Identifying Crime Patterns

Though the **crime analysis process** begins with collecting and managing data, actual analysis of a crime pattern begins with its initial identification.

Pattern identification can be an overlooked and underestimated art. Good crime analysis requires:

- That the *analyst* strive to be the one to identify all crime patterns. If the analyst waits for the rest of the department to notify him of the pattern's existence—or waits for the media to notify everyone—then any brilliant analysis that follows is clouded by the fact that the analyst has allowed the pattern to continue long past its **point of identifiability** (see below).

- That the analyst attempt to identify 100% of identifiable patterns, in all crimes and calls for serve that fall within the analyst's scope of responsibility.

- That the analyst identify each pattern once it reaches its **point of identifiability**. The point of identifiability describes the moment, after *n* incidents, when the analyst can confidently assert that a pattern exists. The actual number of incidents can vary greatly depending on the types of crimes and the strengths of their commonalities. Some patterns will be identifiable after the second incidents; others will require a dozen or more. The idea is simply to identify the pattern as soon as possible, so that the agency can intercede as soon as possible, so that as few people as possible need become victims.

Accomplishing all three of these goals requires the analyst to make information scanning a *daily process*. Furthermore, the analyst must ensure that he has *real-time access* to incident reports. If the point of identifiability is reached upon the third incident, reported at midnight, then the analyst should identify it shortly after arriving for work the next morning.

The analyst may determine the possibility of a pattern based on one of three factors:

1. *Modus Operandi Commonalities*, found through a careful review of incident reports and their narratives;
2. *Exceptional Volume*, found through some brand of **threshold analysis**, either deliberate or unconscious;
3. *Geographic Proximity*, found through **crime mapping**.

Each of these three methods is described in turn below. We also speak briefly about intelligence-based pattern identification, identifying hot spots and criminogens, and cross-jurisdictional pattern identification.

Review of Incident Reports

An analyst reviews daily crime reports. Source: author.

Report review should be a daily process. The usual objections to this statement—lack of time or manpower—become moot once the agency adequately staffs its crime analysis unit. This process is the least scientific of all the means to identify patterns, but it is probably the most effective. Many patterns exhibit neither high volume nor geographic clustering, making the other two methods often ineffective.

Each analyst will develop a method of daily report review that works best for him, but here is a typical example:

1. At least once a day[1], the analyst gathers all incident reports since his last review[2]. For most analysts' minds, a physical written report is the best source, but some analysts work for "paperless" agencies and will have to settle for electronic reports. The key is that the analyst has access to all applicable narratives.

2. The analyst conducts an initial scan of the reports, removing incidents that are unlikely to ever exhibit any patternability, or which do not fall within his scope of responsibility.

3. The analyst reads the remaining reports, paying careful attention to the M.O. factors. At this point, the analyst's memory may alert him to patterns—particularly those with strong M.O. commonalities or SIGNATURES.

4. The analyst compares his current reports with past reports received by the agency. It is at this point that the analyst's experience, knowledge, intuition, and creativity most come into play. It is also at this point that the analyst most benefits from a good RECORDS MANAGEMENT SYSTEM or computerized database with strong querying capabilities and the ability to display the results in matrix form.

Comparing current reports with past reports can be more art than science. The scope and complexity of the analyst's search depends on the nature of the crime and the peculiarities of its various factors.

Assume, for instance, that the analyst has before him a report of a commercial burglary committed between 22:00 and 04:00 the previous night. One or more burglars entered a hair salon in the City Square business district, pried open the register, and stole $400 in cash. They also poured bottles of hair conditioner all over the floor and smeared it on the salon's mirrors and windows.

If the analyst works for a small town in which commercial burglaries are rare, he may decide to review all commercial burglaries in the past year and compare the current incident to it. In city in which commercial burglaries are more common, the analyst may try any or all of the following searches, and more:

