National Institute of Justice

The Research, Development, and Evaluation Agency of the U.S. Department of Justice

NIJ National Institute of Justice

CrimeStat® *III*

User Workbook

Susan C. Smith

Christopher W. Bruce

The National Institute of Justice

Washington, DC

June 2008

About CrimeStat

CrimeStat is a spatial statistics program for the analysis of crime incident locations, developed by Ned Levine & Associates under the direction of Ned Levine, PhD, that was funded by grants from the National Institute of Justice (grants 1997-IJ-CX-0040, 1999-IJ-CX-0044, 2002-IJ-CX-0007, and 2005-IJ-CX-K037). The program is Windows-based and interfaces with most desktop GIS programs. The purpose is to provide supplemental statistical tools to aid law enforcement agencies and criminal justice researchers in their crime mapping efforts. *CrimeStat* is being used by many police departments around the country as well as by criminal justice and other researchers. The new version is 3.0 (*CrimeStat III*).

The program inputs incident locations (e.g., robbery locations) in 'dbf', 'shp', ASCII or ODBC-compliant formats using either spherical or projected coordinates. It calculates various spatial statistics and writes graphical objects to ArcView[AE], MapInfo[AE], Atlas*GIS[TM], Surfer[AE] for Windows, and ArcView Spatial Analyst[(c)].

Copyright

Citation

The program must be cited correctly in any publication or report that uses results from the program. The author's suggested citation is:

> Ned Levine (2007). *CrimeStat*: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.1). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. March.

For More Information

For more information on CrimeStat, visit http://www.icpsr.umich.edu/crimestat.

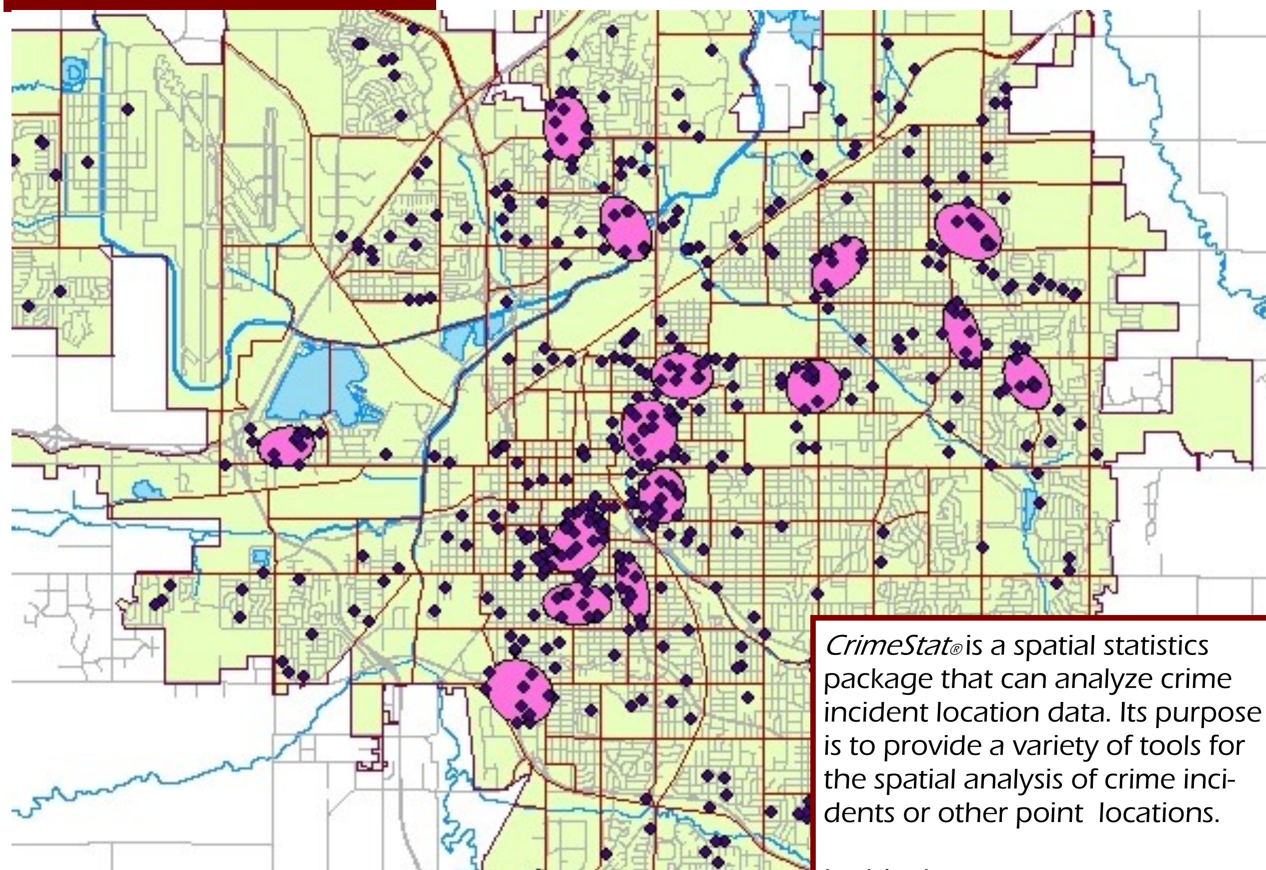
Contact Information

Technical questions relating to the CrimeStat software and/or manual should be directed to:

Dr. Ned Levine
Ned Levine and Associates
crimestat@nedlevine.com

Table of Contents

1 Introduction and Overview



CrimeStat® is a spatial statistics package that can analyze crime incident location data. Its purpose is to provide a variety of tools for the spatial analysis of crime incidents or other point locations.

In this chapter:

- Purpose of CrimeStat III
- Uses of spatial statistics in Crime Analysis
- CrimeStat III as a tool for analysts
- Statistical Routines
- Hardware and Software requirements
- Downloading sample data
- Chapter layout and design

1 Introduction and Overview

1.0 Introduction

To study crime is to study location, geography, neighborhood, environment. When it comes to an understanding of crime and disorder, knowing these is as important as knowing the criminal mind. Not all crimes have an identifiable victim. Not all have stolen property. Not all have a rational motive, nor an identifiable time of day, nor even necessarily a deliberate offender. But nearly all have a location.

In crime analysis, we often identify crime patterns by the physical proximity of the crimes. Long-term problems frequently manifest themselves as "hot spots"; in every jurisdiction, there are businesses, street corners, parks, schools, and other locations known to produce a high volume of everything from street robbery to false alarms. In solving problems, we often find that fixing something about a location, or a type of location, is the key. We allocate our patrols by district, beat, or precinct. We study road networks to analyze everything from response times to probable offender-flight routes. Journey-to-crime, redistricting, response plans, geographic profiling, selective enforcement, automatic vehicle location, probation and parole tracking... crime *is* geography. Policing *is* geography.

1.1 Geographic Information Systems

In the modern era, we study this geography with **geographic information systems**, and the first part of this analysis involves **crime mapping**—**geocoding** incidents or other police-related data and displaying them on a paper or computerized map. There are various types of **thematic maps** like point symbol maps, choropleth maps, and graduated symbol maps that can help tell a “story” using our data. We map crime for a number of reasons:

* To identify **patterns** and **problems**
* To identify **hot spots**
* To serve as a visual aid when disseminating information about patterns and problems
* To show the relationship between geography and other factors
* To look at direction of movement
* To query data by location (e.g., buffers)
* To create patrol districts
* To track changes in crime
* To make maps for police deployment and general police information

Glossary

Geographic information systems: Hardware and software that collects, stores, retrieves, manipulates, queries, analyzes, and displays spatial data. GIS is a computerized fusion of maps with underlying databases that provide information about map objects.

Crime mapping: the application of geographic information systems to crime and police data. Crime mapping helps analyze the “where” factor of crime patterns, series, trends, and problems.

Geocoding: the process of assigning geographic coordinates to data records, usually based on the street address. Geocoding is what turns a list or database of crimes into a map of those crimes.

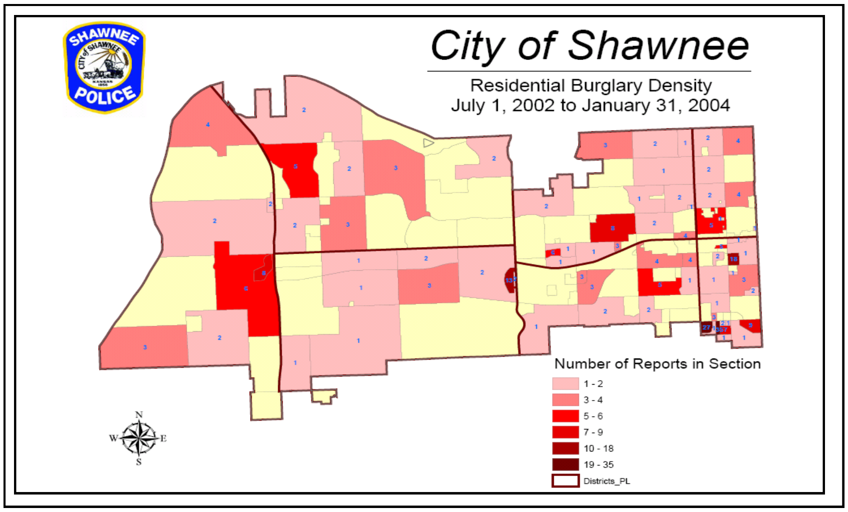
Thematic maps:: a map that shows a “theme” or tells a story rather than simply providing a visual representation of the earth’s surface. Popular thematic maps used in crime mapping are choropleth maps (polygons colored or patterned based on the volume of crime); graduated symbol maps (larger dots at locations with more crimes); and point symbol maps (different colors or symbols based on the type of crime or some other attribute feature).

Patterns:: two or more incidents with direct common causal factors. Patterns are usually short-term phenomena. The most common types of patterns are series, in which the common causal factor is the same offender(s).

Problems : an elusive term that generally refers to a long-term or chronic crime phenomenon based on opportunity rather than on a single offender or group of offenders. The term generally encompasses both the incidents themselves and their underlying causes.

Hot Spots: areas of concentrated crime. Hot spots can be single addresses, parcels, blocks, neighborhoods, or even entire cities, depending on the overall scale of the analysis.

Example: Choropleth map depicting burglary density in the City of Shawnee between July 1, 2002 and January 31, 2004. Using the legend (and labeling provided), the user can identify areas with greater and lesser burglaries. The colored polygons represented "reporting areas" developed by the Shawnee Police Department.



Creating the map is only the first step. The analyst then must *analyze* the mapped data to answer his or her questions: Is there a pattern? What is the nature of the pattern? What are its dimensions? Where are the hot spots for this type of crime? What things might be influencing those hot spots? Where might a serial offender strike next? For most analysts, the predominant paradigm to answer these questions has been *visual interpretation*: simply looking at the map and using common sense, judgment, and knowledge of the jurisdiction and its dynamics. For many crime analysis tasks, visual interpretation works just fine. It can usually identify the spatial concentration of a pattern, allow the analyst to recognize the most serious hot spots, and provide enough information to general reasonable answers to common spatial questions.

But there are times when visual interpretation does not do the job. It cannot easily pick out hot spots among thousands of data points. It cannot detect subtle shifts in the geography of a pattern over time. It cannot calculate correlations between two or more geographic variables. It cannot analyze travel times among complex road networks. And it cannot apply complicated journey-to-crime calculations across tens of thousands of grid cells. For these things, and more, we need **spatial statistics**. And that is where CrimeStat comes in.

Glossary

Spatial Statistics: mathematical techniques that apply descriptive and multivariate statistics, mathematical modeling, and algorithms to spatial data, usually using geographic coordinates.

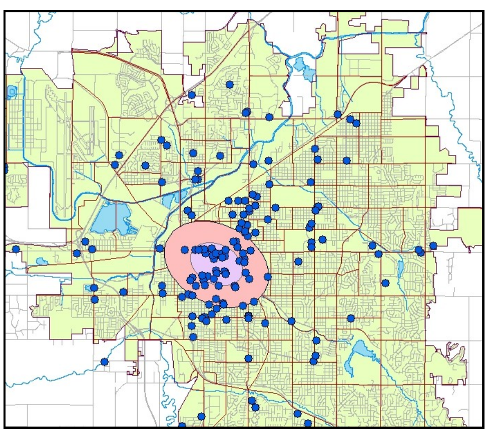
1.2 What is CrimeStat?

CrimeStat is a Windows-based spatial statistics application developed by Ned Levine and Associates of Houston, Texas, under a grant from the National Institute of Justice. Version 1 was introduced in August 1999. The current version, 3.1, was released in March 2007.

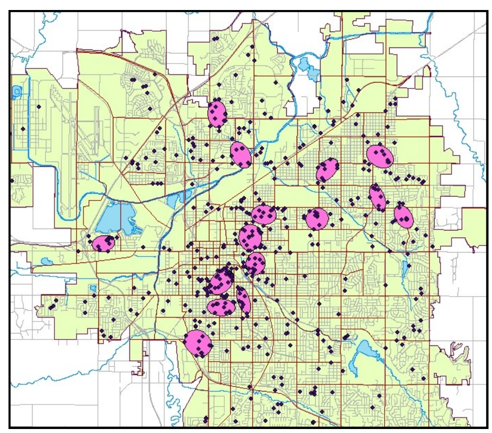
CrimeStat is *not* a GIS. It does not create or display maps of crime. Instead, it reads files geocoded by geographic information systems, like **ArcGIS** and **MapInfo**, and exports its results into formats that those programs can read. For this reason, effective use of CrimeStat requires at least intermediate knowledge of your GIS. You will switch back and forth between the two frequently.

Given the coordinates of crimes (or other types of police data), either individually or aggregated into polygons, CrimeStat can perform a number of calculations and produce a number of types of map layers:

* The mean or center of minimum distance of a group of incidents
* An area representing the standard deviation (either elliptical or rectangular) of a group of incidents
* An area representing the full geographic extent of a group of incidents
* Statistics that measure the spatial relationship between points
* Statistics that measure the level of clustering or dispersion within a group of incidents, including statistics that help identify statistically significant concentrations of incidents
* Distance measurements between points
* Identification of hot spots based on spatial proximity
* Estimation of density across a geographic area through "kernel smoothing"
* Statistics that analyze the relationship between space and time
* Statistics that analyze the movement of a serial offender
* Routines that estimate the likelihood that a serial offender lives at any location in the region, based on journey-to-crime research
* ...And much, much more not covered in this workbook



Standard Deviational Ellipse



Nearest Neighbor Hierarchical Clustering

Glossary

ArcGis: a geographic information system software line made by ESRI for Redlands, California. The ArcGIS family includes ArcReader, ArcView, ArcEditor, and ArcInfo (all different "levels" of the same product) and their associated extensions. ArcView is the most commonly used GIS among law enforcement agencies.

MapInfo a geographic information system produced by Pitney Bowes MapInfo Corporation of North Greenbush, New York. MapInfo is the second most commonly-used GIS in law enforcement.

These different statistics may seem esoteric and academic, but over the course of this book (and its associated training), we will show you how CrimeStat's statistical routines can help you:

* Identify crime patterns and **series**
* Identify the "target area" in which a serial offender is most likely to strike next
* Identify and triage hot spots
* Conduct a risk analysis across your jurisdiction based on known crime locations

There are other useful routines in CrimeStat not included in this workbook that can:

* Create a "geographic profile" in a crime series, to help prioritize investigation of suspected offenders
* Optimize patrol routes and response times

CrimeStat, in short, is valuable for **tactical crime analysis**, **strategic crime analysis**, and **operations analysis**.

1.3 The Need for Spatial Statistics in Crime Analysis

As we noted earlier, most crime analysts interpret maps by simply scanning them with their eyes. This works well for simple maps, or maps with few incidents, but it places too much demand on the limitations of human perception. Certain crime analysis chores, like hot spot identification and spatial forecasting, are enhanced with spatial statistics. Other functions, like **geographic profiling** and density mapping, are impossible without spatial statistics.

Figure 1-1 gives an example of a crime map easy to interpret visually. This eight-incident robbery series exhibits an obvious clustering. We could use CrimeStat to identify the mean center, to show us the series' geographic extent with a convex hull polygon, or to draw a rectangle or ellipse around the majority of the incidents—but a reasonably intelligent person with a crayon could accomplish as much, and far quicker.

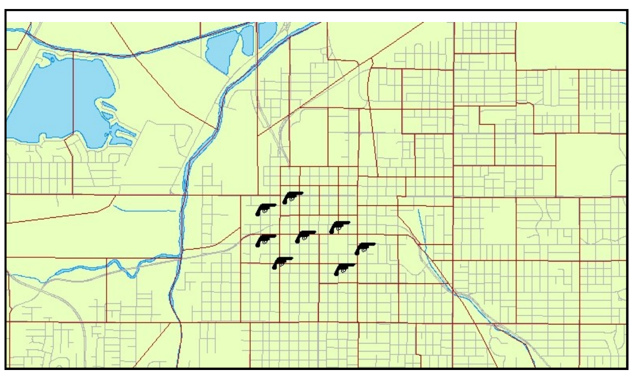


Figure 1-1: Example of a crime map easy to interpret visually

Glossary

Tactical Crime Analysis: the regular search for emerging crime patterns and series, the techniques used to examine and describe the characteristics of these patterns, the methods by which this information is disseminated to a police agency, and the responses that the agency designs and implements.

Strategic Crime Analysis: the analysis of long-term trends or crime problems.

Operations Analysis: the study of a police agency's allocation of resources, such as officer distribution by shift and beat.

Geographic profiling: is a criminal investigative methodology that analyzes the locations of a connected series of crimes to determine the most probable area of offender residence. Typically used in cases of serial murder or rape (but also arson, bombing, robbery, and other crimes), the technique helps police detectives prioritize information in large-scale major crime investigations that often involve hundreds or thousands of suspects and tips.

Figure 1-2 is a different story. Instead of an eight-incident series, we have a 634-incident pin map depicting thefts from vehicles in the spring of 2007. A few factors make visual interpretation inadequate for this map. First, the sheer number of incidents overwhelms the senses. Hot spots seem to appear wherever we look, and the downtown area looks like one huge hot spot. Second, we cannot see multiple incidents at the same location, and 87 locations have more than one incident—two locations have seven! But where our eyes cannot filter through the volume, CrimeStat can. In Figure 1-3, we've used Nearest Neighbor Hierarchical Spatial Clustering to automatically identify areas with exceptional volume.

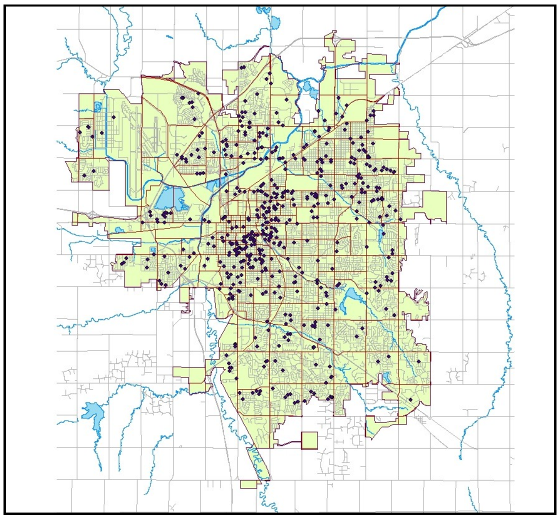


Figure 1-2

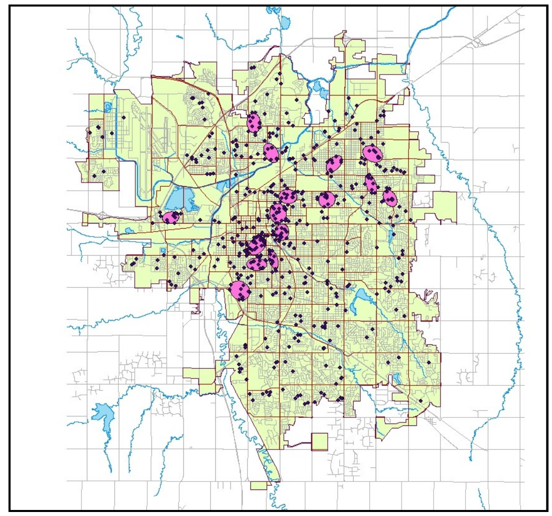


Figure 1-3

Figure 1-4 introduces a problem of a different type. Here, the series of 30 residential burglaries would appear visually manageable, and indeed we can identify its extent and concentrations without much trouble. But there is a subtle shift occurring in the series that we cannot visually identify: the mean location is moving northwestward over time. We would be hard-pressed to identify this trend visually, but CrimeStat's Correlated Walk Analysis routine helps us identify the likely pattern of movement to the next strike in Figure 1-4.

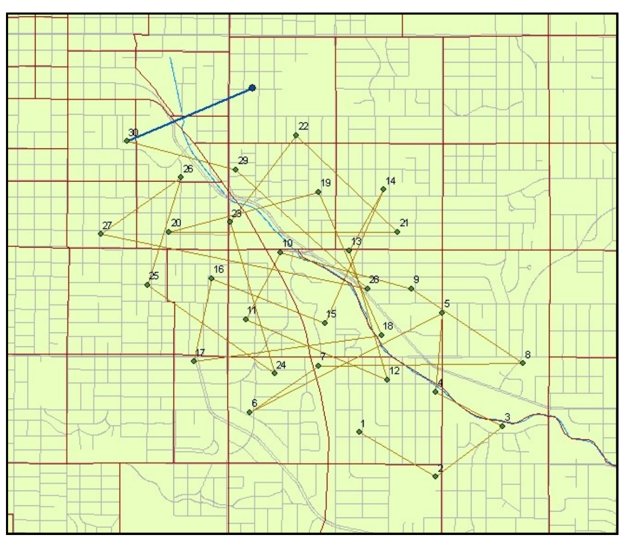


Figure 1-4

That said, CrimeStat is not the only spatial statistics package available to crime analysts. ArcGIS and MapInfo come with tools and scripts that allow various forms of spatial analysis (probably the most well-known of these is ArcView's SpatialAnalyst). Some crime analysts are strong proponents of the ArcView Animal Movements extension to track serial offenders. Geographic profiling software includes Rigel by ECRI and Dragnet from the Center for Investigative Psychology. Finally, analysts could export their crime data, with coordinates, to Microsoft Excel or SPSS and calculate spatial statistics on their own. CrimeStat's virtue is collecting different methods of spatial statistical analysis into a single application that works with multiple geographic information systems, is fairly easy to use given the complexity of the underlying calculations, and is free.

1.4 Hardware and Software Requirements

CrimeStat was developed for the Windows operating system. It will work on machines with Windows 2000, Windows XP, and Vista. Minimum requirements for CrimeStat are 256MB of RAM and a 800MHz processor speed, but an optimal configuration is 1GB of RAM and a 1.6MHz processor. Some of the processes used by CrimeStat, depending on the size of the data file, may required millions of calculations per output. Obviously, more RAM and a greater processor speed will provide a faster CrimeStat experience. Multi-processor machines will also run CrimeStat considerably faster.

Many of CrimeStat's outputs are meant to be displayed in a GIS, and you will likely need a GIS to generate the types of files CrimeStat can read (see Chapter 2). Therefore, analysts who want to get the most use from CrimeStat should also have the latest version of ArcGIS, MapInfo Professional, or whatever GIS application they prefer.

1.5 Notes on this book and this course

This book, and its associated course, are an introduction to the use of CrimeStat in crime analysis. We have chosen the CrimeStat routines and techniques that we think are most valuable to crime analysts and yet still accessible to analysts who are using CrimeStat for the first time. The techniques included in this course represent less than half of the full functionality of the program.

Some techniques we excluded because their complexity requires more attention than we could give in an introductory course (Correlated Walk Analysis, Journey-to-Crime, Crime Travel Demand); others we excluded because they seemed to have limited utility for the typical municipal analyst. The latter point is not meant to be a criticism of the program: CrimeStat was developed for criminologists and researchers as well as crime analysts.

Although we cover some GIS issues in Chapter 2, both this book and this course generally assume that you are already an intermediate or advanced GIS user. This means that you should know how to:

1) Arrange layers in a GIS to create a basemap
2) Geocode data
3) Create thematic maps
4) Import data created by other applications into your GIS
5) Troubleshoot issues associated with projections and coordinate systems
6) Interpret different file types & their associated extensions that make up each GIS layer
7) Modify and update attribute data for your GIS layers
8) Export data, with coordinates, from your GIS to other file types

The GIS screenshots in this book come mostly from ArcGIS 9.2, and the training course that accompanies this workbook also uses ArcGIS. Analysts who use other GIS systems can still follow the steps in the workbook; they will just have to change the output types to their preferred GIS.

1.6 Exploring Lincoln, Nebraska

The lessons and screen shots use GIS, crime, and call-for-service data provided by the Lincoln Police Department in Lincoln, Nebraska. In several cases, to illustrate a particular point, we have created or invented data, so none of the lesson outputs or maps should be taken as an accurate representation of crime in Lincoln.

The data to be used in this workbook is located at the following site:
http://www.icpsr.umich.edu/CRIMESTAT/workbook.html

Before you begin the CrimeStat lessons, it's a good idea to open the Lincoln, Nebraska data layers in your GIS application, assign them the styles and labels that you want, and explore the city a little bit. Refer to Chapter 8 for a list of the Lincoln data layers. Your base map ought to include, at a minimum, streets, citylimit, cityext, streams and waterways (and should also be displayed in a logical order). You can add additional layers if you choose to, however too many layers, albeit "pretty to look at", may lead to confusing maps when the data is added and displayed. **Save this basemap for future use.**

1.7 A Tour of CrimeStat

As we noted earlier, CrimeStat can calculate a wide variety of spatial statistics and, in most cases, generate resulting maps. The rest of this workbook contains step-by-step instructions for importing data and conducting these spatial analysis yourself. Before we begin, however, we thought we'd try a whirlwind tour of the major screens and what's possible with CrimeStat 3.1.

Using CrimeStat is essentially a four-stage process:
1. Enter the data you want CrimeStat to use in the "data setup" screen
2. Choose the type of analysis you want to perform, and enter the associated parameters
3. Click "compute"
4. Close the output window.
5. If the computation results in a map layer, open it in your GIS for further analysis

Chapter 2 covers the details of Step 1. Most of the rest of the book covers the details of Step 2. Steps 3 and 4 are intuitive. Step 5 will be completed on most of the routines (though some routines will not result on a map layer, but rather statistical results that we will interpret.

1.8 Data Setup

The *data setup* screen is where you specify the files on which you want CrimeStat to perform the calculations, and the various parameters of those files. CrimeStat does not *query* data, so you'll need to have already used your GIS or database program to query out only those incidents on which you want to perform spatial calculations.

CrimeStat requires at least one *primary file* and will allow for a *secondary file* that can be compared to the primary file for some types of spatial statistic—for instance, a comparison of homicides (primary file) to poverty rates (secondary file). Finally, a *reference file*—either imported or created in CrimeStat—is used for certain routines, and the *measurement parameters* tab prompts you for the geographic area of your jurisdiction, the length of the street network, and the methods for calculating distance. These parameters are all covered in Chapter 2.

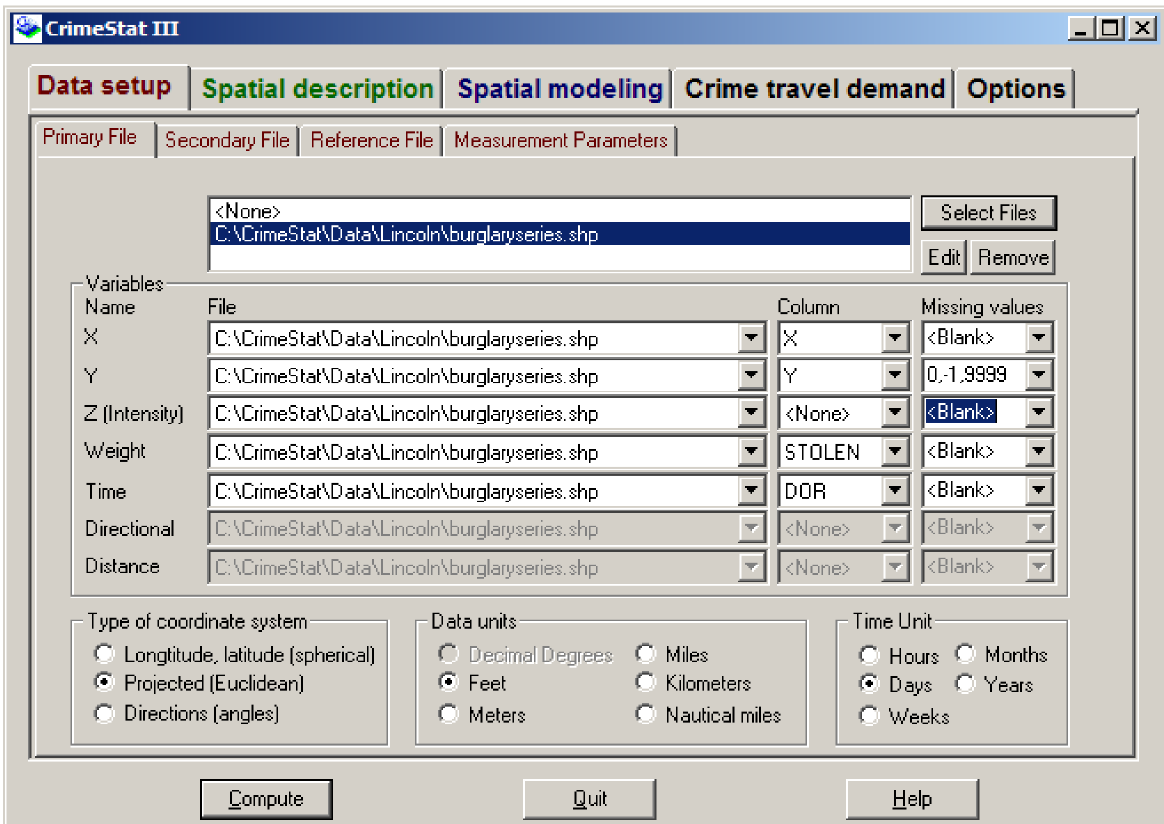


Figure 1-5: the data setup screen.

1.9 Spatial Description

The various "Spatial Description" functions describe your data as it is. The calculations are analogous to descriptive statistics.

On the *Spatial Distribution* screen, we have a number of functions that tell us the central tendency and variance in our data, including the mean center, standard deviation ellipses, and convex hulls. These routines are discussed in Chapter 3.

The *Distance Analysis I* screen has several functions that measure distances between points. *Nearest neighbor analysis* and *Ripley's K* help us determine the significance of the clustering or dispersion of our incidents. *Assign primary points to secondary points* takes the points from one file and connects them to their nearest neighbor in another file. These routines are covered in Chapter 4. *Distance Analysis II* has functions that create matrices of distances between points.

Hot Spot Analysis I and II contains a series of routines that help us identify, flag, and triage clusters in our incident data. These are covered in Chapter 5.

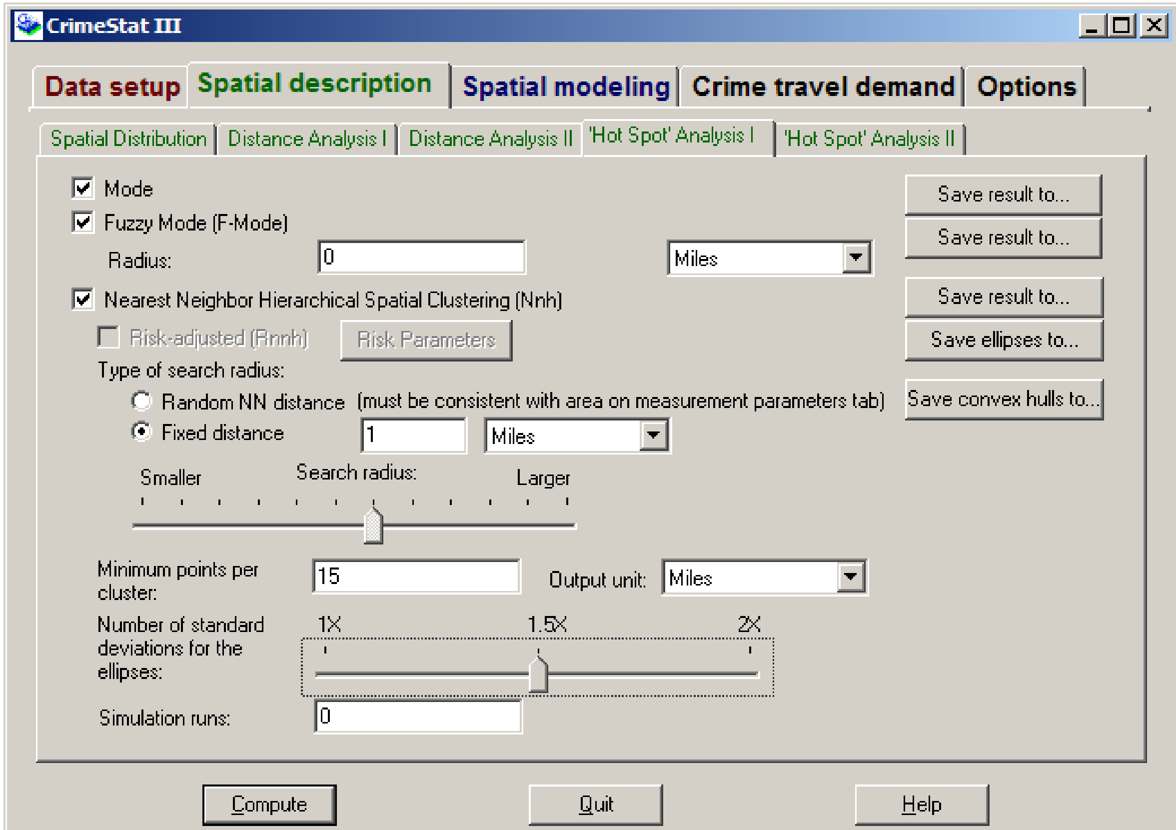


Figure 1-6: the "Hot Spot Analysis I" screen from the Spatial Description section

1.10 Spatial Modeling

Where the Spatial Description screens offer routines to help us analyze our data as it is, the Spatial Modeling screens help us create interpolations and predictions based on our data.

The *Interpolation* tab contains all the options to create a *kernel density estimation*, resulting in a map that you may know as a "density map." These are very popular with crime analysts, and we cover this screen in detail in Chapter 6.

Space-time analysis is about analyzing progression in a series of crimes, including the moving average and correlated walk analyses. We cover the spatial-temporal moving average in Chapter 7.

The *Journey-to-Crime* and *Bayesian Journey-to-Crime Estimation* screens help us determine the likelihood that a serial offender lives in a particular area, based on the locations of his offenses. Although very relevant for crime analysts, the complexity of these routines prohibit their inclusion in this introductory book and class.

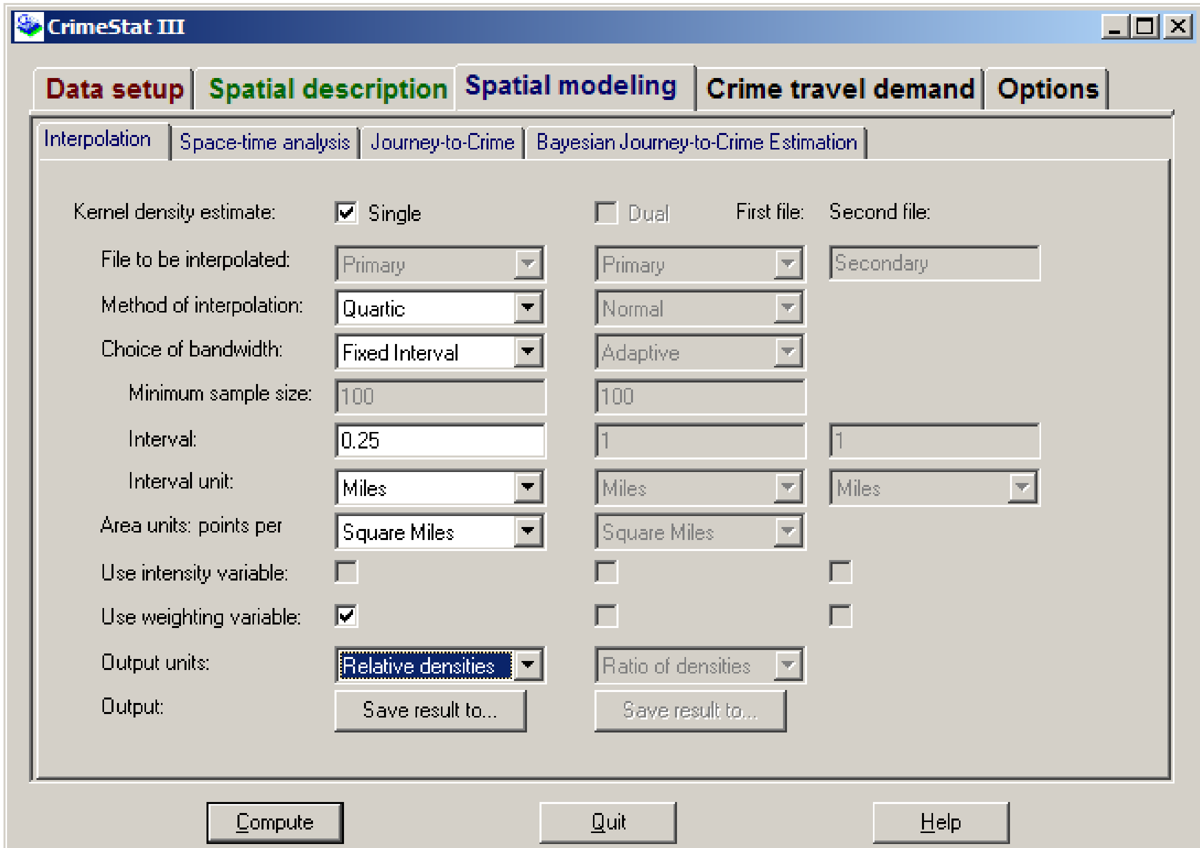


Figure 1-7: the "Interpolation" screen from the "Spatial modeling" module

1.11 Crime Travel Demand

The Crime Travel Demand screens help analyze travel patterns of offenders over large metropolitan areas. It is an emerging and potentially valuable type of analysis, but again the complexity of the routines requires us to save it for a later book and class.

1.12 Summary

Table 1-1 on the following two pages provides a summary of all of CrimeStat's functions and routines. The comments under "Use in Crime Analysis" are based on our own experience and judgment, but other analysts may have found uses for some of the routines that have not occurred to us. We welcome comments and feedback from analysts who have used CrimeStat.

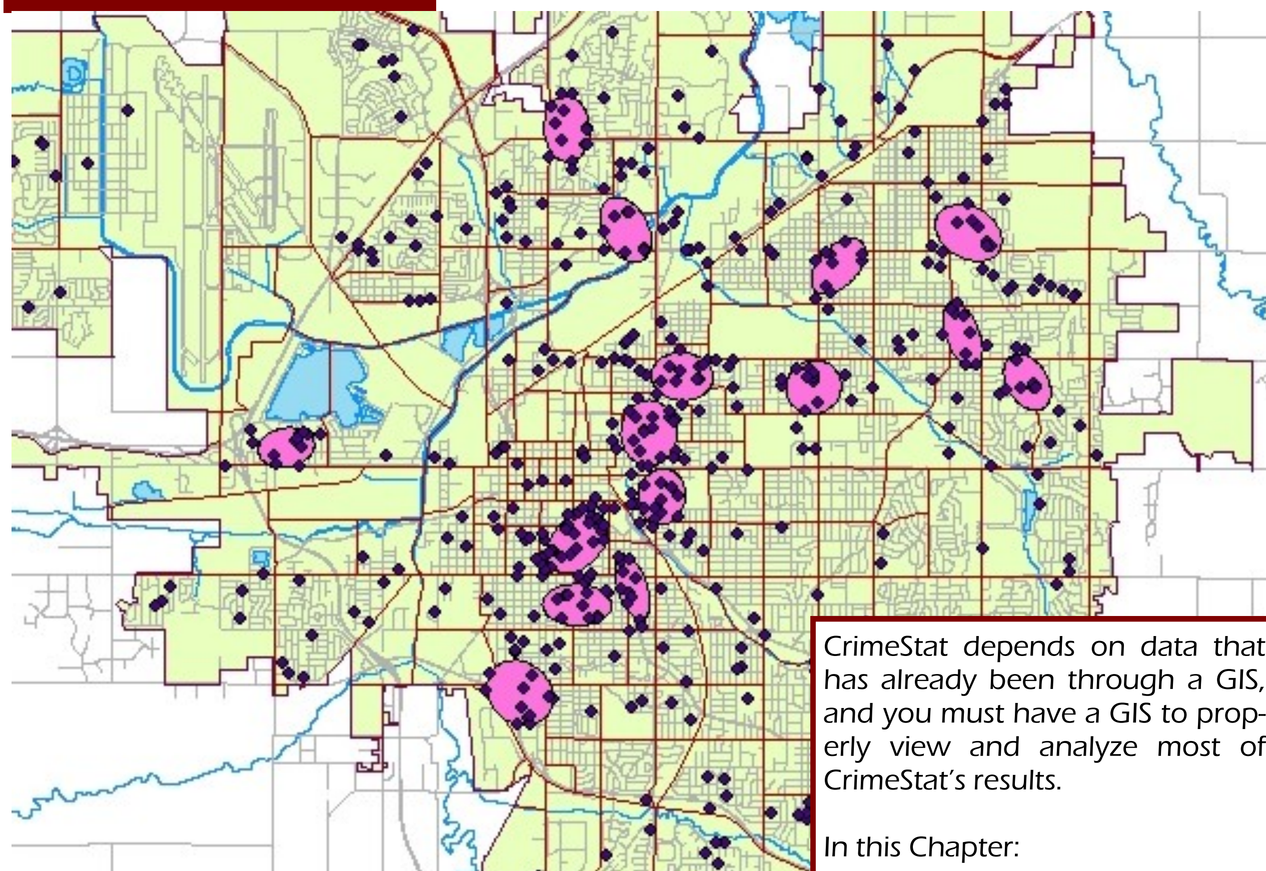
Table 1-1: a summary of the CrimeStat functions and their use in crime analysis

Function	Module & Screen	Description	Use in Crime Analysis	In this book
Mean Center	Spatial Description Spatial Distribution	Identifies the mean centerpoint of a group of points (incidents, offenders' addresses)	Limited use for certain tactical & strategic deployment purposes	Chapter 3
Standard Deviation Ellipse	Spatial Description Spatial Distribution	Creates an ellipse representing one standard deviation around the mean center for a group of points	Predictive value for future incidents in a crime series; helps identify concentration of incidents in a long-term trend or problem	Chapter 3
Median Center	Spatial Description Spatial Distribution	Identifies median center of a group of points	Limited use for certain tactical & strategic deployment purposes	Chapter 3
Center of Minimum Distance	Spatial Description Spatial Distribution	Identifies point at which distance to all other points is minimum	More useful than the median center. CMD of crimes committed by a single serial offender will usually be the best single predictor of where offender lives.	Chapter 3
Directional mean and variance	Spatial Description Spatial Distribution	Similar to mean center but for angular data	Almost none. Crime analysts will rarely if ever have data in this format.	No
Convex Hull	Spatial Description Spatial Distribution	Creates a polygon around the external boundaries of points	Predictive value for future incidents in a crime series; helps identify concentration of incidents in a long-term trend or problem	Chapter 3
Moran's I (spatial autocorrelation)	Spatial Description Spatial Distribution	Helps identify extent of clustering among incidents	Most commonly used measure of spatial autocorrelation that is used for aggregated data (e.g., by census tract of beat).	No
Geary's C (spatial autocorrelation)	Spatial Description Spatial Distribution	Helps identify extent of clustering among incidents	Very little. Rarely tells you anything you don't already know.	No
Moran Correlogram	Spatial Description Spatial Distribution	Calculations Moran's I for different distance intervals	Very little. Can be helpful in deciding on the shape of a kernel function and the bandwidth chosen.	No
Nearest neighbor analysis	Spatial Description Distance Analysis I	Determines whether points are more clustered or dispersed than expected by chance	Some. Although almost all crime data will be more clustered than expected by chance, NNA can help determine which crimes are most subject to clustering; has some limited strategic planning value.	Chapter 4
Ripley's K	Spatial Description Distance Analysis I	An alternate calculation for accomplishing essentially the same thing as NNA.	Little. More difficult to interpret than NNA, and NNA should fit most of your needs in this area.	No
Assign primary points to secondary points	Spatial Description Distance Analysis I	Matches points from one file with their closest neighbors from another file; sums the results.	Can be very valuable if you do not already have a GIS function that does this for you. Can assign crimes to nearest police station, incidents to nearest offender addresses, etc.	Chapter 4
Distance Matrices	Spatial Description Distance Analysis II	Creates a matrix of distances between points in one or two files.	Might be useful for some special projects, but in general, no.	No
Mode (Hot Spot)	Spatial Description Hot Spot Analysis I	Identifies the coordinates with the highest number of points	The function itself is very useful in crime analysis to identify the top "hot addresses," but this is easily done in most GIS programs.	Chapter 5

Table 1-1 (con't): a summary of the CrimeStat functions and their use in crime analysis

Function	**Module & Screen**	**Description**	**Use in Crime Analysis**	**In this book**
Fuzzy Mode	Spatial Description Hot Spot Analysis I	Builds on mode hot spot analysis by creating a radius around each point and capturing all points within the radius	Very useful; takes care of some of the problems with plain modal hot spots, but also has some issues.	Chapter 5
Nearest Neighbor Hierarchical Spatial Clustering	Spatial Description Hot Spot Analysis I	Based on parameters you input, creates ellipses around points of unusually dense volume	One of the most useful routines in CrimeStat with broad uses in strategic analysis. Only hot spot routine that can be adjusted for underlying risk.	Chapter 5
Spatial and Temporal Analysis of Crime	Spatial Description Hot Spot Analysis II	Alternate means for identifying clusters of points	Another very useful cluster-identification routine; should be used in conjunction with NNH.	Chapter 5
K-Means Clustering	Spatial Description Hot Spot Analysis II	Partitions all points into a number of clusters identified by the user	Some strategic value if the goal is to maintain a specific number of hot spots in your analysis. Not as useful as the other two.	No
Anselin's Local Moran	Spatial Description Hot Spot Analysis II	Determines whether polygons have unusually high volume relative to their broader neighborhoods	Some use in strategic analysis. Note: This measure also depends on Moran's I.	No
Kernel Density Estimate	Spatial Modeling Interpolation	Interpolates crime volume across entire region based on crimes at known points; creates risk surface.	Broad strategic and administrative value. One of the most popular types of maps created by crime analysts; CrimeStat gives you control over the process that most GIS-based routines don't offer.	Chapter 6
Knox Index	Spatial Modeling Space-Time Analysis	Calculation that shows relationship between closeness in time and closeness in distance.	Some minor tactical value. Knowing the relationship doesn't generally help much in tactical planning.	No
Mantel Index	Spatial Modeling Space-Time Analysis	Alternate means of calculating relationship of time and distance.	Some minor tactical value. Knowing the relationship doesn't generally help much in tactical planning.	No
Spatial-Temporal Moving Average	Spatial Modeling Space-Time Analysis	Measures changes to the mean center over the life of a series.	Indispensible in analyzing series that "walk' rather than remain in clusters.	Chapter 7
Correlated Walk Analysis	Spatial Modeling Space-Time Analysis	Analyzes movements of serial offender and makes predictions about next offense.	Very valuable for those series in which there is a predictable pattern of space/time movement. Somewhat difficult to set up and interpret, though.	No
Journey to Crime	Spatial Modeling Journey-to-Crime	Estimates likelihood that serial offender lives at any location in the area, based on locations of offenses.	Valuable for prioritizing offenders and offender searches during serial investigations. Actual validity and use under debate now. Routine requires a lot of effort on user's part to set up.	No
Bayesian Journey-to-Crime Estimation	Spatial Modeling Bayesian Journey-to-Crime Estimation	Builds on Journey-to-Crime module. Newest addition to CrimeStat.	See above	No
Crime Travel Demand	Crime Travel Demand	Analyzes offenders' travels across metropolitan area; makes predictions about routes, origins, and destinations	Significant potential strategic value, but still emerging as a technique. Requires extensive understanding and setup on part of user.	No

2 Getting Data into CrimeStat



CrimeStat depends on data that has already been through a GIS, and you must have a GIS to properly view and analyze most of CrimeStat's results.

