

## **ArcGIS Crime Mapping Workbook**

### ***Introduction***

We have now arrived at the workbook section of the textbook. The workbook is designed to provide students with a hands-on method for learning the basics of how to use ArcGIS to conduct crime mapping regardless of individual understanding or experience level with computers or GIS. The workbook is divided into three main sections with each section building on the knowledge learned in earlier sections. The first section, “Getting Started with GIS” takes the student through a review of the basic features of ArcGIS, data sources, data entry, data query and map-making tips all from a crime mapping perspective. In the second section, “Basic Spatial Analysis”, the student begins to conduct spatial analysis of crime data, using techniques such as buffer zones, spatial distribution measures, and distance analysis measures. Importantly, students will not only learn how to perform these analyses but will also learn associated problems and limitations with each technique as well as how to interpret the results from both a theoretical and practical standpoint. The third section introduces the student to more advanced spatial analysis techniques (using Crimestat 2.0) such as hot spot analysis, density analysis, and spatial dispersion analysis. As with other sections, an emphasis will be placed not only on learning how to conduct these techniques but limitations of each technique and how to interpret the results logically. Each chapter is designed to incorporate hands-on exercises to facilitate learning and has accompanying sample data as well as accompanying internet figures and resources for each chapter.

As briefly mentioned above, the workbook is highly integrated with the internet. Specifically, software, sample data, figures and resources for each chapter are available

for student use. Importantly, as this book is not shipped with ArcGIS software, Crimestat 2.0 software, or the sample data used in the workbook, it is incumbent on the student to visit the text webpage in order to fully use the workbook. Moreover, all figures referenced in the text of the workbook are viewable only on the textbook webpage. Finally, this workbook was written to work with ArcGIS 8.2 and should be compatible with newer versions of ArcGIS, but the authors offer no express guarantees. Importantly, as newer versions of ArcGIS are released and their adoption amongst users increases newer versions of the workbook will be written and made available on-line. The webpage associated with the text is [www.mappingcrime.org](http://www.mappingcrime.org), and it provides a wealth of resources related to not only the workbook, but also crime mapping and the spatial analysis of crime in general. Thus, before getting started with the workbook it is strongly recommended that the student visit the text webpage and follow the instructions on how to download the sample data as well as any software (ArcGIS & Crimestat 2.0) that the student doesn't already have access to.

Finally, it is important to state up front that this workbook is in *no way meant to substitute for an in-depth class on the use of a GIS or a class on spatial statistics*. While this workbook will provide students with an introductory understanding of how to use ArcGIS for crime mapping purposes, it is by no means meant to be a comprehensive workbook on how to use ArcGIS and all of its features. Moreover, while the workbook will discuss spatial statistics techniques, it is designed to be an introduction to the use and interpretation of these techniques and not a comprehensive spatial statistics textbook. The primary purpose of this workbook is to provide students with an introductory hands-on experience in conducting spatial analysis of crime events using a GIS. At the end of

each chapter are “further readings” sections that provide students with a list of books and articles that offer a more in-depth coverage of topics discussed in the chapters. It is recommended that students who wish to learn more about topics discussed in these chapters or about the full functionality of Arcview and other GIS products pursue these texts and articles.

## **Chapter 1: Getting Started with ArcGIS**

### **Goals of Chapter 1:**

1. Accessing the textbook webpage
2. Adding data to a view.
3. Examine basic ArcMap tools and features.
4. Save your work

### ***Getting Started***

The first chapter will introduce students to ArcGIS ArcMAP and give them an overview of some of the basic functions and functionality of ArcGIS. While this workbook was written using ArcGIS version 8.2 it should be compatible with newer versions of ArcGIS, although the authors offer no express or implied guarantees to this compatibility. The ArcGIS workbook was written in order to assist those people who have access to the cutting edge software from ESRI how to conduct basic crime mapping. Because ArcGIS is a newer software platform than Arcview 3.3 it has not been as widely adopted amongst those in the crime mapping field and thus the Arcview 3.3 workbook will remain the default workbook included with the text for the foreseeable future. While not as ubiquitous as Microsoft Windows, Arcview 3.3 is arguably the most commonly used GIS software in the world and in the field of crime mapping in particular. While other GIS software such as MapInfo is commonly used in crime mapping, Arcview 3.3

has a very large and well established customer base and thus knowledge in how to use Arcview 3.3 will serve students well in the professional world of crime mapping.

As for the content of chapter 1, students will learn how to open ArcMAP (the mapping program within the ArcGIS program suite), open a view and add layers (map data), explore ArcMAP tools within a view, and saving your work. Importantly, this chapter is designed **only** to give the student an introduction to ArcMAP and **some** of its basic features. This chapter is by no means a substitute for a comprehensive class or textbook explaining the complete functionality of ArcMAP. Students who desire a more in-depth instruction of ArcGIS and ArcMAP in particular are encouraged to review the “Further Readings” list at the end of the chapter.

### *Accessing the Textbook Webpage*

In order for the student to use the workbook they must first access the textbook webpage and download the sample data. The sample data and their accompanying codebooks are designed specifically for students using this text and are easily and quickly downloadable from the textbook webpage. In addition, students who do not have access to a copy of ArcGIS can learn how to acquire a trial copy for educational purposes. Finally, the webpage also assists the student in downloading a free copy of Crimestat 2.0 for use in section III of the workbook.

1. In the browser on your computer type [www.mappingcrime.org](http://www.mappingcrime.org). This will bring you to the main menu of the webpage associated with Spatial Aspects of Crime: Theory and Practice.

2. On the main menu of the webpage click on DATA. This will take you to the page dealing with downloading the data and its associated codebooks.
3. Follow the on screen instructions in order to download and open the sample data.

### ***Obtaining a copy of ArcGIS or Crimestat 2.0***

**ArcGIS:** On the main page of the textbook webpage click on the main menu item labeled “Arcview”. Follow the onscreen instructions on how to obtain a trial copy of ArcGIS

**Crimestat 2.0:** On the main page of the textbook webpage click on the main menu item labeled “Crimestat”. Follow the onscreen instructions in order to download a free and fully functional version of Crimestat 2.0 and its accompanying manual.

### ***Opening ArcMAP***

1. On a Windows desktop click on the Start button and go to the Programs folder and look for the ArcGIS folder. In a typical install of ArcGIS the program is installed within a folder named ArcGIS.
2. Hold the cursor over the ArcGIS folder until ArcMap pops up then move the cursor over ArcMap and click on the program. This will bring up the welcome screen for ArcMap, giving you three options 1) Start using ArcMap with “a new empty map”, the default option, 2) open a “Template” or a ready-to-use layout and base map for a certain geographic area, or 3) open an existing map.
3. Select the default option “A new empty map” and click OK which will take you to the main ArcMap application screen called the “Data View”.

### *Adding data to the View*

In ArcMap map data are called **Layers** and they can represent almost any type of data that has a geographic component. Examples of crime related layers include individual crime locations, city streets, county boundaries, even satellite photos and other images called orthophotos (discussed in Workbook Chapter 2). Importantly, until data is added to the Data View the majority of buttons will not work.

1. In order to add a data layer click on the button with the plus on it. Alternatively you can click on the File menu at the top of the screen and select “Add data”.  
After clicking on the add data button the add data dialogue box will appear.  
Scroll through the folders in the C drive until you come to the “CM Lab Exercises” folder and select this folder.
2. Within the CM Lab Folder click on the Chapter 1 folder. After opening this folder you should see two different files, one named Houston.shp and one named Homicides.shp. While these are Arcview shapefile they are easily opened and used in ArcMap. Importantly, while Arcview 3.3 themes all have three associated files (.shp, .dbf, and .shx) ArcMap layers save all relevant data in a Layer file (.lyr), making it easier to transfer data between computers and save for archival purposes.
3. Double click on the file named Houston.shp. This will add the theme to the table of contents of the Data View window. You will now see a map of Harris County Texas, the county that contains Houston Texas.

4. In order to make the data viewable or not take the cursor and click inside the checkbox to the left of the layer name. You will notice that by clicking in the box and removing the check the data is not longer viewable

**Exercise 1:** Add and make viewable the Homicides.shp theme. You should now see a point theme depicting homicide locations in Houston Texas. Compare your map with Figure 1.1 ([www.mappingcrime.org](http://www.mappingcrime.org)) to determine if your map is correct.

### ***Exploring Selected ArcMap Tools and Features***

In this section we will briefly explore the different functions of several of these buttons that are highly useful in crime mapping. Students are encouraged to explore other buttons on their own.

### ***Editing the Legend***

Editing the legend is a fancy way of saying changing the color, size, fill pattern, and other factors relating to a themes appearance. In this case we will change the color of the Harris County map. This is a particularly important tool in that it will help you to make the map look nice and aesthetically appealing. Moreover, the size and color of symbols is essential in making a map understandable to your reader.

1. Make the Houston layer **active** by clicking on the file name. This should make the file name highlighted in the table of contents. Importantly, if a layer is not active you will not be able to edit its properties. Click once on the colored box beneath the Houston layer, this will open the Symbol Selector box.
2. After the symbol selector box opens you will have various options in terms of the color of the fill, outline, shape and size of the symbols on the map. In this case



since we are editing a polygon layer (see Workbook Chapter 2 for more discussion) we are only going to be altering the color of the fill and the outline of the polygon layer and not its shape as is routinely done with point data.

3. In order to change the color of the Houston layer click on the “Fill Color” drop down menu and select a color of your choice. When you choose a color the symbol color in the “Preview” box will change on the symbol selector main page. Importantly, in order to get the edit to change in the view you must click “OK” on the Symbol Selector page.

**Exercise 2:** Change the color and symbol type for Homicides to blue triangles. Check your work by looking at Figure 1.3 on the textbook webpage.

### ***Zoom In and Zoom Out***

Being able to zoom the view of the map in and out is essential when making crime maps because it allows you to control the viewable area of the map and helps to see areas more clearly. Importantly ArcMap has two different methods for zooming in and out.

1. On the right of the screen should be a floating menu labeled “Tools” containing seven rows of buttons in two different columns. This floating Tool menu contains some of the most often used and most important tools necessary to efficient map making. In the second row are two buttons, the **Fixed zoom in** button has arrows that point inward (left column) the **Fixed zoom out** button arrows pointing outward (right column). When you place the cursor over either of these buttons a yellow box will appear telling you the name of the button.
2. Click twice on the zoom in button and the entire view will zoom in equally. In order to get the view back to its original view you can either click twice on the

zoom out button or click on the **Full extent** button (button that looks like the Earth). The zoom to full extent button is similar to a home button on a web browser in that it will always take the map view back to the original view. This is a particularly useful button when you are zoomed in on a specific feature and want to quickly look at the original view of the map.

**Important Note:** Importantly, clicking on either of these buttons will zoom in or zoom out the view at an equal rate. Thus if you are wanting to zoom in on a specific area or point you will need to use the other two zoom control buttons. In order to zoom in on a specific area you need to use the **interactive zoom in** button that is on the first row of buttons (plus sign in a circle).

1. Click on the interactive zoom in button to activate the zoom in cursor. As you move the cursor over the map the cursor will change to the button symbol (plus sign in a circle).
2. Place the cursor over the lone homicide incident in the extreme north north-west area of Houston. As you click on the homicide incident the view will change its orientation completely, with the homicide being in the center of the view and the view being zoomed in. Click on the zoom to full extent button to reset the view.

In addition to being able to click and zoom in with this button, the interactive zoom in button allows you to draw a rectangle around an area that you wish to zoom in on. This is particularly useful when you want to quickly zoom in on several crime incidents in an area.

3. With the interactive zoom in button activated click and hold the right button on your mouse and draw a rectangle around the two homicide incidents in the

extreme north-east area of Houston. When you release the right button on your mouse the view will zoom in on the two homicides that you selected.

**Exercise 3:** Zoom in and return to the full extent on several different homicide incidents in the view.

### *Identify Features*

One of the more helpful features that ArcMap offers is one called **identify features**. This button allows the user to look at the tabular data associated with any data on the map.

This is particularly important for crime mapping where a user can use the identify button to read a quick view of the tabular data associated with a crime incident during spatial analysis.

1. Make the **identify** features button (“i” in a circle in the sixth row left column) active by clicking on the button. As you move the cursor over the map it will change to the button symbol (“i” in a circle).
2. Make sure that the Homicides layer is active (click on the name) and place the cursor over the lone homicide incident in the extreme north north-west area of Houston and click on the incident. **Importantly, the identify features button will only identify features of active themes.** You should see that the victim was a 32-year-old male Hispanic who was killed on July 24<sup>th</sup> 1992. Whatever information is entered into the table will be viewable in the identify results window.

**Exercise 4:** Try identifying homicide incidents in different parts of the city paying attention to the different motives, weapons used, and victim races.

### ***Measure Tool***

The last feature that we will discuss in this section is the measure tool. This tool is also especially useful for crime mapping in that it can be used to measure the distance between two incidents or any series of incidents.

1. Make the measure tool (ruler with question mark over it in the bottom left corner of the floating toolbar) active. As you move the cursor over the view it will change to a crosshair with a right angle on it.
2. Before using the measure button it is necessary to change the measurement units for the map to Miles. Go to the View menu and select “Data Frame Properties”
3. Select the “General” tab and in the “Units” box in the middle of the menu scroll UP in the “Display” box until you find “Miles” and select it. When you have selected miles click on the OK button at the bottom of the menu. This will change the measurement properties for the map to miles.
4. Making sure that the Homicides layer is active, place the cursor over the lone homicide incident in the extreme north north-west area of Houston. Click the cursor once and draw a straight line to the northern most of the two homicides in the extreme north-east corner of Houston. In the bottom right corner of the screen the segment length and total length are being shown. When you get to the northern most homicide click once and draw a line to the other homicide in this area and then double click on that homicide. The total length should be around 32.55 miles with the segment length being around 1.8 miles.

**TIP:** Clicking on an area or incident once will mark a segment length while double clicking on an area or incident will stop the measuring completely.

**Exercise 5:** Measure the distance from several homicides near the center of Houston. Importantly, you will probably have to use the interactive zoom in button in order to get the view close enough to measure the distances accurately. Note how close together the homicides occur.

### ***Saving your Work***

The last thing you will do before closing ArcMap is to save your work as an ArcMap Documents. All work done in ArcMap is saved in ArcMap document files (.mxd files) much as Microsoft Word saves word documents as .doc files. Importantly, Maps will save any work done (such as editing and analysis) in data views, tables, and layouts. Saving your work as a Map allows you to open your work later without having to re-open layers or other files.

1. Go to the File menu and scroll down to “Save As..”
2. Pick a destination drive and destination folder and save the project as something you can easily remember.

### **Review Questions:**

1. What are map data called in ArcMap?
2. What is the difference between making a theme active and making a theme viewable?

3. When would it be useful to use the identify tool?

### **Further Readings**

1.

## **Chapter 2: Understanding Data Sources**

### **Goals of Chapter 2:**

1. Learn the different types of data and their importance to Crime Mapping.
2. Learn about different sources for crime related data.
3. Understand the problems associated with crime related data.
4. Working with different types of data.

### ***Introduction***

The second chapter is designed to provide students with a practical discussion of some of the issues discussed in Chapter 8 of the textbook. In Chapter 8 several important issues were discussed concerning crime data and the problems associated with it. In Chapter 2 we will explore, from a practical standpoint, how these data issues impact the making of crime maps. Specifically, Chapter 2 discusses the different types of data used in crime mapping, different sources of crime related data, as well as the problems associated with crime related data. Finally, students will learn how to work with different data types in order to get the most out of their crime maps.

### ***Different Types of Data***

In making crime maps it is very important to incorporate different types of data and not just make simple point maps with only crime incident data. Too often crime maps are little more than just point maps depicting crime locations with the possible inclusion of jurisdictional boundaries and police beats. This misses the point of GIS

analysis completely. The beauty of GIS and crime mapping, is that it allows users to analyze complex spatial relationships that cannot be detected using only point maps. If all you want to do is have a simple point map of where crime incidents are occurring in a jurisdiction it would be far easier, and much cheaper, to avoid GIS and simply place pins in a wall map. Incorporating different types of data, such as contextual data about housing units, business locations or population statistics, into crime maps allows users to understand more than just where incidents are occurring, it allows them to understand *why* crimes are occurring where they are.

### ***Point Data***

Point data is data that represents one place in geographic space, such as a physical address. The Homicides layer from Chapter one is an example of point data. While most point data usually represents one specific crime incident at an address, point data can also be used to show different levels of activity at a location, as when using graduated sizes of points. For example, larger sized dots can represent where a location has had multiple crime incidents occur as compared with smaller dots that indicate only one crime has occurred at that location. Figure 2.1 ([www.mappingcrime.org](http://www.mappingcrime.org)) is an example of how graduated symbols are used to represent different amounts of crime that have occurred.

The main type of data that most users will deal with when making crime maps is point data in the form of crime incident locations. However, it is important to know that there are many other types of point data that are also important to understanding crime patterns. This other point data is commonly called **contextual** data. Examples of some kinds of contextual point data include:



School locations

Bar locations

Liquor store locations

Sexually oriented businesses (SOB's)

Bus and train stop locations

ATM Locations

Bank Locations

Abandoned Buildings

Check Cashing Services

All of these data types are important to crime mapping in that they are places that are often related to where criminal events occur. By analyzing crime data in conjunction with contextual point data users may see that a cluster of robberies occurring every Friday are occurring within a one block area of a check cashing service (see Figure 2.2 [www.mappingcrime.org](http://www.mappingcrime.org)). Thus, including this data into your maps can help with understanding why crimes are occurring where they are occurring and how to prevent further occurrences.

**Exercise 1:** List some other types of contextual point data

### ***Polygon Data***

The other major data type that is used when making any kind of map is polygon data. Polygon data is data that represents areas, such as a county boundary, as opposed to a specific geographic location or addresses. In Chapter 1 the Houston layer that you added is a good example of polygon data, in that it is a county boundary for Harris

County Texas broken into census tracts. As with point data, polygon data can contain both crime data and contextual data. Crime data in polygon form is usually aggregate or count data of crime events that occurred in certain areas such as police beats, police districts, census blocks, census tracts or counties. In general contextual data in polygon form is usually aggregate or count data of social variables such as population totals or total housing units at the census block or census tract level. The most commonly used contextual polygon data in crime mapping is census data but there are numerous other types of contextual polygon data such as:

- Residential and Business zone areas

- Tax zone areas

- Property value data

- Neighborhood/Community boundaries

- Property ownership

All of these data categories are important to crime mapping in that they are often associated with crime and crime levels at the area level. By analyzing crime data in conjunction with contextual polygon data users may see that the same cluster of robberies occurring every Friday near the check cashing service are also occurring in an area with high poverty (see Figure 2.3 [www.mappingcrime.org](http://www.mappingcrime.org)). Thus adding contextual data, both polygon and point, provides a much richer view of criminal patterns by showing a more detailed view of the context in which crime is occurring.

### ***Other Data Types***

Two other forms of data that are often used in crime maps are **line data** and **image data**. While line data used in crime mapping is almost exclusively street data used for geocoding address locations and providing road information, it can also include rivers and stream information. While street data is essential for geocoding, street data can also be important in analyzing why crimes are occurring in certain places. As research on Routine Activities, Defensible Space, Journey to Crime, and other place-oriented theories has shown, criminals are heavily influenced by street layouts and traffic networks. Adding detailed street data to the analysis of the robbery cluster example may show that the check-cashing service is located near a major on-ramp, facilitating escape for criminals (see Figure 2.4 [www.mappingcrime.org](http://www.mappingcrime.org)).

Image data usually consists of satellite images of a place or area that have been georeferenced so that it can coordinate with other spatial data. While not necessarily essential to spatial analysis, adding images to a view can greatly enhance a users understanding of the physical area that is being analyzed. In addition, adding image data is an excellent method for improving the attractiveness and comprehension of map data for a maps final audience. The same robbery cluster analysis example with image data added makes the map more easily understandable and also provides some other visual clues as to why crime is occurring in this location (see Figure 2.5 [www.mappingcrime.org](http://www.mappingcrime.org)).

## *Sources of Data*

While it is vitally important to understand the need for different types of data in conducting spatial analysis, equally important is knowing where to acquire this data as well as the potential problems associated with different data sources. Sources of data for crime mapping can be broken down into four general categories, self-generated, public agencies and universities, free resources, and private companies.

As discussed in Chapter 8 of the textbook, **self-generated** data in the form of crime incident locations and aggregated crime totals are the main source of crime data used in crime maps. Incident level data usually consists of calls for service and/or arrest data, though this varies greatly by department. Some police agencies also create other data such as beat districts, station house locations, and jail locations. While self-generated crime data is the most commonly used data form for crime mapping it is also the most problematic of data sources. As was discussed in-depth in Chapter 8 of the textbook, crime data have reliability problems because of issues related to geocoding of crime data and other issues. Specifically, addresses aren't geocoded correctly because of problems such as officer misspellings, officer address gathering practices and incomplete street files. Despite the potential problems associated with crime incident data, it is the only method for gathering data and thus these problems must simply be worked around.

Another major source for data used in crime mapping are **public agencies and universities**. Public agencies charged with gathering data are often excellent sources of data that can be used in crime mapping. Examples include agencies such as tax assessors, which gather data on property values and housing statistics. Other public agencies gather information on such things as zoning areas, liquor licenses, public

assistance, abandoned buildings, and other data that can be important in analyzing crime patterns. Most important about public agencies is that the data they gather is usually available to police agencies and researchers at no cost. Universities are also an excellent source of data due to their responsibilities as both data repositories and training areas. First, many universities have extensive amounts of geographical data, including satellite images, topographical data, census data, and other data available for use at no cost. Secondly, students are often available to help create and gather geographic data such as beat areas, street files, and other pertinent data for police agencies at no cost to the agency. The main problem associated with both public agencies and universities is their bureaucratic slowness. Public agencies and universities have a reputation for getting things done at speeds that can best be described as glacial. Thus, if data or assistance is needed quickly these sources can often be problematic. However, because of their overall data quality and cost, public agencies and universities are both excellent sources of data.