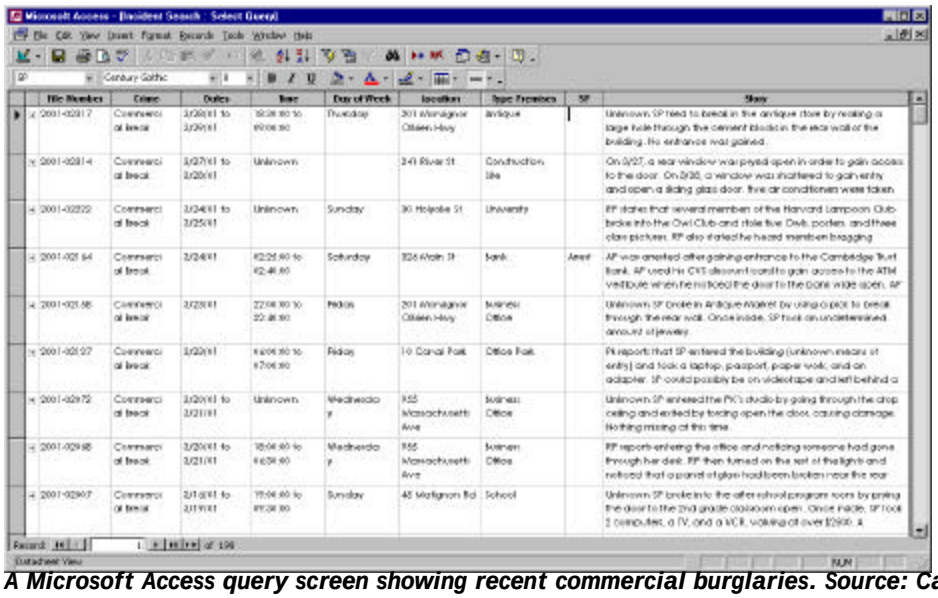
- Commercial burglaries in the City Square district in the past three months
- All commercial burglaries for the past two weeks
- Commercial burglaries in hair salons over the past year
- Commercial burglaries in retail establishments in which cash was stolen, or registers targeted, during the past three months
- Commercial burglaries involving unnecessary vandalism (the crime's signature) over the past three years
- All crimes reported at this location or in this block in the past month

The goal is to broaden the scope of the search as much as possible, without returning more reports than is reasonable for the analyst to review and compare. For crimes with unique signatures (e.g., the burglar writes "Helter Skelter" on the wall with his own blood), it would not be unreasonable for the analyst to search his entire database for similar incidents. For more standard and common crimes (e.g., the theft of a cellular telephone from a car), the analyst does better to limit his search to crimes in the same area that happened fairly recently. The analyst's ability to design effective queries will increase as he better learns both his jurisdiction and the nature of crime committed within it.

---

[1] Some analysts conduct this process two or three times a day, insuring that potential patterns are identified as quickly as possible. Units that are staffed on two or more shifts will also have a greater ability to identify patterns in "real time."

[2] The number of reports the analyst gathers depends on his scope of responsibility, and may range from all calls for service to crimes only to certain crimes only. Analysts in larger departments or in inadequately staffed units will be able to review fewer reports on a daily basis and will have to select a set of TARGET CRIMES.



A Microsoft Access query screen showing recent commercial burglaries. Source: Cambridge (MA) Police Department.

Once the analyst has obtained the results of his query, he then begins to compare the various factors of each past crime with those of the present one, looking for commonalities that would indicate the existence of a crime pattern. Again, there is no scientific formula for the analyst to apply in this process; he must use his own knowledge and experience to decide at which point one or more commonalities indicates the likelihood of a pattern. Such a determination depends, again, on the nature of crime in the analyst's individual jurisdiction. In an area with few commercial burglaries, the analyst may flag a burglary, two months ago, at a restaurant half a mile away in which cash was stolen, while in an area with many burglaries, this may not be enough to indicate a pattern.

Once an analyst believes he has found a pattern, the analysis process begins. Because of the nature of this type of pattern identification, however, much of the analysis work has already been done.

We note that some efforts have been made to automate the review process, automatically scan for similarities among incidents, and identify potential patterns without requiring the analyst to read each report. While sophisticated DATA MINING and NEURAL NETWORK technology is promising, we do not feel that it has developed enough to make the analyst's daily review superfluous, and we know of no police agency that has implemented such an automated process on a regular basis.

Threshold Analysis

Threshold analysis describes the process by which the analyst identifies potential patterns through exceptional volume. The theory behind threshold analysis is that when crime in a particular geographic area reaches a level that is significantly higher than usual, some type of crime pattern is probably to blame.