In this Chapter:

- File formats understood by CrimeStat
- Projection and coordinate system considerations
- Associating your data with values needed by CrimeStat
- Accounting for missing values
- Creating a reference grid
- Measurement parameters
- Getting data out of CrimeStat

2 Getting Data into (and out of) CrimeStat

2.0 Introduction

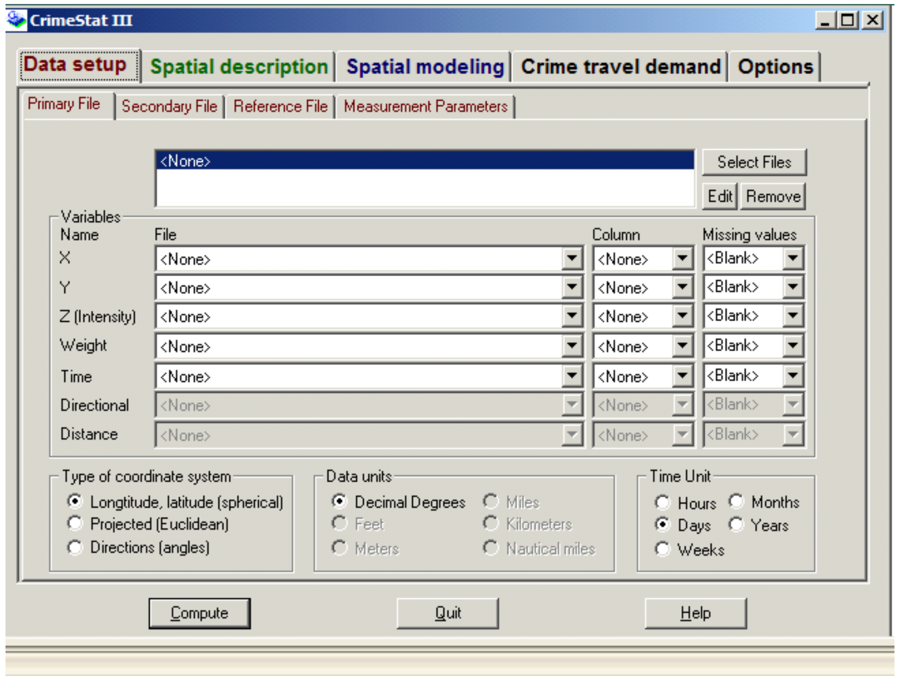


Figure 2-1: The initial CrimeStat data setup screen.

CrimeStat depends on data already created and queried, with geographic coordinates. Most analysts will have to geocode their data in their geographic information systems first, and then open the resulting file in CrimeStat. Analysts who work for agencies in which their records management or computer-aided dispatch systems automatically assign geographic coordinates will be able to import this data into CrimeStat without going through their GIS applications first.

CrimeStat reads files in a large number of formats, including: Delimited ASCII (.txt or .dat), dBASE (.dbf) files, including FoxPro databases, MapInfo attribute tables (.dat), ArcGIS shapefiles (.shp), Microsoft Access databases (.mdb) and ODBC data sources.

Any modern RMS or CAD system should be able to export data to one of these formats, either directly or through an intermediate system like Excel or Access. But for CrimeStat to analyze the data, **X and Y coordinates** must be contained within the attribute data. MapInfo .dat files do not, for instance, contain X and Y coordinates by default—the user must add them to the attribute table.

The one exception to this rule is ArcGIS **shapefiles**. CrimeStat will interpret the geography and automatically add the X and Y coordinates as the first columns in the table.

Glossary

X and Y Coordinates: An *X coordinate value* denotes a location that is relative to a point of reference to the east or west. A *Y coordinate value* denotes a location that is relative to a point of reference to the north or south.

Shapefile: The ESRI Shapefile ("shapefile") is a popular geospatial vector data format for GIS software, developed and regulated by ESRI as an open specification for data interoperability among ESRI and other software products. The ‡shapefile actually refers to a collection of files with ". shp", . "shx", ". dbf", and other extensions on a common prefix name (e.g. "streets.*").

Microsoft Access - [Accidents : Table]

IncNum	IncidentType	DateOfReport	Type	Location	XCOORD	YCOORD
2008000044	MV Accident	01/01/2008	Auto v Auto	2010 S 11TH ST	139816.77	174762.17
2008000071	MV Accident	01/01/2008	Pedestrian	1524 D ST	192668.5	226606.43
2008000074	MV Accident	01/01/2008	Auto v Auto	2455 S 60TH ST	148586.04	181126.62
2008000075	MV Accident	01/01/2008	Auto v Auto	855 S 8TH ST	190643.19	225156.94
2008000082	MV Accident	01/01/2008	Auto v Auto	340 N 56TH ST	150120.46	176752.31
2008000087	MV Accident	01/01/2008	Auto v Auto	4740 S 45TH ST	144759.4	225728.14
2008000090	MV Accident	01/01/2008	Auto v Auto	5002 GREENWOOD ST	164493.52	232881.65
2008000108	MV Accident	01/01/2008	Auto v Auto	2000 W O ST	161710.56	196090.74
2008000136	MV Accident	01/01/2008	Auto v Auto	2220 MANITOU DR	176720.79	195082.35
2008000146	MV Accident	01/01/2008	Auto v Auto	6100 O ST	160110.9	174403.62
2008000156	MV Accident	01/01/2008	Fixed Object	4900 N 27TH ST	163182.41	221745.13
2008000162	MV Accident	01/01/2008	Auto v Auto	821 WASHINGTON ST	135280.74	206007.64
2008000168	MV Accident	01/01/2008	Auto v Auto	437 FLETCHER AVE	158732.37	221403.13
2008000171	MV Accident	01/01/2008	Auto v Auto	3636 N 52ND ST	147063.18	221270.75
2008000174	MV Accident	01/01/2008	Auto v Auto	2229 J ST	141416.25	199348.8
2008000182	MV Accident	01/02/2008	Auto v Auto	6555 O ST	145315.7	205870.23
2008000186	MV Accident	01/02/2008	Auto v Auto	1420 K ST	180141.73	176914.58
2008000188	MV Accident	01/02/2008	Auto v Auto	14TH & N	134052.97	222237.52
2008000195	MV Accident	01/02/2008	Auto v Auto	14TH & N	168515.76	175303.47
2008000210	MV Accident	01/02/2008	Pedestrian	1826 B ST	174453	233006.8
2008000213	MV Accident	01/02/2008	Structure	5523 S 42ND STREET CT	182975.63	191429.24
2008000215	MV Accident	01/02/2008	Auto v Auto	3735 N 56TH ST	135588.87	207387.02
2008000230	MV Accident	01/02/2008	Auto v Auto	4000 PINE LAKE RD	159487.61	225194.97

Figure 2-2. Because it has columns for X and Y coordinates, CrimeStat can read this Microsoft Access table of traffic accidents.

CrimeStat will read, interpret, calculate, and output data in either a spherical or projected **coordinate systems**—you simply have to tell CrimeStat which one your data uses. Most U.S. users will have either:

1) **Longitude (X) and Latitude (Y) data**. For U.S. CrimeStat users, it will be simple to determine if your data is in longitude/latitude format because the X coordinate will be a negative number. If your data is in this format, CrimeStat doesn't need to know anything else. However, it will only read longitude/latitude data in **decimal degrees** (the format used by the overwhelming majority of computer systems), not in degrees/minutes/seconds—if your data is in the latter format, you will have to convert it ahead of time.

2) **U.S. State Plane Coordinates, North American Datum of 1983**. These coordinates are specific to each state and are based on an arbitrary reference point to the south and west of the state's boundaries. CrimeStat will need to know what measurement units (likely feet or meters) that your data uses.

Most CrimeStat calculations require only a single *primary* file, but some routines require a *secondary* file. For the secondary file, the same parameters have to apply.

Glossary

Coordinate Systems: A framework used to define the positions (locations) of points in space either in two or three dimensions

Decimal Degrees: are often used in GIS, web-mapping applications such as Google Map and GPS devices to express latitude and longitude coordinates, rather than expressing geographic coordinates with degrees, minutes, and seconds (dms).

2.1 Entering Your First Primary File

In the next chapter, we're going to analyze the spatial distribution of a (fictional) burglary series in Lincoln, Nebraska. Let's get it set up now.

Step 1: Launch ArcView 9.x and open the basemap project that we created in Chapter 1

Step 2: Open the shapefile at c:\CrimeStat\Data\Lincoln\burglaryseries.shp

Take a few moments to look at the burglary series, including its geography and boundaries. Use the identify tool to view some of the attribute data for the series.

Before we move on, let's check what projection and coordinate system we're using.

Step 3: To view the projection for a layer, open the layer properties (double click the layer name in the TOC) and go to the source tab. The coordinate system for us is listed as Projected Coordinate System: NAD_1983_StatePlane_Nebraska_FIPS_2600_Feet

Step 4: Launch CrimeStat. You should automatically be on the "data setup" screen.

Step 5: Click 

. Change "Type" to "Shape files" and click 

.

Step 6: Navigate to **C:\CrimeStat\Data\Lincoln** and choose the **burglaryseries.shp** file.

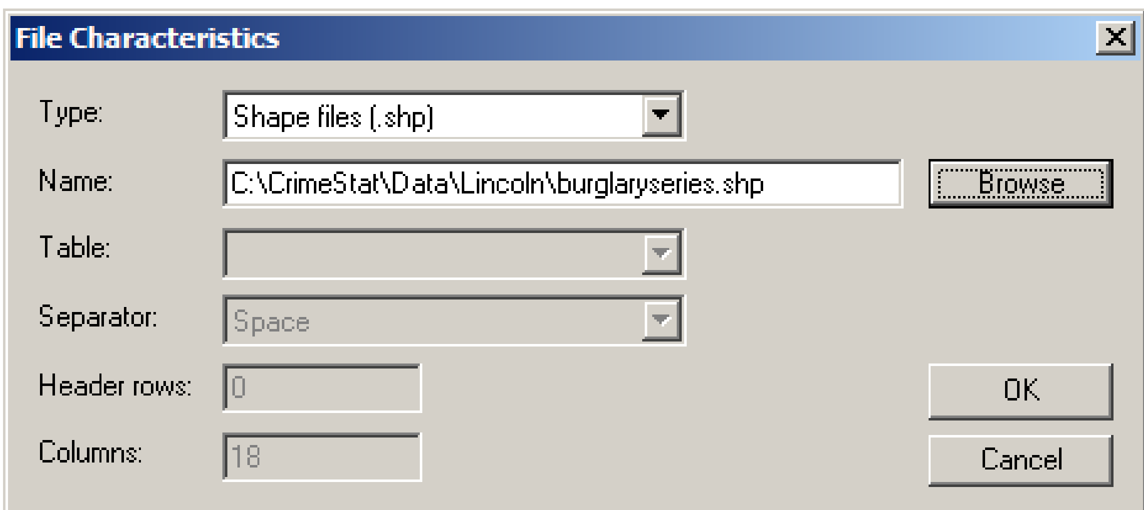


Figure 2-3: Choosing the right file for our first CrimeStat calculations

At this point, we have to tell CrimeStat which fields in our attribute table contain the X and Y coordinates (the first two options under the file name). Because we are using an ArcView shapefile, CrimeStat has already interpreted the file and added fields called "X" and "Y" to the beginning of the field list. In later lessons, when we use different types of files, CrimeStat won't be as helpful.

Step 7: Under the "Column" filed list, add the "X" field for the "X" row and the "Y" field for the "Y" row, as in Figure 2-4.

As we saw in ArcView, our coordinates are in U.S. State Plane Coordinates, with distances measured in feet. The U.S. State Plane Coordinate system is a *projected* coordinate system, so we have to input the appropriate variables into CrimeStat.

Step 8: Under "Type of Coordinate System" (bottom left), choose "Projected (Euclidean"). For "data units" (bottom center), choose "Feet."

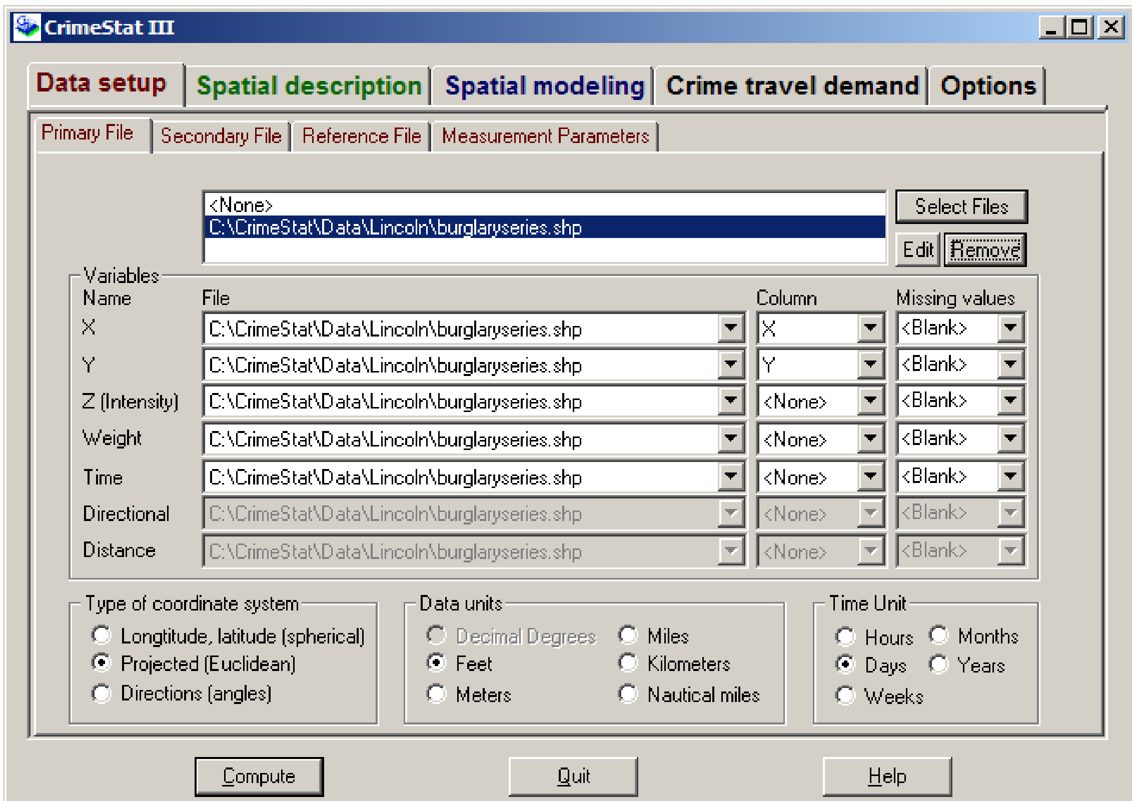


Figure 2-4 The CrimeStat data setup screen with our first file loaded, the X and Y coordinate fields chosen, and the coordinate system and data unit options set.

2.2 Other Settings and Options

At this point, we've entered everything we need to enter to perform our first calculations. CrimeStat only *requires* a file, X and Y coordinates, and information about the coordinate systems. There are, however, a number of other settings, some of which are required for certain routines.

Intensity is an optional field that tells CrimeStat how many times to "count" each point. If we enter no intensity field (if present, often called "FREQ"), the default is to count each point once, which is suitable for most spatial statistics. If instead of a burglary series, we had imported a list of all addresses in our city, with the number of incidents at each address, we would need to use the "intensity" variable. This option will become important in Chapter 6.

Weight, easily confused with "intensity" but very different, is a rarely-used field that allows us to apply different statistical calculations to different points. We do not look at weight in a this workbook.

Time measures are important for several CrimeStat space-time calculations, including the Knox Index, the Mantel Index, and the Spatial Temporal Moving Average, so CrimeStat allows a single time variable, in hours, days, weeks, months, or years. Although we have date and time fields in our current "burglaryseries" file, CrimeStat does not recognize standard "date/time" fields and instead requires its time values as integer or decimal numbers referencing a single origin point. This requires a little work on the part of the analyst, which we will cover in Chapter 7. Also, time must be measured in a consistent manner.

The **missing value** column allows us to account for bad data by telling CrimeStat which records to ignore when it performs calculations on their coordinates. If we don't avail ourselves of this option, and some of our records have zeroes where the X and Y coordinates should be, all of our calculations will jump the rails. So, if even one of our 27 burglaries had zeroes in the coordinate fields, CrimeStat would put the mean center on a runway at Lincoln Municipal Airport—1.8 miles southwest of the mean center's actual location! Similar problems plague zeroes and blank values of intensity, weight, and time.

Fortunately, we can account for several bad data types. The default is "blank," which will exclude blank fields and those with nonnumeric values. Most users should also automatically choose "0." (If you choose anything other than "<None>," CrimeStat will exclude blank fields in addition to whatever else you enter.) CrimeStat gives two other options in its drop-down menus (-1 and 9999), but users can also type directly into these fields, using commas to separate multiple values (so if –1, 99, 999 are missing value codes—along with blanks which are always excluded–you could type in those numbers in the missing value field). Note, however, that CrimeStat does not give you the option to choose a *range* of bad values, so if your bad data isn't *consistent* in its "badness" (i.e., incorrect or missing values have only a few possible codes), you may have to do a bit of data cleaning before you run CrimeStat.

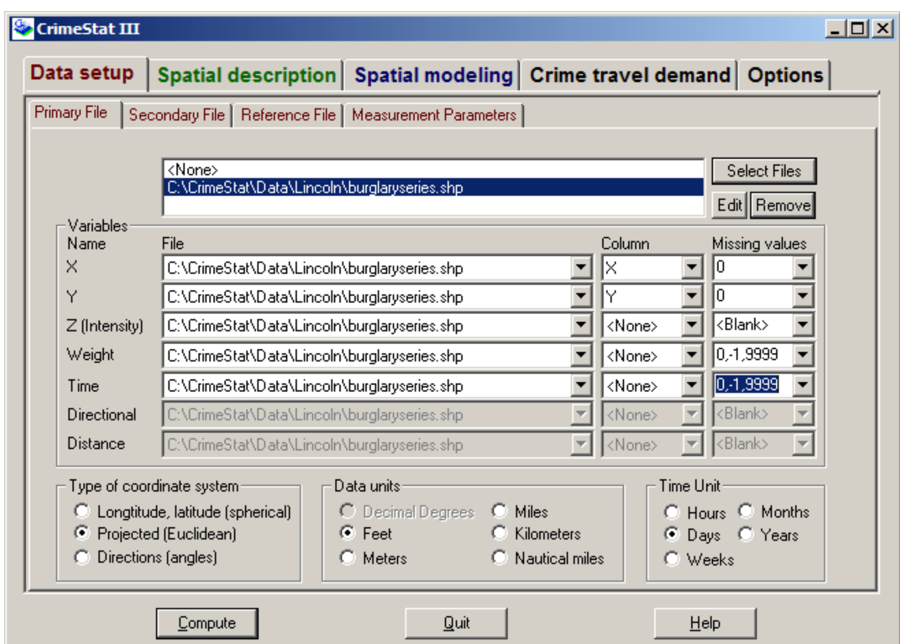


Figure 2-6: The "missing values" field allows us to warn CrimeStat about values that should be ignored.

*   The **directional** and **distance** fields are used if your data uses polar coordinate systems. Crime analysts will rarely, if ever, encounter data in such a format.

*   The **secondary file** screen allows us to enter a second file that we can relate to the first. We will see several examples of secondary files starting in Chapter 6. Secondary files *must* use the same coordinate system and data units as the primary files. All of the other options are the same, except secondary files cannot include a time variable.

2.3 Creating a Reference Grid

For certain calculations, like Kernel Density Interpolation, CrimeStat needs to know the extent of the jurisdiction. The **reference file** is a grid—either imported or created by CrimeStat—that sits over the entire study area. If you want CrimeStat to create the grid for you, you must specify the coordinates at the lower left and upper right extremities of the jurisdiction (they don't have to be exact, as long as they comfortably cover the entire area), and the coordinates must be in the same system as the primary file. We will first use a reference file in Chapter 6, but we will create one here while we're in the data setup screen.

Step 9: Click on the "Reference File" tab. Choose the "create a grid" option.

Since our primary file is in U.S. State Plane Coordinate feet, our grid coordinates will have to be in the same.

Step 10: Enter the values in Figure 2-7 into the "lower left" and "upper right" fields. These coordinates comfortably cover the City of Lincoln and give us some room to spare.

CrimeStat now needs us to determine how many cells the grid will contain. We can specify either a distance for each cell, and CrimeStat will calculate the number of cells needed; or a number of columns, and CrimeStat will calculate the required size of each cell.

Step 11: Choose "by number of columns" and enter 250.

Step 12: Click the "Save" button and give the coordinates the name "LincolnGrid." CrimeStat will save the data to the system registry, so we can load it when we need it.

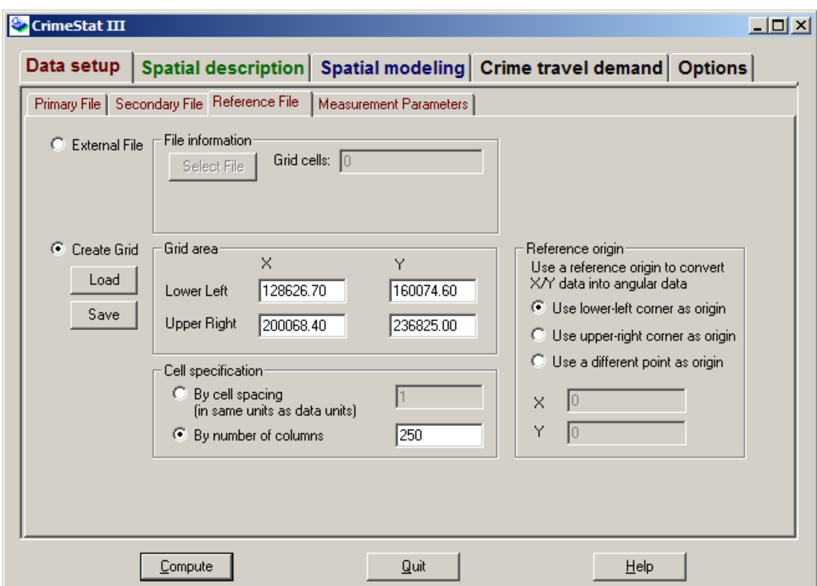


Figure 2-7: Here, we're telling CrimeStat to create a grid with 250 columns over the extent of the jurisdiction. Choosing the number of columns (250 here) is a subjective judgment of the user (the default is 100). The larger number (250) will help keep the output from looking too pixilated. Too many columns, however may bog down the system unnecessarily.

2.4 Measurement Parameters

On the "measurement parameters" screen, we add a few final bits of data necessary for certain routines: the overall geographic area covered by the jurisdiction; the length of the street network; and the way we want to measure distances.

The total **area** of your jurisdiction is fairly easy to obtain from your GIS program. For our Lincoln data, we simply ask ArcView to tell us the area of the "city limits" polygon, which turns out to be 88.19 square miles.

The **length of street network** is a bit more difficult to determine, but most GIS systems can perform the calculation by summing the individual lengths of the streets. For Lincoln, this comes out to be 1283.61 miles.

Finally, the **distance measurement** options allow us to tell CrimeStat how we want to see distances calculated. There are three options, and Figure 2-8 shows the differences among them:

Direct distance, or the literal shortest distance between points ("as the crow flies")

Indirect or *Manhattan distance*, which is measured along a grid (see figure 2-9). Manhattan distance measures make sense for grid cities like the Manhattan section of New York City, Denver, and Lincoln.

Network distance, which measures along the jurisdiction's actual road network. This requires us to specify a file which contains the jurisdictions streets (CrimeStat can read ArcView shapefiles or DBF files here), and to set up a number of other parameters. The result is more accurate distance calculations for routines like Journey-to-Crime, but a much longer processing time. The creator of CrimeStat recommends that "network distance be used sparingly for calculations." (In a grid city, Manhattan distances will often be the same as network distances.)

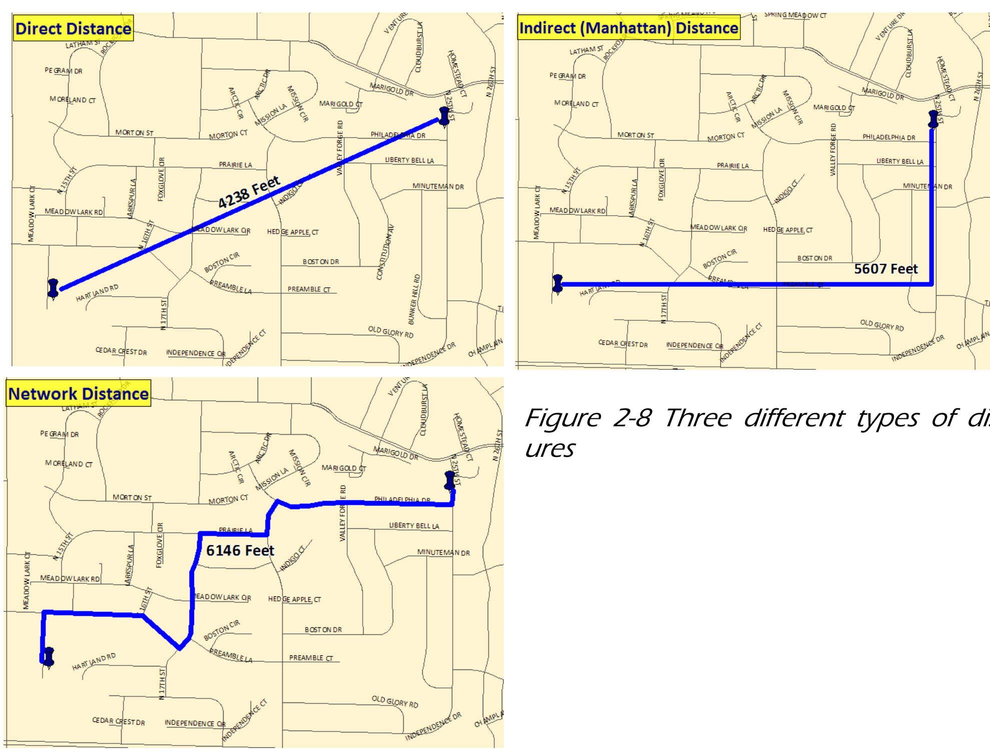


Figure 2-8 Three different types of distance measures

Step 13: Input the area and length of street network as shown in Figure 2-9. Also be sure to selected "Indirect (Manhattan)" are your type of distance measurement.

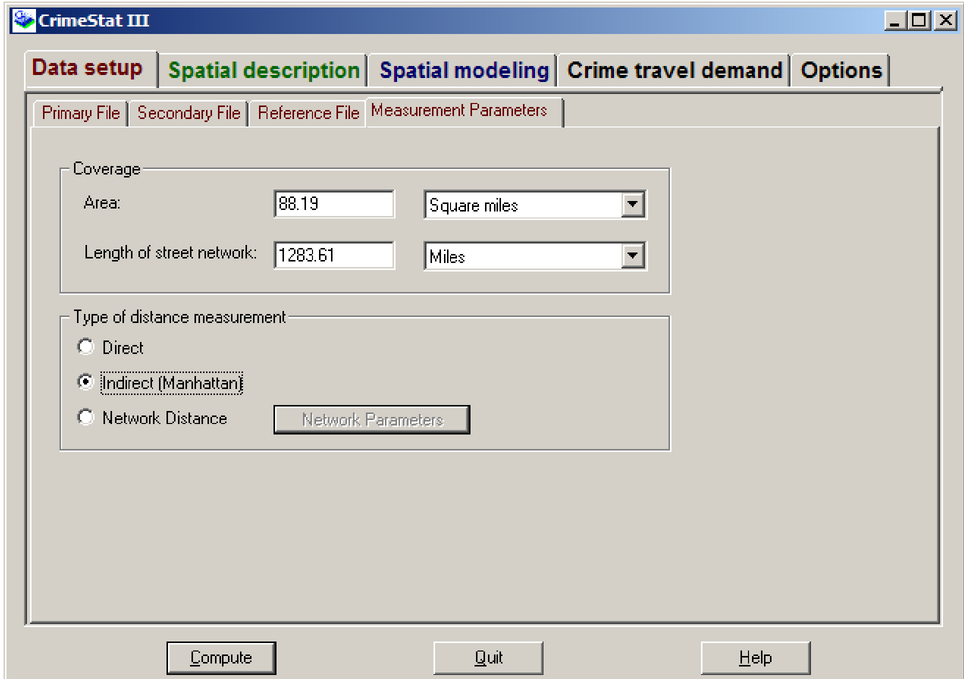


Figure 2-9: Measurement parameters entered for the City of Lincoln

2.5 Getting Data Out of CrimeStat

If a routine results in calculations for a number of records, it will allow you to export the result as a dBASE file (.dbf).

If a routine results in one or more sets of *coordinates*, CrimeStat will allow you to export the result as an ArcView shapefile (.shp), a MapInfo Interchange File (.mif), or an Atlas GIS boundary file (.bna). You can then open or import these files into your GIS and view them. We will do this frequently in the lessons to come. Refer to Figure 2-10 below.

If you are a MapInfo user, note that a .mif is set to work with lat/lon coordinates as a default, but would require changing the name and number for other coordinate systems. Much easier is to export from CrimeStat as an ArcView .shp file and then open it directly in MapInfo (allowed in Version 7.0 and above) or use MapInfo's Universal Translator to convert it to MapInfo format.

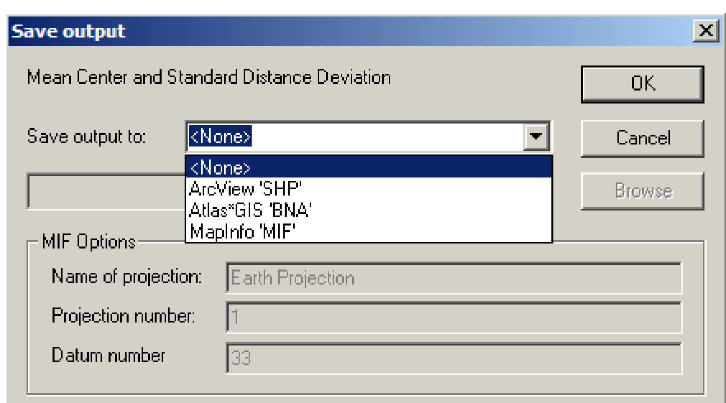
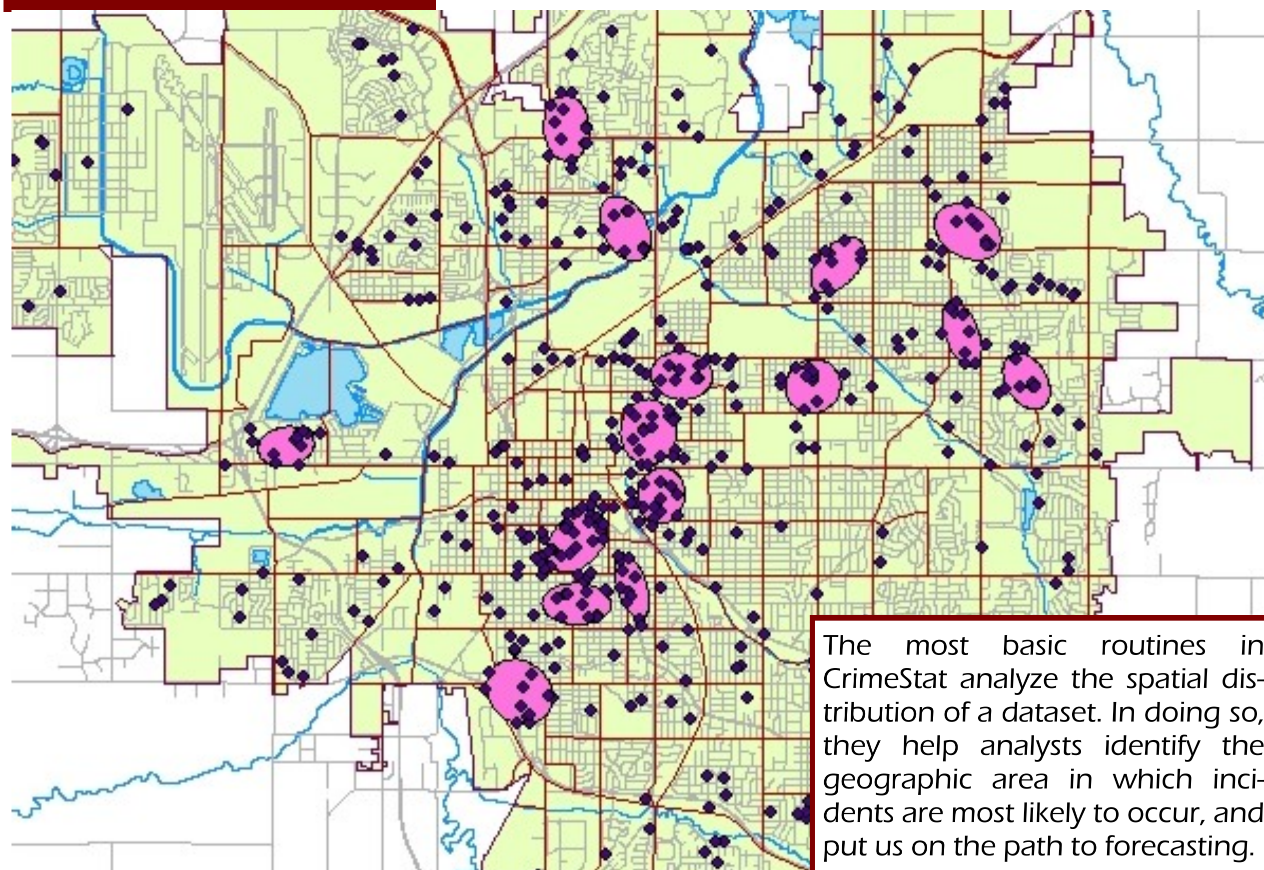


Figure 2-10: CrimeStat allows you to save coordinate outputs into three different formats

3 Spatial Distribution



The most basic routines in CrimeStat analyze the spatial distribution of a dataset. In doing so, they help analysts identify the geographic area in which incidents are most likely to occur, and put us on the path to forecasting.

In this Chapter:

- Spatial forecasting
- Mean and median centerpoints
- Measures of variance
- Analyzing a cluster
- Limitations of spatial distributions

3 Spatial Distribution

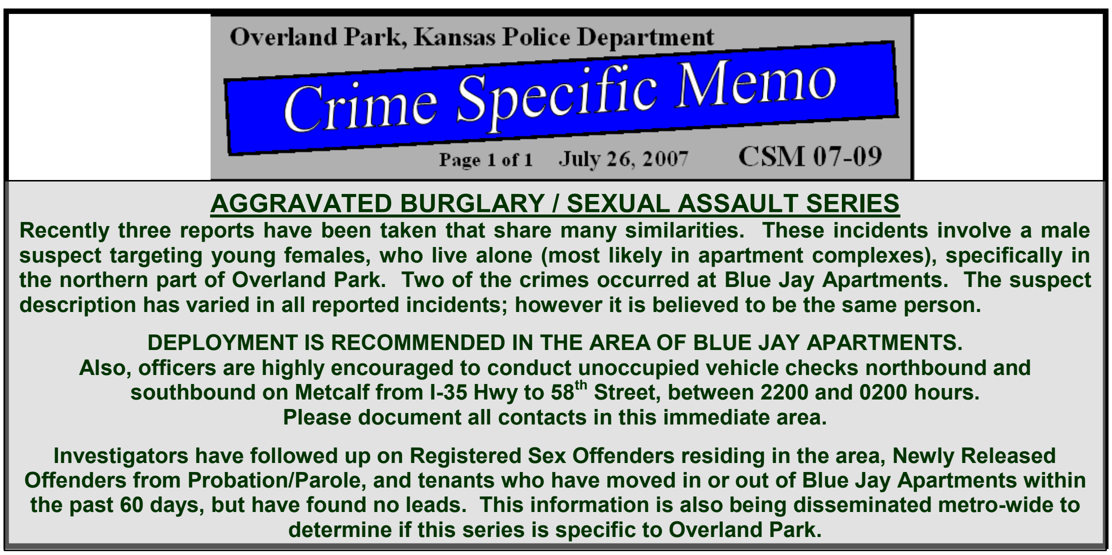
3.0 Introduction

In this chapter, we begin a series of exercises that analyze the various spatial characteristics of crimes and other public safety problems. Some of these techniques apply to series, some to long-term trends, some to both.

Implicit in these exercises is the concept of **forecasting**: identifying the most likely locations and (in some techniques) times of future events. "Forecasting," as a term, comes from meteorology, and as in meteorology, crime analysis forecasting depends on probabilities rather than certainties. Both meteorological and criminological forecasting are part-art, part-science, and both are subject to "chaos theory," in which the beating wings of a butterfly can defeat the most sincere and scientific attempt and prediction.

When a crime analyst's forecast isn't "wrong," it often seems wrong because police activity changes the pattern. Here's an unexaggerated quote received (in various forms) more than once in one of your author's careers: "You said the thief was most likely to strike in the TGI Friday's lot last night between 18:00 and 21:00. Well, I was parked in that lot all night, and nothing happened!"

Because of the possibility—probability, really—of such errors, many analysts insist that they do not forecast. This is nonsense. *Forecasting is inherent in any spatial or temporal analysis*. Just because you avoid the terms "forecast" or "predict" doesn't mean you aren't forecasting. If you describe the spatial dimensions or direction of a crime pattern, you are implicitly suggesting that future events will follow the same pattern. Stating "the burglaries are concentrated in half-mile radius around Sevieri Park" suggests that future burglaries will probably be within a half-mile radius of Sevieri Park. There's no way to avoid it. So we try to get better at it instead.


Overland Park, Kansas Police Department

Crime Specific Memo

Page 1 of 1 July 26, 2007 CSM 07-09

AGGRAVATED BURGLARY / SEXUAL ASSAULT SERIES

Recently three reports have been taken that share many similarities. These incidents involve a male suspect targeting young females, who live alone (most likely in apartment complexes), specifically in the northern part of Overland Park. Two of the crimes occurred at Blue Jay Apartments. The suspect description has varied in all reported incidents; however it is believed to be the same person.

DEPLOYMENT IS RECOMMENDED IN THE AREA OF BLUE JAY APARTMENTS.
Also, officers are highly encouraged to conduct unoccupied vehicle checks northbound and southbound on Metcalf from I-35 Hwy to 58th Street, between 2200 and 0200 hours.
Please document all contacts in this immediate area.

Investigators have followed up on Registered Sex Offenders residing in the area, Newly Released Offenders from Probation/Parole, and tenants who have moved in or out of Blue Jay Apartments within the past 60 days, but have found no leads. This information is also being disseminated metro-wide to determine if this series is specific to Overland Park.

Figure 3-1: Sample crime bulletin from Overland Park, Kansas implicitly suggesting that future events will follow the same pattern. Submitted by Jamie S. May. Printed with permission.

Glossary

Forecasting: identifying the most likely locations and (in some techniques) times of future events.

3 Spatial Distribution

3.1 Spatial Forecasting

Spatial forecasting in tactical analysis is essentially a two-step process

* Identify the target area for the next incident.
* Identify potential targets in the target area.

There is generally an inverse correlation between the predictability of the target area and the availability of potential targets. That is, when the offender prefers very specific targets (e.g., banks that are open on Saturday mornings, fast food restaurants of a particular chain), his next strike will be determined by the locations of those targets. This may take him in any direction. On the other hand, when the target area is highly predictable (e.g., the offender is moving in a linear manner across the city), it's usually because there are plentiful targets (pedestrians, parked cars, houses) distributed throughout the area. Most series fall somewhere in between.

CrimeStat's calculations can help us with Step 1: identifying the target area. Broadly speaking, there are three types of spatial patterns in tactical crime analysis:

* Those that *cluster*: incidents are concentrated in a single area, but randomly distributed throughout that area.
* Those that *walk*: the offender is moving in a predictable manner in distance and direction.
* *Hybrids*: Multiple clusters with predictable walks among them, or a cluster in which the average point walks over time.

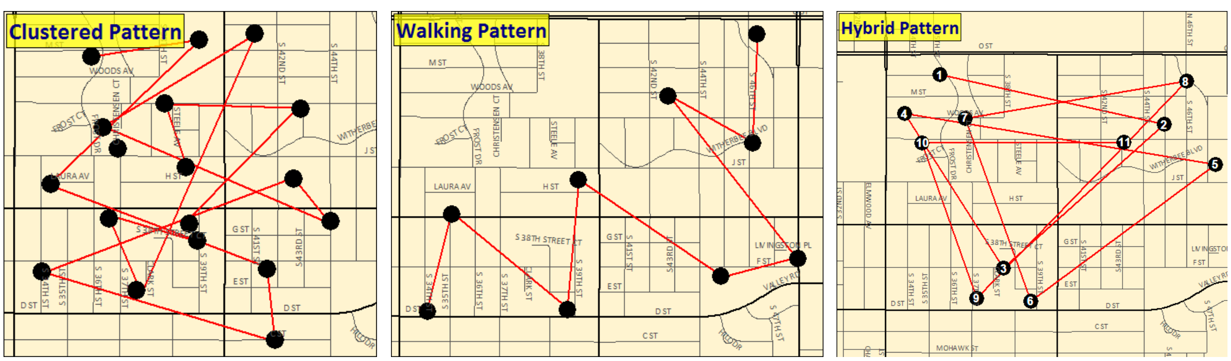


Figure 3-2: Three basic types of spatial patterns in crime series

CrimeStat's spatial distribution routines the focus of this chapter, help you analyze clusters; whereas Spatial-Temporal Moving Averages (Chapter 7) and Correlated Walk Analysis (not covered in this workbook) assist with walking patterns. A combination of the techniques is sometimes necessary for hybrid patterns.

3.2 Spatial Distribution

Faced with a pattern of street robberies, we might reasonably ask three questions:

* What is the *average* location of the robberies?
* In what area are *most* of the robberies concentrated?
* What area serves as the boundary for *all* the robberies?

The answers to all of these questions have some value in tactical response, including planning directed patrols and saturation patrols, establishing deployment points for tactical units, and identifying areas for community notifications.

Each question has several potential answers. In Figure 3-3, we see the various potential calculations plotted for a robbery series.

