e. Effects of victim characteristics on case processing. Describing the characteristics of the victims of various crimes is the first step in looking at the interaction of victims, criminals and the criminal justice system. Examining the effects of the characteristics of the victims on the actions of the criminal justice system is the next step. Two specific questions can be addressed from the PROMIS data:

- How do the characteristics of the victims affect the processing of a case once a defendant is arrested?
- Do these victim characteristics make a difference after controlling for other factors?

In examining the effect of the victim's characteristics on case processing, ideally one would want to examine each person victimized in a criminal incident. However, because only one victim is recorded in the old PROMIS system and this victim is not necessarily the same victim described in each of the codefendant's case records, court cases involving one defendant will be the unit of observation for this study. First, papering and conviction rates will be computed for all the cases involving a victim. These rates will be the dependent variables. Next, the effect of the following independant variables can be tested:

- Personal characteristics of the victim--age; sex; whether employed; whether victim has a physical disability, has used opiates, and/or has a history of chronic alcohol abuse.
- The relationship between the victim and defendant.
- Whether the victim has an arrest record.
- Whether the victim provoked the defendant or participated in the offense.

The variables will be tested separately to see their effects using the crime categories involving personal victims. Next, the variables found to determine prosecution and conviction in the analysis of prosecutor performance can be included in a multiple regression model as control variables to see if the victim's characteristics still have an effect after controlling for these other determinants. The following are hypotheses concerning these relationships.

(1) Victimologists, beginning with Hentig, have hypothesized that the weak are more likely to be victimized--the old, the very young, females, etc.[24] While this hypothesis cannot be tested with arrest data, one would suspect that when the victim is considered defenseless or helpless, this would arouse more anger toward the defendant and more sympathy for the victim. Thus, the following groups would be more likely to have their cases prosecuted: the very young, the very old, females, and persons with a physical disability. This hypothesis would not hold for persons who have used opiates and persons who are chronic abusers of alcohol, since these persons can be seen as responsible for their situation.

(2) The closer the relationship between the victim and the defendant, the less likely the case will be prosecuted or end in conviction. There are several reasons why this might be true: the prosecution believing that the family members or friends who have been victimized will ultimately not be willing to testify may be reluctant to prosecute the case. Another possibility is that the prosecution may believe it is preferable to leave matters between family and friends outside the framework of the criminal courts.

[24] Hans Von Hentig, The Criminal and His Victim (New Haven: Yale University Press, 1948).

(3) If the victim has an arrest record, this will cause his case to result in prosecution or conviction less often. As a witness, he is impeachable if he has had a felony conviction in the past ten years, or a conviction for a misdemeanor involving "moral turpitude"--for example, larceny. It will be assumed that some proportion of the victims with arrest records are impeachable.

(4) If the victim provoked the defendant or participated in the offense, it would be less likely that the case would result in conviction. Mendolsohn constructed a typology of victims based on the degree to which the victim was responsible for his own crime.[25] If a victim is actually labeled as having provoked the defendant or participated in the offense, the prosecutor would probably be less likely to pursue the case.

To complete the research under this area will require five steps:

- A literature search on situational variables associated with various types of criminal behavior should be conducted.
- Situational variables available in PROMIS should be defined.
- Other situational variables available from the case jacket, but not presently recorded in PROMIS, should be examined to see which ones might be worth recording in a small sample study.
- Hypotheses should be developed as to the relationship of the situational variables to the papering and conviction rates.
- Multivariate analysis of the situational variables in terms of their relationship to the papering and conviction rates should be conducted.

6. Predicting Recidivism with PROMIS data

a. Introduction Due to the overcrowding of calendars and the overburdening of prosecutive and adjudicature resources in the urban courts of the United States, there is a need to assign priorities to individual

---

25 Discussed in The Victim and His Criminal by Stephen Schafer (New York: Random House, 1968), pp. 42-43.

cases. Limited resources demand careful answers to these questions: Which cases should receive more attention? What is the most effective way of handling particular types of cases? Such questions force one to examine the goals of the criminal justice system. Cases should be handled to produce what result, what effect?

