

Chapter 4:

Mapping Crime and Geographic Information Systems

The GIS revolution

Although it is tempting to think of geographic information systems (GIS) as a thoroughly contemporary technology, its conceptual roots reach far back. A GIS is based on drawing different spatial distributions of data on paper (or other suitable media) and overlaying them on one another to find interrelated points. Foresman (1998) notes evidence that this model was used at the Angkor Wat temple complex (in today's Cambodia) in the 11th century. Modern geographic information systems can be linked to developments in the 1960s, including land use analysis in the United Kingdom by Coppock (1962), development of the Canadian GIS by Tomlinson (1967), and publication of McHarg's *Design with Nature* (1969).



Early GIS efforts were restricted by the limitations of older computer systems lacking memory and speed, such as the 512k memory of the IBM 360/65, a computer widely used in the 1960s and 1970s (Tomlinson, 1998). This limited the size of data sets and made it difficult to simultaneously manipulate multiple observations or large numbers of variables.



These constraints limited the attractiveness of GIS technology to law enforcement agencies. Weisburd and McEwen (1997) noted that police departments typically lacked the computer resources and the base



What Is a GIS?

A GIS is a computerized mapping system that permits information layering to produce detailed descriptions of conditions and analyses of relationships among variables.

Strictly speaking, any system that permits the representation and analysis of geographic information is a geographic information system. The acronym GIS is understood to refer to computer-based software, generally in the form of a few popular proprietary software packages. Although a prominent component of a GIS, proprietary software does not define a GIS.

Even the acronym "GIS" is the subject of debate, with some arguing that the "s" stands for "system(s)," while others object that this is too narrow, and the "s" should stand for "science."

maps necessary to support a GIS operation. Labor, startup, and operational costs, such as coding incident addresses for the computer, were prohibitive, and the field lacked user-friendly software.

The earliest applications of crime mapping appeared in the mid-1960s, according to Weisburd and McEwen, who cited the work of Pauly, McEwen, and Finch (1967) and Carnaghi and McEwen (1970). Most early crime maps were produced using the SYMAP program developed at Harvard University. Input required punched cards, and output was produced on a line printer that limited detail to the size of the printer typeface. Shading of choropleth maps was accomplished by overstriking printer characters so that they darkened (figure 4.1). All maps produced this way were black and white. Line symbol mapping (of auto thefts, for example) was done on a plotter (figure 4.2). (See Weisburd and McEwen, 1997, for additional information.)

Later, the SYMVU variant of SYMAP used line renditions on a plotter to produce three-dimensional visualizations, like that shown in figure 1.13. Pioneering work done in St. Louis by the police department involved the establishment of a Resource Allocation Research Unit with the objective of improving the efficiency of patrol operations. The unit recognized that fixed boundaries would have to be established for crime mapping purposes. This was done using so-called Pauly Areas, named for (then) Sgt. Glenn A. Pauly, who designated mapping areas similar in size to census block groups. Thomas McEwen then devised a system for geocoding by relating street segments to Pauly Areas. This was the first time in an operational setting that computerized visualization of crime data was recognized as a management tool.

GIS applications in policing took off in the late 1980s and early 1990s as desktop computing became cheaper and software



became more accessible and user friendly. To date, large departments have been more likely to adopt the innovation; however, almost any police agency that wants a GIS can have one. Foresman (1998, figure 1.2, p.11) recognized five ages of GIS development.

The *Pioneer Age*, which lasted from the mid-1950s to the early 1970s, was characterized by primitive hardware and software. The *Research and Development Age* lasted into the 1980s and overlapped the *Implementation and Vendor Age*, which in turn lasted into the 1990s, when the *Client Applications Age* began. The *Local and Global Network Age* followed.

Crime mapping came of age in the Implementation and Vendor period, when computing costs began to fall and software became more immediately useful. Over the past decade we have also seen more examples of police departments commissioning customized versions of software to meet their individual needs.

A survey of police departments conducted in 1997–98 (Mamalian and La Vigne et al., 1999) showed that only 13 percent of 2,004 responding departments used computer mapping. Slightly more than one-third of large departments (those with more than 100 officers) did so, but only 3 percent of small units did. On average, departments had used computer mapping for 3.3 years. Crime analysts were the primary users of mapping, with relatively few patrol officers involved. The types of data most likely to be mapped were:

- Arrests and incidents, including Uniform Crime Reports Parts I and II crimes.

- Calls for service.

- Vehicle recoveries.

The most frequent applications were:

- Automated pin mapping (point data).

- Cluster or hot spot analysis.

- Archiving data.

The GIS perspective

Software

Other information gleaned from the computer mapping survey showed that 88 percent of respondents used off-the-shelf GIS software, such as MapInfo® (about 50 percent), ArcView® (about 40 percent), ArcInfo® (about 20 percent), and others (about 25 percent). Some departments used more than one package. Approximately 38 percent of departments that used mapping had done some kind of customizing, and 16 percent were using global positioning system (GPS) technology.

It would be expected that GIS computer mapping would follow the classic bell curve of innovation adoption (figure 4.3). In the lower left are the early adopters, followed by the early majority, the late majority, and the laggards. A wider bell means a longer process, or greater reluctance to adopt. GIS adoption will likely be a rapid process because the technology is simultaneously becoming cheaper and more powerful.



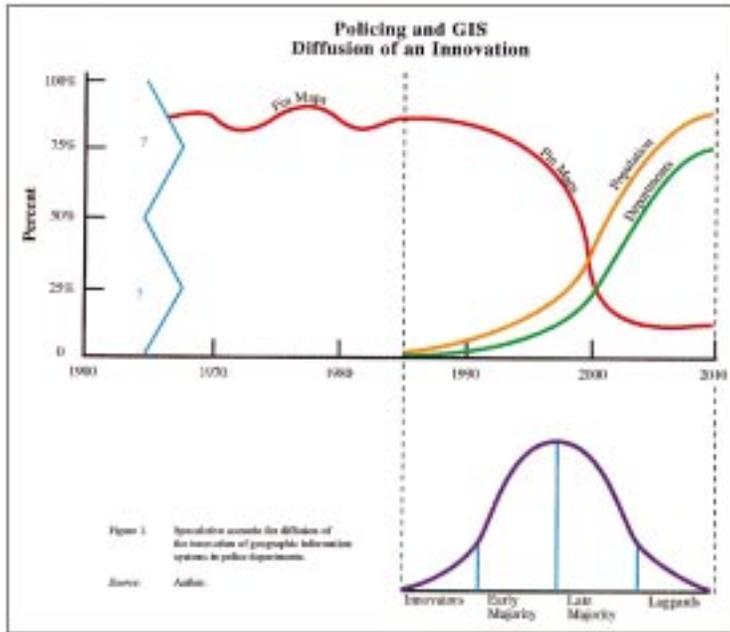


Figure 4.3

Hypothetical bell curves for GIS adoption in police departments.

Source: Keith Harries.

Vector? Raster? Say what?

You will probably hear the terms *vector* and *raster* in mapping conversations. How much do you need to know about them? Enough to understand the jargon and enough to make informed choices about formats.

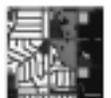
Raster maps store data in the form of a matrix, or grid. A raster map represents information by assigning each pixel, or picture element, a data value and shading it accordingly. The size of a raster data matrix can be determined by multiplying the number of rows by the number of columns in the display. For example, if the display settings on the computer read 640 by 480 pixels, the matrix has a total of 307,200 pixels—each of which would have a data value on a raster-based map.

Vector-based maps are built from digitized points that may be joined to form lines, or vectors, and polygons, or closed shapes. (Digitizing means recording the exact

coordinates of each point along x-y axes.) You will sometimes see the term topology used in connection with vector format, and this refers to the study of geometric forms. This type of analysis is integral to GIS but is largely transparent to users.

Each format has advantages and drawbacks. For example, lines in raster displays may appear jagged if resolution is not high enough. Rescanning an image at a finer resolution greatly increases file size. For example, if the 640- by 480-pixel screen is doubled to 1,280 by 960 pixels, the number of pixels increases four times from 307,200 to 1,228,800 pixels. However, raster processing is quicker.

Vector files are good at showing lines but are labor intensive due to the need to clean and edit vector data (Faust, 1998). Applications that use the vector format include emergency personnel routing and determining whether a suspect could have traveled a particular route in a given



amount of time. Most databases in urban areas use vector format; examples include street centerlines, precinct boundaries, and census geography. Vector files are not very good for managing continuous distributions, such as temperature or land elevation.

Although crime is not continuously distributed (crimes occur at separate points in geographic space), we can estimate values between known points to construct a continuous surface representation. The triangulated irregular network (TIN) data structure is one frequently used way of doing this. In it, points are connected to form triangles, the attributes of which become the basis for surface construction.

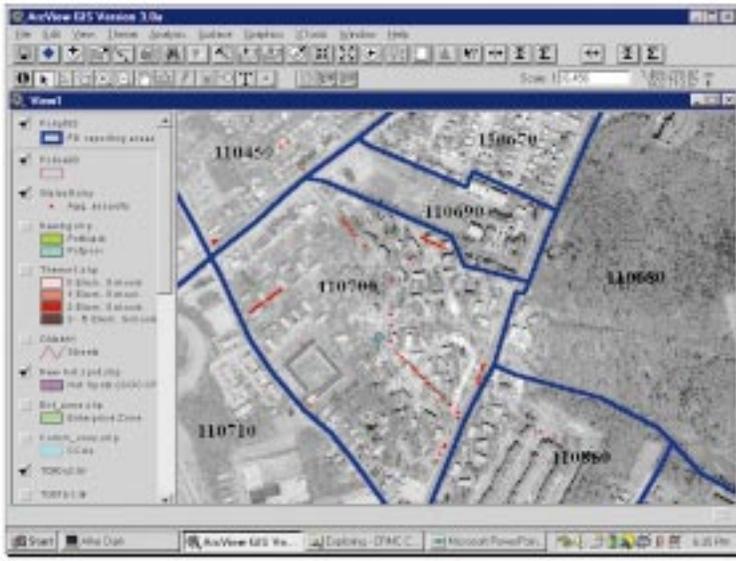
Summing up raster-vector differences, Clarke (1997) cited Bosworth's analogy about the work of composers Mozart and Beethoven. Raster is Mozart (dainty little steps), and vector is Beethoven (big jumps from place to place). Another way of characterizing the difference is to say, "vector systems produce pretty maps, while raster systems are more amenable to geostatistical analysis" (LeBeau, 1995). However, an alternative view argues that whether a vector or raster format is most useful depends on the type of analysis and what it will be used for. For example, crime densities are often calculated using the raster format, even when the point data for crime locations are in vector format. The vector data simply will be converted to the raster format.

It bears repeating that the analyst should know what format is used, why it is used, and the limitations and possibilities of each. A frequently encountered problem involves conversion to another format. Conversion from vector to raster is the simplest. Going from raster to vector, however, means that each line must be converted on a pixel-by-pixel basis and a vector equivalent produced. This is much more time consuming than vector to raster. Users may also convert from one software system to the other (Clarke, 1997).