With the advent of the internet many different types of data can be acquired freely by simple download. The range of **free data** available from these services runs the gamut from satellite images and zoning data to census data and business locations. Sites such as ESRI's Geography Network ([www.geographynetwork.com](http://www.geographynetwork.com)) and GIS Lounge ([www.gislounge.com](http://www.gislounge.com)) are excellent clearinghouses of free GIS data. However, while the range of data available is often great, there are several potentially problematic issues involved with this data. The potential problem for free data sources is the accuracy of the data. The old adage of "you get what you pay for" is often true with free data, in that it is often missing vital information or is in an unusable format. Other problems include

download problems inherent in all Internet transactions as well as the reputability of the individual or group offering the data. While free data is on the whole an excellent source of data, people must be wise consumers of data and be smart about where they get their data.

The final source of data for crime mapping is **private companies**. Companies such as GDT and Urban Data Solutions provide a range of GIS data for a fee. Data available from private companies range from enhanced street files, which are updated frequently, to business data detailing locations of all business in a given area to enhanced census data. Data from private companies is usually excellent, providing accurate, extensive customizable data for end users. The main problem is that to get this accuracy and customized data, end users will have to pay, and depending on the amount of data needed the prices can be steep. However, if cost is not as much a concern as is accuracy and customization, than private companies are an excellent alternative data source to others mentioned above.

### ***Working with Different Data Types***

In this section students will get a hands-on experience in using the different types of data discussed above, including point, polygon, street and image data.

1. Open ArcMap. When the “ArcMap” screen appears select “A new empty map” and click OK.
2. Click on the “Add data” button and maneuver on the C drive to the “CMLab Exercises” folder and than select “Chapter 2”.

3. From the Chapter 2 folder select Houston.shp. You should now be looking at a view with Harris County Texas as the basemap. The first type of data that we will work with is Polygon data, in this case we are going to work with data that is at the census tract level.
4. Right click on the Houston layer and select Properties (bottom selection), opening the Layer Properties dialog box. At the left in the “Show” box are a list of ways in which you can view the polygon data. These options include Single Symbol, Categories, Quantities, Charts, and Multiple Attributes. For the purposes of this lesson click on Quantities.
5. After clicking on Quantities four different sub-fields should appear (Graduated Color, Graduated Symbols, Proportional Symbols, and Dot Density). Click on Graduated Colors.
6. Under the Fields Menu you will see two scroll down menus, one for “Value” and one for “Normalization”. Under the “Value” field scroll down to Homicide. This is a count of the amount of homicides that have occurred in each census tract in Houston between 1985-1994. You’ll notice that the legend has now changed into 5 categories of data with different symbol colors for each different category.
7. Click apply and close the legend editor. Figure 2.6 ([www.mappingcrime.org](http://www.mappingcrime.org)) illustrates how the map has changed dramatically, with high homicide census tracts mostly concentrated near the center of the county.

This map is an example of a **choropleth** map. Choropleth maps are area maps that are shaded according to their level of some phenomenon, in this case homicide frequency (Harries, 1990). In making choropleth maps there are four standardized

options in ArcMap, each with a different method for creating classes of data and each producing a very different map. In addition, ArcMap allows the user to create their own categories with the “Manual” and “Defined Interval” methods. The method used to make the map before you is called **Natural Breaks or Jenks**, and it is the default method for choropleth maps in ArcMap. Natural breaks uses gaps in the distribution of data to establish categories (Harries, 1999). In the map before you ArcMap split the census level homicide data into 5 categories based on what it perceived to be natural breaks in the distribution of homicides at the census tract level.

A second type of classification is called **Equal Interval**. This method divides the distribution based on the area rather than the number of observations (Harries, 1999). Thus if the areas being mapped vary greatly in size the map will be very different from natural breaks and other methods.

**Exercise:** Right click on the Houston layer and select “Properties”, opening the Layer Properties editor. Click on the Classify button opening the “Classification Menu”. Under the Classification Method menu scroll down to Equal Interval. Click OK and then click Apply and close the Layer Properties editor. Figure 2.7 ([www.mappingcrime.org](http://www.mappingcrime.org)) illustrates how the map looks very different from the previous map created with Natural Breaks. The classes of data are very different and the high areas of homicide are much larger, more frequent, and more distributed around Houston.

A second type of classification is **Equal Interval**. This method divides the distribution into groups with an equal range between the groups, such as 0-14.2, 14.2-28.4, 28.4-42.6, 42.6-56.8, etc..



**Exercise:** Right click on the Houston layer and select “Properties”, opening the Layer Properties editor. Click on the Classify button opening the “Classification Menu”. Under the Classification Method menu scroll down to Equal Interval. Click OK and then click Apply and close the Layer Properties editor. Figure 2.8 ([www.mappingcrime.org](http://www.mappingcrime.org)) illustrates again how the map has changed dramatically from the previous map. The majority of the map is in the lowest few categories with the highest areas of homicide being just a few census tracts that are confined to the center of Houston.

A third type of classification is called **Quantile** or equal count. In this classification method, approximately equal numbers of incidents are put in each category (Harries, 1999). Thus, in the case of the Houston data, approximately equal numbers of homicides are placed into each category. Importantly, this method is highly influenced by the number of categories selected. The default in ArcMap is 5 categories, however this can be altered and will have significant impact on the map.

**Exercise:** Open the Layer Properties editor again and go back to the classify button (see above). Scroll down the Method category to Quantile. Click OK and then click Apply and close the legend editor. Figure 2.9 ([www.mappingcrime.org](http://www.mappingcrime.org)) illustrates how the map is very different from the equal interval map and more similar to the equal area map, with a large number of census tracts being in the top groups.

The final method of classification is called **Standard Deviation**. This final method is a statistical measure which classifies data based on the spread of data around the average score (Harries, 1999). First the average observation is determined for the distribution than groups are classified by how much above or below (1 standard

deviation) from the mean they are. This method can be beneficial in displaying extreme scores both above and below the average of the distribution.

**Exercise:** Open the Layer Properties editor again and go back to the classify button (see above). Scroll down the Type category to Standard Deviation. Click OK and then click Apply and close the legend editor. Figure 2.10 ([www.mappingcrime.org](http://www.mappingcrime.org)) displays the results of a standard deviation classification. The standard deviation map is probably the most different of all of the maps that have been made, but potentially the most beneficial as well. In this map the darkest reds show the areas that are highest from the average number of homicides in Houston, while the blue areas are census tracts where homicide is below average in occurrence.

While there is no one correct or best way to classify aggregate data in a choropleth map, there are some suggestions that can help users in making good maps.

### ***Suggestions for Classifying Data***

1. Do not simply trust the default method of the computer (Monmonier, 1996).
2. Use no more than six and not less than four categories in choropleth maps (Harries, 1999).
3. When exploring the data use more classes, but as you move towards presentation use less categories (MacEachren, 1994).

### **Chapter 2 Project: Contextual Data Analysis**

The project for Chapter 2 is to combine different types of data to conduct a theory driven analysis of homicide in Houston. Specifically, using theories discussed in the first part of the text and the sample data provided, students should attempt to conduct an analysis of homicide data in Houston.

### **Point Data**

Homicides.shp: All Homicides in Houston in 1992

Bars.shp: All bars in Houston Texas.

Malls.shp: All malls in Houston Texas.

Liquor.shp: All liquor stores in Houston Texas.

Schools.shp: All schools in Houston Texas.

### **Line Data**

Houstreet.shp: Street file for Houston Texas.

### **Polygon Data**

Houston.shp: Polygon file containing census tract data.

Importantly the accompanying “Houston” codebook will explain to the student the different census data that is contained within the Houston.shp file.

Examples of theory driven analysis that can be done include testing social disorganization using social disorganization variables NDISADV or NINSTABL, which look at the amount of disadvantage and instability in a neighborhood.

### **Review Questions**

1. Why is it important to include other types of data besides crime data when making crime maps?
2. What is contextual data and what are some important types of contextual data for crime mapping?

3. Name the four main types of data.
4. What are the 5 main ways to categorize data in Arcview?

## Chapter 3: Entering Data

### Goals of Chapter 3:

1. Learn the different types of data entry.
2. Learn about problems associated with data used for geocoding.
3. Learn how to geocode data as well as problems associated with geocoding.
4. Learn about GPS and how to enter GPS data.
5. Geocode sample data.

### *Introduction*

One, if not the most important, part of mapping crime data is actually entering the data into the GIS. However, despite its importance to effective crime mapping, data entry and its associated problems receive little discussion in most crime mapping texts. In Chapter 3 students will learn how to both geocode crime data and enter GPS coordinates of crime data as well as the problems associated with each method.

### *Geocoding Data*

**Geocoding** is the process of taking specific street addresses in a table format and matching them to a reference file (street file) containing a range of addresses for a given area. When an address in your data table matches a range of addresses in the reference file they are then given physical coordinates for that address (ESRI, 1998). Thus, in order to geocode crime data you need both a table of crime incidents with associated addresses and a street file with address ranges for the area that contains the crime

incidents. The benefit of geocoding is that it is a fairly quick and effective method for turning address level crime data into points on a map that can then be analyzed.

While geocoding is by far the most popular method for entering crime data into a GIS it is far from a perfect method. When geocoding, the accuracy of address matching is called the **hit rate**. A hit rate is the percentage of crime incidents that match to corresponding street addresses in the reference file. Importantly, the goal of geocoding is to have a 100% hit rate as unmatched addresses are not mapped, leading to inaccurate representations of where crime occurs. Missing crime data caused by low hit rates can lead to serious problems when mapping is being used for tactical or research purposes. However, due to the nature of where crime events occur and standard data gathering practices in police departments, geocoding can lead to low hit rates and inaccurate representations of where crime actually occurs. This is not to say that geocoding is a poor method of data entry for crime data, just that there are specific problems that need to be addressed in the data gathering stage that can help make geocoding more accurate.

One of the most common problems associated with accurate geocoding of crime incidents revolves around where crime incidents occur. Specifically, many crimes occur at places where there are **no associated street addresses**. Examples include crimes that occur in fields, on highways, at intersections, on railways, or on or near water such as a river, lake or ocean (Gebhardt, 1999). While there is no one perfect solution for incidents that occur in these problem areas, suggestions include geocoding to the nearest actual street address, geocoding to the block instead of an address, and using GPS to get accurate coordinates (Gebhardt, 1999).

Another more common problem with data gathering involves **poor data entry** by officers in the field and data entry people. This problem takes on two different forms, officers using common place names instead of street addresses (Wal-mart instead of 1350 Fontaine Rd.) and officers and data entry people misspelling addresses. Because geocoding works by matching actual addresses to a range of street addresses in a reference file, misspellings and common place names will not match against the reference file, thus lowering the hit rate and leaving important crime data off of your maps. Several different solutions exist to this problem. One solution is to use **validated entry** for all addresses (Gebhardt, 1999). In validated entry, all addresses are checked against the reference file during the data entry process, if the address doesn't match an existing address in the reference file the data entry person is alerted to a bad address. While this method makes data entry fairly easy, it requires a lot of advance work and programming to be effective (Gebhardt, 1999). Another solution is to use **forced entry** choices for all addresses (Gebhardt, 1999). Forced entry works by only allowing valid addresses to be entered through the use of drop down menus of all valid addresses. While this method ensures 100% accuracy, it severely limits users and requires a lot of background programming and thus is not always recommended for large databases (Gebhardt, 1999). A final method is the use of **substitution tables** (Gebhardt, 1999). Substitution tables work by allowing programmers to create "other" address names for a valid address in the reference file, which the GIS then searches when looking for matches to addresses. Thus, common place names and common misspellings of addresses can be listed in the substitution tables, increasing the hit rate. This is the easiest of the three methods to implement as it is a standard part of all major GIS packages.

A final problem associated with low hit rates is **poor reference files** used for geocoding. This problem is most common in both rural locations, where street files are often poorly created and maintained, or urban areas that are constantly growing and adding roads. In addition some software programs come with poorly equipped geocoding software, ill prepared to geocode many addresses that are valid addresses (Levine, 2002). Poor reference files make it difficult to match crime locations because they don't contain the full range of addresses for an area. The best solution for reference file problems is to update your reference file several times a year, either through manual updating or through the acquiring of reference files from government agencies or private companies.

### ***Practical Geocoding***

1. Open ArcMap and select "A new empty map" and click OK.
2. Click the "Add data" button and maneuver on the C drive to the "CMLab Exercises" folder and then select "Chapter 3".
3. From the Chapter 3 folder select the following layers:
  - a. Houston.shp.
  - b. Houstreets.shp
4. Make the street theme (Houstreets) active and click on the Tools menu. Under the Tools menu scroll down to Geocoding and when the options appear to the right of Geocoding select "Geocoding Service Manager".
5. In the Geocoding Services Manager box click on the "Add" button on the right and double click on "houstreets.mxs". After returning to the Geocoding Services Manager box click on Close.



### *Selecting File for Geocoding*

1. Under the Tools menu again select “Geocoding”, however this time when the options appear to the right of Geocoding select “Geocode Addresses”.
2. When the “Choose a geocoding service to use..” box appears “Houstreets.mxs” should be highlighted. Click OK.
3. When the “Geocode Addresses: houstreets.mxs” box appears you will need to select the file that we are going to geocode. Next to the “address table” box click on the open file button. This is where you will add the .dbf table to be geocoded.
  - a. Go to Chapter 3 in the CM Lab folder and select Table A.
4. In the Street or intersection box scroll down until you find the variable named INCADD and select it.
  - a. This is the variable that contains the addresses for the incidents.

### *Match Selection and Setting Geocoding Preferences*

In matching addresses ArcMap performs what is called a **batch match**, as opposed to Arcview 3.3 where the user has the initial option of conducting either a batch match or an interactive match. In a batch match ArcMap will attempt to match as many addresses as possible based on geocoding preferences that the user sets before matching. In contrast, in an interactive match the user reviews each address individually and makes all decisions about whether to match an address or not. Whether you choose to conduct a batch match or an interactive match is based on personal choice, with the main factor being the tradeoff of speed vs. accuracy. While batch matching is certainly faster at matching addresses, especially large datasets, it is potentially less accurate due to the setting of preferences. Batch matching is generally recommended for large data sets and

datasets where address entries are known to be accurate (see discussion above), whereas interactive is recommended for datasets that are small or where special care is needed. In this case batch match will be reviewed as it is more commonly used for police data.

As ArcMap conducts a batch match it is essential to set the **geocoding preferences** under the geocoding options editor. This is important because how you set your preferences will help determine your hit rate. Importantly, hit rate is more than just getting a match for an address it is about getting the *correct* match for an address. Thus, in this discussion of setting the geocoding preferences, emphasis will be on setting the preferences conservatively so as to make sure that addresses match correctly.

1. Click on the Geocoding Options box on the lower left. For our purposes, there are three preferences that we will be concerned with: Spelling Sensitivity, Minimum match score, and Minimum score to be considered a candidate. The other preference options are more advanced and beyond the scope of this book.
2. **Spelling Sensitivity:** This setting allows you to adjust how sensitive you want the spelling accuracy to be when looking for address matches. Importantly, 100% requires a perfect spelling match (almost unheard of) to match an address. Suggestions are given to leave the sensitivity at the default 80, in order to ensure correct spelling.
3. **Minimum match score:** This setting determines the minimum score an address in the reference table has to have to be considered a match. For each address ArcMap creates a score from 1-100 showing how close to a match the address is. How you set this score will determine which addresses ArcMap will automatically match. The lower you make the score the fewer addresses will

- match automatically. Importantly, unmatched addresses can be reviewed interactively after the batch match has concluded allowing you to interactively re-match addresses that fall below the criteria you set. Again the suggestion is to leave this score at the default 60.
4. **Minimum score to be considered a candidate:** This score is used to determine the minimum score needed for an address to be considered a candidate for matching. Scores can range from 0-100, with recommendation being to use the default 30 thus allowing a significant number of candidates to be considered.
  5. **Output Fields:** Before starting the geocoding process the last thing that we need to do is select the output fields that we wish ArcMap to create for our geocoded addresses. Here we want ArcMap to create two different fields with the geocoding results, one is a *standardized address* and the other is *X and Y coordinates* for the address. Thus, place a check mark in the boxes next to both of these options.
  6. After reviewing the preferences and output fields click OK.
  7. After returning to the Geocode Address box click on the OK button. This will start the matching process and may take a few minutes depending on the computer you are using to run the exercise.

### ***Interactive Address Re-matching***

1. After the batch matching is finished the **Review/Re-match Address** box will appear. Click on the **Interactive Re-match** button and the **Interactive Review** box will appear. This is where the user will go back and attempt to interactively re-match all of those address that failed to match in the batch match. Importantly,

you can also perform a batch re-match, but unless you change your geocoding preferences your results will not change.

2. When the Interactive Review editor appears you will notice numerous buttons used in interactive re-matching. Below is a description of the Interactive Review editor:

**Address Text Box:** This window located near the top of the editor displays the current address to be matched. Addresses that are being matched are highlighted.

**Candidate Table:** This table displays all of the candidates for the current address from the reference table. The table lists the score for each candidate as well as the address information for each candidate.

**Match Button:** Allows you to match an address to a highlighted candidate from the candidate table. Click this button when you have selected a candidate from the candidate table.

**Unmatch Button:** Allows you to “unmatch” an address that have already matched. Click this only if you feel that you, or the computer, have matched an address incorrectly.

**Next Arrow:** Arrow next to the Record box that allows you to skip an address currently under review. Click this when there are no candidate addresses or no addresses you feel are correct.

**Previous Arrow:** Arrow next to the Record box that allows you to go back to a previous address. Similar to the back button on a web browser.

**Search Button:** Allows you to enter changes you may have made to the address to be matched. When you click this button after making changes to the street address (spelling or street number) the program will attempt to match the newly changed address. Once the process is done new candidates will be shown in the candidate box.

**Geocoding Options Button:** Allows you to view the preferences that you established before the batch match.

**Close Button:** Signifies that you have finished re-matching the addresses.

CLICK THIS BUTTON ONLY WHEN YOU ARE FINISHED WITH ALL OF THE ADDRESSES.

3. When you have re-matched all of the addresses click on the close button.
4. When the Review/Re-match box appears again click on Done. This will create a new layer of the addresses that you just matched. The new layer will be named something like Geocoding results: Geocoding\_result1
5. To re-name the Layer and save it in a file that you will remember make Geocoding results: Geocoding\_result1 active and right click on the Layer name and scroll down to “Save as layer File..”
  - a. Save the new file as something that you remember and in a folder where you will remember.

### ***TIPS for Interactive Re-matching***

1. Try editing the spelling of the street name or number. Oftentimes addresses are spelled incorrectly.

2. Try editing the direction of the street, either adding or removing directions to a street. Sometimes streets have directions that are not added or street files have left the direction off of a street.
3. Pay attention to the score the candidates receive, if the score is high enough or the address appears to match the candidates components despite a low score, you may want to match the address.
4. Whether you match an address interactively is completely up to you, as you CAN match almost any address if you edit it enough. Importantly, there are many ethical issues with how much editing you do in order to achieve a high hit rate.

### ***Adding GPS coordinates***

One of the much-heralded suggestions for improving crime mapping accuracy is to increase the use of GPS measurements instead of relying on geocoding. GPS stands for Global Positioning System, a series of 24 satellites that orbit the Earth and provide exact coordinates for any position. Obviously GPS coordinates are an extremely attractive option because they provide exact position coordinates that do not need to be altered in order to put a crime location into a GIS. All an officer has to do to record a location is look at a GPS receiver and write the coordinates on an incident report. The coordinates can be in any coordinate system, such as decimal degrees or UTM depending on the coordinate system of your basemap. Once the reading has been made the coordinates only need to be put into a table as a .dbf or .txt file so that they can be added to ArcMap. Importantly, GPS receivers have become increasingly cheap, small, and sophisticated, allowing better accuracy for less cost than five years ago. However,

despite their appealing benefits there are several problems with GPS receivers for use in gathering position data. First, for a position to be recorded an officer has to actually be at the physical location to take a reading, as opposed to street addresses which can be recorded over the phone without the need of an officer. Second, despite improvements in GPS technology, reception can be impaired significantly by things such as weather, high buildings, tree coverage, and power supply. While there are technological means to deal with these problems it increases the cost of GPS significantly. Despite the potential problems associated with GPS, it is definitely beneficial in situations where crimes occur on highways, fields, intersections, or waterways and will be increasingly used in the future.

### ***Adding Event Themes***

1. In the map that you currently have open go to the Insert menu and select “Data Frame”. This will allow you to add a new set of layers.
2. Click on the Add Data button and from the Chapter 3 folder add the Post7.shp theme. This data represents an 11 county area in central Kentucky.
3. Once you have added this data click on the **Tools** menu and scroll down to **Add XY Data...**
4. When the Add XY Data menu opens click on the open file button to the right of the “Choose a table from the map or browse for another table” box and proceed to Chapter 3 in the CM Lab folder and select **tableb**.
5. ArcMap will automatically place the variable “X\_Coord” in the X Field box and “Y\_Coord” in the Y field box. However, if for some reason these variables do not show up in these field boxes, then scroll down and manually select them.

6. Click OK and the data should be added to the map of the 11 county area.  
Importantly, this new layer is not really a layer yet. In order to make this a true layer we must save it as a layer.
6. Make “Table B events” active and right click on the Layer name and scroll down to “Save as layer File..”
  - a. Save the new file as something that you remember and in a folder where you will remember.

### **Chapter 3 Project: Geocoding Accuracy**

1. Add the following themes:
  - a. Lexington.shp
  - b. Lexstreets.shp
2. Add the table Geocode.dbf and geocode all of the addresses. Remember to prepare the street file for geocoding (US Streets) and decide whether to use a batch match or interactive match.

