

THE BUFFALO CRIME MAPPING SYSTEM:
A Design Strategy for the Display and Analysis
of Spatially Referenced Crime Data*

Kurt E. Brassel, Jack J. Utano, and Perry O. Hanson, III
Geographic Information Systems Laboratory
State University of New York at Buffalo
Amherst, New York 14226

This paper presents the design strategy for a crime information system with vast capabilities for production of and experimentation with maps. Detailed disaggregated crime information from the city of Buffalo, New York forms the primary data base for the system. This series of crime files, together with socio-economic data from the Census, are integrated into a geographic information system denoted as the Crime Analysis and Research Package (CARP). The Buffalo Crime Mapping System as a part of CARP is a collection of data files and routines designed for versatile display and analysis of spatially referenced crime data. It deals with three types of information: crime, census, and spatial reference information containing various geometrical indices such as block group boundaries and centroids of statistical units. The boundary information is provided by a flexible arc structure. The three major data files are stored as random access files and are interfaced with various display procedures such as choropleth, graduated circle, and dot mapping programs. The user provides the system with a set of instructions as to the desired map type and its display parameters. A driver routine then reads the map type instruction and calls the appropriate mapping routine. This mapping program then proceeds by reading the user parameters, accesses the necessary base files, and produces the final map. At its present stage, the system is batch oriented and does not provide for interactive and/or real-time communication features.

Key Words and Phrases: Geographic Information System, Computer Cartography, Cartographic Data Structures, Information Retrieval, Criminal Data Processing. CR Categories: 3.33, 3.39, 3.73, 3.74, 8.2

1. INTRODUCTION

Information systems of all kinds are becoming more and more important, and their impact influence most spheres of human life. Data concerning population, economic structure, land utilization, pollution, and traffic (to mention but a few) are systematically collected for the purpose of decision-making on various stages of political, scientific and private life. In order to profit from these efforts fully, these data must be made readily available in convenient form to possible users. To fulfill this information need, graphical representation techniques are highly favored, since people are able to overview, integrate and comprehend a large amount of information when displayed efficiently. This calls for development of automated representation techniques for the growing but often neglected quantity of public use data being collected.

In this paper we attempt to present the design strategy for an information system with vast mapping capabilities for both production of and experimentation with maps. Detailed disaggregated crime information from the city of Buffalo, New York forms the primary data base for the system. These series of crime files together with socio-economic data from the Census Bureau are integrated into a geographical information system, de-

noted as the Crime Analysis and Research Package (CARP), developed by the Geographic Information Systems Laboratory at the State University of New York at Buffalo (see Hanson, et al. 1977).

This paper focuses upon a subset component of CARP, the Buffalo Crime Mapping System. This system is designed for the development and implementation of current computer cartographic technology for the automated representation of complex phenomena associated with the geographical crime information system.

In our presentation we will first describe the nature of the crime information base and then discuss the technological aspects of the structure and development of the Buffalo Crime Mapping System, before concluding with a look to its importance and future applications.

2. CRIME INFORMATION AND DISPLAY APPLICATIONS

For the past six years, the Buffalo Police Department (NY) has compiled and maintained detailed incident and arrest records in machine-readable form. In order to appreciate the potential graphic applications of this crime information a brief analysis of the 'information dimensions' in this system is in place. The term information dimension is used here for a group of independent variables of similar quality (Compare Bertin, 1967). Each event in the crime file is described by:

- Location: census tracts/block groups/blocks; precincts; street addresses; police code.

*Research supported by the Research Foundation of the State of New York, Grant Numbers 050-7361A and 50-7289A.

- Time: hour, day, week, month, year.
- Information about the offender (age, race, sex)
- Information about the victim (age, race, sex)
- Type of crime and classification of offense
- Event **status**: offense-arrest-clearance etc.

To assess the information wealth of maps that may be produced from the existing data material, each element of all non-locational information dimensions may be mapped with respect to all location identifiers; some of these information dimensions are quite long (the crime file maintains information about more than 500 classifications of criminal offenses). In addition, this number is magnified when these information dimensions are cross-referenced, thus permitting mapping of an element of one dimension that is qualified by any combination of items of other dimensions (e.g. map of burglaries by blocks on Saturdays, 4-6 p.m. involving white male offenders). Considering this vast wealth of information that may be displayed by various mapping methods there exists an obvious need for a systematic survey of the potential of computer cartography for such display contents of information systems.