Because photographic and satellite images are raster products, there is no choice between formats unless conversion is undertaken. As imagery of both types is used more frequently in crime mapping, we will see mixing of vector and raster technologies. For example, a large-scale aerial photo (raster) might be used as an underlay for point crime data (vector) and patrol area line files (vector). In most cases, the crime analyst will not be aware that different data formats are being used because the importation and manipulation will be seamless (figure 4.4). Although photos and satellite images are in raster format, they can be used to digitize data into vector format. Specific features on an aerial photo, such as the footprints of buildings or physical barriers between neighborhoods, are linear features amenable to vector representation. Aerial photos are often the source of other important base data, such as street centerlines, with data being digitally traced from the photo into a vector system.



Figure 4.4



A map showing a vector format superimposed over a raster format. ArcView®, Baltimore County, Maryland. The heavy blue lines are police reporting areas (vector), triangles are aggravated assaults (vector), and the underlay is orthophotography (raster).

Sources: Reporting area polygons and assault data: Baltimore County Police Department. Orthophotography: Baltimore County, Maryland, GIS Unit, March 1995, pixel size: 1 inch, compilation scale 1 inch=200 feet. Reproduced by permission.

Spatially enabling the data: What is geocoding?

If we break the word *geocoding* into its components, it means coding the Earth—providing geographic reference information that can be used for computer mapping. The history of geocoding is tied to efforts at the U.S. Census Bureau to find ways of mapping data gathered across the country, address by address. In the 1960 Census of Population and Housing, questionnaires were mailed to respondents and picked up from each household by enumerators. In 1970, the plan was to use the mail for both sending and returning surveys—hence references to that census as mail out/mail back. This demanded geocoding capability and, subsequently, the development of an address coding guide (ACG). According to Cooke (1998), the Data Access and Use Labs created to accomplish this were responsible for creating today's demographic analysis industry.

The first geocoding efforts permitted only street addresses to be digitized (admatch), but the capability to show blocks and census tracts was soon added. This demanded that block faces be recognized, and this was done by digitizing the nodes representing intersections. This, in turn, meant that intersections had to be numbered and address ranges had to be reconciled to the correct block faces. The shape of the lines on the map had to be precisely determined and annotated, creating the map's topology. The name given to this new block mapping process was *dual independent map encoding* (DIME) and, when combined with the address matching process, it was referred to as ACG/DIME. By 1980, ACG/DIME had become *geographic base file* (GBF)/DIME. This was followed by a call for a nationwide, seamless, digital map, to be called TIGER, short for topologically integrated geographic encoding and referencing. Census Bureau geographer Robert Marx and his team implemented TIGER for the 1990 Census (Cooke, 1998; Marx, 1986).



Mapping Crime and Geographic Information Systems

TIGER files contain address ranges rather than individual addresses. An address range refers to the first and last possible structure numbers along a block face, even though the physical structures may not exist (figure 4.5). For each chain of addresses between the start node and end node, there are two address ranges, one for odd numbers on the left, the other for even numbers on the right. For a complete explanation, see U.S. Census Bureau (1997).

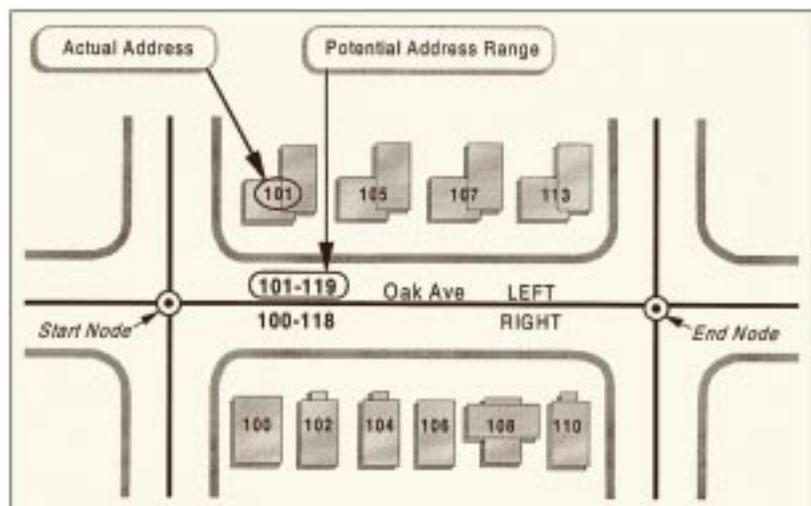
Geocoding is vitally important for crime mapping since it is the most commonly used way of getting crime or crime-related data into a GIS. Crime records almost always have street addresses or other locational attributes, and this information enables the link between the database and the map.

How does the computer map in a GIS know where the data points should be put? It reads the x-y coordinates representing their locations. When crime locations are geocoded, the address is represented by x-y coordinates, usually either in latitude and longitude decimal degrees or in State-plane

x-y coordinates identified by feet or meter measurements from a specific origin. The big headache in working with address data is that those data are often ambiguous and may be erroneously entered in field settings. Common field errors include:

- Giving a street the wrong directional identifier, such as using east instead of west or north instead of south.
- Giving a street the wrong suffix or street type (e.g., “avenue” instead of “boulevard”); providing no suffix when there should be one.
- Using an abbreviation the streets database may not recognize (e.g., St., Ave., Av., or Blvd.).
- Misspelling the street name.
- Providing an out-of-range, or impossible, address. For example, a street is numbered 100 to 30000, but an extra zero is added, accidentally producing the out-of-range number 300000.
- Omitting the address altogether.

Figure 4.5
A map showing TIGER/Line® address range basics.
Source: U.S. Census Bureau, 1997, figure 3-1, pp. 3-9.



As you initially attempt automatic geocoding, street addresses are compared against the existing street file database, and coordinates are assigned to the “hits.” This process is sometimes called *batch matching*. The process is a one-time affair, done automatically. Then, it becomes necessary to deal with the “misses,” those addresses that did not geocode automatically.

Handling misses is done manually. The bad address is displayed with the closest possible matches the database includes. Analysts use these options to select the most likely match. This involves some guesswork and risks geocoding errors. For example, if the address entered is 6256 Pershing Street, and the only reference to Pershing in the database is to Pershing Avenue, then assigning the geocode to “avenue” is not likely to be an error. On the other hand, if the database also contains Pershing Boulevard, Pershing Circle, and other Pershing suffixes, assigning “avenue” could be wrong. This shows how important it is to have standards for entering addresses into a file, whether the system deals with records or computer-assisted dispatching (CAD).

Not all records in large data sets are likely to be successfully geocoded. The title of a section in a chapter in the *MapInfo Professional User's Guide* (MapInfo Corporation, 1995), “Troubleshooting: Approaching the 100% Hit Rate,” hints at this. Some records may not be salvageable for a variety of reasons, including ambiguity in an address that cannot be resolved. Two other issues deserve mention, as well. One is that street addresses are estimated along block faces and may not represent true block face locations. (For more on this, consult technical docu-

mentation.) Second, address matching can be done for locations other than street addresses, such as street centerlines, land parcels, or buildings, depending on the availability of each element in a spatially enabled format.

Surprisingly, there is no minimum standard for geocoding. Maps can be produced and distributed based on a 25-percent *hit rate*. Readers may have no idea that a map represents only a small fraction of all cases. Worse, the missing cases may not be randomly distributed, thus possibly concealing a critical part of the database. For example, in the geocoding process, a person or persons may be inept or may decide to distort the data. If this error originates in the field, it will probably have a geographic bias based on the location of the person making errors. Analysts may consider reporting the hit rate for geocoding to better inform map readers.

Although most map users may not understand the hit rate, a technical footnote reading, “X percent of cases were omitted due to technical problems, but, the police department considers the pattern shown to be representative of the total cases under consideration,” may clarify the information. (Seek legal advice for actual wording.)

Given that there is no minimum standard, the issue becomes: What hit rate is acceptable? This is a subjective decision, but a 60-percent hit rate is unacceptable and may lead to false assumptions. Hit rates this low should raise questions about a crime analysis unit's level of readiness because low hit rates indicate that the base maps in use and/or incoming data are seriously deficient.



A distinction needs to be made between the hit rate and another geocoding measure, the *match score*. The latter is a score derived from matches on each component of the address. If all components of an address are correct—street name, direction, street type—the address will receive a perfect score. Missing or incorrect parts reduce the score. This differs from the hit rate, which is the percentage of all addresses that are capable of being geocoded in either batch or manual mode. Therefore, the hit rate and match score can be used to set acceptable geocoding standards. However, setting the acceptable threshold of either rate too high or too low may result in too few records making the cut or, in the worst case scenario, incidents being given wrong addresses, thus placing crimes on the map where they did not happen.

Like some other aspects of computer mapping, geocoding can be quite involved and demand considerable practice and expertise before you can regard yourself as an expert. The technical procedures used to fix geocoding problems are beyond this document's scope. Readers are referred to the user's guides, online help, and reference guides that accompany software or that are available on a proprietary basis. Asking more experienced GIS users for advice, perhaps in other departments of local government (management information systems, planning, engineering, and so forth), is another possibility. For additional information, see Block (1995).

Data selecting, filtering, and mapping according to useful criteria

The great power of GIS lies in its analytical capability, in addition to its capacity to quickly create maps of large, complex data sets. This analytical power takes more forms than can be described here, but a few examples offer the reader the general idea. (For an excellent overview, see McEwen and Taxman, 1995.) Since procedures vary according to the GIS software used, we offer a broad sketch of some possible analytical approaches. Specific details can be found in user's guides, reference manuals, tutorials, and listservs such as *crimemap@aspensys.com*, where many analysts pose questions and get useful answers on a variety of crime mapping topics, including how-to GIS questions.

Selecting and displaying specific information

Perhaps the most basic analytical task in using geographic information systems is the process of selecting and displaying scientific information. As shown in figure 4.6, when objects, in this case aggravated assault cases, are selected or "highlighted" on the map with the select tool, their corresponding database records are highlighted in color or with a symbol located next to the record. If the analyst wants to bring together the selected records at the top of the table for easier recognition and manipulation, they can be "promoted" (in ArcView, for example) using the appropriate button.



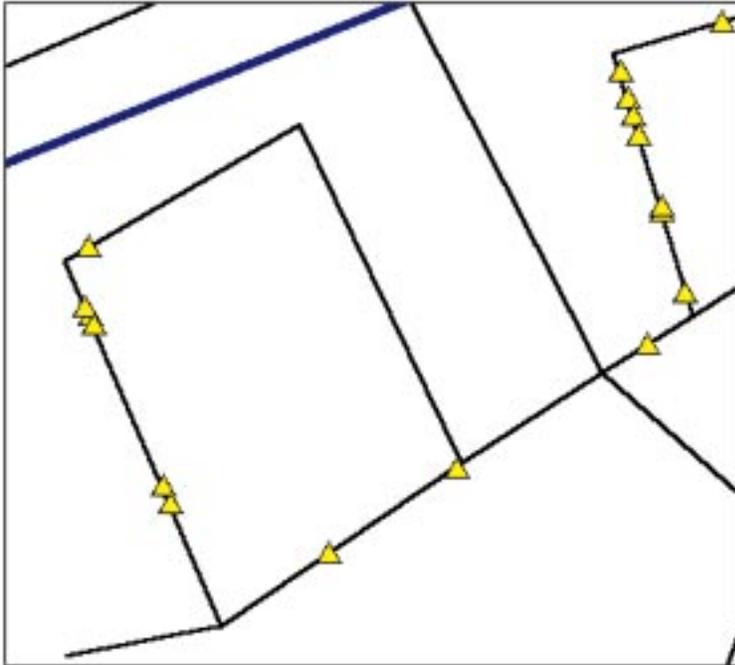


Figure 4.6

A map and table frames from ArcView® showing the relationship between objects selected on the map and their highlighted records in the table.