The analyst can use a statistical method to determine when crime has reached a level that is "significantly higher than usual"—in other words, when crime crosses the *threshold* from average volume to exceptional volume. However, absent any statistical methods, an experienced analyst will employ an almost unconscious method of threshold analysis. An analyst for Gotham City, for instance, might know that auto thefts are very common in the impoverished, urban South Side, but rare on the ritzy North Shore neighborhood. Thus, five thefts in a week in the South Side will raise no eyebrows, but alarm bells ring in the analyst's mind after the same number of thefts on the North Shore. This

informal method, while not scientific, is still perfectly valid since, with either method, the analyst will investigate these crimes further before announcing the existence of any patterns (see below).

It should be noted that many patterns involve only a few incidents and may therefore not set off any threshold alarms. These small patterns can easily be "buried" within high average numbers for a particular area. Because of this, threshold analysis should not be the *only* means by which an analyst seeks crime patterns. The benefit of threshold analysis is that it may catch patterns overlooked in the details of day-to-day report review, and it may benefit understaffed crime analysis units that simply do not have time to effectively use the report review method.

Analysts with many crimes or a large area to review may want to employ a more scientific methodology, which employs the statistical technique known as STANDARD DEVIATION.

To begin this technique, the analyst calculates the average number of *C* crime in *A* areas in *T* time. Like all such queries, the exact scope of this process depends on the analyst's discretion. The analyst may look at robberies in each police sector on a monthly basis, all Part 1 crimes for the entire city on a quarterly basis, total crime in the entire city on a shift basis, and so on. Since setting up the query and running the search may be complex—depending on the analyst's knowledge of querying and the accessibility of his records management system—he should probably settle on some standard that he will use month after month.

Assume that the analyst has decided to study only Part 1 Crimes on a monthly basis[3] in six police sectors. He begins by calculating the monthly averages for the past eight years (the extent of his records management system) and the standard deviation for the range of years. He will unfortunately have to create a spreadsheet for every month for every crime:

Auto Thefts in January

Sector	Average '94-'01	Stand. Dev.	2002	# St. Dev. from Average
1: North Shore	8	1.8		
2: South Side	42	6.4		
3: City Center	20	4.0		
4: Highland Park	17	3.3		
5: Lakeside	12	2.7		
6: West End	33	5.5		

The analyst then calculates the totals for the current year, and notes how many standard deviations away from the average the current year is. By dividing the difference between the average and the current year by the standard deviation.)

Auto Thefts in January

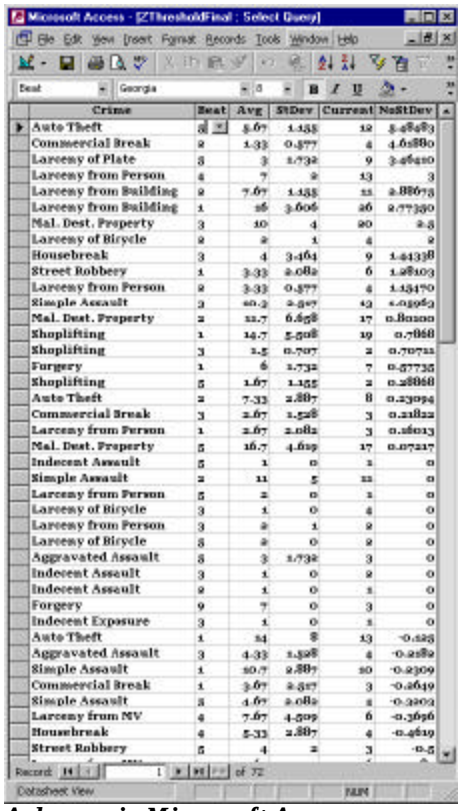
Police Sector	Average '94-'01	Stand. Dev.	2002	# St. Dev. from Average
1: North Shore	8	1.8	7	-0.56
2: South Side	42	6.4	48	+0.94
3: City Center	20	4.0	31	+2.75
4: Highland Park	17	3.3	15	-0.61
5: Lakeside	12	2.7	18	+2.22
6: West End	33	5.5	20	-2.36

[3] Studying totals on a monthly basis may skew results for some crimes, since some years will have more weekend days or other high-volume days than others. We use monthly totals in this example for convenience, but we encourage analysts to pursue alternatives that they think will provide a more accurate comparison.