3. THE BUFFALO CRIME MAPPING SYSTEM

3.1 Overview

The Buffalo Crime Mapping System as part of the Crime Analysis Research Package (CARP) is a collection of data files and routines designed for versatile display and analysis of spatially referenced crime data. A system overview is given in Figure 1. The system may be subdivided according to information type and functional steps where it is dealing with three types of information: crime, census, and spatial reference information. The crime data have been discussed in the previous section. The bulk of the census data are provided by the 1970 federal census and refer to

the Census Bureau statistical units (census tracts, block groups). The spatial reference information consists of various geometrical indices such as block group boundaries and centroids of statistical units.

The three functional steps of the system are data preparation, information storage, and the (application-) processing. The data preparation step includes functions such as data compilation (collection, digitization) and file manipulations in various degrees of complexity. The information storage section consists of three major files. The Multiple Access Crime File (MACF) provides a versatile storage of all single crime events. The Statistical Unit Attribute File (SUAF) includes crime and census data aggregated by block groups and census tracts: it is a location by variable matrix. The Topological Arc Reference File (TOPARF) is a versatile spatial reference file that allows for geometric manipulations and mapping applications. All the essential files are either in an index sequential or random access file organization. In processing we distinguish between data analysis and display functions. The overall system is used as follows: Crime data are modified and put into the Multiple Access Crime File. An aggregation program retrieves subsets of the modified crime information and places it into the location attribute file SUAF. Census and certain spatial reference data are also transferred into SUAF. The file TOPARF is created in some rather sophisticated steps from spatial reference information. Once created the three files are available for user processing. Data analysis as a rule makes use of disaggregated or aggregated crime and census information (MACF + SUAF) whereas mapping routines are in general based on aggregated crime/census information (SUAF) and the spatial reference file (TOPARF). Further discussion of the system emphasizes the functions that are important for information display. The

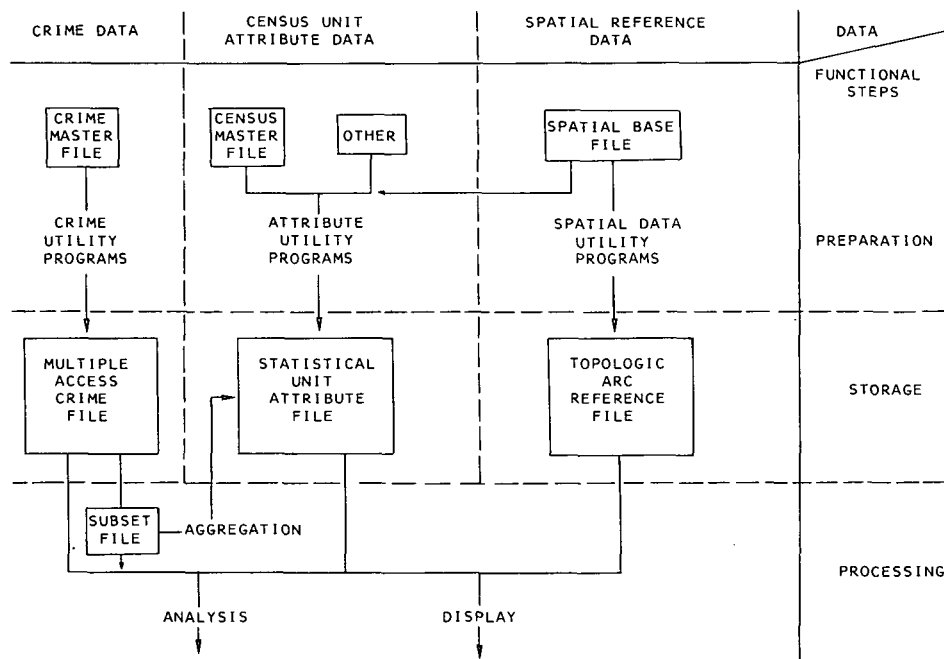


Figure 1.

Overview of the Buffalo Crime Mapping System

several sections deal with the following structural units outlined in Figure 1:

- The Buffalo criminal data record structure;
- Data aggregation and storage of thematic information;
- Organization of the spatial reference file;
- Display programs and interface to base files.