Feat	Loc_cod	Date_gnd	Date_wcd	Time_wcd	Dist	Vic_cn	Veh_cn	Att
0570	0450	19961210	19961210	2043	20	1	0	2
0570	0450	19961207	19961207	0140	10	1	0	
0570	0450	19961207	19961207	0140	10	1	0	
0570	0450	19961207	19961207	0140	10	1	0	1
0570	0420	19961206	19961206	1455	20	1	1	1
0570	0450	19961204	19961204	1924	20	2	0	1
0570	0450	19961122	19961122	1400	30	1	0	
0570	0450	19961115	19961115	1730	20	1	0	3
0570	0430	19961107	19961107	1730	20	1	0	1
0570	0450	19961107	19961107	1730	20	1	0	1
0570	0450	19961103	19961103	1135	20	2	0	1
0570	0450	19961031	19961028	0220	20	1	0	3
0570	0450	19961027	19961027	1545	30	2	0	
0570	0410	19961005	19961005	1822	10	1	0	
0570	0450	19960924	19960924	1602	10	2	0	
0570	0450	19960910	19960910	1215	10	1	0	1
0570	0450	19960828	19960828	1540	60	1	0	

Source: Keith Harries.

Mapping time

A more sophisticated process for selecting information uses various criteria and is referred to as “filtering” or “querying.” Sometimes, you will see Structured Query Language (SQL), which is a specialized programming tool for asking questions of databases. If the conditions you set are satisfied, then certain cases are included

or retained in a new data set, as a subset of the main data. The new file can then be saved, mapped, or manipulated in any way. For example, the condition “Time is greater than or equal to 1500 hours and time is less than or equal to 2300 hours” (written in computerese as *time* >= 1500 & *time* <= 2300) would select all cases in the 8-hour shift between 3 and 11 p.m.



Many variations are possible on the “mapping time” theme. In figure 4.7, maps of domestic disputes in Charlotte-Mecklenburg, North Carolina, show change over a decade. The units of analysis were the central points, or *centroids* (see the section “Centroid display” later in this chapter) of 537 response areas. Idrisi software (see appendix) was used to generate the two surfaces based on the square roots¹ of the data values for each response area, resulting in a clear picture of substantial growth over the time period. In 1984, domestic disputes occurred mainly in the north and west, as shown in green. By 1993, calls had increased in number and geographic coverage, particularly on the east side and in the southwest.

Time geography is viewed through a different lens as shown in figure 4.8. Using methods similar to those used to produce figure 4.7, seven daily maps were constructed, followed by an eighth map to denote the day of the week with the highest frequency of calls for each of the 537

response areas. Maps such as those in figures 4.7 and 4.8 could be used to allocate resources and to coordinate domestic violence prevention efforts.

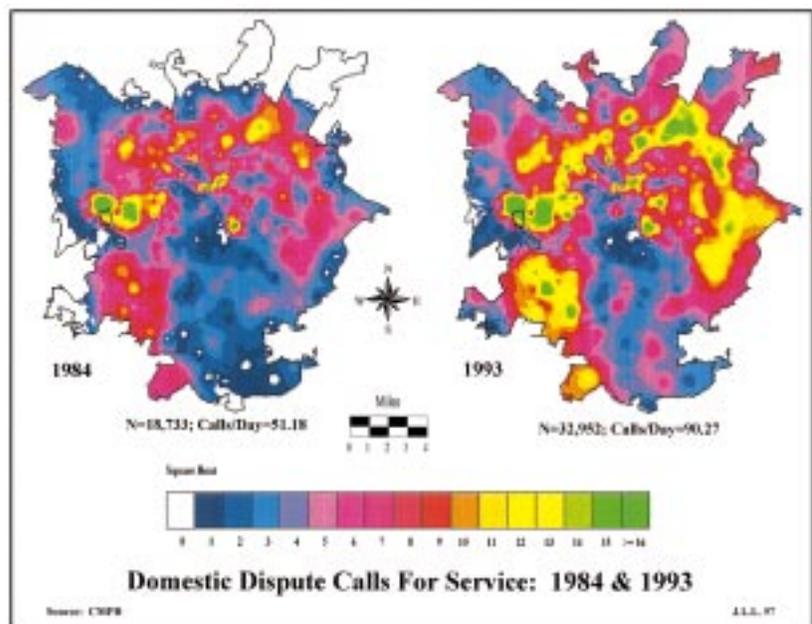
Mapping space

Because crimes usually affect some neighborhoods more than others, maps may focus on certain beats, posts, patrol areas, communities, census tracts, neighborhoods, or other units. What is the geography of crimes in terms of council districts? Could this information be used by the police department to anticipate political firestorms? Attention may not be confined to such official areas, but may involve informal or ad hoc areas, such as a 500-yard radius around a drug market, bus stop, or automatic teller machine, for temporary investigative purposes. Provided that the boundary files for the official areas are in the computer, queries can be addressed to them. For example, you could compute the rate of vandalism incidents per 1,000 housing units, per unit of

Figure 4.7

Maps showing domestic dispute calls for service, 1984 and 1993, Charlotte-Mecklenburg, North Carolina.

Source: Map by J. LeBeau, Southern Illinois University, with data from the Charlotte-Mecklenburg, North Carolina, Police Department. Reproduced by permission.



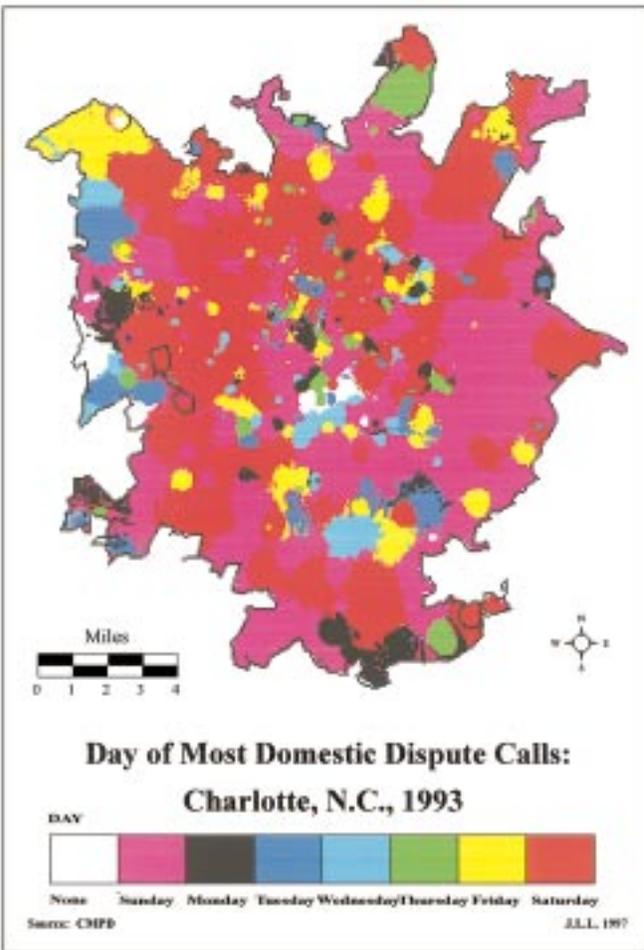


Figure 4.8

A map showing the day of the week with the most domestic dispute calls for service, Charlotte-Mecklenburg, North Carolina, 1993.

Source: Map by J. LeBeau, Southern Illinois University, with data from the Charlotte-Mecklenburg, North Carolina, Police Department. Reproduced by permission.

the general population, or per unit of the male youth population. A few alternative base maps are shown in figure 4.9.

Mapping incident types and modus operandi

Conditions, or filters, can be used to refine searches at any level the analyst chooses. For example, the most obvious filter would isolate all crimes of a specific type. However, filtering can isolate crimes by time of day, by neighborhood, and by modus operandi (MO). Conditions can be set to specify all the desired criteria, possibly resulting in the isolation of a cluster of incidents that could be linked to the same perpetrator. In figure 4.10, for example,

rape has been selected. These incidents are shown without identifying victims by using a large symbol to make the location of each somewhat vague. The analyst must trade some precision to accommodate the overriding need to protect victims.

Mapping attributes of victims and suspects

Mapping by characteristics of victims, suspects, or both can also be useful and is easily accomplished. For example, where have females been assaulted? Is there evidence that a cluster of burglaries has occurred at homes occupied by elderly persons?



Figure 4.9
Some alternate base maps for use in crime mapping.
Source: Keith Harries.

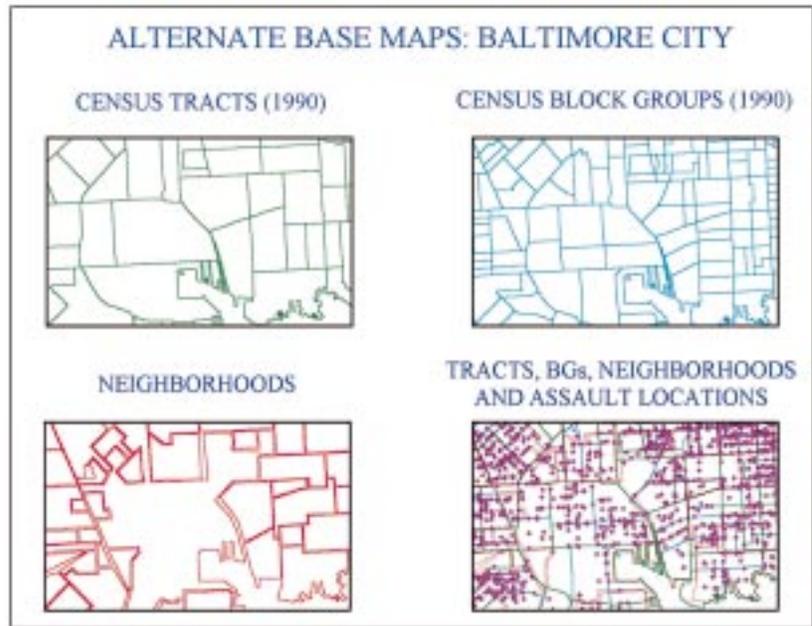


Figure 4.10
A map showing the distribution of rape in the city of Cambridge, Massachusetts. Large symbols were used to avoid presenting the specific locations of the victims.
Source: Cambridge, Massachusetts, Police Department. Reproduced by permission.



Mapping other recorded characteristics

Possibilities for filtering are unlimited. You can set as many time, space, victim, suspect, MO, or other filters as you wish, given data availability. For example, where are the armed robberies occurring? What is the pattern of robberies at gunpoint within a 1-mile radius of drug markets? Where are juvenile offenders

victimizing elderly persons at gunpoint during hours of darkness? Are spousal assaults randomly distributed, or are they clustered?

The only limitation is the availability of geocoded or geocodable information in the database. A potentially useful map might reflect the relationship between persons on probation or parole and the types of



crimes they have committed. For example, a rash of robberies occurs in a neighborhood: Where are the robbery probationers and parolees in that area? Do the MOs match up?

