Part I

GIS and Basic Spatial Analysis Tasks

1 Getting Started with ArcGIS: Data Management and Basic Spatial Analysis Tools

A *Geographic Information System* (GIS) is a computer system that captures, stores, manipulates, queries, analyzes, and displays geographically referenced data. Among the diverse set of tasks a GIS can do, mapping remains the primary function. The first objective of this chapter is to demonstrate how GIS is used as a computerized mapping tool. The key skill involved in this task is the management of spatial and aspatial (attribute) data and the linkage between them. However, GIS is beyond mapping and has been increasingly used in various spatial analysis tasks as the GIS software becomes more capable and also friendlier to use for these tasks. The second objective of this chapter is to introduce some basic spatial analysis tools in GIS.

Given its wide use in education, business, and governmental agencies, ArcGIS is chosen as the major software platform to implement GIS tasks in this book. Unless pointed out otherwise, all studies in this book are based on ArcGIS 9.0. All chapters are structured similarly, beginning with conceptual discussions to lay out the foundation for methods, followed by case studies to get readers acquainted with techniques. Section 1.1 offers a quick tour of spatial and attribute data management in ArcGIS. Section 1.2 illustrates the typical process of GIS-based mapping in case study 1A: mapping the population density pattern in Cuyahoga County, Ohio. Section 1.3 surveys basic spatial analysis tools in ArcGIS, including queries, spatial joins, and map overlays. Section 1.4 illustrates some of the spatial analysis tools in case study 1B: extracting census tracts in Cleveland and deriving a polygon adjacency matrix. The polygon adjacency matrix defines spatial weights often needed in advanced spatial statistical studies such as spatial cluster and regression analyses (see Chapter 9). The chapter is concluded with a brief summary in Section 1.5.

It is assumed that readers have some basic GIS knowledge equivalent to one introductory GIS course. This chapter is not intended to review or cover all ArcGIS functions. Instead, it serves as a quick warm-up to prepare readers for more advanced spatial analysis in later chapters.

1.1 SPATIAL AND ATTRIBUTE DATA MANAGEMENT IN ArcGIS

Since ArcGIS is chosen as the primary software platform for this book, it is helpful to have a brief overview of its major modules and functions. *ArcGIS* was released in 2001 by Environmental Systems Research Institute, Inc. (ESRI) with a full-featured graphic user interface (GUI) to replace the older versions of *ArcInfo* based on typed commands. ArcGIS contains three major modules: ArcCatalog, ArcMap, and ArcToolbox. *ArcCatalog* views and manages spatial data files. *ArcMap* displays, analyzes, and edits the spatial data as well as attribute data. *ArcToolbox* contains various functions for managing and analyzing data, including managing map projections, converting between data formats, and implementing commands that are available from the older ArcInfo system. In ArcGIS 9.0, ArcToolbox is accessed in either ArcMap or ArcCatalog. Most, but not all, of the functions from the older ArcInfo system are available in the new ArcGIS system. For some tasks, we still need to access the old ArcInfo command interface. For example, Appendix 1 discusses how to import and export ASCII files in ArcInfo Workstation, and case study 2 (Section 2.3.2) uses ArcInfo Workstation to compute network distances.

1.1.1 Map Projections and Spatial Data Models

GIS differs from other information systems because of its unique capability of managing geographically referenced or spatial (location) data. Understanding the management of spatial data in GIS requires the knowledge of the *geographic coordinate system* with longitude and latitude values and its representations on various *plane coordinate systems* with x, y coordinates (map layers). Transforming the Earth's spherical surface to a plane surface or between different plane coordinate systems is a process referred to as *projection*. In ArcGIS, ArcMap automatically converts data of different coordinate systems to the first dataset added to the frame in map displays, commonly referred to as *on-the-fly reprojections*. However, this may be a time-consuming process if the dataset has a large size. It is a good practice to use the same projection for all data layers in one project. Two map projections are commonly used in the U.S.: *Universal Transverse Mercator* (UTM) and the *State Plane Coordinate System* (SPCS). Strictly speaking, the SPCS is not a projection; instead, it may use one of three different projections: Lambert Conformal Conic, Transverse Mercator, and Oblique Mercator. For the least distortion, north–south oriented states or regions use a Transverse Mercator projection, and east–west oriented states or regions use a Lambert Conformal Conic projection. Some states (e.g., Alaska, New York) may use more than one projection. For more details, one may refer to the Understanding Map Projections PDF file on the ArcGIS CD-ROM from the ESRI.

To check the existing projection for a spatial dataset in ArcGIS, one may use ArcCatalog by clicking the layer > choose Metadata > Spatial, or use ArcMap by right-clicking the layer > Properties > Source.

Projection-related tasks are conducted under ArcToolbox > Data Management Tools > Projections and Transformations. Under Projections and Transformations, the Define Projection tool creates a new projection definition file (PRJ) containing the projection parameters, or modifies an existing one (if incorrect). The Define Projection tool only labels a dataset with its correct coordinate system; it does not change the coordinates. If the spatial data are vector based, choose Feature > Project (also under Projections and Transformations) to actually transform the coordinate systems from one projection to another and generate a new layer. The tool provides the option to create a new coordinate system, use a predefined coordinate system, or import a coordinate system from an existing geodataset. If the spatial data are raster based, choose Raster > Project Raster.

Traditionally, a GIS uses either the vector or the raster data model to manage spatial data. A *vector GIS* uses geographically referenced points to construct spatial features of points, lines, and areas, and a *raster GIS* uses grid cells in rows and columns to represent spatial features. The raster data structure is simple and relatively easier to model. The vector data structure is used in most socioeconomic applications and is used in most applications in this book. Most commercial GIS software can convert from vector to raster data, or vice versa. In ArcGIS, the tools are available under ArcToolbox > Conversion Tools.