3.2 The Buffalo Criminal Data Record Structure

Figure 2 provides an overview of the various modifications performed to upgrade the utility of the crime information. The initial modification step in the system entails the reorganization of the logical file structure of the Buffalo Police Master Crime Files. Geocoding the crime files becomes the first step. The crime files were used as input to an address-coding, computer program package provided by Frank Rens of the Medical Center of the State University of New York at Buffalo (Rens, 1975). This particular software uses a GBF/DIME file to match the street name and number in the police files with a 1970 census tract, block group, and block identifier. Once this address-coding process is complete, the crime records are compatible with the 1970 census data.

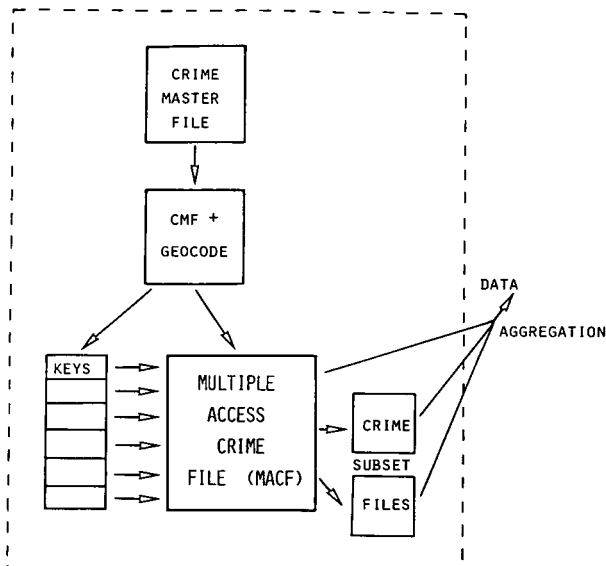


Figure 2. The Crime Record Structure

Upon completion of the geocoding phase, the files still retain their sequential organization by chronological identification (CD) number. Two forms of random access file organizations, actual key and indexed-sequential, are used to restructure the crime files into a more usable and manageable format. In an actual key structure, the crime records are stored and retrieved on the basis of a pre-determined relationship between the key of the record -- the CD code -- and the direct address of the disk location where the record is housed.

A data record in the index-sequential file contains only the particular crime variable code and

the actual key information (block number and record number) that points to the actual complete crime record in the actual key data base. In addition, other index-sequential files may be constructed that contain keys identifying and linking crime records according to time of offense or street location to the crime data base. These subsidiary index-sequential files act only as table pointers to the actual key data base. To address the secondary file only a key derived from the data record (crime type, for example) is needed and immediately the appropriate position of the actual key file for extracting crime records that match the pre-specified key is determined.

To obtain these multiple access paths to the actual key data base, the laboratory utilized one additional key configuration associated with the index-sequential file organization. The feature contains a partial key option where the user is permitted to pass only a portion of the full crime key to one of the index-sequential key files. By combining a number of specific crime variables to form one multiple index-sequential key file, extracting records based on a complete or partial collection of these crime keys may be achieved in a tractable manner.

The augmented capability generated by this multiple key organization equips the researcher with the flexibility to search the crime data base along three separate lines of inquiry. For instance, the user may key on any or all of the three crime variables simultaneously. Whether it be burglary incidents within a certain time period along a specified street, burglaries during a given time interval, or simply all burglary records, the user need only specify a partial or complete crime key to retrieve any of these types of queries in one access step.

Given the overall system design of the Crime Analysis Research Package, the utility attributed to this initial data preparation stage becomes apparent. The eloquence of this data structure may be further illustrated in the following example. Given the construction of the actual key data file for the 1975 Buffalo crime file with a corresponding index-sequential key file aligned by crime type, the researcher is able to call all the homicide records at a fraction of the CPU time it would require under a cumbersome sequential procedure. Here, the crime type code for homicide is first passed to the index-sequential key file. This file is immediately positioned to a contiguous subset of the index-sequential file where the homicide keys reside. The block/record location that is appended to each key record is then retrieved sequentially from that section where the homicide records are stored and are, in turn, used as key location identifiers to point to and extract the appropriate homicide records from the actual key data file. The retrieved records form a subset sequential disk file. Thus, for the user interested in utilizing the Mapping System to generate choropleth maps of homicide in Buffalo, he is not burdened with addressing the entire crime data base; instead he works with a considerably smaller subset file which is then aggregated into a number

of statistical units according to user-opted crime conditions. This data aggregation component is described next.

3.3 Data Aggregation and Storage of Attribute Information

The Statistical Unit Attribute File (SUAF) is created from crime and census information. Crime data are extracted first from the multiple access crime file MACF by user-opted crime keys and then the contents of the subset file are transformed to obtain frequency profiles of various crime variables for future mapping and data analysis (Figure 3).