Remember, the theory of standard deviation tells us that in a normal distribution, about 68% of all variables (in this case, the number of auto thefts in January) fall within one standard deviation of the mean, 95% fall within two standard deviations of the mean, and 99.7% fall within three. Applying this theory to real world examples, the analyst might use the following terminology:

"Cold"	Crime is more than 2 SDs below the mean
"Cool"	Crime is between 1 and 2 SDs below the mean
"Normal"	Crime is between 1 SD below the mean and 1 SD above the mean
"Warm"	Crime is between 1 and 2 SDs above the mean
"Hot"	Crime is more than 2 SDs above the mean

Applying this to our example, we see that auto thefts in January in the North Shore, South Side, and Highland Park neighborhoods are within one standard deviation of the mean (either way) and are thus unexceptional. City Center and Lakeside, however, are both experiencing exceptionally high levels of crime ("hot"), which indicates a possible pattern at work. The West End, meanwhile, is exceptionally cold. Perhaps there is some factor driving auto theft down.



A dynamic Microsoft Access query showing the hottest crimes and beats for any given time period, allowing the analyst to perform threshold analysis daily. Source: Cambridge (MA) Police Department.

High volume, of course, does not necessarily mean that every auto theft in City Center or Lakeside is part of a crime pattern, nor does it necessarily mean that a pattern truly exists. The analyst must now take a closer look at auto thefts in January in City Center and Lakeside to determine if there is any indication of a pattern other than high volume. Only by studying the various factors involved in each crime will the analyst be able to make a confident determination.

The problem with this method is that calculating these totals month after month can be an exhausting process with all but the best records management systems. (Even if the analyst uses a spreadsheet program to perform the actual calculations, he still has to import the data from his RMS into the spreadsheet.) If the analyst decides that he wants to change a factor (for instance, the date range), the totals must be recalculated.

Analysts who know advanced querying techniques—and who can directly query the data in their systems—will be able to design dynamic queries that allow them to specify parameters every time they run the query. Detailing advanced SQL statements is beyond the scope of this article, however.

Analyst who *do* devise some dynamic querying method may theoretically be able to conduct threshold analysis on a daily basis, but the nature of this method of pattern identification suggests a less frequent application: once a week or once a month (for each type of crime) should suffice.

One final word about thresholds: since crime is subject to long-term trends, a strict comparison of averages and standard deviations may not be entirely valid. Consider the following totals for auto thefts in January on the South Side:

Year	Total
1994	50
1995	49
1996	48
1997	44
1998	40
1999	38
2000	36
2001	33
Average	*42*
St. Dev.	*6.4*
2002	**48**

Though the 2002 total of 48 incidents falls comfortably within one standard deviation of the mean, it is notable that this total reverses an eight-year trend of decreases in auto theft in City Center, and it represents a 45% increase from the previous year.

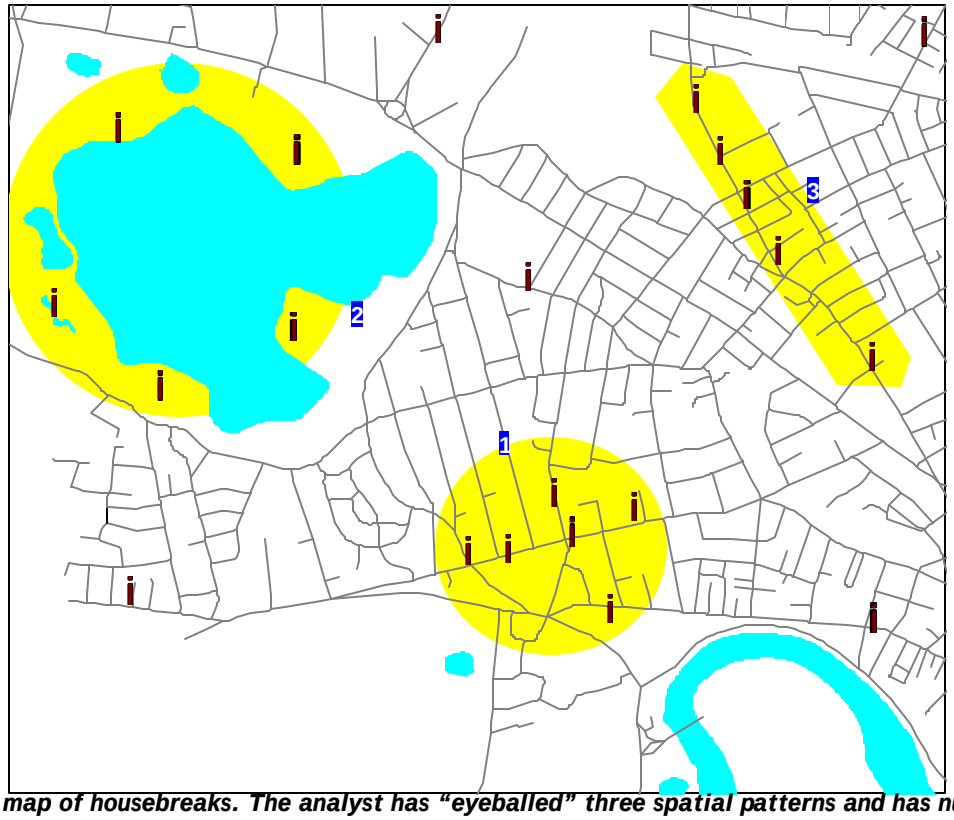
Analysts whose jurisdictions are subject to such trends may find it more valuable to compare averages and standard deviations in *rate of change*, rather than in absolute numbers.

