Visualization of Criminal Activity in an Urban Population
Alex Breuer, Joshua Hursey, Tonya Stroman & Arvind Verma
Indiana University- Bloomington

Abstract
Crime is a multi-dimensional and complex activity. In order to understand its nature one has to comprehend not only its spatio-temporal dimensions but also the nature of crime, the victim-offender relationship, the role of guardians and history of similar previous incidents. Visualization of all the dimensions is a formidable task due to the limitations of the present methods. Both for the police department and criminal justice researcher the need to deal with a vast amount of data is a pre-requisite to the task of controlling the crime. This paper presents an interactive visualization technique intended to present the viewer with an accurate and intuitive view of the criminal activity in a cityscape. The technique employs many different visualization elements, which taken together presents a useful methodology to handle the important associated factors of crime. The paper also describes the software technique and discusses avenues for future investigation.

Keywords  Crime  Visualization  Spatial-Temporal

Introduction
Crime mapping is a powerful tool of investigation and management for criminal justice practitioners. Its utility has been suggested in supporting problem-solving and community policing efforts improving law enforcement and criminal justice operations to apprehending and convicting offenders (LaVigne and Wartell, 1998). Crime mapping has also led to some powerful investigative techniques such as ‘geographical profiling’ (Rossmo, 1995) to track serial offenders and understand the movement patterns of offenders. The utility of crime maps in analyzing crime data (Hirschfield and Bowers, 2001), in understanding neighborhood ethnic composition (Brimicombe, Ralphs and Sampson, 2001), in analyzing drug markets (Voltz, 1999), and in understanding homicide hot spots (Adams, 2001) have all been well established. The COMSTAT process adopted by New York police further demonstrated its utility in creating intelligence based management system (Walsh and Vito, 2004). However, crime maps have generally not gone beyond displaying the spatial dimensions (Weisburd and McEwen, 1997; O'Kane, Fisher and Green, 1994). Commonly, these maps use street networks and colored symbols to depict the nature of criminal incidents. By adding orthogonal images and surrounding landmarks mappers are able to add more information to the visualization. However, these methods are unable to depict and incorporate simultaneously other factors and dimensions of crime thus reducing the utility of visualization.
Crime indeed has several dimensions. Brantingham and Brantingham (1981) suggest that there are four major dimensions that constitute the criminal event: the law which defines the act to be criminal; the motivated offender; the victim or the target and the coming together of all three at a geographical convergence. Further, the environment surrounding a criminal event is as important as the ‘awareness space’ of the motivated offender (Brantingham and Brantingham, 1993). Obviously, the convergence that takes place at the geographical location is itself facilitated by several factors. For example, ‘routine activities’ brings the victim and offender in close proximity in the absence of guardianship at a specific place (Cohen and Felson, 1979). Similarly, the nature of ‘hot spot’ (Sherman and Buerger, 1989) that makes the place criminogenic and the situational factors (Clarke, 1992) which provide opportunities for the offender to operate and play an important role in facilitating crime at a particular spot. Since criminal event is an interconnected incident of several dimensions, investigators and analyst need more information than the simple geographical location of the incident. The ability to solve crimes and develop preventive methods is dependent upon the capability of answering the six W’s about the criminal incident: Where and when did it take place, what is the nature of incident, who has done it to whom and why. Thus, it is not only necessary to know where the crime took place, the spatial coordinates, street address and neighborhood information but also the time period, victim-offender characteristics, relation to previous and neighboring incidents as well as knowledge of police response. The visualization technique therefore needs to display more than the limited geographical information about the criminal incident for it to be a useful tool for the investigators.