Earlier versions of ESRI's GIS software used the *coverage* data model. Later, *shapefiles* were developed for the ArcView package. Since the release of ArcGIS 8, the *geodatabase* model has become available and represents the new trend of object-oriented data model. The object-oriented data model stores the geometries of objects (spatial data) also as attribute data, whereas the traditional coverage or shapefile model stores spatial and attribute data separately. In any case, spatial and attribute data in socioeconomic analysis often come from different sources, and a typical task is to join them together in GIS for mapping and analysis. This involves attribute data management as discussed below.

1.1.2 Attribute Data Management and Attribute Join

A GIS includes both spatial and attribute data. Spatial data capture the geometry of map features, and attribute data describe the characteristics of map features. Attribute data are usually stored as a tabular file or table. ArcGIS reads several types of table formats. Shapefile attribute tables use the *dBase* format, ArcInfo Workstation uses the *INFO* format, and geodatabase tables use the *Microsoft Access* format. ArcGIS can also read several text formats, including comma-delimited text and tab-delimited text. Appendix 1 discusses how to import and export data in ASCII format in ArcGIS. Both tasks are important for advanced analysts who use GIS along with other software (e.g., SAS) or write programs for more complex computational tasks.

For basic tasks of data management, some can be done in both ArcCatalog and ArcMap, and some can be done only in ArcCatalog or ArcMap. Creating a new table or deleting/copying an existing table is done only in ArcCatalog (recall that ArcCatalog is for viewing and managing GIS data files). A table is created in ArcCatalog by right-clicking the folder where the new table will be placed > New. A table can be deleted or copied in ArcCatalog by right-clicking the table > Delete (or Copy).

Adding a new variable to a table (either a "field" in a shapefile attribute table or dBase file or an "item" in an ArcInfo Workstation INFO file) can be done in both ArcCatalog and ArcMap. Deleting an existing item in an INFO file can be also done in both ArcCatalog and ArcMap, but deleting a field in a dBase file can only be done in ArcMap. For example, for adding a field to a shapefile attribute table, one may use ArcCatalog > right-click the shapefile > Properties > click Fields and type the new field name into an empty row in the Field Name column and define its Data Type. One may also add a field in ArcMap by opening the table > Options > Add Field.

TABLE 1.1
Types of Relationships in Combining Tables

Relationship	Match		Join or Relate in ArcGIS
One to one	One record in the destination table	One record in the source table	Join
Many to one	Multiple records in the destination table	One record in the source table	Join
One to many	One record in the destination table	Multiple records in the source table	Relate
Many to many	Multiple records in the destination table	Multiple records in the source table	Relate

For deleting a field in ArcMap, one needs to open the table > right-click the field > Delete Field. Updating values in a table is done in ArcMap: open the table > right-click the field and choose Calculate Values. In addition, in ArcMap, basic statistics for a field can be obtained by right-clicking the field and choosing Statistics.

In GIS, an *attribute join* is often used to link information in two tables based on a common field (key). The table can be an *attribute table* associated with a particular geodataset or a *stand-alone table*. In an attribute join, the field name of the key does not need to be identical in the two tables to be combined, but the data type must match. There are various relationships between tables in a join: one to one, many to one, one to many, and many to many. For either a one-to-one or many-to-one relationship, a *join* is used to combine two tables in ArcGIS. However, when the relationship is one to many or many to many, a join cannot be performed. ArcGIS uses a *relate* to link two tables while keeping the two tables separated. In a relate, one or more records are selected in one table, and the associated records are selected in another table. Table 1.1 summarizes the relationships and corresponding tools in ArcGIS.

A join or relate is accessed in ArcMap. Under Table of Contents, right-click the spatial dataset or the table that is to become the destination table, and choose Joins and Relates > Join (or Relate) > choose "Join attributes from a table" in the Join Data dialog window. A join is temporary in the sense that no new data are created and the join relationship is lost once the project is exited if the active project is not saved. The result of a join can be preserved by exporting the combined table to a new table.

Once the attribute information is joined to a spatial layer, mapping is convenient in ArcGIS. In ArcMap, right-click the layer and choose Properties > Symbology to invoke the dialog window. In the dialog, one can select a field to map, choose colors and symbols, and plan the layout. Map elements (scale, north arrow, legends) can be added to the draft map by clicking Insert from the main menu bar.

1.2 CASE STUDY 1A: MAPPING THE POPULATION DENSITY PATTERN IN CUYAHOGA COUNTY, OHIO

For readers without much exposure to GIS, nothing demonstrates the value of GIS and eases the fear of GIS complexity better than the experience of mapping data from the public domain with a few clicks. This section uses a case study to illustrate how spatial and aspatial information are managed and linked in a GIS, and how the information is used for mapping. The procedures are designed in a such a way that most functions introduced in Section 1.1 will be utilized.

A GIS project begins with data acquisition. In many cases, it uses existing data. For socioeconomic applications in the U.S., the *Topologically Integrated Geographical Encoding and Referencing* (TIGER) files from the Census Bureau are the major source for spatial data, and the decennial census data from the same agency are the major source for attribute data. Both can be accessed at the website www.census.gov. Advanced ArcGIS users may download the TIGER data directly and use the TIGER conversion tool to extract needed spatial data. The TIGER conversion tool can be accessed under ArcToolbox > Coverage Tools > Conversion > To Coverage > Advanced Tiger Conversion (or Basic Tiger Conversion). The conversion process may be time-consuming, and resulting files may require further editing work. Fortunately, many TIGER-based spatial layers are already processed and available in ArcGIS formats (shapefiles or coverage) in public domains. Some are contained in the data CDs that come with the ArcGIS software from the ESRI. If the spatial data are in the coverage interchange format (e00), the file can be converted to a coverage following ArcToolbox > Coverage Tools > Conversion > To Coverage > Import from Interchange File. In this case study, the spatial data are downloaded in shapefiles from the ESRI website.

Although readers can download the data as instructed, all data needed for the project are provided in the enclosed CD for convenience:

1. Shapefile `tgr39035trt00`
2. dBase file `tgr39000sf1trt.dbf`

Throughout this book, all computer file names and variable names (also some command lines that are specific to the projects) are in the `Courier New` font.