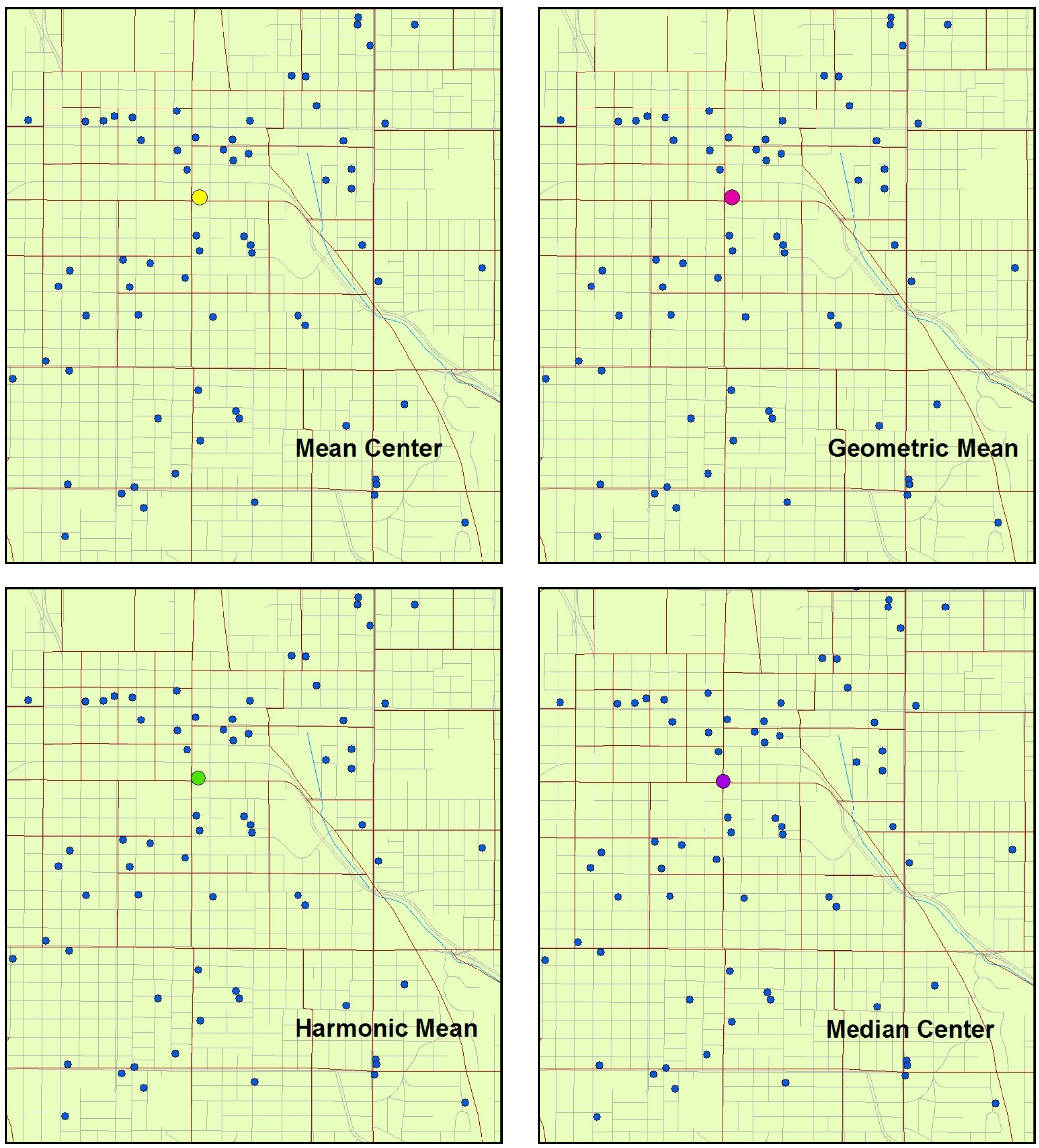


Figure 3-3

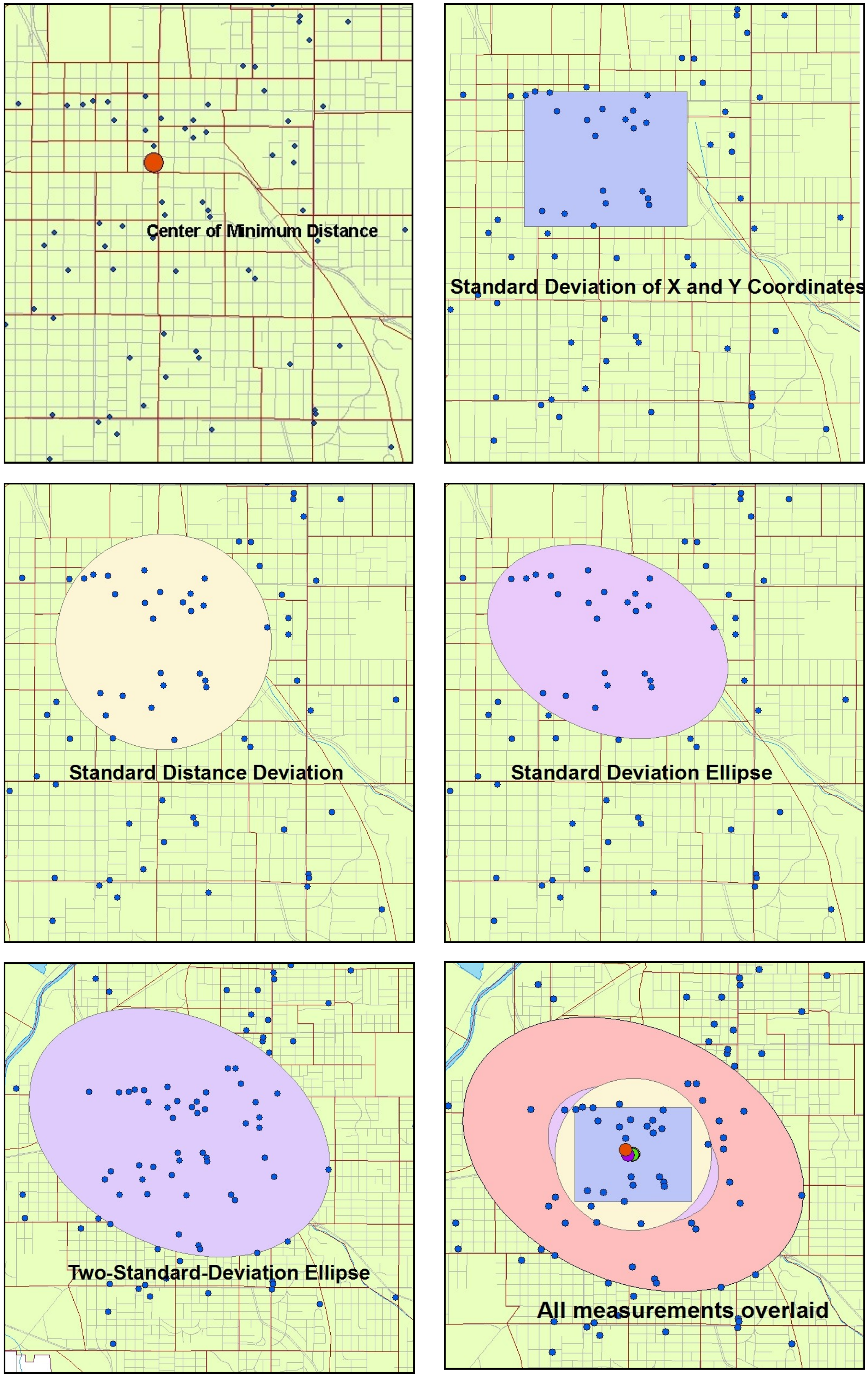


Figure 3-3 (con't)

Now that we see all of the measures of spatial distribution on the map, let's take a look at what they mean.

The **mean center** is the intersection of the mean of the X coordinates and the mean of the Y coordinates. It is the simplest of the statistics and has been calculated for years by analysts who plot their incidents on Cartesian planes. The mean center can be weighted, although that would not make sense in a crime series where every incident has a weight of one.

The **mean center of minimum distance** represents the point at which the sum of the distance to all the other points is the smallest. It is a very useful measure in journey to crime analysis and is used to define the center of many of the hot spot routines.

The **median center** is the intersection between the median of the X coordinates and the median of the Y coordinates. As with the median in non-spatial statistics, it is useful if outliers are wreaking havoc with your mean center.

The **geometric mean** and **harmonic mean** are two alternate measures of the mean center. For crime analysis purposes, the mean center almost always suffices.

In practice, the mean center, geometric mean, and harmonic mean usually end up at the same location, and the differences are esoteric enough that a crime analyst can generally rely on the mean center. In fact, none of these measures of central tendency will differ greatly. In our robbery series example, even with a couple of outliers, all five locations end up in the same block—good enough for police operations purposes!

The various polygons seek to measure the concentration, rather than the specific center point, of the series.

A rectangle shows the **standard deviation of the X and Y coordinates**. Like the mean center, analysts have been drafting these on Cartesian planes since before desktop GIS systems. The rectangle encloses the area in which four lines intersect: one standard deviation above the mean on the X axis, one standard deviation below the mean on the X axis, one standard deviation above the mean on the Y axis, and one standard deviation below the mean on the Y axis.

The **standard distance deviation** circle approaches the standard deviation calculation differently. It calculates the linear distance from each point to the mean center point, then draws a circle around one standard deviation from the center point.

The **standard deviational ellipse** improves on the standard distance deviation by accounting for skewed distributions, minimizing the amount of "extra space" that would appear in some standard distance deviational circles. CrimeStat calculates ellipses that are both one and two standard deviations from the mean center.

Finally, the **convex hull polygon** encloses the outer reaches of the series. No point lies outside the polygon, but outliers increase the size of the polygon.

Normally, the single standard deviational ellipse measure would include nearly half the cases whereas the two standard deviational ellipse would incorporate around 90% of the incidents, though the actual percentages will vary from case-to-case. In our robbery series example, the standard deviation rectangle surrounds 53 percent of the incidents; the standard distance deviation circle surrounds 84 percent; the single standard deviation ellipse surrounds 68 percent; and the two standard deviation ellipse surrounds 95 percent. The convex hull, by definition, encompasses 100 percent.

3.3 Analyzing a Cluster

In Chapter 2, we prepared the burglaryseries.shp file for analysis. We're ready to calculate the spatial description of this series. Before beginning, make sure you have followed the instructions in Chapter 2.

Also open the burglary series shapfile in ArcView and be ready to add additional layers.

Step 1: Click on the "Spatial Description" tab.

Step 2: Select all of the checkboxes from "Mean center and standard deviation" through "Convex Hull" except for "Directional Mean and Variance."

Step 3: For each of the checked boxes, click on the "Save Result To" to the right, choose "ArcView 'SHP,'" click "Browse," set the directory to C:\CrimeStat\Data\Lincoln, and name the file "burglaryseries" (see Figure 3-4).

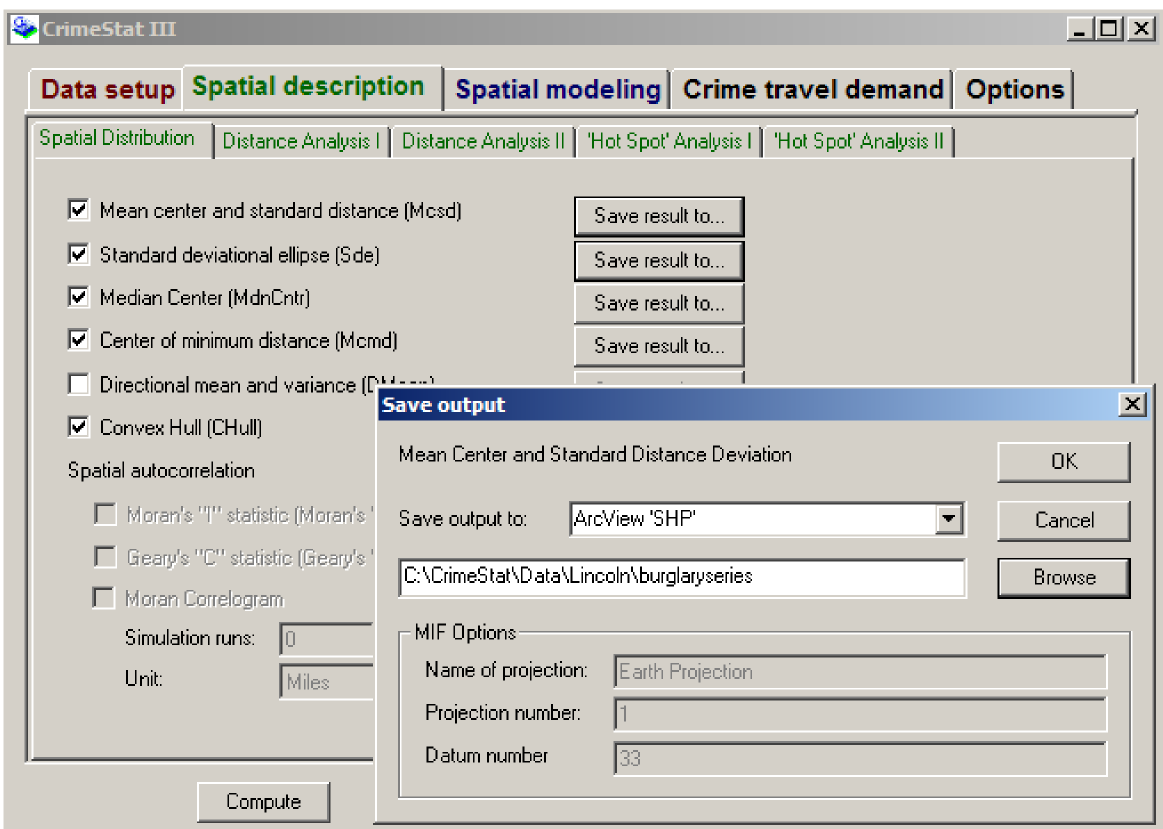


Figure 3-4

Step 4: Click "Compute."

CrimeStat will show you the results of some of the calculations (Figure 3-5; note there are five tabs at the top of the screen), and it will save 10 ArcView shapefiles to the C:\CrimeStat\Data\Lincoln directory. All the files will end with "burglaryseries" but begin with a different code corresponding to the calculation, namely:

2SDE: Two standard deviation ellipse
CHull: Convex Hull
GM: Geometric Mean
HM: Harmonic Mean
MC: Mean Center
Mcmd: Mean center of minimum distance
MdnCntr: Median Center
SDD: Standard Distance Deviation
SDE: One standard deviation ellipse
XYD: Standard deviation of the X and Y coordinates

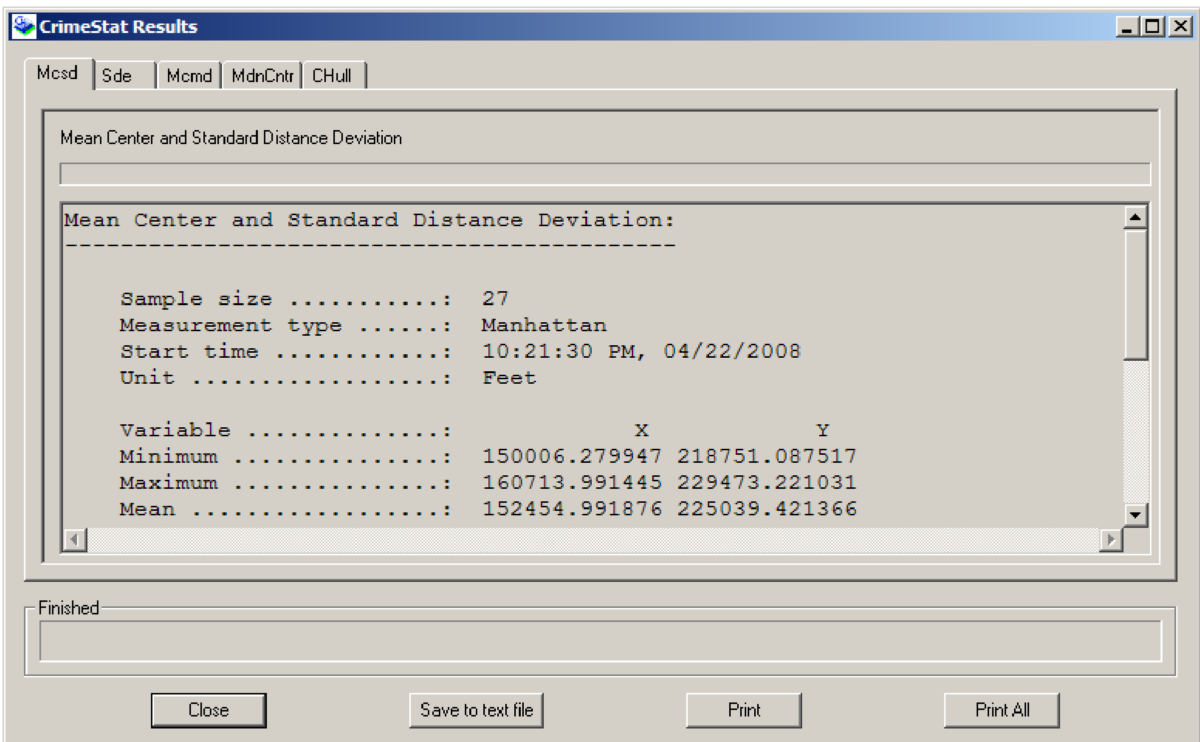


Figure 3-5

Step 5: Open each of these shapefiles in ArcView, one at a time, making sure to format each one appropriately. Change the symbology as appropriate for each layer. When you are finished, your map should look similar to Figure 3-6.

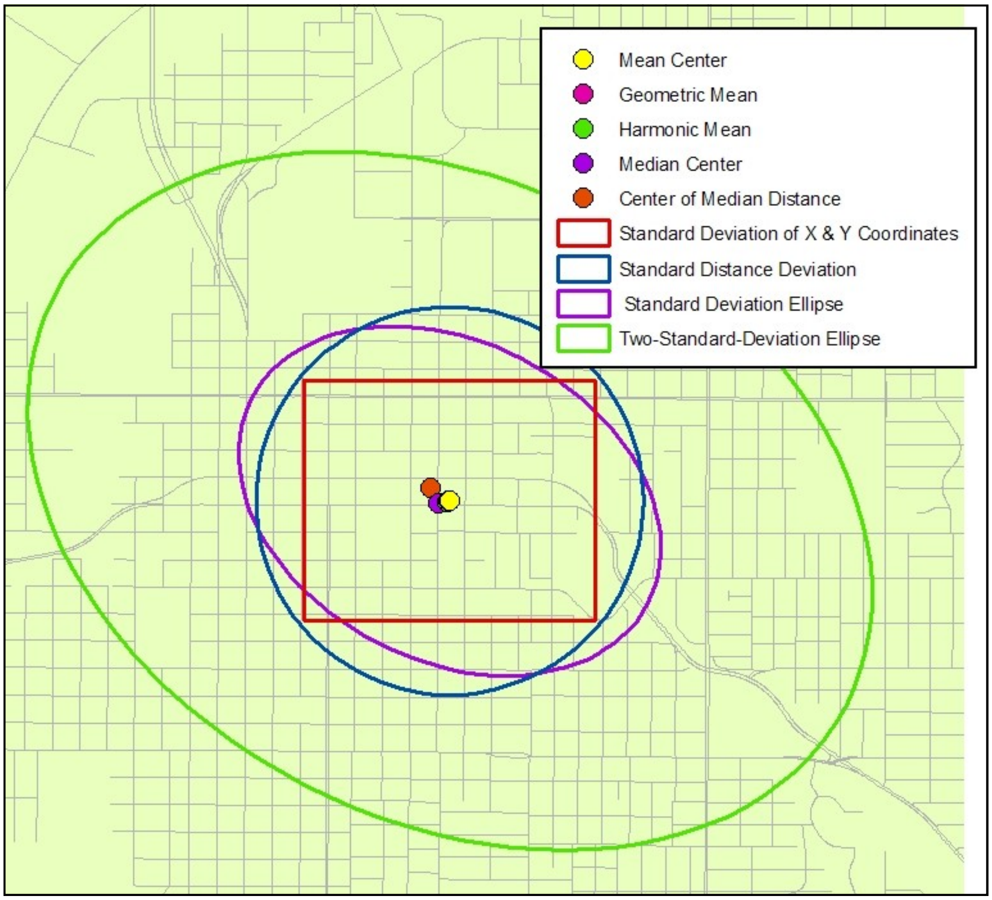


Figure 3-6

You should see that the geometric mean, harmonic mean, mean center, mean center of minimum distance, and median center vary only slightly from each other (you may need to zoom in quite a bit to see all of the points). The measures of dispersion contain various percentages of the total incidents.

As you look at the map, consider the following questions.

1. If the agency decides to place an unmarked patrol car in the area to respond quickly to any alarms or reports of burglaries, where would you suggest that they station it?

2. In what area would you predict the next offense is likely to occur?

3. In what area would you predict the next offense will almost certainly occur?

4. If the agency wanted to suppress the offender by saturating the area with patrol officers, in what area would you recommend they concentrate?

5. If the agency wanted to station "scarecrow cars" in the area to deter the offender, where would you recommend that they station them?

6. If the agency wanted to alert residents about the series, encouraging potential future victims to lock doors and hide valuables, in what area should they call or leave notices?

These are our answers. If yours differ slightly based on your analytical judgment, that's fine; there are no absolute "rights" and "wrongs" here.

1. An unmarked car stationed to respond quickly to incidents would probably best be stationed at one of the mean center calculations (in this case, it doesn't matter much, but the mean center of minimum distance would minimize response times). That would be about halfway along West Harvest Drive.

2. Assuming we are correct that the series doesn't "walk," the next offense is most likely to occur within a single standard deviation of the mean center. The standard deviation "rectangle," the single standard deviation ellipse, and the standard distance deviation all provide good estimates of the densest concentration.

3. The next offense will almost certainly occur within the two standard deviation ellipse or the convex hull polygon, unless the offender does something completely new for his next burglary. These of course have the disadvantage of being larger and thus harder to concentrate resources within.

4. Saturation patrol would best be concentrated in the single standard deviation ellipse, because it encloses a majority of the incidents and conforms best to the street geography in this part of the city.

5. This is a bit of a trick question because it's not fully answerable with these CrimeStat calculations. But note that the standard distance deviation encompasses a neighborhood with few entries and exits. An offender would almost certainly have to pass cruisers stationed at four locations: the intersection of NW 12th St and W Highland Blvd; NW 13th St and W Fletcher Ave; NW 1st St and W Fletcher Ave; and NW 1st St and W Highland Blvd.

6. To reach all potential targets, the two standard deviation ellipse would be the best choice.

3.4 Cautions & Caveats

At this point, you might be thinking, "This is all great, but I could have drawn this by hand and done just as good a job!" You are probably correct. But we would point out the following:

* Your hand drawing wouldn't account for multiple incidents at a single location. CrimeStat does.
* This series has a limited number of points. If analyzing a larger series, or a year's worth of crime, your ability to visually interpret the points would suffer significantly.
* While you may be able to draw a decent centerpoint and ellipse by hand, CrimeStat's calculations are more precise, and there's always virtue in better precision.

There are, however, cases in which the latter statement is not true. If a pattern appears in multiple clusters, as in Figure 3-7, CrimeStat's calculations will be "correct" mathematically but will not accurately represent the pattern's geography. And, as we've seen, if the pattern "walks," the various measures of spatial distribution will describe the pattern's past, but not it's future.

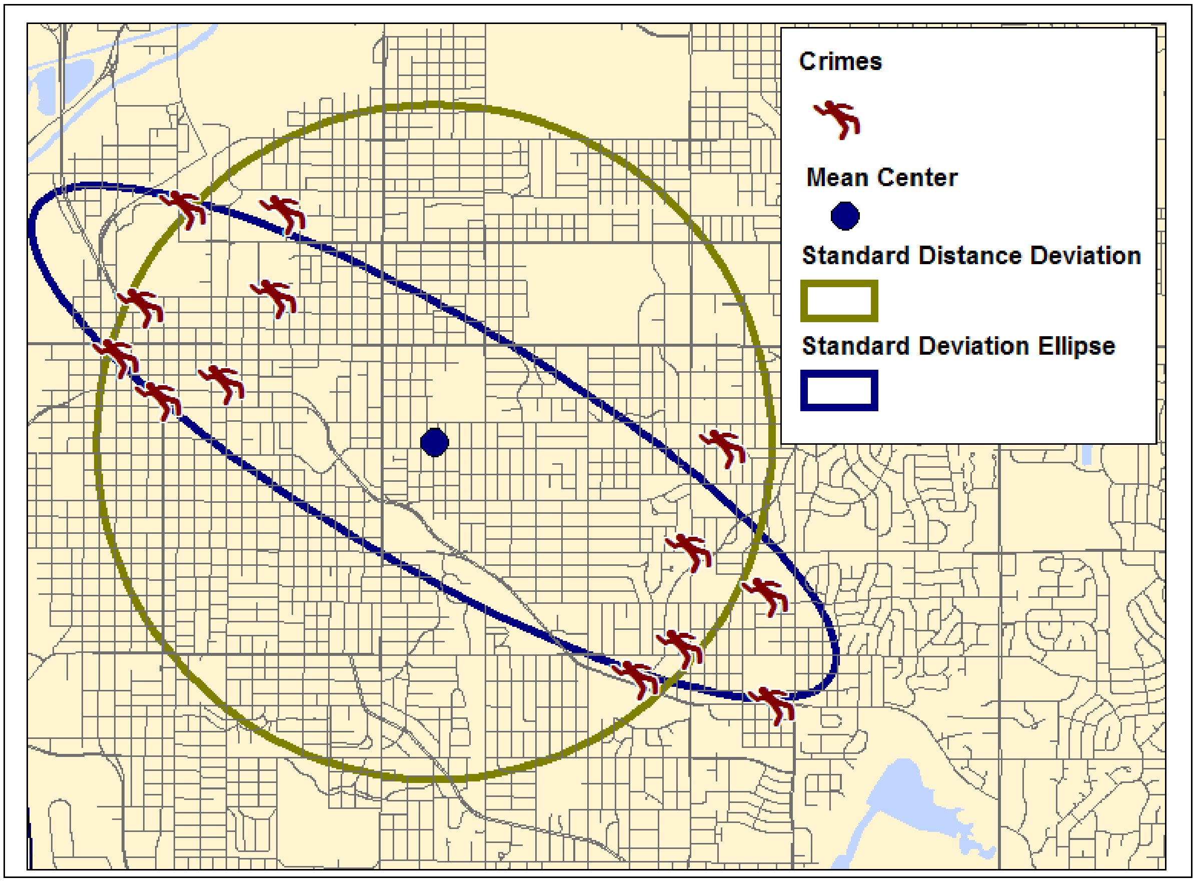
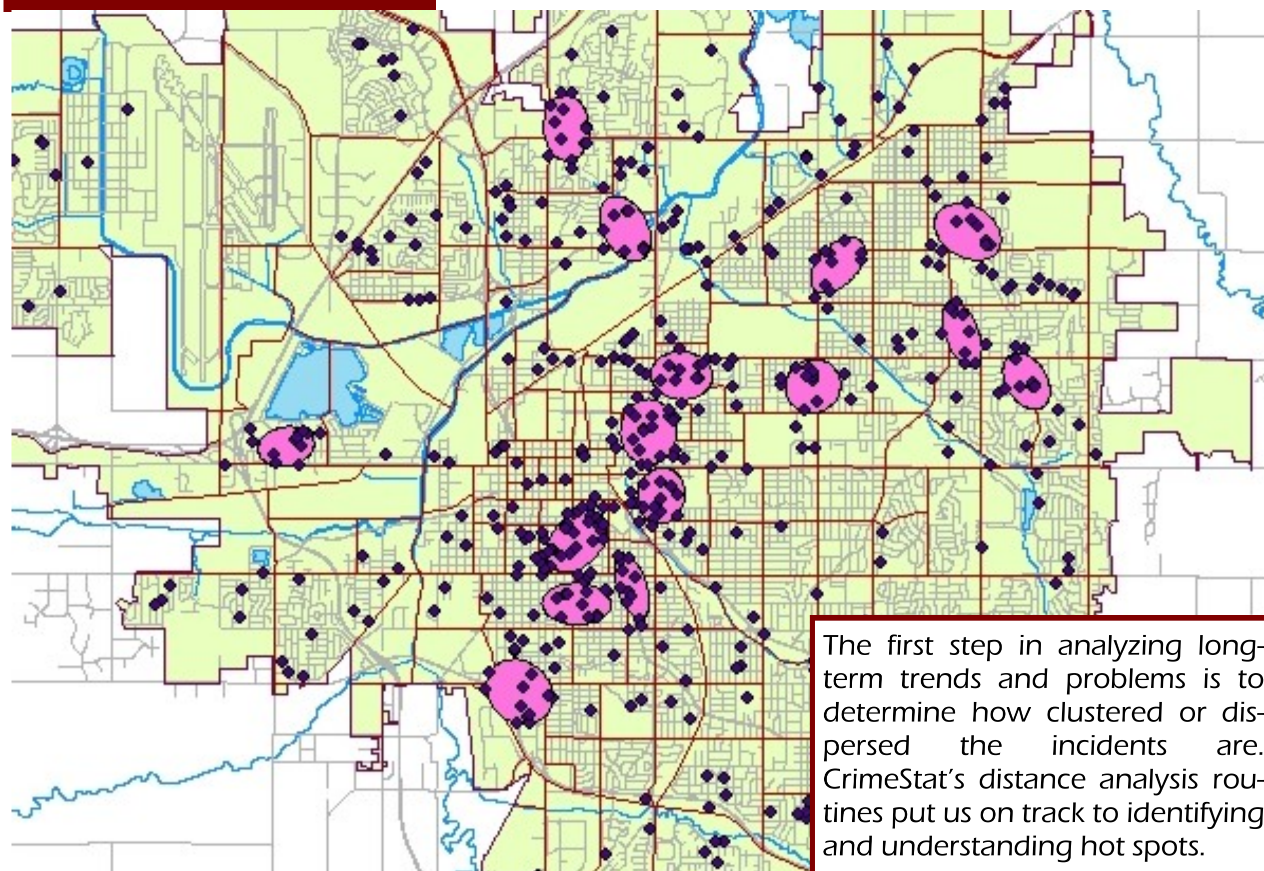


Figure 3-7: an unhelpful spatial distribution. The mean center, standard deviation ellipse, and standard distance deviation circle are technically correct, but they miss the point of the pattern, which is that it appears in two clusters. The analyst in this case would probably want to create a separate dataset for each cluster and calculate the spatial distribution on them separately.

Like all CrimeStat routines, measures of spatial distribution must be used in conjunction with the analyst's own critical thinking, experience, and judgment to produce a operationally-relevant result.

4 Distance Analysis



The first step in analyzing long-term trends and problems is to determine how clustered or dispersed the incidents are. CrimeStat's distance analysis routines put us on track to identifying and understanding hot spots.

In this Chapter:

- Nearest neighbor analysis
- Comparing relative clustering and dispersion for multiple offense types
- Assigning points from one dataset to their nearest neighbor in another dataset

4 Distance Analysis

4.0 Introduction

In this chapter, we turn from short-term crime series to long-term crime trends. We use **distance analysis** techniques to answer questions about the dispersion of incidents, and hot spot analysis to identify areas where crimes concentrate. In doing so, we unleash the full power of the software. While analysts could manually work out certain spatial description calculations, or simply use visual interpretation, the techniques in this chapter would be functionally impossible without a program like CrimeStat.

4.1 Nearest Neighbor Analysis

If you scattered crime randomly across the jurisdiction, probability would produce some small clusters and some wide gaps, but there would be an average distance between events. CrimeStat can compare the actual average distance between points and their nearest neighbors with this “expected” distance for a random distribution. The resulting calculations indicate whether your incidents are significantly clustered or dispersed (Figure 4-1)

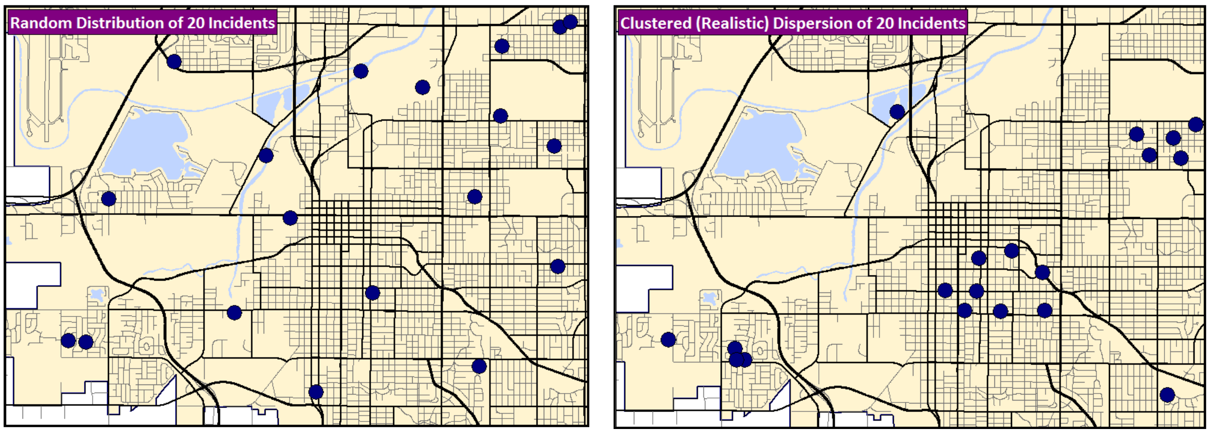


Figure 4-1: A random distribution (left) compared to a significantly clustered distribution (right)

CrimeStat offers two primary measures for distance analysis, both of which have several options and sub-calculations. These are:

1. Nearest neighbor analysis
2. Ripley’s K statistic

Nearest neighbor analysis (NNA) measures the distance of each point to its nearest neighbor, determines the mean distance between neighbors, and compares the mean distance to what would have been expected in a random nearest neighbor distribution. The user can control whether to compare each point to its single nearest neighbor or to run the routine against the second-nearest, third-nearest, and so on. The user can also define whether the distance is ***direct*** (**standard nearest neighbor analysis**), ***indirect*** (**linear nearest neighbor analysis**) or ***based on a network*** (see page 22).

Glossary

Distance Analysis: statistics for describing properties of distances between incidents including nearest neighbor analysis, linear nearest neighbor analysis, and Ripley's K statistic.

Nearest Neighbor Analysis (NNA): measures the distance of each point to its nearest neighbor, determines the mean distance and compares the mean distance to what would have been expected in a random distribution.

NNA produces a calculation called the **nearest neighbor index** (NNI). In the NNI, a score of 1 would indicate absolutely no discrepancy between the expected distances in a random distribution and the measured distances in the actual distribution. Scores lower than 1 indicate that incidents are more clustered than would be expected in a random distribution, and scores higher than 1 indicate the incidents are more dispersed than would be expected in a random distribution. Refer to Figure 4-2 below for a sample output from CrimeStat. Significance levels are offered for the NNI.

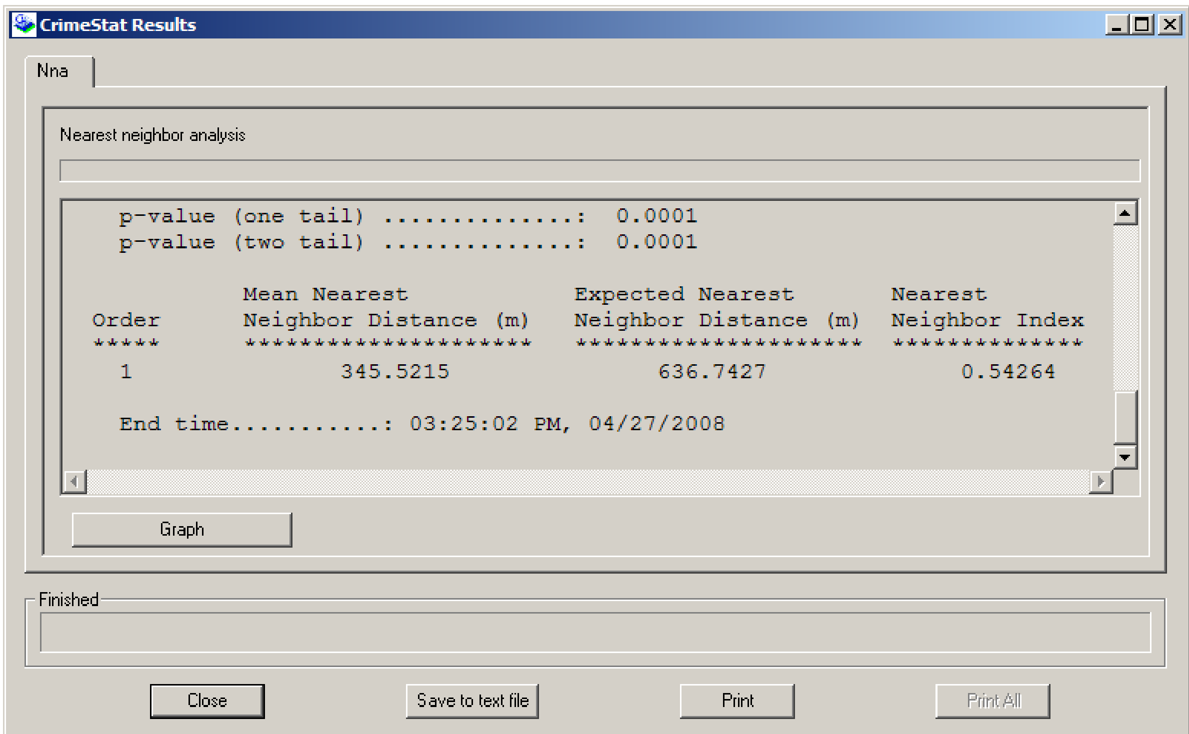


Figure 4-2

It will come as no surprise to most crime analysts that many crime types—perhaps all—show statistically significant degrees of clustering. The average geographic area simply does not provide the conditions necessary for a truly random allocation of incidents. Housebreaks will not occur in locations with no houses, and will naturally be clustered in dense population centers. Business crimes cannot occur where there are no businesses, and will thus be concentrated in commercial areas. "Hot spots" crop up for a variety of reasons all over the place. There are both tactical and strategic uses to knowing the degree of clustering a data set.

The primary value for analysts is to conduct distance analysis for several crimes and compare the results to each other. This will help the analyst determine which offenses are *most* clustered into "hot spots," and which are more spread across the jurisdiction. But because distance analysis has fairly limited utility for crime analysts, the more complicated of the two measures—Ripley's K—is not covered in this workbook.

4.2 Comparing distances for three offenses

In these lessons, we look at the Nearest Neighbor Index for three offense types in Lincoln, Nebraska in 2007: robberies, residential burglaries, and thefts from automobiles. To begin we must set up the data:

Step 1: Open a fresh CrimeStat session. On the "Data setup" tab, click 

, "choose a shapefile", and navigate to **c:\CrimeStat\Data\Lincoln\robberies.shp**. Choose "X" for the X-Coordinate and "Y" for the Y-Coordinate. At the bottom of the screen, specify you are using a "Projected" coordinate system with the data units in "Feet." Your screen should look like Figure 4-3.

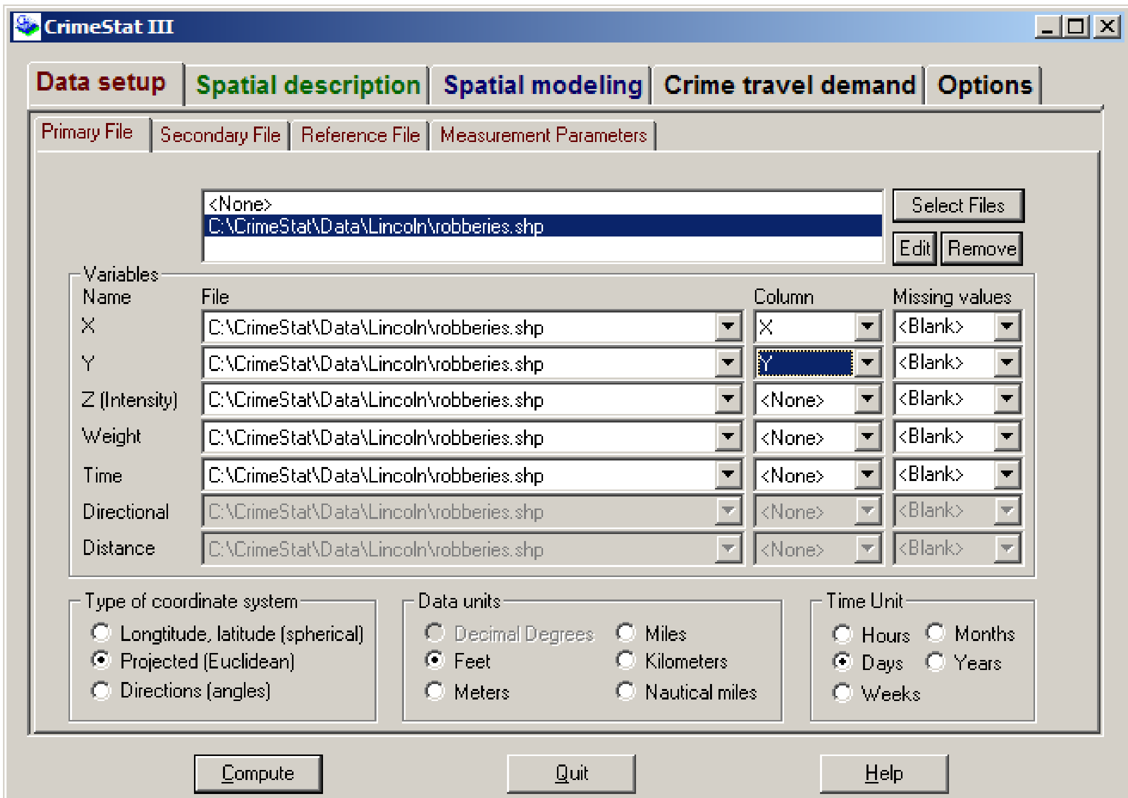


Figure 4-3

Step 2: Click on the "Measurement Parameters" tab. To calculate NNA, CrimeStat needs to know the square mileage and street distance of the jurisdiction. Input the figures you used in Chapter 1 (Area of 88.19 square miles and Length of street network of 1283.61 miles). For now, use a "Direct" distance measurement, as in Figure 4-4 (top of next page).

Glossary

Nearest Neighbor Index (NNI): a calculation produced using Nearest Neighbor Analysis which provides a score that the user can use to determine whether the incidents are clustered more or less than they would be in a random distribution.

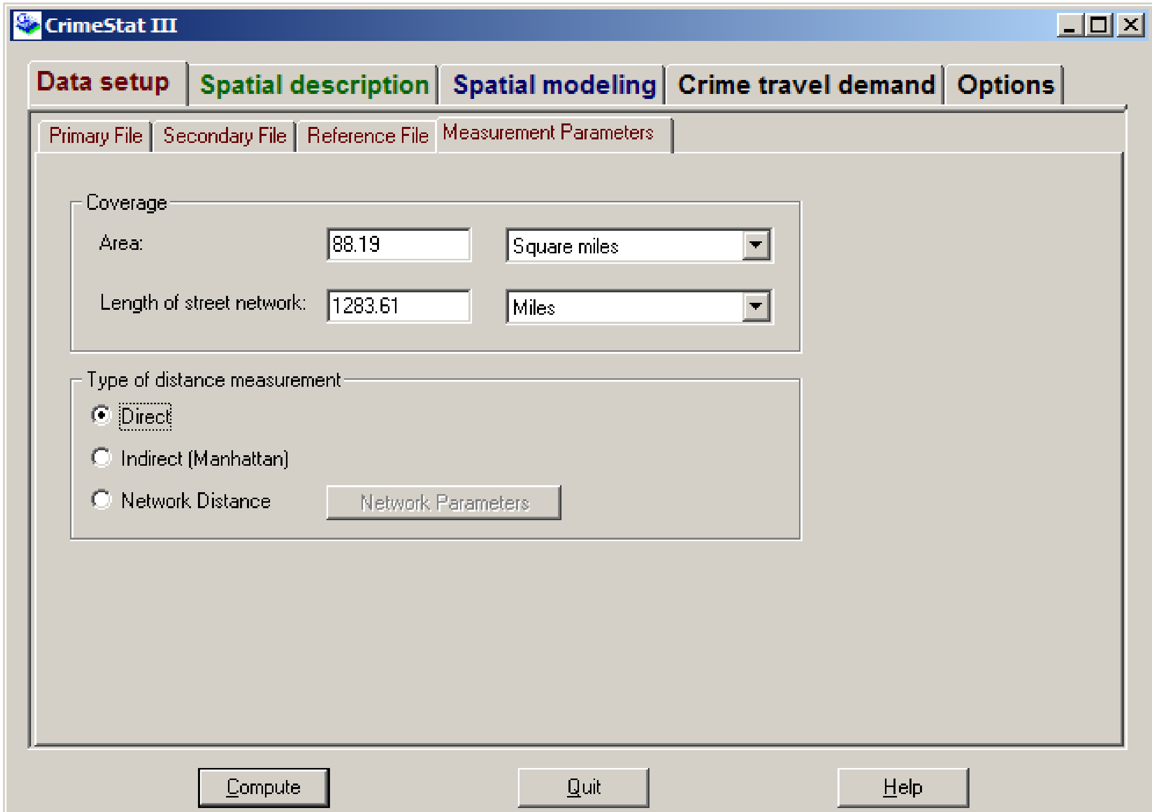


Figure 4-4

Step 3: Click on the "Spatial description" tab and the "Distance Analysis I" sub-tab. Check the "Nearest neighbor analysis (Nna)" box at the top of the screen. Leave the default settings for now and click "Compute."

CrimeStat should ultimately tell you that the expected mean distance between nearest neighbors (in a random distribution) is 1874.1078 feet, the actual average minimum distance between nearest neighbors is 1066.86. This results in an NNI of 0.56926, which indicates the actual distribution is clustered. This result is significant at a 99.99% confidence level (Figure 4-5, next page).

NOTE: Changing the distance measurement from Direct to Indirect will result in a different result, as would changing it to a Network Distance. Of special note is that if you are only calculating the distance to the "first" nearest neighbor, you will want to choose "Direct", however if you want two or more nearest neighbors, you will need to choose either Indirect (Manhattan) or Network Distance.