Year	Total	Change
1994	50	
1995	49	-1
1996	48	-1
1997	44	-4
1998	40	-4
1999	38	-2
2000	36	-2
2001	33	-3
Average Change		*-2.43*
St. Dev. In Change		*1.27*
2002	**48**	**15**

Applying this to 2002 totals, we see that the increase of 15 auto thefts in City Square from 2001 to 2002 represents a whopping 11.8 standard deviations above the average change, indicating the probability of a pattern at work.

Crime Mapping

The crime mapping method of pattern identification involves creating pin maps or thematic maps of various crimes, and seeking geographic hot spots or clusters. This method is valid *only* for crime patterns that exhibit geographic clustering. Many crime patterns involve crimes that are *not* geographically close, so, like threshold analysis, the crime mapping method should not be the only means by which an analyst seeks patterns. However, also like threshold analysis, mapping crime may help catch patterns lost in the detail of daily report review, and it may be beneficial for understaffed agencies that cannot effectively use daily report review.



An analyst's map of housebreaks. The analyst has "eyeballed" three spatial patterns and has numbered them for further analysis. Automated programs would have caught the cluster in #1, but may have missed the pattern around the lake in #2 and the linear pattern in #3.

Trying to find crime patterns through mapping can be a daily process, but most analysts will do fine to view a map of each crime type once per week. The analyst must use his intelligence and experience to determine the time range to display. For many low-volume crimes, the analyst may want to view the past six to twelve months to find patterns, while for more common crimes, a monthly or bi-weekly time range may suffice.

Identifying patterns through crime mapping requires that the analyst achieve as close to a 100% geocoding HIT RATE as possible, which involves vigorous editing of both map data and crime data. Low hit rates obscure potential patterns.

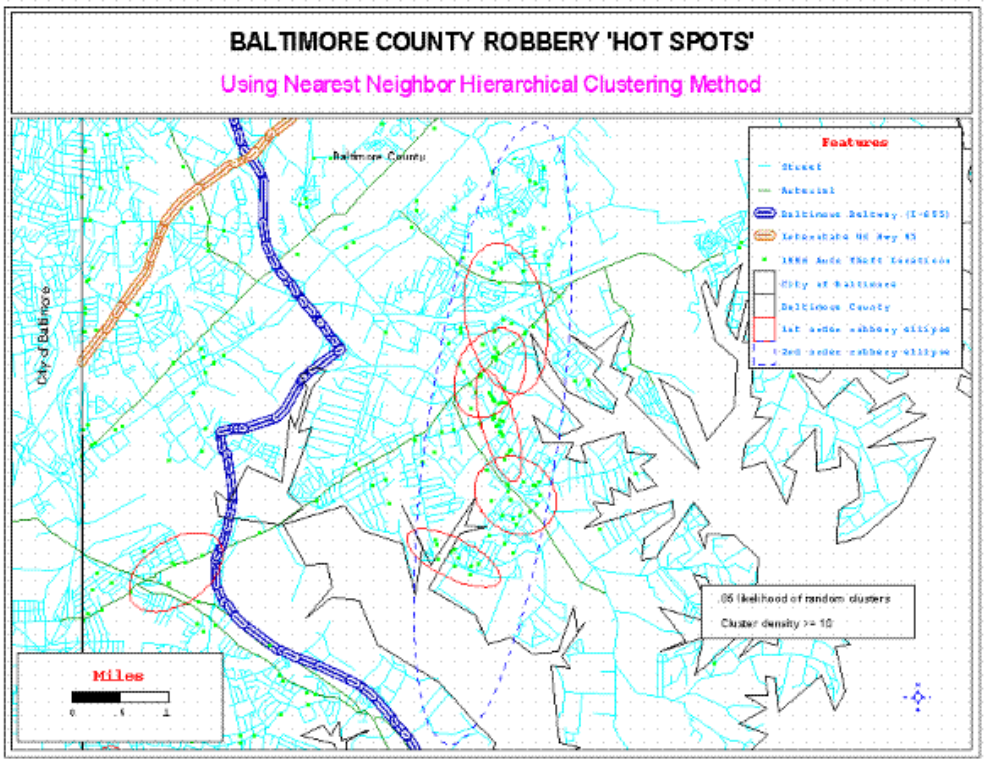
As with threshold analysis, analysts identifying potential patterns through crime mapping have both informal and scientific methods at their disposal. The informal method involves simply "eyeballing" the map and flagging clusters or other spatial patterns for further analysis. The scientific method uses advanced GIS tools such as STAC[4] and CrimeStat[5] to identify clusters for you. One advantage to the latter method is that your eyeball cannot see multiple points at the same location ("stacked" on top of each other, they look like a single point from a bird's-eye GIS view) unless you use a THEMATIC MAP (either a GRADUATED SYMBOL MAP or a CHOROPLETH MAP) or a point dispersion tool. Automated methods *will* account for multiple incidents at the same location and may circle some hot spots that your eyeball misses.

