Introducing GIS analysis

GIS analysis lets you see patterns and relationships in your geographic data. The results of your analysis will give you insight into a place, help you focus your actions, or help you choose the best option.

In this chapter:
- What is GIS analysis?
- Understanding geographic features
- Understanding geographic attributes
Geographic information system (GIS) technology is now about 30 years old. However, for the most part, people are still using it only to make maps. GIS can do much more. Using GIS for analysis, you can find out why things are where they are and how things are related. By learning to use GIS for analysis, you can get more accurate and up-to-date information, and even create new information that was unavailable before. Having this information can help you gain a deeper understanding of a place, make the best choices, or prepare for future events and conditions.

So why aren’t more people using GIS for analysis? One reason is that GIS use has only recently become widespread, so it’s still new to many people. Because of this, many organizations are only now finishing building their GIS databases (a process that took a long time in the past but is becoming faster with the huge amount of geographic data now available). Another reason is that using GIS for analysis has been difficult and cumbersome. Now, new easy-to-use software employing graphic interfaces is removing that obstacle. A third reason more people aren’t using GIS for analysis is that they don’t know what they can do with GIS beyond making maps and creating reports. Or, if they do, they don’t know how to go about it. While geographic data is becoming more widely available and GIS software is becoming easier to use, to do effective GIS analysis you still need to know how to structure your analysis and which tools to use for a particular task.

That’s where this book comes in. You may not be aware of it, but if you make maps today, you are in effect doing analysis already. One of the goals of this book is to help you build better maps—maps that clearly and accurately present the information you need from your data. We also want to introduce you to some of the basic analysis concepts and tasks that—while useful in their own right—are the building blocks for more advanced analysis.

In this book, we’ve identified the most common geographic analysis tasks people do every day in their jobs:

- Mapping where things are
- Mapping the most and least
- Mapping density
- Finding what’s inside
- Finding what’s nearby
- Mapping change

The book is organized in three parts. In this chapter, you’ll learn what GIS analysis is, and what it can do for you. You’ll also review some basic GIS concepts: what geographic data is and how it’s stored, and more about data values, their use, and interpretation. Chapters 2 through 4 present key map-building concepts. They focus on ways of presenting geographic data to best see the patterns of how things are distributed. The later chapters focus on map query and map-based analysis tasks that let you look at geographic relationships.

In the next decade, the use of GIS analysis will grow. A new type of user will emerge—the spatial scientist. A significant number of GIS users will emerge as advanced modelers. Our goal is to help you expand your analytical GIS skills and sophistication. To do that, ESRI plans to add another book to this series covering more advanced analysis concepts and methods.
GIS analysis is a process for looking at geographic patterns in your data and at relationships between features. The actual methods you use can be very simple—sometimes, just by making a map you’re doing analysis—or more complex, involving models that mimic the real world by combining many data layers.

The chapters in this book follow the process you go through when performing an analysis.

**Frame the question**

You start an analysis by figuring out what information you need. This is often in the form of a question. Where were most of the burglaries last month? How much forest in each watershed? Which parcels are within 500 feet of this liquor store? Being as specific as possible about the question you’re trying to answer will help you decide how to approach the analysis, which method to use, and how to present the results.

Other factors that influence the analysis are how it will be used and who will use it. You might simply be exploring the data on your own to get a better understanding of how a place developed or how things behave; or you may need to present results to policy makers or the public for discussion, for scientific review, or in a courtroom setting. In the latter cases, your methods need to be more rigorous, and the results more focused.

**Understand your data**

The type of data and features you’re working with help determine the specific method you use. Conversely, if you need to use a specific method to get the level of information you require, you might need to obtain additional data. You have to know what you’ve got (the type of features and attributes, discussed later in this chapter), and what you need to get or create. Creating new data may simply mean calculating new values in the data table (see ‘Working with tables’ later in the chapter) or obtaining new layers.