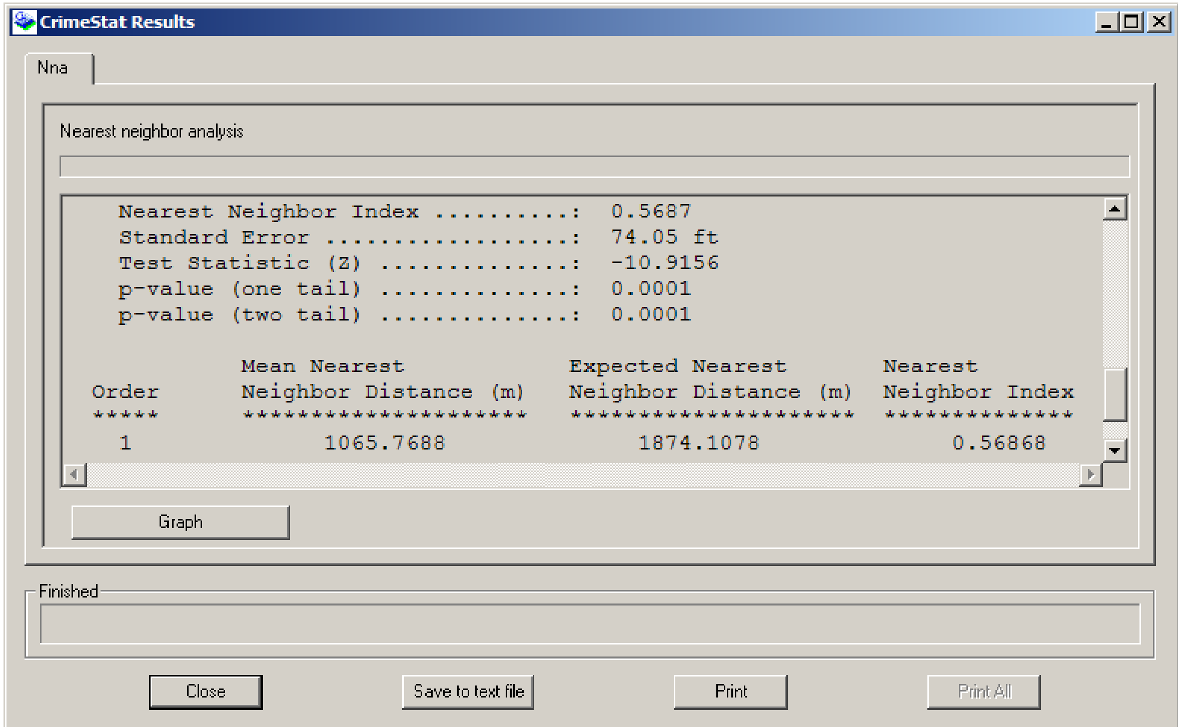


Figure 4-5

Step 4: Repeat Steps 1-3 for the **C:\CrimeStat\Data\Lincoln\resburglaries.shp** and **C:\CrimeStat\Data\Lincoln\theftfromautos.shp** files. The results should be as follows:

Crime	Actual	Expected	NNI
Robberies	1066.8578	1874.1078	0.56926
Residential Burglaries	348.7187	636.7427	0.54766
Thefts from Autos	236.2937	447.2314	0.52835

The results indicate that all three crimes show a significant level of clustering—none are close to a "random" distribution (you will probably get this result no matter what crime you try). But among the three, thefts from autos are the *most* clustered, and robberies are the least. What implications does this knowledge have for you, as a crime analyst, in how you approach your analysis of these crimes? What implications does it have for the types of maps that best represent these crimes?

Step 5: Open each of these layers in your Lincoln ArcView project. Do the results above seem consistent with what you can observe?

This is a bit of a trick question, since although you might be able to give an answer for robberies and perhaps even residential burglaries, the sheer number of thefts from automobiles overwhelms the eye. The crime seems to blanket the city and on visual interpretation shows no clustering at all—and yet it's the most clustered of the three offenses. This is why programs like CrimeStat are necessary.

A few other notes on NNA:

We have been using the default values, which computes a single nearest neighbor. You can change this to any number you like, although it shouldn't be higher than the total number of offenses in your dataset, and the developer of CrimeStat recommends it not be higher than 100 in any case. The result will perform the same calculations (but only a significance test on the first one) for the second, third, fourth, and subsequent nearest neighbors—up to the limit you've specified. There is limited utility for crime analysts in doing this. Refer to Figure 4-6, below.

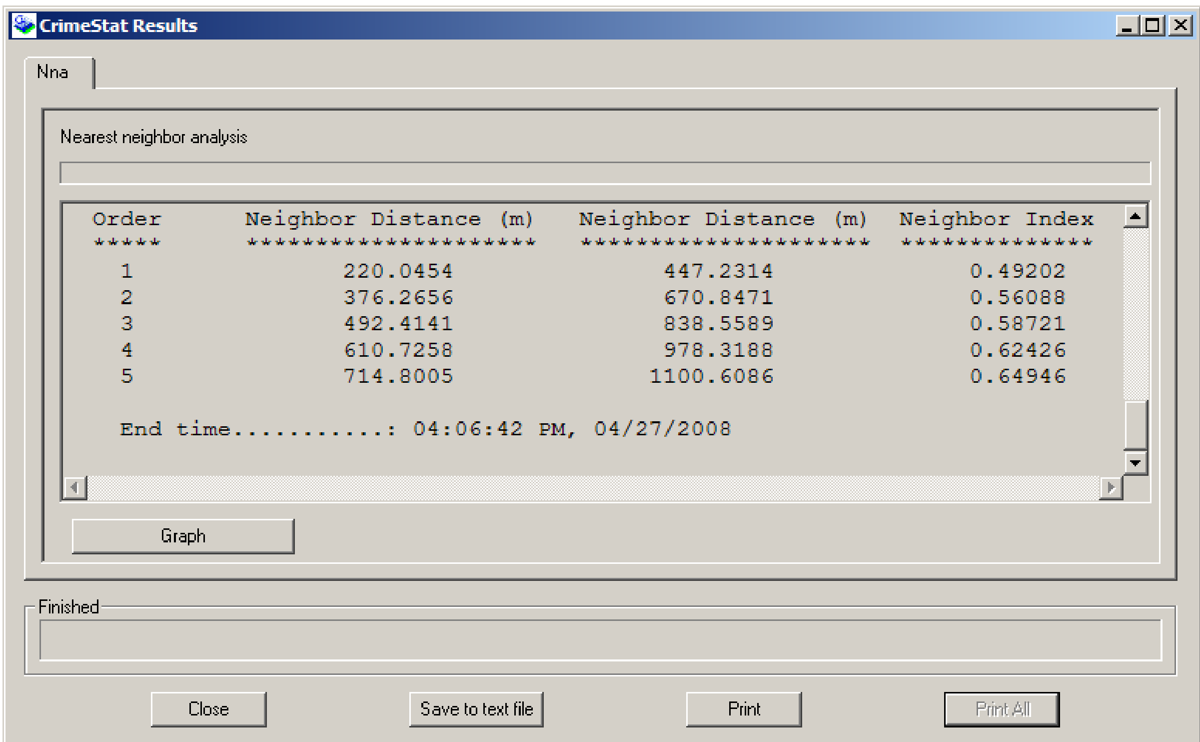


Figure 4-6: thefts from autos computed to the fifth nearest neighbor, indicating that incidents are still fairly clustered at that level.

"Nearest neighbors" of incidents close to the borders of your jurisdiction may, in fact, be in neighboring jurisdictions, in which case NNA overestimates the nearest neighbor distance. CrimeStat allows you to compensate for this "edge effect" with the "Border correction" option. If you use this option, CrimeStat assumes that another incident lurks just on the edge of the jurisdiction's border and calculates the nearest neighbor accordingly (thus probably underestimating the true nearest neighbor distance). However, CrimeStat does not use the real borders of the jurisdiction but instead assumes a rectangular or circular border depending on which option you choose. The analyst will have to decide whether to use border correction. An analyst working for a jurisdiction in a large metropolitan area may in fact be surrounded by similar crimes, and it would make sense to compensate for this problem. Lincoln, Nebraska is surrounded by smaller jurisdictions and open prairie, so it would make little sense here.

All of our examples have shown crimes that are more clustered than expected by chance. Far rarer are crimes more *dispersed* than expected by chance, but it may show up occasionally. If you were to use NNA for a crime series, for instance, an NNI of greater than 1 might indicate that the offender was deliberately spacing his crimes evenly to avoid detection.

4.3 Assigning Primary Points to Secondary Points

The final option on the "Distance Analysis I" screen will assign points in the primary data file to points in a secondary data file, then sum these assignments. There are two ways of doing so:

1. *Nearest neighbor assignment*. This method assigns each point in the primary file to the nearest point in the secondary file. For instance, if the primary file contained incidents of juvenile disorder, and the secondary file contained locations of schools, you could find out which schools were closest to the most disorder incidents. Keep in mind that the "nearest neighbor" may vary depending on the measurement parameter you choose on the "Data Setup" screen.

2. *Point-in-polygon assignment*. In this routine, CrimeStat interprets the geography of a polygon file (such as police reporting areas) and calculates how many points fall within each polygon. You can use only ArcGIS shapefiles as polygon files for this routine.

The distinction between the two is important. In the simplified diagram in Figure 4-7, showing precinct boundaries and locations of station houses:

Point 1 is in Precinct 2's boundaries and is also closest to the Precinct 2 station house.
Point 2 is in Precinct 1's boundaries but is actually closest to the Precinct 3 station house.

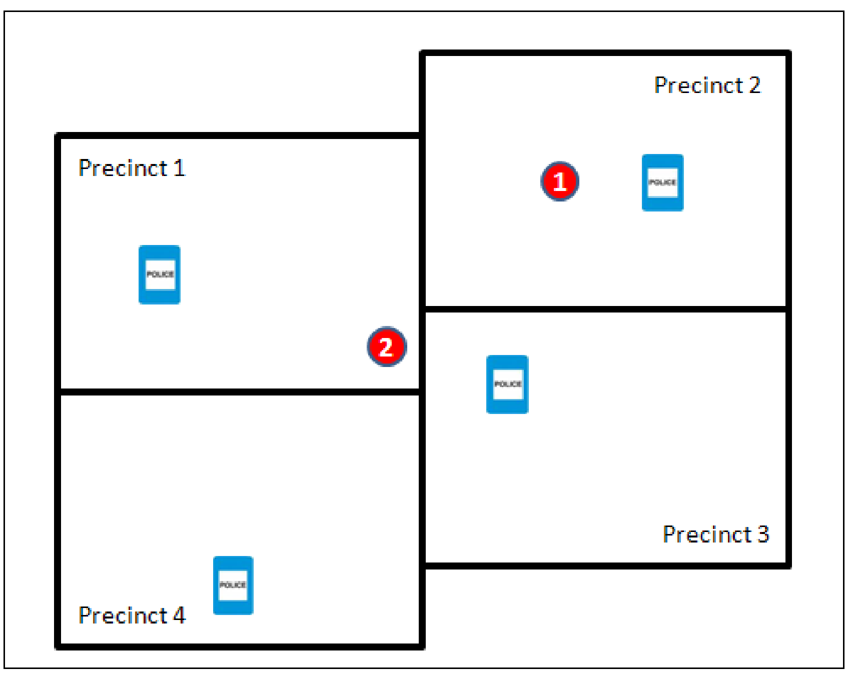


Figure 4-7

In other words, nearest neighbor assignment would assign Point 2 to Precinct 3, whereas point-in-polygon assignment would assign it to Precinct 1.

Most GIS systems will perform point-in-polygon assignments (both ArcGIS and MapInfo do it quite easily), but very few perform nearest neighbor assignments without special scripts. So for the purposes of this lesson, we will focus on nearest neighbor assignment.

In this lesson, we assume that Lincoln has a growing problem with residential burglaries committed in the afternoon. Research into the problem has suggested that the offenders are students on their way home from school. The police department wants to know which schools are closest to the most burglaries.

Step 1: Return to the "Data setup" screen and "Remove" any existing files. Choose "Select files" and find the **C:\CrimeStat\Data\Lincoln\afternoonhousebreaks.dbf** file. Set the X co-ordinate as "X" and the Y coordinate as "Y." The data is projected in feet.

Step 2: Click on the "Secondary file" tab and select the **C:\CrimeStat\Data\Lincoln\schools.dbf** file. Set the X coordinate as "CentroidX" and the Y coordinate as "CentroidY."

Step 3: Click on the "Spatial description" tab and the "Distance Analysis I" sub-tab. Check the "Assign primary points to secondary points" box. Click on the "Save result to" box, navigate to **C:\CrimeStat\Data\Lincoln** and name the file **schoolcounts.dbf**. Leave the other options at their defaults. Click "Compute."

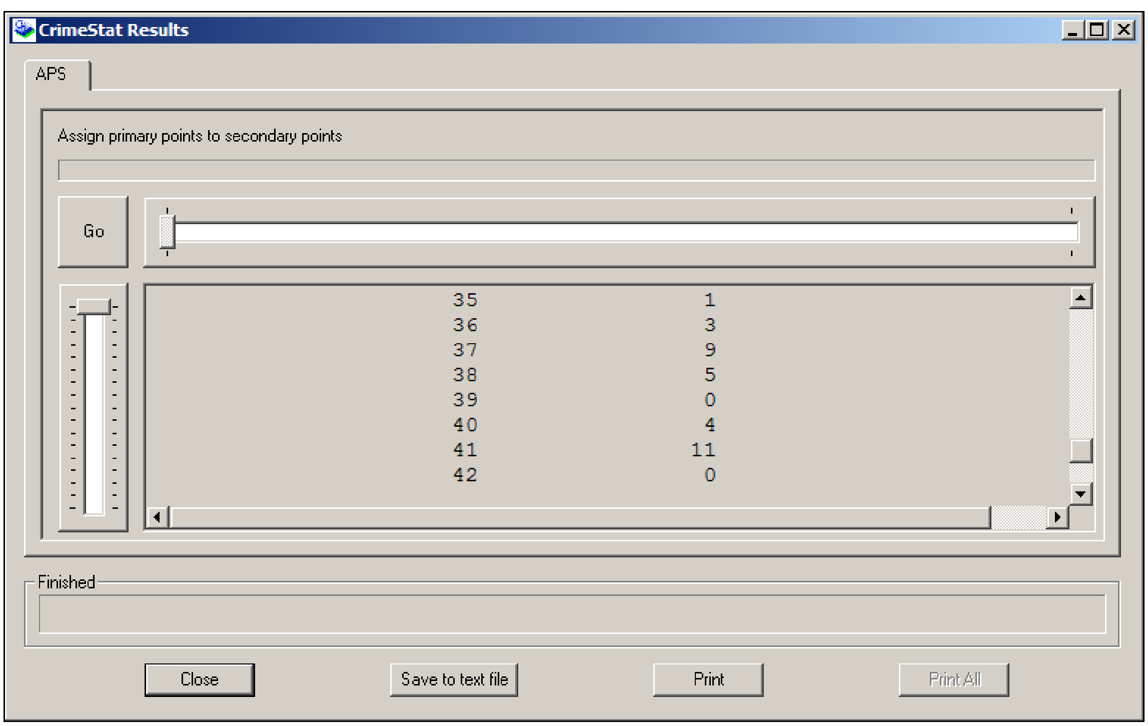


Figure 4-8

Unfortunately, CrimeStat's output table is not immediately helpful. Instead of showing us something useful from our secondary file, such as the school's name, it simply gives us the point number. We can see that Point 41 has the highest number of afternoon burglaries that are nearest to it (11), but we need to return to the DBF file to determine that Point 41 corresponds to the McPhee Elementary School.

However, when we open the **schoolcounts.dbf** file (see Figure 4-9), we see all of our original data, plus the new "FREQ" field, by which we can sort:

Microsoft Excel - schoolcounts.dbf

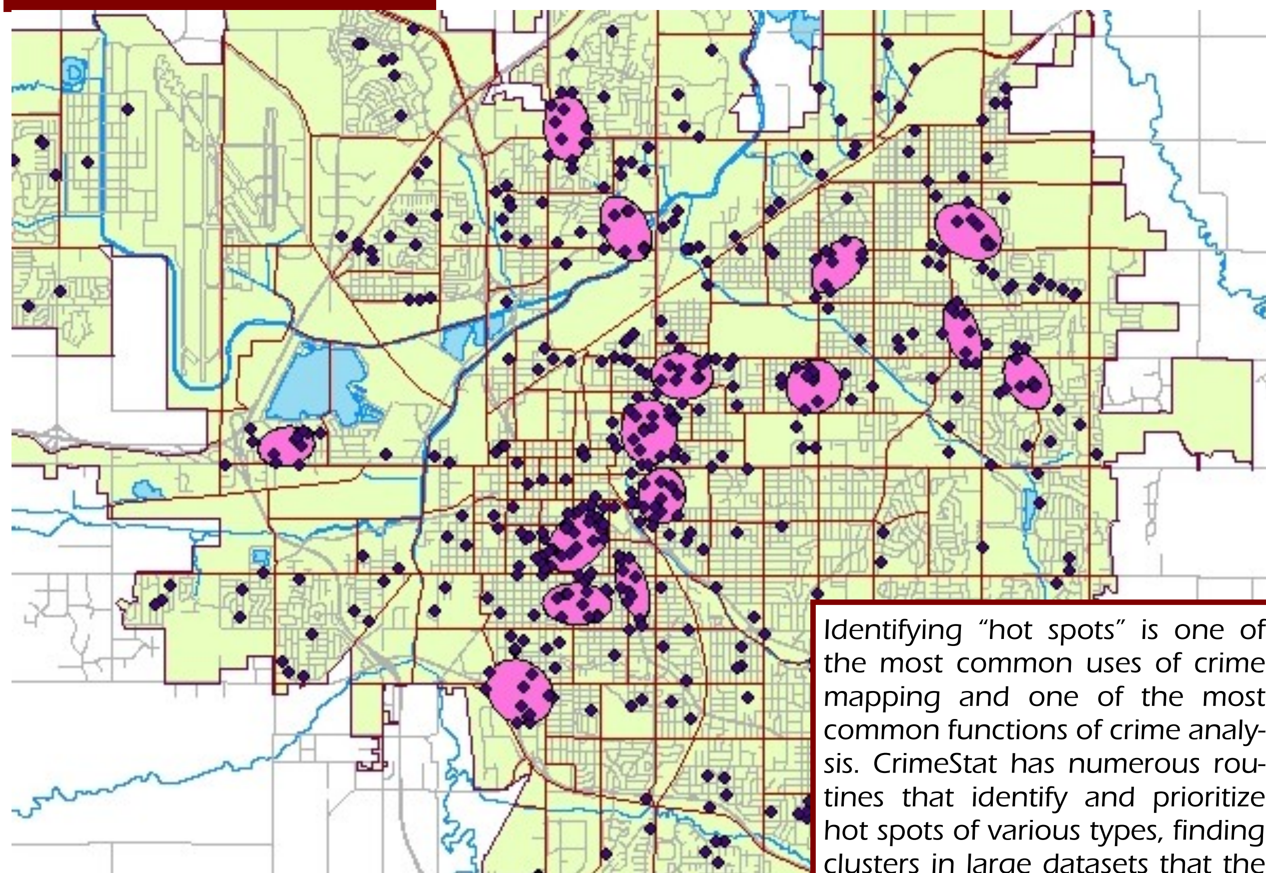
G2 | 11

	A	B	C	D	E	F	G
1	NAME	LINK	SHAPE_LENG	SHAPE_AREA	CENTROIDX	CENTROIDY	FREQ
2	McPhee Elementary School	Schools/McPhee.pdf	1200.055586	90008.143947	161021.039530	202313.451492	11
3	Lincoln High School	Schools/Lincoln_High.pdf	4432.813155	1021924.183620	163893.404514	202866.012584	9
4	Everett Elementary School	Schools/Everett.pdf	1793.711504	190814.622981	159517.692285	200702.767667	8
5	Huntington Elementary School	Schools/Huntington.pdf	1795.093123	198996.521580	173118.316667	215661.930181	7
6	Sacred Heart School		250.001521	3750.015752	167296.418684	207286.501316	6
7	Elliott Elementary School	Schools/Elliott.pdf	1648.986600	169504.563638	164529.167333	204718.680059	5
8	St.Theresa's School		869.651864	40476.747472	169345.340617	203076.874977	5
9	Fredstrom Elementary School	Schools/Fredstrom.pdf	3213.672768	514898.106942	151439.185615	225369.856943	4
10	Arnold Elementary School	Schools/Arnold.pdf	3739.111988	544746.141324	135023.040177	220014.777814	4
11	Belmont Elementary School	Schools/Belmont.pdf	2021.227059	247593.498331	160170.285988	217338.984850	4
12	Riley Elementary School	Schools/Riley.pdf	2704.534835	429507.287501	174297.113662	209308.747885	4
13	Lakeview Elementary School	Schools/Lakeview.pdf	2372.012227	318654.499131	150098.431058	206406.849463	4
14	Park Middle School	Schools/Park.pdf	2267.039615	296633.585132	157551.972080	202287.093693	4
15	Maxey Elementary School	Schools/Maxey.pdf	3326.830465	653995.853420	183697.003494	186088.479489	4
16	Northstar High School	Schools/Northstar_High.pdf	7815.901232	3218743.496210	167551.422453	225201.230541	3

Figure 4-9

The assignment routine is a utility, rather than analytical one, and is aimed at aggregating data for use in other routines (e.g., Moran's "I", Moran Correlogram, Anselin's local Moran). However, knowing a little about how distance analysis works helps us better understand what happens in Chapters 5 and 6, with hot spot analysis and kernel density estimation

5 Hot Spot Analysis



Identifying "hot spots" is one of the most common uses of crime mapping and one of the most common functions of crime analysis. CrimeStat has numerous routines that identify and prioritize hot spots of various types, finding clusters in large datasets that the eye would miss.

In this Chapter:

- Summary of different hot spot routines
- Mode and fuzzy mode
- Nearest neighbor hierarchical spatial clustering
- Spatial and Temporal Analysis of Crime

5 Hot Spot Analysis

5.0 Introduction

In Chapter 4, we looked at some routines that tell us how "clustered" our incidents are, compared to what we would expect in a random distribution. We also saw that this statistic offered limited utility to the analyst. In this chapter, we look at routines that do something a lot more useful: actually identify hot spots.

A **hot spot** is a spatial concentration of crime. Crime analysts and criminologists have rarely tried to come up with a definition more precise than that. For the purpose of these routines, we might define a hot spot as *a geographic area representing a small percentage of the study area which contains a high percentage of the studied phenomenon.* The scale of the "spot" depends on the scale of the study area: in a city, a hot spot might be an address, building, complex, parking lot, block, park, or some other relatively small area. In a state, however, a hot spot might be an entire city.

CrimeStat has a number of routines to study different types of hot spots, although they all meet this basic definition. The routines are:

Mode: identifies the geographic coordinates with the highest number of incidents

Fuzzy Mode: identifies the geographic coordinates, plus a user-specified surround ing radius, with the highest number of incidents

Nearest-Neighbor Hierarchical Spatial Clustering: builds on the nearest neighbor analysis (NNA) that we saw in Chapter 4 by identifying clusters of incidents

Spatial & Temporal Analysis of Crime (STAC): an alternate means of identifying clusters by "scanning" the points and overlaying circles on the map until the densest concentrations are identified

K-Means Clustering: the user specifies the number of clusters and CrimeStat positions them based on the density of incidents

Aneslin's Local Moran statistic: compares geographic zones to their larger neighborhoods and identifies those that are unusually high or low

Kernel Density Interpolation: a spatial modeling technique discussed in Chapter 7

The techniques are organized into two screens: "Hot Spot Analysis I" and "Hot Spot Analysis II." The developer of CrimeStat explains that this division is simply for space reasons rather than any inherent separation between the sets of techniques.

In these lessons, we will cover the first four types of hot spot identification. We will summarize all of them at the end of the chapter.

5.1 Mode and Fuzzy Mode

Mode is the simplest of the various hot spot techniques. It just counts the number of incidents at the same coordinates.

In real life, even two incidents rarely share the exact same coordinates. For example, if a shopping mall parking lot has 20 thefts from vehicles, the likelihood of two of them happening at the same precise point in the parking lot is very small. However, in the police records management system, all 20 incidents will probably have the same address, which will geocode to the same point on the mall property. Hence, all 20 will share the same geographic coordinates even though they didn't actually happen at that specific point. Mode hot spot identification, in short, takes advantage of the inherent inaccuracy involved in police records management and GIS.

We'll use mode to identify the top hot spots for thefts from automobiles in Lincoln, Nebraska.

Step 1: Start a new CrimeStat session. On the "Data Setup" screen and the "Primary File" subscreen, click "Select Files" and add **c:\CrimeStat\Data\Lincoln\theftsfromautos.shp**. Assign the X coordinate to "X" and the Y coordinate to "Y." The coordinate system is "Projected" with data units in "Feet." (You may still have this all set up from the previous lesson.)

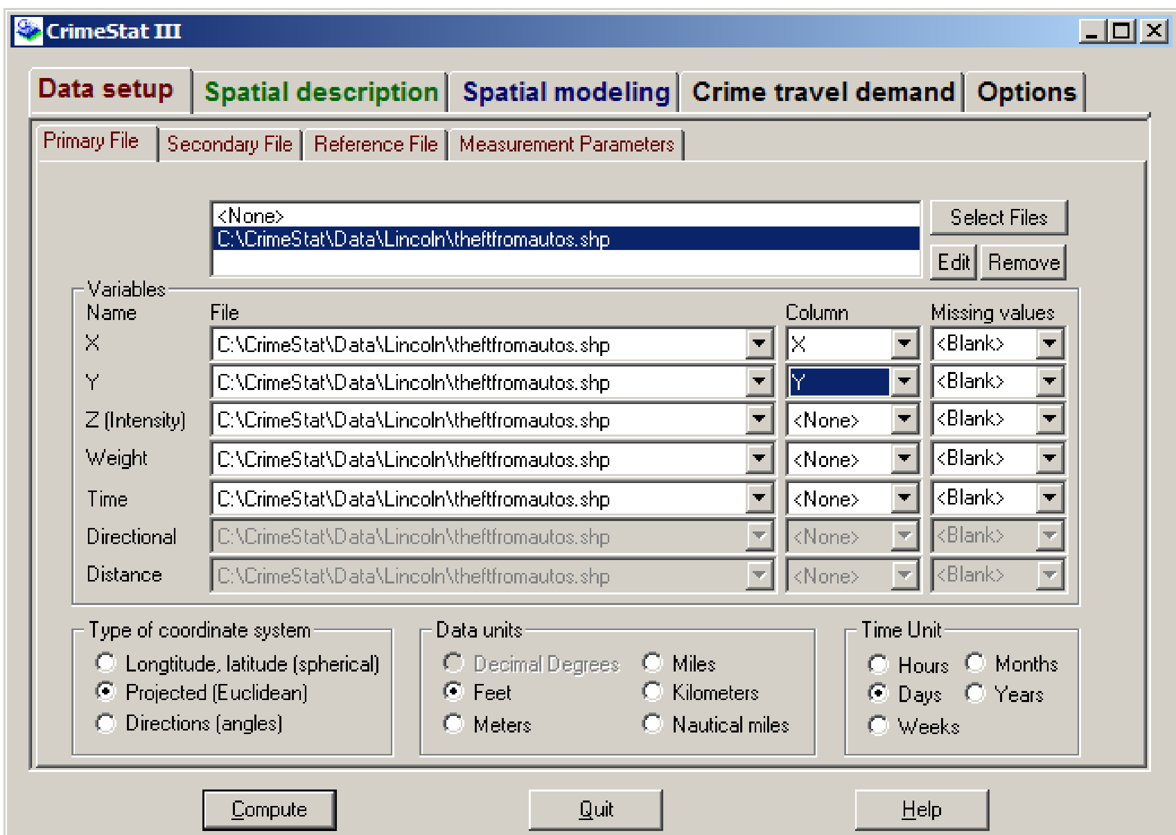


Figure 5-1: the data setup screen for thefts from automobiles in Lincoln

Step 2: Click on the "Spatial description" tab and then the "Hot Spot Analysis I" sub-tab. Check the "Mode" box and note there are no options. (Check your Measurement Parameters and make sure you have the Lincoln data entered and "Direct" chosen for the type of distance measurement). Click "Compute."

CrimeStat produces a simple list of X/Y coordinates, starting with the top 45, and how many incidents are at each pair of coordinates. In order to visualize these incidents on our map, we would have to use the "Save result to..." option to save them to DBF format and import them to our GIS, but we will save that for the fuzzy mode.

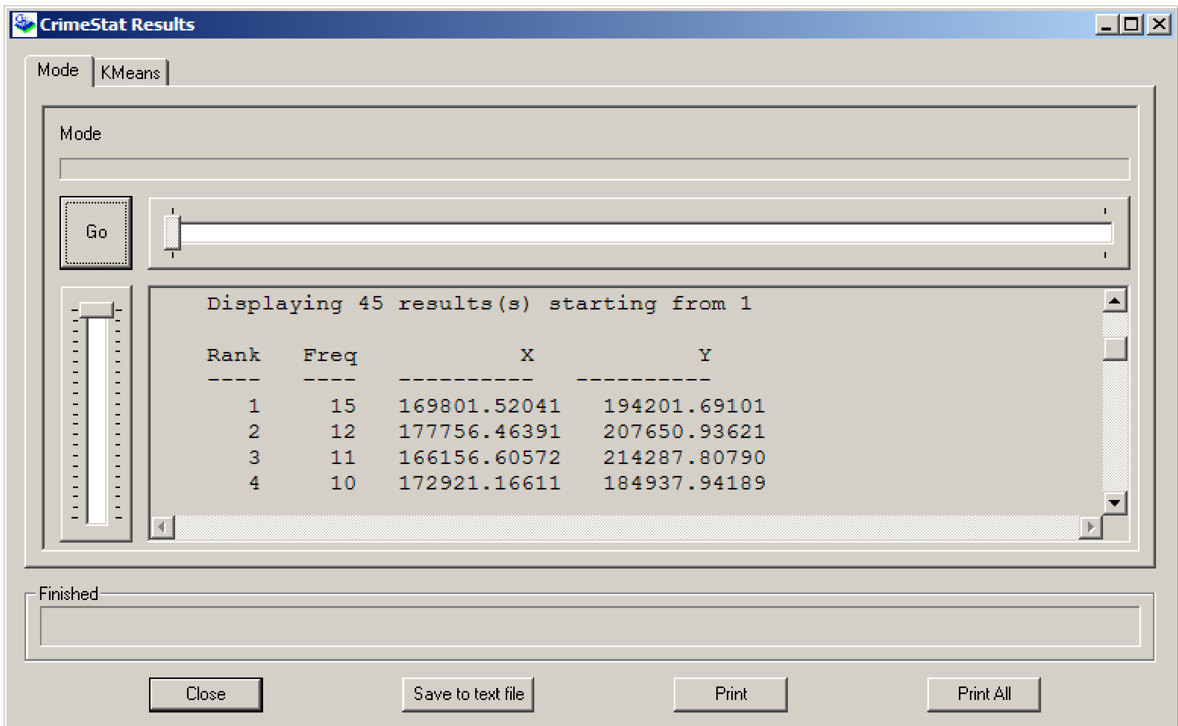
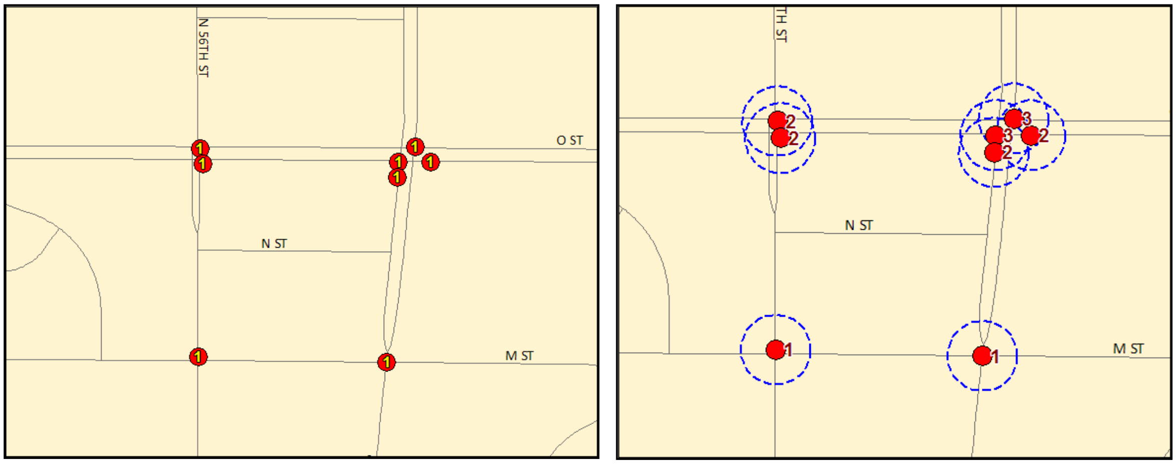


Figure 5-2: the result of the mode hot spot identification for thefts from automobiles

Analysts performing the steps to this point may have already realized that there a number of other ways to achieve the same result using Access queries, Excel Pivot Tables, or tools already inherent in their GIS systems. However, it's nicely output as a DBF file in the routine, which is then easily imported into a GIS program. Also, the decreasing order of the tabular count is made explicit in the output table. With other applications, you'd have to first aggregate the data and then sort it to get the same results, not to mention then importing it into CrimeStat. With CrimeStat, it's a two-step process whereas with other applications there will be more steps.

The *fuzzy mode* allows you to specify a search radius around each point and includes the incidents within that radius in the count. It is the only viable option for agencies that are achieving ultra-accurate geocoding using GPS capture or digitizing (using these methods, two points will almost never have the same coordinate pairs). It's also beneficial for analysts who want to group nearby points into a single point—for instance, if different stores at a shopping mall have different addresses, but they are so close together as to be regarded as essentially the same point.



Figures 5-3 and 5-4: accidents at several intersections. The agency has been ultra-accurate in its geocoding, assigning the accidents to the specific points at the intersections where they occur. The mode method (left) would therefore count each point only once, whereas the fuzzy mode method (right) aggregates them based on user-specified radiuses.

Step 3: Leave the data setup the same, but this time un-check "Mode" and check "Fuzzy Mode." Specify a search radius of 500 feet. Click the "Save result to…" box and save it as a DBF file to **c:\CrimeStat\Data\Lincoln\HS.dbf** (CrimeStat will automatically add an "FMode" to this when it saves). Click "Compute."

Where before your top hot spot had 15 incidents now the top hot spot has 28.

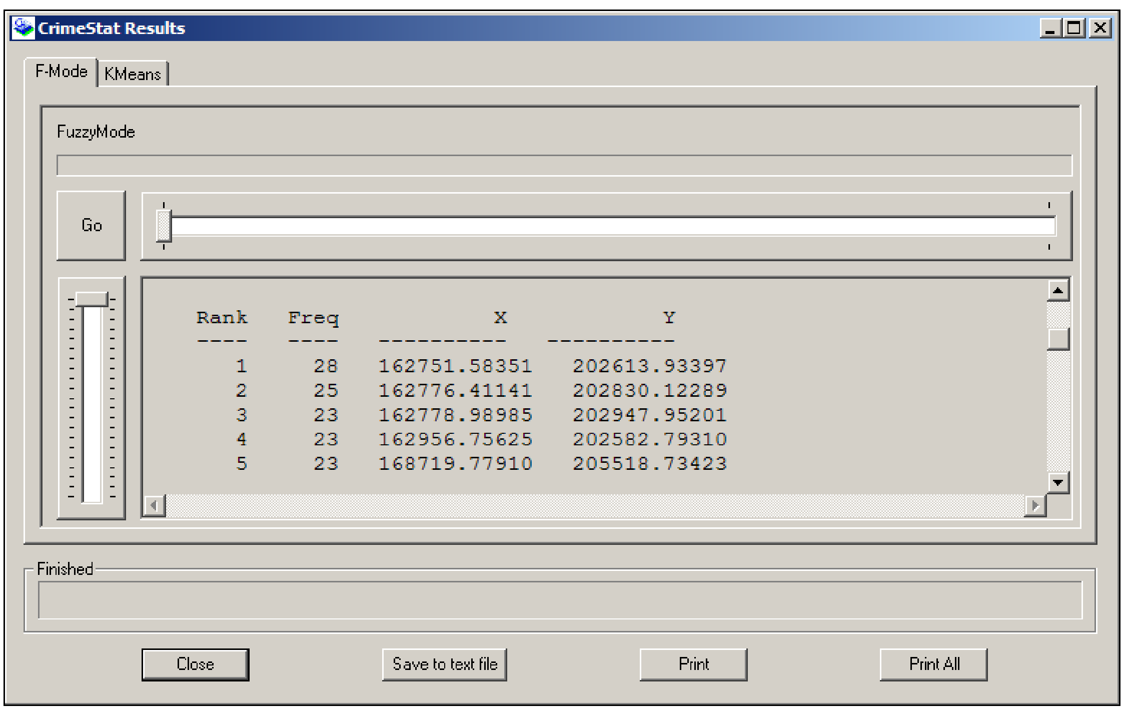
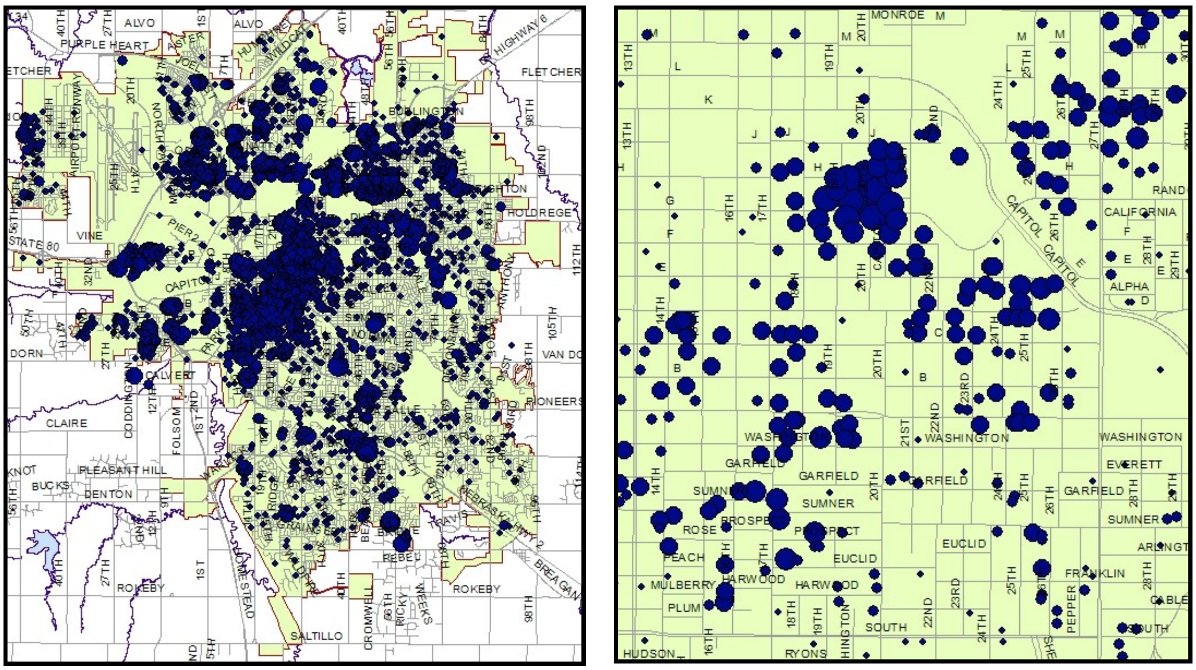


Figure 5-5: the result of a fuzzy mode analysis on thefts from automobiles

Step 4: Switch to your Lincoln ArcView project, open the DBF you just created. Create a proportional symbol map based on its "Freq" field. Does it help make sense of the large number of points in front of you?

Your map ought to make one disadvantage of the fuzzy mode method apparent. Because CrimeStat calculates a radius and count for each point, points close together will fall in one another's radiuses. In areas of very dense concentration, this will result in multiple hot spots that are all counting each other. Still, this method works well to identify the single top hot spot. In Lincoln, for thefts from automobiles, this turns out to be around the area of South 20th Street and G Street.



Figures 5-6 and 5-7: Figure 5-6 shows a citywide view of Lincoln with the hot spots identified with the dots appearing to almost or completely overlay other dots using the fuzzy mode method. Figure 5-7 shows a zoomed image of the identified hot spot area from Figure 5-6.

5.2 Nearest Neighbor Hierarchical Spatial Clustering

In Chapter 4, we used nearest neighbor analysis (NNA) to determine if a particular crime was more clustered than might be expected by random chance. CrimeStat indicated, unsurprisingly, that all three crimes we tested showed clustering. **Nearest neighbor hierarchical clustering (NNH)** takes this analysis to the next logical level by actually identifying those clusters.

At its default settings, CrimeStat compares the distance between pairs of points to the distance expected in a random distribution of points in the jurisdiction's area, and it clusters those groups of pairs that are unusually close together. This creates a number of "first-order" clusters. CrimeStat then conducts the analysis again on the first-order clusters, and circles *clusters* that are unusually close together, creating "second-order" clusters.

It continues establishing more levels of clusters (hence, the "hierarchical" in the term) until it can no longer find any clusters. In practice, the routine often stops after identifying second-level or third-level clusters.

CrimeStat will create both standard deviation ellipses and convex hulls (see Chapter 3) around the hot spots it identifies. You can save these in the GIS format of your choice and import them in.

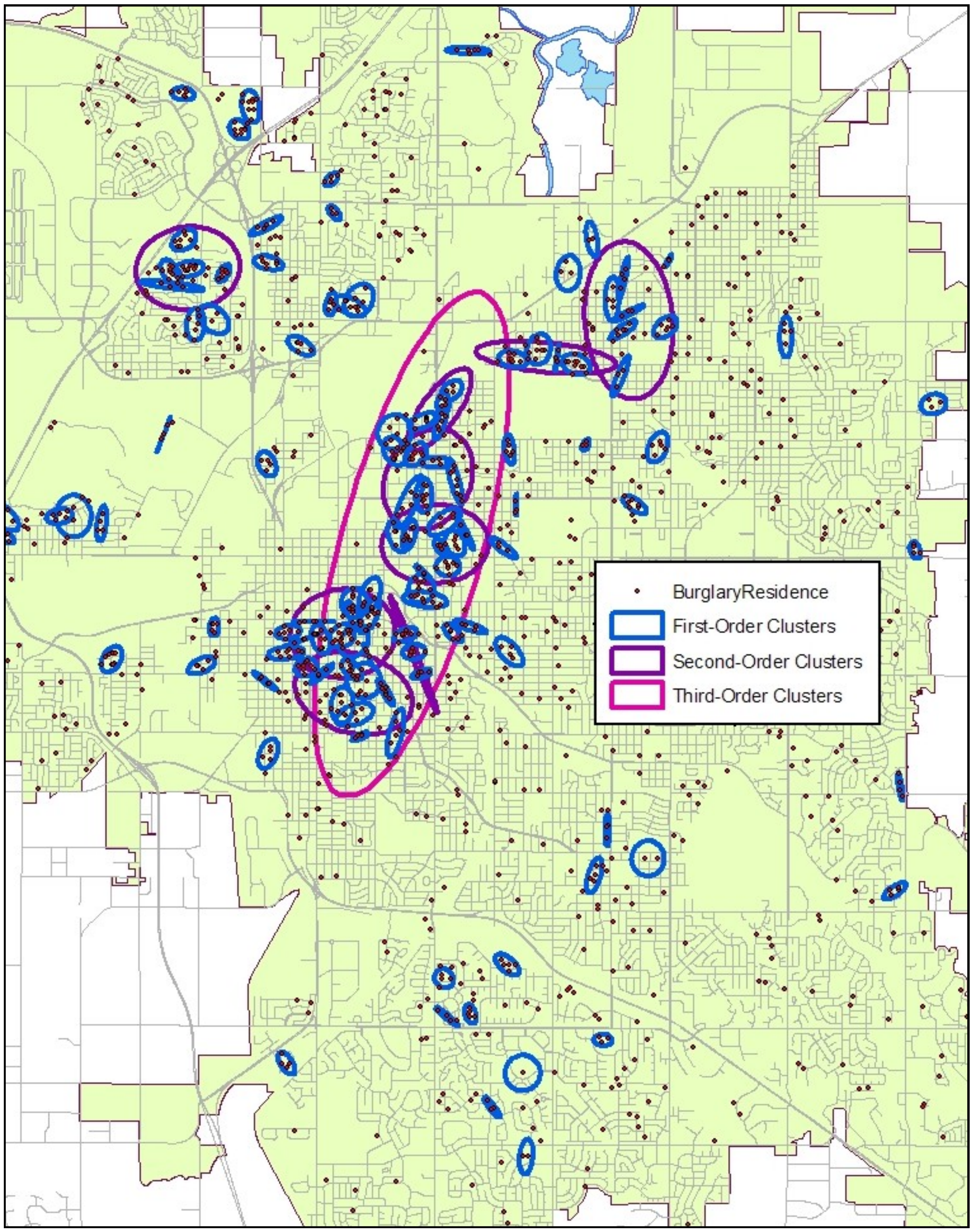


Figure 5-8: Residential burglaries in Lincoln with first, second, and third-level clusters identified by standard deviation ellipses.

There are a number of options that the user can tweak when running NNH.

* Instead of basing the threshold distance on the expectations of a random distribution, the user can specify a *fixed distance*. Although it looks like a minor setting, switching to a user-defined threshold substantially alters the nature of the technique: instead of an objective measure based on probability, it becomes a subjective measure based on the analyst's own judgment. You usually have to play with different settings before you find one that works.

* Even if two points are related in a "cluster," an analyst probably won't be interested in seeing that cluster among thousands of data points. The *minimum points per cluster* option allows you to reduce the number of identified clusters by specifying a minimum number of incidents within each one. The default is set at 10.