The following is a step-by-step guide for the process:

1. *Downloading spatial data*: Visit the ESRI's website for the Census 2000 TIGER/Line Data at http://www.esri.com/data/download/census2000_tigerline/. Select the state (Ohio) and the county (Cuyahoga) and download the census tract 2000 layer shapefile in WinRAR ZIP file. The unzipped shapefiles[1] share the name `tgr39035trt00`.[2]
2. *Transforming to UTM*: In ArcCatalog, we may check the projection for the shapefile `tgr39035trt00` and find out that it uses the geographic coordinate system. In ArcToolbox, choose Data Management Tools > Projections and Transformations > Feature > Project to invoke the dialog. In the dialog, select `tgr39035trt00.shp` as the Input Dataset, name the Output Dataset `cuyautm.shp`, and define the Output Coordinate System as UTM (zone 17; units, meters). Here we import the definition file from an existing dataset to define the output coordinate system: click the graphic icon under "Output Coordinate System" to activate the Spatial Reference Properties dialog > select Import > choose `clevbnd`. Figure 1.1 shows the dialog windows for the task. Click OK in both windows to execute.

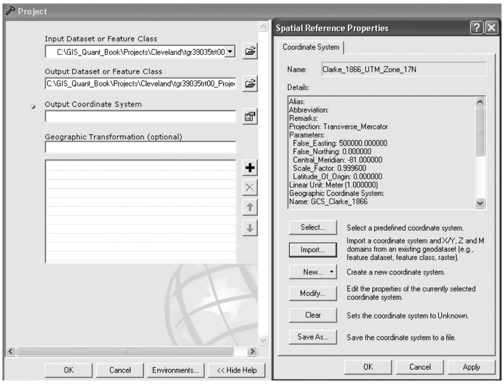

FIGURE 1.1 Dialog windows for projecting a spatial dataset.

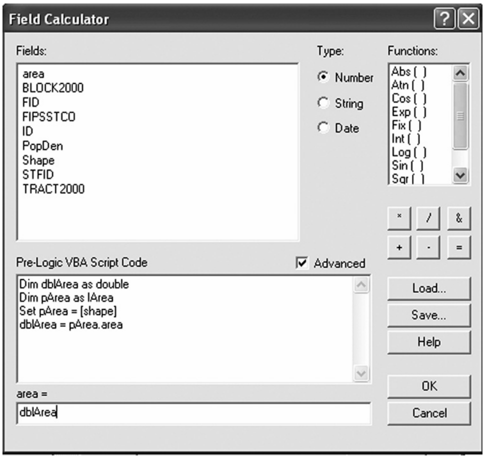

FIGURE 1.2 Dialog window for updating area in shapefile.

3. *Updating area for the shapefile*[3]: In ArcMap,[4] open the attribute table of `cuyautm` and add a new field `area` by clicking the Options button, selecting Add Field, and choosing the data type Double. Right-click the field `area` and choose Calculate Values. In the dialog, check Advanced and type the following Visual Basic (VBA) statements in the first text box:

```
Dim dblArea as double
Dim pArea as IArea
Set pArea = [shape]
dblArea = pArea.area
```

Type `dblArea` in the text box directly under the field `area`. Click OK to update areas. Figure 1.2 shows a sample dialog window for updating the areas.

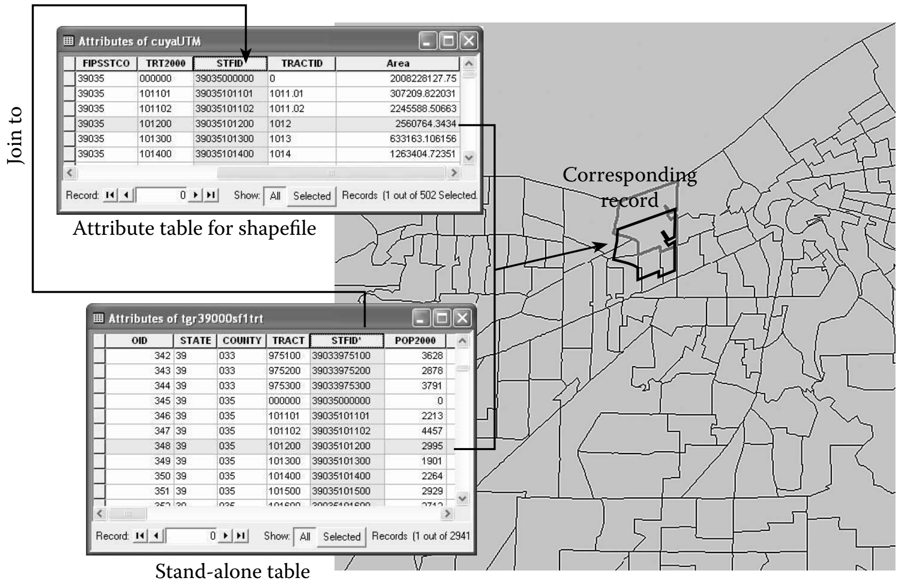

FIGURE 1.3 Attribute join in ArcGIS.