Using GIS to measure from maps: Aggregating data

Why measure? In the most general sense, measurement is the foundation of scientific analysis, and it lies behind any quantitative analytical statement. For example, what is the crime rate? To answer this we have to know the base of the rate. Do we want it per 1,000 persons, per reporting area, or per patrol district? To calculate this rate we must know how many crime incidents have occurred, and, if we are calculating a population-based rate, how many persons there are per unit area. This value, the base of our rate, is also known as the denominator, because it is the bottom of the fraction used to calculate the rate. Therefore:

$$\text{density} = \frac{\text{number of incidents per area}}{\text{population per area}}$$

Here, “number of incidents per area” is the numerator (top of the fraction) and “population per area” is the denominator. If there are 428 incidents and the population expressed in thousands is 3.7, the rate is $428/3.7$, or 115.68 per 1,000 persons. We can check that the calculation is correct by multiplying the rate by the population to reproduce the original incident count:

$$115.68 \times 3.7 = 428.$$

A GIS program would do these calculations for you, but analysts need to know how to provide appropriate instructions before anything useful can be produced.

An application of density analysis is shown in figures 4.11, 4.12, and 4.13. First, the density of burglar alarm calls was mapped using 48,622 locations for alarm calls in 1990 in Charlotte-Mecklenburg, North Carolina. Using the ArcView Spatial Analyst extension, a grid was used to generate the surface shown in figure 4.11. Peaks occurred near the central business district, along major transportation arteries, and in the industrial northwest area. A similar map (see figure 4.12) was prepared to show the density of the 10,288 burglaries reported in 1990, focusing on both the central business district and a radial, highway-oriented distribution. In the final phase, the burglary density surface was subtracted from the alarm density surface, and a query tool was used to select parts of the surface where burglary density exceeded alarm call density (see figure 4.13). This map directs the analyst to areas where displacement may be occurring and suggests areas for possible interventions. Such interventions could include additional alarm installations or other target-hardening measures (see figures 3.11–3.15).

The various types of measurement now available in GIS programs are too numerous to describe. However, a few types of measurement will be outlined to provide a sense of what can be done.

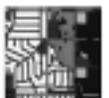


Figure 4.11

A map showing the density of burglar alarm calls, Charlotte-Mecklenburg, North Carolina, 1990.

Source: Map by J. LeBeau, Southern Illinois University, with data from the Charlotte-Mecklenburg, North Carolina, Police Department. Reproduced by permission.

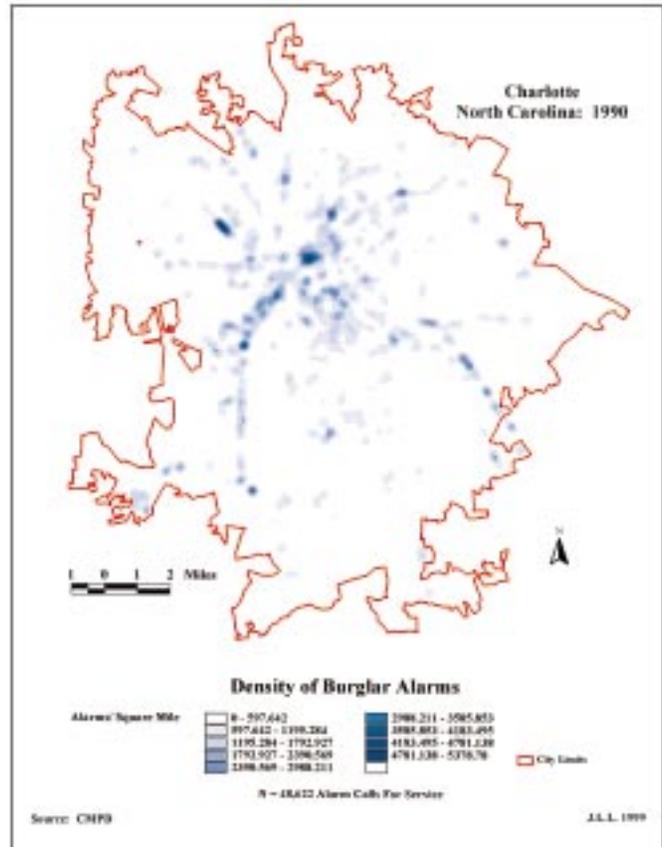
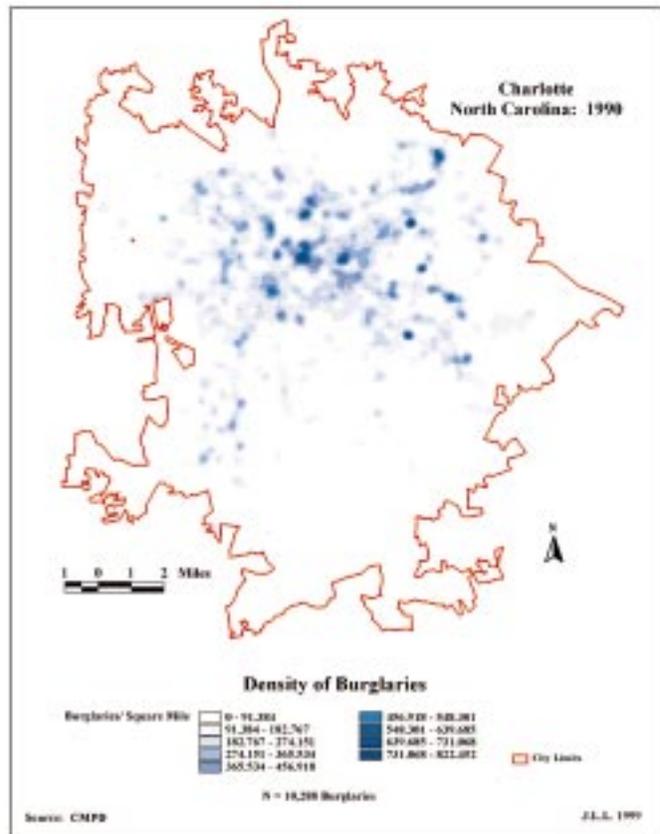


Figure 4.12

A map showing the density of burglaries, Charlotte-Mecklenburg, North Carolina, 1990.

Source: Map by J. LeBeau, Southern Illinois University, with data from the Charlotte-Mecklenburg, North Carolina, Police Department. Reproduced by permission.



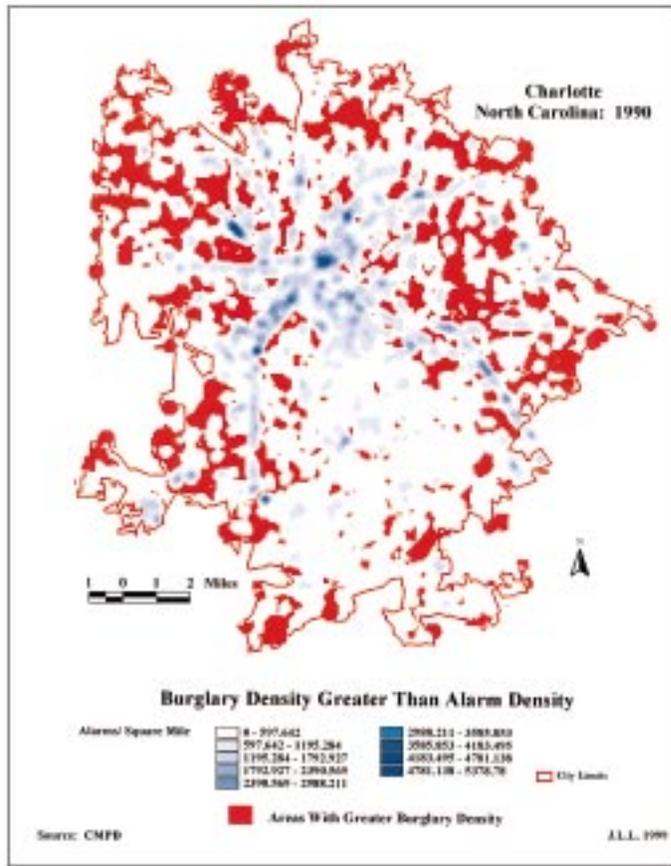


Figure 4.13

A map showing areas with burglary density greater than alarm call density, Charlotte-Mecklenburg, North Carolina, 1990.

Source: Map by J. LeBeau, Southern Illinois University, with data from the Charlotte-Mecklenburg, North Carolina, Police Department. Reproduced by permission.

Count incidents in areas

Although the primary need for counting will be to total crime incidents, counting other objects or events could be useful, too. It may be helpful, for example, to know how many and what types of alcoholic beverage licenses (beer, liquor, restaurant, liquor store, and so forth) are in specific areas. The information could serve as a gross index of alcohol availability, although it would not necessarily indicate where or how much alcohol is consumed or by whom. Or perhaps the local building inspection agency supplies a list of addresses of code violations. These could be mapped and sorted by any relevant areas, such as neighborhood association jurisdictions or police operations areas.

GIS software makes such identification easy. Users may find it helpful to write down the operation they want to do to clarify the steps—especially if they anticipate several steps of filtering, measuring, or other manipulations. In fact, this exercise is helpful for any kind of analysis. For example, you might want to count incidents of spousal abuse within patrol districts, which in some programs might be expressed in SQL something like this:

```
count spousal abuse.object
within patrol district.object
```

This tells the program to evaluate each spousal abuse incident (“object”) and determine in which patrol district it occurred. Additional instructions may ask the program to group incidents by patrol



district by listing in a new table or file each patrol district identification number and the number of spousal abuse incidents that occurred there.

Measure areas and distances

Area measurements are especially useful for determining how many crimes occur per unit area. This is not to be confused with crime measurements by population size or density. Generally, though not necessarily, crime densities will reflect population densities because population density is an expression of crime potential. More people means more potential victims and offenders.

Distance measurements are also simple. They require the use of ruler or tape software tools and, as with area measurements, the units can easily be changed.

Measure inclusion and overlap

Areas of interest in policing do not always fit together neatly. Police districts, precincts, patrol areas, and so forth, may not match school districts, council districts, census tracts, neighborhoods, community conservation districts, and officially designated hot spots. GIS tools allow users to measure overlaps between areas or small enclaves in large areas because any incidents found in a specific area can be electronically identified. All the crimes (or drug markets, liquor licenses, parolee addresses, injury accidents, and so forth) in a specific area can be selected and separated as a new data set for special analysis. How are drug arrests divided among council districts or neigh-

borhood association areas? One GIS package, for example, includes the following functions: *contains*, *contains entire*, *within*, *entirely within*, *intersects*. These capabilities are typical of this type of analytical function.

Centroid display

A *centroid* is an area's center defined as the halfway point on its east-west and north-south boundaries. However, the centroid will not always be inside the area. For example, an area may be L-shaped, in which case the centroid theoretically would fall outside the area. A centroid is generally used as the point where labels will be located by default and where statistical symbols will be placed.

Centroids can approximate the geographic midpoints of areas, which may in turn (theoretically) approximate the most accessible points in the areas. Normally, centroids are hidden, but they can be displayed on request. If you have area objects without centroids, the centroid (x,y) function can be used to generate them. Centroids are used infrequently in crime analysis and are typically used as surrogates for other conditions, such as accessibility.

For surface-fitting purposes, the data values that apply to tracts, block groups, or blocks could be assigned to their centroids, thus reducing areas to points for computational convenience. Given the common use of grid-based, surface-fitting algorithms, however, this type of centroid application is unlikely.