**Project Answer:** Check the accuracy of your geocoding against Figure 3.1 at [www.mappingcrime.org](http://www.mappingcrime.org)

### **Review Questions**

1. What are the two main ways for entering crime data into Arcview?
2. List the associated problems with data used for geocoding.
3. What is the Hit Rate
4. Name and explain two methods for increasing hit rates.



5. Explain the difference between Batch matching and interactive matching of addresses.
6. Explain the benefits and problems associated with using GPS for data acquisition.

### **Further Readings**

1. Gebhardt, C. (1999). Geocoding: Improving Your Hit Rate. Presentation given at the Crime Mapping Research Centers Annual Conference, Orlando, FL. Available on-line at <http://iaca.net/resources/FAQs/hitrate.html>

## **Chapter 4: Querying and Merging Data**

### **Goals of Chapter 4:**

1. Learn the importance of querying and merging data for crime mapping.
2. Learn the different ways to query and select layers.
3. Learn how to merge data.
4. Query and Merge sample data.

### ***Introduction***

By now you have learned much about the basics of using ArcMap for crime mapping, from how to use some of its basic tools to data quality issues to entering data into ArcMap. In chapter 4 we are going to focus on how to manipulate data for analytic and organizational needs. Specifically, you will learn how to query and merge data in order to create layers that contain only data that you need.

### ***Why Query and Merge Data?***

Most police reports contain numerous fields of information relating to the crime incident, including: type of crime, date and time of occurrence, victim and offender information, weapon use and type and incident location. While all of this information is extremely important to conducting analysis of crime series and crime patterns, unless it is manipulated correctly all of this information is potentially lost in crime mapping applications. As we learned in chapter 3, when crime information from police reports is

geocoded and displayed on a map it is based solely on the location of the incident, ignoring all of the other information contained in the report. When a months worth of crime is geocoded and mapped it only provides users with a point map view of all crimes in that jurisdiction for that month. Thus, if using this same map a user wanted to know the location of robberies only, or robberies in relation to drug crimes the map would be virtually worthless. Importantly, maps based on information such as crime type, offender motive, and weapon use are vitally important to both criminal justice practitioners and researchers. Maps displaying spatial patterns of specific crimes are necessary for tactical planning and problem solving by practitioners and for research projects by researchers. Fortunately, ArcMap makes the creation of these customized maps easy through the use of the **query** and **merge** functions.

A **query** is a simple operation in which the user is allowed to define what features in a layer they want to display on a map. An example would be creating a query in which robberies are selected from a month of crimes in order to allow the user to see the location of all robbery locations. This function, which can be conducted on any information contained in a data table, is one of the more common functions used by crime analysts in their daily functions. Common queries relating to crime and crime reports include:

- Crime Type
- Offender Characteristics (Race, Gender, Age range, etc..)
- Offender Motive
- Weapon Type
- Date of Occurrence

- Time of Occurrence

The other function that is commonly used in organizing and customizing data for crime maps is the merge function. **Merge** is a function that allows users to combine several different layers of the same type together to create a new layer. This type of operation is commonly used in crime maps when combining multiple crime themes from different time periods into one single theme for a longer period of time. An example would be merging robbery layers from each month in a given year to create a layer containing all robbery incidents for an entire year. It is important to note that only layers of the same type can be merged. Thus you can only merge two point layers or two polygon layers, not a point layer and a polygon layer. In the context of crime maps', merging of layers is most commonly done to combine layers based on date of occurrence, type of crime, motive, or weapon used.

### *Querying Layers*

1. Open ArcMap and select "A new empty map" and click OK.
2. Click the "Add data" button and maneuver on the C drive to the "CMLab Exercises" folder and then select "Chapter 4".
3. From the Chapter 4 folder select the following layers:
  - a. Houston.shp.
  - b. Homicides.shp
4. Make both of these layers viewable and make Homicides active as well.

5. Under the Selection menu, scroll down to “Select by Attributes..” This will open the Select by Attributes Box, which will allow the user to perform various different queries of the data..

Within the select by attributes box are several important areas. First is the **Fields** box, which lists all the different fields of information that you can choose to query. Second is the **Unique values** box, which contains all of the values of the selected fields.

Specifically, when a field is selected the values of that field will show up in the Unique values box. Third is, for lack of a better term, the operators area. This is the area between the Fields box and the Unique values box that contains the different operations that can be performed in the query. For example when querying months you can choose all months greater than (>) 4 (April), all months less than (<) 4 or only incidents in the 4<sup>th</sup> month (=). The final area is the query text box, which shows the users the query that they are currently creating.

### *Creating a Query*

1. After opening the Select by attributes box the first thing you need to do is select a **field** to query. In this example we will select all female homicide victims. Importantly, all queries must contain three things, a field, an operator, and a value.
2. Scroll down the fields box until you find the field Vgender and double click on it and it should appear in the “query text box” at the bottom. Importantly you **must double click** on the field name or it will not appear in the “query text box” and will not be available for query.

3. Next we will pick an **operator function**, in this case equals. Click once on the “=” sign. When you click on the equals sign it should appear in the query text box after Vgender.
4. Finally we will select the **value** for the query. In this case double click on the “F” value. As with the fields, values must be double clicked in order for it to be included in the query. The equation in the query text box should now read as follows:

**“Vgender” = ‘F’**

5. When your query is complete click on the “Apply” button. This button will run your query for you, selecting all values that match your query and coloring them light blue on the map.
6. Figure 4.1 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows you what the map should look like if queried correctly. After you close the query text box your map should have 71 incidents colored blue or “selected”. You can check to see that you have this correct by opening the table (right click on the homicides layer and select “Open Attribute table”). On the bottom under the table it will say 71 of 536 selected, denoting that 71 of the 536 records were selected by your query. In lay terms, your query for female victims produced 70 incidents that had female victims.

### ***Converting, Clearing, or Querying the Query***

After you have conducted the query you can either convert the query to a layer, which will allow you to save it and manipulate it at a later date, clear the query allowing you to conduct a new query, or make a query of the selected features.

**Convert the query to a layer:**

1. Right click on the “homicides” layer and scroll down to Selection.
2. When the menu pops out to the right of selection scroll down and select “Create Layer From Selected Features”. You should now have a new layer entitled “homicides selection” added to your map.

**Clear the query:** Under the “Selection” menu scroll down and select “Clear Selected Features”. Alternatively you can right click on the Homicides layer and scroll down to the “Selection” menu and select “Clear Selected Features”.

**Query of the selected features** (query results).

1. Open the Select by Attributes box and select a new field. Importantly, make sure that the layer selected for the query is “Homicides” (If you have cleared the previous query then you will need to redo it for the rest of this exercise). If it is not selected then go to the “Layer” menu at the top of the Select by Attributes box and scroll down to that layer. In this example we are going to query white victims from the female homicide victims query to determine how many white female homicide victims there were.
2. Select a Field, in this case Vrace, for victim race.
3. Select an operator, in this case “=”.
4. Select a value, in this case “W”. Your equation should now look like the following:

**“Vrace” = ‘W’**

5. Before clicking on “Apply” go to the “Method” dropdown menu at the top and choose “Select from current selection”. Once you have done this click on apply. This will make your selection from the selected features (female homicide victims). Close the Select by Attributes box after your query is complete.
6. Figure 4.2 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows what a correctly queried map should look like. The number of queries should have decreased significantly, so that only 22 incidents are selected. Again to check the queries open the associated table.

### ***Merging Themes***

1. With ArcMap already open make the current layers unviewable (de-select the check boxes to the left of the layer names), add the following themes to the view.
  - a. 1993.shp
  - b. 1994.shp
2. Under the Tool menu scroll down and select “**Geoprocessing Wizard**”. This will open the Geoprocessing Wizard, which is designed to walk you through the process of merging layers. The Geoprocessing wizard has five different geoprocessing functions that allow you to create or alter layers. The different functions include, Dissolve, **Merge**, Clip, Intersect, and Union. For our purposes we are going to focus on the Merge function, although other functions may prove beneficial in other crime mapping applications.
3. Select the “**Merge layers together**” option on the main screen and click “Next”.



4. In the first box of the second screen where it states “Select at least two layers to merge”, select both the 1993 and 1994 point layers.
5. In the “Specify the output shapefile or feature class” text box, click on the open file icon and choose a location to save the new theme and an appropriate name, such as 9394.shp
6. To finish the merge click on the Finish button. You should now have a new layer named “9394” added to the view. Make it both viewable and active and then open its attribute table. Figure 4.3 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows you what a correctly merged map should look like. Scroll through the table and notice that the table now includes both the data on all 1993 and 1994 homicides in Houston.

**TIP:** While any two point or polygon themes can technically be merged, it is important to merge only those layers that share similar fields in order for all of the information to be included in the attribute table.

#### **Chapter 4 Project: Merge and Query of Data**

1. Create a new view and add the following themes:
  - a. Houston.shp
  - b. 1985.shp
  - c. 1987.shp
  - d. 1990.shp
2. Merge the three different point themes into one new theme and name it 8590.shp.
3. Query the new theme in order to find all **robbery homicides** committed during the **summer months** (May-September). In order to find out the Field and Value

information to create this query see the “Homicides” codebook you downloaded from the textbook webpage.

**Project Answer:** To check the accuracy of your query see Figure 4.4 on-line at [www.mappingcrime.org](http://www.mappingcrime.org)

### **Review Questions**

1. What is a query and how is it important to crime mapping?
2. What is a merge and how is it important to crime mapping?
3. What are the 3 necessary parts of a query?
4. How can you check how many incidents were selected by your query?
5. After you make a query what are your main options as to what to do?
6. What happens if you try to merge two different types of themes, such as a point and a polygon theme?

## Chapter 5: Understanding Layouts and Map Design

### Goals of Chapter 5:

1. Understand the importance of scale.
2. Understand how to use Layouts
3. Understand basic map composition
4. Understand basic Map Design for Crime Maps
5. Design, Print, and Export a map.

### *Introduction*

While all of the previous chapters have been important in providing you with a basic understanding and sense of familiarity with ArcMap and making crime maps, none may be as important to your actual creation of intelligible maps than Chapter 5. In Chapter 5 we are going to explore the art/science of map design and making informative crime maps for audiences of different kinds. This chapter is so important for students, because without a solid understanding of map design your maps will be at least ineffective and at worst unethical and deceitful. Ultimately, how well a map speaks to the audience depends on how well a map is designed and produced.

### *Scale and Map Design*

In map design one of the more important factors to determine is the scale of the map. **Scale** is the degree of reduction in the map as compared to reality (Monmonier,

1993). Maps are designed to be miniature representations of reality and the scale on a map is what tells us how much smaller than reality the map is. The scale of a map is usually expressed as a ratio of the distance measure on the map in relation to the distance measure on the ground (Monmonier, 1993). Thus, a scale with a ratio of 1:15,000 inches would indicate that one inch on the map is equal to 15,000 inches on the ground.

In general there are two kinds of maps, **large scale** maps which show small areas and **small scale** maps which show large areas. An example of a large scale map would be a map showing crime in a 6-block area of a city, whereas a small scale map would be a regional map of the United States showing crime victimization by state. The main importance of scale to map design is the level of detail that is available in a map and its impact on the interpretation and meaning of a map (Harries, 1999). Specifically, large scale maps are capable of showing much more detail and are thus inherently easier to interpret, whereas small scale maps have less detail and tend to be more abstract and difficult to interpret. Figure 5.1 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows the pattern of homicide incidents in relation to bar locations in Houston Texas from two different scales and illustrates the impact of scale on the level of understanding and detail in a map. The smaller scale map makes it appear that all homicides occur within feet of bars, whereas the large scale map makes it clear that while there is a definite relationship between bars and homicide locations it is not as strong as it appears in the small scale map.

From a practical standpoint when making a map, the level of scale used will depend largely on the purpose of the map and the subsequent data available for the map. For instance if you need to make a map showing the pattern of robberies for a six beat section of a jurisdiction, your scale will be large by definition. This large scale map will

thus be able to show a great deal of detail of the six beat area, including road names, bar locations, and ATM locations. In contrast, if you are showing the level of robbery victimization by county for an entire state your scale will be small and your level of detail will be more abstract. If you were to try to show roads and road names in addition to bar and ATM locations, your map would be overwhelming and difficult to understand. Importantly, changing the scale of your map in ArcMap is very easy, as the “zoom in” and “zoom out” buttons control it. As you zoom in on a map location, the scale, which is displayed in the right corner of the view, gets larger.

### ***Map Elements***

Before we discuss how to design and print a map using a “Layout” we need to discuss the four necessary elements of a map. While arguments can be made that certain maps do not need to contain all four elements of a map, in general these four simple elements improve the effectiveness and comprehension of a map. First, every map must have a **title**. Titles can vary from basic to complex, but serve the purpose of informing the reader of the purpose and content of the map. A second necessary element is the **legend**, which acts as a table of contents for the reader listing the different data and graphical features of the map. In general legends contain the symbols and colors that are used to define geographic features in the map as well as a text definition of what the symbol or color represents. A third necessary element is a **scale**. As discussed above, a scale is important because it tells the reader the relationship between the map measurement units and the measurement units in the real world. Importantly, scales also allow readers to measure distances on a map, which can be a primary function of some

crime maps dealing with journeys to crime and travel distances. A final necessary map element is a **compass**, which informs the reader of the orientation of the map (North and South). Other elements of a map that are not necessarily essential but can be beneficial are **credits**, which inform the reader where the data came from, and **neatlines**, which surround the body and other elements of the map to give it a cleaner appearance.

### ***Map Layout***

1. Open ArcMap and select “A new empty map” and click OK.
2. Click the “Add data” button and maneuver on the C drive to the “CMLab Exercises” folder and then select “Chapter 5”.
3. From the Chapter 5 folder select the following themes:
  - a. Houston.shp.
  - b. Homicides.shp
4. Make both of these viewable and edit their appearance. Make the Houston layer clear and make the Homicides into red stars at 12 pt size. In ArcMap, map design takes place in a Layout.
5. Under the View menu select **Layout View**. This will bring up the “Template Manager” box. Alternatively, you can select the page button in the bottom left corner of the map (next to the small globe).

## ***Editing Map Elements***

### ***Adding a Legend***

When looking at the layout you should instantly notice none of the map elements that should be included with a good map are on the layout. In ArcMap the user is charged with adding all of the different map elements that they want. Importantly, this allows the user the complete control of customizing the look of the map.

### ***Legend Items***

1. Under the Insert menu select “Legend”. This will bring up the Legend Wizard, which will guide you through designing an effective and attractive looking legend.
2. First you will choose the map layers that you want to include in your legend. In the left column are ALL of the layers that are in the data map that you are working on, while the right column shows the Legend items you have chosen to include. As a rule every layer that is visible on the finished map should be included in the legend. In this case there are only two layers on the map (homicides and Houston) both of which have been included in the legend.
3. If for some reason you want to remove a layer from the legend simply click on it and then click on the arrow pointing towards the Map layers column. Conversely, if there are map layers you want to add to the legend click on them and then click the arrow pointing towards the Legend column.

### **Legend Tips:**

- A. **Legend Order:** The order of the legend can be altered simply clicking on the Legend item and clicking either the up or down arrow to the right of the Legend Item column until the item is in the desired position. In crime maps point layers (crime incidents) usually are at the top of the legend followed by line layers (roads) and polygon layers (jurisdiction boundaries).
- B. **Items included in the Legend:** Only items that are “viewable” in the view mode are included in the legend.
- C. **Layer names:** If the name of the layer is not to your liking than you need to return to the Data view and change it. Once in the data view, click on the Layer name until it is highlighted and then simply re-name it to whatever you want. In all cases Layer names should be capitalized.

### ***Legend Title***

1. Once you have all of the Legend items you want and in the order you want click on the Next button at the bottom of the page. This will take you to the Legend Title page of the legend wizard.
2. In general, the legend is best left named “Legend”, but if for some reason you want to re-name it simply highlight the word legend in the “Legend Title” box and type in whatever you want as a title.
3. If you want to change the font and the justification (alignment) of the legend that can also be easily done in this section of the Legend wizard.



4. Once you have finished with the Legend Title, click on the Next button.

### ***Legend Frame***

The legend frame gives the user the option creating a frame around the legend that will set it apart from the rest of the map. Importantly, ArcMap give the user several customization options.

Border: This is the outline around the legend. To add a border click on the dropdown menu and select a border size that you find attractive.

For this exercise select the .05 Point option

Background: This is the background color/shading of the legend. To edit the background simply click on the dropdown menu and choose a color that you find attractive.

For this exercise select Grey 10%

Drop Shadow: This is the shading behind the legend that gives the legend a 3-D appearance. To edit the Drop Shadow simply click on the dropdown menu and choose a color that you find attractive.

For this exercise select Grey 20%

Once you have made these changes click on the Next button.

### ***Item Size and Spacing***

The last two sections of the Legend Wizard allow the user to further customize the look of the legend by editing the size and shape of symbol patches as well as the spacing of the items in the legend. While this is a great feature for the advanced user, for

our purposes it is beyond the scope of our discussion. Thus, for these last two sections of the wizard, simply click on Next and then Finish.

### ***Altering the size and location of a Map Element***

Once your legend has been created you will probably want to move it around the map to a place where it does not obstruct the view of the items on the map. Importantly, moving and altering the size of all of the map elements is easily done.

1. With the pointer tool click **once** on any of the basic map elements (compass, scale, title, legend). Four squares should appear at the corners of the map element. These four squares represent the outline of the map element and can be re-sized to your liking.
2. Place the pointer on any one of the squares. This should change the icon to a two-headed arrow.
3. Click and hold the square and move the square inward or outward to change the size to your liking.
4. Once you have the map element the size you want you can change its location by clicking once on the map element and placing the pointer over the main body of the map element.
5. When the pointer tool changes to a four-headed arrow click and hold to drag the map element to wherever you want in the map layout.
6. If you want you can also alter the size and location of the view frame (main map), using the same procedures described above.

**TIP:** When trying to determine the correct size of a map element remember that map elements should not be so large as to distract from the view frame, nor should they be so small as to be difficult to find or understand.

### ***Editing the Scale***

While in the data mode we can alter the scale of the map, which is the only other map feature whose properties and appearance in the layout are controlled completely in the data mode.

1. Select the interactive “zoom in” button and click in the center of the view.
2. After the map has zoomed in minimize the view and open up the layout for view1.

Notice how the legend and scale have changed in accordance with the alterations made in the view mode. Whatever scale you are in the data mode will be the scale for the layout.

### ***Adding a Compass and Scale bar***

As with the legend, all of the other map elements must be added to the layout by going to the Insert menu and selecting a map element. Importantly, in creating a north arrow and scale bar the user simply chooses a style that they like from the options available in the North Arrow or Scale Bar selector menus and clicks OK.

**TIP:** There is no conventional scale bar or compass style, but try to choose a style that complements the map rather than distracts. Remember that scales and compasses are there to inform, not to look pretty or distract from the message that map is trying to convey.

### *Adding a Map Title*

1. Under the Insert menu select “Title”. A small title box should now appear at the top of the map layout with the word title highlighted.
2. Simply type the name for your map into the title box and it should appear. When you hit enter the highlight color should disappear.
3. In order to alter the font type and font size, simply double click on your title, bringing up the properties box.
4. On the bottom right corner of the Text Properties click on the Change Symbol button.
5. The user is now able to scroll through a variety of options for font types and font sizes.
6. When you have selected a font type and size that you like simply click on the OK button, which will take you back to the Text properties box. Now click on OK and your new font type and font size should appear.

### *Adding Other Elements to the Layout*

In addition to editing basic map elements, ArcMap also allows you to add other elements to the layout such as charts, tables and pictures. These elements often help to convey complex or supplemental information not easily conveyed in the map. In crime mapping these elements are used quite often. **Charts** are often added to provide tabular views of crime distributions that are being displayed in the map. An example would be a point map of all homicides for a twelve-month period with a chart showing the monthly

distribution of homicides. **Tables** are also used quite often in crime maps to provide detailed information about the map elements. An example would be a point map of robbery locations with a table providing detailed information about suspect information and motive. Finally, **pictures** are also used quite a bit in crime maps to give them a more professional appearance. An example would be including a .gif image of a department or universities logo above the legend to inform the reader of the source of the map. In addition to charts, tables, and pictures ArcMap also allows you to add basic map elements such as data frames, compasses, scale bars, and legends. With the exception of adding a data frame to display another map, this is primarily used when a user wants to design a layout completely from scratch.

1. In order to add these other elements select either “Picture” or “Object” from under the Insert menu. **Picture** will allow you to add images to the layout, whereas **Object** will allow you to add charts and other documents to the layout.
2. For this example select Picture. This should take you to the Chapter 5 folder within the CM Lab exercises folder. However, if you are not immediately taken there then maneuver to this folder.
3. Find HPD.gif and select it and then click OK.
4. An image of the Houston Police Department badge will now appear on the layout. In order to change its position and size follow the directions above. Figure 5.2 ([www.mappingcrime.org](http://www.mappingcrime.org)) provides an example of a map layout with the Houston Police Department badge contained within it.

### ***Printing and Exporting a Layout***

Once you have your layout like you want it you are down to the printing or exporting stage. Most times you will simply print out the map, but there are times when you may need to export the map as an image for use on the internet or in powerpoint presentations. To **print a map** all you need to do is simply click on the printer icon at the top of the page. This will allow you to print to your default printer or select another printer to print from.

### ***Exporting a map***

1. Select the File menu and click on “Export”..
2. After you have selected a folder to save the image to you need to select the image file type. The default file type is .jpg, but you can choose from a wealth of other options by clicking on the “Save as type” drop down menu.
3. In the “File Name” box type an appropriate name for the map and click “Export”.