One criterion for judging the effectiveness of the criminal justice system is its ability to reduce crime. Basically, there are two ways of accomplishing this goal. One is to reduce the number of persons who commit a crime for the first time and the other is to reduce recidivism among those who have already committed a crime. Since court cases involve arrested persons (some proportion of whom, it may be assumed, have indeed committed a crime), an appropriate way of deciding how to assign priorities in case handling is to use the goal of reducing recidivism. However, actions taken to reduce recidivism may produce an increase in the number of persons committing a crime for the first time. A recent article by Cook points out that if recidivism is reduced by providing effective vocational rehabilitation in prisons, for example, any deterrent effect of incarceration on the general public may be lost.[26]

If the reduction of recidivism through effective policies of the criminal justice system is set as a goal, three basic research questions emerge:

- When a person is screened for prosecution, what is the best prediction that can be made as to whether he or she will recidivate?
- How do the actions of the police, the prosecutor, the defense counsel and judge during the processing of a case affect recidivism, controlling for the presonal characteristics of the defendant and the defendant's criminal history?

---

26 Phillip Cook, "The Correctional Carrot: Better Jobs for Parolees," Policy Analysis, 1, No. 1 (Winter 1975), p. 49.

- For persons sentenced to a period of incarceration, do any of their experiences with the corrections system affect the probability of recidivism, controlling for personal characteristics, characteristics of the defendant's criminal history, and the actions taken during the processing of a case?

The first question is aimed at developing a tool which will be useful to prosecutors in determining which cases have defendants with a greater liklihood of recidivating. The second question allows the different actors in the handling of a case to evaluate their decisions with respect to particular types of defendants. If it can be determined empirically, what affect particular decisions made by the prosecutor or judge have on particular types of defendants, policy can be revised (within the existing moral and legal framework) to maximize the probability that the defendant will not recidivate. As for the third question, the effect of the defendant's experiences within the corrections system on recidivism can be used to evaluate treatment programs, again with the goal of reducing recidivism.

The information needed to answer these questions can be derived from the empirical data on defendants as they pass through the court system, available in PROMIS. In addition, sentencing data from the court (which have not always been available in PROMIS in the past) and data from D.C. Department of Corrections can be used to complete the analysis. This section describes how the analysis of recidivism will be structured to make the best use of the data available in Washington, D.C.

b. Previous research. Although recidivism is a topic which has received a great deal of attention by criminologists, the focus in terms of prediction has most often been on who is likely to commit another crime

after incarceration or after receiving a particular sentence. Parole studies have been numerous, beginning as early as 1923, with Warner's study of the success or failure of parolees in Massachusetts.[27] The general form of these early studies was to analyze the association between characteristics of the criminal, the crimes for which he had been convicted, his institutional experiences, and recidivism or "parole failure." The methodology of the early studies was simple, with multiple regression analysis, which controls for the influence of many independent variables simultaneously, only appearing as late as the 1950's. This method of looking at the success or failure of a given group of parolees became known as the "base expectancy" approach. At first, the scores developed added a point for each characteristic a defendant possessed which was positively associated with recidivism. Later, the results of multiple regression analyses were used to form weighted scores predicting parole success or failure.

PROMIS currently contains such a weighted score for the defendant based on the results of a parole study in California.[28] The score is computed based on information collected at the initial screening of the case by the prosecutor and is weighted as follows:

---

27 Early parole studies are discussed and summarized in Hermann Mannheim and Leslie T. Wilkins, Prediction Methods in Relation to Borstal Training (London: Her Majesty's Stationary Office, 1955).

28 Donald M. Gottfredson, "The Base Expectancy Approach," in Norman Johnston, et al. (eds.) The Sociology of Punishment and Correction, 2nd ed. (New York: Wiley, 1970), pp. 807-13.

Question	Weight
Arrested in past 5 years	+ 10
Number of previous arrests (if greater than one)	+ 5
Number of previous arrests for crimes against persons (if greater than one)	+ 5
First arrest for auto theft	+ 2.5
Indication of opiate use at any time	+ 5
Indication of alcohol abuse	+ 2.5
Alias ever used	+ 2.5
Was present job held for less than 6 months	+ 2.5

It was originally added to PROMIS in order to provide an index of the seriousness of the defendants' histories in misdemeanor cases and has not been tested to see how well it predicts the future behavior of defendants in Washington, D.C. One phase of the research is to revise the score, possibly adding information about the effect of the characteristics of the current case. It should be remembered that, although in addressing the first research question the same approach as in previous parole studies is being used, the population being studied is different. Rather than incarcerated persons, PROMIS contains data on persons arrested. The purpose in trying to predict their behavior is to allow the prosecutor to concentrate more resources on the persons most likely to recidivate.