Derivative measures: How to create new indicators

Derivative measures are new variables created by manipulating information in one or more databases. The rate calculations discussed in the section “Using GIS to measure from maps” are a simple form of derivative measures that divide a crime count by a population measurement to produce a population-based rate. Generally, if you can express a sought-after relationship using ordinary mathematical logic, then it can be calculated in a GIS. There you will typically find an array of *operators* (+, -, =, *, “like,” “contains,” “and,” “or,” “not,” and so forth), *aggregates* (average, count, minimum, maximum, sum, and so forth), and *functions* (area, centroid, distance, perimeter, day, year, and so forth). These provide substantial versatility in general analysis or in creating derived measures.

Getting a bit more fancy

The complexity of potential derivative measures is unlimited. For example, you might want to create a quality-of-life measure for community areas that goes beyond mapping income or real estate values. This index could include such variables as crime rate, education levels, dropout rate, drug addiction measures, and incivility reports. Provided the underlying data can be geocoded, or joined to a geocoded table, they can be mapped. Then most, if not all, mathematical manipulations can be done in the GIS.

Apples and oranges

How do you combine variables measured in different units, such as dollars, years, or population? The quickest approach is to combine data in the GIS using overlays and then to use “logical operators” such as “greater than” or “less than” to reselect groups using your criteria. A more in-depth process, but one that leads to greater familiarity with the data, involves converting the data values into ranks (*ordinal scale* measurement) before making any calculations. Although you will lose the *ratio level* measurement this way, you gain the overwhelming advantage of being able to work with any units of measurement. Another advantage of this process is that conversion to ranks smooths the effects of poorly measured data by intentionally making them less precise. The conversion eliminates some of the “phony” precision of data that are inherently subject to error.

A problem that can generally be overcome is that GIS software is typically weak in statistical (as compared with purely mathematical) tools and may not be able to convert data to an ordinal scale.² This could be done by using a statistical software package such as S+, SAS, or SPSS. Or if the number of measures and areas is not too large, the work could be done manually. Then ranks are summed to generate the index, after taking care to organize ranks so the lower numbers always represent either the best or the worst, but not a mix of both. The resulting numbers will indicate a cumulative rank, or relative status, that can be mapped (figure 4.14). Crime data can then be overlaid on the index map to show a possible relationship with, for example, social dysfunction. For

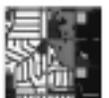
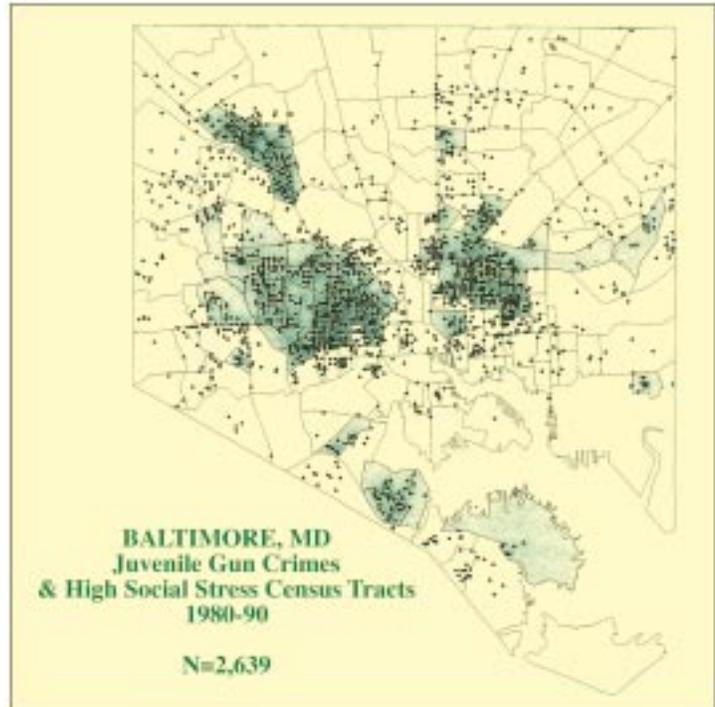


Figure 4.14

A map based on an Urban Pathology Index.

Source: Harries and Powell, 1994, figure 2, p. 54. Adapted by permission.



additional explanation, see Harries and Powell (1994).

GIS as a tool for data integration and exploration

A GIS is an ideal tool for bringing together various databases that share common geography. This function will become more useful as the importance of data integration is increasingly recognized. Not only is there a need for more data integration, but there is also a need for recognition that most data used in policing about land use, street centerlines, liquor establishments, bus routes, schools, subway stops, and so forth are likely to come from sources outside the police department. Finding these types of data and adapting them for crime analysis often take considerable initiative and may also demand attention to data quality. This raises the

issue of *metadata*. This term refers to data about data. Metadata provide information about the databases that you use. (See “Minimum for Federal Geographic Data Committee-Compliant Metadata.”) Metadata standards are developed under the auspices of the Federal Geographic Data Committee (FGDC), a unit that coordinates development of the National Spatial Data Infrastructure (NSDI). (See the appendix for additional resources.)

All data with common geography can be overlaid. These layers may be manipulated—moved up or down, added or removed (permanently or temporarily), or made to become visible or invisible only when the map is shown at a specified scale. As noted on page 92, “What Is a GIS?,” a fundamental concept in GIS is *layering*. The various forms of this process can provide a GIS with much of its power and flexibility.



Minimum for Federal Geographic Data Committee-Compliant Metadata**Identification_Information:**

Citation:

Citation_Information:

Originator: This is the originator

Publication_Date: 000000

Title: This is the title

Description:

Abstract:

A concise abstract of the data

Purpose:

(Types of projects that use the data, models, general use, illustrations, specific type of analysis, and so forth. Be specific if the data are geared toward a narrow set of applications.)

Time_Period_of_Content:

Time_Period_Information:

Single_Date/Time:

Calendar_Date: 000000

Currentness_Reference: Publication date of sources

Status:

Progress: Complete

Maintenance_and_Update_Frequency: Unknown

Spatial_Domain:

Bounding_Coordinates:

West_Bounding_Coordinate: -109.9959

East_Bounding_Coordinate: -108.0022

North_Bounding_Coordinate: 37.0671

South_Bounding_Coordinate: 35.9330

Keywords:

Theme:

Theme_Keyword_Thesaurus: none

Theme_Keyword: None

Access_Constraints: none

Use_Constraints:

Limitations of data for use at certain scales and date ranges and for use with other data

Metadata_Reference_Information:

Metadata_Date: 19950628

Metadata_Contact:

Contact_Information:

Contact_Organization_Primary:

Contact_Organization: Bureau of Land Management

Contact_Address:

Address_Type: mailing address

Address: PO Box 0047 DWO

City: Denver

State_or_Province: CO

Postal_Code: 80225-0047

Contact_Voice_Telephone: 555-5555

Metadata_Standard_Name: FGDC Content Standards for Digital Geospatial Metadata

Metadata_Standard_Version: 19940608

Source: <http://www.blm.gov/gis/meta/minimum.html>

Combining data in geographic space provides opportunities for data exploration and analysis that are lacking when geographic data are missing. An analyst may want to see how robbery locations relate to the locations of convenience stores. Although this information may be in different databases, it can be brought together in GIS and the locations subjected to the necessary analysis. For example, buffer zones could be constructed at a specified distance around each convenience store, and the number of robberies in each zone counted. Then the percentage of robberies proximal to stores could be calculated to provide an indication of the importance of this type of store as a robbery target. The possibilities offered by this type of *spatial analysis* are virtually unlimited. They include hot spot analysis, stolen auto recovery directions and distances, delineations of gang turfs, calculations of area-specific rates, the construction of crime or other “surfaces,” network analyses, boundary determinations, and others mentioned elsewhere.

Hot spots

Definition in geographic space

The term *hot spot* has become part of the crime analysis lexicon and has received a lot of attention. What are hot spots? How do we recognize them?

A hot spot is a condition indicating some form of clustering in a spatial distribution. However, not all clusters are hot spots because the environments that help generate crime—the places where people are—also tend to be clusters. So any

definition of hot spots has to be qualified. Sherman (1995) defined hot spots “as small places in which the occurrence of crime is so frequent that it is highly predictable, at least over a 1-year period.” According to Sherman, crime is approximately six times more concentrated among places than it is among individuals, hence the importance of asking “wheredunit” as well as “whodunit.” (See the appendix for hot spot-related resources.)

A great deal of confusion surrounds the hot spot issue, including the distinction between *spaces* and *places*. Block and Block (1995) pointed out that a place could be a point (such as a building or a classroom) or an area (such as a census tract or a metropolitan region). However, the former generally are regarded as places, and the latter, with their greater area, are spaces.

Sherman’s definition notwithstanding, there is currently no widely accepted definition of a hot spot. Indeed, a rigid, absolute definition may not be possible. Except for programs with procedures that self-define hot spots, such as the Spatial and Temporal Analysis of Crime (STAC) program (Block, 1995), jurisdiction-specific procedures to define hot spots may make the most sense because they will fit local conditions. In Baltimore County, Maryland, for example, hot spots are identified according to three criteria: frequency, geography, and time. At least two crimes of the same type must be present. The area is small, and the timeframe is a 1- to 2-week period. Hot spots are monitored by analysts until they become inactive (Canter, 1997).



In many cases, analysts may not be able to define hot spots but may know one when they see it. This makes comparisons difficult both within and between jurisdictions.³ Furthermore, meaningful time-based analyses are problematic, because hot spot definition criteria may not be used consistently over time.

Wide interjurisdictional and intrajurisdictional variations in environments also make the application of absolute definition criteria tricky. For example, the size and shape of city blocks vary widely. West of the Appalachian Mountains, city layouts are usually dictated by the rectangular land survey system, and blocks tend to be fairly regular and rectangular. In the east, where metes-and-bounds surveys prevailed, blocks are more likely to be irregular in shape and size. Densities also vary greatly. Can the same definition criteria be applied in low-density areas as in high-density areas? Crime-prone populations are found in both environments. Can hot spots exist in very low-density suburbs? Residents would probably think so.

Hot spots and scale

Are hot spots purely a function of scale? Some argue that any set of points in geographic space can be made into a hot spot if the scale is modified enough. At extremely small scale,⁴ all the crime incidents in an entire metropolitan area appear to be a hot spot (figure 4.15, upper left). As scale increases, points become more dispersed (figure 4.15, upper right and lower left) until, at the largest scale, individual points can be isolated (figure 4.15, lower right). The level of resolution in the absence of absolute criteria makes it possible to manipulate the presence or absence of hot spots. However, absolute criteria are difficult to apply in urban environments (Brantingham and Brantingham, 1995).

Generally, the hot spot concept is applied to street crime rather than white-collar crime, organized crime, or terrorist crime. That a few white-collar crimes might overwhelm street crime in their economic impact tends to be ignored. This may be

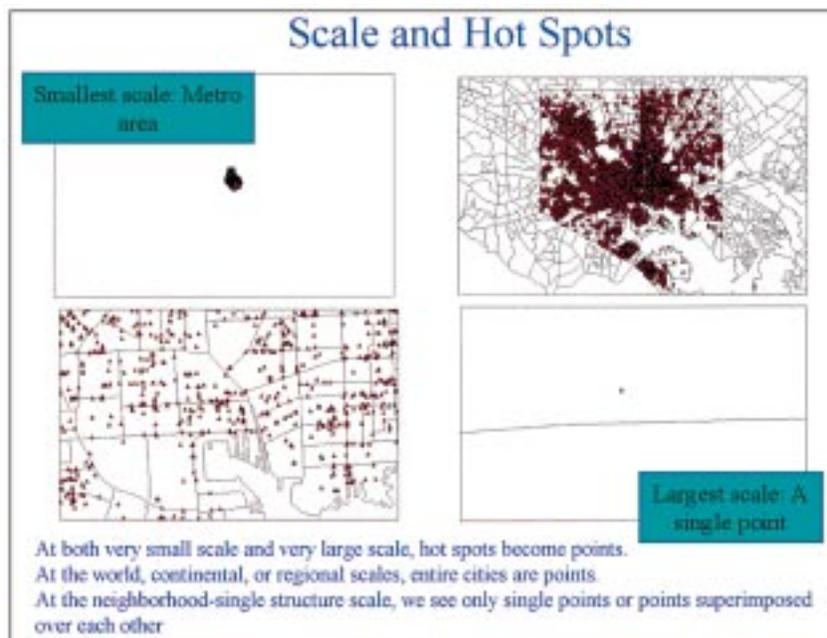


Figure 4.15

An illustration of the scale-dependency of hot spots.