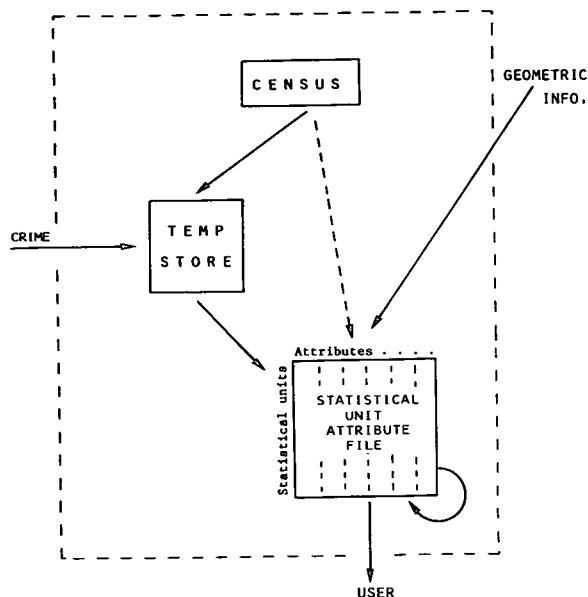


Figure 3. The Aggregation and Storage of Attribute Information

The logical structure of SUAF represents a two-dimensional matrix, location by variable, in random access format. In our current implementation, this random access matrix houses a number of crime and statistical variables (henceforth called attributes) by the various statistical units of the region in question. In the Buffalo case, the location dimension includes the city block groups, census tracts, precincts, and the total city. Given the particular random access structure used for SUAF, storage requirements, in terms of disk space, are dynamically allocated so that the amount of variables housed in the file is not restricted.

Principal application of SUAF may be distinguished between three cases: 1) retrieval by attributes; 2) single call retrieval, and 3) retrieval by locations. A typical use of such a file in a mapping environment is the retrieval of one variable, i.e. one information item for the units of the entire region (i.e. block groups or census tracts) for the production of one map (e.g. choropleth map). For such an application it is desirable to be able to retrieve the user-requested information

in a single read step. Other mapping routines may require single cell access in the location by variable matrix. Finally, a file user may also be interested in retrieving several or all variable values for one location.

Besides crime and census information a number of geometrical indices are stored in SUAF. These indices, such as centroids, surface area, area perimeter, min-max x-y coordinates, inscribing and circumscribing circles, etc., are also pertinent to the statistical units. Another type of information in the SUAF file are various pointers such as hierarchical pointers to higher/lower level statistical units in SUAF itself, or pointers to the geographic base file; the latter will be discussed in the next section.

Raw frequency counts by statistical units represent the basic information stored in SUAF. The user is now able to interface this file with the mapping and data analysis components of the system to obtain various maps of the frequency counts. In many applications, however, it is necessary to convert these new counts in some sort of standardized form to obtain more reliable representations of the data. Accordingly, the Lab has written a versatile procedure that is designed to perform various manipulations for any variable in SUAF (ratio, average, sum computations, etc.). The results may then be stored as new attributes in SUAF.

SUAF is thus a versatile storage file containing various types of attribute information for each of the pre-defined statistical units. Together with the spatial reference file, they represent the essential data source for all mapping procedures.

3.4 Organization of the Spatial Reference File

Spatial base files include all information necessary for spatial allocation of thematic data via location identifiers. They should be organized in such a way as to minimize storage requirements, maximize information content, and provide a multitude of graphic applications with only a minimum of conversion manipulations. The TOPological Arc Reference File (TOPARF) forms the base file for the Buffalo Mapping System. TOPARF currently contains record descriptions for the block group and census tract outlines of the city; its topological properties allow for multiple use of this data structure including manipulations where neighborhood functions are of importance (compare Peucker and Chrisman, 1975).