The disadvantage to automated methods is that they assume an equal opportunity surface (not accounting, for instance, for large fields where auto thefts could not occur) and generally identify only

[4] Spatial & Temporal Analysis of Crime, developed by Drs. Carolyn and Richard Block, available free from the Illinois Criminal Justice Information Authority

[5] Developed by Ned Levine & Associates with a grant from the National Institute of Justice. See http://www.icpsr.umich.edu/NACJD/crimestat.html.

clusters. Linear or ring-shaped patterns are often overlooked. By "eyeballing" the map, an analyst can merge the spatial patterns of crime with his own knowledge of the jurisdiction and achieve a much more sophisticated, if "unscientific," result.



A map showing clusters of robbery in Baltimore County, identified with CrimeStat. Source: Ned Levine,* CrimeStat II: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 2)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC, May 2002, chapter 6, p. 227.

Finally, geographic clusters only signify *potential* patterns. The analyst needs to take a further look at each cluster to determine if it exhibits commonalities other than geographic proximity. This can be done from within the GIS application or through a separate database query.

Intelligence-Based Pattern Identification

Analysts must accept that no method or combination of methods will ever allow them to identify 100% of crime patterns (which is why we speak in terms of *identifiable* patterns instead of total patterns). A crook who robs a bank on Friday, breaks into a convenience store on Monday, and steals a car on Thursday has probably successfully eluded detection by varying his M.O. so considerably. The only way the analyst would like these crimes is if the suspect was described in detail in each incident, or if the analyst receives specific INTELLIGENCE (from e.g., a detective, a parole officer, or an associate) that this criminal is active and may be committing these crimes. We might call such a method *intelligence-based pattern identification*.

Upon the arrest of an offender and the collection of evidence, the analyst may engage in a kind of reverse pattern identification, tying likely crimes to an offender to aid in prosecution.

For example, assume that an analyst receives a call from the parole officer for Scott Driscoll. The parole officer says that Driscoll hasn't made any appointments in several weeks, and that he may be