The previous research relating to the third research question of predicting recidivism after correctional treatment follows in the tradition of these same parole studies. However, the second question concerning the effect of actions taken during case processing on recidivism is a relatively new area. Sentencing decisions are the only actions which have been

thoroughly researched. Hood and Sparks summarize these studies as having four basic conclusions:

- Probation is likely to be at least as effective in preventing recidivism as an institutional sentence.
- Fines and discharges are much more effective than either probation or imprisonment for first offenders and recidivists of all age groups.
- Longer institutional sentences are no more effective in preventing recidivism than shorter ones.
- The offenders most likely to improve are the "medium risks."[29]

Although many have recognized the potential in studying the effects of actions taken during case processing, the data needed to do such a study have generally not been available. The President's Commission on Law Enforcement and Administration of Justice called for such work to be done when it wrote in 1967:

> A question to be explored is whether the rearrest probabilities and the crime-type distributions become worse for those who are processed further through the court system. If that is the case, it may result either from differences among individuals who reach the various stages or from the treatment itself. Unfortunately, data to examine such basic questions do not now exist . . .[30]

The hypothesis that rearrest probabilities increase the further a defendant moves into the criminal justice system can be tested with PROMIS data.

---

[29] Roger Hood and Richard Sparks, Key Issues in Criminology: (New York: McGraw-Hill, 1970), pp. 186-191.

[30] Marvin Wolfgang, et. al., The Sociology of Crime and Delinquency (New York: Wiley, 1970), p. 113.

c. A model of recidivism. In an attempt to structure the analysis to predict recidivism, a model was developed (See Figure 1). This is a simplified model, with categories of variables represented by boxes on the diagram. The probability of a defendant recidivating is the dependent variable in the model, with the independent variables to the left. (The measurement of recidivism is discussed in detail in Section 6d.) There are seven factors which can be seen as possibly influencing recidivism: the personal characteristics of the defendant; his previous criminal history; actions taken by the police, prosecutor, defense counsel, and judge during the criminal court process; and finally, the defendant's experiences within in corrections system.

The structure of the model hypothesizes interrelationships among the factors influencing recidivism. The personal characteristics of the defendant are hypothesized as having an effect on the defendant's criminal history as well as an independent effect on recidivism. Similarly, the defendant's criminal history may affect the actions taken by the police, prosecutor, defense counsel and judge, as well as having an effect on the probability of recidivism. Feedback relationships are anticipated between the actions taken by the police, prosecutor, defense counsel and judge, and the defendant's criminal history. The defendant's criminal history is eventually determined by the actions of the criminal justice agents; and as a criminal history accumulates, this has an effect on the actions taken by the police, prosecutor, defense counsel and judge. Since a case moves from the police to the prosecutor, defense counsel and judge, and finally into the corrections system (if there is a period of incarceration imposed), the actions of an agent at each stage in the criminal justice process affect