Source: Keith Harries.



because white-collar crime does not cause the same kind of community fear and anxiety as street crime. Similarly, if a city experienced several terrorist bombings or school shootings within 1 year, it is considered a hot spot that defies the normal hot spot definition. There is a qualitative aspect to hot spots; they refer only to limited crime types.

Hot spots in time

Just as hot spots can be described geographically, they can also be defined using time-related criteria. An important question is: How long is a hot spot “hot”? The answer requires defining an incident accrual rate within the spot, based on units of time. Related decisions are needed to determine whether the hot spot’s “temperature” is measured according to all confirmed crimes, all calls for service, specific crimes, or other conditions. In a GIS framework, hot spots (and/or incidents within hot spots) can be color coded or otherwise symbolized according to their age.

An approach to monitoring hot spots over time is shown in figure 4.16. This map shows Devil’s Night arson hot spots in Detroit in 1994 and 1997, using the STAC program developed by the Illinois Criminal Justice Information Authority. Although not a mapping program itself, STAC can be used with most popular GIS packages. (For additional details on STAC, see the appendix and figure 4.20.)

Definition and measurement

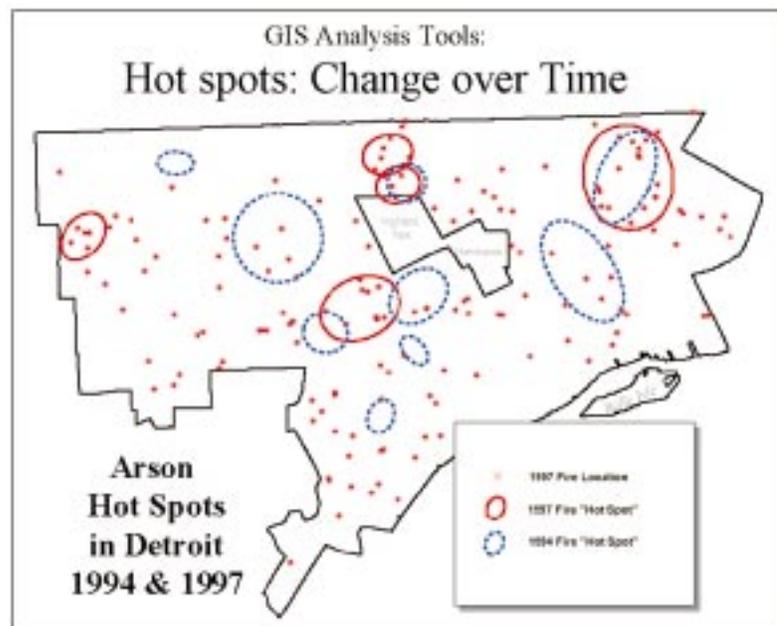
What is a hot spot? Perceptions and definitions vary widely. Some analysts may see a hot spot as any cluster that looks interesting. Others define hot spots using rigid, detailed criteria. A study by Buerger, Cohn, and Petrosino (1995) found that the latter group initially used the following relatively formal criteria:

- Not more than one standard linear street block (one side of the street only).

Figure 4.16

A map showing hot spot changes over time: Detroit, 1994 and 1997.

Source: Martin, Barnes, and Britt, 1998, figure 3, p.10. Reproduced by permission.



- Not more than half a block from an intersection.
- No closer to another hot spot than one block.

The Buerger group further identified three principal definition-related issues:

- **Public space.** Hot spots were initially limited to one side of the street, raising the question of how *street curtilage* (public space in front of private properties) would be treated. Common sense dictated that if a patrol car was across the street, technically outside the hot spot, it should be considered in the hot spot, so the definition was modified to include both sides of the street.
- **Intersections.** Ambiguities surrounded the definition of an intersection. The term eventually came to include not only the street, but also adjacent sidewalks and buildings. Even when a hot spot did not technically include all four corners of an intersection, it was found that the best viewpoint for seeing around a corner might be on the other side of the street, outside the hot spot. Thus, all four corners of intersections came to be included in hot spots.
- **One-block exceptions.** Irregular blocks with large open spaces contained some hot spots, making exceptions to the one-block rule.

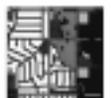
In practice, hot spots are defined in numerous ways, some with rigid criteria, like those above, and others with a more flexible approach. None is right or wrong. Both approaches have pros and cons, and an informal cost-benefit analysis can

determine the ideal criteria in individual locations. The sharply defined criteria may omit many commonsense exceptions (but allow greater comparability in space and time); softer rules permit easy adaptation to local variation (but make comparisons difficult).

Hot spot mapping

A detailed presentation of hot spot mapping methods is beyond the scope of this guide. However, an investigation sponsored by the Crime Mapping Research Center at the National Institute of Justice in 1998 can offer some tips. This assessment found that most hot spot analysis methods fall into one of five categories: *visual interpretation*, *choropleth mapping*, *grid cell analysis*, *cluster analysis*, and *spatial autocorrelation* (Jefferis, 1999; see also Canter, 1995).

- **Visual interpretation.** The survey showed that, of the police departments that do computer mapping, 77 percent conducted hot spot analyses. Of these, 86 percent identified hot spots visually, and 25 percent used a program to perform this task (Mamalian and La Vigne et al., 1999). The problems presented by the visual approach include overlapping points, points stacked on top of one another, making it impossible to see how many incidents are represented⁵ (that is, only one appears at any given location). Most serious, perhaps, is that readers' interpretations of point data vary, resulting in different interpretations of the same patterns.
- **Choropleth mapping.** Areas are shaded according to their data values, by either rate or frequency. The



caveats mentioned in chapter 2 still apply—class interval selection methods will affect the appearance and interpretation of the map (see figures 2.15 and 2.16), as will color choices, shading levels, and size variation among the polygons. The latter elements tend to draw attention to the largest areas, particularly when they have higher data values.

- **Grid cell analysis.** A grid is superimposed over a map. Points within cells, or within a designated radius from the centers of the cells, are assigned to the cells. The size of cells is variable and affects the outcome of the analysis. Small cells present higher resolution, at the cost of more computer resources. With larger cells, resolution suffers, but computation is easier. What is the advantage of grid cell analysis over a pin map? First, adding points to the grid solves the problem of “stacked” data points, which occurs when multiple incidents occur at the same location or nearby locations. Second, the points

are transformed into a smooth surface, generalizing the data. (For related methods, see the appendix: ArcView Spatial Analyst Extension, Idrisi, Vertical Mapper, the U.S. Department of Justice Criminal Division Hot Spot Slider.) Several examples of grid cell analysis have been illustrated in figures 4.11, 4.12, and 4.13. Another map of this type is shown in figure 4.17, which depicts hot spots in the United Kingdom city of Nottingham and police perceptions of hot spots. This map was produced using the custom program known as SPAM (Spatial Pattern Analysis Machine), which links to MapInfo for the finished map (see appendix). Variations of surface mapping include three-dimensional renditions, as noted in chapter 1. One key to readers’ perception of three-dimensional maps is the degree of vertical exaggeration in the map. In the examples shown in figures 4.18 and 4.19 of Salinas, California, quite different types of data (firearm crimes and gangs) are shown in three dimensions.

Figure 4.17

A map showing domestic burglary hot spots and police perception in the Meadows area of Nottingham, United Kingdom.

Source: Ratcliffe and McCullagh, 1998, figure 1, p. 48. Reproduced by permission.

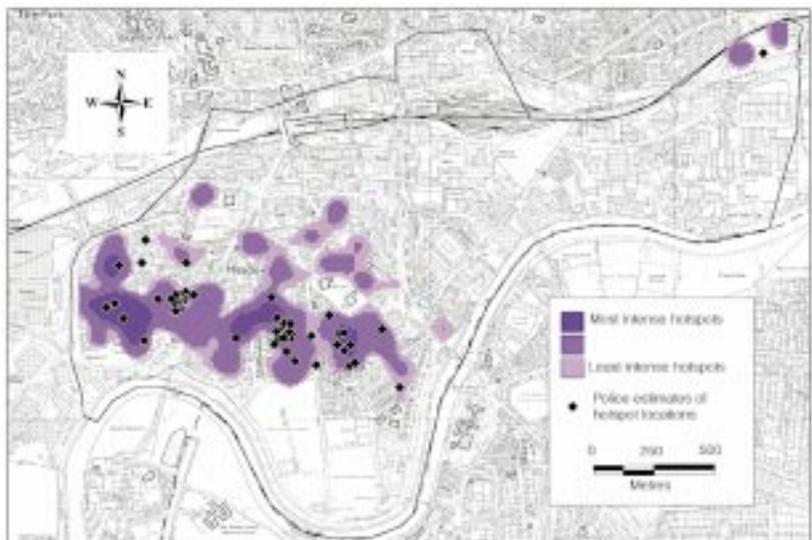




Figure 4.18

A map showing firearm incidents in Salinas, California.

Source: Salinas, California, Police Department. Reproduced by permission.

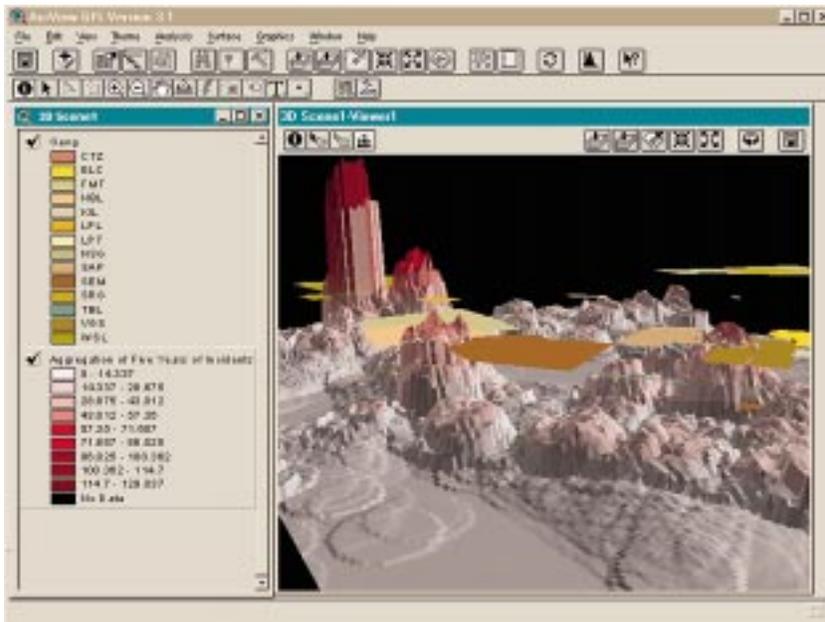


Figure 4.19

A 5-year analysis of incident data with an overlay of identified gang turfs in Salinas, California.