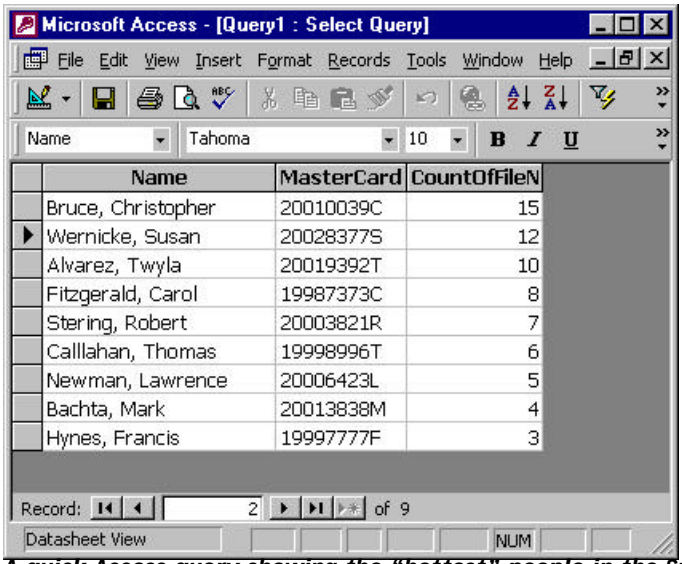
involved in criminal activity in the analyst's town, where he lives. The analyst conducts some research on Driscoll and finds that he was arrested and convicted a year ago for breaking into cars and stealing cellular telephones. In the past month, the rate of cellular telephone thefts from cars has been slightly higher than average, but this is a frequent crime in Gotham City, and the thefts have not exhibited enough commonalities for the analyst to confidently aver the existence of a pattern.

The analyst issues an alert to patrol officers with the parole officer's warning. Two days later, an officer spies Driscoll running away from a parking lot with a cellular telephone in his hand. The officer stops Driscoll to question him, and a witness arrives who relates that she saw Driscoll steal the phone from a car. Driscoll is arrested.

A search of Driscoll's apartment reveals a dozen or so cellular and digital phones. The analyst is able to search his database and find the incidents in which these phones were stolen, thus aiding detectives and prosecutors. Driscoll has been a cunning crook, varying his days, times, locations, and M.O., but the intelligence the analyst received allowed him to identify a pattern where he previously could not.

Hot Spots and Criminogens

HOT SPOTS (locations with multiple incidents) and CRIMINOGENS (individuals connected to many incident reports)[6] are easier to identify than other types of CRIME PHENOMENA. Though the analyst may identify both of them through the regular review process, it also behooves him to periodically run a database query that simply counts the number of incidents attached to a particular person or location (which could be an exact address, a street, a reporting area, a census block, a grid cell, or other "location" type) in any given time period. By sorting these query result in descending order by the total, the analyst can easily see who the "hottest" people and what the "hottest" spots are. A more careful review of each incident will then help the analyst discern the nature of the problem.



A quick Access query showing the "hottest" people in the Summer of 2001

The analyst may periodically want to create such queries to find, for instance, the "hottest" types of property stolen, the "hottest" makes and models of cars stolen, or the "hottest" weapons used in violent crimes. Such queries may supplement ongoing pattern and trend analysis, even if they do not, *per se*, constitute pattern identification.

---

[6] Records management systems based on MASTERCARDS greatly facilitate this process.

Cross-Jurisdictional Pattern Analysis

Any of the methods described above can be used to identify crime patterns that cross jurisdictions, assuming that neighboring agencies have found some way to share crime data. Thus far, most efforts have focused on cross-jurisdictional crime mapping, such as the San Diego area ARJIS[7] project. Maryland has developed a system known as the Regional Crime Analysis System (RCAS) in which multiple agencies submit crime reports to a combined database that allows detailed M.O. review. The 19-year-old VICAP system is a cross-jurisdictional pattern identification system, though generally only for murder. Other systems, such as the NLETS Telecommunications system, allow analysts to see detailed crime descriptions from other jurisdictions, but sporadically. All in all, cross-jurisdictional pattern identification remains an infrequent, irregular process.

Analysts without the benefit of automated cross-jurisdictional systems would do well to at least contact analysts from surrounding agencies on a regular basis to review crimes that are frequently subject to cross-jurisdictional patterns. The exact list will differ by region of the world, but commercial burglaries, commercial robberies, murder, rape, arson, and confidence games frequently fit this definition.

Conclusion

No single method will ensure that the analyst identifies *all* identifiable patterns. Good analysts will use a combination of these methods on a regular basis: report review every day; crime mapping perhaps every week; threshold analysis perhaps every month; and intelligence-based methods and cross-jurisdictional methods as the situation permits.

Methods of identifying crime patterns presented above will, and should, consume a large part of the analyst's day. It is not unreasonable to expect an analyst to spend a quarter to a third of his time seeking out patterns of crime, particularly if he engages in careful and thorough daily report review. The payoff is the peace of mind that comes from knowing that all identifiable crime patterns have been found.

[7] Automated Regional Justice Information System, http://www.arjis.org/