* By default, the routine groups pairs of points that have less than a 50% chance of being randomly allocated. In other words, there is a 50% chance that the points are indeed clustered and not a statistical fluke. By adjusting the *search radius bar*, you can adjust the threshold distance and therefore the associated probability (this works only for the "Random NN distance" option). At the furthest position to the left, CrimeStat uses the smallest distance but offers a 99.999% confidence that the pair of points are truly clustered; at the furthest position to the right, it uses the largest distance but offers only a 0.1% confidence that the pair are actually clustered. The sliding scale points represent multiples (to the right) and fractions (to the left) of the random nearest neighbor distance. Clearly, to the "right" is applying a looser grouping distance whereas to the "left" is applying a tighter one. But, it's all based on the random nearest neighbor distance. This is applicable to pairs of points. However, a cluster is usually more than two points so that the likelihood of, say, three points being clustered is much smaller than for a pair of points. In practice, the "minimum number of points" setting greatly reduces the likelihood that CrimeStat will identify "false positives" in its NNH analysis. We recommend leaving the bar at its default setting.

* If you're outputting your results as ellipses, you'll need to specify the *number of standard deviations for the ellipses*. A single standard deviation is the default, and the norm, but it makes small ellipses that can be hard to view at a small map scale. On the other hand, two standard deviations tends to exaggerate the size of the hot spot. We recommend leaving it at one unless the ellipses are too hard to see, and increasing it to 1.5 if so.

* You may also consider outputting a convex hull versus an ellipse, as the convex hull has a greater accuracy. The convex hull has a much higher density than the ellipse. Even though the ellipses are more elegant graphical objects, the convex hulls are actually more accurate than the ellipses because the latter are abstract objects whereas the hulls are defined by the data.

In general, the official CrimeStat manual warns, "the user may have to experiment with several runs to get a solution that appears right."

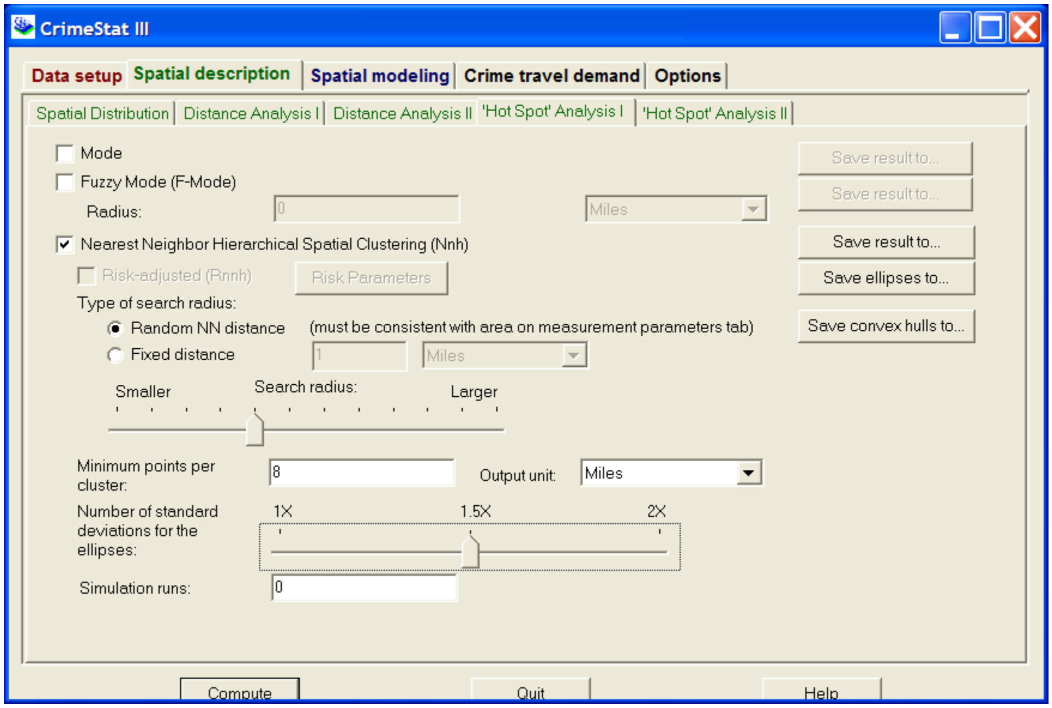


Figure 5-9: the Hot Spot Analysis I screen with the various NNH options.

Step 5: Go back to the “Data setup” screen and click on the “Measurement Parameters” sub-tab. Enter 88.19 square miles for the area and 1283.61 miles for the street network. Choose an **indirect distance** measure.

Step 6: Return to the “Spatial description” tab and the “Hot Spot Analysis I” sub-tab. Uncheck the Fuzzy Mode and check the “Nearest-Neighbor Hierarchical Spatial Clustering” box. Leave it at random NN distance, but adjust the minimum number of points per cluster to 8. Adjust the size of the ellipses to 1.5 standard deviations.

Step 7: Click the “Save ellipses to…” button, choose ArcView SHP format, and browse to **c:\CrimeStat\Data\Lincoln**. Name the file “LFA” (CrimeStat will automatically tag it with further descriptors). “Save” and “OK” these screens.

Step 8: Click on the “Save convex hulls to…” button, choose ArcView SHP format, and browse to **c:\CrimeStat\Data\Lincoln**. Name the file “LFA” (CrimeStat will automatically tag it with further descriptors). “Save” and “OK” these screens.

Step 9: Click “Compute.” CrimeStat should create 36 ellipses—34 first-order clusters and two second-order clusters. It saved the first-order ellipses as **NNH1LFA.shp**, the second-order ellipses as **NNH2LFA.shp**, the first-order convex hulls as **CNNH1LFA.shp**, and the second-order convex hulls as **CNNH2LFA.shp**.

Step 10: Add these new shapefiles to your ArcView project. You may note that some of the first-order ellipses seem to surround single points—these indicate locations where multiple incidents are geocoded to the same coordinates. You would not have been able to identify these visually. (Such locations will not have convex hulls, because the hot spots are single points and a convex hull must connect points in an area.)

Step 11: Return to CrimeStat and experiment with different NNH settings, including entering a fixed distance. How do your different settings change the nature and number of hot spots? (Hint: before taking the time to load new shapefiles, check the "results" screen after you click "Compute." It will tell you how many clusters it identified. If it didn't identify any, or too many, you may want to try a different setting.) Try the routine on the residential burglary and robbery files as well.

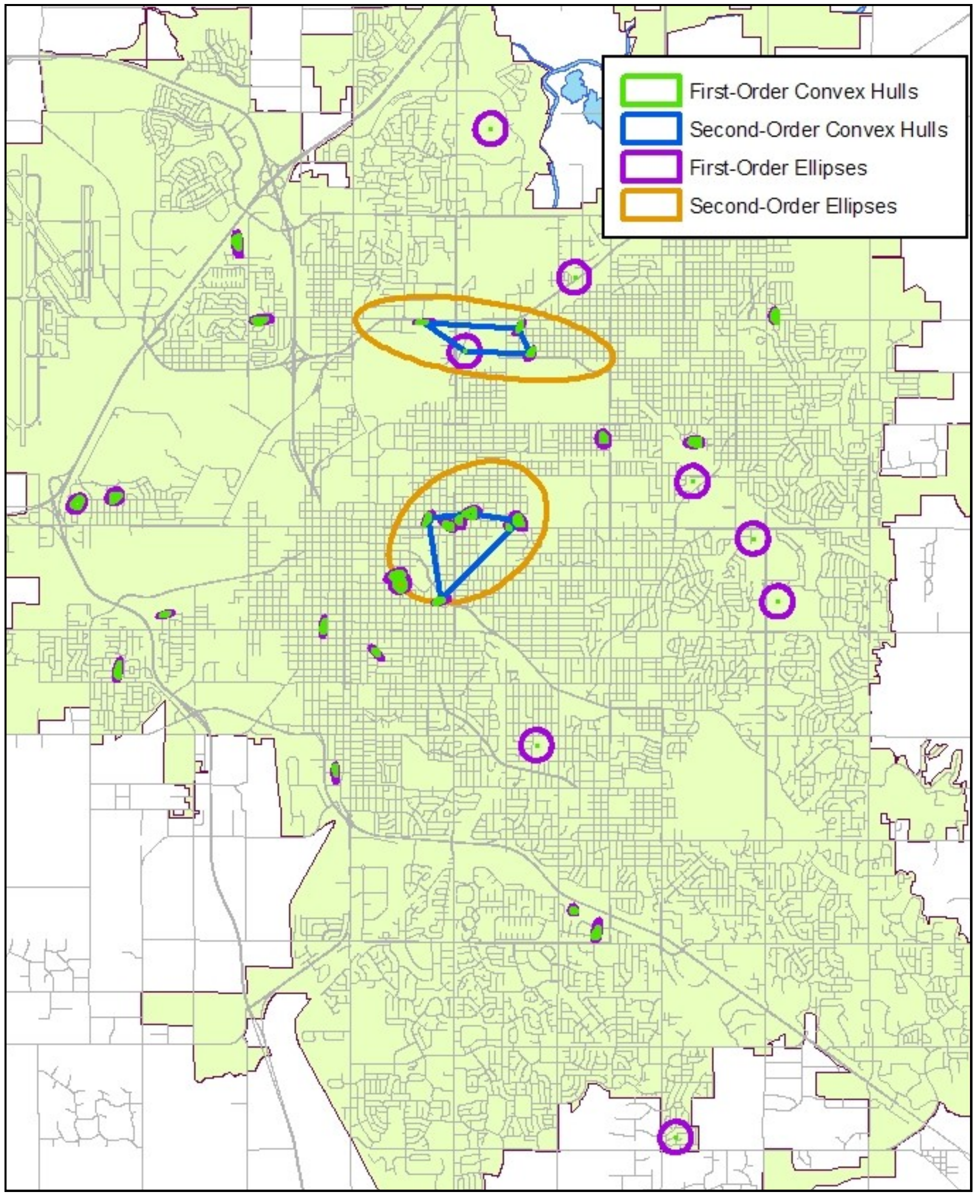


Figure 5-10: the four polygon files uploaded into an ArcView project. Most analysts find the ellipses more intuitive than the convex hulls.

Looking at the pinmap of thefts from automobiles, you may wonder why CrimeStat didn't identify hot spots in any number of other areas that would seem to show very dense clustering. But remember, the NNH routine accounts for the fact that it's *normal* to have points, points everywhere. It selects out those that are closer than would be expected on the basis of chance.

5.3 Spatial and Temporal Analysis of Crime (STAC)

STAC was originally developed as a separate program for the Illinois Criminal Justice Information Authority by Richard and Carolyn Block. CrimeStat integrated it in Version 2. Like NNH, STAC produces ellipses and convex hulls, but uses a different method for generating them.

Essentially, STAC's algorithm scans the data by overlaying a grid on the study area and applying a search circle to each node of the grid (the size is specified by the user). The routine counts the number of points in each circle to identify the densest clusters. It aggregates circles that share the same points, so hot spots may vary in size.

STAC offers some of the same parameters as NNH, including the search radius, the minimum number of points per cluster, and the number of standard deviations for the ellipses.

Step 12: Un-check the NNH option on the "Hot Sot Analysis I" screen and switch to the "Hot Spot Analysis II" screen. Check "Spatial and Temporal Analysis of Crime (STAC)."

Step 13: Click the "STAC Parameters" button. Set an initial search radius of 0.25 miles and 10 minimum points per cluster. We do not have a reference file loaded for the area boundary, so select the "From data set" option. Click "OK." (Note: for Scan Type, we choose "Rectangular" since Lincoln's streets are on a grid; for irregular jurisdictions, the user would choose "Triangular")

Step 14: Click the "Save ellipses to..." button, choose ArcView SHP format, and save the file at **C:\CrimeStat\Data\Lincoln** with the file name **LFA**. "OK" out of this screen and do the same with "Save convex hull to..."

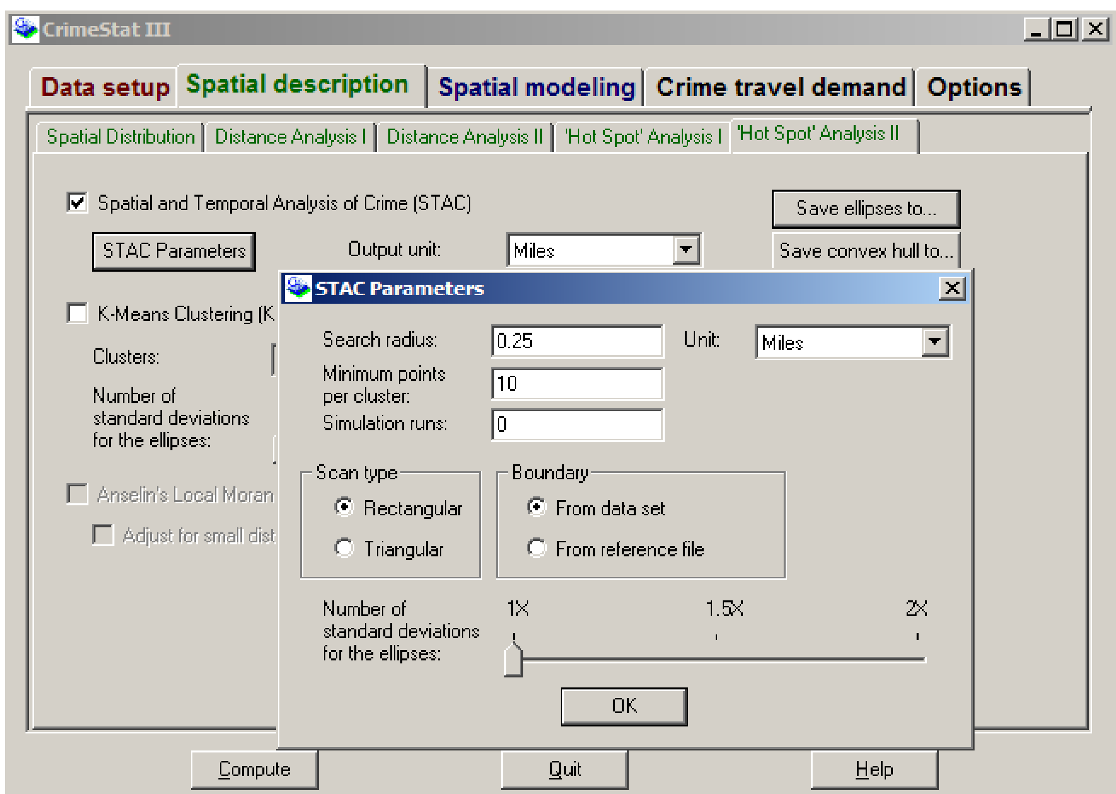


Figure 5-11: The "Hot Spot Analysis II" screen with the STAC parameters.

Step 15: Click "Compute." STAC should identify seven clusters. It will save the ellipse layer as STLFA.shp and the convex hull layer as CSTLFA.shp. Open these layers in your Lincoln ArcView project. How do they compare to the hot spots identified with Nearest Neighbor Hierarchical Clustering?

Step 16: Try adjusting the STAC parameters and see how that changes your hot spots. Or try STAC with either the robbery or residential burglary layers.

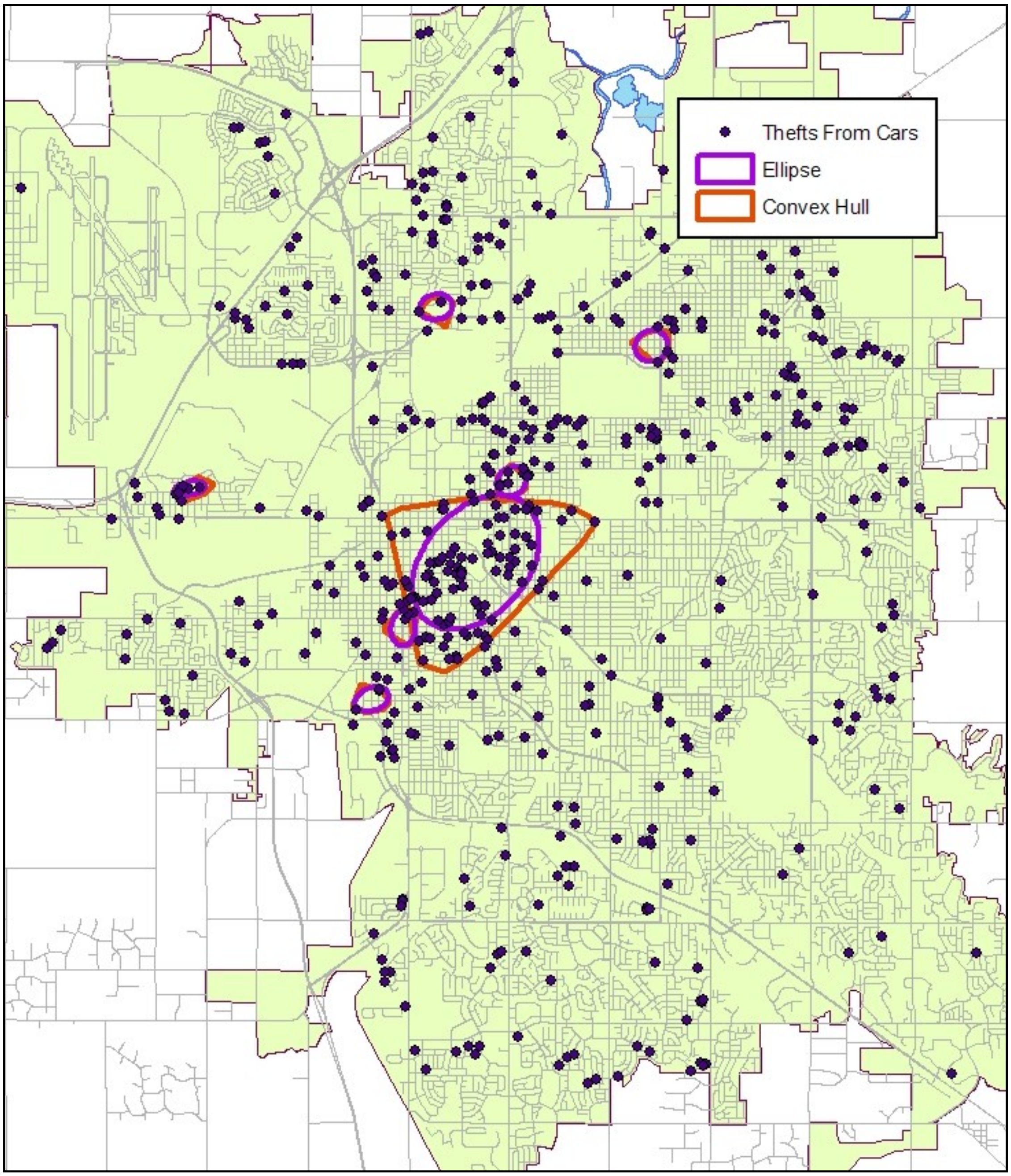
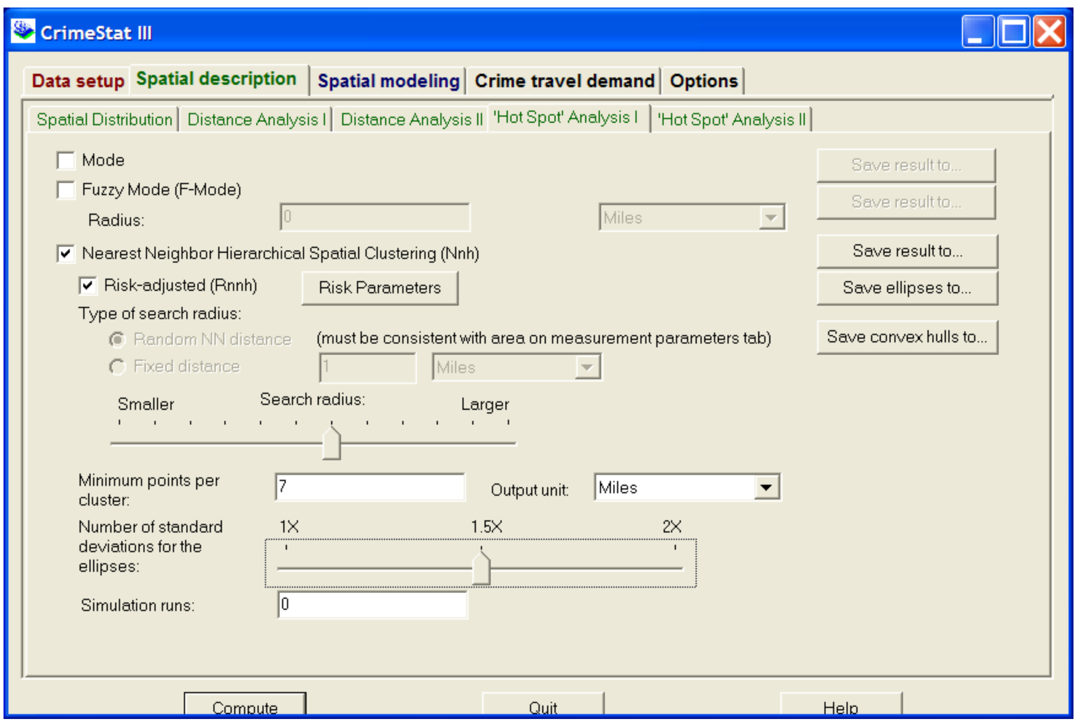


Figure 5-12: Hot spots for thefts from automobiles in Lincoln, as identified by STAC.

It's worth noting that all of these hot spot identification methods have identified clusters based on *volume* rather than *risk*. For most crime analysis purposes, volume is how we want to do it—but not always. Consider that residential burglaries are clearly going to be densest in the areas with the highest population. Three burglaries in a rural neighborhood might signify a hot spot, while 20 burglaries in the middle of town may be nothing worth sounding the alarm about.

CrimeStat does offer one technique to normalize your hot spots this way: the Risk-Adjusted Nearest Neighbor Hierarchical Spatial Clustering routine (RNNH). It relies on a secondary file containing some kind of denominator—perhaps the number of houses when studying housebreaks, or the number of parking spaces when studying thefts from cars. A census block layer with population totals is a common example, although not suitable for all crimes. We have not had time to explore RNNH in this chapter, and it's rare in crime analysis overall, but analysts may want to consult the relevant sections of the CrimeStat manual for more information.

NOTE: For those that would like a challenge at this point, you can run the Risk Adjusted Nearest Neighbor Hierarchical Spatial Clustering routine (RNNH) and compare the results with and without the Risk Adjusted option. For your secondary (denominator) file, use the **censusblocks.dbf**. You will need to select "Households" for the "Z" (Intensity) field. See Figure 5-13 below for suggested parameters on the Spatial Description tab.

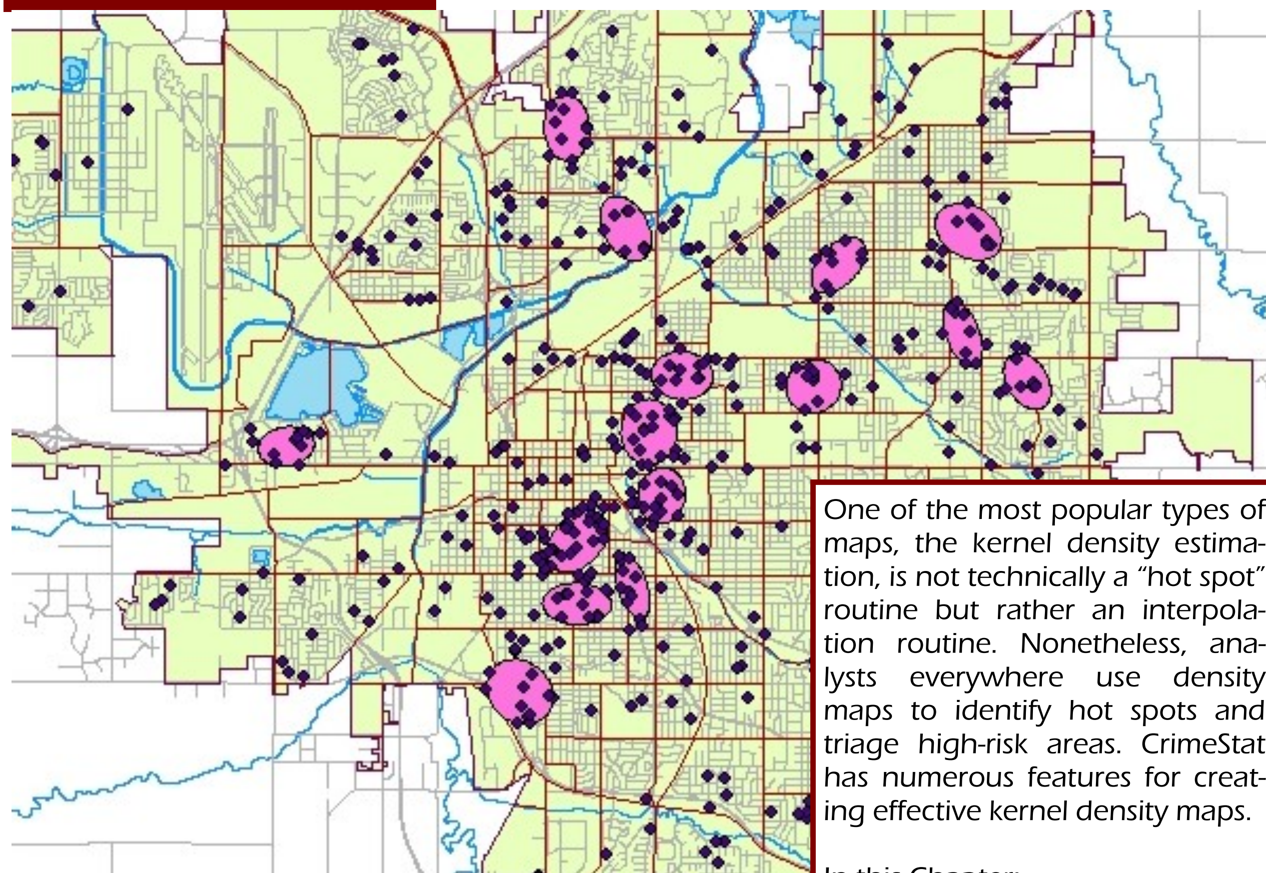


While CrimeStat's hot spot identification techniques are not as subjective as circling areas with a magic marker, there is still a considerable amount of subjectivity in them. An analyst could adjust the parameters in Fuzzy modes, NNH, and STAC to create as many or as few "hot spots," of as large or small a size, as he or she desires. But as we saw at the beginning of the chapter, "hot spots" themselves are a somewhat subjective concept. There's nothing inherently wrong with subjectivity; the key is to understand these methods well enough that you can "triage" your data for tactical and strategic targeting. Practice and experimentation will allow you to achieve that understanding.

5 Hot Spot Analysis

Technique	Description	User Options	Advantages	Disadvantages
Mode	Identifies the specific coordinate pairs with the most points	None	Easy to understand Easy to compute	Can be easily queried outside CrimeStat Will not make hot spots out of points unless they're literally right on top of each other—does not work with certain agency geocoding methods
Fuzzy Mode	Counts points within a user-specific search radius around each point; identifies those with the highest volume	Size of search radius	Easy to understand and compute Combines points that are nearby even if they don't share the same exact coordinates	Usually results in multiple overlapping hot spots
Nearest Neighbor Hierarchical Spatial Clustering (NNH)	Identifies clusters where points are closer together than would be expected on the basis of random chance.	Size of search radius (and probability of error) Minimum points per cluster Number of standard deviations for ellipses	If left on "Random NN distance," based on statistical probability—the most "objective" of the options. Hierarchy method identifies both large and small hot spots; works at the tactical and strategic levels Produces the most concentrated clusters of all hot spot routines.	Multiple settings can render results meaningless if user doesn't know what he's doing.
Risk-Adjusted NNH	Works like NNH, but is based on a rate (using a user-specified denominator) rather than simple volume	Rate/Risk variable Type of bandwidth	Only hot spot method that considers underlying surface risk	Difficult to get suitable data For most crime analysis purposes, we want the straight volume
Spatial and Temporal Analysis of Crime (STAC)	Scans the data with a series of overlapping circles; identifies those with the highest volume	Size of circles Minimum points per cluster Scan type Boundary parameters Number of standard deviations for ellipses	Fast and intuitive Parameters give analyst large control over results	Multiple settings can render results meaningless if user doesn't know what he's doing.
K-Means Partitioning Clustering	User specifies the number of hot spots he wants to see; CrimeStat tests various positions and sizes	Number of clusters Separation between clusters Number of standard deviations around ellipses	Good for certain tactical and strategic projects in which the number of areas to be targeted is already known	Clusters may not make practical sense depending on user settings
Anselin's Local Moran Statistic	Looks for hot zones in relation to the overall neighborhood they sit within	Adjust for small distances Calculate variance	Looks for outliers based on neighbors rather than overall volume	Some use in crime analysis Requires data already aggregated into polygons

6 Kernel Density Estimation



One of the most popular types of maps, the kernel density estimation, is not technically a "hot spot" routine but rather an interpolation routine. Nonetheless, analysts everywhere use density maps to identify hot spots and triage high-risk areas. CrimeStat has numerous features for creating effective kernel density maps.

In this Chapter:

- How kernel density estimation works
- Understanding different interpolation methods
- Guidelines for kernel size and bandwidth
- Creating and mapping a kernel density estimation
- Uses and misuses of kernel density

6 Kernel Density Estimation

6.0 Introduction

We do not claim to have undertaken extensive research in the subject, be we feel confident in stating that, aside from basic pin maps, kernel density maps (also known, with varying degrees of accuracy, as surface density maps, continuous surface maps, density maps, isopleths maps, grid maps, and "hot spot" maps) are the most popular map type in crime analysis. It's rare to open a crime analyst's problem profile or annual crime report and not be confronted with one of these multi-colored plots.

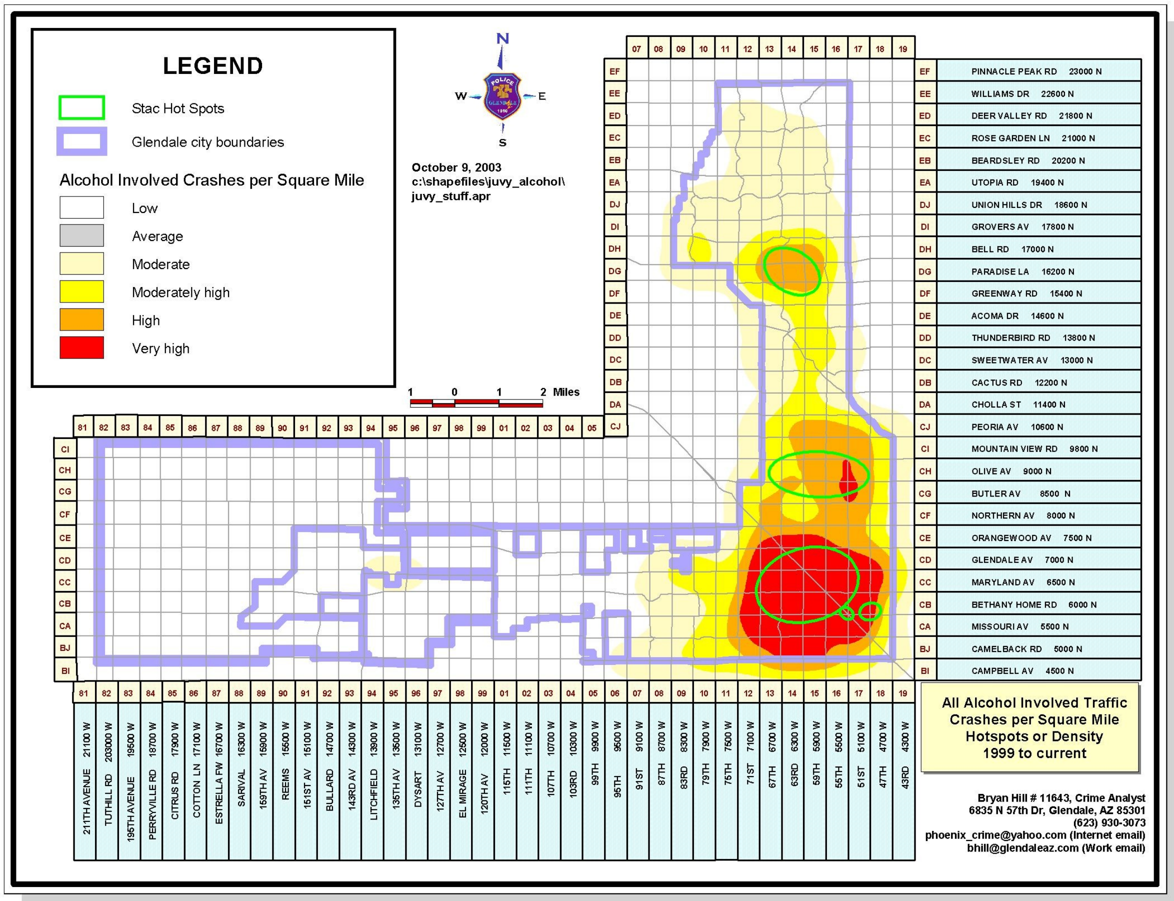


Figure 6-1: a kernel density map showing alcohol-involved traffic accidents in Glendale, Arizona.

Where the hot spot maps we studied in Chapter 5 were based on known, actual volumes of incidents at specific locations, **kernel density estimation** (KDE, also KDI: kernel density interpolation) *generalizes* data over a larger region.

In non-crime analysis scenarios, interpolated maps are invaluable for estimating values over a large region from which only samples have been taken. Temperature is a good example. Color-coded isotherm maps show temperatures for all parts of a geographic area, but it is functionally impossible to know the exact temperature at all points on the surface of the Earth. Instead, to create an isotherm map, the user samples temperatures at a few dozen to a few thousand locations (depending on the map scale) and then, based on these known temperatures, *estimates* the temperatures at the other points in the area.

The question always arises, then, about how accurate this method is for crime analysis. After all, we are not "sampling" our crime data; we generally map our entire population of data. We don't need to estimate. Unlike temperature, not every point on the Earth's surface has a number of crimes.

KDE maps for crime analysis, then, are best viewed as "risk surfaces." Although not every point on the Earth's surface has crime, every point on the Earth's surface does have a *chance* of crime. KDE estimates this probability based on locations where crimes have occurred in the past.

6.1 How KDE Works

With KDE, every point on the map has a **density estimate** based on its proximity to crime incidents (or whatever it is we're mapping). Because CrimeStat cannot literally calculate the density estimate for every point (there an infinite number), it overlays a grid on top of your map and calculates the density estimate for the centerpoint of each grid cell. The specific number of cells in the grid is defined by the user on the "Reference File" screen under Data setup.

.

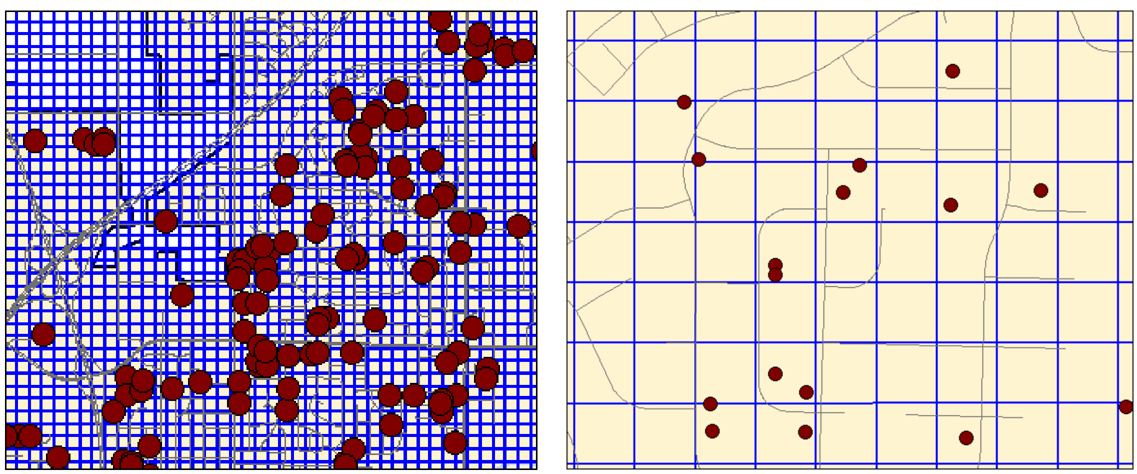


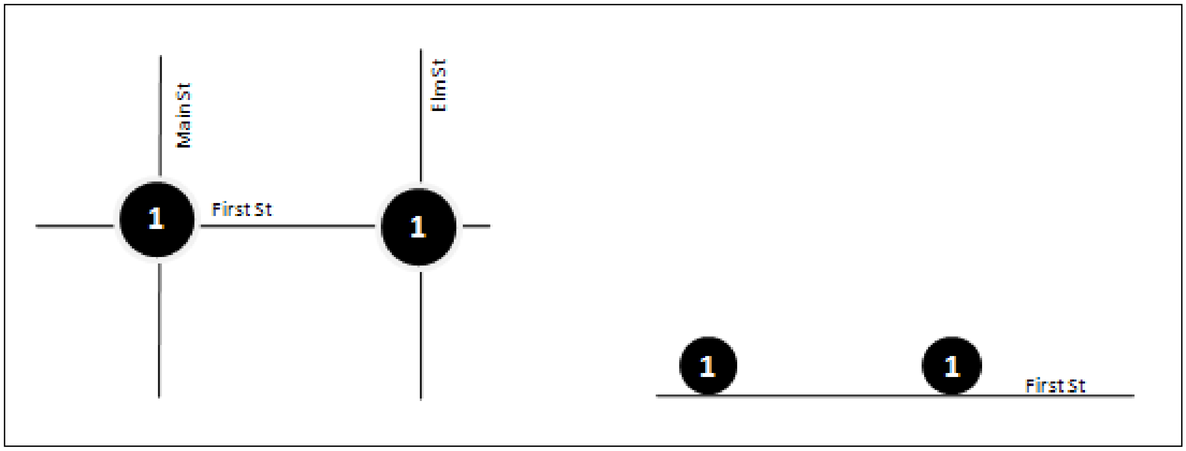
Figure 6-2: a grid overlaid on Lincoln theft from automobile data, at two different scales

CrimeStat measures the distance between each grid cell centerpoint and each incident data point and determines what weight, if any, the cell gets for that point. It then sums the weights received from all points into the density estimate.

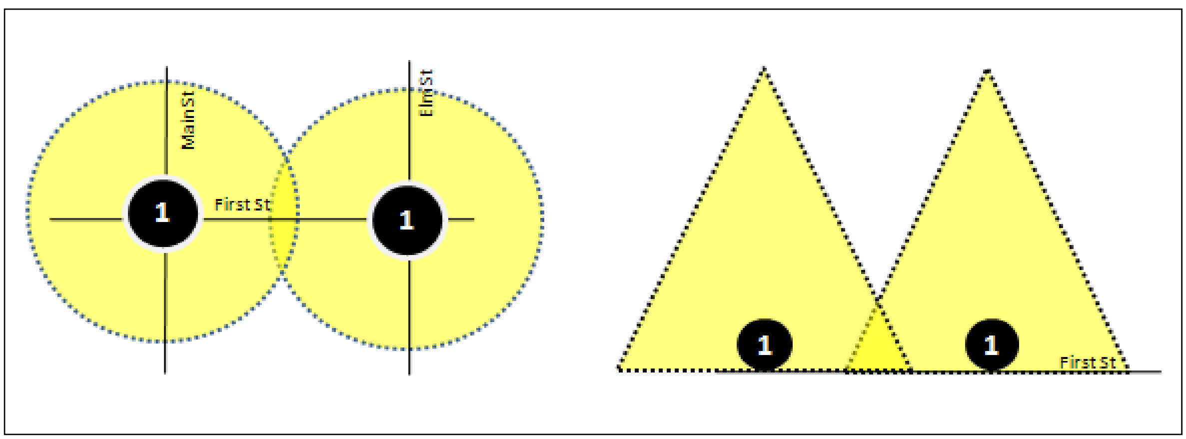
The specific weight each cell receives depends on:

* The distance from the grid cell centerpoint to the incident data point
* The size of the radius around each incident data point
* The method of interpolation

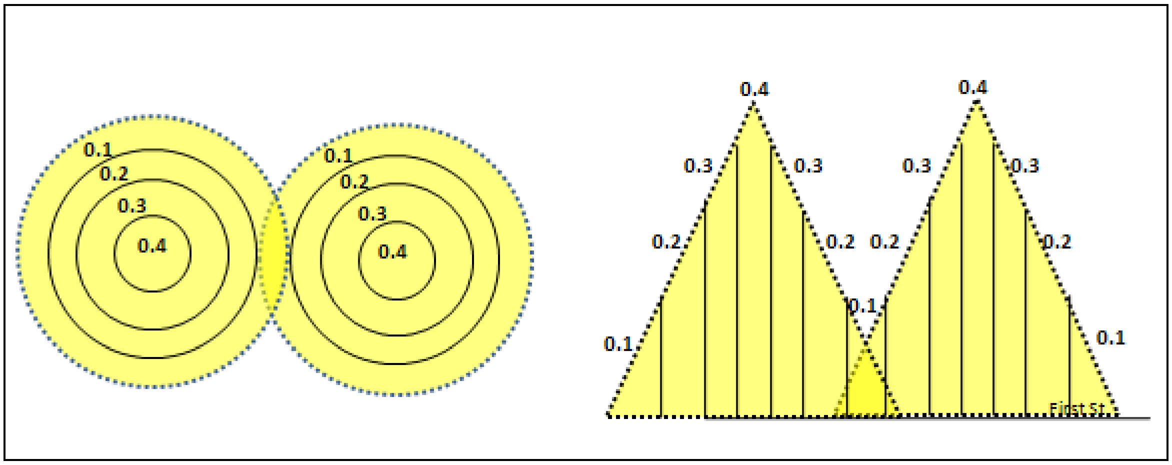
A series of conceptual illustrations help explain the latter two points. Assume we have points along First Street with robberies (we're showing the phenomenon in two dimensions and three dimensions). Each point with a robbery begins with a density of 1, or 100%.



KDE places a symmetrical surface called a **kernel** *over each data point. The size of the kernel is specified by the user and the shape is specified by the method of interpolation. The value represented by the point (1) is then "smoothed' throughout the area of the kernel. In this example, we use a triangular or conic kernel.*

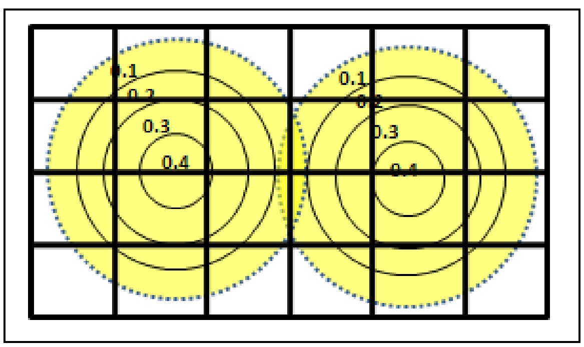


With a triangular kernel, the smoothing proceeds from the initial point and decreases in a linear relationship to the edges of the kernels. (We represent it with concentric circles below, but the smoothing is actually a constant.)



It might help to think of each incident as a tall column of clay and the kernel as a cup that you press down over the clay, forcing it to spread out and take the shape of the cup. The weight each point on the map receives depends on the height of the clay above it.

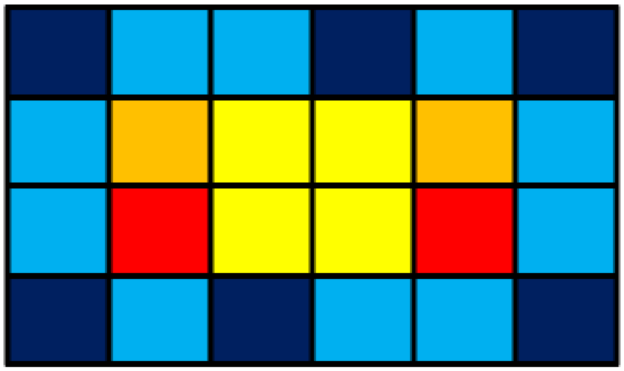
Now that we know the size of the radius and the specific weight any point will receive if and where it falls within the radius, we overlay the grid cells.

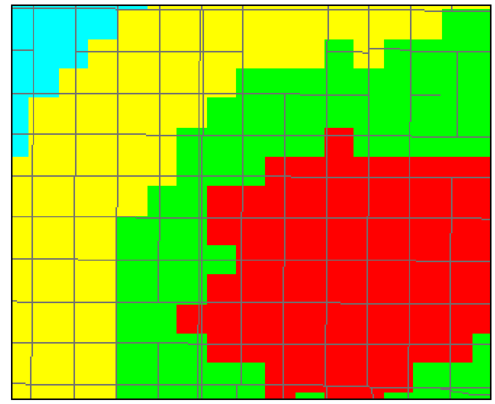


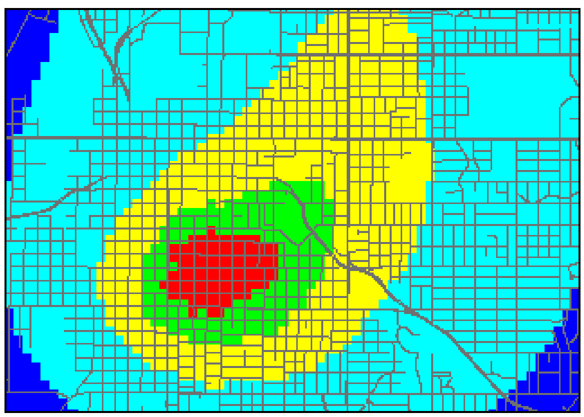
0.0	0.1	0.1	0.0	0.1	0.0
0.1	0.3	0.2	0.2	0.3	0.1
0.1	0.4	0.2	0.2	0.4	0.1
0.0	0.1	0.0	0.1	0.1	0.0

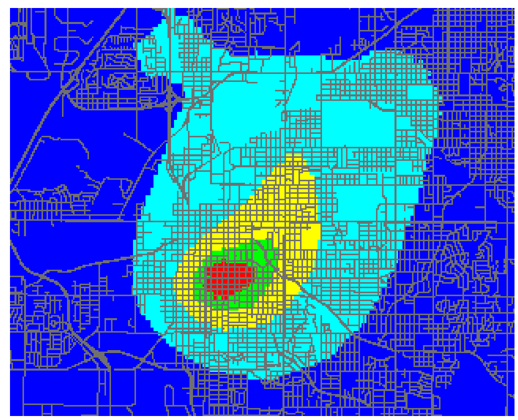
(If multiple points appear at the same location, which of course is frequent, then multiple kernels will appear at the same location, and the aggregated totals for grid cells that overlap those kernels will be higher.)