Due to the inherently geospatial nature of the data, crime visualization borrows heavily from cartographic design, and most designs are based in a geospatial substrate. Current approaches to geospatial crime visualization have addressed the problems of representing crime density information with techniques from cartography utilizing choropleth maps, graduated symbol maps and topographical maps. Generally, the temporal dimension of crime in most analyses is neglected in favor of the spatial dimensions. One reason is the inherent difficulty of visualizing space and time together. The approaches to representing temporal data in a geospatial substrate are not as well developed as the former techniques.
One current approach to display the time component is to represent crime data in an artificial 3-D perspective, in which the third dimension represents time. This approach suffers from the problems of projecting a 3-D model into a 2-D projection, such as occlusion and distortion of data. The two dimensional computer screens or any hard copy makes it impossible to display more than the three spatial dimensions. Therefore, another attempt by the researchers is to display the data through creative use of colors and symbols that add more information about the nature of crime incidents (Bonneau, Ertl and Nielson, 2006). Alternately, one can use a series of separate maps where the incidents are shown over different periods of time. Such a technique is limited since it is difficult to comprehend information from separate windows and fails to provide comprehensive information that accrues from viewing space and time together.

More advanced computational methods such as Virtual Reality Modeling Language (VRML) have been used to add temporal information to the spatial dimension in an interactive manner (Lodha and Verma, 1999). However, this demonstration has suffered from two limitations: the project was not directly geared towards the development of spatio-temporal dimensions but more concerned with enhancing the ability of the investigators to visualize the crime data in a new manner. Second, it did not develop a system of visualization that could easily be adopted by investigators. That is, criminal justice practitioners could realize the capabilities of VRML but had no way of knowing how to adopt it to their own problems and data sets. Nevertheless, packing greater and greater information is difficult if not impossible by the techniques mentioned above. In order to deal with these problems we develop and present another technique that enables several dimensions of the information to be visualized simultaneously by the researcher.

Our objective is to transform the large, complex and detailed dataset that is available about crime incidents, and present it in a visually appropriate and non-overwhelming manner. We believe such visualizations will enable crime analysts to easily interpret the data and find patterns for solving the crimes. It will also enable police officers to find historical hotspots of crime, and tract crime trends for a specified area. Accordingly, in this paper, as an illustration we set out to visualize the following dimensions of the criminal incidents:
- Crime rate and density in a specific area
- Number of arrests made for a specific crime
- Trend of a type of crime of a given time range
- Top 5 or 10 most dominant crime types
- Locations of dense criminal activity
- Location of significant landmarks (e.g., Police Stations, Fire Departments, Hospitals and Schools) in relation to areas of crime

These factors constitute not only the usual spatio-temporal dimensions but some other associates that are useful for crime analysis. We designate this visualization technique as ‘Multiple Interrelated Views’ where the various components of crime desired by the researcher are viewed in an overlapping manner. This enables the researcher to obtain all or any part of the information in an easy to comprehend manner. The technique seeks to provide the ability to transition between viewing the data representation for a specific location as well as a larger region. This suggests the need for:

1] A zoom and pan feature for map interactions
2] The ability to zoom into the data
3] Present statistical data for specific crimes
4] Represent significant landmarks prominently on the map

These four components help establish a common reference system and easy correlation between crime incidents and incident response resources. We also incorporated the capability to represent the trends of specific crimes for a fixed time period. This helps determine the changes in crime rates over the particular time period.

**Data and its transformation**

Our data was provided by the Indianapolis police department representing selected criminal incidents in Indianapolis between Jan. 1, 2000 to Dec. 31, 2000. For each criminal incident, the dataset provided: street address, Geographic Information System (GIS) coordinates, timestamp (date & time), and criminal code reference. The data was brought into the ArcGIS 9.1 software where using the street address we were able to map the criminal incident to a specific location on the map. We used census tracks to define a spatial grouping for the crimes. The census boundary was used as a reference
to display the density and rate of crime. Once focused on a specific area we could represent the information of that area at a higher resolution. The GIS software also allows us to access the locations of various landmarks such as the location of Police Stations, Fire Stations, Hospitals, and Schools along with primary streets and their intersections. These features help orient the viewer when presented with the map at different scales.