**Choose a method**

There are almost always two or three ways of getting the information you need. Often, one method is quicker and gives you more approximate information. Others may require more detailed data and more processing time and effort, but provide more precise results. You decide which method to use based on your original question and how the results of the analysis will be used. For example, if you’re doing a quick study of assaults in a city to look for patterns, you might just map the individual crimes and look at the maps. If the information will be used as evidence in a trial, though, you might want a more precise measure of the locations and numbers of assaults for a given time period.

**Process the data**

Once you’ve selected a method, you perform the necessary steps in the GIS. In this book, we give you some of the concepts behind what the GIS is doing, so you can better interpret the results. We also give you some context for choosing parameters that might be required during the analysis.

**Look at the results**

The results of the analysis can be displayed as a map, values in a table, or a chart—in effect, new information. You need to decide what information to include on your map, and how to group the values to best present the information. You must also decide whether charts would help others easily see the information you’re presenting.

Looking at the results can also help you decide whether the information is valid or useful, or whether you should rerun the analysis using different parameters or even a different method. GIS makes it relatively easy to make these changes and create new output. You can compare the results from different analyses and see which method presents the information most accurately.
As you can see, the type of geographic features you’re working with affect all steps of the analysis process. Spending some time up front looking at your data—and figuring out how it can be analyzed—will make the process run smoothly. Following is a discussion and definitions of the different types of geographic data, how they’re represented in the GIS, and how you’ll be working with them.

You need to be aware of the different types of geographic features, different ways they’re represented, and a little bit about map projections and coordinate systems.

**TYPES OF FEATURES**

Geographic features are either discrete, continuous phenomena, or summarized by area.

**Discrete features**

For discrete locations and lines, the actual location can be pinpointed. At any given spot, the feature is either present or not.

**Continuous phenomena**

Continuous phenomena such as precipitation or temperature can be found or measured anywhere. These phenomena blanket the entire area you’re mapping—there are no gaps. You can determine a value (annual precipitation in inches or average monthly temperature in degrees) at any given location.

Continuous data often starts out as a series of sample points, either regularly spaced (such as sampled elevation data) or irregularly spaced (such as weather stations). The GIS uses these points to assign values to the area between the points, in a process called interpolation. Some noncontinuous data is treated as continuous—for a given place—in order to create maps showing how a quantity varies across the place. For example, you could create a map of land values by interpolating the center points (centroids) of all the parcels in the city.

Continuous data can also be represented by areas enclosed by boundaries, if everything inside the boundary is the same type, such as a type of soil or vegetation. Of course, since the data varies continuously across the landscape, the boundaries really indicate where things are more similar than not—they’re not really definitive as they are with discrete areas such as parcels, where the boundary is legally defined.
Features summarized by area

Summarized data represents the counts or density of individual features within area boundaries. Examples of features summarized by area include the number of businesses in each ZIP Code, the total length of streams in each watershed, or the number of households in each county (obtained by summing the number of households in each census tract). The data value applies to the entire area, but not to any specific location within it.

A lot of data already comes summarized by area, especially demographic data, which consists of totals (total population, total households, and so on) or a percentage based on a category (percent over 65 years old, or percent Hispanic). Some business data is also aggregated to these boundaries or to ZIP Codes, area codes, or other boundaries.

You may have other data you want to summarize by area. If the features are already tagged with a code assigning them to the area, it’s just a matter of doing statistics on the data table—for example, sum the total revenue for all businesses in each ZIP Code. That figure is then assigned to each area boundary when you join the two tables. You can then use that attribute to map the areas and look for patterns.

If the features aren’t tagged with the codes for the areas by which you want to summarize them, the GIS lets you overlay the areas with the features to find out which ones are within each area and tag them with the appropriate code. You’ll read more about this in chapter 5, ‘Finding what’s inside.’
TWO WAYS OF REPRESENTING GEOGRAPHIC FEATURES

Geographic features can be represented in the GIS using two models of the world: vector and raster.

With the vector model, each feature is a row in a table, and feature shapes are defined by x,y locations in space (the GIS connects the dots to draw lines and outlines). Features can be discrete locations or events, lines, or areas. Locations, such as the address of a customer, or the spot a crime was committed, are represented as points having a pair of geographic coordinates.