When displaying the final result on a map, the grid cells are color-coded based on the density. At large scales (zoomed in), the grid cells might be obvious, but at smaller scales, the distribution takes on a smoother appearance. Specifying a large number of cells at the outset will make the final result appear smoother at a larger scale.









Figures 6-3 to 6-6: our hypothetical KDE followed by an actual KDE at decreasing scales.

6.2 KDE Parameters

There are, naturally, many parameters to understand when running a KDE in CrimeStat. As with hot spot analysis, many of these parameters depend on the analyst's own experience and judgment, and sometimes only experimentation can lead to the "correct" settings.

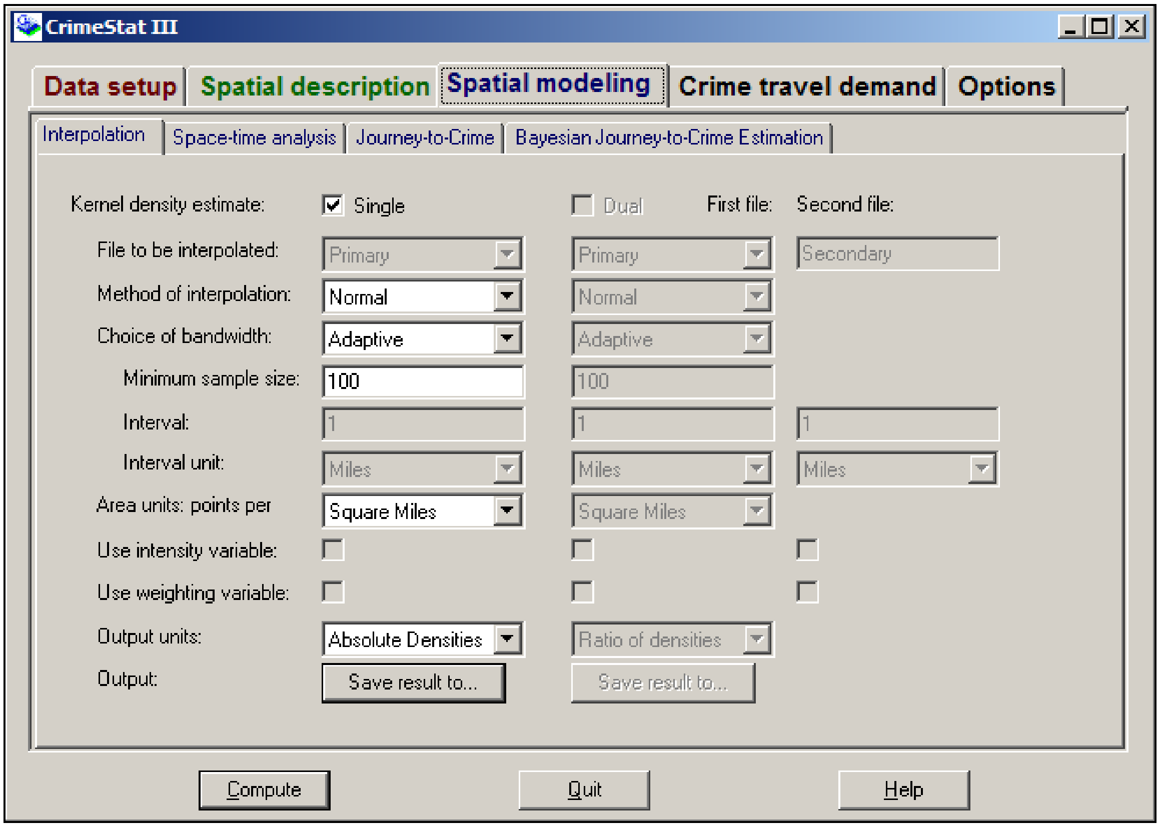
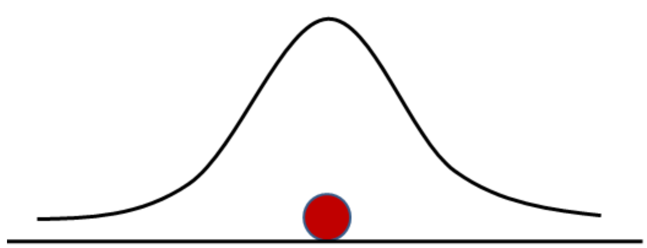


Figure 6-7: the Kernel Density Estimate screen in CrimeStat

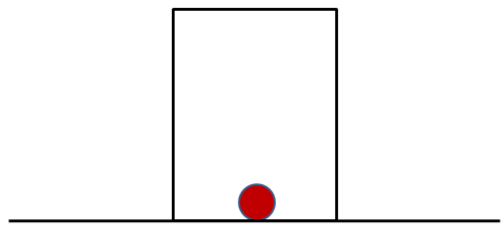
First, CrimeStat will allow a *single* or *dual* kernel density estimate. Single estimates identify hot spots for most crime analysis purposes. Dual estimates can help normalize data for population or other risk factors or calculate change from one time period to the next.

In our example above, our kernel was shaped like a cone. The size of the kernel is referred to as its **bandwidth**, and it can be specified to some degree by the user. The shape of the bandwidth is specified by the *method of interpolation*. CrimeStat offers five interpolation methods:

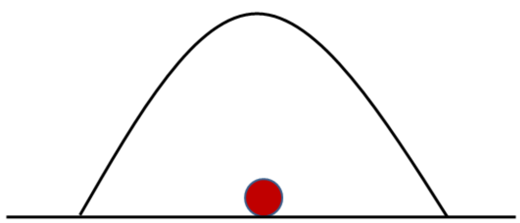
1. *Normal distribution*. A normal distribution kernel, represented by a bell curve, peaks above the data point, declines rapidly for one standard deviation, and then enters a less dramatic rate of decline for subsequent standard deviations. Unlike the other methods (including the conical one above), the normal distribution does not have a defined radius—it continues across the entire reference grid. This means that every point on the reference grid gets a value from each kernel (even if the value is so small that it rounds to 0).



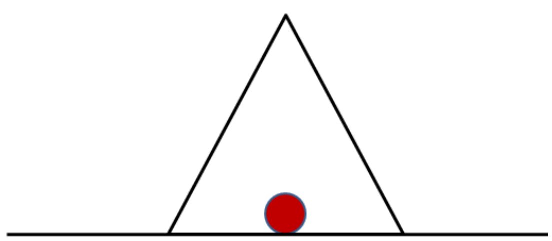
2. Uniform (flat) distribution. Represented by a cylinder, a uniform distribution has a fixed radius, but all points within the radius receive an equal density weight.



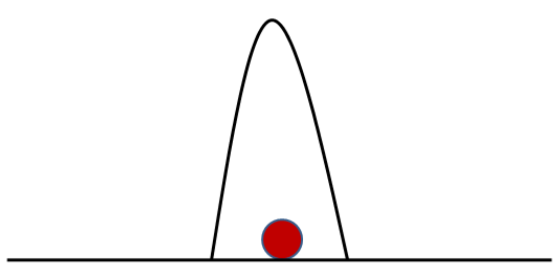
3. *Quartic (spherical) distribution*. A quartic distribution is another curve, but more gradual in its initial stages than a standard deviation curve. Density is highest over the point and falls of gradually to the limits of the radius.



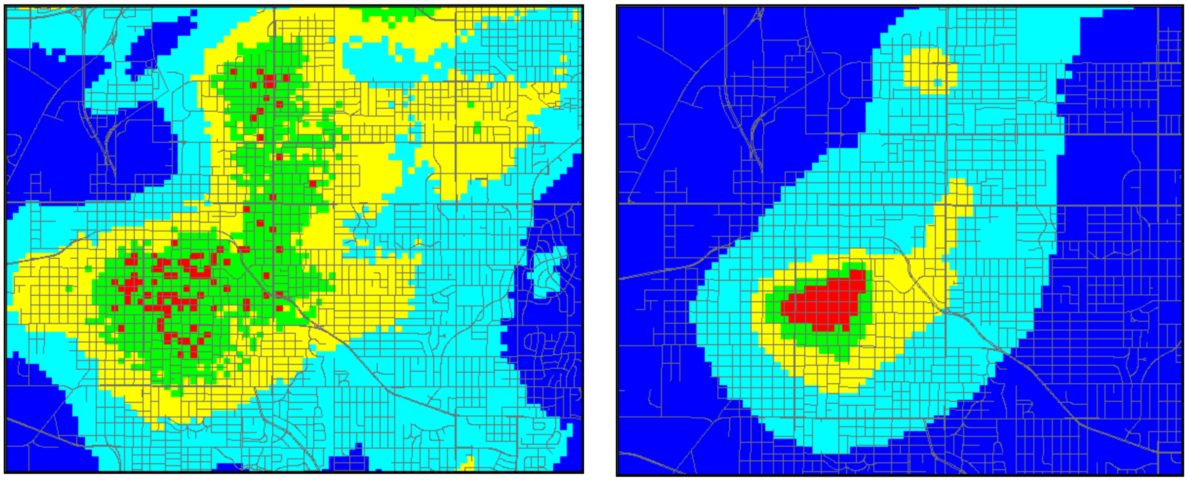
4. *Triangular (conical) distribution*. As we've seen, the triangular distribution peaks above the point and falls off in a linear manner to the edges of the radius.



5. *Negative exponential distribution*. A negative exponential distribution is another curve, but one that falls off rapidly from the peak to the specified radius.



Each method of interpolation will produce slightly different results. Triangular and negative exponential functions tend to produce and emphasize many small hot and cold spots and thus produce a "mottled" appearance on your map. Quartile, uniform, and normal distribution functions tend (in that order) to smooth the data more.



Figures 6-8 and 6-9: residential burglaries in Lincoln smoothed with the negative exponential distribution method (left) and the normal distribution method (right).

The *choice of bandwidth*, *minimum sample size*, and *interval* parameters are all related, and they all work together to specify the size of the bandwidth. If you choose an "adaptive" bandwidth, CrimeStat will adjust the size of the kernel until it's large enough to contain the minimum sample size. If you choose a "Fixed Interval" bandwidth, you specify the size.

Visually, it does not matter which *output unit* you choose, since the grid cells will remain the same relative to each other and will therefore "map" the same. The output unit does make a difference in terms of your legend, and how you explain the results.

Absolute densities are simply the sum of all the weights received by each cell—but re-scaled so that the sum of the densities equals the total number of incidents. This setting is the default, and it will suffice for most crime analysis purposes.

Relative densities divide the absolute densities by the area of the grid cell. Thus when explaining the map, you can say that the red represents X points per square mile rather than X points per grid cell.

Probabilities divide the density by the total number of incidents. The result is the chance that any incident occurred in that cell.

There isn't any firm guidance on what interpolation method or bandwidth size to use. It depends largely on the nature of the crime and what the analyst wants to display. But to help you out, consider what KDE is doing: interpolating locations of known crimes across a larger region, and thus creating a "risk surface." The assumption is that if there are a lot of incidents at one location, the locations around it have a higher risk of those incidents.

A few moments' thought reveals that this statement is truer for certain types of incidents than others. For instance, consider a map of motor vehicle collisions. Certain intersections will emerge as "hot spots" for collisions, but those intersections will usually be hot spots for a particular reason: congestion, bad timing on the stop lights, bad intersection design, no stop sign where one is needed, and so on. Sections of road 2,000, 200, or even 20 feet away from the intersection may not have the same conditions and therefore have no higher risk of collisions than any other point on your map. KDE would seem to have limited utility for motor vehicle collisions. (That said, many researchers feel KDE can be very useful, especially in large metropolitan areas due to the high volume of collisions.)

But consider a subset of motor vehicle collisions: those caused by drunk driving. Although a general area might be high-risk for drunk driving because of the proximity of bars, liquor stores, fraternities, and so on, the specific location of the collision has more to do with where the bleary-eyed, swerving driver finally encountered another car, or nodded off and struck a utility pole, than it has to do with anything about the exact location. So for drunk driving collisions, it does make sense to smooth the risk over a larger area.

Some incident types would seem to fall somewhere in between. Street robberies depend on a motivated offender encountering a suitable target at a particular place and time. If the robber prowls an area looking for a target, then the specific location doesn't matter, but if the robber stays at a particular location waiting for a victim to appear, then it does.

Thus, when deciding which parameters to use for a particular dataset, it makes sense to ask two questions:

Across how great an area is this incident likely to have an effect? Then adjust the interval distance (the bandwidth size) accordingly.

How much of this effect should remain at the original location, and how much should be dispersed throughout the bandwidth interval? Then adjust the method of interpolation accordingly.

Table 6-1 on the following page offers suggestions for different types of crime, but we hasten to add that these suggestions are based on our own reasoning and analytical experience, and not on any formal research. The nature of these phenomena may very well differ in your jurisdiction.

Table 6-1: interpolation method and interval possibilities for different types of crime

Incident Type	Interval	Interpolation Method	Reasoning
Residential burglaries	1 mile	Moderately dispersed: quartic or uniform	Some burglars choose particular houses, but most cruise neighborhoods looking for likely targets. A housebreak in any part of a neighborhood transfers risk to the rest of the neighborhood.
Domestic violence	0.1 mile	Tightly focused: negative exponential	Domestic violence occurs among specific individuals and families. Incidents at one location do not have much chance of being contagious in the surrounding area.
Commercial robberies	2 miles	Focused: triangular or negative exponential	A commercial robber probably chooses to strike in a commercial area, and then looks for preferred targets (banks, convenience stores) within that area. The wide area may thus be at some risk, but the brunt of the weight should remain with the particular target that has already been struck.
Thefts from vehicles	0.25 mile	Dispersed: uniform	If a parking lot experiences a lot of thefts from vehicles, your GIS will probably geocode them at the center of the parcel. This method ensures that the risk disperses evenly across the parcel and part of the surrounding area (which probably makes sense)—but not too far, since we know that parking lots tend to be hot spots for specific reasons.

Use similar considerations to determine your own initial values for whatever type of offense or other phenomenon you're studying.

6.3 Creating a KDE

To create our first KDE, we'll use the last example in the table and study thefts from motor vehicles in Lincoln in 2007.

Step 1: Start a new CrimeStat session. On the "Data setup" screen, choose "Select file," choose an "ArcView SHP" file, and add the **C:\CrimeStat\Data\Lincoln\theftfromautos.shp** file. "Save" and "OK" to return to the main screen. Set the X coordinate to "X" and the Y coordinate to "Y." The data is "Projected" in "Feet." (All of this may be set up from the previous lesson.)

Step 2: We now have to create our reference grid. Click on the "Reference File" tab. If you created a reference grid already in Chapter 2, you should be able to click "Load" and retrieve it. Otherwise, set up the grid with the coordinates in Figure 6-10 (next page), with 250 columns.

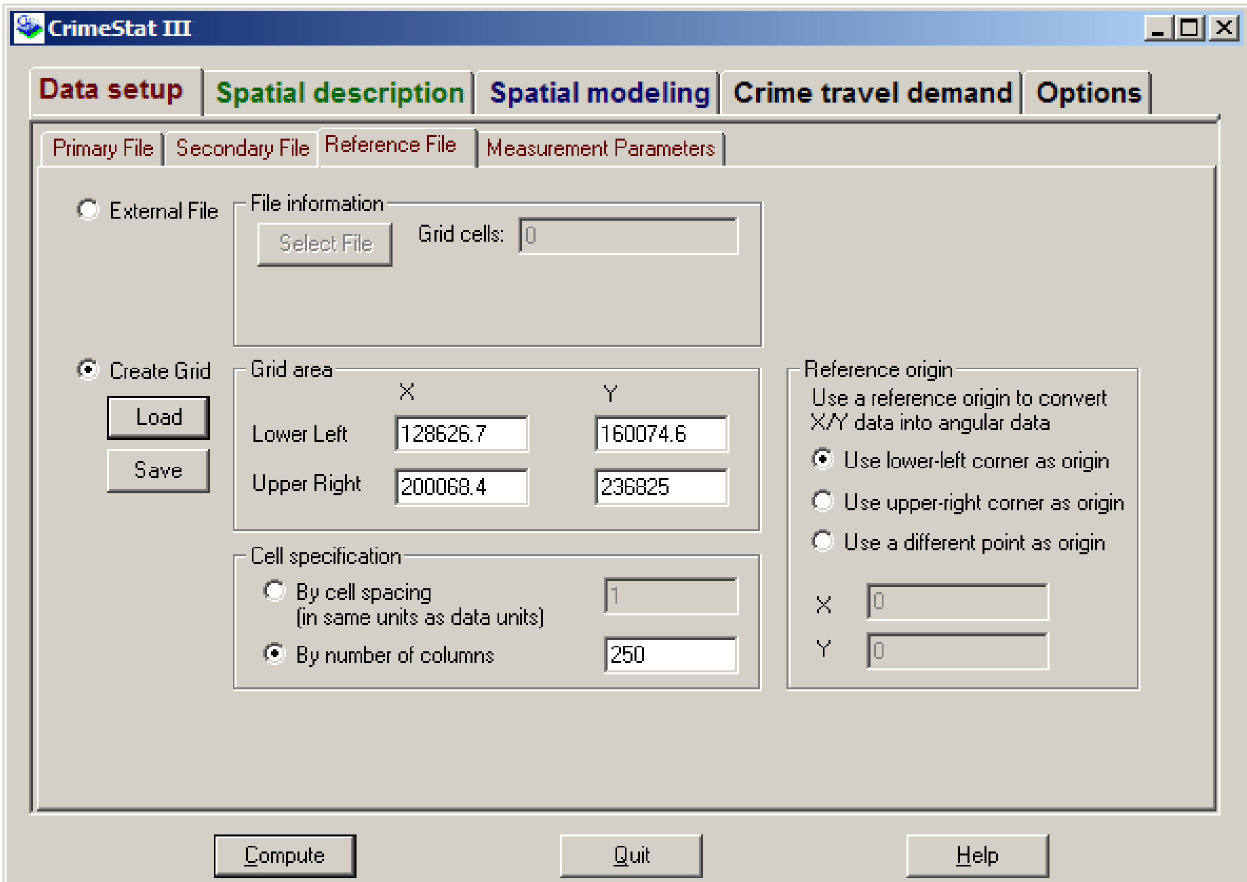


Figure 6-10: settings for our reference grid

We have set the boundaries of the grid area to fall comfortably outside the actual territory covered by the city of Lincoln. If you set the boundaries too tightly, it can result in an "edge effect" in which cells at the edge of the reference grid receive higher-than-warranted values, particularly when using a normal interpolation method.

Step 3: Click on the "Spatial modeling" tab and then the "Interpolation" sub-tab. Click the check box for a "Single" KDE. Choose a "Uniform" interpolation with a "Fixed Interval" bandwidth, and set the bandwidth interval to 0.25 miles (Figure 6-11, next page).

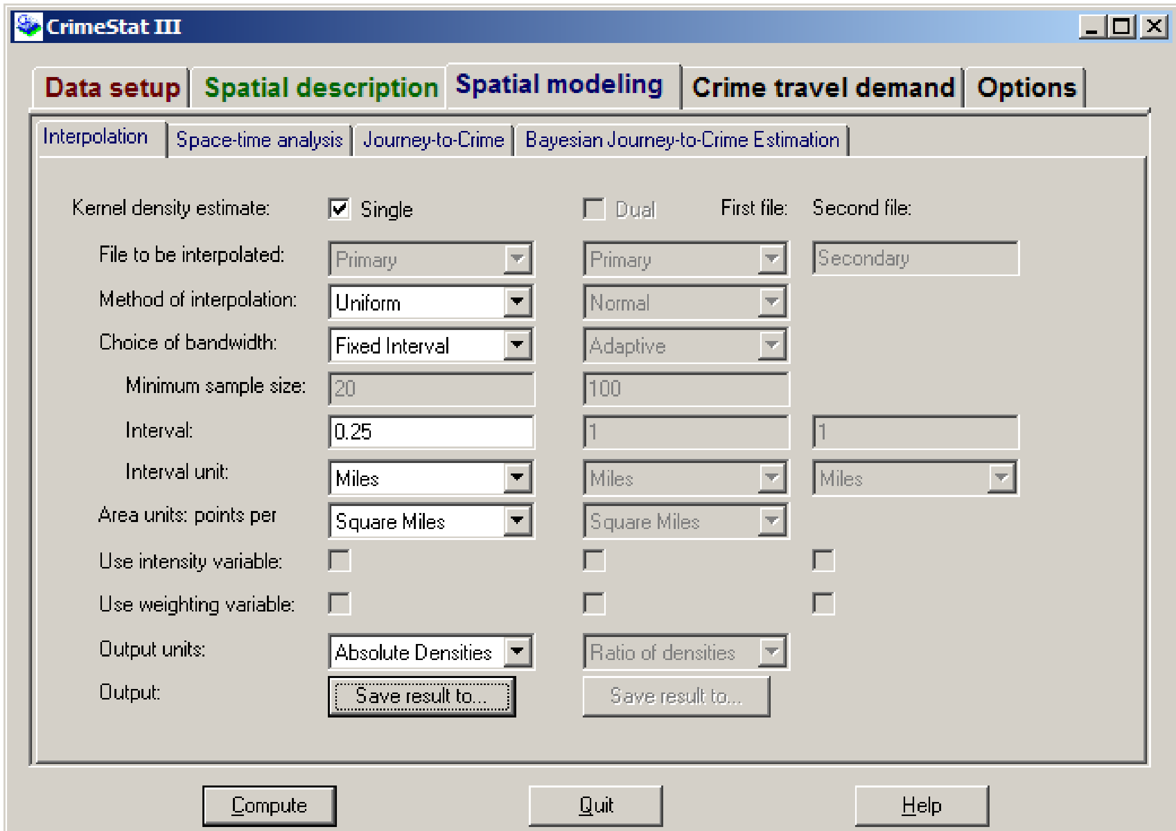


Figure 6-11: setting up the KDE parameters

Step 4: Click the "Save Result to..." button. Save the output to an ArcView SHP file at **C:\CrimeStat\Data\Lincoln** with a file name of **LFA**. (CrimeStat will automatically put a "K" in front of this.) "OK" this box and click "Compute."

It will take a few minutes to run the calculations and output the file (the time depends on your memory and processor speed). When you are finished, you will have a polygon layer comprising 67,250 square cells: 250 columns and 269 rows. The layer has three pieces of attribute data, the most important of which is a Column titled "Z," which contains your density estimate.

Step 5: Open the "KLFA" shapefile in your Lincoln ArcView project. To color and display the results, you will need to create a choropleth map based on the "Z" field. We find that five ranges, using the "Equal Ranges" routine tends to result in the best display. Use a sensible, intuitive color scale for your ranges, such as red to dark blue. Arcview will offer you a series of options for your color ranges / ramps from which you can choose.

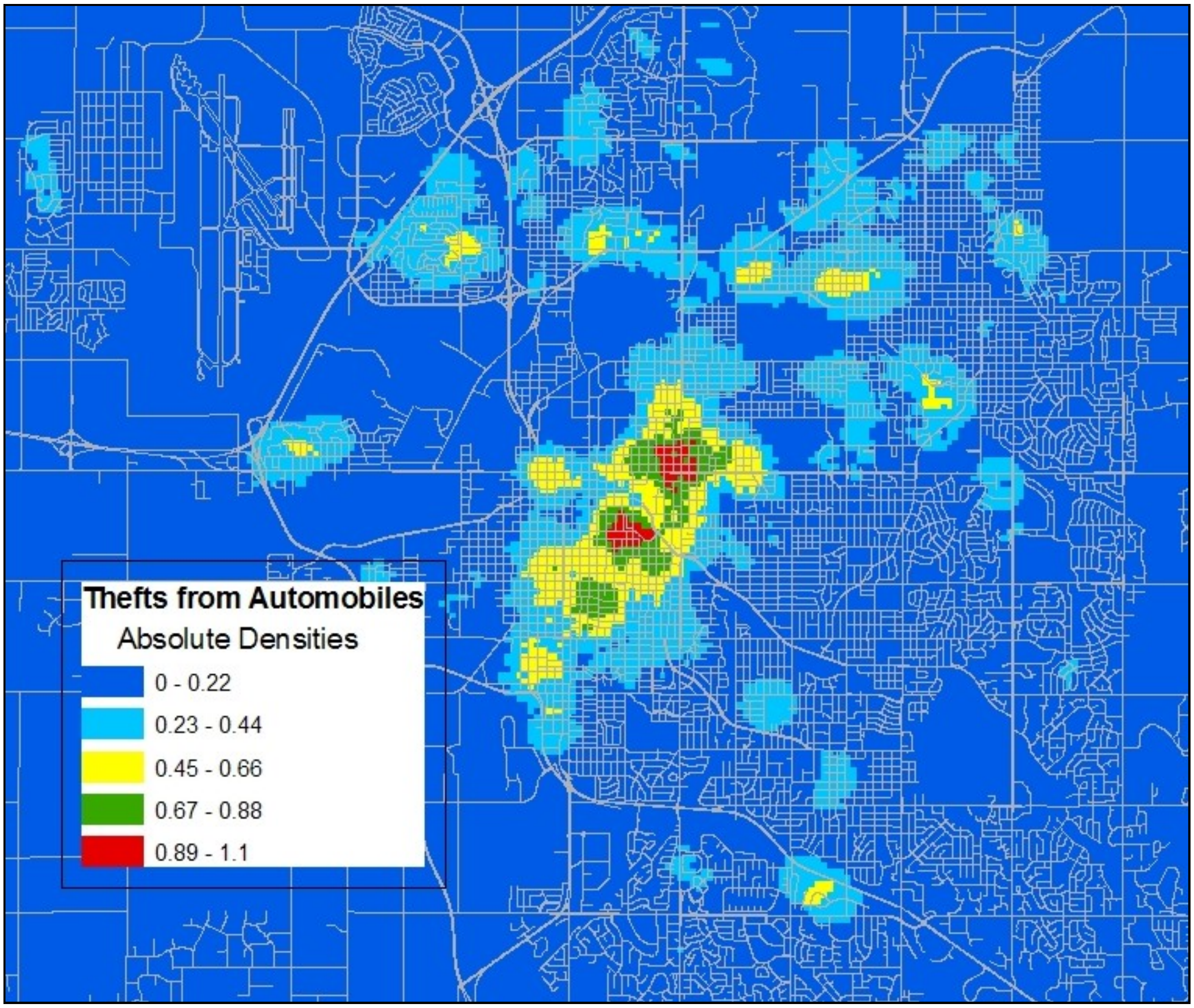


Figure 6-12: the results of our KDE for thefts from autos in Lincoln.

Step 6: Experiment! Try different settings for this file, and/or repeat Steps 1-5 with the robbery and residential burglary files (adjusting the initial settings in the manner that makes most sense to you). Make sure you understand how each parameter affects your data.

6.4 Dual KDE

A dual KDE is simply a kernel density estimation based on two files, one primary and one secondary. The results of the two KDEs are then added to, subtracted from, or expressed as a ratio of each other.

If we wanted to analyze two crimes at one time, for instance, we could assign one as the primary file and one as the secondary file. We could then tell CrimeStat to add the densities for the two files. However, it would be somewhat easier just to include both crimes in the database query from which we produce the original primary file.

The more common use of a dual KDE is to normalize for risk. One of the fundamental problems with a single KDE is that it assumes a *uniform risk surface*. Hot spots are based entirely on volume. But 60 residential burglaries in a neighborhood with 600 houses is much worse than 60 residential burglaries in a neighborhood with 6,000 houses. By having CrimeStat divide the housebreak density by this risk variable (number of houses), we can produce a better estimation of relative risk.

Although normalizing data this way is generally a good idea, there are four things to keep in mind:

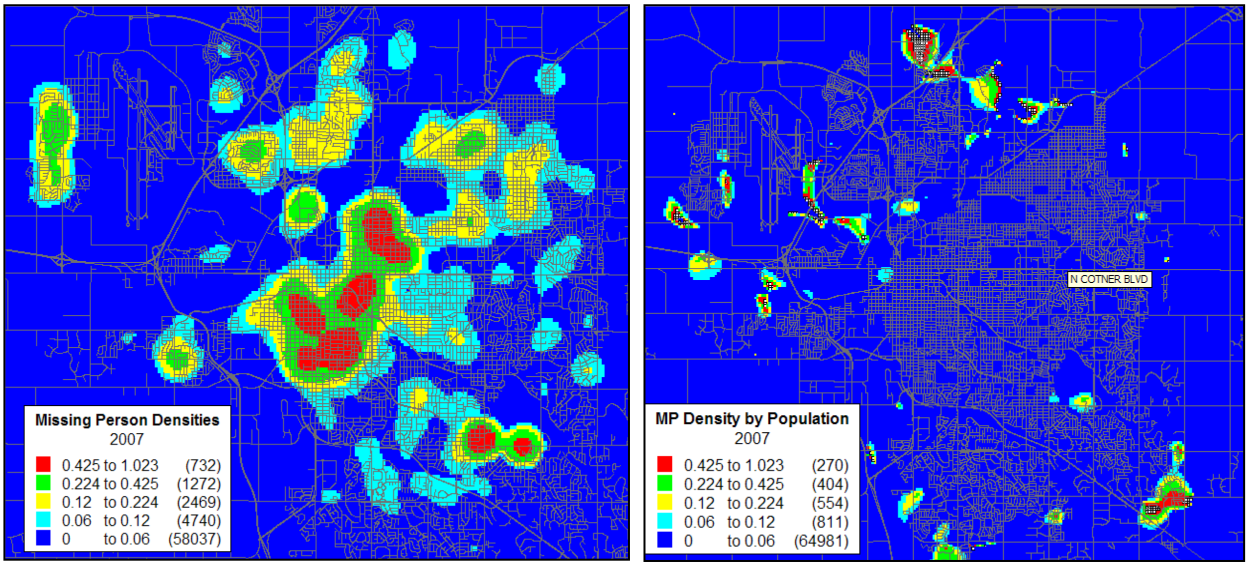
1) Sometimes you want a normalized volume, sometimes you don't. It depends on the ultimate uses of the analysis. For certain tactical and strategic interventions, you want to target the areas of the highest volume, regardless of the underlying risk.

2) Data to use for the secondary file is very hard to come by. Yes, we have census population data, which we'll use in a moment, but using population as a denominator only makes sense for a limited number of crimes that occur primarily at residences. Normalizing commercial burglary or commercial robberies would require data on the number of businesses (or, even better, the total employment levels at those businesses), and normalizing auto thefts or thefts from automobiles would best be done with data on the number of parking spaces. None of this data is easily obtained and it may require extensive research on the part of the crime analyst.

3) CrimeStat requires point data for the secondary file and interpolates it just like the primary file. To normalize by population data contained in census block boundaries, we include a file containing the centerpoints of those census blocks. Thus, CrimeStat does not read whether the primary point is "in" a particular census block. Instead, it smooths the population from each census block centerpoint. Instead of an interpolation normalized by underlying geography, you end up with an interpolation normalized by another interpolation.

4) You cannot use a different interpolation method (kernel) for the numerator and denominator, but you can use an adaptive bandwidth, one for the numerator and one for the denominator each file. Thus, the bandwidth size and interpolation method that you choose may have to be a compromise between what works best for each of the files.

With all of these conditions in mind, if we desire normalized data, and we have suitable data to serve as a denominator, and we handle the interpolation properly, CrimeStat will produce significantly different results than when we rely on volume alone, as we see in Figures 6-13 and 6-14.



Figures 6-13 and 6-14: missing persons data in Lincoln, by volume (left) and relative risk based on population density (right).

The parameters for the secondary file are the same as we discussed for the primary file until the "Output Units" option. There are six options for the final density calculation:

1) *Ratio of densities*: the default option and the most common. It divides the density in the primary file by the density in the secondary file. This is best used for the type of "normalization" we discussed above.

2) *Log ratio of densities*: a logarithmic function that helps control extreme highs and lows in your data. The CrimeStat manual suggests this function for strongly skewed distributions in which most reference cells have very low densities but a few have very high densities.

3) *Absolute difference in densities:* subtracts the secondary file densities from the primary file densities. This option is valuable if your primary file has crimes for one time period and your secondary file has the same type of crime for another time period. The resulting map will show how the crime changed from one time period to the next.

4) *Relative difference in densities:* like the relative density option for the primary file, this option divides the primary and secondary file densities by the area of the cells before subtracting them. It will result in the same ratios, and thus the same map, as the absolute difference.

5) *Sum of densities:* adds the two densities together—useful if you want to show the combined effects of two types of crime.

6) *Relative sum of densities:* divides the primary and secondary files by the area of the cells before adding them.

We will create a dual KDE that analyzes relative risk of residential burglary, using the ratio of densities option.

Step 7: On your "Data Setup" screen, "Remove" the larceny from automobiles file and "Select Files" to add the **C:\CrimeStat\Data\Lincoln\resburglaries.shp** file. Assign the X and Y coordinates to the "X" and "Y" fields, and make sure the coordinate system is set to "Projected" with the data units in "Feet." By now, you should be a pro at this.

Step 8: Click on the "Secondary File" tab. Click "Select Files." Choose a DBF file and "Browse" to find the **C:\CrimeStat\Data\Lincoln\censusblocks.dbf** file. (This is a file we created. It contains the X and Y coordinates of the centerpoint of each Lincoln census block, plus the total population and number of households.) Set the X variable to the "X" field, the Y variable to the "Y" field, and the Weight variable to the "Households" field.

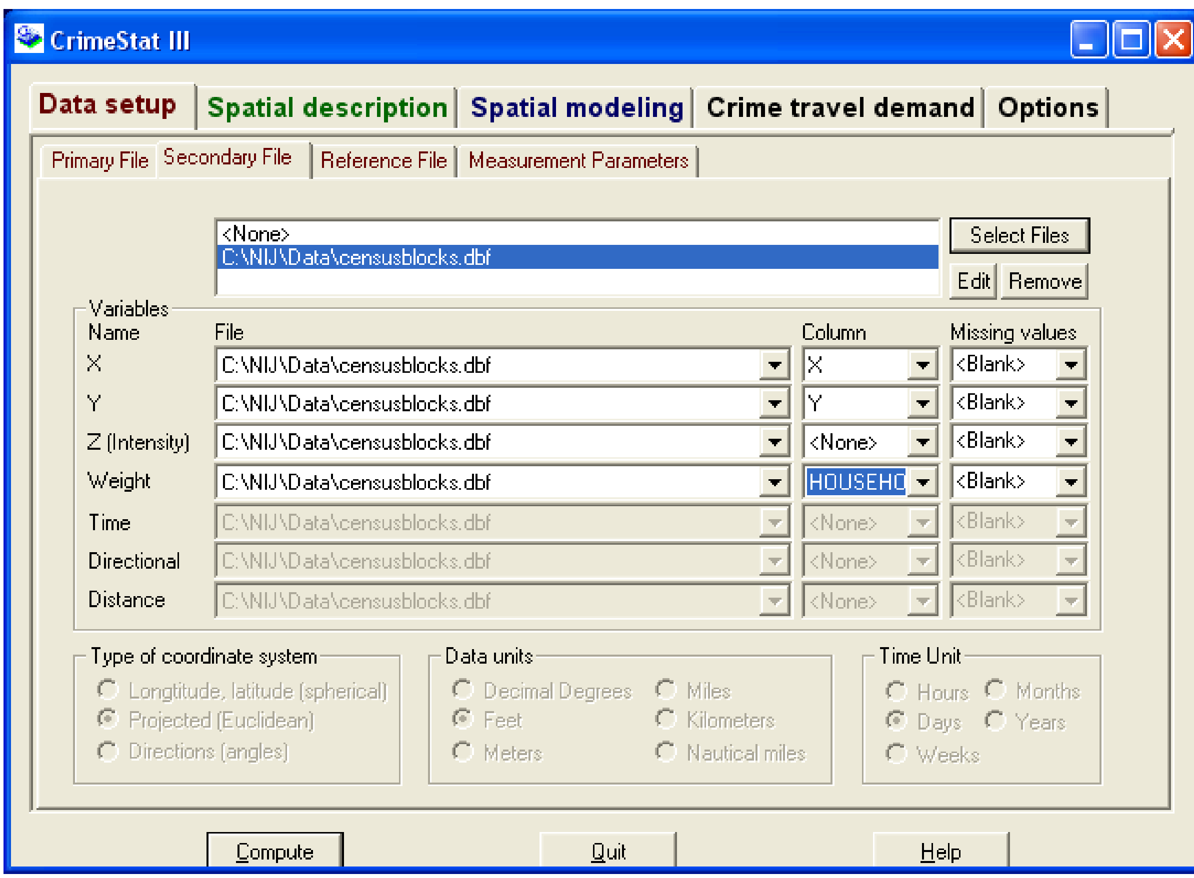


Figure 6-15: the data setup screen for the secondary file

Step 9: Click on the "Spatial Modeling" tab and the "Interpolation" sub-tab. Check the "Dual" box. Set the options as shown in Figure 6-16. Make sure you check the "Use Weighting Variable" option.

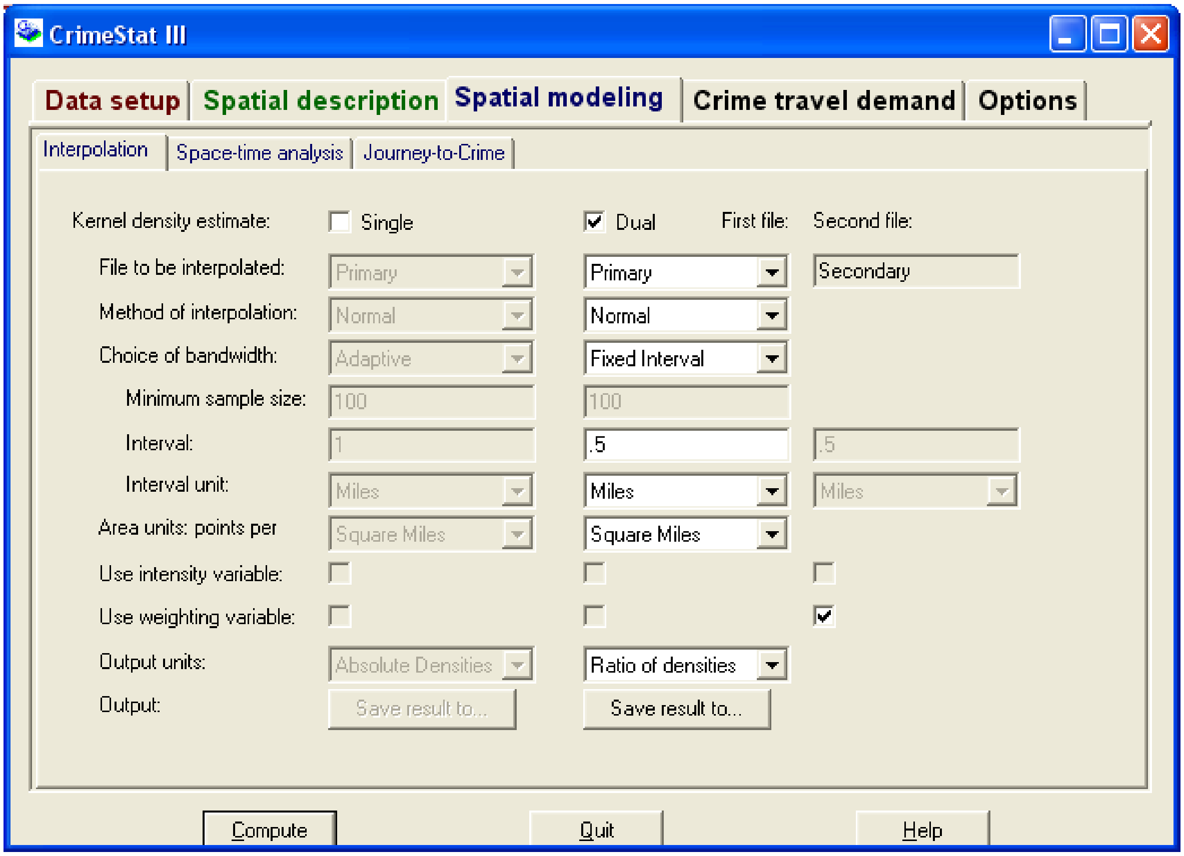


Figure 6-16: Setting our options for a dual KDE for residential burglaries

If we were simply analyzing housebreak volume, we would probably use a uniform or quartic method of interpolation, but the normal method works better for household data, as it distributes some risk across the entire jurisdiction while still keeping higher densities near the centroids of the census blocks. We were not comfortable extending the most significant portion of the risk beyond 0.5 miles, however.

Step 10: Click the "Save Result to..." button and save it as a SHP file at **C:\CrimeStat\Data\Lincoln** with a file name of **resburglaries**. CrimeStat will automatically add a "DK" to that. Now click "Compute".

Step 11: Open the "DKresburglaries" layer in your ArcView project and create a choropleth map based on the density ratio. Your result should look similar to Figure 6-17. Compare the result to Figure 6-18, which shows a KDE based on pure volume of data; what do the differences tell you about residential burglary risk in Lincoln?

NOTE: By changing the Method of Interpolation, the Choice of Bandwidth, the Interval (and Interval unit) and the Output Units, you can drastically change your output. Spend some time looking at the various outputs by changing one parameter at a time (when you change several items at once, it is harder to determine what effect each parameter has on the result).

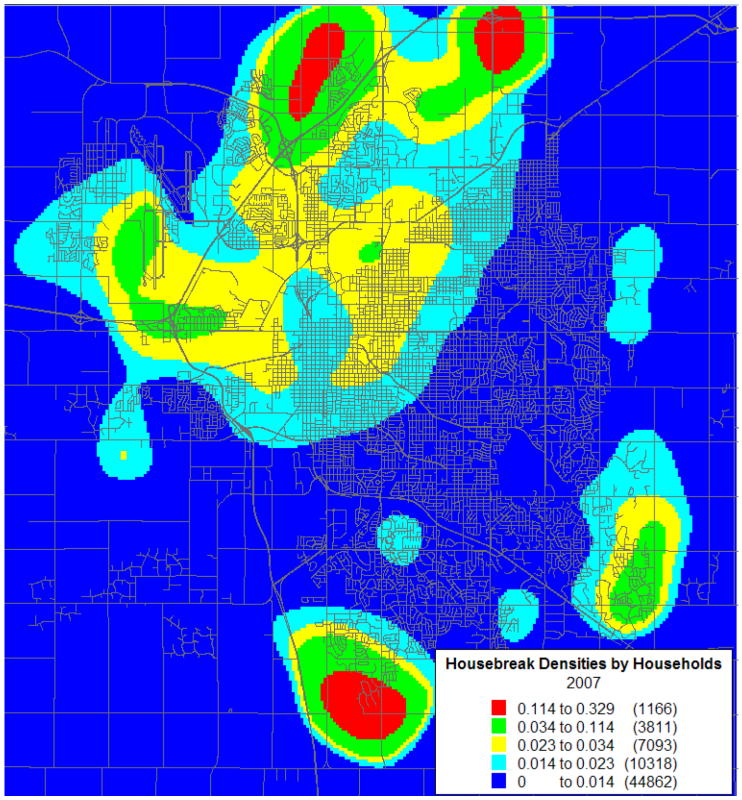


Figure 6-17: Lincoln's residential burglary density, normalized by the number of households

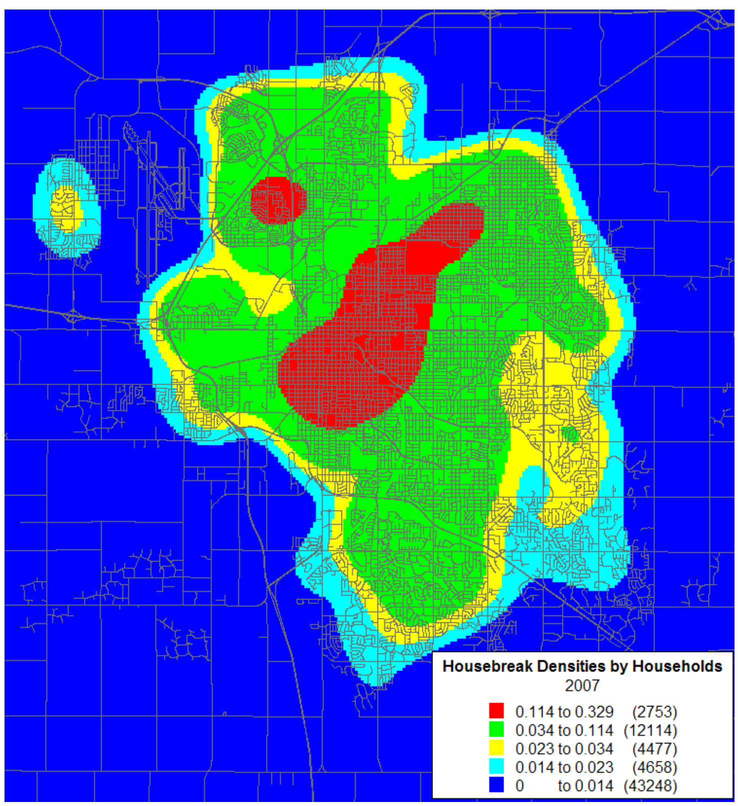


Figure 6-18: Lincoln's residential burglary density, based on raw volume.