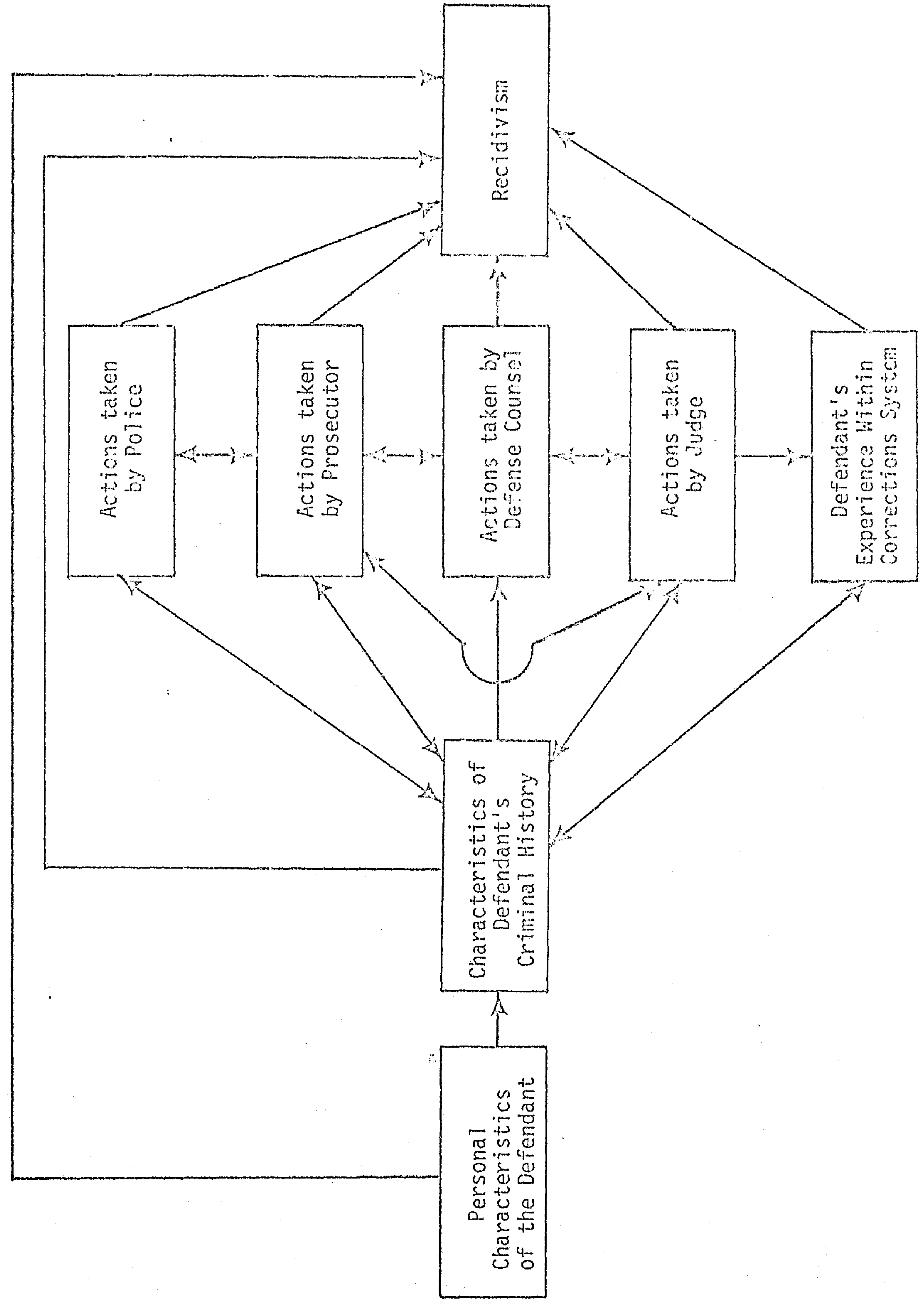


Figure 1. RECIDIVISM MODEL

the actions of the agent at the next stage. In addition, the actions of the prosecutor, defense counsel and judge are interrelated, with the actions of each having an effect on the actions of the others. Finally, whether the defendant enters the correctional system is dependent upon the sentencing decision of the judge. His time in the correctional system becomes part of his criminal history. The defendant's criminal history may affect the actions taken by any of the criminal justice agents at any point during the process.

Using this model, the analysis can be structured to answer the questions posed for predicting recidivism. The personal characteristics of the defendant and certain characteristics of his criminal history, including the most recent offense, are the pieces of information available when a case against a defendant is brought by the police to the prosecutor for screening. A separate analysis of these variables will yield the best prediction of recidivism from the data available. Additional variables available from corrections data may also suggest items which should be collected at screening due to their ability to predict recidivism.

The second research question is, given the propensity toward recidivism shown by the above analysis, what actions taken by the criminal justice system during the processing of a case can reduce this basic propensity? In many studies, the only focus is on the influence of the personal characteristics and/or criminal history of the defendant on recidivism. Since these are essentially fixed characteristics when the defendant enters the system with a new case, it is difficult to make any policy recommendations other than to concentrate more resources on those most likely to recidivate. On the other hand, if it can be determined that certain actions of the prosecutor are successful with a certain group of defendants, for instance, first

offenders under 25, this finding could be used to recommend specific actions to be taken in the handling of these individuals.

The third question about the effects of incarceration can be addressed by analyzing only persons sentenced to some period of incarceration. All of the other preceding variables can be included in the analysis, in order to see if any treatment by the correctional system makes a difference.

d. Measuring the dependent variables of recidivism. There are three problems in defining the dependent variables of recidivism in this analysis: what event is to be considered a failure--a rearrest, a reprosecution or a reconviction; how can the seriousness of the recidivistic events be accounted for?