The block group and census tract outlines consist of straight line segments that are defined by a series of digitized points (x,y coordinates). Points that are members of one single polygon outline (block group or census tract) are called cartographic points and points that are common to three or more polygon outlines are called topological nodes. A chain of line segments joining two nodes is called an arc. TOPARF currently consists of approximately 1000 directly accessible records, each describing a unique arc. A direction quality is assigned to the arcs, where they are interpreted as being directional vectors. The initial direction (forward, backward) of an

arc is assigned arbitrarily, but once defined, it has to be used consistently for the remaining reference to that arc. Arcs and their associated arc records are illustrated in Figure 4, where the arcs are labelled by Arabic numbers, polygons (areas) by upper case letters, and nodes by Roman numerals. The record in Figure 4 describes arc 1. It is defined as being directed from a node I (start node or FROM node) to a node II (end node or TO node). Given this direction, arc 1 is separating polygon A (AreaLeft AL) from polygon B (Area Right AR). In addition, the record includes information regarding the neighboring arcs; these arcs are defined as being neighbors when they are delimited by the same start or end node of the present arc record. Four pointers are used to indicate the directional relationship of the present arc with its neighbors; they are defined as being in a forward left (FL), forward right (FR), backward left (BL), and backward right (BR) direction. Finally, the arc record includes information on the number and x-y coordinate values of the cartographic points. The arc pointers (FL,FR,BL,BR) point to other arc records in TOPARF in such a way as to permit a 'migration' along polygon outlines without the need of accessing additional pointer or coordinate files.

	Area left	Area right	FROM node	TO node	forward left	forward right	backward right	backward left	cartographic points
Arc 1 :	A	B	I	II	2	3	4	5

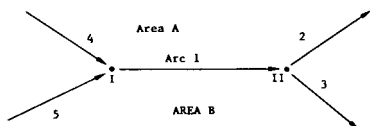


Figure 4. Arc Record Organization

Figure 5 gives an overview of the several files and procedures involved in creating the Buffalo Base File TOPARF. The base information for the topological reference file is a DIME file of the city of Buffalo (Bureau of the Census, 1970), and a file including segment coordinates of the same area. A first step connects the independent segments of the Buffalo DIME file into chains of polygon outlines. The segment coordinates are then attached to these chains resulting in a coordinate chain file. A display routine MAPDIME allows for plotting any subdivision of the region on a separate sheet and labels the complete set of topological identifiers. From the edited chains the arc file TOPARF is created. The box labelled 'Digitizer' indicates the possibilities for other strategies to create a chain file in case DIME files are not already available.

3.5 Display Software and Their Interface to the Base Files

The logical structure inherent in any generalized or versatile mapping system should include at least rudimentary procedures for handling three basic components in automated map design: data

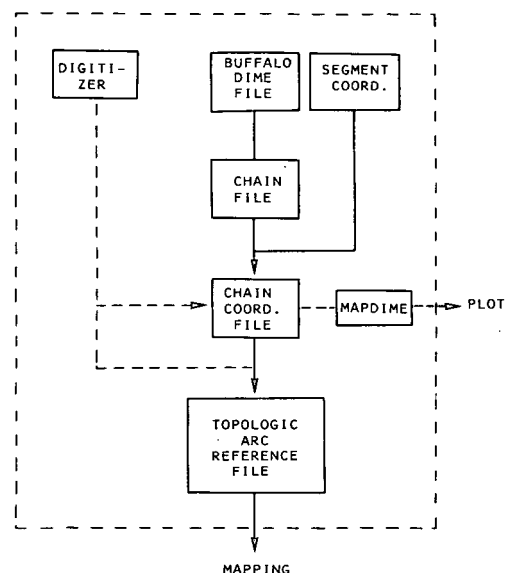


Figure 5. Organization of the Spatial Reference File

preparation, data manipulation, and mapping step capability. Systems containing all three design facets may be considered as a multiple step software package in contrast to single step mapping routines where data preparation and manipulation are evolved independently from the mapping system. In addition to this multiple step capability, automated mapping systems should usually include software facilities for the production of various types of maps. Such systems may be designated as multiple product systems as opposed to single product packages where, although the data generation and manipulation components are incorporated into the system, the mapping package is able to produce only one kind of map (e.g. choropleth). The first two components of the Buffalo Crime Mapping System have been given thorough treatment in previous sections; this section focuses upon the mapping phase and highlights the multiple product facet of the system.