6.5 Uses and Cautions

Kernel density estimation is, essentially, another "hot spot" technique, but unlike the techniques reviewed in Chapter 5, KDE's hot spots are part theoretical. Although KDE maps are attractive and fairly intuitive, it's important to keep in mind:

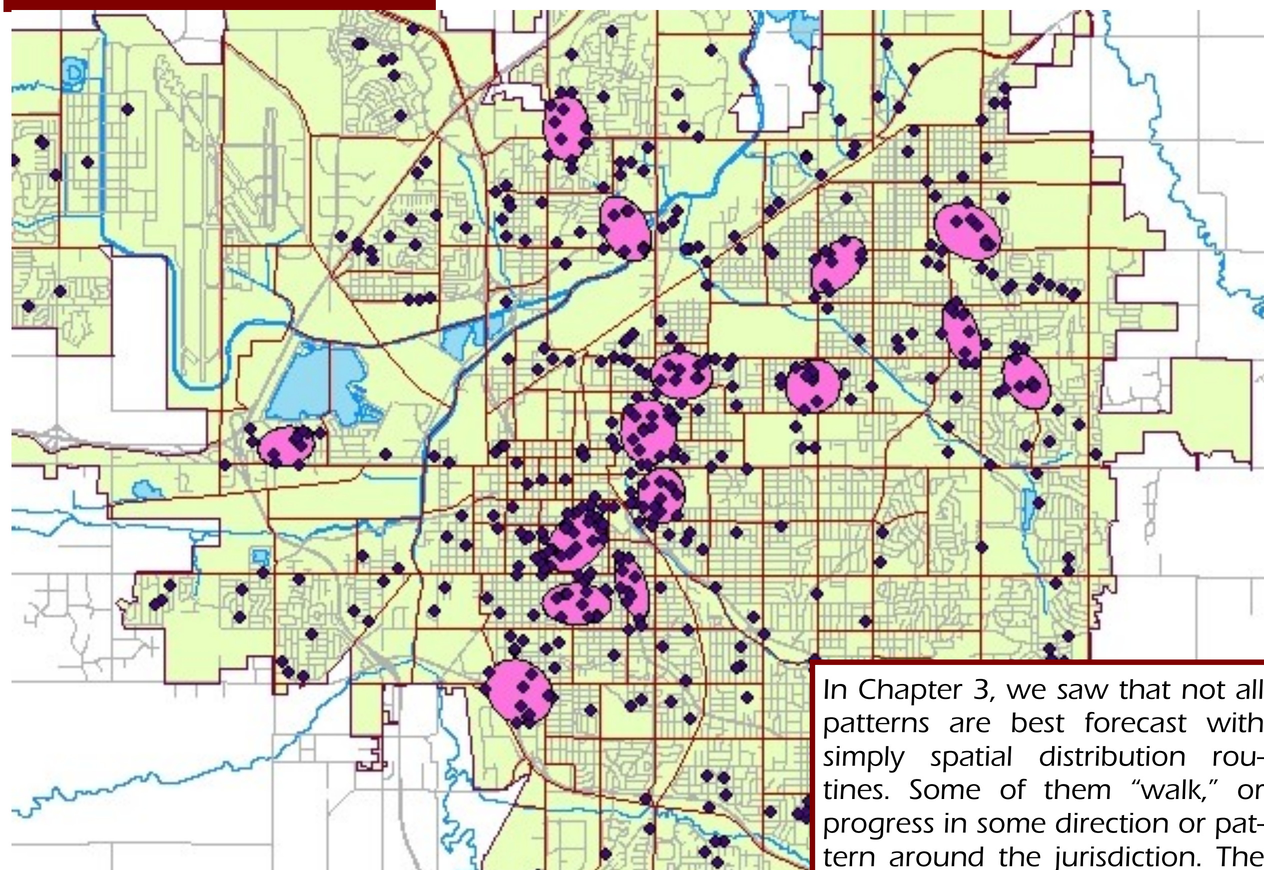
KDE maps are *interpolations*: incidents did not occur at all of the locations within the hottest color.

* Unless you have CrimeStat calculate a dual KDE, KDEs assume a uniform risk surface. This is usually not the case. Residential burglaries cannot occur where there are no residences, and if your jurisdiction has only two banks, they are the only points at risk for bank robberies, no matter what the KDE interpolates between them. Real-life natural barriers such as rivers and highways will not stop kernel radiuses from extending into them, and you will therefore find density estimates in lakes, forests, and fields.

* Intelligent control of the parameters can minimize these issues, but it's always best to interpret a KDE as a relative risk surface *for those suitable targets that may existing within them.* With all those caveats in mind, KDEs can be a powerful and easily-understood tool for tactical and strategic analysis and planning.

7

Spatial Temporal Moving Average



In Chapter 3, we saw that not all patterns are best forecast with simply spatial distribution routines. Some of them "walk," or progress in some direction or pattern around the jurisdiction. The Spatial Temporal Moving Average can help us identify and analyze those types of patterns.

In this Chapter:

- Understanding the Spatial-Temporal Moving Average
- Using a time variable in CrimeStat

7 Spatial Temporal Moving Average

7.0 Introduction

For this final chapter, we will return to a tactical example, covering a single routine: the **spatial-temporal moving average (STMA)**.

In Figure 7-1, we have 18 points representing a fictional convenience store robbery series that hits the city of Lincoln between May 3, 2007 and September 6, 2007. We have used the "Spatial Distribution" routines to capture the mean center, standard distance deviation, and the standard deviation ellipse.

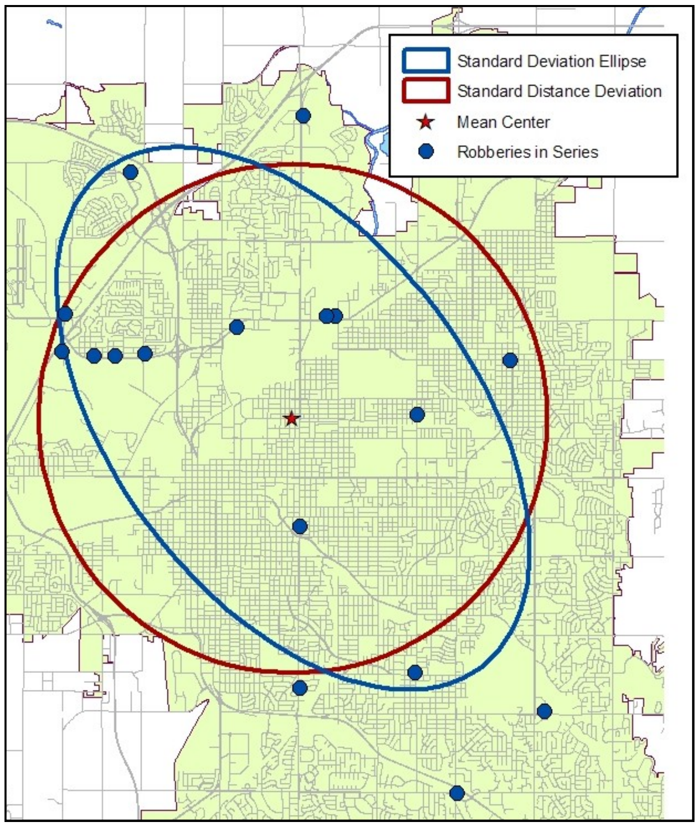


Figure 7-1: a convenience store robbery series with the standard deviation ellipse, standard distance deviation, and mean center.

These calculations helped us back in Chapter 3 when we were studying a residential burglary series, but a look at the map for our convenience store robbery series shows that there's something "off" about them. The mean center does not appear very close to any of the robberies. The ellipses cover such a large swath of downtown Lincoln, most of which has no robberies, that they're useless for tactical deployment purposes.

What happened? If you'll recall from Chapter 3, we indicated there are two basic types of crime patterns, spatially speaking: those that cluster, and those that walk. This one walks. The spatial-temporal moving average will show us just how.

The spatial-temporal moving average calculates the mean center at each point in the series, thereby tracking how it moves over time. The user specifies how many points are included in each calculation using the "Span" parameter.

7 Spatial Temporal Moving Average

Assume we have a moving pattern of 10 incidents and we tell CrimeStat to calculate the spatial-temporal moving average with a span of 3. For each point, CrimeStat calculates the average for that point and the two points on either side of it in the sequence. At the first and last points, it only calculates the moving average for two points, since there is no point before the first one and no point after the last one.

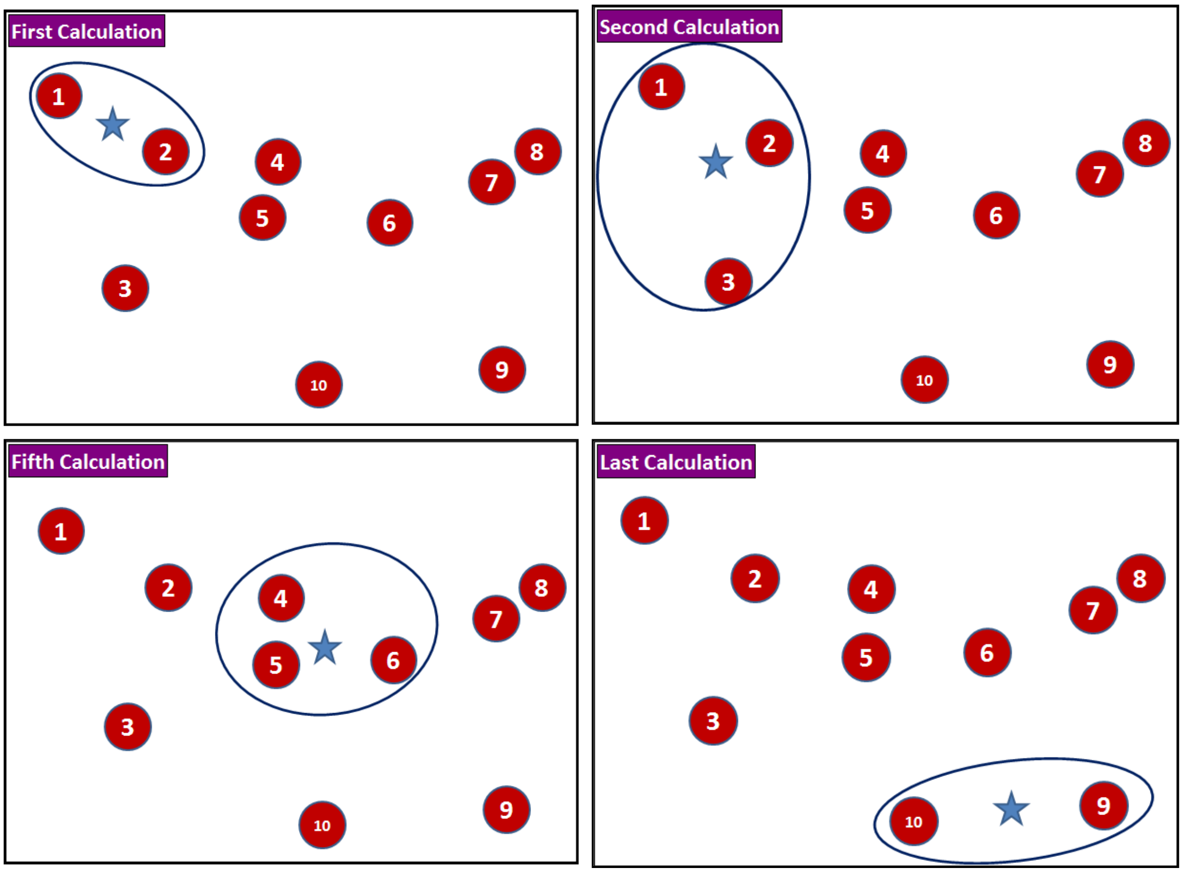


Figure 7-2: points included, and results, in four out of 10 moving average calculations.

The final result is a series of moving average points tied together with a path.

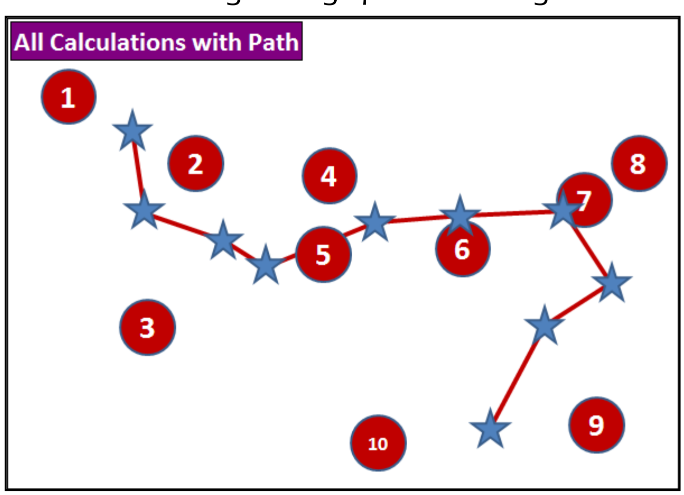


Figure 7-3: all 10 moving average calculations with the path between them

The *span* is the only parameter in the spatial-temporal moving average calculation. We, and the CrimeStat developers, recommend an odd number because it causes the center observation to fall on an actual incident, with an even number on both sides of it. The default of five is generally adequate, although you might want to raise it for very large series. If you go too high, your moving averages will begin to approximate the actual average for the series, and you won't see much movement at all. If you go too low, you'll simply be viewing changes from one incident to the next rather than a true moving average.

Step 1: Start a new CrimeStat session. On the "Data setup" screen choose "Select Files" and load the **C:\CrimeStat\Data\Lincoln\CSRobSeries.shp** file. Set the X coordinate to "X" and the Y coordinate to "Y."

For the first time, we'll be adding a "Time" setting to one of our files. All of the "Space-Time" analysis routines require it; STMA needs it so it will know how the incidents are sequenced. CrimeStat will not accurately calculate actual date/time fields like "06/09/2008" or "15:10." Instead, it requires actual numbers. It doesn't matter where the numbers start as long as the intervals are accurate, so if your data goes from June 1, 2008 to July 15, 2008, you could assign "1" for June 1, "2" for June 2, "31" for July 1," and so on—or you could assign "3000" for June 1 and "3031" for July 1. It's really only the intervals that matter.

Microsoft makes this easy for us. It stores dates as the number of days elapsed since January 1, 1900 and times as proportions of a 24-hour day. In either Access or Excel, we can convert date values to these underlying numbers, so June 1, 2008 becomes 39600, and 15:10 becomes 0.6319. For this file, we have already used Excel to figure the Microsoft date from the actual date, and the field is labeled "MSDate."

Step 2: Set the "Time" row value to "MSDATE." Make sure the "Time Unit" is set to "Days."

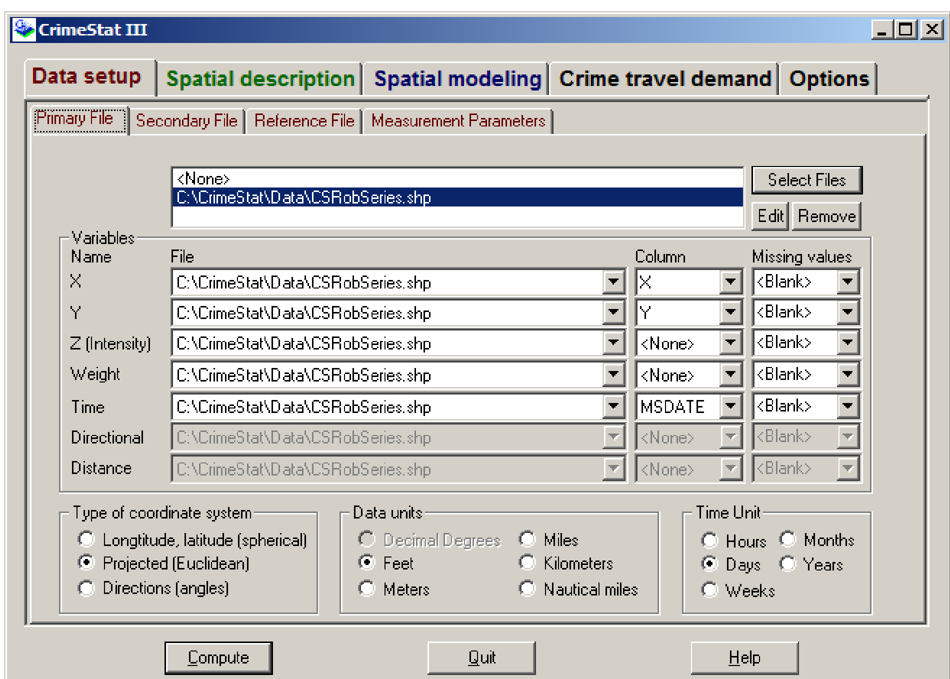


Figure 7-4: data setup for space-time analysis

Step 3: Click on the "Spatial Modeling" tab and then the "Space-time analysis" sub-tab. Check the "Spatial-temporal moving average" box, and leave the span at 5 observations. CrimeStat outputs the moving averages themselves to a DBF file and the path (which it calls a "graph") to a GIS file (such as an ArcView SHP file).

Step 4: Click the "Save output to..." button and save it as a ".dbf" file to **C:\CrimeStat\Data\Lincoln** with the name **CSRobSeriesMA**. Next choose "Save" and "OK".

Step 5: Click the "Save graph" button, choose an "ArcView SHP" file, and save it to **C:\CrimeStat\Data\Lincoln** with the name **CSRobSeriesMA**. "Save" and "OK."

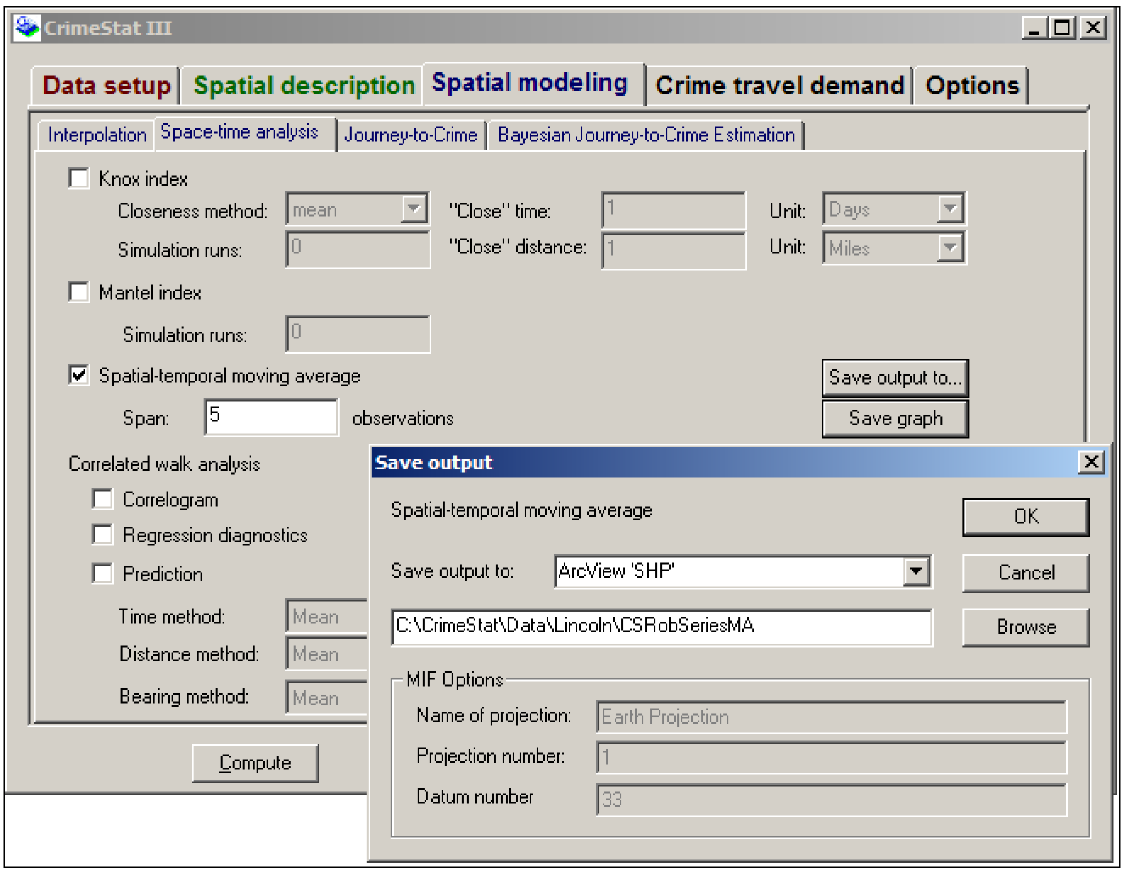


Figure 7-5: preparing to run the spatial-temporal moving average

Step 6: Click "Compute." CrimeStat will run the calculations and save the files into the indicated locations. Switch to your Lincoln ArcView project and load the files.

You will have to create points, based on the X and Y coordinates that CrimeStat saved, to display the mean centers from the CSRobSeriesMA.dbf file. To do this, right click on the file layer in ArcView and choose "Display XY Data". You will then receive a box that says "Display XY Data" at the top. For the "X" Field, you will need to change it from "SPAN" to "Mean X" and the "Y" Field from "SPAN" to Mean Y"

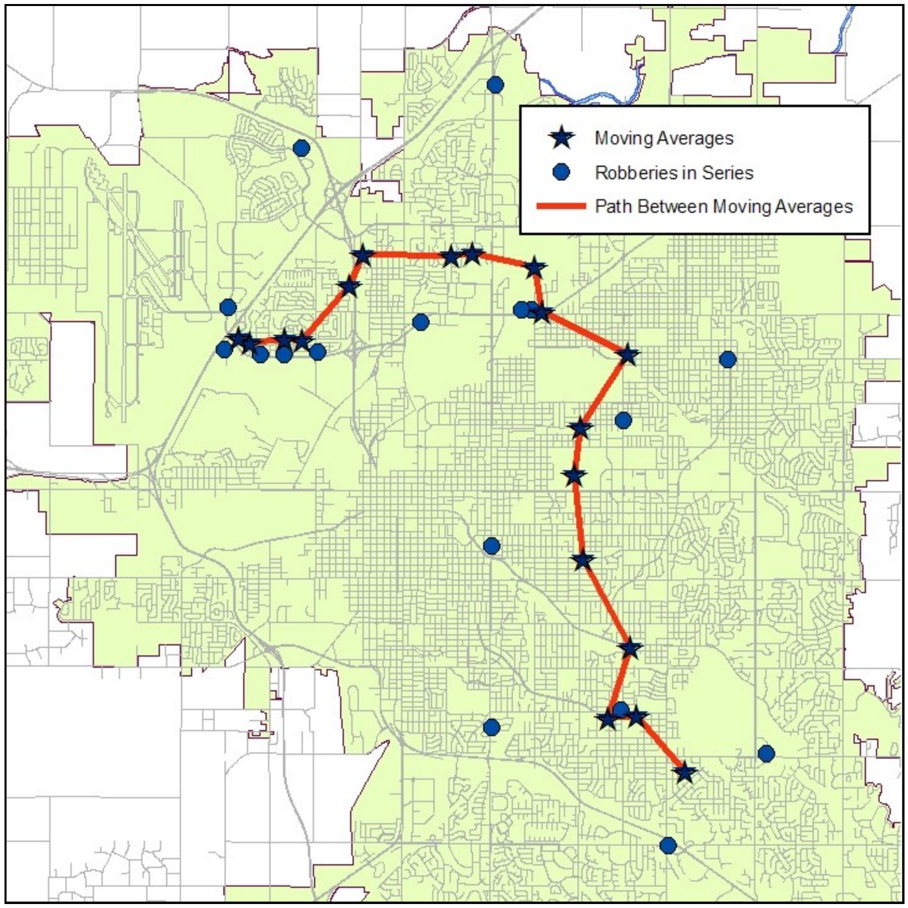


Figure 7-6: The robbery series with our moving averages and path

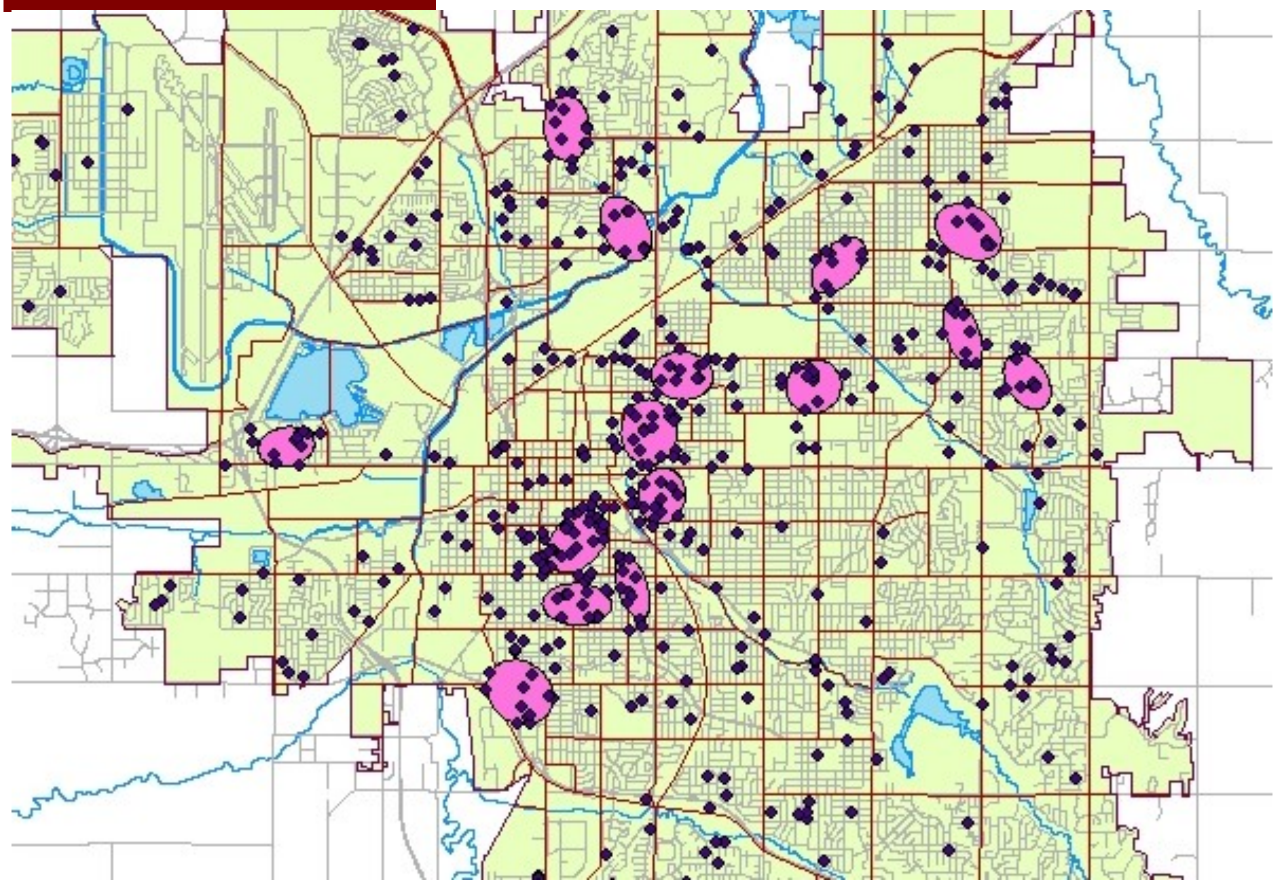
The spatial-temporal moving average gives us a better sense of what's going on in our robbery series. The offender is moving around the city as he commits his robberies, subtly shifting his mean center northward and then westward. Knowing this fact greatly reduces (and shifts) the area in which we will focus our tactical action to the northwest part of the city. **As an added activity, you can run STMA again using the burglaryseries as your primary file.**

STMA does not predict where the next offense will occur (although it can help us make a reasonable guess). For certain types of series, **correlated walk analysis (CWA)**, too complex to cover in this manual, can assist with that. For now, a good rule of thumb is to run the STMA routine on your crime series, and if there isn't a visible pattern of movement, use the spatial distribution techniques to make your forecasts.

Glossary

Correlated walk analysis: .A routine that analyzes the spatial and temporal sequencing of incidents, including the distance, bearing, and time interval, to predict the most likely location of the next incident.

Appendix A
Data Library



LINCOLN POLICE CODING FOR INCIDENT REPORTS

[RD_] Lincoln Police Department reporting district

[ADDRESS] location of occurrence

[INC_] five digit Lincoln Police incident code, containing more detail than the UCR crime type

[TYPE_CODE] four digit code number corresponding to the UCR crime type.

[CALL_TYPE] text definition corresponding to the UCR type code.

[LOC_CODE] two digit Lincoln Police Department location code for the type of premise

[CASE_NUMBE] Lincoln Police Department's unique case number for each incident.

[STATUS_] two digit code for the current investigative status of the case

[TIME_] time reported

[DATE_] date reported

[DAY_FROM] beginning day of week for incidents occurring over a range of days/times

[DATE_FROM] beginning date for incidents occurring over a range of days/times

[TIME_FROM] beginning time of day for incidents occurring over a range of days/times

[DAY_TO] ending day of week for incidents occurring over a range of days/times

[DATE_TO] ending date for incidents occurring over a range of days/times

[TIME_TO] ending time for incidents occurring over a range of days/times

[VICTIM] name of victim

[OFFICER] name and employee number of officer assigned

[LOSS] dollar amount of theft loss

[DAMAGE] dollar amount of property damage

[COMMENTS] summary comments

[ALC] Y/N toggle for alcohol involvement

[DRUG] Y/N toggle for drug involvement

[GANG] Y/N toggle for gang involvement

[BX7] one character code for point of entry

[BX14] one character code for method of entry

LINCOLN POLICE CODING FOR INCIDENT REPORTS (CON'T)

[BX15] one character code for method of vandalism

[BX17] one character code for projectile type

[BX18] one character code for weapon type

[BX13] one character code for victim activity at time of occurrence

[BX19] one character code for how weapon used

[BX8] one character code for injury severity

[BX20] one character code for source of weapon

[ADDRESS2] same as [ADDRESS] with leading zeroes removed

[URL] hyperlink to the Lincoln Police Department case file on this case

[Rpt_Date] same value as [DATE_] in serial format (yyyymmdd)

[RPT_DOW] day of week derived from [DAY_FROM] (Sunday = 1)

[From_Date] same value as [DATE_FROM] in serial format (yyyymmdd)

[FROM_DOW] day of week derived from [DAY_FROM] (Sunday = 1)

[From_Time] numeric field with value derived from [TIME_FROM]

[To_Date] same value as [DATE_TO] in serial format (yyyymmdd)

[To_DOW] day of week derived from [DAY_TO] (Sunday = 1)

[To_Time] numeric field with value derived from [TIME_TO]

[Rpt_Time] numeric field with same value as [TIME_]

[CVLEGEND] generalized field derived from [CALL_TYPE] collapsing values into broader categories

[SYMBOLOGY] generalized field derived from [CALL_TYPE] collapsing values into broader categories (slightly broader than [CVLEGEND])

[iwGeoName] address locator used to geocode this record

[Status] geocoding status of this record

[Score] geocoding match score for this record

[Side] Side of street for this record if geocoded to streets

[NEIGHBORHOOD] name of neighborhood association, if any, in which this record falls

[TEAM_AREA] Lincoln Police Department team area in which this record falls

UCR CODE NUMBERS AND ABBREVIATIONS

CODE #	ABBRV	DEFINITION
1101	MURDER	MURDER
1102	MURDER	MURDER (DEATH BY NEGLIGENCE)
2101	RAPE	RAPE (GUN)
2102	RAPE	RAPE (KNIFE)
2103	RAPE	RAPE (OTHER DANGEROUS WEAPON)
2104	RAPE	RAPE (NO WEAPON)
2201	ATT RAPE	ATT RAPE (GUN)
2202	ATT RAPE	ATT RAPE (KNIFE)
2203	ATT RAPE	ATT RAPE (OTHER DANGEROUS WEAPON)
2204	ATT RAPE	ATT RAPE (NO WEAPON)
3101	ROB GUN	ROBBERY (GUN-ST,ALLEY, ETC)
3102	ROB GUN	ROBBERY (GUN-COMMER.HOUSE)
3103	ROB GUN	ROBBERY (GUN-GAS STATION)
3104	ROB GUN	ROBBERY (GUN-CONVEN. STORE)
3105	ROB GUN	ROBBERY (GUN-RESIDENCE)
3106	ROB GUN	ROBBERY (GUN-BANK)
3107	ROB GUN	ROBBERY (GUN-MISC)
3201	ROB KNIFE	ROBBERY (KNIFE,ST,ALLEY)
3202	ROB KNIFE	ROBBERY (KNIFE-COMMER.HOUSE)
3203	ROB KNIFE	ROBBERY (KNIFE-GAS STATION)
3204	ROB KNIFE	ROBBERY (KNIFE-CONVEN. STORE)
3205	ROB KNIFE	ROBBERY (KNIFE-RESIDENCE)
3206	ROB KNIFE	ROBBERY (KNIFE-BANK)
3207	ROB KNIFE	ROBBERY (KNIFE - MISC)
3301	ROB ODW	ROBBERY (ODW - ST, ALLEY, ETC)
3302	ROB ODW	ROBBERY (ODW - COMMER. HOUSE)
3303	ROB ODW	ROBBERY (ODW - GAS STATION)
3304	ROB ODW	ROBBERY (ODW - CONVEN STORE)
3305	ROB ODW	ROBBERY (ODW - RESIDENCE)
3306	ROB ODW	ROBBERY (ODW - BANK)
3307	ROB ODW	ROBBERY (ODW - MISC)
3401	ROB NO WEAP	ROBBERY (NO WEAP - ST,ALLEY, ETC)
3402	ROB NO WEAP	ROBBERY (NO WEAP - COMMER HOUSE)
3403	ROB NO WEAP	ROBBERY (NO WEAP - GAS STATION)
3404	ROB NO WEAP	ROBBERY (NO WEAP-CONV. STORE)
3405	ROB NO WEAP	ROBBERY (NO WEAP - RESIDENCE)
3406	ROB NO WEAP	ROBBERY (NO WEAP - BANK)
3407	ROB NO WEAP	ROBBERY (NO WEAP - MISC)
4111	ASLT GUN	ASSAULT (UNKNOWN - GUN)
4112	ASLT GUN	ASSAULT (KNOWN - GUN)
4121	ASLT KNIFE	ASSAULT (UNKNOWN - KNIFE)
4122	ASLT KNIFE	ASSAULT (KNOWN - KNIFE)
4131	ASLT ODW	ASSAULT (UNKNOWN - OTHER DANG. WEAP
4132	ASLT ODW	ASSAULT (KNOWN - OTHER DANG.WEAP)
4141	ASLT NO WEAP	ASSAULT (UNK - NO WEAP - AGGRAV)

UCR (con't)

4142	ASLT NO WEAP	ASSAULT (KNOWN-NO WEAP-AGGRAV)
4151	ASLT NO WEAP	ASSAULT (UNK - NO WEAP - NOT AGGRAV
4152	ASLT NO WEAP	ASSAULT (KNOWN-NO WEAP-NOT AGGRAV)
4211	ASLT PO/GUN	ASSAULT PO (UNK - GUN)
4212	ASLT PO/GUN	ASSAULT PO (KNOWN - GUN)
4221	ASLT PO/KNIF	ASSAULT PO (UNK - KNIFE)
4222	ASLT PO/KNIF	ASSAULT PO (KNOWN - KNIFE)
4231	ASLT PO/ODW	ASSAULT PO (UNK-OTHER DANG. WEAP)
4232	ASLT PO/ODW	ASSAULT PO (KNOWN - OTHER DANG.WEAP
4241	ASLT PO/N WP	ASSAULT PO (UNK-NO WEAP-AGGRAV)
4242	ASLT PO/N WP	ASSAULT PO (KNOWN-NO WEAP-AGGRAV)
4251	ASLT PO/N WP	ASSAULT PO (UNK-NO WEAP-NOT AGGRAV)
4252	ASLT PO/N WP	ASSAULT PO(KNOWN-NO WEAP-NOT AGGRAV
5111	BURG RD FE	BURGLARY RD FE
5112	BURG RN FE	BURGLARY RN FE
5113	BURG RU FE	BURGLARY RU FE
5121	BURG RD NFE	BURGLARY RD NFE
5122	BURG RN NFE	BURGLARY RN NFE
5123	BURG RU NFE	BURGLARY RU NFE
5131	BURG RD ATT	BURGLARY ATTEMPT RD FE
5132	BURG RN ATT	BURGLARY ATTEMPT RN FE
5133	BURG RU ATT	BURGLARY ATTEMPT RU FE
5211	BURG NRD FE	BURGLARY NRD FE
5212	BURG NRN FE	BURGLARY NRN FE
5213	BURG NRU FE	BURGLARY NRU FE
5221	BURG NRD NFE	BURGLARY NRD NFE
5222	BURG NRN NFE	BURGLARY NRN NFE
5223	BURG NRU NFE	BURGLARY NRU NFE
5231	BURG NRD ATT	BURGLARY ATTEMPT NRD FE
5232	BURG NRN ATT	BURGLARY ATTEMPT NRN FE
5233	BURG NRU ATT	BURGLARY ATTEMPT NRU FE
6001	LPP $200 +	POCK-PICK ($200 - UP)
6002	LPP $50-199	POCK-PICK ($50 - 199)
6003	LPP $1-49	POCK-PICK ($0 - 49)
6004	LPP ATTEMPT	POCK-PICK (ATTEMPT NO-LOSS)
6101	LPS $200 +	PURSE SNATCH ($200 - UP)
6102	LPS $50-199	PURSE SNATCH ($50 - 199)
6103	LPS $1-49	PURSE SNATCH ($0 - 50)
6104	LPS ATTEMPT	PURSE SNATCH (ATTEMPT - NO LOSS)
6201	LSL $200 +	SHOPLIFTING ($200 - UP)
6202	LSL $50-199	SHOPLIFTING ($50 - 199)
6203	LSL $1-49	SHOPLIFTING ($0 - 49)
6204	LSL ATTEMPT	SHOPLIFTING (ATTEMPT - NO LOSS)
6301	LFA $200 +	LAR FROM AUTO ($200 - UP)
6302	LFA $50-199	LAR FROM AUTO ($50 - 199)
6303	LFA $1-49	LAR FROM AUTO ($0 - 49)
6304	LFA ATTEMPT	LAR FROM AUTO (ATTEMPT - NO LOSS)

UCR (con't)

6401	LAA $200 +	LAR AUTO ACCESS ($200 - UP)
6402	LAA $50-199	LAR AUTO ACCESS ($50 - 199)
6403	LAA $1-49	LAR AUTO ACCESS ($0 - 49)
6404	LAA ATTEMPT	LAR AUTO ACCESS (ATTEMPT - NO LOSS)
6501	LSB $200 +	STOLEN BIKE ($200 - UP)
6502	LSB $50-199	STOLEN BIKE ($50 - 199)
6503	LSB $1-49	STOLEN BIKE ($0 - 49)
6504	LSB ATTEMPT	STOLEN BIKE (ATTEMPT - NO LOSS)
6601	LFB $200 +	LAR FROM BUILDING ($200 - UP)
6602	LFB $50-199	LAR FROM BUILDING ($50 - 199)
6603	LFB $1-49	LAR FROM BUILD ($0 - 49)
6604	LFB ATTEMPT	LAR FROM BUILDING (ATTEMPT-NO LOSS)
6701	LCOM $200 +	LAR FR COIN OPR MACH($200-UP)
6702	LCOM $50-199	LAR COIN OPR MACH($50 - 199)
6703	LCOM $1-49	LAR COIN OPR MACH($0 - 50)
6704	LCOM ATTEMPT	LAR COIN OPR MACH(ATTEMPT-NO LOSS)
6801	LFF $200 +	LARCENY FIELD ($200 - UP)
6802	LFF $50-199	LARCENY FIELD ($50 - 199)
6803	LFF $1-49	LARCENY FIELD ($0 - 49)
6804	LFF ATTEMPT	LARCENY FIELD (ATTEMPT-NO LOSS)
6901	LAR $200 +	LARCENY OTHER ($200 - UP)
6902	LAR $50-199	LARCENY OTHER ($50 - 199)
6903	LAR $1-49	LARCENY OTHER ($0 - 49)
6904	LAR ATTEMPT	LARCENY OTHER (ATTEMPT - NO LOSS)
7100	AT CAR	AUTO THEFT - CAR
7200	AT TRK/BUS	AUTO THEFT - TRUCK/BUS
7300	AT MC	AUTO THEFT - MOTORCYCLE
7350	AT VEH THEFT	AUTO THEFT - OTHER
7400	OSAT	OUTSIDE AUTO THEFT
7410	OSCR	OUTSIDE CRIME REPORTED
7501	CHILD A/N	CHILD ABUSE (PHYSICAL)
7502	CHILD A/N	CHILD ABUSE/NEG (UNSANITARY)
7503	CHILD A/N	CHILD ABUSE/NEG (MENTAL)
7504	CHILD A/N	CHILD ABUSE/NEG (UNATTENDED)
7505	CHILD A/N	CHILD ABUSE (UNATTEND IN VEHICLE)
7506	CHILD A/N	CHILD ABUSE/NEG ABANDONMENT
7508	CHILD A/N	CHILD ABU/NEG (VICTIM UNK)
7509	CHILD A/N	CHILD ABUSE/NEG (OTHER)
7510	CHILD STEAL	CHILD STEALING/KIDNAPPING
7511	CONT MINOR	CONTRIBUTE TO DELINQ OF MINOR
7600	ADULT ABUSE	ADULT ABUSE/NEGLECT
7900	BOMB THREAT	BOMB THREAT
7950	BOMB DEVICE	BOMB DEVICE-NO EXPLOSION
7980	BOMB EXPLODE	BOMB EXPLOSION
8100	ARSON	ARSON
8201	PHONE OBSC	TELEPHONE OBSCENE CALL
8202	PHONE THREAT	TELEPHONE THREATENING CALL
8209	PHONE OTHER	TELEPHONE OTHER CALL

UCR (con't)

8310	SUICIDE-GUN	SUICIDE (BY FIREARMS)
8320	SUICIDE-HANG	SUICIDE (HANGING)
8330	SUICIDE-KNIF	SUICIDE-KNIFE/CUTTING
8340	SUICIDE-OD	SUICIDE (OVERDOSE)
8350	SUICIDE-GAS	SUICIDE (GAS OR ASPHYXIATION)
8360	SUICIDE-JUMP	SUICIDE (JUMPING)
8390	SUICIDE-OTHE	SUICIDE (OTHER)
8410	ATT SUI-GUN	ATTEMPT SUICIDE (GUN)
8420	ATT SUI-HANG	ATTEMPT SUICIDE (HANGING)
8430	ATT SUI-KNIF	ATTEMPT SUICIDE (KNIFE/CUT)
8440	ATT SUI-OD	ATTEMPT SUICIDE (OVERDOSE)
8450	ATT SUI-GAS	ATTEMPT SUICIDE (GAS OR ASPHYXIATE)
8460	ATT SUI-JUMP	ATTEMPT SUICIDE (JUMPING)
8490	ATT SUIC-OTH	ATTEMPT SUICIDE (OTHER)
8601	POSS NARCO	POSSESSION OF NARCOTICS/DRUGS
8602	SELL NARCO	SELL/DELIVER NARCOTICS/DRUGS
8609	NARCO-OTHER	NARCOTICS/DRUGS-OTHER (POSS & SELL)
8610	NARC-LICENSE	NARC STATE LICENSE VIOL
8701	GAMBLE-HOUSE	GAMBLING HOUSE
8702	GAMBLE-SPORT	GAMBLING SPORTS EVENTS
8709	GAMBLE-OTHER	GAMBLING-OTHER
8801	CCW	CARRYING CONCEALED WEAPON
8802	ILLEG WEAPON	ILLEGAL WEAPONS
8809	WEAPON-OTHER	WEAPONS OTHER OFFENSES/VIOLATIONS
9011	FORG CHKS PV	FORGERY (BANK CHECK - PV)
9012	FORG CHKS BV	FORGERY (BANK CHECK - BV)
9013	FORG CHKS GV	FORGERY (BANK CHECKS - GV)
9021	FORG CC PV	FORGERY (CREDIT CARDS - PV)
9022	FORG CC BV	FORGERY (CREDIT CARDS - BV)
9023	FORG CC GV	FORGERY (CREDIT CARDS - GV)
9031	FORG/OTH PV	FORGERY (OTHER - PV)
9032	FORG/OTH BV	FORGERY (OTHER - BV)
9033	FORG/OTH GV	FORGERY (OTHER - GV)
9111	FRAUD FTP PV	FRAUD (FAIL TO PAY - PV)
9112	FRAUD FTP BV	FRAUD (FAIL TO PAY - BV)
9113	FRAUD FTP GV	FRAUD (FAIL TO PAY - GV)
9121	FRAUD ISC PV	FRAUD (INSUF CHECKS - PV)
9122	FRAUD ISC BV	FRAUD (INSUF CHECKS - BV)
9123	FRAUD ISC GV	FRAUD (INSUF CHECKS - GV)
9131	FRAUD NAC PV	FRAUD (NO ACCT CHECKS - PV)
9132	FRAUD NAC BV	FRAUD (NO ACCT CHECKS - BV)
9133	FRAUD NAC GV	FRAUD (NO ACCT CHECKS - GV)
9141	FRAUD CHK PV	FRAUD (OTHER CHECKS - PV)
9142	FRAUD CHK BV	FRAUD (OTHER CHECKS - BV)
9143	FRAUD CHK GV	FRAUD (OTHER CHECKS - GV)
9151	FRAUD CC PV	FRAUD (CREDIT CARDS - PV)
9152	FRAUD CC BV	FRAUD (CREDIT CARDS - BV)
9153	FRAUD CC GV	FRAUD (CREDIT CARDS - GV)