**Timestamp**

By using the timestamp, we could group crimes into temporal categories. We incorporated the facility to let the viewer determine the granularity of the categories. For example, if the viewer requests a single year of data, then the categories may be represented by month. If they request a single day, then the categories may be hours. By restricting the time range, the viewer also limits the amount of data that needs to be processed, since we need not process data outside the specified time range.

**Crime Code**

The crime code classification allows us to group crimes together into a hierarchy of categories. For this dataset we were able to represent the following primary categories: Assault, Burglary, Homicide, Larceny, Rape, Robbery, and Vehicle Theft. Each primary category has one or more sub-categories. By representing crime in a hierarchical manner we could present the viewer with an overall view of the crime density or rate in an area, and then allow them to focus their attention on a specific type of crime.

**Visualization**

The objective of the visualization was to represent the spatiality of, trends in, and correlations between crimes, while always preserving a common reference system. From our data analysis we were able to extract the visuals and incorporate in the technique the facility to allow a viewer the desired information in an appropriate and non-overwhelming manner. Each of these visuals was represented in a distinct overlapping window that ensured the information was available if desired.

**Spatial Representation**

We began with a 2D geographic map of Indianapolis generated with GIS software which overlays the landmarks (noted above) onto the visual field. This map allows for a common reference system for the viewer. To this map we added a hillshade representation of the crime data.
The hillshade representation, seen in Figure 1, expresses the density (or rate) of criminal incidents in a specific area in relation to adjacent areas, over the entire time range specified. The color representation ranges from high crime density in red to low crime density in green. The shading adds depth to the 2D image without tiling the map into the 3D plane. Areas of higher elevation correlate to areas of high crime density. The city map is then overlaid on the hillshade representation, which allows the landmarks to standout more prominently without being obfuscated by the hillshading. For each zoomed view, this spatial representation is recreated for the zoomed area. The zoomed area is then overlaid upon the full map of the city, emanating from the center of the mouse click. This allows the viewer to move between views of the entire city, and associate the current view with a region of the city.

**Crime Relationship Representation**

In our visualization, we sought to preserve the hierarchical nature of the crime code classification system. By using a treemap visualization technique (Shneiderman, 1992), seen in Figure 2, we were able to preserve the hierarchy and allow the viewer to zoom into the data by clicking on the primary categories.
If a user clicks on a primary category then they are presented with another window, on top of the previous, that presents the subcategories for that category in a treemap format. The user is able to ascend the hierarchy by clicking on the treemap window under the current window. The color-coding matches the color-coding selected by the user. We allow the user to define the color-coding used in the visualization. This helps them readily identify those crime categories that they are most concerned with and recognize them with a color that is common in all the windows.

**Temporal Representation**

We explored many different techniques for expressing trends in crime for a given area. The most obvious appeared to be the line graph that provides the crime trend for a specific time period, as seen in Figure 3.
Apart from the line we wanted to explore a different representation that might assist the viewer in identifying trends more readily. The best option we had was a crime clock where the trend is presented radially. The advantage with this representation is that for the daily period of time it has a familiar look and is more comprehensible. The clock visualization is shown in Figure 4.

Each slice of the clock represents a time segment in which crimes are aggregated. The number of incidents for the time segment is in red so as to stand out, and each time segment is labeled appropriately. By placing the color coded name for the category in the middle of the clock, we allow the viewer to easily identify the nature of the crime. The shading of the time segments is meant to visually indicate the progression of time around the circle in a clock-wise manner. For each category, there is one clock for visualization. In order to depict all of the clock visualizations at the same time, we employ a fisheye
distortion technique (Bederson, 2000). This allows the viewer to bring into focus one-clock visualization at a time, and the others become pushed to the periphery.

**Combination of Visual Elements**

When all of the visual elements are taken together, we produce an interactive visual environment in which the viewer can explore the criminal incident dataset. Each visual element depicts the relationship between the density or rate of criminal activity to time, space, or each other. A snapshot of the visualization can be seen in Figure 5.