The schematic relationship between the data base, the mapping software, and the system user is provided in Figure 6. It outlines the system's

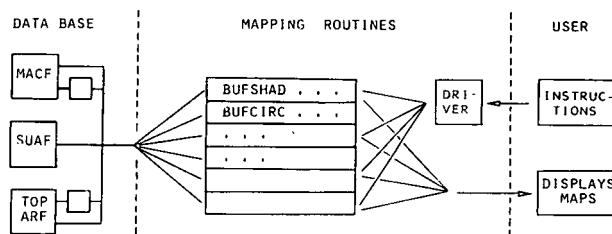


Figure 6. Use of the Buffalo Crime Mapping System

structural components necessary for the production

of maps from a user's frame of reference. The user initially provides the system with a set of instructions as to the desired map type(s) and its respective display parameters (the various types of mapping routines implemented and planned for the system are described below). Map parameters include user options such as map scale, shading symbolism, class interval determination, legend information, and the type(s) of data to be accessed from the data base. User instructions may be stacked and the system is designed for multiple map type production and multiple copy production in a single run. A driver program then reads the map type instruction and calls the appropriate mapping routine. This mapping program then proceeds by reading the user map parameters, accesses the necessary base files, and produces (a) final map(s). At its present stage, the system is batch oriented and does not provide for interactive and/or real-time communication features.

To assess the application capabilities of the Buffalo Crime Mapping System, the range and type of maps that may be produced is discussed. In general, potential map productions are a function of three major ingredients in any automated map production environment. Maps may be classified according to the data types they are designed for, the display devices used for their production, and the cartographic representation methods chosen.

DATA TYPES. As with most computerized mapping packages, two sets of data, thematic and geometric, are used to generate maps in this system. However, since different types of maps may be produced given the several types of data in the two sets, the user is restricted to using particular data set combinations for map production. For instance, specific thematic data (aggregated-disaggregated) require particular spatial-referencing information,

whereas certain geometric data (centroids-segments-area outlines) govern the kinds of thematic display possible.

According to the kinds of data set interaction permissible in this system, the mapping programs are designed to access the three data base files (MACF, SUAF, TOPARF) to extract the user-opted thematic and geometric information. In the present version, however, the user is responsible for creating any data set not already stored in the data base files.

DISPLAY DEVICES. Plotter, line printer, and CRT displays represent the primary output media used in automated mapping environments. The routines presently implemented in this system are designed for the CALCOMP 913 drum plotter. It permits high resolution vector representations and color variation. Line printer programs may easily be incorporated in the system; however, additional hardware devices are presently not available on the host computer.

CARTOGRAPHIC METHODS. Given the availability of the respective software and the hardware capacity of the host computer, there is no restriction as to the implementation of any of the traditional cartographic techniques into the Buffalo Crime Mapping System. However, the thematic and geometric data base requirements for a particular mapping program must be compatible with those in our system.

To illustrate computerized cartographic applications we discuss here briefly the mapping software planned for implementation into the system.