### ***Map Design***

The last section of Chapter 5 is designed to provide readers with a discussion of the major factors that need to be considered when designing a map as well as provide some suggestions for how to best design a map. It is important to note that this section is not a substitute for a more complete study of cartography and map design. Those readers interested in a more in-depth discussion of the topics discussed here are encouraged to read the books in the “Further Readings” section and seek out instruction at a local College or University.

## ***Audience and Medium***

The two biggest factors to consider when beginning your map design are your map audience and the medium that the map will be read in. Both of these factors should be considered in-depth before beginning map design because of their impact on color, symbolization, and labeling. **Audience** refers to the main readers of your map. Importantly, different audiences have very different needs, requirements, and levels of comprehension of a map and thus map design will be different depending on your audience. Table A (see page 279 of the textbook) outlines the four major audiences for crime maps as well as basic needs and suggestions for color, symbolization and labeling of maps for these audiences. As you can see, the needs of the map and the sophistication of the readers impacts color, symbolization, and labeling in different ways. **Research** maps are geared toward conveying complex theoretical concepts and facts and assume an audience that is sophisticated in terms of map comprehension. Because of their sophisticated orientation, research maps should avoid excessive symbolization and labeling and use grayscales and blue to red color scales to show differentiations. In contrast, maps for **Public** consumption need to be geared toward ease of understanding by anyone, a sort of least common denominator approach. Thus, these maps should emphasize non-confusing color schemes and labeling of major geographic features, with particular attention being paid to the size and shape of symbols so as to avoid offending, confusing, or misleading readers. Maps for **officers** present a different purpose, as they are designed to provide quick and easy depictions of crime series and patterns to officers who are busy yet knowledgeable of the geography of the area they patrol. Officer maps should thus focus on black and white color schemes, simple symbolization and basic

labeling to facilitate officer comprehension while patrolling. Finally, **police manager** maps are a combination of research and officer maps in that they are often portraying difficult concepts, but must do so in a quick and easy manner. Maps for police managers should use gray scales and blue to red color ramps as well as avoid excessive symbolization and labeling. While far from hard and fast rules, these suggestions should be reviewed before beginning map design and altered to fit your specific needs.

In addition to the above suggestion based on audience type, users should also consider the suggestions given in Table B (See page 279 of the textbook), which describes the impact of **medium** on color, symbolization and labeling. **Medium** is the method in which the map will be read and consists of normal print, large print, overheads, and electronic media. As with audience, the medium in which the map will be read impacts color, symbolization, and labeling. Notice that maps that will be printed out on **plotters** need more intense colors, larger symbols, and larger and more detailed labeling than maps printed on normal paper because of the large size of the final product. Similarly, maps that are printed out on **overheads** also need darker colors, larger sized symbols, and darker and thicker labels because of the tendency of light colors and lettering to wash out on overheads. In contrast, one of the major issues for printing a map out on **normal paper** is whether or not the printer is color and its subsequent impact on color choices. Specifically, if a map is to be printed out in black and white then the symbolization and color scheme should de-emphasize colors to avoid problems with differentiating between symbols. Finally, maps used in **electronic media** should use simple colors readable on different monitors and use easy to read fonts as labels may be difficult to read depending on monitor resolution.



## *Map Colors*

In making crime maps, or any maps for that matter, the three biggest issues to deal with are choosing colors and symbols that fit best with your map as well as labeling geographic features on the map. **Color** is extremely important in map making because it allows the map designer to convey a host of information quickly and easily. Color has three basic components **Hue**, which is the actual color of a color, **Value**, which refers to the lightness or darkness of a color, and **Saturation**, which refers to the color's intensity, brilliance, or purity of a color (Monmonier, 1996). Changing the hue, value, and saturation of a color can dramatically alter its appearance and its usefulness in a map. In crime mapping applications color can be used to show crime intensity or density in an isoline map, different values in an area or choropleth maps, and to differentiate between incident types in a point map. However, while color can be beneficial to understanding differences and intensities on a map, it can also be used to mislead and confuse the reader. Harries (1999) suggests that color can be used to make dangerous or politically charged areas appear less threatening through the use of "cool" colors that make the reader think the area is non-threatening. Other examples include giving violent crimes such as homicide emotionally charged colors such as red, which is usually associated with blood (Harries, 1999). Finally, if color schemes are poorly chosen they can easily confuse the reader, making maps hard to understand. Examples include using bright multi-color spectrums to show differences in crime rates at the census tract level in an area map.

### *Suggestions for Color*

- Don't go overboard with colors just because they are available in ArcMap.
- Use color ramps (blues to reds) to show differences in intensity in isoline maps.
- Use darker colors to denote high values and light colors to denote low values (Harries, 1999).
- Gray scales are always a good choice when unsure about color choices.

### *Symbolization*

The other important factor that crime mappers will deal with in map design is the choice of **symbols** for their map. As with color choices, symbolization is vitally important because it allows a user to convey a large amount of information quickly and easily. There are three basic symbols in all maps, **Points**, which are used to mark specific locations such as crime incidents, **Lines**, which are used to symbolize roads or rivers, and **Areas**, which are used to symbolize the form and shape of jurisdictions or boundaries. Symbolization can make things more clear for a reader such as when different symbols are used to denote different crime types or when graduated points are used to denote differences in amount of crimes at a location or in an area. However, symbol size and shape can also confuse and mislead just as easily. If symbols are used that are too big for the scale of the map, the reader will be misled as to the location of the crime, the value of incidents being portrayed, or its importance in relation to other information on the map. Similarly, making all non-violent crime symbols larger than the symbols used for violent crimes can mislead readers into believing violent crime is not as big a problem. Other potential problems include the emotional connection to certain

symbol types, such as using little chalk outlines to portray homicide locations as in Figure 5.3 ([www.mappingcrime.org](http://www.mappingcrime.org)). Map makers must always be cognizant of the audience of their map and the potential to mislead, confuse, or disturb the reader through the choices they make as to symbolization.

### ***Suggestions for Symbolization***

- Symbol size should match map scale.
- Symbol shapes should be logical and not confuse or mislead.
- Use graduated symbols to denote differences in amount and different symbol types to denote differences in crime types.
- Use color when necessary and make choices that are logical and not distracting.

### ***Labeling***

The last of the three major issues in map design is labeling. **Labeling** involves identifying places and geographic features on a map (Monmonier, 1993). As with color and symbol choices, label choices can help facilitate understanding of maps as well as confuse and mislead readers. The biggest issues with labeling revolve around font type and size, proximity to the place/feature and the amount of labeling. Poorly chosen font size and font type can to reader confusion as when labels are too small to read or so large they overlap different feature making it difficult to associate a label with the correct feature (see Figure 5.4, [www.mappingcrime.org](http://www.mappingcrime.org)). This is closely related to proximity, which deals with the association of labels with their given place or geographic feature. If a label is too close or not close enough to a place or feature it can lead to confusion as to the correct place that is being labeled. Figure 5.5 ([www.mappingcrime.org](http://www.mappingcrime.org)) provides an

example of poor proximity labeling can lead to confusion in map interpretation. Finally, too much and too little labeling can also lead to confusion for readers. Too little labeling can leave readers wondering which places are which, whereas too much labeling can overwhelm a map and make interpretation next to impossible. Figure 5.6 ([www.mappingcrime.org](http://www.mappingcrime.org)) provides examples of how too little and too much labeling can confuse the interpretation of a map. Importantly, as with choosing colors and symbols, labeling is part art, in that a map maker must attempt to get the right feel for how much, how close and what size and type a label should be. In crime maps features that are usually labeled include major roads, jurisdictional boundaries (Counties), major cities and towns, landmarks, and in some cases specific incidents.

### ***Labeling Suggestions***

- Font size should fit with the scale of the map. The larger the scale the more detailed labeling that is possible.
- Do not use strange looking fonts, stick with fonts such as Times New Romans.
- Labels should not be right on top of a feature so as to obscure the symbol you are labeling.
- Labels should not be on top of each other. Make sure that labels of different features do not overlap each other.
- Strike a balance between too much and too little labeling.

### ***Other Issues in Map Design***

In addition to color and symbolization choices, map makers must also consider several other important issues when designing a map. One important issue is internal organization. **Internal Organization** is the alignment of the different map elements in the layout. It should be remembered that the main focus of the map is the map itself and not the map elements. Thus, make the map elements complementary to the map frame and not distracting. A final important aspect is the **focus of attention** of the map. Focus of attention is the idea that people read maps like they would a page in a book, and thus the main focus of the layout should be the main map (Harries, 1999). Ideally the key issue to be addressed in a map would be closest to the center of the page as possible.

### **Chapter 5 Project: Designing Effective Maps**

1. Create a new theme and add the following themes:
  - a. Houston.shp
  - b. Homicides.shp
  - c. Bars.shp
  - d. Schools.shp
2. Create two different layouts, taking special consideration of the suggestions given in this section.

**Map A:** Portrait map showing point locations of all homicide incidents for August through December (Homicide Codebook from [www.mappingcrime.org](http://www.mappingcrime.org)) and their relation to school locations. In addition you must insert the HPD symbol to the map.

Refer to the suggestions in this section to help you in designing your map.

Remember you must query the homicides theme to get only those incidents involving incidents from August thru December.

**Audience:** Public

**Medium:** Large format print

**Map B:** Landscape map showing homicide rates by census tract for all of Houston and their relation to bar locations.

**Audience:** Research

**Medium:** Powerpoint presentation (electronic media).

Check your map design against the authors maps on-line at [www.mappingcrime.org](http://www.mappingcrime.org)

### **Review Questions**

1. Explain the difference between large scale and small scale and discuss their importance to map design.
2. Where does map design take place in ArcMap.
3. What two map features are controlled through a dynamic link to the view mode?
4. What four editing changes can be made to the basic map elements?
5. What are two of the most important things to consider when designing a map?
6. Audience and medium most directly impact what aspects of map design?

### **Further Readings**

*Map Design Readings*

1. Dent, B.D. (1993). *Cartography: Thematic Map Design*. Dubuque, IA: William C. Brown. Second Edition.
2. Monmonier, M. (1993). *Mapping it Out: Expository Cartography for the Humanities and Social Sciences*. Chicago: University of Chicago Press.
3. Monmonier, M. (1996). *How to Lie with Maps*. Chicago: University of Chicago Press.

## **Section II: Basic Spatial Analysis**

Now that you have a basic understanding of how to use ArcMap to enter, query, manipulate, and display crime data you will begin to learn how to perform spatial analysis. While many researchers and police departments create maps that are little more than point maps of crime incidents when analyzing crime patterns, spatial analysis is vital to determining true spatial patterns. Specifically, in order to determine if crime incidents are truly random or clustered in space, spatial analysis must be conducted. Section II begins the process of informing students how to assess spatial patterns by introducing students to several basic spatial analysis techniques commonly used to assess spatial patterns. Chapter 6 discusses proximity analysis techniques such as buffer analysis and theme selection. Chapter 7 introduces students to spatial distribution measures such as mean center, standard deviational ellipse and spatial autocorrelation. Finally, Chapter 8 discusses distance analysis measures such as nearest neighbor analysis. As mentioned earlier, the workbook is not designed to be a comprehensive spatial statistics textbook as such an undertaking is far beyond the scope and purpose of this book. Rather, the workbook is designed to introduce students to spatial analysis techniques, their importance to crime analysis, when to use them, and how to interpret their results. For more comprehensive and technical discussions of the techniques discussed in these and subsequent chapters students should pursue the readings in the “Further Readings” section of each chapter.



## Chapter 6: Proximity Analysis

### Goals of Chapter 6:

1. Understand proximity analysis and its importance to crime mapping.
2. Understand how to conduct Buffer analysis
3. Understand how to conduct geographic selection
4. Perform proximity analysis exercises.

### *Introduction*

Up to this point the workbook has focused almost solely on providing students with an understanding of how to use a GIS to enter, manipulate, query, and display crime data. While developing a basic understanding of ArcMap is important to effective crime mapping, equally as essential is understanding how to perform spatial analysis of crime data to assess crime patterns and distributions. Chapter 6 introduces the student to basic spatial analysis through its discussion of proximity analysis techniques. Specifically, students will learn how to perform buffer analysis and geographic selection to answer specific questions about crime data.

### *What is Proximity Analysis and why use it?*

**Proximity analysis** is a term used to describe several different techniques used to assess the proximity of crime incidents to other geographic features. Proximity analysis techniques are commonly used to determine how many crimes occur within a set distance of a specific location or class of locations. An example of this type of analysis is

determining how many drug related crimes occur within 500 feet of schools, an increasingly popular analysis with the increase in drug free zones around schools. While proximity analysis does not involve any spatial statistics and is rather basic in its creation and interpretation, it is an important part of spatial analysis because of its utility in answering everyday questions concerning crime incident locations and patterns.

However, because proximity analysis is rather simple in its analysis functions it is limited in its utility. Specifically, proximity analysis cannot determine randomness or clustering of incidents or provide any other useful statistical information, it is simply a set of techniques for determining the proximity of incidents to geographic features or locations.

The main methods of proximity analysis are Buffer zones and geographic selection. **Buffer zones** are usually circles that are drawn around specific geographic features or locations at user defined distances. The distance used for creating a buffer zone is not a constant and is usually dependent on either an investigative or analytical purpose (Harries, 1999). The example above involving drug free zones around schools would involve a buffer zone of 500 feet around every school in a jurisdiction, but this distance could easily be increased to any distance the user desired. Because the buffer zone distance is user defined and guidelines are non-existent for their creation, careful thought must go into the investigative or analytic purpose behind the creation of the buffer zone. As will be made clear in completing the exercises, the distance used in buffer zone creation greatly impacts the results of proximity analysis. Because of its ease of creation and interpretation, buffer zone analysis is one of the more common methods of basic spatial analysis. Specific examples of buffer zone analysis include:

- **Sex Offender Registration:** Used to determine if registered sex offenders are living too close to restricted areas such as school or parks.
- **Drug Free Zones:** Used to determine if crimes are occurring within drug free zones and are thus susceptible to stiffer penalties.
- **Alcohol Related Crimes:** Used to determine the proximity of alcohol related crimes to alcohol related establishments such as bars and liquor stores.
- **Probationers/Parolees:** Used to determine if probationer/parolee residences violate probation/parole conditions by being too close to specific locales (schools, bars, etc..).

The other major proximity analysis technique is geographic selection.

**Geographic selection** is a technique that allows the user to select layer features that are within a specified distance of a different layers features. Specifically, a user can select all homicides that are within .25 miles of a school or within 1000 feet of a bar location.

There are three main differences between geographic selection and buffer zone analysis.

First, in geographic selection no graphic zones are actually drawn as with buffer zone analysis. While some find this beneficial because it makes it easier to analyze the selected features (they are yellow as with a query), others would rather have the graphic zones for visual impact. A second difference is that as with queries (Chapter 4), a new layer can be created from the results of a geographic selection. This can be highly beneficial for analysis and visualization and is a feature that is often used by researchers. Finally, more advanced queries can be performed with the results of a geographic

selection. Once a geographic selection has been conducted further geographic selections can be conducted with the results or advanced queries can be conducted with the geographic selection results through the query builder (Chapter 4). For example if a user performs a geographic selection for all homicides with a half mile of schools they can then perform a query on the results in which they find all victims under age 18. This is a particularly valuable tool for looking at specific crimes or offender types in conjunction with proximity to specific geographic features.

### ***Buffer Zone Analysis***

1. Open ArcMap and select “A new empty map” and click OK.
2. Click the “Add data” button and maneuver on the C drive to the “CMLab Exercises” folder and then select “Chapter 6”.
3. From the Chapter 6 folder select the following layers:
  - a. Houston.shp.
  - b. Homicides.shp
  - c. Bars.shp
4. Make all three of these layers viewable and edit their appearance to make the homicides easily distinguishable from the bars.
5. Under the “Tools” menu select “**Buffer Wizard...**”.
6. When the “Buffers wizard” appears select “The features of a layer” and then scroll thru the layer options selecting “Bars”. For this example “Bars” will be the geographic feature around which buffers will be created. Importantly, any layer that is added to the view can be used to create buffers.

7. Click on the “Next” button in the bottom right corner. This advances the wizard to the next stage.
8. When the “How do you want to create buffers?” page appears select “At a specified distance” and make sure that the box to the right of it has a 1 in it. This is the distance field and allows the user to enter any distance they desire. It is important to note that this field must be edited in combination with the “Distance units” field at the bottom of the page. Thus, for analytic purposes if you wanted to make buffers that were 1000 feet from all bar locations you would first need to change the “at a specified distance” box from 1 to 1000 and then alter the “distance units” box from miles to feet.
9. Make sure the “Distance units” is **miles** for this example and click on the Next button.
10. When the next page appears make sure that “Dissolve barriers between buffers” is selected “yes”. This is the default option and it simply dissolves the circles between the buffers if they should overlap. While this is a personal choice whether to dissolve the buffers or not, most people find it more visually attractive to dissolve the circles.
11. The final stage is deciding “Where you want he buffers to be saved?”. The default is “**In a new layer**” which makes the buffers into a new layer that can be edited just like a regular layer. The second option is “**in an existing editable layer**” which adds the buffers to an existing layer of your choice. The final option is to save the buffer “**As graphics layer in the data frame**”, which simply creates buffers in the form of graphics which can be deleted from the data view.

It is recommended that users choose either “as graphics layer in the data view” or the default option “in a new layer” as these are the easiest to work with. For this example choose the “as graphics layer in the data view” option.

12. When you have selected your output choice simply click on the “Finish” button in order to calculate the buffers, which will appear on the screen when ArcMap is finished. Figure 6.1 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows you what your map should look like when you have finished.

**Interpretation:** You should now see circles of 1 mile in diameter around every bar location in Houston. In order to use this analytically you can zoom in on specific areas to count the number of homicides within the buffers. If you zoom in on the city center you will notice that all homicides are within 1 mile of a bar and that the buffers are difficult to see because there are so many bars close together. This illustrates that importance of choosing a buffer distance wisely as this has little practical impact other than to tell the student that there are a lot of bars in Houston, so many that Homicides are certain to occur near a bar.

**TIP:** You can make the buffers editable and easier to see by changing the output to “in a new layer”.

1. Go back and conduct the exercise again but this time change the output to “in a new layer” and save it in “My Documents” as Buffer1.
2. Once the buffer zone has been added to the data view click on the layer color in order to edit its appearance.
3. Change the fill color to “no color”, this will make the buffers clear.

4. Next increase the “Outline Width” to 2 by clicking on the up arrow next to the outline width. This will change the weight of the buffer zones making them darker in appearance.
5. Finally, in order to add color to the buffer zone lines select a color of your choice in the “Outline Color” menu.
6. Once you are done with these three alterations to the buffer click on the OK button. You should now be able to see the buffer much more clearly and thus be able to analyze the relationship of homicides to bars more easily.

### ***Geographic Selection***

1. Under the view menu select “Data Frame” in order to create a new data view for which to conduct the next exercise. After you have added a new data frame add the following layers from Chapter 6 in the “CMLab Exercises” folder.
  - a. Houston.shp
  - b. Homicides.shp
  - c. Schools.shp
2. Make all three of these layers viewable and edit their appearance to make the homicides easily distinguishable from the schools.
3. Under the “Selection” menu select “Select by Location”.
4. When the “Select By Location” box appears you will notice several different drop down menus that allow you customize your geographic selection. The first menu is the “I want to” menu, which determines what kind of location selection you will be doing. For our purposes, and for most commonly conducted analyses that

- are performed, the operation conducted is “select features from”. This operation simply allows you to select features from a current layer or layers based on some kind of geographic relationship.
5. The next menu is “the following layers” menu, which allows the user to pick which layers they want to select features from. For this example place a check mark in the “Homicides” layer.
  6. The next drop down menu is the “That” menu. This menu allows you to choose the geographic relationship that is to be selected on, such as “are within a distance of”. Under this menu there are 11 different options allowing the user to have a wide range of choices in their selection methods. For this example scroll down and “**Are Within Distance Of**”. Thus, for this example we are selecting all homicides that are within a specified distance of something, which will be determined in the next two menus. .
  7. The next drop down menu is the “**the features in this layer**” menu, which allows the user to chose the layer that you want your active layer (Homicides) to be selected if they are within a specified distance of. In this case select Schools, because we want to select all homicides within a specified distance of school locations. Importantly, you can choose from any other layer that has been added to the view.
  8. The final aspect of the geographic selection is the “apply a buffer to the features” menu, which allows the user to determine the distance to be used for the selection. In this case type a 1 in the first box and scroll down to miles in the second box. Thus our query will select all homicides that occurred within 1 mile of schools. It



- is important to note that you can type smaller distances than 1 mile such as .1 mile. Moreover, you can also change the distance to any units you desire (feet, kilometers, etc..) simply by scrolling through the drop down in the second box..
9. When you are ready click on Apply. This will select all of the homicides within 1-mile of schools and make them yellow. Figure 6.2 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows you what your map should look like after the theme selection is complete.

### ***Options for Theme Select Results***

- A. **Create new Layer:** You can create a new layer of just the selected homicides by right clicking on the homicide layer and selecting “Create Layer from Selected Features” under the Selection menu (See Chapter 4 for more instructions). This will allow you to manipulate and edit a new layer and is beneficial if you need to analyze or view only these selected features.
- B. **Clear Selections:** You can clear the current selections by clicking on the Selection menu and scrolling down to “Clear Selected features”. This will clear all selected features, allowing you to return your view to before the selection process begun. This is recommended when you want to start the geographic selection process over from the beginning.
- C. **Advanced Query:** You can perform a more advanced query by either using the Select by Attributes builder (see Chapter 4) or using the geographic selection option again. In order to query from the selected results the user must choose the correct layer in the “the following layers” option. In this example the layer would be “Homicides Selection”. This is a particularly useful option as it allows the user to first select all incidents within a specific distance of a geographic feature and then query those results

further based on some other important information such as offender type. An example would be to geographically select all incidents within 500 feet of schools and then query those results to find all drug related crimes, thus providing all drug crimes within 500 feet of schools.