![Figure 5: Final Rendering Snapshot](image)

**Technical Specifications**

*Programming tools*
The following tools and programming languages were utilized in developing the visualization:

- ArcGIS Software (ArcSDE API)
- Java, Perl, C/C++ Programming Languages
- Relational Databases (SQL queries, PostgreSQL)
- OpenGL or other interactive visualization environments (Fry and Reas, 2006)

**Pipeline**

The visualization is accumulated in a pipelined manner (as seen in Figure 6), developing the data and visual representations as it progresses from the raw data to the final rendering.

![Figure 6: Rendering Pipeline](image)

There are four main stages: Raw Data, Data Table Transformations, Visual Mapping and View Transformations.

**Raw Data**

This is the raw data representation. The data is saved in a relational database, and is accessed using standard SQL queries. ArcGIS uses this database to aggregate the data when rendering the spatial representation. The combination of ArcGIS and the relational database generate a map with associated metadata from the database. The development of

---

1 Due to the sensitive nature of the data set, the dataset and associated source code cannot be released publicly, but is available upon request and approval.
the *Rich Map* is the first aggregation of data, which contains all of the data represented by both the ArcGIS and database systems. This *Rich Map* can be saved to a database, as to alleviate the need for ArcGIS in the pipeline.

**Data Table Transformations**

In this section of the pipeline, the system is attempting to refine the data set based upon the parameters set by user and their interactions with the system. For example, the user will select a time range over which they wish to view the data set. The *Interactive Refinement* stage will reduce the *Rich Map* to a subset of the data included in this range. Similarly this stage would also reduce the *Rich Map* to just the data that is currently in zoom. This is all to prepare the subset of the data that the user is concerned with for the visual elements to work at a later stage in the pipeline. This is the stage where the user has influence upon the data that is displayed.

**Visual Mapping**

This takes the constrained *Rich Map* of data, and imposes a color mapping upon it in the *Color Mapper* component. The *Color Mapper* component takes the user defined color-to-crime mapping and adds it to the *Rich Map*. This is all of the information that the different visual elements need to produce their respective visual elements.

**View Transforms**

In this stage each individual visual element is able to interact with the reduced and enhanced *Rich Map* dataset. For this visualization this stage includes the visual elements: *Timeline*, or the temporal representations; *Treemap*, or the crime relationship representation; and *Isobar*, which is the spatial representation. These are described in more detail below. Each visual element module produces one or more images for the *Rich Map* that are presented and which represent the dataset desired by the viewer.

**Final Render**

Once all the visual elements are produced, they are combined with the base map and legend to produce the final rendering of the image. As the user interacts with the final rendering of the image they can influence the pipeline but this can be exercised only at the initial stage. For example, if the user changes the color-coding, then the pipeline is re-executed from the *Visual Mapping* stage, as the *Data Table Transformations* stage is not directly affected by this interaction. An example of a final rendering is shown in Figure 5.
Spatial Representation

There are several choices for algorithms to interpolate a function over a nonlinear surface defined by discrete measurements. For example, one could use polynomial splines, Levenberg-Marquardt, or Delaunay triangulation. Since the surface need not be differentiable, Delaunay triangulation is an appealing choice due to its simplicity relative to the other options (Okabe, Boots and Sugihara, 1992), and that it is included in popular GIS tools. In order to draw the temporal crime map, it is necessary to have geospatial crime density information. Although this could be accomplished by imposing an arbitrary grid over the map, and computing crime density based on those cells, such a grid is inherently arbitrary. It is important to use natural boundaries in order to comprehend the crime density data meaningfully. Furthermore, it is important to have an appropriate level of granularity so the geospatial subdivisions must not be too large or small. For computing crime density, we used crimes per police precinct for the crime topology, and census tracts for other experimental maps. We observed that the police precinct divisions produced appropriate density information, which was more granular than the census tracts, although this may not necessarily hold generally.