4. *Downloading attribute data*: The corresponding attribute dataset is downloaded at the same site by selecting the layer "2000 census tract demographics (SF1)." The attribute dataset `tgr39000sf1trt.dbf` is in dBase format for the whole state of Ohio, where `SF1` stands for Summary File 1 (i.e., the 100% count census data based on the short form).[5]
5. *Extracting attribute data for Cuyahoga County*[6]: In ArcMap, add the census data `tgr39000sf1trt.dbf` and open it. Click the Options tab at the lower right corner of the table > choose Select By Attributes > input the Structured Query Language (SQL) statement `county='035'` and click Apply (see Section 1.3 for more discussion on query in ArcGIS). All records in Cuyahoga County are selected. Click the Options tab again > Export. Name the extracted data file `cuya2k_popu.dbf`. If desirable, the table may be simplified by deleting all fields except for `STFID` and `POP2000` needed for the project.
6. *Joining spatial and attribute data*: Right-click the layer `cuyautm` > choose Joins and Relates > Join to join the table `cuya2k_popu.dbf` to the shapaefile `cuyautm` based on the common key `STFID`. STFID is the unique ID for each tract containing the codes for the state (two digits), county (three digits), and tract (six digits). Figure 1.3 shows how the spatial and attribute data are joined and related to the graphic elements on the map.
7. *Adding and calculating population density*: Right-click the layer `cuyautm` again and choose Open Attribute Table to examine the resulting combined table. In the combined table, a field name is now identified by its source table name and the original field name. For example,

cuyautm.area represents the field area from the attribute table of shapefile cuyautm, and tgr39000sf1trt.STFID represents the field STFID from the table tgr39000sf1trt.dbf (some long field names may be truncated). In project instructions hereafter in this book, *we omit the source table name in most cases when referring to field names* (e.g., in calculation formulas, table joins, and others) unless we wish to emphasize the source table.

Click the Options tab > Add Field to add a new field popuden. The added field will be placed in the combined table as the last field of cuyautm, but ahead of the first field of cuya2k_popu.dbf. Right-click the field popuden, choose Calculate Values, and input the formula 1000000*[POP2000]/[area]. In the formula, both fields POP2000 and area are entered by clicking the field names from the top box to save time and minimize chances of errors. In project instructions hereafter in this book, we simply write the formula such as popuden=1000000*POP2000/area for similar tasks. Note the map projection unit is meter, and the formula calculates population densities as persons per square kilometer.

8. *Mapping population density pattern*: Right-click the layer cuyautm > choose Properties > Symbology > Quantities > Graduated Colors to map the field popuden. Experiment with different classification methods, number of classes, and color schemes. From the main menu bar, choose View > Layout View to preview the map. Also from the main menu bar, choose Insert > Legend (Scale Bar, North Arrow, and others) to add map elements.

   A population density map for the study area is shown in Figure 1.4. The large tract on the north is Lake Erie. The map uses customized breaks for density classes.

1.3 SPATIAL ANALYSIS TOOLS IN ArcGIS: QUERIES, SPATIAL JOINS, AND MAP OVERLAYS

Many spatial analysis tasks utilize the information on how spatial features are related to each other in terms of the location. Spatial operations such as queries, spatial joins, and map overlays[7] provide basic tools in conducting such tasks.

Queries include attribute (aspatial) queries and spatial queries. An *attribute query* uses information in an attribute table to find attribute information (in the same table) or spatial information (features in a spatial data layer). Attribute queries are accessed in ArcMap: either (1) under Selection from the main menu bar > choose the option Selection by Attributes or (2) in an opened table, choose the Options tab > Selection by Attributes. Both allow users to select spatial features based on a query expression in SQL using attributes (or to simply select attribute records from a stand-alone table). Step 5 in case study 1A already used this function. Another option under Selection in the main menu bar is Interactive Selection Method, which uses a pointer to select features on the screen (in either a map or a table).

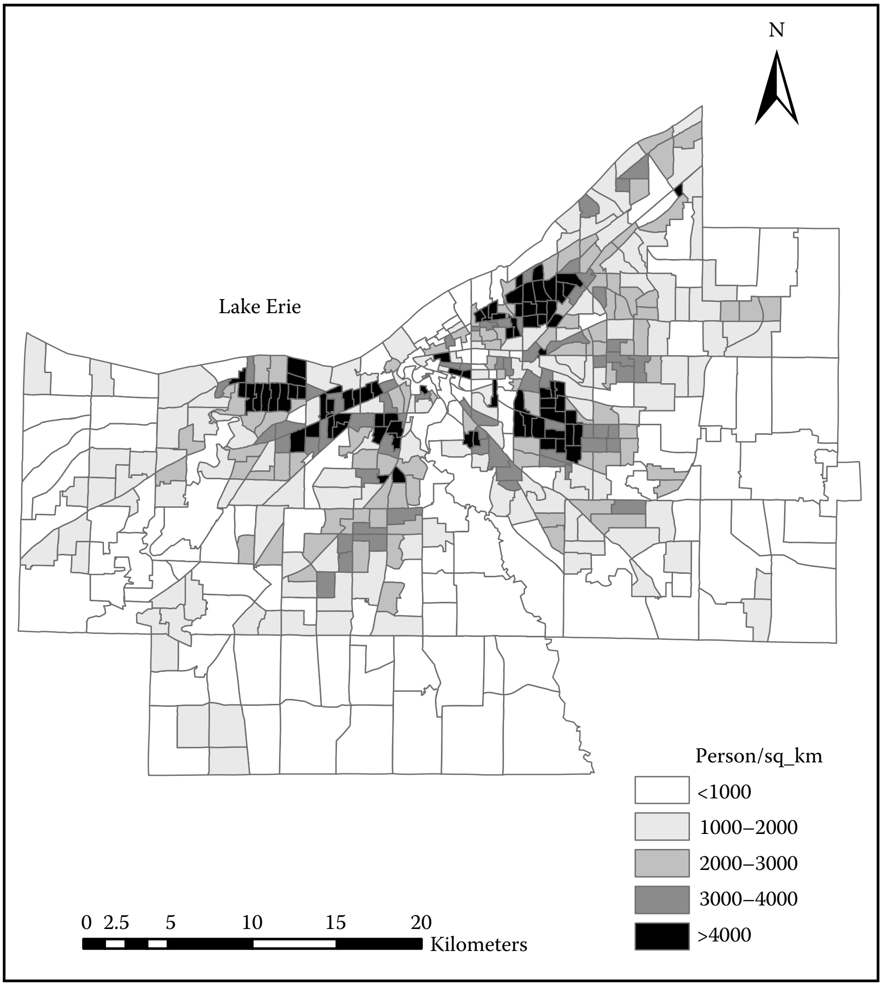

FIGURE 1.4 Population density pattern in Cuyahoga County, Ohio, 2000.