**Interpretation of Geographic Selection Results:** The results of a geographic selection are interpreted in the same manner as buffer zone results. In fact the results are easier to interpret simply because they are selected features and thus an accurate count of the incidents can be gained by simply right clicking on the layer that was selected from and choosing the “Open Attribute Table” option after the geographic selection is complete and looking at the data line under the attributes table. In the example above, 296 of the 536 homicides that occurred in Houston occurred within 1 mile of a school.

### **Exercises for Chapter 6**

**Buffer Zone Exercise:** Conduct two different buffer zone analyses to determine the extent that homicides occur near bars.

**Map A:** Buffer zone of 1-mile around bar locations.

**Map B:** Buffer zone of 1000 feet around bar locations.

Compare your results to the correct analysis on the Textbook webpage  
[www.mappingcrime.org](http://www.mappingcrime.org)

**Geographic Selection Exercise:** Conduct two different geographic selection exercises to determine the extent that homicides occur near schools.

**Map A:** Geographic selection of all homicides within .5 miles of schools.

**Map B:** Geographic selection of all homicides within .5 miles of schools AND the victim age is under 18. See the Code Book in Appendix C for victim age information and see Chapter 4 for information on how to conduct queries.

Compare your results to the correct analysis on the textbook webpage

[www.mappingcrime.org](http://www.mappingcrime.org)

### **Review Questions**

1. What is Proximity Analysis?
2. How is proximity limited in its utility?
3. What are some common uses of Buffer Zone analysis?
4. List the three main differences between buffer zone analysis and geographic selection analysis.

## Chapter 7: Spatial Distribution Analysis

### Goals of Chapter 7:

1. Understand spatial distribution analysis and its importance to crime mapping.
2. Understand mean center analysis
3. Understand standard deviation analysis
4. Understand spatial autocorrelation analysis
5. Perform spatial distribution analysis exercises.

### *Introduction*

While Chapter 6 introduced students to spatial analysis of crime data through a discussion of proximity analysis techniques, these techniques were fairly basic and without statistical significance. Chapter 7 however introduces students to more advanced spatial analysis through a discussion of first order spatial analysis techniques. **First order analysis techniques** are a group of techniques that describe the general spatial pattern of crime incidents such as where the distribution is centered, how far it spreads, and in what direction it disperses (Levine, 2002). Thus, first order analysis techniques provide a good *general* description of the distribution of crime incidents as to its average location, movements, and directional change through time (LeBeau, 1987). Specific techniques to be discussed include centographic measures such as mean center and standard deviation analysis as well as spatial autocorrelation analysis. As mentioned previously, this Chapter is not intended to provide an in-depth discussion of the assumptions and statistical methods behind each technique. Rather it is designed to

provide the student a basic introduction into how to use these techniques to analyze crime distributions. Students seeking more in-depth discussion concerning these techniques are encouraged to pursue the readings in the “Further Readings” section at the end of the Chapter.

### *Mean Center Analysis*

One of the most commonly used first order techniques is **mean center** analysis. The mean center is by far the simplest technique used to describe a spatial distribution and is akin to the mean of a series of numbers in regular statistics (Levine, 2002). In technical terms the mean center is the average location or center of gravity of the spatial distribution of a phenomenon (Levine, 2002). Le Beau (1987) describes the mean center of a distribution of incidents as “a synthetic point or location representing the average location of a phenomenon”. In other words, the mean center provides the user with a point that is the approximate center of a distribution of crime incidents. The utility of mean center analysis is its ability to compare the distribution of different crimes or the same crime at different times. Figure 7.1 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows two different maps with the results of mean center analyses. Map A shows the mean centers for both robbery and burglary incidents in Houston. The mean center for robbery is close to the city center whereas the mean center for burglary is closer to the suburb area, indicating that the spatial distribution of robbery and burglary are very different. Map B provides the mean centers for robbery incidents for two separate months. In this case the mean centers are almost on top of each other, indicating that the distribution of robbery incidents stayed fairly consistent spatially between the two months.

While the mean center is a good measure of the spatial distribution of incidents it is subject to measurement problems based on the distribution of the incidents. Specifically, the mean center is prone to being influenced heavily by spatial outliers or odd shaped incident patterns. Figure 7.2 shows two maps representing the results of two different mean center analyses ([www.mappingcrime.org](http://www.mappingcrime.org)). Map A shows the results of a mean center analyses where the distribution of incidents is L shaped. In this analysis the mean center is located in the middle of the L shaped distribution outside of the actual incident pattern. Map B shows the results of a mean center analysis for a bi-modal distribution of incidents, or incident pattern with two different areas of concentration. In this case the mean center is in between the two clusters of incidents, which while technically is the center of the distribution, is misrepresentative of the actual distribution of incidents. In both examples the mean center results placed the mean center of the distribution outside of the actual distribution. This illustrates the point that users must be aware of the nature of their distribution for the mean center to be a credible measure (Stephenson, 1980). In the examples above the user should possibly use a different spatial distribution technique, or in the case of the bi-modal distribution conduct mean center analyses on both parts of the distribution.

### ***Performing Mean Center Analysis in Crimestat***

The directions provided on how to conduct the various statistical analyses within Crimestat are abbreviated instructions. For a complete set of instructions as well as the rationale and theory behind each of the various techniques please download and read the Crimestat 2.0 manual from the textbook webpage ([www.mappingcrime.org](http://www.mappingcrime.org)) Follow the directions in Appendix B in order to install Crimestat 2.0 on your computer.

1. Double click on the Crimestat shortcut on your desktop or wherever you have placed the program. When the Crimestat splash screen appears click anywhere in the screen to advance to the main menu.
2. When the main menu appears you will see four main tabs, Data setup, Spatial description, Spatial modeling, and Options. These are the four main sections of Crimestat, each of which has numerous subheadings that will be explored more fully as necessary (for a full explanation of Crimestat and its many uses download the Crimestat 2.0 manual from the textbook webpage, [www.mappingcrime.org](http://www.mappingcrime.org))
3. **Data Setup:** In order to conduct any type of analysis within Crimestat you must first enter data in to the program. In the case of a mean center analysis all that is necessary is for data to be added under the **Primary File** tab.
4. With the Primary File selected, click on the “Select Files” button on the top right corner of the screen. This will allow you to navigate to the data file that you wish to use for analysis. In all analyses conducted in Crimestat the Primary File is the default data file. Importantly, some analyses that will be conducted later in the workbook will require data to be entered into the secondary file as well as the Reference File.
13. Navigate on the C drive to the “CMLab Exercises” folder and than select “Chapter 7”. Within Chapter 7 select “Homicides.dbf” and click OK taking you back to the main menu. Under the Variable column all of the different variables will all now have Homicides.dbf, and its path, as the file name. Notice that the default file type is .dbf, indicating that the file is a database file and not a shapefile that is used in Arcview and ArcMap. While .dbf is the default, if you

scroll down the file type you will see that Crimestat will allow you to input several different file types as well as make connections to outside database managers.

14. In the variable section are four different columns, name, File, Column, and Missing values. **Name** provides the name of the different variables in Crimestat, such as X, Y, Weight, Intensity, Time, Directional, and Distance. **File** provides the name of the file that is currently the primary file for the analysis. **Column** controls the selection of the variable within the file being analyzed and **Missing values** allows users to define how missing values are handled in their data.
15. **Defining Variables:** After selecting your data file to be analyzed the next step is to inform Crimestat which variables within your data file contain the X and Y coordinates. This is an essential step as Crimestat uses X and Y coordinates to perform the analyses, and thus unless variables containing these coordinates have been identified Crimestat will not be able to perform the analysis. In this example, for the X variable scroll down in the column field until you find the variable named X\_COORD and for the Y variable select Y\_COORD.
16. **Setting Coordinates and Data Units:** The final step in setting up the Primary file is to select the coordinate system and data units. Crimestat is capable of performing analysis on data using several different types of coordinate systems including Spherical, Euclidian, and Angles and their corresponding data units decimal degrees, feet, and meters. For our purpose make sure that spherical coordinates and decimal degrees are selected.



17. **Mean Center Setup:** Click on the Spatial Description tab and then click on the Spatial Distribution subtab. You should see that Crimestat provides several different techniques that cover spatial distribution, more than will be discussed in this workbook (for more discussion on these other techniques see the Crimestat manual).
18. Place a check mark in the box to the left of the line stating “**Mean Center and Standard Distance (Mcsd)**”. This means that we have chosen to conduct a mean center and standard distance analysis. Importantly, in conducting this analysis several different centographic analyses will be conducted, some of which will be discussed more fully in this Chapter.
19. Click on the **Save Results To** button to the right of the check box. In the “Save Output to” box, select Arcview SHP and then click on the browse button (Crimestat lets the user select the method of the output from Arcview Shapefiles to Mapinfo MIF files, for our purposes always use Arcview SHP files). Navigate to the Chapter 7 folder within the CMLab Exercises folder and name the file **mean**. Click OK.
20. When you return to the Spatial Distribution screen click on the “Compute” button in the lower left hand side. This will start Crimestat conducting the mean center analysis.
21. **Results:** When Crimestat finishes with the mean center analysis a results screen will appear showing you in a text format the results of the analysis. Notice that Crimestat has computed several different centographic measures, including the Mean center, Geometric mean, Harmonic mean, Standard distance deviation and

the Standard deviation of the X and Y coordinates. The Geometric and Harmonic Mean are two alternative centrophobic measures that discount extreme values better than the mean center. They are not often used in crime analysis but are included in Crimestat and calculated when the mean center is calculated. The Standard distance deviation and Standard deviation of X and Y coordinates are standard deviation measures and will be discussed more later in the Chapter.

**TIP:** If you notice that there is no information for the mean or the results of the three different means are very different than you probably have a problem with either your primary file setup or your data.

22. Close the results box by clicking on the close button, minimize Crimestat and open ArcMap and select “A new empty map” and add the Houston.shp layer from Chapter 7 in the CMLab Exercises folder.

23. **Adding the Mean Center Shapefile:** Click on the **Add Data** button in the view and navigate to the Chapter 7 folder where you saved the results of the mean center analysis. Look for a file named “**Mcmean.shp**”. The **MC** in front of the word mean (the name you used to save the results) stands for mean center. Every time Crimestat performs an analysis in which output is generated as a shapefile, it places a prefix in front of the name you use to save the results. This prefix is designed to make it easier for the user to identify results. Other prefixes generated during mean center analysis include:

**GM:** Geometric Mean

**HM:** Harmonic Mean

**XYD:** Standard Deviation of the X and Y coordinates

**SDD:** Standard Distance Deviation

24. After you have added the Mcmean.shp file to the view make it active. As Figure 7.3 ([www.mappingcrime.org](http://www.mappingcrime.org)) illustrates, the mean center should be fairly close to the middle of the map near the city center.

**Interpretation of Mean Center Results:** The mean center in the city center indicates that the approximate middle point of homicides in Houston is the near the city center. Importantly, if we were to conduct mean center analyses for different crimes we could determine how similar the distributions were. Moreover, mean center analysis could also be conducted on homicide distributions for other years in order to determine the degree that the distribution of homicides moves over time.

**TIP:** The color, shape, and size of the mean center can be edited in the same manner as any other symbol within ArcMap.

***Standard Deviation Analysis:***

Another important first order statistics used in analyzing crime distributions is the standard deviation. Measures of **standard deviation** are essentially two-dimensional versions of a basic standard deviation measure in statistics. The utility of standard deviation measures in analyzing crime distributions is that they can describe the dispersion, orientation, and shape of the distribution of crime incidents (Levine, 2002). As with mean center analysis, the main use of standard deviation analyses is to compare the distributions of different crimes or the same crime at different times. By using the various standard deviation analysis techniques the user can compare the dispersion, orientation, and shape of different crime distributions. Crimestat calculates three

different measures of standard deviation, the standard deviation of X and Y coordinates (XYD), the standard distance deviation (SDD), and the standard deviational ellipse (SDE), all of which will be discussed separately.

The **standard deviation of X and Y coordinates** (XYD) provides a simple measure of the dispersion of a crime distribution. Its output is a rectangle, with rectangles that are longer than they are tall having a more east west orientation. Conversely output that is taller than it is wide has a more north south distribution. Figure 7.4 shows the simulated results of XYD analysis for both auto theft and car jacking incidents in Houston ([www.mappingcrime.org](http://www.mappingcrime.org)). The box representing the XYD results for auto theft is wider than that of the box representing the XYD results of car jacking, indicating that the dispersion of auto theft is more east west than that of car jacking which is a more north south dispersion. Unfortunately, one of the problems associated with XYD analysis is that it is unable to provide any information about the shape of the distribution, as it is solely a measure of the dispersion of a distribution.

A second measure of standard deviation is the standard distance deviation or standard distance as it is often called. The **standard distance (SDD)** is the closest to a two dimensional equivalent to the standard deviation of the three standard deviation measures included in Crimestat. This measure is considered a more useful measure of the dispersion of a spatial distribution from the mean center than XYD analysis (Le Beau, 1987). Its main improvement of XYD analysis is that it provides a single summary statistic of the dispersion of crime incidents as opposed to two separate statistics that are created in XYD analysis (dispersion of X and dispersion of Y) (Levine, 2002). As with other standard deviation analyses its benefit to analyzing crime distributions is that it can

compare the dispersion of two different crime types or the distributions of the same crime at two different time periods. Figure 7.5 shows the simulated results of SDD analysis for both auto theft and car jacking in Houston ([www.mappingcrime.org](http://www.mappingcrime.org)). As opposed to XYD analysis the results of SDD analysis are ellipse that show the dispersion of the two distributions. In this case, Auto theft with the bigger ellipse has a wider dispersion of incidents than does car jacking with its much smaller ellipse, indicating that car jacking incidents occur in a smaller area than auto thefts.

The final standard deviation measure to be discussed is **standard deviational ellipse (SDE) analysis**. Standard deviational ellipse is similar to standard distance (SDD) except that it is better able to represent distributions that are skewed in one direction or another, a common occurrence with crime patterns (Levine, 2002). As opposed to SDD, standard deviational ellipse analysis is capable of handling skewed distributions, thus providing a measure of the distributions shape and orientation. As we know, few crime distributions are perfectly symmetrical, most are skewed in one direction or another based on various factors such as built environment, criminal opportunities, and other factors. Thus, SDE is an excellent measure for assessing the spatial dispersion, orientation and shape of crime distributions at different time periods or the distributions of two different crime types. Figure 7.6 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows simulated results of SDE analysis for car jacking and auto theft in Houston. The ellipse for the auto thefts is larger than that of car jackings and is pointed more in an east west direction than the smaller more north south ellipse representing car jackings. This indicates that the distribution of auto thefts is more dispersed and more differently shaped than the distribution of car jackings.

### *Performing Standard Deviation Analyses*

1. If you have left Crimestat open from performing mean center analysis than go to the spatial distribution subtab under the spatial description tab. If however, you have closed the Crimestat than follow the directions above for performing the mean center analysis. Importantly, if you have already completed the above exercise than the YXD and SDD analysis have already been completed and the results should be saved in Chapter 7.
2. Once you have selected “homicides.shp” for the primary file and completed the rest of the setup as described above for mean center analysis, go to the spatial distribution subtab and place a check mark in the box next to “Standard Deviation Ellipse”. This will tell Crimestat to conduct a SDE analysis.

**TIP:** If you still have a check mark in the box next to “Mean center and Standard Distance” you may want to deselect it. Crimestat is capable of running multiple analyses at once and will thus re-perform the mean center analysis again. While this is not a problem, it may take more time to complete the SDE analysis.

3. Click on the “Save Results to” box next to the SDE analysis and follow the instructions from above, saving the results as a .SHP file in the Chapter 7 folder. Importantly, save the file as “ellipse” in order to make it easier to identify the results when you add them to the view.
4. Click OK and when you return to the main screen click compute.
5. When the analysis is finished close the results window and minimize Crimestat. Open ArcMap and create a new view and add Houston.shp (If ArcMap is already

- open from creating the mean center than simply make the mean center not viewable.
6. Add the following layers, all of which should be contained in the Chapter 7 folder:
    - a. Xydmean.shp: Standard deviation of the X and Y Coordinates
    - b. Sddmean.shp: Standard distance deviation.
    - c. Sddellipse.shp: Standard deviational ellipse.
  7. As illustrated in Figure 7.7 ([www.mappingcrime.org](http://www.mappingcrime.org)), as you make them viewable you will see that they overlap and are fairly consistent spatially. To make it easier to see the three different analyses make all three themes clear and their outlines thicker and of different colors.

**Interpretation of Standard Deviation Results:** As discussed above all three techniques provide a measure of the dispersion of the crime distribution, in this case homicides in Houston Texas. The different analyses all point to the fact that homicide is fairly dispersed in Houston with a slightly more east west orientation than north south. Moreover, the SDE analysis indicates that the shape of the distribution is slightly skewed but not much. Importantly, the SDE and SDD results are almost identical, indicating only a slight skew in the distribution towards the upper right corner.

### ***Spatial Autocorrelation Analysis***

The last spatial distribution methods that will be discussed are used for determining the degree of spatial autocorrelation present in a crime distribution. Spatial autocorrelation is an extremely important idea for the analysis of crime distributions,

because it provides information on the degree to which the locations of crime incidents are related to each other (Levine, 2002). When the location of one crime incident (robbery location) is related to the location of another crime incident (separate robbery location) it is termed **positive spatial autocorrelation**. Generally, incidents that exhibit positive spatial autocorrelation are termed clustered. Thus robbery incidents that are clustered together in a specific section of a city exhibit positive spatial autocorrelation. Conversely, when the location of one crime incident (robbery location) is unrelated to the location of another crime location (separate robbery location) it is termed **negative spatial autocorrelation**. Crime incidents that exhibit negative spatial autocorrelation are termed spatially independent or uniformly dispersed. Thus, robbery incidents that are uniformly distributed throughout a city would exhibit negative spatial autocorrelation. It is important to note that in cities most social phenomenon such as restaurants locations, population centers, and ethnic groups exhibit positive spatial autocorrelation (Levine, 2002). This is important for crime research in that crime distributions may exhibit spatial autocorrelation simply because everything in the city exhibits positive spatial autocorrelation. While this measure is not used a great deal by practioners conducting crime analysis it is an important statistical tool for those conducting research into spatial patterns. Specifically, spatial autocorrelation measures are often used to determine the degree to which aggregate level data, such as crime rates at the census block or census tract level, are clustered within a city. In addition, these measures can also be used to determine the difference in spatial autocorrelation between two different crime types or the same crime at two different time periods.



The two main tests for spatial autocorrelation are **Moran's I** and **Geary's C**, both of which are used to determine the degree of spatial autocorrelation in aggregate data and both of which are included in Crimestat. Importantly, because these measures are for determining spatial autocorrelation in aggregate data, point data that has been used thus far for analysis will not work. Thus, in order to test for spatial autocorrelation using Moran's "I" or Geary's "C", incident level data must be aggregated to a zone level such as census tracts, police districts, etc. **Moran's I** applies to zones that have associated continuous variables attached to them (census tracts with homicide rates) and it works by calculating the value of the variable at one location with the value of that variable at all other locations (Anselin, 1992). The output of a Moran's "I" test is similar to a correlation coefficient, in that it varies from  $-1.0$  to  $+1.0$  where positive values indicate positive spatial autocorrelation and thus clustering, and negative values indicate negative spatial autocorrelation and thus spatial randomness (Anselin, 1992). When nearby values have similar values (high homicide rate next to high homicide rate) the "I" value is higher, conversely when nearby values have dissimilar values (high homicide rate next to low homicide rate) the "I" value is low (Levine, 2002). Thus, an analysis of homicide rates at the census track level that produced an "I" value of  $.875$  would indicate very high positive spatial autocorrelation and thus clustering of census tracts with high homicide rates. Conversely, an "I" value of  $-.236$  would indicate negative spatial autocorrelation and that census tracts with high homicide rates were randomly distributed near census tracts with low homicide rates.

The other test for spatial autocorrelation is **Geary's C**. Although both of these tests provide accurate measures of spatial autocorrelation Geary's C is more sensitive to

differences in small neighborhoods and works by looking at the squared deviations of observations (Levine, 2002). Because the measures are determined using different methods the output and subsequent interpretation is also slightly different. Geary's C ranges from 0-2 with values of less than 1 indicating positive spatial autocorrelation and values greater than 1 indicating negative spatial autocorrelation. Thus if an analysis of homicide rates at the census tract level produced a "C" value of .658 it would indicate positive spatial autocorrelation or clustering of census tracts that have high homicide rates. Conversely if the same analysis produced a "C" value of 1.034 it would indicate negative spatial autocorrelation, or that census tracts with high homicide rates were near census tracts with low homicide rates.

### ***Performing Moran's I and Geary's C***

Because these measures are so similar in their analysis, how to conduct and interpret them will be discussed together.

1. Open Crimestat and add as the primary file Houston.shp from the Chapter 7 folder within the CMLab Exercises folder. This is the basemap for Houston, Texas, which contains various amounts of data that have been aggregated at the census tract level.
2. For the X and Y values scroll down and select X\_Coord and Y\_Coord respectively. In this case the X and Y coordinates are for the center of each census tract, as opposed to the actual location of each incident as in all other analyses performed so far.

3. **Intensity Variable:** Unlike other analyses, both Moran's "I" and Geary's "C" require what is called an **intensity variable** in order for them to be calculated correctly. As mentioned in the explanation of these statistics, both techniques work by analyzing values at the zone level. The intensity variable serves as the value for this analysis. For this example scroll down and select **Homrate** (directly after Dblack). This variable is the homicide rate for each census tract in Houston, Texas.
4. Click on the spatial distribution subtab under the spatial description tab. Click in the boxes next to both Moran's "I" and Geary's "C". When you do this you will notice that two boxes named "**Adjust for small distances**" become active with check marks. This is a statistical weight that attempts to ensure the statistical validity of the measures by taking into account the small distances between the zones. It is recommended that adjustment for small distances be used when analyzing zones that are very close together such as census tracts.
5. After this is complete click the Compute button.