The ArcGIS tools provide all the database functionality to produce the static features of the maps we used for our design. Indeed, one could automate the map generation process using the ArcSDE programming API. Most data processing steps only require simple database operations; the only GIS specific data processing operations are spatial joins and geocoding. ArcGIS also offers creation of triangulated irregular networks for crime topology synthesis. For the crime topology we found that color categories were best assigned by Jenk's natural breaks, which ArcGIS supports. Finally, ArcGIS can rasterize map information, which would allow another program to associate points on the map with database entries for interactive tree and temporal map drawing. These features of spatial representation may be seen in Figure 1.

Crime Relationship Representation

For this visual element we used a treemap that was generated with a popular library written in Java (Plaisant and Shneiderman, 2004). The treemap can be zoomed in as the user clicks on the different primary nodes in the tree. The labeling is done in the library to produce display windows needed for this interactive visualization. As the user
zooms into the treemap by clicking on primary nodes, another treemap is overlaid upon the parent to represent a stack hierarchy allowing the user to return to a previous, higher view of the tree. If the user hovers the mouse over a node in the treemap it pops up with a small ‘tooltip’-like window showing the specific data for that node, and it’s children. This is a feature currently built into the library.

**Representation themes**

All of the temporal representations were presented close together in the final rendering, as to directly correlate them to each other. The line graph representation was developed in Microsoft Excel, but could also be produced with other graphics packages such as GNUplot. The actual number associated with the density or rate of crime was presented at the edge of the interior of the image in red, as to standout in the image. The textual representation of the time slice was placed near the interior of the graphic. The interior circle was associated with the color-coded crime and name of the crime for quick identification. The time slices were ordered in a clock-wise manner and their background shading in degrading gradients of grey, as to express the clock-wise ordering of the information. One such clock was presented for each crime category in the dataset.

By producing one temporal clock for each crime category in the dataset, the system produced more clocks that can be easily displayed on the final rendering. Therefore, we used a fisheye menu technique that can allow the user to focus upon one clock, and move the other clocks into the periphery of the menu. We also provided an interactive legend that could explain all of the different visual elements. We also provided interactive buttons to allow the user to restrict the time range explored, and color-coding of crimes in the final rendering.

**Conclusion**

The work doesn’t stop with the final rendering. The user needs to be able to interact with the data to properly explore the face of crime in the city. If the visual element distorts the data in any way, the user will interpret inaccurate data, and be lead to make improper correlations from the visualization. Thus, while developing this application we attempted to make sure that the user is able to intuitively interact with
each of the visual elements and that the visual elements accurately represent the data that they are acting upon.

The advantage of using this visualization technique is the ability to create simultaneous multiple overlapping windows for depicting the different dimensions of crime. Apart from the spatio-temporal dimensions, information about the nature of crime, victim-offender characteristics, neighborhood, previous incidents and police action can all be displayed in overlapping interactive windows for the viewer. As the fisheye visual suggests the windows could be made to hide behind other windows so that the viewer can focus upon the needed information. The facility to add colors and symbols of choice, of selecting time and spatial cone of resolution further adds capabilities of adding more information to the incident and area under study. Since the computer programming is not too challenging the technique can easily be adopted by most researchers.

A further area of research is to embed this technique with the VRML for panning into the area for a variety of images. VRML has been shown to present space and time together in an interactive manner. The capability of zooming into the area and viewing the image from a variety of angles is a special feature of VRML. By incorporating the two techniques the visualization becomes dynamic and extremely versatile in displaying almost all the features of criminal incidents.

REFERENCES


Bonneau, Georges-Pierre; Ertl, Thomas and Nielson, Gregory M. (eds.) 2006. *Scientific Visualization: The visual extraction of knowledge from data* Berlin: Springer.

Brantingham Paul J; Brantingham Patricia L. 1981. *Environmental Criminology* Beverley Hills, Ca: Sage


Fry, Ben and Reas, Casey 2006. *Processing 1.0 (BETA)* [http://processing.org/](http://processing.org/)

Hirschfield, A and Bowers KJ. 2001. The development of a social, demographic and land use profiler for areas of high crime *British Journal of Criminology*, 37 (1), pp. 103-120.