Compared to other information systems, one unique feature of GIS is its capacity of *spatial query*, which finds information based on location relationship between features from different layers. The option Selection by Location under Selection in the main menu bar searches for features in one layer based on its spatial relationship with another layer. The spatial relationships are defined by "intersect," "are within a distance of," "completely contain," "are completely within," etc.

Selected data by either an attribute query or a spatial query can be exported to a new dataset: (1) spatial features are saved in a layer by right-clicking the source layer and choosing Data > Export Data, and (2) attribute data are saved in a table by clicking the Options tab in the opened table > Export.

While an attribute join utilizes a common field between two tables, a *spatial join* uses the locations of spatial features, such as overlapping or proximity between two layers. We use the terms *source layer* and *destination layer* to differentiate the two layers by their roles: attributes from the source layer are processed and transferred to the destination layer. If one object in the source layer corresponds to one or multiple objects in the destination layer, it is a *simple join*. For example, by spatially joining a polygon layer of counties (source layer) with a point layer of school locations (destination layer), attributes of each county (e.g., county code, name, administrator) are assigned to schools that fall within the county boundary. If multiple objects in the source layer correspond to one object in the destination layer, two operations may be performed: summarized join and distance join. A *summarized join* summarizes the numeric attributes of features in the source layer (e.g., average, sum, minimum, maximum, standard derivation, or variance) and adds the information to the destination layer. A *distance join* identifies the closest object (out of many) in the source layer from the matching object in the destination layer and transfers all attributes of this nearest object (plus a distance field showing how close they are) to the destination layer. For example, one may spatially join a point layer of geocoded crime incidents (source layer) with a polygon layer of census tracts (destination layer), and generate aggregated crime counts in census tracts (summarized join); or spatially join a point layer of bus stations (source layer) with a point layer of census block centroids (destination layer) and identify the nearest bus stop to each block (distance join).

There are a variety of spatial joins between different spatial features (Price, 2004, pp. 287–288). Table 1.2 summarizes all types of spatial joins in ArcGIS. Spatial joins are accessed in ArcMap in a manner similar to that of attribute joins: right-click the source layer > choose Joins and Relates > Join. In the Join Data dialog window, choose "Join data from another layer based on spatial location," instead of "Join attributes from a table."

Map overlays may be broadly defined as any spatial analysis involving modifying features from different layers. The following reviews some of the most commonly used map overlay tools (available with any ArcGIS license): Clip, Intersect, Union, Buffer, and Multiple Ring Buffer. A *Clip* truncates the features from one layer using the outline of another. An *Intersect* overlays two layers and keeps only the areas that are common to both. A *Union* also overlays two layers but keeps all the areas from both layers. A *Buffer* creates areas by extending outward from point, line, or polygon features over a specified distance. A *Multiple Ring Buffer* generates buffer features based on a set of distances. In ArcGIS 9.0, the above map overlay tools are grouped under different tool sets through ArcToolbox > Analysis Tools: Clip is under the Extract tool set, Intersect and Union are under the Overlay tool set, and Buffer and Multiple Ring Buffer are grouped under the Proximity tool set. Other map overlay tools used in the projects of this book include Erase (Section 1.4.2, step 3), Near (Section 2.3.2, step 2), Point Distance (Section 2.3.1, step 2), Dissolve (Section 4.3.1, step 2), and Append (Section 4.3).

One may notice the similarity among spatial queries, spatial joins, and map overlays. Indeed, many spatial analysis tasks may be accomplished by any one of the three. Table 1.3 summarizes the differences between them. A spatial query only finds the information on screen and does not create new datasets (unless one chooses to export the selected records or features). A spatial join always saves the result in a new layer. There is an important difference between spatial joins and map overlays.

TABLE 1.2
Types of Spatial Joins in ArcGIS

Source Layer (S)	Destination Layer (D)	Simple Join	Distance Join	Summarized Join
Point	Point		For each point in D, find its closest point in S, and transfer attributes of that closest point to D	For each point in D, find all the points in S closer to this point than to any other point in D, and transfer the points' summarized attributes to D
Line	Point		For each point in D, find the closest line in S, and transfer attributes of that line to D	For each point in D, find all lines in S that intersects it, and transfer the lines' summarized attributes to D
Polygon	Point	For each point in D, find the polygon in S containing the point, and transfer the polygon's attributes to D	For each point in D, find its closest polygon in S, and transfer attributes of that polygon to D	
Point	Line		For each line in D, find its closest point in S, and transfer the point's attributes to D	For each line in D, find all the points in S that either intersect or lie closest to it, and transfer the points' summarized attributes to D
Line	Line	For each line in D, find the line in S that it is part of, and transfer the attributes of source line to D		For each line in D, find all the lines in S that intersect it, and transfer the summarized attributes of intersected lines to D
Polygon	Line	For each line in D, find the polygon in S that it falls completely inside, and transfer the polygon's attributes to D	For each line in D, find the polygon in S that it is closest to, and transfer the polygon's attributes to D	For each line in D, find all the polygons in S crossed by it, and transfer the polygons' summarized attributes to D
Point	Polygon		For each polygon in D, find its closest point in S, and transfer the point's attributes to D	For each polygon in D, find all the points in S that fall inside it, and transfer the points' summarized attributes to D
Line	Polygon		For each polygon in D, find its closest line in S, and transfer the line's attributes to D	For each polygon in D, find all the lines in S that intersect it, and transfer the lines' summarized attributes to D
Polygon	Polygon	For each polygon in D, find the polygon in S that it falls completely inside, and transfer the attributes of source polygon to D		For each polygon in D, find all the polygons in S that intersect it, and transfer the summarized attributes of intersected polygons to D

TABLE 1.3
Comparison of Spatial Query, Spatial Join, and Map Overlay

Basic Spatial Analysis Tools	Function	Whether a New Layer Is Created	Whether New Features Are Created	Computation Time
Spatial query	Finds information based on location relationship between features from different layers and displays on screen	No (unless the selected features are exported to a new dataset)	No	Least
Spatial join	Identifies location relationship between features from different layers and transfers the attributes to destination layer	Yes	No	Between
Map overlay	Overlays layers to create new features and saves the result in a new layer	Yes	Yes (splitting, merging, or deleting features to create new)	Most

A spatial join merely identifies the location relationship between spatial features of input layers and does not change existing spatial features or create new features. In the process of a map overlay, some of the input features are split, merged, or dropped for generating a new layer. In general, map overlay operations take more computation time than spatial joins, and spatial joins take more time than spatial queries.