**Interpretation of Moran's "I":** In looking at the results the two main things to be concerned with are the Moran's "I" and the normality significance (z) scores. In this case the Moran's "I" is .075480 with a Z score of 31.749654, indicating positive spatial autocorrelation. As mentioned before high and positive "I" values indicate positive spatial autocorrelation. Thus according to the Moran's "I" with adjustment for small distances, homicide rates at the census tract level **are** clustered. In practical terms, census tracts with high homicide rates are near census tracts with high homicide rates and

census tracts with low homicide rates are near other census tracts with low homicide rates.

**Interpretation of Geary's "C":** As with the Moran's "I" results the two key values to consider in interpreting Geary's "C" are the "C" value and the Z value. As mentioned before, values under one indicate positive spatial autocorrelation, while values over one indicate negative spatial autocorrelation. Thus according to Geary's "C" with adjustment for small distances, homicide rates at the census tract level **are not** clustered. Rather, in contrast to the results of the Moran's "I" census tracts with high homicide rates are not near other census tracts with high homicide rates. The differing results between the Moran's "I" and the Geary's "C" can probably be attributed to the more sensitive nature of Geary's "C" at the neighborhood level. As mentioned above, Geary's "C" is more sensitive to small neighborhood differences and is probably picking up some of the census tracts with high homicide that are near census tracts with low homicide rates. (See Figure 7.8 for visual inspection of the distribution)

**TIP:** In determining spatial autocorrelation in aggregate level data it is always wise to run both Moran's "I" and Geary's "C" and to thoroughly analyze the distribution you are analyzing. Analyzing the spatial differences **visually** can help with the interpretation of the results when there are disagreements between the measures (See Chapter 2 to learn how to create maps using aggregate data). In the example above, Geary's "C" appears to be more sensitive to the neighborhood differences in homicide rates at the census tract level.

## Exercises for Chapter 7

**1. Mean Center Analysis:** Conduct two different mean center analyses to determine the extent that the distribution of homicide changed between 1989 and 1991.

**Themes required:** Houston.shp, 1989.shp, 1991.shp

**Map Product:** Mean center for both 1989.shp and 1991.shp.

Compare your results to the correct analysis on the Textbook webpage

[www.mappingcrime.org](http://www.mappingcrime.org)

**2. Standard Deviation Analysis:** Conduct two different standard deviation analyses using all three different standard deviation techniques (XYD, SDD, SDE). Determine the degree to which the distribution of homicides changed from 1989 to 1991.

**Themes required:** Houston.shp, 1989.shp, 1991.shp

**Map Product:** XYD, SDD, and SDE analyses for both 1989.shp and 1991.shp.

It may be visually easier to create three separate maps one for each different standard deviation analysis.

Compare your results to the correct analysis on the textbook webpage

[www.mappingcrime.org](http://www.mappingcrime.org)

**3. Spatial Autocorrelation Analysis:** Conduct two Moran's "I" and Geary's "C" analyses, one on adult homicide rates and one on juvenile homicide rates (See appendix C for code book information for Houston.shp).

**Themes required:** Houston.shp

**Map A:** Adult homicide rate with Moran's "I" and Geary's "C" values

**Map B:** Juvenile homicide rate with Moran's "I" and Geary's "C" values

Compare your results to the correct analysis on the textbook webpage

[www.mappingcrime.org](http://www.mappingcrime.org)

### **Review Questions**

1. Define what the mean center is and list its benefits as well as its associated problems.
2. List the three different types of standard deviation analysis and discuss their differences.
3. Define spatial autocorrelation and discuss the difference between positive and negative spatial autocorrelation.
4. How do the uses for each of the spatial analysis techniques differ?

### **Further Readings:**

1. Cliff, A. and J. Ord (1973). *Spatial Autocorrelation*. Pion: London.
2. Cressie, N. (1991). *Statistics for Spatial Data*. New York: J. Wiley and Sons, Inc.
3. Ebdon, David (1988). *Statistics in Geography*. Blackwell: Oxford.
4. Langworthy, R. H. and E. Jefferis (1998). "The utility of standard deviational ellipses for project evaluation". Discussion paper, National Institute of Justice: Washington, DC.
5. LeBeau, James L. (1987). "The methods and measures of centrography and the spatial dynamics of rape", *Journal of Quantitative Criminology*, 3 (2), 125-141
6. Levine, N. (2002). *Crimestat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v2.0)*. Ned Levine and Associates, Houston, TX and

the National Institute of Justice, Washington, DC. Available on-line at:

<http://www.icpsr.umich.edu/NACJD/crimestat.html>

7. Stephenson, L.K. (1980). Centographic Analysis of Crime. In Georges-Abeyie, D. E., and Harries, K. (eds.), *Crime a Spatial Perspective*, Columbia University Press, New York, pp. 146-155.

## **Chapter 8: Distance Analysis**

### **Goals of Chapter 8:**

1. Understand distance analysis and its importance to crime mapping.
2. Understand Nearest Neighbor Analysis
3. Understand Edge Effects
4. Understand K-Order Nearest Neighbor Analysis
5. Perform nearest neighbor analysis exercises.

### ***Introduction***

The final analysis technique that will be discussed in the basic spatial analysis section is distance analysis. Distance analysis is in a class of techniques called second order techniques. Whereas first order statistics such as the mean center and standard deviations informed users as to the shape, orientation, and dispersion of a distribution of crime incidents, second order statistics provide information on the clustering of crime distributions. Specifically, second order statistics provide information about whether or not there are areas within an overall distribution of crime incidents that are more concentrated or clustered than other parts of the distribution (Levine, 2002). These measures are most commonly used in crime mapping to compare the clustering of the distributions of two different crime types, compare distributions of the same crime at two separate time periods, or as a precursor to hot spot analysis.



### *Nearest Neighbor Index (Nna)*

One of the most popular and easier to use distance analysis measures is the **Nearest Neighbor Index (Nna)**. The Nna compares the distance between actual crime locations and the distances of a random distribution to determine if the incidents are closer than would be expected by chance (Levine, 2002). The actual index is a comparison of the average distance between a crime incident and its closest neighbor, with a distance between two incidents that would be expected by chance (Levine, 2002). When an Nna is conducted it does not produce any output that can be plotted on a GIS, but rather produces an index value based around a value of 1.0. Those crime distributions where incidents are clustered more than would be expected by chance have Nna index values less than 1.0, with values closer to 0 showing increased clustering. Conversely, those crime distributions that are spatially dispersed or non-clustered would have Nna values above 1.0. Importantly, Nna analysis alone cannot show you which areas of the distribution may be clustered, only that there are areas where incidents are closer than would be expected by chance. As discussed above, Nna analysis is primarily useful in determining the difference in the clustering of two different crime types or the difference in clustering of the same crime at two separate time periods.

For example, a crime analyst wanting to determine if assaults are more clustered in January than in February could use Nna to give a measure of both crime distributions. If the January assaults have an Nna value of .3584 and the February assaults have an Nna value of .5691 then it would appear that January assaults are more concentrated than February assaults and are thus more likely to have “hot spots”.

While Nna analysis can be useful for understanding the amount of clustering within a crime distribution, there are two main problems associated with it. The first problem relates to the method for which nearest neighbors are selected and is termed **edge effects**. In the real world, crime is often committed on the border of artificially created police jurisdiction, yet when it is analyzed rarely is crime data from two or more jurisdictions analyzed. The impact of this is that the nearest neighbor of a crime incident may be across a jurisdictional border where no data exists, causing the Nna program to select another incident from within the study area that may not be as close. Figure 8.1 ([www.mappingcrime.org](http://www.mappingcrime.org)) illustrates this problem of edge effects. This is termed the edge effect in that crime incidents that are near a study areas border may increase the nearest neighbor index value artificially (Levine, 2002). A solution to this problem is to use a modified search area that limits the area searched to just the study area. Crimestat solves this problem by providing two different edge effects search options, which will be discussed more fully in the next section.

The other problem with Nna analysis is that it is only an indicator of first order randomness, or the distance between one crime location and its nearest neighbor. Thus, Nna ignores the possible importance of the distance between a crime location and its second nearest neighbor, and third nearest, etc..or what is called **K-order nearest neighbor**. (Levine, 2002). This is an important issue because it limits the ability of Nna analysis to that of just first order analysis and prevents it from being useful in judging the degree of spatial randomness of the entire distribution. On the other hand K-order nearest neighbor analysis allows for just such a use, providing an indication of the degree of clustering for the entire distribution and in essence a measure of spatial randomness for

a distribution (Levine, 2002). Thus, the main benefit of K-order Nna is that it provides a more complete picture of the entire distribution than a simple Nna analysis. K-order Nna is used in the same manner as a simple Nna, in that it is useful in comparing the distribution of two different crime types or the distributions of the same crime at two different times. As with edge effects, Crimestat allows for the computation of K-order Nna, the interpretation and analysis of which will be discussed more in the next section.

### ***Performing Nearest Neighbor Analysis***

1. Open Crimestat and proceed to the Data setup tab and the primary Field subtab. Navigate to the Chapter 8 folder within the CMLab Folder and add Homicides.shp as the primary file (See Chapter 7 for more assistance).
2. Click on the Spatial Description Tab and the **Distance Analysis** subtab.
3. Put a check mark in the empty box next the **Nearest Neighbor Analysis (Nna)**. Below this box are several other boxes related to both Edge Effects and K-order Nna. For our purposes, Edge Effects will be discussed now, with K-order Nna being discussed later in this section.

**Border Correction:** By default Crimestat does not take into account Edge Effects, thus the None option is selected in the border correction section. However, Crimestat allows the user to choose from two different Edge effects search methods, Rectangular and Circular. (For more information about the exact properties of these two edge effects corrections methods see pages 180-183 of the Crimestat Manual). ***The default uncorrected method is fine for most calculations,*** however if you feel your distribution

may be subject to edge effects, it is suggested that you try computing the Nna values with and without considering Edge Effects.

4. After you have determined whether or not to use any border correction method, click on the Compute button.
5. Crimestat calculates 10 different statistics in performing the Nna analysis, the most important for interpretation is the Nearest Neighbor Index, which is at the bottom of the output.

**Interpreting Output:** For the sample data of Houston Homicides the Nearest Neighbor Index was .43001, indicating that the distribution of homicide incidents in Houston is more clustered than would be expected by chance. Thus, if this analysis was to be used to determine if “hot spot” analysis should be conducted, it should be conducted because it is likely that there are homicide “hot spots” in Houston. Moreover, this Nna result could now be used as a baseline for comparison to other crime distributions or the distribution of homicide at a different time period to determine if it is more or less clustered.

### ***K-Order Nearest neighbor Analysis***

There are two main differences in conducting a K-order Nna and a first order Nna, the number of nearest neighbors to be computed and the output created by the analysis. In order to fully analyze the output of K-order Nna it is recommended that the output be exported to a spreadsheet program and a graph of the data made.

1. With the Homicides.shp as the primary data click on the distance analysis subtab within the spatial description tab.
2. Put a check mark in the empty box next the **Nearest Neighbor Analysis (Nna)**.

3. Under the Nna box is a box that says “number of nearest neighbors to be computed”. Replace the default “1” with “100”. While any number of nearest neighbor analyses can be conducted, Crimestat recommends that no more than 100 be computed. Importantly, anything over “1” nearest neighbor calculation is technically a K-order Nna.
4. Click on the “Save Results to” box and save the output as a .dbf file. Navigate to the Chapter 8 folder within the CMLab folder and name the file korder.
5. Click OK and when you return back to the main distance analysis screen click the compute button.

**Analyzing the output:** The main difference between the output of a Nna analysis and a K-order Nna analysis is that in a K-order analysis you have 100 different Nna index values instead of just one. As you scroll down through the Nna index values you can see that the first Nna index value is identical to that produced in the simple Nna analysis (.43001) and that the other 99 values range from .51840 to around .56120, indicating that there is clustering throughout the entire distribution. In order to make analysis and interpretation easier we are going to import the results into Microsoft Excel and graph the Nna index values.

6. Open a spreadsheet program such as Microsoft Excel. When Excel opens click on the Open File button and navigate to the Chapter 8 folder within the CMLab exercises folder.
7. Change the file type to be opened from “All Microsoft Excel Types” to .dbf files by scrolling down the file type box. Double click on **Nnakorder** file to open the

- K-order output in Excel. You should now see four different columns of data 101 rows in length: ORDER, NNDISTANCE, EXPECTEDNN, and NNINDEX.
8. Highlight the **NNINDEX** column by clicking on the “D” above NNINDEX.
  9. With the NNINDEX column highlighted, click on the **Insert** menu and select **Chart**.
  10. When the Chart Wizard appears, select “Line” under Chart type, which will change the chart sub-type views on the right. Select the chart sub-type from the upper right corner and click Next.
  11. The next window will provide you with a preview of the chart, click Finish to complete the chart.

**Interpretation of Results:** The graph shows a line representing the Nna index values over the 100 nearest neighbors that were computed. Figure 8.2 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows the way the graph should look. The line of Nna values starts out at around .4000 and then immediately jumps up to over .5000 then makes a fairly steep climb to over .5500 where it then smoothes out for the rest of the distribution. Because all of the values are below 1.0 it indicates that there is clustering throughout the entire distribution. Thus, the graph reinforces the analysis that the distribution of homicides exhibits spatial clustering on a whole.

### **Exercises for Chapter 8**

1. **Simple Nna:** Conduct two different Nna analyses to determine if there are differences in the clustering of homicides between 1989 and 1991.

**Themes required:** Houston.shp, 1989.shp, 1991.shp

Compare your results to the correct analysis on the Textbook webpage

[www.mappingcrime.org](http://www.mappingcrime.org)

2. **K-order Nna:** Conduct two different K-order Nna analyses to determine if there are differences in the clustering of homicides between 1989 and 1991.

**Themes required:** Houston.shp, 1989.shp, 1991.shp

Compare your results to the correct analysis on the Textbook webpage

([www.mappingcrime.org](http://www.mappingcrime.org))

### Review Questions

1. Explain what distance analysis is
2. How is distance analysis beneficial to crime mapping?
3. What does Nearest neighbor Analysis determine?
4. How is Nearest Neighbor Analysis useful in analyzing crime distributions?
5. What are Edge Effects?
6. What is K-order Nna analysis and how does it differ from a simple Nna analysis.

### Further Readings

1. Clark, P.J. and F.C. Evans (1954). "Distance to nearest neighbor as a measure of spatial relationships in populations". *Ecology*, 35, 445-453.
2. Cressie, N. (1991). *Statistics for Spatial Data*. New York: J. Wiley and Sons, Inc.
3. Ebdon, David (1988). *Statistics in Geography*. Blackwell: Oxford.

4. Levine, N. (2002). Crimestat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v2.0). Ned Levine and Associates, Houston, TX and the National Institute of Justice, Washington, DC. Available on-line at:  
<http://www.icpsr.umich.edu/NACJD/crimestat.html>



### **Section III: Advanced Spatial Analysis**

The final section of the workbook is devoted to providing the student with an introduction to more advanced spatial analysis techniques. The techniques discussed in Section II provided information on the shape, orientation, and dispersion of distributions as well as the general clustering of crime distributions. While this information is of primary importance in assessing spatial patterns of crime, in many situations we need to see where clusters of incidents are occurring. It is this clustering of incidents that Section III will primarily focus on. Specifically, the chapters in this section will provide students with an understanding of the primary techniques used in determining where high concentrations of crime are occurring and if these areas are experiencing change over time. Chapter 9 discusses “Hot Spot” analysis in general covering the main issues involved in this type of analysis as well as several major techniques used for determining “Hot Spot” locations. Chapter 10 introduces students to density mapping and kernel density interpolation, one of the best methods for determining high concentrations of crime. Finally, Chapter 11 discusses spatial dispersion mapping and dual kernel density analysis, techniques used for assessing changes in spatial patterns of crime over time. As mentioned several times throughout the text, the workbook is not designed to be a comprehensive spatial statistics textbook as such an undertaking is far beyond the scope and purpose of this book. Rather, the workbook is designed to introduce students to spatial analysis techniques, their importance to crime analysis, when to use them, and how to interpret their results. For more comprehensive and technical discussions of the

techniques discussed in these and subsequent chapters students should pursue the readings in the “Further Readings” section of each chapter.

## Chapter 9: Hot Spot Analysis

### Goals of Chapter 9:

1. Understand “Hot Spot” analysis and its importance to crime mapping.
2. Understand major issues surrounding “Hot Spot” analysis
3. Know the different methods of “Hot Spot” analysis
4. Understand the Fuzzy Mode analysis technique
5. Understand the Nearest Neighbor Hierarchical analysis technique
6. Understand the Risk Adjusted NNH technique
7. Perform “Hot Spot” analysis exercises

### *Introduction*

One of the more popular current trends in policing is the use of hot spot analysis to assist in management decisions. Specifically, hot spot analysis has become a vital aspect of Problem Oriented Policing strategies and is used by many police agencies to help direct increasingly scarce resources and prioritize patrol function so as to target areas of increased need. Other uses of hot spot analysis included determining the differences in concentrations of different crime types and changes in concentration of a crime at different points in time. Despite increased use by police agencies in designing patrol strategies and resource allocation, much confusion exists as to what a hot spot really is. Harries (1999) defines a hot spot as “..a condition indicating some form of clustering in a spatial distribution.” In contrast, Sherman (1995) defines hot spots as “

small places in which the occurrence of crime is so frequent that it is highly predictable, at least over a 1-year period.” A more statistical definition of hot spots comes from Levine (2002), who states that hot spots are groups of crime incidents that are spatially closer than would be expected by chance. While there is no widely accepted definition of a hot spot, most crime analysts and academics would agree that hot spots are areas where crime is more concentrated than in the rest of a crime distribution. These hot spots can be either places (building or bar location) or spaces (census tract or neighborhood) and are usually consistent over some period of time (Block and Block, 1995).

While confusion is rampant about the theoretical definition of a hot spot, even more confusion exists as to the practical definition of a hot spot, or how to define a hot spot for analysis purposes. Much of the confusion surrounding the practical definition of hot spots revolves around the impact of time, scale, number of crimes, and risk in determining if a hot spot actually exists. **Time** is a key element in hot spot determination as a unit of analysis. Specifically, how long of a time period will you analyze to determine if an area is a hot spot and how long must an area be hot to be considered a hot spot? Is a bar that has a brawl resulting in 12 arrests a hot spot despite the fact that there has never been any arrest at the bar before or after the brawl, or must there be sustained problems at the location over a longer period of time for it to be a hot spot? Moreover what period of analysis do you use in determining hot spots, one day of data, one week, one month, one year? The length of time that is analyzed will significantly impact the determination of a hot spot. As the bar example from above shows, if you were to use one day or one week of crime data the bar would probably be a hot spot, but if you were to use a month or a years worth of data it may not be. Most jurisdictions require an area

to be an active problem area for at least 2 weeks and sometimes longer before it is considered a hot spot. As for the length of time used for analysis, jurisdictions vary greatly and most are impacted in part by the amount of crime they have. Those with more crime use shorter periods of analysis than those with smaller crime amounts. As there is no standard procedure for determining a hot spot, the best suggestion is to use your own knowledge and experience with the crime distributions you area analyzing to assist in time related issues.

A second major issue in the determination of hot spots is the impact of **scale** on hot spot identification. As figure 9.1 shows, when scale is smallest (county level) all crime incidents appear to be a hot spot, yet as scale increases (individual address) crimes become more dispersed and hot spots dissolve ([www.mappingcrime.org](http://www.mappingcrime.org)). Many jurisdictions use the neighborhood or block level to determine hot spots. In this scheme if a given number of crimes occur within a one or two block area they will consider it a hot spot, but if it is more spread out they will not. While this method may work fine for large cities, rural areas, small towns and cities that are more spread out will have to modify their scale or search area for hot spot determination. As with the impact of time on hot spot determination, the best suggestion is to use local knowledge and experience to guide the determination of the appropriate scale for analysis.

A third key element in hot spot determination is the **number of crimes** necessary for an area to be considered a hot spot. As with time and scale, number of crimes can have a significant impact on the determination of hot spots especially when considering jurisdiction size and crime frequency. With crime frequency being much lower and search areas much smaller in small and rural jurisdictions than in large urban

jurisdictions, the threshold for an area to be considered a hot spot is also much lower. While in large cities a minimum of 10 incidents for a hot spot may be appropriate, it is too high a minimum number for most small and rural jurisdictions. As Paulsen (2001) demonstrated, 10 crime incidents represented approximately 24% of the average weekly number of crime incidents in a small town, far more than is necessary for an area to be considered a hot spot. In contrast, 10 incidents represented only around 3% of all weekly crimes in a larger jurisdiction, a more reasonable percentage of total crimes for a hot spot. Thus, when determining the appropriate number of crimes necessary for a hot spot, individual jurisdictions need to analyze their crime distributions and establish minimum amounts that will show true clusters of crime.

The final key issue in determining hot spots is the ignorance of **risk** when determining hot spots. Hot spots are generally determined based solely on the volume of crime incidents in a geographical area, ignoring the importance of the relative number of victims/targets in those areas (Levine, 2002). High volume crime areas, especially in large cities, are often a function of high population densities, meaning these areas may not really be experiencing an increased risk of crime. High volume areas of crime may be high simply because there are more people living/working in those areas, and in fact they may be less risky than other areas with fewer crimes but also less people. The danger in using measures based solely on volume, is that police may focus interventions in areas where crime volume is high but crime risk isn't, leaving areas with a higher crime risk devoid of police interventions. Importantly, one of the methods that will be discussed in this section is a risk-adjusted measure for determining hot spots of crime.