Source: Salinas, California, Police Department. Reproduced by permission.

■ **Cluster analysis.** Cluster analysis methods depend on the proximity of incident points. Typically, an arbitrary starting point (“seed”) is established. This seed point could be the center of the map. The program then finds the data point statistically farthest from there and makes that point the second seed, thus dividing the data points into two groups. Then distances from each seed to other points are repeatedly cal-

culated, and clusters based on new seeds are developed so that the sums of within-cluster distances are minimized. (For related methods, see the appendix: STAC, SaTScan, SpaceStat, and Geographic Analysis Machine (GAM.) Figure 4.20 illustrates hot spots derived from the STAC method, which performs the functions of radial search and identification of events concentrated in a given area (Levine, 1996).

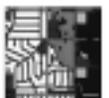
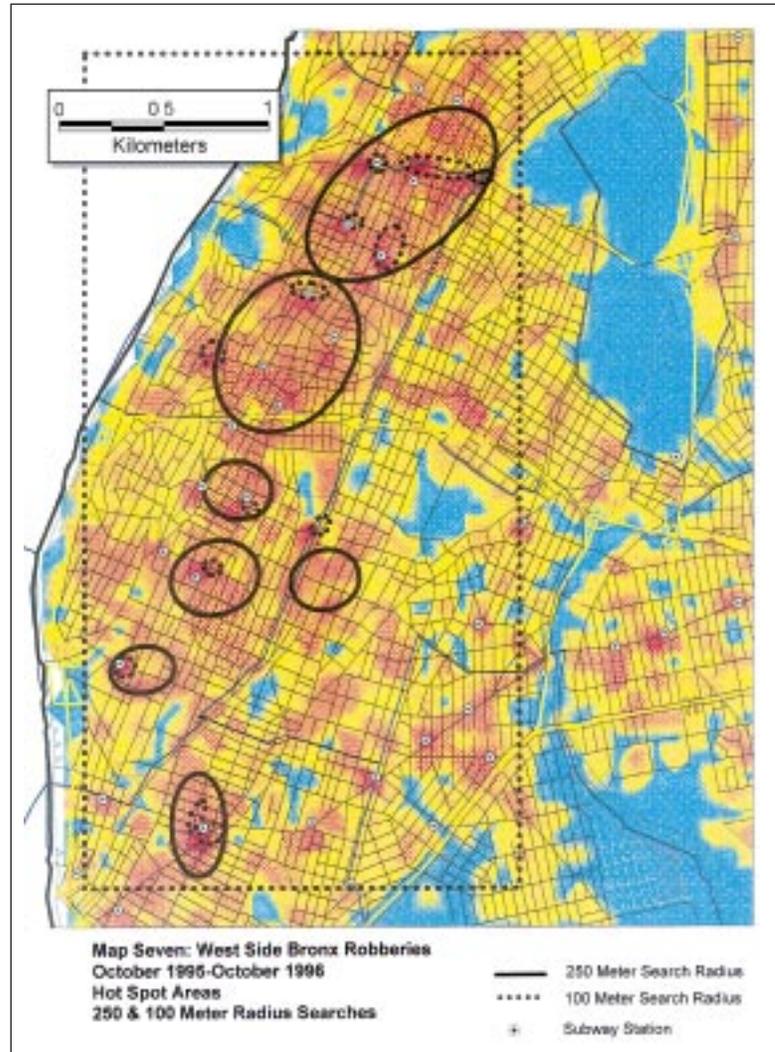


Figure 4.20

A map showing robbery hot spots, using STAC, West Bronx, New York.

Source: Block and Block, 1997, figure 7. Reproduced by permission.



■ **Spatial autocorrelation.** This concept relies on the idea that events that happen in different locations may be related. In a crime hot spot, for example, underlying social and environmental processes generate crimes in a small area. Multiple events, such as the presence of drug markets, may have similar causes. This means that statistical measures of this condition, known as autocorrelation, can serve as hot spot indicators (Roncek and Montgomery, 1995).

All methods of hot spot mapping should produce similar maps if there are underly-

ing and recognizable point clusters. Something is wrong if a method produces clusters where visual inspection indicates there are none. However, analysis should recognize that some methods involve user-defined search criteria, and variations in those criteria, such as differences in cell sizes or search radii, can affect outcomes.

Buffering: Meaning and applications

A *buffer* is a zone around an object, such as a school or intersection, that has some investigative or analytical significance. For



example, drug-free school zones may be defined using a 1,000-yard radius. Such buffers can be drawn around schools and overlaid on large-scale aerial photographs so that field officers can easily recognize the zone's boundaries, even without demarcating signs. Hardcopy maps can be given to patrol officers as an aid in recognizing the zones. Buffering tools in GIS programs make this a relatively simple task (figure 4.21).

Techniques for selecting objects can be used to identify certain types of events. For example, what are the characteristics of calls for service within 1 mile of high schools? Calls for service can be identified and separated into a new data set if they are within the 1-mile buffer.

Buffers are shown as circles if the location buffered is a point or street address, but buffers do not have to be circles. For large

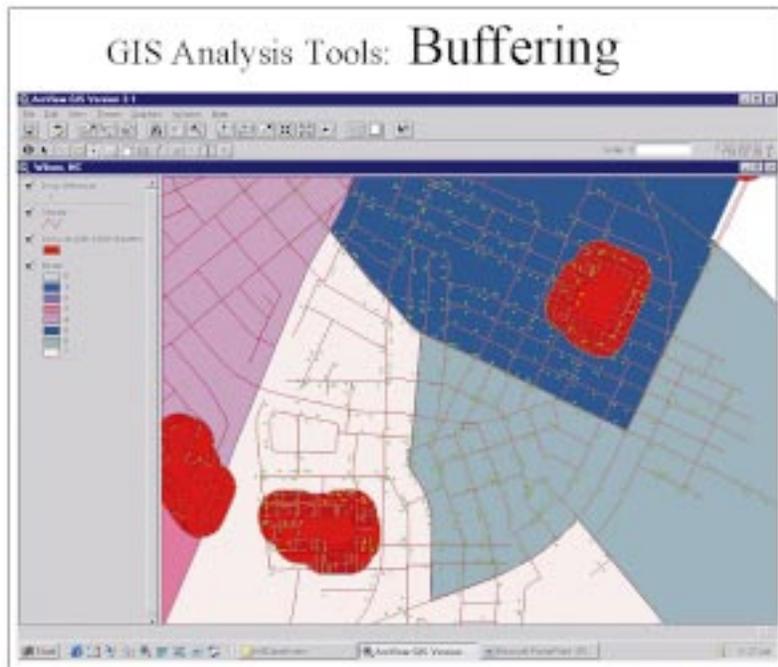


Figure 4.21

An example of buffering used in conjunction with the mapping of related data. Red areas represent 1,000-foot drug-free school zone buffers around schools, large polygons represent police beats, and point data represent drug offenses.

Sources: J. Stith and the Wilson, North Carolina, Police Department. Reproduced by permission.

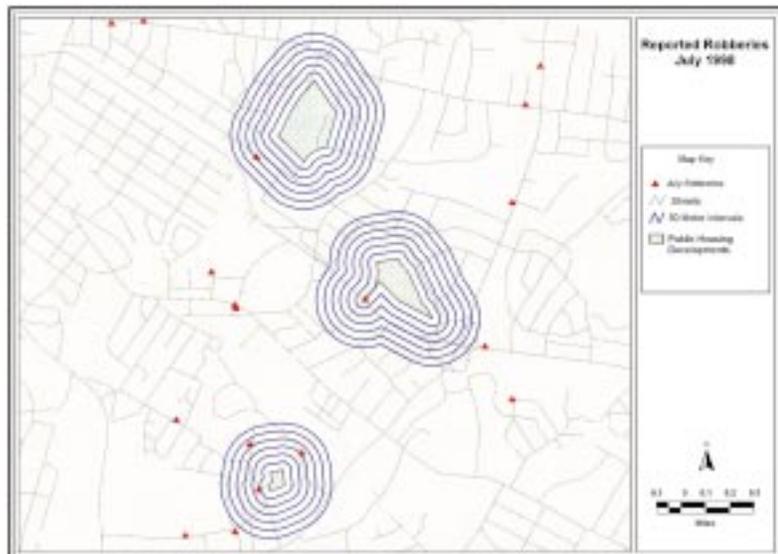
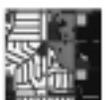


Figure 4.22

A map showing polygon buffers around public housing in Baltimore, Maryland.

Source: Hyatt and Holzman, 1999. Reproduced by permission.



polygons like school campuses, parks, apartment complexes, or industrial plants, buffers can mirror the shape of the polygon. In figure 4.22, public housing properties were buffered to evaluate the relationship between public housing and the surrounding neighborhood. The underlying question was whether crime in public housing was committed mostly by residents or by persons from the surrounding community. Analysis of incidents in the buffer zones can help determine the answer, using data on the residential addresses of victims and offenders. The same areas could be represented either by a circle buffer (if it is represented as a point or address) or as a polygon buffer (if the area is mapped to match its actual footprint).

In a community policing example, questions may arise about the quality of street or neighborhood lighting. Analysts can consult with city engineers to learn about the illuminated radius of various street-

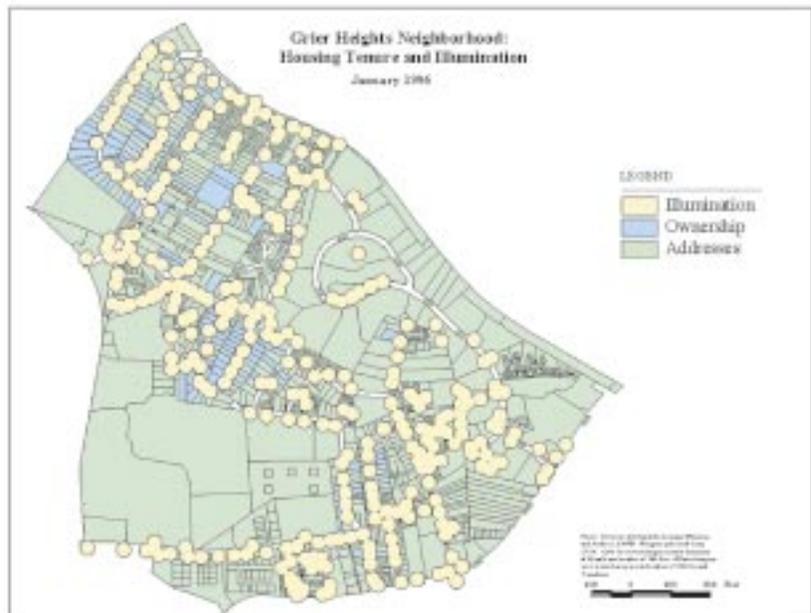
lights and their coordinates in the community. Then, using the buffer tool in GIS, circles of the appropriate radii can be drawn around each light location to create a basis for assessing community concerns about lighting quality (figure 4.23).

Also in the context of community policing, an extremely disruptive phenomenon is house fires. This is especially true in older neighborhoods with substandard row houses, where the likelihood of fire spreading from one home to adjacent houses is great. In one study, buffers were drawn around residences where fire-related injuries occurred in the previous 2 years. This, along with census data indicating risk, helped establish zones that were appropriate for the distribution of smoke alarms. (This may seem like a fire department function, but economic and social disruptions can contribute to conditions in which crime flourishes, making both fire and police functions relevant, as with arson cases. This helps demonstrate

Figure 4.23

The Grier Heights neighborhood in Charlotte, North Carolina, showing street light buffers and housing tenure in 1996.