1.4 CASE STUDY 1B: EXTRACTING CENSUS TRACTS IN THE CITY OF CLEVELAND AND ANALYZING POLYGON ADJACENCY

The following datasets are needed for the project and are provided in the CD:

1. Shapefile `cuyautm` contains the census tracts in Cuyahoga County, Ohio.
2. Coverage `clevbnd` defines the boundary for the city of Cleveland, OH.

Each coverage in the CD is in ArcInfo interchange file (.e00) format, which needs to be converted back for use in ArcToolbox by selecting Coverage Tools > To Coverage > Import From Interchange File. Dataset 1 is a product from case study 1A, but provided in the CD so that one may start this project independent of case study 1A. The coverage `clevbnd` for the city boundary is found in a public domain.

1.4.1 Part 1: Extracting Census Tracts in Cleveland

In many cases, GIS analysts need to extract a study region that is contained in a larger area. This part of the project extracts the census tracts in Cleveland from Cuyahoga

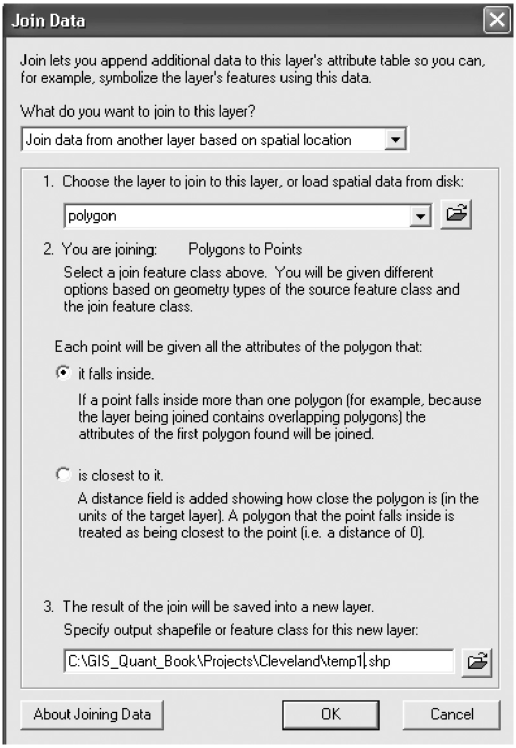

FIGURE 1.5 Dialog window for spatial join.

County. Overlaying the two layers shows that `clevbnd` is slightly off the census tract boundaries in `cuyautm` near the borders, and `cuyautm` contains more geographic details. Although the census tracts in `cuyautm` do not exactly match the boundary of `clevbnd`, their centroids are within the boundary of `clevbnd`. The plan is to identify the centroids of `cuyautm` that fall inside `clevbnd` and use this information to extract a polygon layer of census tracts for the city. If one simply uses `clevbnd` to clip `cuyautm`, not all geographic details of tracts in `cuyautm` would be preserved.

1. *Generating county census tract centroids*: Activate ArcToolbox in ArcMap > choose Data Management Tools > Features > Feature To Point. In the dialog, choose `cuyautm` as Input Features, name `cuya_pt` for Output Feature Class, and check the option Inside. A shapefile `cuya_pt` for all tract centroids is generated.
2. *Identifying tract centroids inside the city boundary*: Right-click the destination layer `cuya_pt` and choose Joins and Relates > Join. In the dialog, choose the option "Join data from another layer based on spatial location," the source layer "`clevbnd` polygon," and the option "it falls inside," and name the output shapefile `tmp1`. Figure 1.5 shows the dialog window for a spatial join. Open the attribute table for `tmp1` and note that `clevbnd_id` = 1 for all tracts inside the city boundary, 0 for those outside.

3. *Attaching information of identified tract centroids to the polygon layer*: Add the layer cuyautm, right-click it and choose Joins and Relates > Join > choose the option "Join attributes from a table" and tmp1 as the source table, and use STFID as the common key (in both the destination layer cuyautm and the source table tmp1).
4. *Extracting census tracts in the city*: Open the attribute table of cuyautm > click the Options tab > choose Select by Attributes > input the selection criterion tmp1.clevbnd_id = 1. All polygons within the city are selected and highlighted. Right-click the layer cuyautm and choose Data > Export Data > make sure that "Selected features" is selected in the top box, and name the output clevtrt. The shapefile clevtrt contains all census tracts in Cleveland.

The above instructions use a spatial join. As explained in Section 1.3, one may also use a spatial query tool (Selection by Location) or a map overlay tool (ArcToolbox > Analysis Tools > Overlay > Identity) to accomplish the same task. For example, a spatial query tool can be used to obtain the same result in one step: choose Selection from the main menu bar > Selection by Location > use the dialog to "select features from cuyautm that have their center in clevbnd polygon" > export the selected features to a shapefile clevtrt.

1.4.2 Part 2: Identifying Contiguous Polygons

Deriving the *polygon adjacency matrix* is a very important task in spatial analysis. For example, in Chapter 9, both the area-based spatial cluster analysis and spatial regression utilize the matrix to define *spatial weights* in order to account for spatial autocorrelation. Adjacency between polygons may be defined in two ways: (1) *rook contiguity* uses only common boundaries to define adjacency, and (2) *queen contiguity* includes all common points (boundaries and vertices) (Cliff and Ord, 1973). For the rook contiguity, one may simply use the ArcInfo Workstation command PALINFO to generate the polygon adjacency matrix. This case study uses the queen contiguity to define polygon adjacency in order to illustrate some of the basic spatial analysis tools discussed in Section 1.3.