Overall in coming up with a practical definition for hot spot determination the best suggestion is presented by Harries (1999) who suggests that “jurisdiction specific procedures to define hot spots make the most sense because they will fit local conditions”. Importantly, in creating these local specifications, practitioners and researchers *must* be careful to consider the impact that time, scale, crime amount, and risk have on their creation of hot spots. A sound definition will factor in all of these key elements along with fitting local crime distributions in order to maximize valid hot spot identification.

### ***Hot Spot Methods***

Almost as problematic as the confusion over the definition of a hot spot is the number of methods used to determine if and where hot spots exist in spatial distributions. While an exhaustive discussion of the various methods used for hot spot determination is beyond the scope of this book, most methods can be grouped into one of five categories. Research conducted by the Crime Mapping Research Center found that hot spot analysis techniques fall into five basic categories: visual interpretation, choropleth mapping, grid cell analysis, point pattern analysis, and spatial autocorrelation (Jefferis, 1999). Of these five methods the most simplistic is the **visual** method. The visual method involves officers/crime analysts simply looking at point maps and determining areas that appear to be hot spots. While this is obviously the cheapest and quickest method, problems exist such as subjective interpretation by analysts and overlapping of points both of which lead to problems with validity of hot spot locations (Jefferis, 1999).

The second method commonly used is termed **choropleth mapping** and involves aggregating crime data to beats or census tracts and creating color shaded maps based on either crime frequency or crime rates (see Chapter 2). While this method can be useful for showing areas where crime is high in the aggregate, many of the problems discussed in Chapter 2 regarding the making of choropleth maps will impact its utility for hot spot analysis.

The third method for determining hot spots is termed **grid cell analysis**. Grid cell analysis works by placing a grid of equal size cells over the study area and aggregating the crime locations within each cell with cluster analysis then being performed on the aggregated cell counts (Jefferis, 1999). This type of analysis is often termed density mapping and will be discussed more in-depth in the next chapter on density maps and kernel density interpolation.

The fourth method commonly used for determining hot spots of crime is **spatial autocorrelation** and is generally conducted on crime data aggregated to block groups. Anselin and Bera (1996) define spatial autocorrelation as “the coincidence of value similarity with locational similarity”. In other words, areas of similar crime totals will be clustered together spatially. The main difference between this type of analysis and choropleth mapping is that clustering is determined by statistical means such as Anselin’s Local Moran statistic (See Crimestat manual for further discussion of Anselin’s Local Moran).

The final method used for hot spot analysis is termed **point pattern analysis** and is one of the more common statistical methods used for determining hot spots (Mamalian, C. and N. La Vigne, 1999). Point pattern analysis methods generally involve the analysis



of the proximity of points from an arbitrarily placed seed point (Harries, 1999). These techniques seek to determine if crime locations are closer together than would be expected by chance. Numerous different programs exist for analyzing point patterns including STAC, SaTScan, Geographic Analysis Machine (GAM), and of course Crimestat, which includes 6 different point pattern techniques. Because of their combination of popularity, statistical inference, and visual output the discussion of hot spot techniques will focus on these point pattern techniques. Specifically, we will focus on the fuzzy mode, Nearest Neighbor Hierarchical clustering (NNH), and Risk adjusted NNH methods.

### ***Fuzzy Mode***

Within point pattern analysis there are several different types of methods, the simplest and most intuitive is the point location method, of which the fuzzy mode is one type (Levine, 2002). Point location analysis works by counting the number of incidents occurring at different locations; those locations with the most incidents are defined as hot spots (Levine, 2002). Thus, if looking at a distribution of robberies, those locations where multiple incidents occurred will be defined as hot spots. The major problem with standard point location methods, such as the mode, is that they are very limited in the hot spots they find. In particular, hot spots are really hot locations or specific locations with more than one incident. While this is useful in many situations, for certain crimes types few incidents occur at the same location, but they do occur very near each other. An example would be burglaries. If all six houses in a neighborhood were burglarized in the same week point location analysis would not classify this as a hot spot despite it being a

higher than average incidence of burglary. In order to deal with some of the problems of the mode and other point location methods the **fuzzy mode** was created. The fuzzy mode is a variation of the point location method in that it allows “users to define a small search area around each location to include events that occur around or near that location” (Levine, 2002). Thus, the user can define a search radius of 100 meters (or whatever they choose) and the routine will find all incidents that occurred at each location *and* within 100-meter radius of each location. In using the fuzzy mode the user must be careful how they select the search radius, as increasing it too much (1 mile) will make many hot spots because of points being counted multiple times.

### ***Performing a Fuzzy Mode Analysis***

1. Open Crimestat and proceed to the Data setup tab and the primary Field subtab. Navigate to the Chapter 9 folder within the CMLab Folder and add Homicides.shp as the primary file (See Chapter 7 for more assistance).
2. Click on the Spatial Description Tab and the “**Hot Spot**” **Analysis I** subtab.
3. Put a check mark in the empty box next to **Fuzzy Mode (F-Mode)**. Below this box are two other boxes out to the right of the word Radius that are used for defining the search radius used in the fuzzy mode.
4. **Distance Box**: The first box is the distance box and is used for defining the distance of the search radius. For this discussion, type 100 in this box. Importantly, you can select any number and distance unit you like but the larger the search radius the less utility the fuzzy mode has.
5. **Distance Units Box**: The second box is the distance units box and allows the user to select the distance units that they wish to use for the search radius. The

- choices include miles, nautical miles, feet, kilometers, and meters. For this discussion select meters.
6. After you have set the search radius parameters click on the “Save result to” box. Save the output as a .dbf file (your only choice) and name the file fuzzy and save it in the Chapter9 folder. Click Ok when finished.
  7. When you are done click Compute. Crimestat will create a .dbf file that contains four different variables:
    - a. **Rank:** This is the rank order of the hot spot with those with the most incidents being ranked first followed by the second ranked location, etc.
    - b. **Hits:** This variable tells you the number of incidents contained at each location.
    - c. **X:** The X coordinate of the hot spot
    - d. **Y:** The Y coordinate of the hot spot.
  8. **Adding Output to ArcMap:** Open ArcMap and add Houston.shp to the data view from the Chapter 9 folder.
  9. Under the Tools menu select “Add XY Data”
  10. When the add XY Data menu opens click on the Open file button to the right of the blank box at the top. Navigate to the Chapter 9 folder within the CMLab Exercises folder (where you saved the Fuzzy analysis) and add the “Fmodefuzzy.dbf” file.
  11. When you return to the Add XY Data menu make sure that X is in the X Field and Y is in the Y Field boxes and then click OK.

12. You should now see approximately 469 points added to the view. In order to tell which are the hot spots we need to change these points to graduated symbols.
13. Right click on Fmodefuzzy Events and scroll down and select “Properties”
14. Click on the “Symbology” tab and then choose Quantities from the categories on the left hand side. Under Quantities select “Graduated Symbols”. In the Fields Value menu scroll down and select Freq. This should make five classes of symbols of varying sizes (4-12 point sizes). Click apply and you will now be able to see the hot spots on the map. Figure 9.2 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows you what your map should look like upon completing the analysis.

**MAP TIP:** In order to make the hot spots stand out more, change the colors of each different category so that they are all different colors.

**Interpretation of Output:** Interpreting the output for the fuzzy mode, and most point pattern analysis, is very simple and straightforward. Those circles that are largest are the areas that have the highest concentration of crime. In this case they represent the specific locations and areas within 100 meters that have the highest incidence of homicide in Houston. On this particular map the two top hot spots are in the Southwest area of Houston (see Figure 9.2).

### ***Nearest Neighbor Hierarchical Clustering (Nnh)***

The second point pattern hot spot analysis method we will discuss is Nearest Neighbor Hierarchical Clustering (Nnh), which is a type of hierarchical clustering method (Levine, 2002). Hierarchical clustering methods work by grouping incidents together based on some criteria, and then grouping them again and again into second and

third order groups until all the points are grouped together or the grouping criteria fails (Levine, 2002). In the case of Nearest Neighbor Hierarchical clustering, there are two criteria for grouping crime incidents. The first criterion is the nearest neighbor distance, which is set by using a confidence interval. This criteria works by the user selecting a desired confidence interval (.10, .05, .01. etc..) to ensure the points are not grouped by chances. Thus, a user selects a desired confidence level in order to ensure that points are not grouped together based on chance, with this confidence interval determining the size of the search radius. The smaller the confidence interval (.01, .001) the smaller the search radius used for the clustering and the lower probability that crime incidents are grouped by chance. The second criterion is the minimum number of crime incidents necessary for the formation of a hot spot. Here the user defines the minimum number of points necessary for a hot spot to be found. Thus in Nnh clustering, hot spots are created when a minimum number of points cluster together at a better rate than the user defined criteria for chance. For example, if the confidence interval was .01 and the minimum points was 10, then any hot spots that were found would indicate that we are 99% sure that 10 incidents were clustered together not by chance but by some real spatial association.

This method has both advantages and disadvantages as compared to other point pattern methods. In terms of advantages, Nnh is able to determine small concentrations of incidents based on the ability to control the confidence interval level. Smaller confidence intervals (.01, .001, etc..) create smaller grouping areas and are thus able to more clearly cluster small concentrations of crime (Levine, 2002). Second, groupings of hot spots can be found through the second and third order clustering that occurs in Nnh

(Levine, 2002). In reality hot spots are often close together in space, with Nnh analysis clusters of hot spots can be found indicating not only small areas of crime concentration but more global indicators of crime concentrations. In terms of disadvantages one of the biggest problem with Nnh analysis is it is subject to manipulation by users through the selection of confidence intervals and minimum points (Levine, 2002). When users alter these two criteria very different results can be created, leading to inconsistent results in analysis.

### *Performing Nnh Analysis*

1. Open Crimestat and proceed to the Data setup tab and the primary Field subtab. Navigate to the Chapter 9 folder within the CMLab Folder and add Homicides.shp as the primary file (See Chapter 7 for more assistance).
2. Click on the Spatial Description Tab and the “**Hot Spot**” Analysis I subtab.
3. Put a check mark in the empty box next to **Nearest Neighbor Hierarchical Spatial Clustering (Nnh)**. Below are several boxes related to setting up the search radius (confidence interval) and the minimum points per cluster.
4. **Setting the Confidence Interval:** Setting the confidence is vitally important to Nnh analysis because it will determine the size of the search radius. As the confidence interval gets smaller (sliding the arrow to the left toward .00001) the search radius gets smaller and the chance of grouping crime incidents together by chance gets smaller. However, along with the chance of grouping points by chance getting smaller, the chance of finding real hot spots decreases with a smaller search radius. Moreover, the size of the hot spots that are found will also decrease in size with the selection of smaller confidence intervals. Thus, users

should think carefully of the consequences before choosing the confidence interval. For the purposes of **this exercise**, select the .05 confidence interval (fifth notch from the left).

5. **Setting Minimum Points:** As with setting the confidence interval, selecting the minimum points per cluster is vitally important to your final output. As discussed earlier in this chapter, the number of crimes in a hot spot is dependent on several factors such as crime frequency and jurisdiction size. Ultimately, the decision on the minimum number of points per cluster is one that is based on local judgment and knowledge of the data being analyzed. It is important to remember that the goal of this type of analysis is to create hot spots that are meaningful in size and thus not simply two incidents that are close together. Levine (2002) suggests that users start with the default setting (10 points) and alter the setting if there appears to be too many or too few clusters. For the purpose of this exercise leave the minimum number of points at the default 10 incidents.
6. Click on the “Save Results to” box and navigate to Chapter 9 folder. Save the results as a .shp file and name the results homicide. Click Ok when you are done.
7. When you return to the Hot Spot Analysis I screen click Compute.
8. When the analysis is finished scroll through the results to see how many, if any, hot spots were found and what order they are. In this case 7 hot spots were generated, six first order clusters and 1 second order hot spot. In looking at the results of the individual hot spots created, if you scroll to the right you can see how many crime incidents are within each hot spot.

9. **Adding the Output to ArcMap:** Open ArcMap and add the following layers from the Chapter 9 folder.

- a. Houston.shp
- b. Nnh1homicide.shp: First order hot spots
- c. Nnh2homicide.shp: Second order hot spot.

**MAP TIP:** In order to make the hot spots more appealing visually follow the following steps.

1. Click on the desired hot spot layer symbol to open up the Symbol Selector editor.
2. Change the fill color to “No Color. This will make the symbol empty and colorless on the view.
3. Change the Outline Width to 2.0. This will increase the size of the lines around the hot spots.
4. Change the Outline Color to red. It is important to choose a color which will stand out, such as red to make interpretation easier.
5. When you are done with your changes click OK.

**Interpreting the Output:** Figure 9.3 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows you what your map should look like. As with Point location analysis results, hot spots created by Nnh analysis is fairly direct and easy to interpret. In the exercise at hand there are six first order hot spots (clusters of crime incidents) that are centered mostly around the city center, with the exception of one hot spot in the southwest area of Houston. There is also a second order hot spot which encompasses parts of five of the hot spots in the city center



area, indicating that there are not only individual concentrations of crime in this area, but that it has a generally higher concentration of crime than the rest of Houston.

### ***Risk Adjusted Nnh Analysis (Rnnh)***

While Nnh analysis is an improvement over the Fuzzy mode method of hot spot analysis, it still suffers from the ignorance of risk in its creation of hot spots. As discussed earlier in the chapter, Nnh analysis is a hot spot method that creates hot spots based on volume only and not on any actual increased risk of victimization. Thus, Nnh hot spots may be a function of high population density and may not represent the most criminally risky areas of a city, just the area of a city where most of the activity, both criminal and non-criminal occurs. While the volume method of hot spot creation is fine for many police patrol practices, dealing with increased risk of victimization is of primary concern in many police interventions. Thus, Rnnh is an important analysis for understanding the relative volume of crime incident hot spots in relation to a baseline variable such as the underlying population of a city. Risk adjusted Nnh works by combining a kernel density interpolation (to be discussed in-depth in Chapter 10) with an Nnh analysis to produce clusters of points that are closer than would be expected on the basis of a baseline variable such as population or total crimes (Levine, 2002). It creates these risk-adjusted clusters by adjusting the search radius based on the distribution of the baseline variable (Levine, 2002). Thus, when looking at robbery hot spots with the baseline variable being population, Rnnh would cluster robbery locations that are closer together than would be expected based on the underlying population where the incidents are occurring. In a sense Rnnh hot spots are like hot spot rates in that the baseline

variable controls for the differences in population where the incidents occur to produce a more accurate depiction of high crime concentrations.

### ***Performing Rnnh Analysis***

1. Open Crimestat and proceed to set up the primary file just as you did for the Nnh analysis, *including checking the Nnh box for analysis and setting the preferences the same as for the Nnh analysis.*
2. In addition to setting up the primary file, a secondary file and a reference file must be set up for an Rnnh analysis to work. Thus click on the Data Setup tab and the **Secondary File** subtab.
3. For the secondary file navigate to the Chapter 9 folder and select the Houston.shp file. Set the X and Y variables as you did the primary file and for the **Intensity** variable scroll down and select population. This intensity variable is what will serve as the baseline variable for the Rnnh analysis.
4. Click on the Reference file and type the following coordinates in the Grid Area boxes.

	<u>X</u>	<u>Y</u>
Lower Left	-94.91	30.20
Upper Right	-95.95	29.51

5. When you are finished typing in the coordinates click on the save button and name the grid file "**Houston**". We are saving this file for later use in Chapter 10 and Chapter 11 where we will use grid cell analysis extensively.

6. Return to the Hot Spot Analysis I subtab under the Spatial Description Tab and place a check mark in the box next to **Risk-adjusted**. This makes the Nnh analysis a Risk-adjusted analysis instead.
7. Click on the **Risk Parameters** box and put a check in the box next to Use Intensity variable. This tells Crimestat which variable to use as the baseline variable in the analysis. Leave all of the other settings at the default. (For a more complete discussion of these different settings see Chapter 10).
8. Click Ok on the settings of the risk parameters box and then click on the “Save Results to” button when you return to the main Hot Spot Analysis I menu. Save the file as a .shp file in the Chapter 9 folder and name it **risk**. Click Ok and then click Compute at the main menu
9. When Crimestat has finished computing the Rnnh scroll through the results to see the number of hot spots that were created. Notice that **no hot spots were created with the preferences set as they were**. This means that there were no homicide hot spots of when clustering incident locations on the basis of the baseline variable of population.
10. Reset the preferences so that the minimum number of points per cluster is 5 and then click Compute. After Crimestat is finished you will see that one hot spot has been found this time with six incidents.
11. Open the output in ArcMap as you did the Nnh results. The prefix for Rnnh output is Rnnh. Figure 9.4 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows you what your map should look like for the Rnnh analysis.

**Interpretation of Results:** As Figure 9.4 illustrates, only one hot spot was found using the Rnnh method of hot spot analysis. The small single hot spot is in the southwest corner of Huston, contained within the lone hot spot in the southwest part of Houston from the Nnh analysis. The results indicate that there is only one area where homicides are more concentrated than would be expected on the basis of the population distribution.

## **Chapter 9 Exercises**

1. **Fuzzy Mode Hot Spot Analysis:** Conduct two different fuzzy mode analyses to determine if there are differences in the clustering of homicides between 1989 and 1991. Use 100 meters as the search radius.

**Layers required:** Houston.shp, 1989.shp, 1991.shp

Compare your results to the correct analysis on the Textbook webpage  
[www.mappingcrime.org](http://www.mappingcrime.org)

2. **Nnh Analysis:** Conduct two different Nnh hot spot analyses to determine if there are differences in the clustering of homicides between 1989 and 1991. Use the following settings. Try adjusting the minimum number of points to both higher and lower numbers and see how it impacts the number of hot spots created.

**Search radius: .05 level**

**Minimum Points: 10**

**Layers required:** Houston.shp, 1989.shp, 1991.shp

Compare your results to the correct analysis on the Textbook webpage  
[www.mappingcrime.org](http://www.mappingcrime.org)

3. **Rnnh Hot Spot Analysis:** Conduct two different Rnnh hot spot analyses to determine if there are differences in the clustering of homicides between 1989 and 1991.

**Search Radius:** .05 level

**Minimum Points:** 5

**Risk Parameters:** Select use intensity variable and use default for other settings.

**Layers required:** Houston.shp, 1989.shp, 1991.shp

Compare your results to the correct analysis on the Textbook webpage [mappingcrime.org](http://mappingcrime.org)

4. **Hot Spot Comparison:** Compare the results of the hot spot analysis for the 1989 data using the three different techniques to determine how much difference there is in how they create hot spots.

Compare your results to the correct analysis on the Textbook webpage

[www.mappingcrime.org](http://www.mappingcrime.org)

### **Review Questions**

1. Define what a Hot Spot is.
2. What are some of the major issues in determining where hot spots are occurring?
3. List the main ways in which Hot Spots are determined and some of the associated problems with each.
4. What is the best way to determine hot spots?
5. List the different methods for determining hot spots from point data.
6. Briefly describe the way in which Point Locations methods determine hot spot locations.

7. Compare Fuzzy mode analysis with Nnh analysis in how they create hot spots.
8. How does a risk adjusted Nnh analysis differ from a standard Nnh hot spot analysis?

### **Further Readings**

1. Block, C.R. (1994). "STAC hot spot areas: a statistical tool for law enforcement decisions". In *Proceedings of the workshop on Crime Analysis Through Computer Mapping*. Criminal Justice Information Authority: Chicago, IL.
2. Block, R. and C. R. Block (1995). "Space, place and crime: Hot spot areas and hot places of liquor-related crime". In John E. Eck and David Weisburd, *Crime Places in Crime Theory*, Rutgers Crime Prevention Studies Series, Criminal Justice Press: Newark.
3. Buerger, M.E., E.G. Cohn, and A.J. Petronison (1995). Defining the "hot spots" of crime: Operationalizing theoretical concepts for field research. In: J. E. Eck and D. Weisburd, (eds.), *Crime and Place*. Monsey, NY: Criminal Justice Press; and Washington, DC: Police Executive Research Forum, pp. 47-62.
4. Grubestic, T. H. and A. T. Murray (2002). "Detecting hot spots using cluster analysis and GIS". Paper presented at Annual Conference of the Crime Mapping Research Center, Dallas, TX.

<http://www.ojp.usdoj.gov/cmrc/conferences/01conf/Grubestic.doc>

5. Harries, K. (1999). Mapping Crime: Principle and Practice.  
<http://www.ncjrs.org/html/nij/mapping/pdf.html>
6. Jefferis, E. (1998). "A multi-method exploration of crime hot spots". Crime Mapping Research Center, National Institute of Justice: Washington, DC.  
<http://www.ojp.usdoj.gov/cmrc/whatsnew/hotspot/intro.pdf>
7. Knox, E.G. (1988). "Detection of clusters". In P. Elliot (ed), Methodology of Enquiries into Disease Clustering, London School of Hygiene and Tropical Medicine: London.
8. Levine, N. (2002). Crimestat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v2.0). Ned Levine and Associates, Houston, TX and the National Institute of Justice, Washington, DC. Available on-line at:  
<http://www.icpsr.umich.edu/NACJD/crimestat.html>
9. Sherman, L. W. and D. Weisburd (1995). "General deterrent effects of police patrol in crime 'hot spots': A randomized controlled trial". *Justice Quarterly*, 12, (4) 625-648.
10. Weisburd, D. and L. Green (1995). "Policing drug hot spots: The Jersey City drug market analysis experiment". *Justice Quarterly*, 12 (4), 711-735.