Source: E. Groff and Charlotte-Mecklenberg Police Department. Reproduced by permission.



the breadth of data that crime analysts need access to.)

In Tornado Alley, the high-frequency tornado region in the Plains States, a community safety concern is the audibility of warning sirens. Because the decibel output of sirens is known and varies with weather conditions, a GIS can be used to map the audible zone of each siren. Areas where sirens cannot be heard can be targeted by public safety agencies for special action during tornado warnings. This type of map could also serve as a blueprint for locating new sirens.

Data, data, everywhere: What's an analyst to do?

Mapping outside data

Although most data used in crime analysis are generated and used within one department, the need to integrate information from other agencies is becoming more important. Unfortunately, outside data are often in an incompatible format. What do you do when you get a *delimited* or *fixed field ASCII* file, for example? Table 4.1, "Census data in ASCII," shows data almost exactly as they are presented on the U.S. Census Bureau Web site, with only slight editing to adjust the spacing. The Federal Information Processing Standard (FIPS) codes attached to all census geographic areas⁶ make up the first two columns. Starting in the left-hand column, the State of Missouri FIPS code is 29. In the next column, the city of St. Louis FIPS code is 510. Next are the tract numbers, literal tract numbers, tract populations (P00010001), tract median family

incomes in 1989 (P107A001), and tract per capita incomes in 1989 (P114A001).

Most GIS software can handle data compiled in a variety of formats, although some variations may generate headaches. MapInfo will open the following formats: dBase, Lotus®, Microsoft Excel®, delimited ASCII, some raster files (.tif, .pcx, and so forth.), AutoCAD® (.dxf), and others, using either the *Open* or *Import* command. ArcView expects tabular data to be in dBase (.dbf), Info, or delimited text (.txt) format. One solution to the somewhat limited data conversion repertoires of some GIS programs is to launder files through a more versatile spreadsheet program and then transfer them to the GIS in a more compatible format.

To do this, the user imports the foreign data spreadsheet into Lotus, Microsoft Excel, or Microsoft Access®, and then exports it in a GIS-compatible format. This way, a fixed-field ASCII file can be converted into a delimited ASCII or dBase format. This is done by parsing the fixed-field file in the conversion program, and then outputting it in a delimited format. Parsing is a process of instructing the program how to read the fixed-field data by identifying the variables in each field and dragging field delimiters to appropriate locations. For example, the analyst instructs the program that the case number is in columns 1–10, the address is in columns 11–30, and so forth. Delimited means that each data field is separated from the next by a character such as a comma or a tab. With delimiters, it does not matter if data values have different widths, as in the sequence 3.5, 14.276. When the program recognizes the



delimiter as a cue, it moves to the next value.

You may receive data that consist of x-y coordinates, without the points them-

selves. In such situations, the coordinates are used to generate the points in the GIS, using a Create Points command that allows users to select a preferred symbol and an appropriate projection. (For information

Table 4.1. Census Data in ASCII Tab-Delimited Format as Acquired Directly From U.S. Census Bureau Web Site

(<http://www.census.gov>)

See text for explanation

(Selected census tracts for St. Louis, MO)

FIPS. ST	FIPS. CO.	FIPS. TRACT90	STUB.GEO	P0010001	P107A001	P114A001
29	510	1011	Tract 1011	000002867	000032281	14262
29	510	1012	Tract 1012	000003287	000041559	15743
29	510	1013	Tract 1013	000004446	000036234	15401
29	510	1014	Tract 1014	000003075	000028322	11859
29	510	1015	Tract 1015	000003763	000021519	9787
29	510	1018	Tract 1018	000003773	000025197	9027
29	510	101899	Tract 1018.99	000000028	000000000	13929
29	510	1021	Tract 1021	000002820	000032625	15883
29	510	1022	Tract 1022	000006608	000040897	17131
29	510	1023	Tract 1023	000001834	000031157	14262
29	510	1024	Tract 1024	000002550	000034545	13226
29	510	1025	Tract 1025	000002253	000034911	13158
29	510	1031	Tract 1031	000003429	000039677	19250
29	510	1034	Tract 1034	000002200	000037222	13316
29	510	1036	Tract 1036	000001617	000033105	13240
29	510	1037	Tract 1037	000002883	000035461	14024
29	510	1038	Tract 1038	000004213	000039194	15466
29	510	1039	Tract 1039	000001229	000028375	11771
29	510	1041	Tract 1041	000003153	000030313	12649
29	510	1042	Tract 1042	000004076	000030660	14218



about map projections, see chapter 1.) After the points are generated, they can be imported as a new layer on the map. Data generated in the field, perhaps from patrol cars using global positioning system technology, can be treated in the same manner.

Any database with an address or geographic reference included can be mapped, provided the corresponding digital base map is available. For example, you may want to map census tract data. The census data are available, but the map of tracts is not. In this case, the map is readily available to download off the Web, but non-census tract maps must be acquired elsewhere.

Data warehousing and data mining

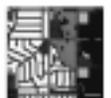
Police departments generate volumes of information. A single call for service ultimately results in its own pile of paper, and computer files tracking all calls for service grow rapidly. *Data warehousing* and *data mining* provide sophisticated ways of storing and accessing information.

A data warehouse is a megadatabase that stores data in a single place instead of storing them in project files or throughout the local government or government agency. Government agencies have been slow to do this because agency politics tend to create an attitude oriented more toward defending departmental turf than toward sharing data. A data warehouse could assist with crime analysis efforts, which often demand data from diverse sources, such as the health, housing, traffic, fire protection, liquor licensing, and planning agencies.

Law enforcement is primarily a local government activity, which often leaves police agencies at the mercy of data managers overseeing city or county information technology functions. Ideally, data warehouses consolidate all jurisdictional databases and permit use of data from any agency according to quality control standards. Data mining, as the label suggests, involves digging nuggets of information out of vast amounts of data with specialized tools. These tools are typically called exploratory data analysis (EDA), which, in the context of mapping, can become exploratory *spatial* data analysis (ESDA) tools. An IBM software engineer (Owen, 1998) identified these as the factors that brought data mining to the attention of the business community:

- The value of large databases in providing new insights is recognized.
- Records can be consolidated with a specific audience or objective in mind.
- Cost reductions are achieved with large-scale database operations.
- Analysis is being *demassified* (futurist Alvin Toffler's term) meaning that the information revolution permits the creation of specialized custom maps for specific audiences.

Chapter 2 discussed hypothesis testing. That discussion now comes full circle because data warehousing and data mining make hypothesis testing even more practical. Queries can be addressed to large arrays of data, increasing the reliability of responses. However, this is truer for historical questions than for current data.



Cautions

Although GIS makes mapmaking relatively easy, it is not necessarily easy to make good maps. The fundamental problem is that fancy fonts, tables, maps, or diagrams can dress up almost any data. However, just because data look good does not mean they are.

As a rule of thumb, do not blindly accept the default settings in GIS programs. Defaults apply to such steps as the selection of class intervals in choropleth maps (see “Classifying map information” in chapter 2), as well as the colors, symbol types, and sizes.

Rule of Thumb

Default settings in GIS programs should not be accepted blindly.

Boilerplate maps, produced regularly to show specific needs such as weekly precinct or division crime patterns, can be fine tuned so they consistently convey the intended message. Problems are more likely to arise with specialized maps. A checklist may be a useful reminder of the most important map elements and criteria:

- **Need.** Is a map needed for this message or analysis? Could the job be done as well or better with another approach, such as a table, a narrative, a chart, or conversation?
- **Data source.** Are the data reliable? If there are questions about data quality, how can the audience be alerted? (By using a map subtitle or map footnote?)

Are data so poor that mapping and analysis should not be done? Who decides if this is the case?

- **Scale.** Is the appropriate area shown? Can the map be enlarged without compromising the message?
- **Scope.** Is the map trying to show too much? Too little? Can more context be added to better inform the reader?
- **Symbols.** Would icons convey the message more convincingly than abstract default symbols? Note that some icons are awkward shapes and may have a minimum size, below which they become meaningless.
- **Color.** Misused color detracts from an otherwise excellent map. Excessive use of bright colors may hurt the eye and repel the reader. Illogical color gradations may be confusing. Think in terms of drawing the eye to important areas (normally higher data values) by using more intense color tones. Think about how the map will be used. Color may be irrelevant if the map is to be distributed by fax or printed in a document that will not be reproduced in color. In such cases, color can be counterproductive. Even black-and-white (gray-scale) shades should be chosen carefully if the document is to be faxed. Gray-scales should not be too subtle (variations will be lost), and cross-hatched shading should be coarse (lines relatively far apart) or they will not hold up through the fax process.
- **Lettering.** Are the default font style and size appropriate? Consider, for



example, whether the user may decide to make the map into an overhead transparency. Will the lettering be legible in that medium?

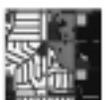
- **Methodology.** What opportunities does your software offer? (Hot spot identification? Buffers? Filtering?) Have those opportunities been taken advantage of? Or have unnecessarily glitzy methods only created confusion?
- **Privacy.** Will this map reveal information about individuals who may be subject to privacy restrictions? Most data are in the public domain, including arrest records and court documents. Exceptions to this include the practice of protecting the identities of rape or sexual assault victims, and the identities of juveniles.

Summary

Chapter 4 has explained:

- The impact GIS has had on crime mapping.
- How GIS is used in law enforcement agencies.

- How GIS affects what we can do and how we do it.
- The difference between vector and raster formats.
- What geocoding is.
- How we can filter data and make useful maps according to specific criteria.
- How GIS can be used to measure information on maps.
- The meaning of derivative measures and how they are created and used.
- What hot spots are and how they are defined, measured, and mapped.
- What buffering is in the context of GIS.
- How large databases can help with mapping and analysis.
- What data warehousing and data mining are.
- Some factors that demand the exercise of caution in the mapping process.



Notes

1. Data transformations using square roots or logarithms smooth data and reduce problems associated with having both very large and very small values in a distribution. The effect may change the distribution to a more normal, or bell-shaped, curve.
2. The need for more integration of statistical tools is gradually receiving more recognition, and the problem should diminish over time. For example, a new package known as Regional Crime Analysis GIS (RCAGIS) developed for the Baltimore-Washington, D.C., region has an embedded statistical package, CRIMESTAT. Also, the S+ package by MathSoft® (<http://www.mathsoft.com/splus/>) interfaces with ArcView.
3. Hot spots may be split by jurisdiction boundaries in such a way that they fail to meet hot spot criteria on either side of the line.
4. Reminder: Small-scale maps show large areas, large-scale maps show small areas.
5. This problem of overlapping points can be fixed with software manipulation.
6. Regions of all nations have FIPS codes. Greater London, U.K., for example, is UK17. For details, see: <http://webcentral.bts.gov:80/itt/T%26T/resource/fipscode.pdf>.

Whats Next in Chapter 5?

- Current events in crime mapping.
- How to apply analytical crime mapping.
- Criminal intelligence.
- Crime prevention.
- Courts and corrections.
- Public information.
- Resource allocation and planning.
- Census geography and analysis.
- Crime mapping applications and improved effectiveness.