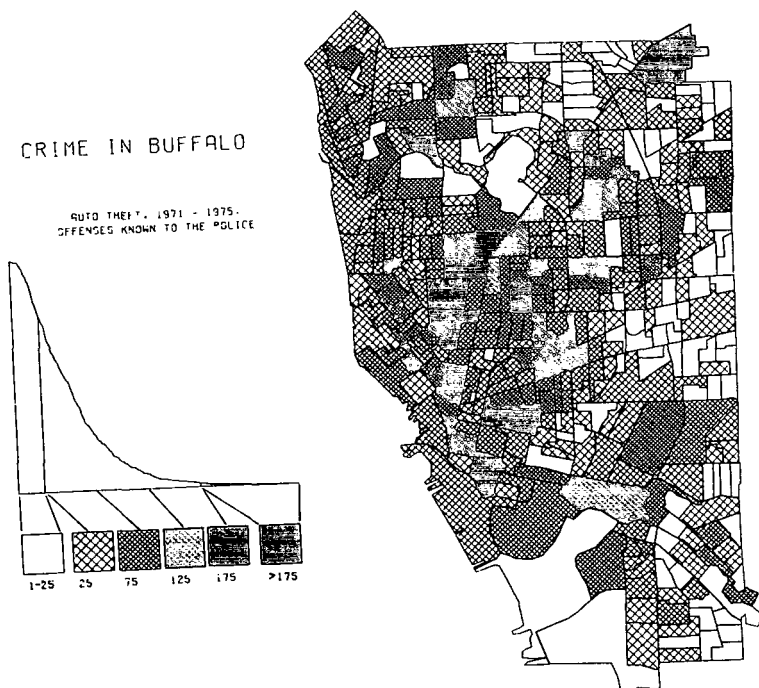


Figure 7. Continuous Shading Choropleth Map

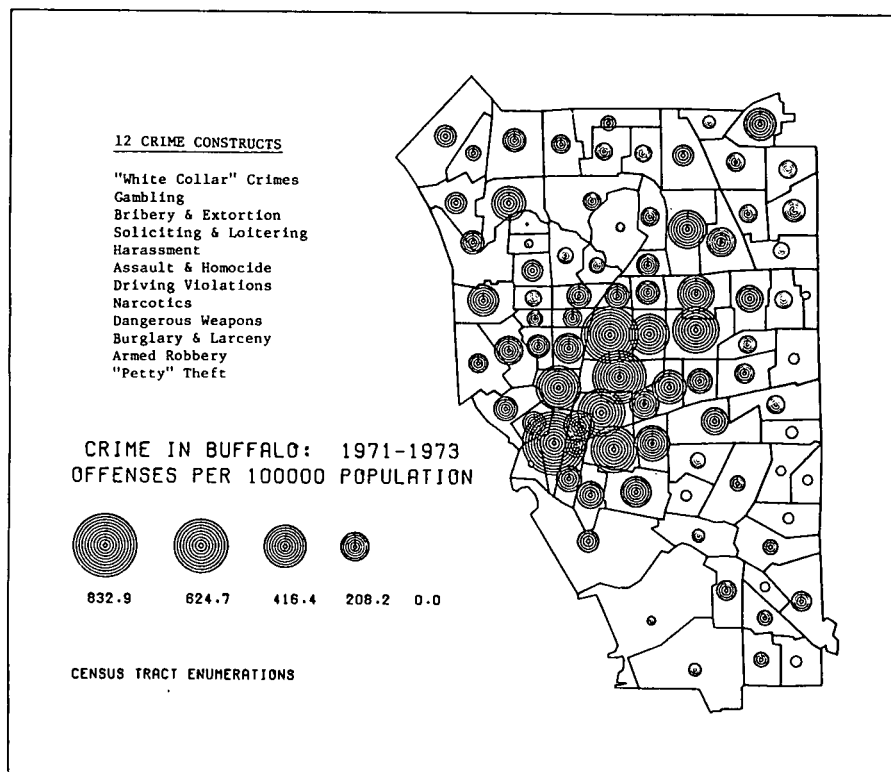


Figure 8. Graduated Circle Mapping

Choropleth Mapping. The program retrieves a set of attributes from SUAF (one item for each statistical unit), and the polygon outlines from TOPARF or an intermediate base file. This mapping routine provides for title and legend information, a shading symbol mechanism, an algorithm to choose class intervals, and various output parameters. At present the program uses a modified version of Tobler's continuous shading algorithm (Tobler, 1974). It has been modified to provide more flexibility in assigning grey tones to variable values (Figure 7).

Graduated Circle Maps. The program retrieves centroids and attributes from SUAF and represents one variable on one map. Options include titles and legends, minimum circle size, maximum circle size, density of circle symbol, type of base map (block group outlines, census tract outlines; full lines or dashed). The graduated circles are shaded by concentric circles, where the radius increment can be chosen by a user option (Figure 8).

Ring diagrams. The program maps two variables or a subset variable using the same algorithm as the graduated circle routine; one variable defines the outside radius of the ring, the second an inner radius, the area between the inner and outer radius is shaded. A repetitive use for rings in different colors allows for multiple variable representation.

Graduated rectangle maps. This program is used to map two or more variables over space. One variable defines the horizontal length of the rectangle,

the other the height. Shading of the rectangles can be used to express a third - preferably nominal - variable.

Frame diagrams. The routine is a rectangular version of the ring diagrams. Attributes may be expressed by the horizontal and vertical dimensions of the frame, or the frame width. Multiple concentric frames in different shadings and/or colors are possible.

Graduated segmented circle maps (pie charts). Several attributes are used to define the overall size of a segmented circle and its subdivisions. The single sectors are represented in different shadings and/or color. The routine will also allow to map segmented rings rather than filled circles.

A variation of this program allows for tilting the circle to an ellipse to produce a pseudo-3-d effect. The ellipse may then be lifted on a stem (height proportional to an additional variable) into the third dimension.

Point symbol mapping. Figurative symbols for inventory mapping may be graduated in size according to variable values. Includes an algorithm for optimum placement within the polygon in question.

Contour mapping. A version of the program CONTUR (by F. Rens) will be implemented that access the variables and centroids in SUAF, interpolates a regular grid and constructs contour lines.