## **Chapter 10: Density Mapping**

### **Goals of Chapter 10**

1. Understand Density Mapping analysis and its importance to crime mapping.
2. Understand the Kernel Density Analysis technique
3. Understand the Dual Kernel Density analysis technique
4. Perform Kernel Density Interpolation analysis exercises

### ***Introduction***

An alternative method to “Hot Spot” analysis for determining areas with high concentrations of crime is to create a density map or kernel density interpolation map of crime. Kernel density interpolation is one of the new cutting edge spatial statistical techniques for analyzing point patterns in order to determine high crime areas (Levine, 2002; McLafferty, Williamson, and McGuire, 2000). Whereas hot spot analysis techniques work by clustering incidents together that are closer than would be expected by chance, kernel density interpolation works by using grid analysis to estimate the density of crimes across an entire surface area. The first step in kernel density interpolation is to place a grid with equal sized cells over the entire study area. A circular “kernel” with a constant search radius is placed over the center of each grid cell and crime incidents are weighted based on their distance from the center of this kernel. Those incidents that are farther away from the center are weighted less than those closer to the center. Using this analysis technique crime incidents are weighted within each grid cell



in the entire search area providing a measure of crime density for the entire search area. Thus, every grid cell has a respective crime density measure regardless of whether or not an incident has occurred in the grid cell, providing a smooth and constant picture of crime density for an entire area. The results for each grid cell ( $Z$  value) can then be easily visualized using contour maps, 3-D maps, or surface maps (Levine, 2002).

This method of determining high concentrations of crime holds several advantages over other “hot spot” techniques, the first of which is the constant density map that is created (Levine, 2002; McLafferty, Williamson, and McGuire, 2000). Whereas hot spot analysis techniques such as Nnh provide information about crime incidents themselves (their relative spatial relationship to each other), interpolation provides information about an entire study area. Thus, through interpolation you can get a complete picture of the density of crime for each and every location within a search area, not just a picture of which incidents are most closely related. A second advantage of kernel density interpolation is the output that is generated. In hot spot analysis techniques such as STAC, Nnh, Rnnh, and others, ellipses are drawn around the crimes that have been determined to be closer than would be expected by chance. The problem with this is that few crime distributions are elliptical and thus these hot spots are not the true shape of the high concentration areas but merely convenient output methods. However, with kernel density interpolation high crime areas are represented in their true shape, which is often odd shaped and irregular. Thus, kernel density interpolation can provide a much more accurate visualization of true high concentration areas of crime. A final advantage of kernel density interpolation over other high concentration of crime methods is that it is statistically a better “hot spot” identifier than other techniques

(Levine, 2002). As Levine states (2002), “the kernel estimate..is a continuous surface; the densities are calculated at *all* locations; thus, the user can visually inspect the variability in density and decide what to call a ‘hot spot’ without having to define arbitrarily where to cut-off the ‘hot spot’ zone”. While kernel density interpolations are still subject to the subjective setting of several different parameters, they provide a more objective and statistically valid measure of high concentrations of crime than other measures.

### ***Performing a Single Kernel Density Interpolation***

1. Open Crimestat and proceed to the Data setup tab and the Primary File subtab. Navigate to the Chapter 10 folder within the CMLab Folder and add Homicides.shp as the primary file (See Chapter 7 for more assistance).
2. Select the **Reference File** subtab. If you saved the reference file coordinates during the Rnnh analysis setup then simply click on the **Load** button, select the Houston file and click Load. If you did not save the reference file coordinates from the Rnnh analysis than enter the following coordinates:

	<u>X</u>	<u>Y</u>
Lower Left	-94.91	30.20
Upper Right	-95.95	29.51

3. When you have finished entering the coordinates click on the **Spatial Modeling** tab and then select the **Interpolation** subtab and place a check mark in the box next to the word **Single**.

**Selecting Parameters for Kernel Density Interpolations:** As with Nnh and Rnnh analysis there are several different parameters in conducting a kernel density interpolation that have an important impact on the final output. In the case of an interpolation the parameters of importance are the **interpolation method** and **bandwidth** selection and size.

**Interpolation Methods:** Crimestat contains five different interpolation methods: Normal, uniform, quartic, triangular, and negative exponential. These methods of interpolation are the methods used for creating the kernels used in weighting the crime incidents within each grid cell. As there are advantages to each of these different methods selection of an interpolation method is going to be based on the spread of the crime distribution as well as trial and error. Levine (2002) suggests that the user should start with the default method and then adjust according to how the density surface looks in relation to the actual crime distribution. The following descriptions of these methods are from the Crimestat manual (Levine, 2002)

**Normal:** Weighs all points in the study area, with near points being weighed more highly than distant points. Produces a smoother distribution. This is the default for Crimestat.

**Uniform:** Weighs all points within the circle equally. Produces a smoother distribution.

**Quartic:** Weighs near points more than far points, with a gradual fall off of far points. Produces a smoother distribution.

**Triangular:** Weighs near points more than far points but the fall off is more rapid than with quartic. Produces a spikier distribution by emphasizing the peaks and valleys between crimes.

*Negative Exponential:* Weighs near points much more highly than far points. Produces a spikier distribution by emphasizing the peaks and valleys between crimes.

\*\*\*For the purposes of this exercise we are going to use the default **normal** distribution. As mentioned above, trial and error and analysis of your crime distribution are the best methods for determining the correct interpolation method.

**Bandwidth Selection:** The other major parameter in conducting interpolations is the choice of a bandwidth. The bandwidth is the search radius that is used in determining how close crime incidents are to the kernel function. In general larger bandwidths will create a smoother distribution whereas smaller bandwidths will create a spikier distribution that emphasizes the peaks and valleys between crime incidents. As with interpolation methods, selecting a bandwidth is not standardized and depends a lot on trial and error and how much emphasis a users wants to place on differences between points. Crimestat contains two different bandwidth choices, fixed and adaptive. The following descriptions of these bandwidths are from the Crimestat manual (Levine, 2002)

**Fixed Interval:** In this method the user selects the exact bandwidth size they want to use in the analysis. If a user chooses to use a fixed interval bandwidth they must select the measurement unit (squared miles, squared nautical miles, squared feet, squared kilometers, and squared meters) as well as actual measurement distance. Once this interval has been chose it will be used in each grid cell.

***Adaptive Bandwidth:*** In contrast to fixed intervals, which use the same search radius throughout the analysis, adaptive bandwidths adjust the radius so that a minimum number of points are found, thus providing constant precision over the entire study area. Thus, in areas with many crime incidents a small bandwidth will be used, whereas in areas with few crimes a large bandwidth will be used. In selecting an adaptive bandwidth the user has the option to change the minimum number of points selected by the bandwidth radius. As with choosing a fixed interval distance, smaller numbers of points will create a spikier distribution and larger numbers will create smoother distributions. Figure 10.1 shows how the changing the minimum number of points effects the distribution, with the 25 point interpolation being spikier than the 100 point interpolation. In Crimestat the default bandwidth is adaptive with a minimum number of points of 100.

\*\*\*For the purposes of this exercise we are going to use the default **adaptive bandwidth** with 100 minimum points selected. As mentioned above, trial and error and analysis of your crime distribution are the best methods for determining the correct bandwidth.

**Output Units:** This parameter allows the user to select the output units for the density estimate. The default is squared miles and in most situations this is the correct output unit to use. For the purposes of this exercise we will use **squared miles**.

**Calculate:** The final parameter that the user must select is how the density will be calculated. The two options that are provided in Crimestat are densities and probabilities.

**Densities** provides a measure of the density of crime over the study area. In contrast

**probabilities** renders the output as a probability that a crime will occur at any one location. Importantly, the selection of this parameter has no impact on the way the distribution will look visually, it only impacts the interpretation.

\*\*For the purposes of this exercise we will select the default **densities**.

4. Once all of the parameters have been selected, click on the **Save Results to** button. Name the file **density** and save the file as a .shp file in the Chapter 10 folder.
5. After you return to the main Interpolation screen click on Compute. After Crimestat has finished computing the Interpolation results close the results window and open ArcMap.
6. After you have created a new view add the interpolation output .shp file from the Chapter 10 folder. The prefix for all single interpolation outputs is K. Thus the output should be kdensity.shp. The output should appear as a big grid that covers the entire Houston area. In order to view the density results we have to change the variable that is being viewed to the **Z value**.
7. Right click on the symbol under the kdensity layer and scroll down and select properties. Select the symbology subtab and then select “Quantities” and under quantities select “Graduated colors”..
8. Under the fields menu scroll down the Value drop down menu and select Z and click OK. This variable contains the values for each grid cell and will reveal the density of crimes for the Houston area.

**MAP TIP:** In order to make the grid lines disappear click on each individual color symbol and change the outline color to :”No color”. After doing this for each different color in the grid it should eliminate the annoying grid look to the results.

**MAP TIP:** In order to see the basemap over the interpolation results drag the basemap (Houston) layer above the interpolation result layer (kdensity) in the table of contents.

**Interpretation of Results:** Figure 10.2 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows what your results should look like for the Kernel Density Interpolation. The darker the color on the map the higher the concentration of crime incidents. Thus, in the example the highest area of homicides is in the city center area. Importantly, the distribution is almost like a bulls eye with lower areas of homicide density radiating out from the city center.

### ***Dual Kernel Density Interpolation***

In addition to a single kernel density interpolation routine there is a second interpolation method called a dual interpolation. **Dual kernel density interpolation** works in the same manner as a single kernel density interpolation, with the exception that in a dual kernel density interpolation a secondary file containing other point data is also used. However, while single kernel density interpolations provide information about the volume of crime incidents, dual kernel density interpolations provide information about crime risk density and changes in crime density. Similar to a Rnnh analysis, a dual kernel density interpolation can be used to create a measure of the density of crimes based on an underlying variable such as population in order to show the risk of that crime occurring. As with Rnnh analysis this type of analysis is very useful in many police interventions as it provides a more accurate measure of the true risk of crime occurrence and not just a

measure of crime volume. The other main use of a dual kernel density interpolation is for measuring changes in crime patterns over time. Importantly this dual kernel density interpolation routine will be discussed in-depth in Chapter 11 on Spatial Dispersion mapping.

### ***Performing a Dual kernel Density Interpolation***

1. Set up the primary and reference file exactly as you did for performing the exercise above for the single kernel density interpolation. As stated above, the vast majority of a dual kernel density interpolation analysis is identical to a single kernel density interpolation.
2. Click on the Secondary File tab. Navigate to the Chapter 10 folder and add the **Houston.shp** file. Set the coordinate values and make “**population**” an intensity value. Population will thus be the variable used to create the risk adjusted density map for homicides showing the density of homicide risk in Houston.
3. Click on the **Spatial Modeling** Tab and select the **Interpolation** subtab.
4. Place a check in the box next to the word **Dual**. By default the primary file will be the first file and the secondary file will be the second file, but this can be changed for analysis purposes.

**Setting of Parameters:** The setting of parameters is identical to that for single kernel density interpolation, with the exception of the Calculation parameter. Thus, the only discussion of setting parameters will be of the calculation parameter. ***For this exercise*** leave the interpolation method and bandwidth as the default measures.



**Setting the Calculation parameter:** As opposed to a single density interpolation where there were only two calculation options, densities and probabilities, a dual kernel density interpolation gives the user four separate options. The different calculations include: ratio of densities, log ratio of densities, differences in densities, and sum of densities.

***Ratio of Densities:*** First file divided by the second file. Can be used to provide an estimate of risk for a crime if the first file contains crime locations and the second file contains census data with population as an intensity variable.

***Log ratio of Densities:*** This calculation takes the natural logarithm of the density ratio. This is used primarily when the majority of study cells have very low densities but a few have very high densities (Levine, 2002). The result will be one in which the spikes will not be as pronounced.

***Difference in Densities:*** The first file minus the second file. This calculation is commonly used for determining differences in density patterns of crime between two separate time periods. The result will be a density map showing where crime has decreased and increased between the two time periods.

***Sum of Densities:*** The density estimate for the first file plus the density estimate for the second file. This is used for combining the density estimates of two separate crime distributions.

\*\* For the purpose of this exercise select ratio of densities, as we want to create a density estimate of homicide risk in Houston.

5. Click on the **Save Results to** button and navigate to the Chapter 10 file. Save the file as a .shp file and name it **dual**.
6. When you return to the main interpolation menu click the Compute button. When Crimestat finishes with the calculations close the results window, minimize Crimestat and open ArcMap.
7. Add the following layers to the view.
  - a. Houston.shp: Basemap of Houston found in the Chapter 10 folder.
  - b. DkDual: dk is the prefix for dual kernel density interpolation output.
8. Make dkdual viewable and follow the same procedure for making the density map as with the single density interpolation map. Figure 10.3 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows you what your map should look like for the Dual kernel Density analysis. Don't forget to use the map tips from above in order to make the density map easier to view.

**Interpreting of Results:** As with a single kernel density interpolation the darker the color the higher the risk of a homicide. Figure 10.3 shows the results of the single kernel density estimate next to the risk estimate created through the dual kernel density estimate. Notice how the area of highest concentration of crime has changed shape, becoming more teardrop shaped. In addition the rest of the distribution has also changed to reflect the impact of population on the risk of homicide victimization.

## Chapter 10 Exercises

1. **Single Kernel Density Interpolation:** Make two different maps of the 1989 homicide distribution with two different bandwidths in order to see how bandwidth impacts density maps.
  - a. Map A: Use an adaptive bandwidth with a sample size of 100
  - b. Map B: Use a fixed interval bandwidth with an interval of 200 meters.

Compare your results to the correct analysis on the Textbook webpage

[www.mappingcrime.org](http://www.mappingcrime.org)

2. **Dual Kernel Density Interpolation:** Make two different risk maps, one for 1989 data and one for 1994 data to assess the difference in homicide risk between the two years. For both maps choose the interpolation and bandwidth method of your choice using your knowledge of the crime distribution and kernel density methods.
  - a. 1989 Map:
  - b. 1994 Map:

## Review Questions

1. How is kernel density interpolation used in crime mapping?
2. Briefly describe the process of how kernel density interpolation is conducted.
3. List and explain the advantages that kernel density interpolation has over other techniques.

4. List the different interpolation methods that can be used in performing a kernel density interpolation and briefly describe the distribution that each creates.
5. Explain the differences in the two bandwidth options in Crimestat. Why would you use one over the other?
6. What are the two ways a dual kernel density interpolation can be used for crime mapping?

### **Further Readings**

1. Farewell, D. (1999). "Specifying the bandwidth function for the kernel; density estimator". <http://www.iph.cam.ac.uk/bugs/documentation/coda/node44.html>
2. Kelsall, J.E. and P.J. Diggle (1995). "Kernel estimation of relative risk", *Bernoulli*, 1, 3-16.
3. Levine, N. (2002). *Crimestat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v2.0)*. Ned Levine and Associates, Houston, TX and the National Institute of Justice, Washington, DC. Available on-line at: <http://www.icpsr.umich.edu/NACJD/crimestat.html>
4. Mc Lafferty, S., D. Williamson, and P.G. Mc Guire (2000). *Identifying Crime Hot Spots Using Kernel Smoothing*. In V. Goldsmith, P.G. McGuire, J.H. Mollenkopf, and T.A. Ross (eds.), *Analyzing Crime Patterns: Frontiers of Practice*. Sage Publications, Thousand Oaks, CA.

## **Chapter 11: Spatial Dispersion Mapping**

### **Goals of Chapter 11**

1. Understand Spatial Dispersion Mapping and its importance to crime mapping
2. Understand Map Calculations with Dual Kernel Density Interpolation
3. Perform Spatial Dispersion Map Exercises

### ***Introduction***

The final chapter in the workbook will discuss one of the newer and more important crime mapping tools, spatial dispersion mapping. Spatial dispersion is important to both crime mapping practitioners and researchers because it lets them assess changes in crime density over time or differences in crime density between two different crime types. By measuring differences in crime density over time users can see how crime distributions are evolving within a jurisdiction, which is particularly useful for redrawing police beats and re-assigning work load. Figure 11.1

([www.mappingcrime.org](http://www.mappingcrime.org)) shows the result of a spatial dispersion analysis of homicides committed in January with those committed in February. The blue areas show where Homicide decreased in February compared to January, while the red areas show where Homicide has increased from January to February. Common time periods for this type of analysis include yearly, seasonally, monthly, and weekly changes in crime densities, as well as analyzing differences in crime patterns based on the time of day of the crime. Spatial dispersion analysis is also beneficial in other types of analysis, particularly when

assessing the impact of police interventions. Using spatial dispersion mapping users can assess how successful an intervention was at reducing crime in a specific area as well as showing areas where the crime may have diffused or spread to because of the intervention. Finally, this type of analysis can also allow users to make comparisons in the crime density patterns of two different crime types, showing where one crime type is denser than another crime type.

The process involved in creating a spatial dispersion map involves conducting a **dual kernel density interpolation** with two different distributions of crime locations. This is the same process that was discussed at length in Chapter 10. Because the dual kernel density interpolation conducted in Chapter 10 was for producing maps of risk, there are some subtle changes that need to occur in order to create a spatial dispersion map. The main differences between using dual kernel density interpolation for producing risk maps and spatial dispersion maps revolves around the data calculations conducted. In using a dual kernel density interpolation to produce a risk map the secondary file must contain census data on population and coordinate information on the median centers of the census data. In contrast, when performing spatial dispersion mapping the secondary file must be point locations for a crime distribution. For example when conducting a spatial dispersion analysis of the change in homicides from January to February, the primary file would contain February homicide data while the secondary file would contain the January homicide data. The second difference when using dual kernel density interpolation for spatial dispersion mapping is the change in the calculation used for the density map. In creating a density risk map the calculation used is the ratio of densities, whereas in spatial dispersion mapping it is the **difference in densities**. The difference in

densities subtracts the second file from the first file. Using the example from above, that would be subtracting the January homicide density from the February homicide density. Importantly, all of the other issues surrounding kernel density interpolations (interpolation method, bandwidth size) still apply to spatial dispersion mapping. Thus, when using dual kernel density interpolation to produce spatial dispersion maps it is suggested that you review the discussion in Chapter 10 concerning these issues.

### ***Performing Spatial Dispersion Mapping using Dual Kernel Density Interpolation***

1. Open Crimestat and click on the Data Setup tab. For the **Primary file** select janjun.shp, which contains all homicides in Houston from January to June. Make sure to enter the X and Y coordinates for this file and leave all other fields blank.
2. Click on the **Secondary File** tab and select the file julydec.shp, which contains all homicides in Houston from July to December.
3. Click on the **Reference File** tab and load the Houston reference file. If you have failed to save the reference coordinates from previous exercises involving kernel density interpolation see Chapter 10 for the correct reference grid coordinates.
4. Click on the **Spatial Modeling** tab and the **Interpolation** subtab. Place a check mark in the box next to the word Dual.
5. **Setting the parameters:** As you have learned in the last two chapters setting the parameters for an interpolation involves much trial and error and a strong knowledge of your crime distribution. While students are encouraged to experiment with using different parameters to conduct interpolations, for the purposes of this exercise use the parameters specified below.

- a. **Interpolation Method:** Use the default Normal.
  - b. **Bandwidth:** Use the default Adaptive with 100 minimum point sample size.
  - c. **Calculation:** Use difference in densities.
6. Click on the Save Results to button and navigate to the Chapter 11 folder. Save the results as a .shp file and name it change. The prefix for the output is dk. Thus the your output file for the spatial dispersion analysis will be **dkchange.shp**.
7. When Crimestat finishes computing the dual kernel density interpolation close the results window and minimize Crimestat.
8. Open ArcMap and add the following layers:
  - a. **Houston.shp:** Houston basemap
  - b. **dkchange.shp:** Output file for change in density of homicide.
9. Make the dkchange theme viewable and open up the Symbology editor as in Chapter 10 in order to add the Z value. **Importantly**, instead of using the usual red monochromatic color scheme select a **Blue to Red color scheme** in the color ramp drop down menu. This will make it easier to interpret the output.  
  
Remember to use the map tips from Chapter 10 to remove the grid lines and overlay the Houston basemap on top of the density output.

**Interpretation of Results:** Figure 11.2 ([www.mappingcrime.org](http://www.mappingcrime.org)) shows the results you should have achieved. The darker blue areas show the areas where homicide decreased the most from the period of January to June to the period of July to December. In other words, in the second half of the year homicide incidents decreased the most in the dark blue areas followed by the lighter blue areas. In contrast the darkest red areas show



where homicide incidents increased the most from the January to June period to the period of July to December. In other words, in the second half of the year homicides increased the most in the dark red areas followed next by the lighter red areas. Finally the white area in between the two color extremes is where homicide stayed fairly stable between the two time periods.

### **Chapter 11 Exercises**

**Spatial Dispersion Map:** Perform a spatial dispersion analysis on the change in homicide density from the period of 1986-89 to the period of 1990-94 to determine if the pattern of homicide changed much. Use all of the default parameters for the analysis.

Compare your results to the correct analysis on the Textbook webpage

[www.mappingcrime.org](http://www.mappingcrime.org)

### **Review Questions**

1. List the different uses for spatial dispersion mapping for both practitioners and researchers.
2. What type of calculation is necessary for conducting a spatial dispersion map using kernel density interpolation?
3. What is the main difference between conducting a dual kernel density interpolation to produce risk maps and one to produce spatial dispersion maps?

## Further Readings

1. Farewell, D. (1999). "Specifying the bandwidth function for the kernel; density estimator". <http://www.iph.cam.ac.uk/bugs/documentation/coda/node44.html>
2. Kelsall, J.E. and P.J. Diggle (1995). "Kernel estimation of relative risk", *Bernoulli*, 1, 3-16.
3. Levine, N. (2002). *Crimestat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v2.0)*. Ned Levine and Associates, Houston, TX and the National Institute of Justice, Washington, DC. Available on-line at: <http://www.icpsr.umich.edu/NACJD/crimestat.html>
4. Mc Lafferty, S., D. Williamson, and P.G. Mc Guire (2000). *Identifying Crime Hot Spots Using Kernel Smoothing*. In V. Goldsmith, P.G. McGuire, J.H. Mollenkopf, and T.A. Ross (eds.), *Analyzing Crime Patterns: Frontiers of Practice*. Sage Publications, Thousand Oaks, CA.