Lines, such as streams, roads, or pipelines, are represented as a series of coordinate pairs.

Areas are defined by borders, and are represented as closed polygons. They can be legally defined, such as a parcel of land; administrative, such as counties; or naturally occurring boundaries, such as watersheds. When you analyze vector data, much of your analysis involves working with (summarizing) the attributes in the layer’s data table.

With the raster model, features are represented as a matrix of cells in continuous space. Each layer represents one attribute (although others can be attached), and most analysis occurs by combining the layers to create new layers with new cell values.

The cell size you use for a raster layer will affect the results of the analysis and how the map looks. The cell size should be based on the original map scale and the minimum mapping unit. Using too large a cell size will cause some information to be lost. Using a cell size that is too small requires a lot of storage space, and takes longer to process, without adding additional precision to the map.

Soil types represented as a raster layer

Raster layer of vegetation. A larger cell size (right map) shows the patterns, but some detail is lost.
While any feature type can be represented using either model, discrete features and data summarized by area are usually represented using the vector model. Continuous categories are represented as either vector or raster, and continuous numeric values are represented using the raster model.

You might also represent discrete features as raster when you’re combining them with other layers in a model, since raster is particularly good for this kind of analysis.

MAP PROJECTIONS AND COORDINATE SYSTEMS

All the data layers you're using should be in the same map projection and coordinate system. Otherwise, you won’t be able to draw them on top of each other, or combine them to see relationships, such as what features are inside an area, or what features are near another feature.

A map projection translates the locations on the globe (which is almost a sphere) onto the flat surface of your map. All map projections distort the shapes of the features being displayed, as well as measurements of area, distance, and direction. In general, if you’re mapping a relatively small area, such as a town or county, this distortion is negligible. It may be more of a concern if you’re mapping a large area such as a state, country, or the entire world, because the curvature of the earth begins to come into play.

A coordinate system specifies the units used to locate features in two-dimensional space, and the origin point of those units.

If you’re using an established GIS database, chances are the data you’re using is already in the same coordinate system and projection. If you’re collecting data from various sources, though, you’ll want to check this. Several issues are involved in choosing a map projection and coordinate system, including where on the globe the area you’re mapping is located, how large the area is, and whether you need precise measurements of distance or areal extent. Several of the references at the end of the book contain information on how to choose a coordinate system and map projection, and how to project your data.
Each geographic feature has one or more attributes that identify what the feature is, describe it, or represent some magnitude associated with the feature. The type of analysis you do depends partly on the type of attributes you’re working with.

**TYPES OF ATTRIBUTE VALUES**

Attribute values include:

- Categories
- Ranks
- Counts
- Amounts
- Ratios

**Categories**

Categories are groups of similar things. They help you organize and make sense of your data. All features with the same value for a category are alike in some way, and different from features with other values for that category. For example, you can categorize roads by whether they’re freeways, highways, or local roads, and crimes by whether they’re burglaries, thefts, assaults, and so on.

Category values can be represented using numeric codes or text. Text values are often abbreviations, to save space in the table.

**Ranks**

Ranks put features in order, from high to low. Ranks are used 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, though, that the section that passes through a mountain gorge has a higher scenic value than the section near a dairy farm.

Streams ranked by recreation value

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 can assign ranks based on another feature attribute, usually a type or category. For example, you’d assign all soils of a certain type the same suitability for growing a particular crop.
Counts and amounts

Counts and amounts show you total numbers. A count is the actual number of features on the map. An amount can be any measurable quantity associated with a feature, such as the number of employees at a business. Using a count or amount lets you see the actual value of each feature as well as its magnitude compared to other features.

Ratios

Ratios show you the relationship between two quantities, and are created by dividing one quantity by another, for each feature. For example, dividing the number of people in each tract by the number of households gives you the average number of people per household. Using ratios evens out differences between large and small areas, or areas having many features and those having few, so the map more accurately shows the distribution of features.

Two special ratios are proportions and densities. Proportions show you what part of a total each value is. 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. Proportions are often presented as percentages (the proportion multiplied by 100). Densities show the distribution of features or values per unit area. For example, by dividing the population of a county by its land area in square miles, you’d get a value for people per square mile. Density is the subject of chapter 4.