UCR (con't)

9161	FRAUD CON PV	FRAUD (CON GAME - PV)
9162	FRAUD CON BV	FRAUD (CON GAME - BV)
9163	FRAUD CON GV	FRAUD (CON GAME - GV)
9171	FRAUD	FRAUD (CONSUMER FRAUD - PV)
9172	FRAUD	FRAUD (CONSUMER FRAUD - BV)
9173	FRAUD	FRAUD (CONSUMER FRAUD - GV)
9174	FRAUD	FRAUD (VEH TITLES)
9175	FRAUD	FRAUD (VEH MISREPRESENTATION)
9181	FRAUD	FRAUD (OTHER - PV)
9182	FRAUD	FRAUD (OTHER - BV)
9183	FRAUD	FRAUD (OTHER - GV)
9190	FRAUD	FRAUD-FAIL RET RENT ITEM
9195	LAR AS BAIL	FRAUD - LARCENY AS BAILEE
9211	EMBEZZLEMENT	EMBEZZLEMENT
9300	COUNTERFEIT	COUNTERFEITING US CURRENCY
9411	VAND/RESID	VANDALISM (ALL RESID. NOT VEH)
9422	VAND/COM	VANDALISM (COMMERCIAL=NO VEH)
9433	VAND/SCHOOL	VANDALISM (SCHOOLS-NOT VEH)
9443	VAND/VEH	VANDALISM (ALL VEHICLES)
9444	VAND/GOVT	VANDALISM (GOV'T PROP-NO VEH)
9455	VAND/PARKS	VANDALISM (PARK PROP-NO VEH)
9488	VAND/DRV BY	VANDALISM (DRIVE BY SHOOTING)
9499	VAND/OTHER	VANDALISM (ALL OTHER)
9500	MS PERS	MISSING PERSON ADULT
9510	MS PERS JUV	MISSING PERSON JUVENILE
9520	JUV RUNAWAY	MISSING PERSON JUV RUNAWAY
9611	PROSTITUTION	PROSTITUTION - SOLICIT
9612	PROSTITUTION	PROSTITUTION - PANDERING
9621	PORNOGRAPHY	PORNOGRAPHY
9631	VICE/OTHER	VICE - OTHER
9711	SODOMY	SODOMY
9722	IND EXPOSURE	INDECENT EXPOSURE
9733	MOLEST	MOLEST/FONDLING
9744	RAPE(STAT)	RAPE (STATUTORY)
9755	SEX OF/OTHER	SEX OFFENSE (OTHER - NOT CLASSIFIED
9766	INCEST	INCEST
9800	LOST/STOLE	LOST OR STOLEN ARTICLE
9810	FALSE INFO	GIVE FALSE INFORMATION
9910	DEATH	DEATH-NATURAL CAUSE
9940	DEATH ACCID	DEATH ACCIDENTAL (NON-TRAFFIC)

STATUS CODES

10 Unfounded
30 Inactive
40 Cleared-Adult
41 Cleared-Juvenile
42 Cleared by exception-Adult
43 Cleared by expection-Juvenile
50 Open-Active Followup

M.O. Field Values for Incident Reports

POINT OF ENTRY

FIELD	VALUE	MEANING
BX7	A	Door
BX7	B	Overhead Door
BX7	C	Sliding Door
BX7	D	Window
BX7	E	Skylight
BX7	F	Roof/Top
BX7	G	Wall
BX7	H	Transom
BX7	I	Hood
BX7	J	Trunk
BX7	K	Gas Cap
BX7	O	Other
BX7	U	Unknown

METHOD OF ENTRY

FIELD	VALUE	MEANING
BX14	A	Broke Glass
BX14	B	Hid in Building
BX14	C	Kicked
BX14	D	Removed Door
BX14	E	Removed Glass
BX14	F	Removed Screen
BX14	G	Pried
BX14	H	Slipped Lock
BX14	I	Drilled Lock
BX14	J	Twisted/cut Knob
BX14	K	Cut Hole
BX14	L	Already Unlocked
BX14	M	Smash and Grab
BX14	N	No force, open
BX14	O	Other
BX14	P	Cut Screen
BX14	Q	Used Key
BX14	U	Unknown

VANDALISM

FIELD	VALUE	MEANING
BX15	C	Cut/Slashed
BX15	H	Kicked/Hit W-item
BX15	I	Item Thrown
BX15	L	Pulled/Twisted
BX15	M	Multiple Methods
BX15	P	Paint/Writing
BX15	S	Shot/Projectile
BX15	T	Scratched

PROJECTILE

FIELD	VALUE	MEANING
BX17	A	BB/Pellet
BX17	B	Bullet
BX17	C	Marble/Bearing/Shot
BX17	D	ROCK/BRICK

TYPE OF WEAPON

FIELD	VALUE	MEANING
BX18	A	Handgun
BX18	B	Shotgun
BX18	C	Rifle
BX18	D	Multiple Firearms
BX18	E	Knife/Cutting Inst
BX18	F	Club/Blunt Object
BX18	G	Explosive/Incendia
BX18	H	Hands/Feet or Teet
BX18	I	Gas/Exhaust
BX18	J	Pills/Drugs
BX18	M	Multiple Types
BX18	N	None
BX18	O	Other

HOW WEAPON USED/TYPE OF SUICIDE

FIELD	VALUE	MEANING
BX19	A	Implied, Not Seen
BX19	B	Threatened Only
BX19	C	Fired/Detonated
BX19	D	Stabbed/Cut With
BX19	E	Thrown
BX19	F	Hit/Kicked
BX19	G	Bit
BX19	H	Choked/Asphyxiate
BX19	I	Scratched
BX19	J	Jumped/Fellr Latc
BX19	K	Hanging
BX19	L	Overdose
BX19	M	Electrocuted

INJURY SEVERITY

FIELD	VALUE	MEANING
BX8	M	Minor/Superficial
BX8	N	None
BX8	S	Aggravated/Serious

WEAPON WAS:

FIELD	VALUE	MEANING
BX20	A	Brought By Suspect
BX20	B	Taken from Victim
BX20	C	Previously at Scene

VICTIM WAS:

FIELD	VALUE	MEANING
BX13	A	Closing Business
BX13	B	Opening Business
BX13	C	Making Bank Deposit
BX13	D	Taking out Garbage
BX13	E	Jogging/Walking
BX13	F	Riding Bicycle
BX13	G	Making a Delivery
BX13	H	Inv. in Prostitution
BX13	I	On Vacation
BX13	J	At funeral/wedding
BX13	K	At School
BX13	L	At Work
BX13	O	At Other Location
BX13	P	Present At Time

Location Codes

PUBLIC OPEN AREAS
01 STREET
02 ALLEY
03 SIDEWALK
04 PARKING LOT (GOV'T OWNED)
05 PARKING GARAGE (GOV'T OWNED)
05 LOT OR FIELD (GOV'T OWNED NOT PARK)
68 EMERGENCY TELEPHONE
69 INTERNET / CYBERSPACE / WEB
77 BIKE PATH / RECREATIONAL TRAIL
78 SKYWALK SYSTEM
PRIVATE OPEN AREAS
07 COMMERCIAL PAY PARKING LOT
08 COMMERCIAL PARKING GARAGE
09 LOT OR FIELD (NOT YARDS/LAWNS)
37 CONSTRUCTION SITE
96 CEMETARY
PARKS / RECREATION
10 CITY PARKS
10 CITY PARK PROPERTY / FACILITIES
77 BIKE PATH / RECREATIONAL TRAIL
LIQUOR ESTABLISHMENTS
11 PRIVATE CLUB
12 TAVERN / BAR / LOUNGE
13 LIQUOR STORE (OFF SALE ONLY)
RETAIL BUSINESSES
14 APPLIANCE STORE
15 AUTOMOBILE DEALER
15 AUTOMOTIVE PARTS OR SUPPLIES
16 BICYCLE DEALER / PARTS / SUPPLY
17 DISCOUNT STORE
17 DEPTARTMENT STORE
18 DRUGSTORE / PHARMACY
19 ELECTRONICS / STEREO
20 FURNITURE STORE
21 GROCERY / FOOD STORE
22 HARDWARE STORE
23 PAWN SHOP
23 SECOND HAND / CONSIGNMENT STORE
24 RESTAURANT
25 GAS / SERVICE STATION (PRIMARY BUSINESS)
26 GUN SHOP
26 SPORTING GOODS STORE
27 JEWELRY STORE
28 OTHER RETAIL BUSINESS
29 AMUSEMENT HALL / POOL HALL / ARCADE
36 CONVENIENCE STORE
38 VIDEO STORES
39 CHECK CASHING / ADVANCE
SERVICE INDUSTRY
30 APPLIANCE REPAIR
31 AUTO REPAIR
32 ELECTRONIC REPAIR
33 PLUMBING / ELECTRIC / HEATING / AIR ETC.
34 OTHER REPAIR SERVICE
OFFICE PROFESSIONAL BUILDING
35 OFFICE / PROFESSIONAL BUILDING
97 MEDICAL / DENTAL / DOCTOR / CLINIC
WHOLESALE
40 WHOLESALE DRUGS / MEDICAL
41 WHOLESALE AUTO PARTS
42 WHOLESALE LIQUOR
43 WHOLESALE ELECTRONIC
44 WHOLESALE OTHER
MANUFACTURING
50 MANUFACTURING / FACTORIES
TRANSPORTATION SERVICES
55 BUS / RAILROAD TERMINALS AND FACILITIES
55 STAR TRAN
55 RAILROAD TRACKS / PROPERTY
56 AIRPORT
SERVICE FACILITIES
57 LIBRAY
58 MUSEUM
UTILITIES
60 LINCOLN ELECTRIC SYSTEM
60 LINCOLN WATER SYSTEM
60 AQUILA
60 CABLE TV FACILITIES
60 TELEPHONE / CELL FACILITIES
OTHER BUSINESSES
61 BANKS & FINANCIAL INSTITUTIONS
62 STORAGE FACILITIES
63 SPORT FACILITIES
64 MOVIE THEATER
64 AUDITORIUM / THEATER /CONCERT HALL
PUBLIC BUILDINGS
70 COUNTYCITY BUILDING
71 COUNTYOWNED FACILITY
72 CITYOWNED FACILITY
73 STATE OWNED FACILITY
74 FEDERALOWNED FACILITY
INSTITUTIONS
75 JAIL / PRISON
75 YOUTH DETENTION CENTER
75 STATE CORRECTIONAL FACILITY
75 REGIONAL CENTER
75 DRUG / ALCOHOL RESIDENTIAL TREATMENT

Location Codes (con't)
75 DETOX CENTER
75 CORNHUSKER PLACE DETOX
76 RESIDENTIAL INSTITUTION
76 NURSING HOME
76 GROUP HOME
76 FREEWAY STATION / CEDARS
RELIGIOUS INSTITUTIONS
96 CHURCH / SYNAGOGUE / TEMPLE
96 MONASTARY / CONVENT / PARSONAGE
96 OTHER RELIGIOUS FACILITIES (EXCEPT SCHOOLS)
MEDICAL FACILITIES
97 HOSPITAL
97 MEDICAL / DENTAL OFFICES
97 DENTIST / DOCTOR OFFICE
97 MEDICAL CLINIC
SCHOOLS
87 PRESCHOOL
88 PUBLIC ELEMENTARY SCHOOL
89 PUBLIC MIDDLE SCHOOL
90 PUBLIC HIGH SCHOOL
91 UNIVERSITY / COLLEGE / COMMUNITY COLLEGE
92 PRIVATE ELEMENTARY
93 PRIVATE MIDDLE SCHOOL
94 PRIVATE HIGH SCHOOL
95 PRIVATE BUSINESS / TRADE SCHOOL
47 UNL ACADEMIC BUILDING
48 UNL SUPPORT FACILITY
RESIDENCES / DWELLINGS
80 SINGLE FAMILY
81 MOBILE HOME
82 DUPLEX
83 MULTI UNIT / APARTMENT (36)
84 MULTI UNIT / APARTMENT (7+)
85 BOARDING HOUSE
86 HOTEL / MOTEL
45 UNL APARTMENT
46 UNL RESIDENCE HALL
47 OFF CAMPUS HOUSING (GREEK HOUSE)
47 UNL FRATERNITY / SORORITY
UNKNOWN LOCATIONS
99 UNKNOWN

Incident Codes

CODE INCIDENT CLASSIFICATION

01111 ABANDONED BICYCLE
01222 ABANDONED VEHICLE
55111 ABDUCTION/KIDNAPPING
55222 ABDUCTION/KIDNAPPING ATTEMPTED
03000 ACCIDENTS P.D. NOT REPORTABLE
03021 ACCIDENTS P.D. NOT REPORTABLE H&R
03111 ACCIDENTS P.D. REPORTABLE
03121 ACCIDENTS P.D. REPORTABLE H&R
03211 ACCIDENTS INJURY
03221 ACCIDENTS INJURY H&R
03311 ACCIDENTS FATALITY
03321 ACCIDENTS FATALITY H&R
31333 ACCIDENTS OUTSIDE OUR JURISDICTION
09900 ADULT ABUSE
92333 AIRCRAFT EMERGENCY
14000 ALCOHOL DRUNK
14100 ALCOHOL DWI
14300 ALCOHOL BUY/PROCURE FOR MINORS
14400 ALCOHOL MINOR IN POSSESSION
14500 ALCOHOL MINOR ATTEMPT TO PURCHASE
14600 ALCOHOL USING FALSE I.D.
14710 ALCOHOL POSSESSION IN PARK
14720 ALCOHOL CONSUMPTION IN PUBLIC
14730 ALCOHOL OPEN CONTAINER IN PUBLIC
14900 ALCOHOL LIQUOR LICENSE VIOLATION
14800 ALCOHOL SALE TO MINOR
14999 ALCOHOL UNL POLICY VIOLATION
31223 OUTSIDE O.P.S. ALCOHOL/CHEMICAL TESTING
44000 ANIMAL ALL ANIMAL CALLS OTHER THAN DOGS
44111 ANIMAL DOG BARKING
44144 ANIMAL DOG BITE TO PERSON
44133 ANIMAL DOG INJURED/DEAD
44155 ANIMAL DOG LICENSE VIOLATION
44122 ANIMAL DOG RUNNING LOOSE
44199 ANIMAL OTHER DOG CALL
45111 ARSON
05000 ASSAULT NON DOMESTIC
05100 ASSAULT DOMESTIC
05200 ASSAULT VICTIM UNDER PROT. ORDER
05300 ASSAULT OF POLICE OFFICER
41000 AUTO THEFT
41000 UNAUTHORIZED USE OF MOTOR VEHICLE
31500 AUTO THEFT STOLEN ELSEWHERE, FOUND IN LINCOLN
47333 BOMB EXPLOSION
47444 BOMB SUSPICIOUS PACKAGE/DEVICE

Incident Codes

CODE INCIDENT CLASSIFICATION

47222 BOMB THREAT
47111 BOMB UNKNOWN CALL TYPE
09000 BRIBERY
06000 BURGLARY
09100 CHILD ABUSE/NEG ABANDONMENT
09400 CHILD ABUSE/NEG LOST OR FOUND CHILD
09500 CHILD ABUSE/NEG PHYSICAL ABUSE
09200 CHILD ABUSE/NEG UNATTENDED IN VEHICLE
09300 CHILD ABUSE/NEG UNATTENDED, OTHER
09700 CHILD ABUSE/NEG UNSANITARY CONDITIONS
09600 CHILD ABUSE/NEG VERBAL OR EMOTIONAL
09800 CHILD ABUSE/NEG COMPULSORY EDUCATION
09000 CHILD ABUSE/NEG OTHER
40166 CITY LICENSE INVESTIGATION
48000 CRIME PREVENTION PRESENTATION
48100 CRIME PREVENTION OTHER THAN PRESENTATION
11100 DEATH ACCIDENTAL
11200 DEATH NATURAL
11510 DEATH MANSLAUGHTER/MURDER
11520 DEATH MANSLAUGHTER BY NEGLIGENCE
11530 DEATH JUSTIFIABLE HOMICIDE
11300 DEATH SUICIDE
11999 DEATH UNK CAUSE
13000 DISTURBANCE DOMESTIC
12401 DISTURBANCE BETWEEN NEIGHBORS
12311 DISTURBANCE WILD PARTY
12341 DISTURBANCE FIREWORKS
12500 DISTURBANCE SMOKING IN PUBLIC
12100 DISTURBANCE PUBLIC URINATION
12000 DISTURBANCE ALL OTHER
51000 EMBEZZLEMENT
15111 ESCAPEE FROM LPD CUSTODY
31811 ESCAPEE OTHER THAN FROM LPD
40181 ESCORT FUNERAL
40182 ESCORT MILITARY
40183 ESCORT MONEY TRANSFER
40184 ESCORT HAZ. MATERIAL
40185 ESCORT OTHER
47333 EXPLOSION OF BOMB/INCENDIARY DEVICE
16000 ACCIDENTAL EXPLOSION OF BUILDING, GAS PIPE, ETC
52000 EXTORTION
04111 FALSE ALARMS EMPLOYEE
04222 FALSE ALARMS CUSTOMER
04333 FALSE ALARMS MECHANICAL
04444 FALSE ALARMS OTHER
04999 FALSE ALARMS UNK CAUSE

Incident Codes

CODE INCIDENT CLASSIFICATION

21000 FALSE INFORMATION GIVEN TO OFFICER
21500 FALSE INFORMATION CITIZEN VICTIM
16555 FIRE ALARM
16666 FIRE FALSE ALARM
16000 FIRE NO ALARM
12341 FIREWORKS DISTURBANCE
12344 FIREWORKS ILLEGAL POSS/SALE BY ADULT
12345 FIREWORKS ILLEGAL POSS/SALE BY JUVENILE
12346 FIREWORKS POSS/SALE LEGAL FRWKS BY JUVENILE
17100 FORGERY CHECKS
17200 FORGERY CREDIT CARDS
17400 FORGERY COUNTERFEIT MONEY/COINS
17300 FORGERY ALL OTHER
17500 FORGERY FALSE INSURANCE CERTIFICATE
54222 FOUND ARTICLES OTHER THAN ABAND. BIKE OR VEH.
29200 FRAUD CHECKS INSUF/NO ACC/NOT FORGED
29500 FRAUD CON GAMES/DECEPTION/FALSE PRETENSE/OTHER
29300 FRAUD CREDIT CARDS/ATM/BANK CARD
29600 FRAUD FAIL TO RETURN RENTED/BORROWED ITEM
29100 FRAUD FAIL TO PAY
29400 FRAUD IMPERSONATION
18110 GAMBLING BET OR WAGER ON SPORTS
18120 GAMBLING BET OR WAGER OTHER THAN SPORTS
18200 GAMBLING TAMPER WITH SPORTS EVENT
18300 GAMBLING OPERATE/PROMOTE/ASSIST GAMBLING
18400 GAMBLING BUY/SELL/POSSESS EQUIPMENT
24122 HAZARDS NONTRAFFIC (WIRES,TREES,ETC)
28000 HAZARDS TRAFFIC HAZARDS OTHER
28333 HAZARDS TRAFFIC SIGN MISSING
28444 HAZARDS TRAFFIC SIGNAL MALFUNCTION
24133 HEALTH/SANITATION VIOLATIONS OF HEALTH CODES
56100 INHALE INTOXICANT/HARMFUL VAPORS
25211 JUVENILE MISSING PERSON/RUNAWAY
22000 JUVENILE NOT OTHERWISE CLASSIFIED
55111 KIDNAPPING
55222 KIDNAPPING ATTEMPTED
23100 LARCENY POCKET PICK
23200 LARCENY PURSE SNATCH
23300 LARCENY SHOPLIFTING
23310 LARCENY GAS FROM SELF SERVICE PUMP
23400 LARCENY FROM BUILDING
23500 LARCENY FROM MOTOR VEHICLE
23600 LARCENY MOTOR VEH. ACCESSORIES
23700 LARCENY FROM COIN OPERATED MACH.
23800 LARCENY STOLEN BIKE

Incident Codes

CODE INCIDENT CLASSIFICATION

23900 LARCENY OTHER OR FROM OPEN AREA
24144 LITTERING
54112 LOST CELL PHONE
54111 LOST OR STOLEN ITEM
56100 MEDICAL INHALE INTOXICANT OR HARMFUL VAPORS
56366 MEDICAL SUICIDE ATTEMPT
56000 MEDICAL EMERGENCY (OTHER)
56466 MENTAL INVESTIGATION
56400 MENTAL INVESTIGATION W/EPC
24000 MISC OTHER
25111 MISSING PERSON ADULT
25211 MISSING PERSON JUVENILE,(INCLUDES RUNAWAY)
31700 MISSING PERSON OUTSIDE JURISDICTION
27000 NARCOTICS OTHER
27111 NARCOTICS POSSESSION
27222 NARCOTICS SALE/DELIVER
27333 NARCOTICS MANUFACTURE
27444 NARCOTICS OVERDOSE
27555 NARCOTICS NO TAX STAMP
31811 OUTSIDE ESCAPEE
31223 OUTSIDE O.P.S. CHEMICAL TESTING
31333 OUTSIDE O.P.S. ACCIDENT
31222 OUTSIDE O.P.S. OTHER
31444 OUTSIDE TRANSPORT PERSON OR PRISONER
31100 OUTSIDE CRIME ALL REPORTS OUTSIDE OUR JURISDICTION
72111 PARADE
82111 PARKING BLOCKING/INTERFERE WITH STREET
82122 PARKING BLOCKING ALLEY
82133 PARKING BLOCKING DRIVE
82222 PARKING 24 HOUR
82333 PARKING SIDEWALK SPACE
82444 PARKING SNOW EMERGENCY
82666 PARKING TRUCK OR TRAILER IN RESIDENTIAL
82888 PARKING TOW FOR UNPAID TICKETS
82000 PARKING OTHER
39100 PORNOGRAPHY
24166 PROPERTY DAMAGE UNINTENTIONAL,NON TRAFFIC
39210 PROSTITUTION
39220 PROSTITUTION PROMOTE/ASSIST
72122 PROTEST/MARCH
24255 PROTECTION ORDER VIOLATION
05200 PROTECTION ORDER VIOLATION WITH ASSAULT
35111 PROWLER
36000 RECEIVE STOLEN PROPERTY
31500 RECOVER STOLEN VEHICLE STOLEN O.P.S.
37000 ROBBERY

Incident Codes

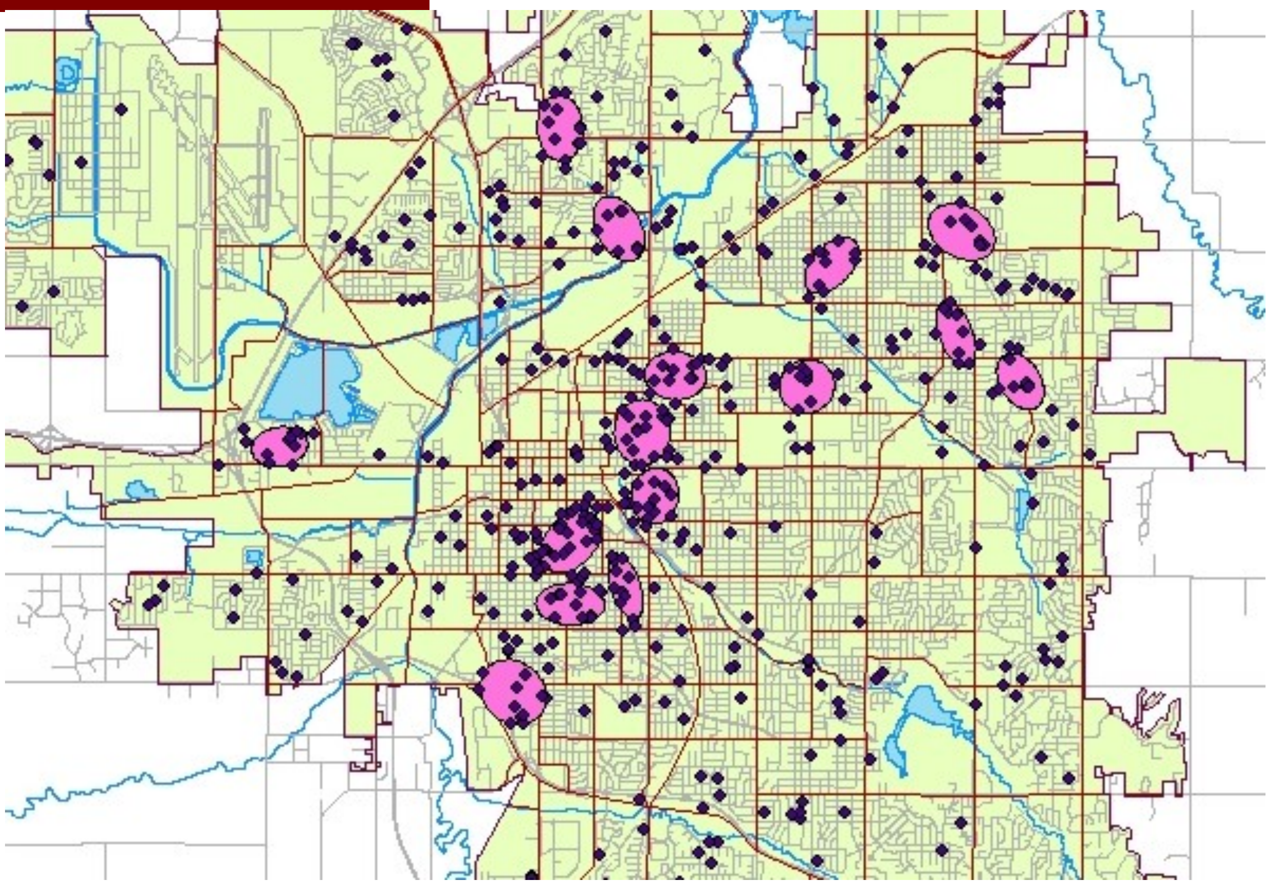
CODE INCIDENT CLASSIFICATION

25211 RUNAWAY OR MISSING JUVENILE
38100 SEX OFFENSE RAPE
38180 SEX OFFENSE RAPE, ATTEMPTED
38600 SEX OFFENSE RAPE, STATUTORY
38800 SEX OFFENSE INCEST
38700 SEX OFFENSE INDECENT EXPOSURE
38500 SEX OFFENSE MOLEST/FONDLING
38300 SEX OFFENSE SEXUAL ASLT WITH AN OBJECT
38200 SEX OFFENSE SODOMY
38900 SEX OFFENSE OTHER
24222 SOLICIT/SELL WITHOUT LICENSE
40111 SS ASSIST CITIZEN
40112 SS ASSIST MOTORIST
40222 SS CHECK WELFARE OF PERSON
40122 SS CIVIL STANDBY
40144 SS LOCKED OUT
40155 SS OPEN DOOR FOUND BY OFFICER
40133 SS OTHER
40333 SS BUILDING/ROOM LOCK/UNLOCK (UNL)
40444 SS DIGNITARY PROTECTION
24222 SOLICIT/SELL WITHOUT A LICENSE
35555 STALKING
24233 STRIKES LABOR DISPUTES
11300 SUICIDE
56366 SUICIDE ATTEMPT
35222 SUSPICIOUS PERSON
35333 SUSPICIOUS VEHICLE
35444 SUSPICIOUS ITEM OR PACKAGE
61222 TELEPHONE OBSCENE CALLS
61000 TELEPHONE OTHER
61009 TELEPHONE OTHER/PROT ORDER
61111 TELEPHONE THREATENING CALLS
61119 TELEPHONE THREATENING CALLS/PROT ORDER
94102 TRAFFIC DIRECTION
94444 TRAFFIC IMPROPER REGSISTRATION
94000 TRAFFIC OTHER
94666 TRAFFIC SPEED SURVEY
94555 TRAFFIC SUSPENDED DRIVER
82888 TRAFFIC UNPAID TICKETS
94449 TRAFFIC WHEEL TAX VIOLATION
94777 TRAFFIC RADAR TRAILER
31444 O.P.S. TRANSPORT PERSON OR PRISONER
24244 TRESPASSING
41000 UNAUTHORIZED USE OF MOTOR VEHICLE
42200 VANDALISM BY DRIVE BY SHOOTING

Incident Codes

CODE INCIDENT CLASSIFICATION
42100 VANDALISM BY GRAFFITI
42000 VANDALISM OTHER
70111 WARRANT SERVICE
08111 WEAPONS CONCEALED
08222 WEAPONS DISCHARGE IN CITY
08333 WEAPONS FELON IN POSSESSION
08555 WEAPONS ILLEGAL WEAPON
08000 WEAPONS OTHER
08999 WEAPONS UNL POLICY VIOLATION

Appendix B
Glossary



Glossary

Absolute densities are simply the sum of all the weights received by each cell—but re-scaled so that the sum of the densities equals the total number of incidents.

Aneslin's Local Moran statistic: compares geographic zones to their larger neighborhoods and identifies those that are unusually high or low

ArcGIS: a geographic information system software line made by ESRI for Redlands, California. The ArcGIS family includes ArcReader, ArcView, ArcEditor, and ArcInfo (all different "levels" of the same product) and their associated extensions. ArcView is the most commonly used among law enforcement agencies

Bandwidth refers to the shape and size of the kernel is ; can be specified to some degree by the user. The shape of the bandwidth is specified by the *method of interpolation*

Mean center of minimum distance represents the point at which the sum of the distance to all the other points is the smallest

Convex hull polygon encloses the outer reaches of the series. No point lies outside the polygon, but outliers increase the size of the polygon

Correlated Walk Analysis A routine that analyzes the spatial and temporal sequencing of incidents, including the distance, bearing, and time interval, to predict the most likely location of the next incident.

Crime mapping: the application of geographic information systems to crime and police data. Crime mapping helps analyze the "where" factor of crime patterns, series, trends, and problems.

Distance analysis techniques answer questions about the dispersion of incidents, and hot spot analysis to identify areas where crimes concentrate.

Forecasting: identifying the most likely locations and (in some techniques) times of future events

Fuzzy Mode: identifies the geographic coordinates, plus a user-specified surrounding radius, with the highest number of incidents

Geocoding: the process of assigning geographic coordinates to data records, usually based on the street address. Geocoding is what turns a list or database of crimes into a map of those crimes

Geographic information system: Hardware and software that collects, stores, retrieves, manipulates, queries, analyzes, and displays spatial data. GIS is a computerized fusion of maps with underlying databases that provide information about map objects

Geographic profiling: an investigative methodology that determines the most probable areas of offender residence or "home base" based on the locations of crimes in a series. Geographic profiling is used to triage tips and prioritize investigations

Geometric mean / Harmonic mean are two alternate measures of the mean center

Hot spot: areas of concentrated crime. Hot spots can be single addresses, parcels, blocks, neighborhoods, or even entire cities, depending on the overall scale of the analysis

Glossary

Journey-to-crime: literally, the distance an offender travels, and the routes he takes, between his home or "base of operations" and crime location. Research has identified the typical distance ranges of journeys for various types of crimes

Kernel density estimation (KDE, also KDI: kernel density interpolation) *generalizes* data over a larger region

Kernel density maps (also known, with varying degrees of accuracy, as surface density maps, continuous surface maps, density maps, isopleths maps, grid maps, and "hot spot" maps) are the most popular map type in crime analysis

Kernel Density Interpolation: a spatial modeling technique

K-Means Clustering: the user specifies the number of clusters and CrimeStat positions them based on the density of incidents

MapInfo: a geographic information system produced by Pitney Bowes MapInfo Corporation of North Greenbush, New York. MapInfo is the second most commonly-used GIS in law enforcement

Mean center is the intersection of the mean of the X coordinates and the mean of the Y coordinates. It is the simplest of the statistics and has been calculated for years by analysts who plot their incidents on Cartesian planes. The mean center can be weighted, although that would not make sense in a crime series where every incident has a weight of one

Median center is the intersection between the median of the X coordinates and the median of the y coordinates. As with the median in non-spatial statistics, it is useful if outliers are wreaking havoc with your mean center

Mode: identifies the geographic coordinates with the highest number of incidents

Nearest neighbor hierarchical clustering (NNH) takes analysis to the next logical level by actually identifying the clusters

Nearest-Neighbor Hierarchical Spatial Clustering: builds on the nearest neighbor analysis (NNA) by identifying clusters of incidents

Operations analysis: the study of a police agency's allocation of resources, such as officer distribution by shift and beat

Pattern: two or more incidents with direct common causal factors. Patterns are usually short-term phenomena. The most common types of patterns are series, in which the common causal factor is the same offender

Probabilities divide the density by the total number of incidents. The result is the chance that any incident occurred in that cell

Problem: an elusive term that generally refers to a long-term or chronic crime phenomenon based on opportunity rather than on a single offender or group of offenders. The term generally encompasses both the incidents themselves and their underlying causes

Relative densities divide the absolute densities by the area of the grid cell. Thus when explaining the map, you can say that the red represents X points per square mile rather than X points per grid cell

Series: two or more crimes of the same or similar type, committed by a single offender

Spatial & Temporal Analysis of Crime (STAC): an alternate means of identifying clusters by "scanning" the points and overlaying circles on the map until the densest concentrations are identified

Standard deviation ellipse improves on the standard distance deviation by accounting for skewed distributions, minimizing the amount of "extra space" that would appear in some standard distance deviation circles. CrimeStat calculates ellipses that are both one and two standard deviations from the mean center

Standard distance deviation circle approaches the standard deviation calculation differently. It calculates the linear distance from each point to the mean center point, then draws a circle around one standard deviation from the center point

Spatial statistics: mathematical techniques that apply descriptive and multivariate statistics to spatial data, usually using geographic coordinates

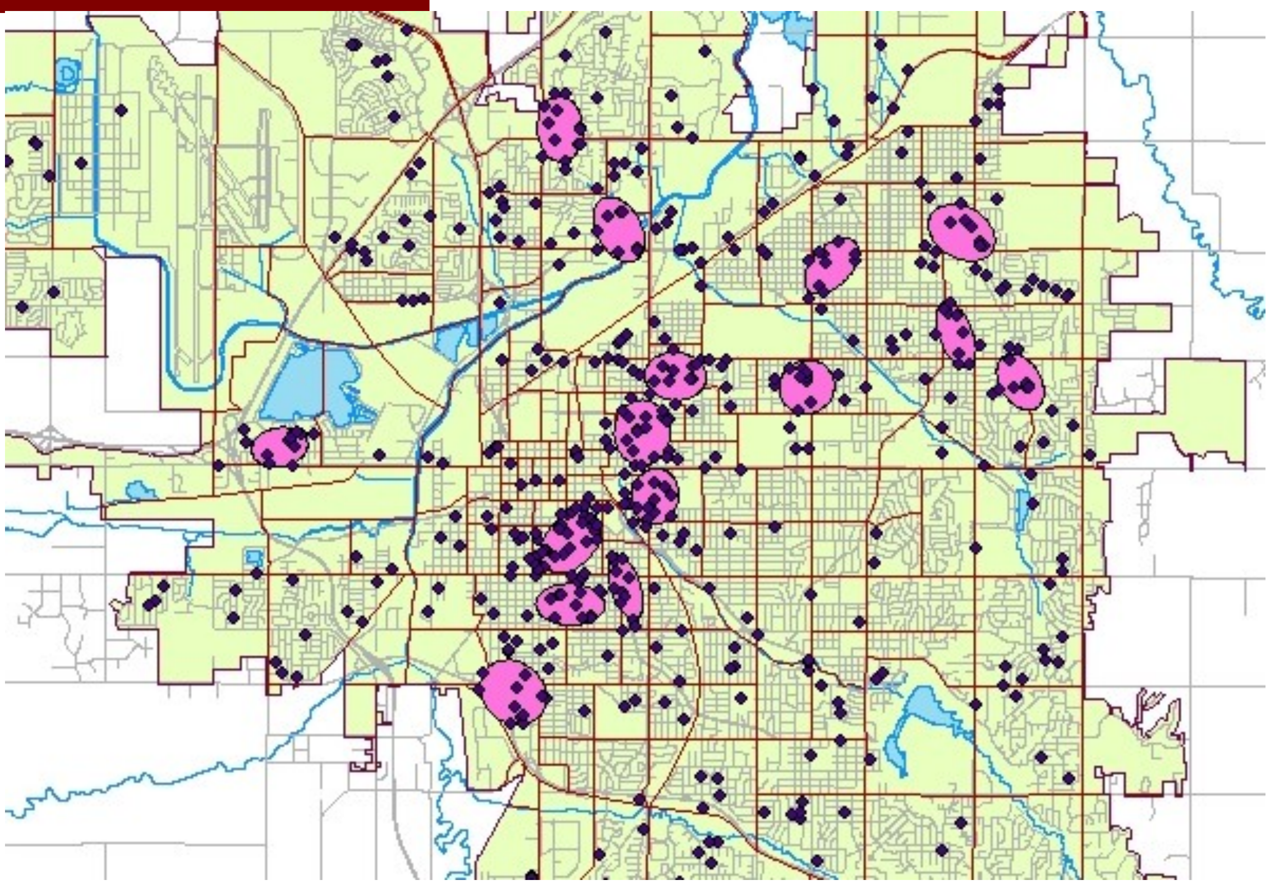
Strategic crime analysis: the analysis of long-term trends or crime problems

Tactical crime analysis: the regular search for emerging crime patterns and series, the techniques used to examine and describe the characteristics of these patterns, the methods by which this information is disseminated to a police agency, and the responses that the agency designs and implements

Thematic map: a map that shows a "theme" or tells a story rather than simply providing a visual representation of the earth's surface. Popular thematic maps used in crime mapping are choropleth maps (polygons colored or patterned based on the volume of crime); graduated symbol maps (larger dots at locations with more crimes); and point symbol maps (different colors or symbols based on the type of crime or some other attribute feature)

C

Appendix C
GIS Functions



CrimeStat III (CSIII) requires specific data formats and elements from ArcGIS map layers. CSIII can either read .shp files directly or data from the .dbf file of the map layer. Regardless of which option is used, there are data elements that must be included in all data files used by CSIII and there are optional data elements needed depending on the routines to be used. To meet these requirements there are tasks in ArcGIS that must be carried out on data in preparation for analysis. In addition, after data are analyzed in CSII, ArcGIS is necessary for joining output, as well as, visualizing that output.

Knowing how to do many, if not all, of these tasks will keep the course on schedule. If you are not familiar with ArcGIS, or many of the functions listed in the attachment, then I highly recommend you purchase *Getting to Know ArcGIS Desktop: The Basics of ArcView, ArcEditor, and ArcInfo Updated for ArcGIS 9* by Robert Burke, Eileen Napoleon and Tim Ormsby, ISBN: 158948083X.

General ArcGIS Functions Required for use with CrimeStat III.

These are general tasks that make using ArcGIS less cumbersome for exploring, visualizing, sampling and preparing data for use in CSIII.

Loading and Displaying Data. Loading data into the GIS, setting layer sequence and displaying layers appropriately will be the foundation for which to prepare data for use in CSIII, as well as, viewing output from the analysis results. For example, once data it loaded map layers will likely need to be rearranged for viewing, it might be necessary to see what attribute data is included or how to change characteristics of the map layer.

Navigating and Exploring Spatial and Attribute Data. These tasks are necessary for getting to know [exploring] the data in graphical and map forms. This provides opportunity for visualizing trends in data that will help guide observation selection and parameter specification for analysis in CSIII. For example, data distributions can be viewed in a histogram allowing for a change the thematic map classifications to see how the distributions change in direction, concentration, etc... over space.

Manipulating the Database. These tasks are required for preparation of attribute data values for use in CSIII. For example, creating new fields, setting field parameters and populating/manipulating those new fields with values to be used with the functions in CSIII. For example, a column of data may need summarized and recoded to be used as an intensity value in CSIII.

Manipulating Geometry. These functions are required to create new or aggregated geometric features, such as points, lines or polygons, to be used in various functions of CSIII. For example, assigning of points to polygons for enumerating observation counts within a boundary, join a boundary's centroids with its boundary after analysis in CSIII, or combine neighboring jurisdictions.

Selecting and Exporting Data. These functions will be required to for sub-setting data for specific or comparison analysis. For example, selection of observations within particular jurisdictions for sub regional analysis. Saving [exporting] these sub-sets of data off into new files for use in CSIII will be necessary. Selection can be done both spatially and with attributes through queries. Using the Spatial and Attribute Query dialog boxes will be used frequently. You should be more familiar with the different types of spatial queries than the attribute queries.

Specific ArcGIS Functions Required for use with CrimeStat III.

These are specific functions in ArcGIS that will be required for preparing data for use in CSIII or for formatting output from CSIII for visualization in ArcGIS. Some functions require multiple steps that are not laid out in a sequence in documentation what CSIII expects as input. Therefore, some notes have been provided under those functions.

Point-in-Polygon. This function is required as there are instances when the distribution of boundary [census tracts, for example] centroids will be analyzed using weighted values. Thus, the resulting output from CSIII will need to be reassigned back to the polygon is belongs to for thematic visualization purposes and other spatial analysis in the GIS.

Creation of a new column in that attribute table of a layer. This function is necessary for the addition of X and Y coordinate values of a polygon centroid or a point when only the .dbf file is opened in CSIII. If a .shp file is opened in CSIII it automatically recognizes the X and Y columns that are stored in the geometry file of the set of files that make up a map layer. However, if the .dbf file is opened the user must specify a column that contains a pair of latitude and longitudinal values of the observations.

Selecting a sub-set of geographic objects. This function is primarily for spatial sampling and sub-setting data for specific or comparison analysis in CSIII. Often, patterns in a single jurisdiction need to be separated from the larger set of observations for more micro scale analysis or comparison with a neighboring jurisdiction, which requires the selection of that jurisdiction and being exported out a separate map layer. As mentioned above, the use of Spatial and Attribute Query dialog boxes will be used frequently. Familiarity with the different types of spatial queries will be helpful.

Calculation of the X and Y coordinate values from the centroid of a polygon or directly from a point. This function is fundamental because to conduct analysis on aerial units (polygons) the centroid must be created. For CSIII this is done by creating an X column for the longitude and a Y column for the latitude. These columns should be of data type Double with a precision of 16 and a scale of 8, that is, there should be 8 numbers to the left and right of the decimal point. However, an actual point can be made from the X and Y columns so as to have a .shp file that contains an actual geometric point representing the centroid.

Note: Instructions on how exactly to do this, including code to calculate the values, are in the Help file. Open the Help file and using the *Find* tab and search using the term "Making Field Calculations." In the results section look for the listing of the same term and double-click on that listing. In the dialog window click on the entry that states "Adding the x,y coordinates in a point layer to a new field" or "Adding the x,y coordinates of a polygon layer to a new field." The former calculates the coordinate pair of each observation while the latter calculates the coordinate pair for the centroid of a polygon [boundary].

Centroid creation from a polygon. This will be required if a .shp file is to be directly opened in CSIII as opposed to a .dbf file that contains centroids of a boundary file.

Polygon attribute assignment to a centroid. This is only required if a centroid layer is going to be used directly in CSIII as a .shp file. Otherwise, if a layer is going to be opened as a .dbf file, then columns that have captured the X & Y coordinates of the polygon centroid will be necessary.

Coding or recoding data for missing values. This function uses the Field Calculator in the attribute display window by right-clicking on the field that is to be recoded. This is necessary to keep CSIII from treating values other than missing. CSIII recognizes several different values as missing. For example, zeros can not be used as they are recognized a legitimate coordinates, which are the Greenwich Mean for longitude and the Equator for latitude. Thus they need to be recoded to have a value that CSIII will recognize as a missing value.

Calculating area for polygons and lengths for lines. These are necessary for calculating the size of the study area and length of a street network as it relates the observations under study. Guessing can cause analysis results to be inefficient in that too much area lends itself to skewed analysis results as observations might be grouped into once section of the area, which makes the group appear as a cluster when it really is not. Too little area may cause the converse.

Note: The best way to do this is to create a new field that will hold the values for wither the area of each polygon boundary or line segment feature and calculate the appropriate value. Instructions on how exactly to do this, including code to calculate the values, are in the Help file under "Making Field Calculations." Open the Help file and using the *Find* tab and search using the term "Making Field Calculations." In the results section look for the listing of the same term and double-click on that listing. In the dialog window click on the entry that states "Updating Area/Length for a Shapefile." Once this has been done the values can be added up by opening the attribute file and right-clicking on the column that holds the area/length values and select 'Summarize' for get the sum.

Symbolizing and Classifying Data. There are several options for displaying attribute values of geographic features. In ArcGIS there are five methods for symbolizing data under which there are several techniques for classifying the attribute values. The 'feature' and 'quantities' methods will primarily be used.

Buffer Creation. This technique is primarily for spatial sampling and sub-setting data for analysis in CSIII. Unlike selecting from within a set of boundary features, buffers can be created off of any geographic feature (i.e. point, line or polygon) so that data can be subset in ways that are unique to the area and patterns can be analyzed in proximity to specific geographic features.

Identifying Minimum Bounding Rectangle (MBR) Coordinates. This is necessary so that area can be calculated for analysis routines in CSIII, as well as, some output functions. In some routines CSIII uses the MBR of the point data set as the MBR if values are not specified. This, however, may introduce scaling problems when conducting analysis and geographic output results may be unnecessarily large. Also, to be activated, there are some routines that require MBR coordinates.

Note: Currently, the only way to do this is to manually place the mouse cursor in the lower left and upper right map window corners, write down the coordinates and type into CrimeStat III under the Reference File tab in the Grid Area input section.