Here we focus on identifying contiguous polygons for one exemplary tract. Similar to step 4 in Part 1, select the tract with TRACTID = '1038' from the layer clevtrt and export the selected feature to a shapefile zonei. This part of the project is to identify neighboring tracts for zonei based on the queen contiguity. Figure 1.6 shows the areas around the sample tract with their TRACTID values. Based on the queen contiguity, there are six neighboring tracts (1026, 1028, 1029, 1035, 1036, and 1039) for tract 1038. If based on the rook contiguity, tract 1028 would not be included as a neighboring tract.

The following implements the task of identifying contiguous tracts for zonei:

1. *Buffering tract*: In ArcToolbox, choose Analysis Tools > Proximity > Buffer. Buffer a small distance (say, 30 m)[8] around the Input Features zonei, and name the Output Feature Class zonei_buff.

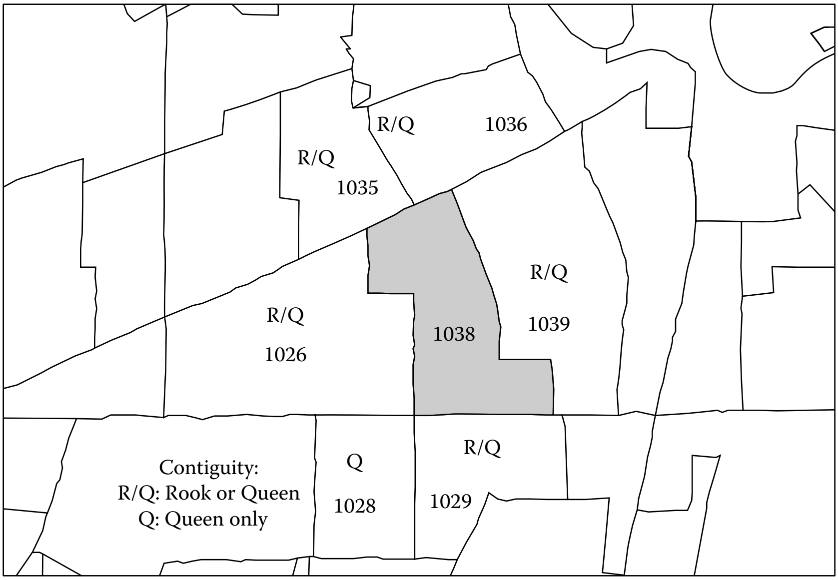

FIGURE 1.6 Rook contiguity vs. queen contiguity.

2. *Clipping the buffer from study area*: In ArcToolbox, choose Analysis Tools > Extract > Clip. Choose `clevtrt` as the Input Features and `zonei_buff` as the Clip Features, and name the Output Feature Class `zonei_clip`.
3. *Extracting neighboring polygons*: In ArcToolbox, choose Analysis Tools > Overlay > Erase. Choose `zonei_clip` as the Input Features and `zonei` as the Erase Features, and name the Output Feature Class `zonejs`. The shapefile `zonejs` contains all neighboring tracts of `zonei` based on the queen contiguity.

Figure 1.7 illustrates the process. The layer `zonei` is a single tract 1038. By buffering, the output layer `zonei_buff` contains a single polygon (the original tract is shown inside only for reference). By clipping the buffered zone from the study area, `zonei_clip` has seven polygons, including the original tract. The original tract `zonei` is removed by erasing it from `zonei_clip`. The output is a six-tract layer `zonejs` (with TRACTIDs shown in Figure 1.7).

Deriving a polygon adjacency matrix for the study area requires iterations of the same process on all tracts. An Arc Micro language (AML) program Queen_Cont`.aml` is provided in the CD for this task, which is based on Shen (1994).

At the end of the project, one may use ArcCatalog to delete unneeded data to save disk space, but keep the layers `cuyautm`, `cuya_pt`, and `clevtrt`, which will be used in future projects.

1.5 SUMMARY

The following summarizes major GIS and spatial analysis skills learned in this chapter:

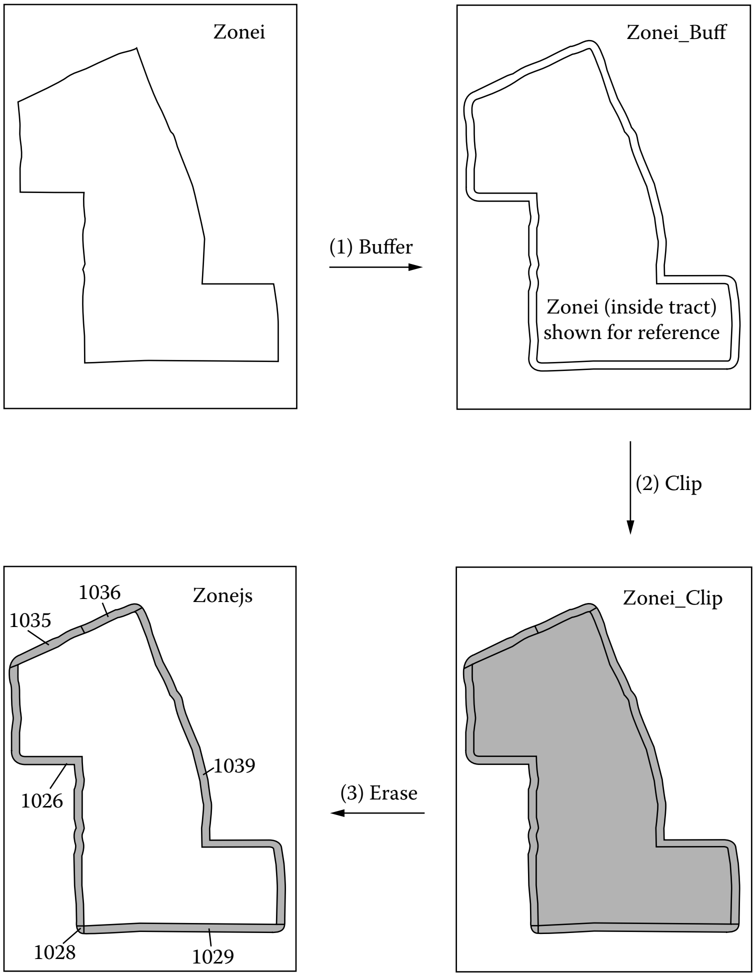

FIGURE 1.7 Workflow for defining queen contiguity.