Introducing GIS analysis • 17
Continuous and noncontinuous values

Categories and ranks are not continuous values—there are a set number of values in the data layer, and more than one feature may have the same value. Usually there is at least one feature having any given value. Mapping categories is discussed in chapter 2, while mapping ranks is discussed in chapter 3.

Counts, amounts, and ratios are continuous values—each feature potentially has a unique value anywhere in the range, between the highest and lowest values. That’s important to realize, because knowing how the values are distributed between the highest and lowest values will help you decide how to group them for presentation, so you can see the patterns. Classifying continuous values is discussed in chapter 3.

WORKING WITH DATA TABLES

An important part of GIS analysis is working with the tables that contain the attribute values and summary statistics. Three common operations you perform on features and values within tables are selecting, calculating, and summarizing.

Selecting

You select features in order to work with a subset, or to assign a new attribute value to just those features—for example, assigning a specific rank to several different categories.

To do this, you select the rows in the data layer’s attribute table that pertain to those features. You select them using a query, which is usually in the form of a logical expression:

\[
\text{select attribute} = \text{value}
\]

For example, to select only commercial parcels, you’d specify:

\[
\text{Select Landuse = COM}
\]

where Landuse is the attribute name, and COM is the value for commercial.

Besides “equals,” other commonly used logical operators include greater than (>, less than (<), and not equal (<>).

You can also string several expressions together to select features meeting several criteria. For example, to find commercial parcels larger than 2 acres, you’d specify:

\[
\text{Select Landuse = COM and Acres > 2}
\]

If you wanted to select features meeting at least one of several values or criteria, use “or.” For example, to select commercial and industrial parcels, you’d specify:

\[
\text{Select Landuse = COM or Landuse = IND}
\]
Calculating

You can calculate attribute values to assign new values to features in the data table. You first add a new field to the table, then assign the values for that attribute to each feature. You can assign values directly, such as with ranks, or assign values based on existing fields, such as with ratios. For example, you could select all soils of a certain type, give them a rank based on how suitable they are for growing certain crops, then do the same for other soil types in the area:

Select Soil = 288

then,

Calculate Rank = 2

<table>
<thead>
<tr>
<th>Acnen</th>
<th>Soil Code</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.28C</td>
<td>1.78B</td>
<td>2</td>
</tr>
<tr>
<td>15.28C</td>
<td>2.36F</td>
<td>2</td>
</tr>
<tr>
<td>25.28F</td>
<td>2.31D</td>
<td>2</td>
</tr>
<tr>
<td>16.28F</td>
<td>2.36F</td>
<td>2</td>
</tr>
<tr>
<td>46.17F</td>
<td>2.36F</td>
<td>2</td>
</tr>
</tbody>
</table>

Or, you could calculate the average number of people per household in each census tract by dividing the total population by the number of households in each tract:

Calculate People per household = Population / Households

Summarizing

Another way of working with tables is to summarize the values for specific attributes to get statistics. In some cases you get a single value, such as a total or an average (mean); in other cases, you create a new table, listing some value for each type (category), plus a count of features. This is known as a frequency. It includes the number of features of each type and, optionally, an additional statistic, such as a total or a value.

Land value of parcels within a floodplain

<table>
<thead>
<tr>
<th>Landuse</th>
<th>Key of Parce</th>
<th>Total Sft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>1</td>
<td>61,046</td>
</tr>
<tr>
<td>Multi Family</td>
<td>1</td>
<td>127,059</td>
</tr>
<tr>
<td>Public</td>
<td>1</td>
<td>140,042</td>
</tr>
<tr>
<td>Farm</td>
<td>1</td>
<td>43,083</td>
</tr>
<tr>
<td>Single Family</td>
<td>43</td>
<td>43,083</td>
</tr>
<tr>
<td>Vacant</td>
<td>20</td>
<td>81,468</td>
</tr>
</tbody>
</table>

Amount of each land-use type within a floodplain