3-d mapping. A version of TRID (by F. Rens, modi-

fied by C. Reed) is being implemented. The program interpolates a regular grid and displays traverses along column, rows, or diagonals. Options include variation of viewing azimuth and height, horizontal and vertical scale, continuous or histogram version. As an alternative the program CENVUE (Tobler, 1973) is considered to be interfaced to the system. Produces 3-d block diagrams on a polyhedron (rather than grid) base.

Fence maps. Attributes assigned to statistical units can be represented by a pseudo 3-d view of the area bound by a fence where the fence height is proportional to the variable value. Options include viewing direction, planimetric and vertical scale, etc.

Dot maps. The mapping is based either on single crime events or on aggregated data. The dots are placed either at a location near the crime occurrence or in random fashion within the block groups (or census tracts). The latter procedure needs a random dot placing algorithm which avoids super-position of dots. Mapping of multiple variables in various colors is planned.

Cartograms. Implementation of area distortion cartograms (Tobler) or non-contiguous cartograms is planned. Both will be based on the display of one attribute for each statistical unit.

4. CONCLUSIONS

The Buffalo Mapping System as part of the Crime Analysis and Research Package (CARP) will be a powerful tool to generate several types of maps for research and production purposes. At the present time it is being implemented on the CDC CYBER 173 at the State University of New York at Buffalo. The Buffalo Criminal Data Record system is fully implemented where the Multiple Access Crime Files for the years 1971-1975 have been constructed. The data aggregation routines together with the aggregation file SUAF is operational. The spatial reference file TOPARF is available for mapping purposes in a preliminary form, and a number of the mapping routines have been implemented. In essence, then, the data base structure is completed and present emphasis is on implementing the mapping routines and extending the options of the currently operational display programs.

This paper has been concerned primarily with the design concepts and the technological aspects of the present mapping system. Future concern, however, will not be directed to the particular technical structure of the system, but to its versatile mapping potential based on multiple input and multiple display capabilities. This future research will provide for a systematic production and comparison of various types of maps using several cartographic methods.

After the stages of implementing the data base structure and incorporating the various mapping routines in the system, a phase of systematic computer mapping will follow. That phase will not just be restricted to the physical production of maps but will include a thorough evaluation of these various products as far as design, read-

ability, and applicability is concerned. Such an evaluation will include surveys with various user groups.

ACKNOWLEDGMENTS

The authors want to express their thanks to the Buffalo Police Department for their interest and cooperation in providing the data files. Mr. Frank Rens, Mr. Carl Reed, and Ms. Donna Peuquet have made some of their programs available to the system. Among the students who contributed through programming efforts: Bruce Eaton, Barry Glick, Wendy Ormont, and Tim Young. The latter also drew some figures of this report. The project has greatly benefited from suggestions by Dr. D. Marble. Their contributions are gratefully appreciated. The project has been supported through grants from the Research Foundation of the State of New York and the SUNY Buffalo Institutional Funds Committee.

REFERENCES

- Bertin, Jacques, 1967. Semiologie Graphique. Hague, Netherlands: Mouton. 426 pgs.
- Brassel, K., 1975. "Efficient Visualization of Complex Data Sets" Progress Report of a Research Project, unpublished manuscript, 16 pgs.
- Bureau of the Census, 1970. Census Use Study, Report No. 4: The DIME Geocoding System, 38 pgs.
- Hanson, Perry O., "A Geographic Information System for Criminal Research in Buffalo, New York", Proceedings of the Middle States Division, Assoc. of American Geographers, (8,1974) 51-55.
- Hanson, Perry O., K. Brassel and J.J. Utano, 1977. A Crime Analysis Research Package (CARP): A Strategy for the Display and Analysis of Spatially Referenced Crime Data, Technical Papers R-77/4, Geographic Information Systems Laboratory, State Univ. of New York at Buffalo, 13 pgs.
- Hanson, Perry O. and Jack J. Utano, "The Utilization of Police Information Systems - A Research Strategy" forthcoming in Review of Public Data Use.
- Laboratory for Computer Graphics and Spatial Analysis, Lab-Log, Harvard University, Graduate School of Design, January, 1977.
- Peucker, T.K. and N. Chrisman, April 1975. "Cartographic Data Structures" in The American Cartographer, Vol. 2, No. 1, pp 55-69.
- Rens, Frank, 1975. Private correspondence.
- Tobler, Waldo, 1973. "Choropleth Maps Without Class Intervals?", Geographical Analysis, Vol.5, 262-265.