1. Spatial data formats in ArcGIS and conversions between them
2. Map projections and transformations between them
3. Attribute data management (creating, editing, and deleting datasets and fields)
4. Attribute join (including joining attribute data with spatial data)
5. Mapping attributes
6. Spatial joins
7. Attribute and spatial queries
8. Map overlay operations (clip, buffer, intersect, union, and erase)

Other noticeable tasks include finding spatial and attribute data in the public domain, updating areas in a shapefile, and extracting centroids from a polygon layer to create a point layer. Projects in other chapters will also utilize these skills.

This chapter also discusses key concepts such as different relationships between tables (one to one, many to one, one to many, and many to many), various spatial joins, and differences among spatial queries, spatial joins, and map overlays.

For additional practice of GIS-based mapping, readers can download the census data and TIGER files for a familiar area and map some demographic (population, race, age, sex, etc.) and socioeconomic (income, poverty, educational attainment, family structure, housing characteristics, etc.) variables in the area.

APPENDIX 1: IMPORTING AND EXPORTING ASCII FILES IN ArcGIS

For a small dataset in ASCII (text) format, it is fairly easy to convert it to ArcGIS-readable formats. One may use Microsoft Excel to open the file, add a top row containing field headings, and save the file as a CSV file (comma-delimited text file). CSV files are readable in ArcGIS. One crucial shortcoming in this process is the lack of control on defining field types and formats. For example, all values for census tract codes or STFID in an ASCII file appear to be numeric. A CSV file generated from the ASCII file automatically defines the field as numeric and makes it infeasible to join the data to a GIS layer extracted from TIGER files, which usually defines the field as character.

Conversely, in ArcMap, one may open a table (either a feature attribute table or a stand-alone table) and choose Options > Export to export the table in dBase format. The dbase file can then be read in Microsoft Excel and saved as an ASCII file. Alternatively, one may also use ArcToolbox > Spatial Statistics Tools > Utilities > Export Feature Attribute to Ascii, which exports feature class coordinates and selected attribute values to a space-, comma-, or semicolon-delimited ASCII text file. However, only one variable (along with the coordinates) is exported at a time.

Microsoft Access is also a commonly used package for attribute data management, including tasks of importing and exporting ASCII files. The following discusses how to import and export a large ASCII file through ArcInfo Workstation.

1. *Converting ASCII to INFO file*:
   a. In ArcCatalog, click on the directory or workspace that is to contain the new file, choose File > New > INFO table to create a new INFO file (say, `ninfo`), and define all fields (name, data type, and format).
   b. Invoke the ArcInfo command interface in Windows by choosing Start > All Programs > ArcGIS > ArcInfo Workstation > Arc. Type `w …` to navigate to the right workspace (e.g., `w c:\Quant_GIS\proj1`), and then type `tables` to activate the TABLES module.
   c. In TABLES, type `select ninfo` to choose the newly defined INFO file.
   d. Say `tfile` is the text file to be converted. Type `add from tfile` to append all data. The converted table can be viewed by typing the command `list`.
   e. Type `quit` to exit TABLES and type `quit` again to exit ArcInfo Workstation.

2. *Exporting INFO file or ArcGIS coverage feature tables to ASCII*: Conversely, an INFO file or an ArcInfo feature table (.PAT or .AAT) can be exported to an ASCII file as follows:
   a. In ArcInfo Workstation, navigate to the workspace and type `tables` to activate TABLES module.
   b. In TABLES, type `select ninfo` (e.g., `ninfo` is the file name) to choose the file.
   c. Type `unload tfile1` to export all items of the data to an ASCII file (say, named as `tfile1`). One may limit the fields to be exported by a command, `unload yfile item1 item2` (say, the field names to be unloaded are `item1 and item2`), or further specify the format by adding the option COLUMNAR, such as `unload yfile item1 item2 columnar`. The option `COLUMNAR` may be very useful as the output ASCII file is space delimited and each field takes exactly the same space as defined in the INFO file. The UNLOAD command without the option exports to a comma-delimited ASCII file and encloses values from character fields with single quotes (‘’).

NOTES

1. A layer in the shapefile format has at least three files (.dbf, .shp, .shx) associated with it. Some contain additional files (.prj, .sbx, .avl, .xml). For convenience, the remainder of this book uses the singular term *shapefile* to refer to multiple files associated with one shapefile layer.
2. In the file name, `tgr` indicates its source from TIGER files, `39` is the state's FIPS code, `035` is the county code, and `trt00` stands for census tracts in 2000.
3. Alternatively, one may use ArcToolbox > Spatial Statistics Tools > Utilities > Calculate Areas to implement the task (a new output feature will be created).
4. Unless specified otherwise, ArcMap is the work environment in all project instructions.
5. This dBase file does not contain all census variables. For complete census data, visit the 2000 Census website at http://www.census.gov/main/www/cen2000.html. Processing the files (e.g., SF1, SF3) requires understanding the 2000 Census data structure and using some data analysis software (e.g., SAS, Access).
6. One may omit this step and join the table `tgr39000sf1trt.dbf` directly to the layer `cuyautm`. The resulting layer automatically excludes records of other counties.
7. Some may consider these traditional GIS terms outdated. We use the terms to emphasize distinctive tasks accomplished by these spatial operations.
8. The buffer distance must be larger than the fuzzy tolerance (i.e., about 1 m for zonei), but small enough not to go beyond the immediate neighboring polygons.