapping continuous data, you'll typically draw a perspective view using a range of colors or shades. You can also draw contour lines over the view, and label them to give more definition to the surface and show data values.

You can display individual locations or areas using a single color, or you can use colors based on categories. The bottom view shows not only where the employment centers are (the clusters of taller pillars), but also which types of businesses have the largest numbers of employees.



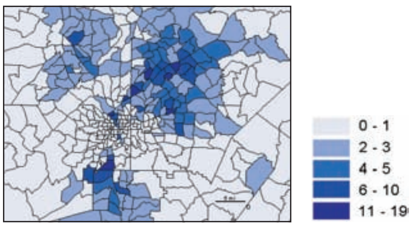
The height of the pillars represents the number of employees at each business.



Color coding the pillars by type of business shows which types have the most workers.

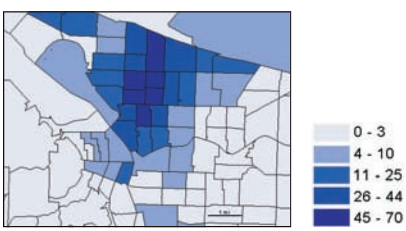
If your map presents the information clearly, you can compare different parts of the map to see where the highest and lowest values are. Looking at the transition between where the least and most are—for example, seeing where change is rapid or gradual—can give you further insight into relationships between places.

You'll want to see whether values cluster or are evenly distributed. In this map, the Asian-American population is clustered in three areas. A store owner selling to this population would focus on these areas for an ad campaign.

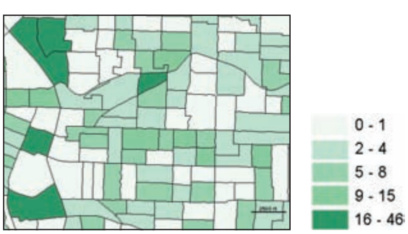


Percentage Asian-American, by census tract

Values may be concentrated in one place, or scattered across an area. In the top map (below), the African-American population is concentrated in five census tracts, and gradually diminishes outward from these. In contrast, the Hispanic population in the bottom map is fairly evenly distributed. A business targeting these customers could site a new store accordingly.

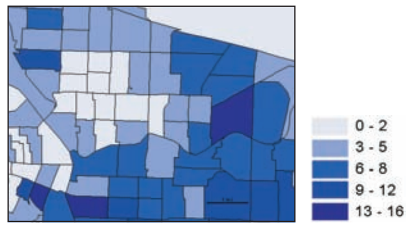


Percentage African-American, by census tract

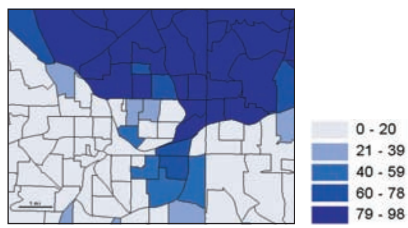


Percentage Hispanic, by block group

The relationships between the locations of the features with high and low values can help you understand how people or phenomena behave. There may be a gradual trend across the area, from low to high, as shown in the top map (percentage Native American), or there may be a definite line dividing areas of high and low values, as appears in the bottom map (percentage White). Sociologists might gain insight from these maps about the different ways ethnic groups mix in these places.

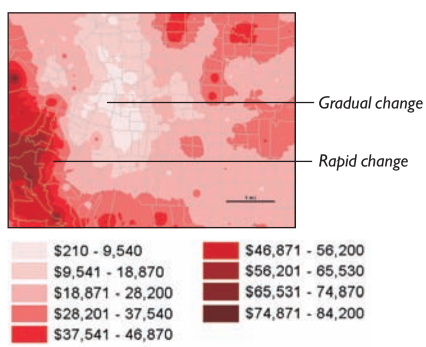


Percentage Native American, by census tract



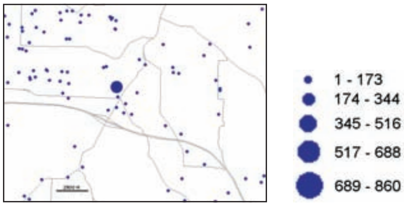
Percentage White, by census tract

The change in values across the area may be abrupt in some areas and gradual in others. This surface map of household income might help a marketing firm delineate areas for targeted advertising.



Median household income

Some features may stand apart from surrounding features. On the map below, most businesses have few employees, with one exception. A transportation planner would want to know where exceptions to the pattern of workers are.



Businesses by number of employees

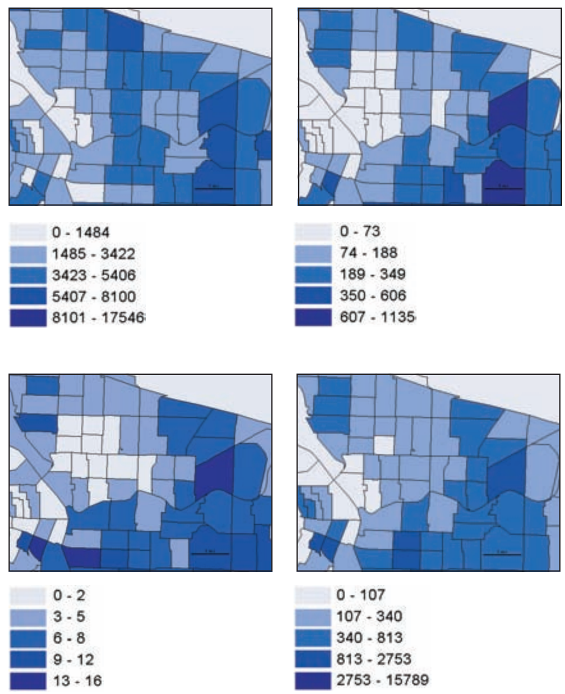
If you're summarizing by area, the areas you use can affect the patterns you see on the map. Using a few large areas may obscure subtle patterns. Conversely, using many small areas may present too much local variation to see the overall patterns.

Often, data can be summarized at several levels. The one you use can affect the patterns you see. In these maps, the high poverty levels shown in the block groups in the upper left (left map) are averaged out when mapped by census tract, as shown in the map on the right.

Poverty rate by block group (left) and by census tract. Note that the block groups with a high poverty rate in the upper left are subsumed when the data is aggregated and mapped by census tract.

Keep in mind that data collected for small areas can be summarized up to larger ones, but the reverse is not true. For example, if you know the number of high school students in each block group, you can sum the students for each block group in a tract to get the total for the tract. But if you start with the number of students in each tract, you can't divide them up to each block group in the tract.

To really understand what's going on in a place, you may want to display several maps showing related information. For example, to understand the distribution of the Native American population in a region, you'd want to create maps showing total population by census tract, and the population, percentage, and density of the Native American population.



The maps of total population and Native American population (upper left and right) show the numbers of people. The map of percentage Native American (lower left) shows the proportion of the total population, while the density map (lower right) shows the distribution of the Native American population in this area.