Defining how far a defendant must move into the criminal justice system before being considered a recidivist is important. As Blumstein and Larson point out, if arrests are used, someone may be included as a recidivist who did not actually commit another crime; whereas if convictions are used, many persons who did commit crimes will not be included.[31] Since most of the predictive studies of recidivism have been with parole data, reconviction is usually the measure used. Rather than choosing one definition or another, rearrests, reprosecutions and reconvictions will each be analyzed separately as dependent variables. It is expected that each of these measures will yield slightly different results, but the direction of the relationships is hypothesized to be the same. For instance, if the age of the defendant has a negative effect on the probability of rearrest,

[31] Alfred Blumstein and Richard C. Larson, "Problems in Modeling and Measuring Recidivism," Journal of Research in Crime and Delinquency, 8, No. 2 (July 1971), pp. 124-125.

it would also be expected to have a negative effect on the probability of reconviction.

Many researchers have dealt at least theoretically if not empirically, with the problem of the seriousness of the recidivistic act. If a defendant can move from committing armed robberies to an occasional misdemeanor this can be seen as improvement in some sense. Glaser makes this point when he states:

> Recidivism is measured in terms of one rearrest, reconviction or reimprisonment, although those thus classified as recidivists differ tremendously in the immediacy, extent and seriousness of their renewed criminal behavior.[32]

An example of a study which did take the seriousness of the subsequent crime into account was conducted by Locke, et al. The study analyzed pretrial release in Washington, D.C., by using a criterion of "dangerousness" based on certain elements of the offense.[33] PROMIS contains the Sellin-Wolfgang score of each offense, which is a measure of the seriousness of an offense based on how many persons were injured, hospitalized, the amount of property taken, etc. The score was developed by asking citizens and criminal justice practitioners to rate the seriousness of crimes and then isolating the elements of the crime which were associated with higher ratings. The seriousness score can be used as a scalar dependent variable in a regression analysis, since it ranges from 0 to 80 and higher. For those persons who recidivate, the dependent variable can be the seriousness

32 Daniel Glaser, Pretrial Release with Supportive Services for "High Risk" Defendants, Monograph, National Center for Crime and Delinquency, 1973, p. 22.

33 Locke, J.W., et al., "Compilation and Use of Criminal Court Data in Relation to Pre-Trial Release of Defendants," National Bureau of Standards, Technical Note 535 (Washington: Government Printing Office, 1970).

score of the second case.

The problem of measuring the frequency of recidivistic acts involves giving each person an equal amount of time to recidivate. This problem has led to the application of various approaches of operations research to recidivism, such as the use of failure rate analysis.[34] The disadvantage of using failure rate analysis with PROMIS data is that it becomes difficult to include the richness of the control variables available for each defendant. Since a failure rate is a group measure computed as failures over total exposure time, programs are analyzed without controlling for the "mix" of persons in each program. Another approach to equalizing each defendant's chance to recidivate is to set a specified time period for recidivism. Either a person recidivates or does not in the time period set, regardless of his behavior after the time period.

There is still another complication in giving defendants a fixed period of time to recidivate: each person must be able to recidivate. If someone is incarcerated, he or she cannot be compared to a person who is living in the community. Corrections information can be used to provide data on when defendants are in jail or prison. Whatever the period of time each defendant is given to recidivate, time must only be counted for each day the defendant is "free."

In summary, there will be three types of dependent variables:

- Counts of the number of rearrests, reprosecutions, and reconvictions for the panel of defendants as they are followed in PROMIS.

34 Stephen Stollmack and Carl M. Harris, "Failure-rate Analysis Applied to Recidivism Data," Operations Research, 22, No. 6 (November 1974), pp. 1192-1205.

- Dichotomous measures of whether or not a person recidivates by being rearrested, reprosecuted, or reconvicted within a given period of time.
- Measures of the seriousness of the subsequent case based on the characteristics of the previous case.

One of the contributions of the study will be in analyzing the differences, if any, among the results obtained from using these various measures.

e. Independent variables. Since data have already been collected in PROMIS, there is not the same process of choosing independent variables as there is in most research. Of course, there is some choice as to the structuring of the variables. Fortunately, some information is available for each of the factors identified in the model. In addition to PROMIS, supplemental information is being sought from the D.C. Superior Court on sentencing decisions and from the corrections department in regard to treatment of persons incarcerated. The sentencing data will have to be hand coded since they are in free-format, literal form on the court's tape. The D.C. Department of Corrections has a data system (CRISYS) which can be used for two purposes: first, to provide information on persons incarcerated which will be used as independent variables in the analysis, and second, to yield information on whether a person is in jail or out in the community and "eligible" to recidivate. The independent variables available for testing in the model are shown in Table 4.

f. Hypotheses. Since there are three research questions, there are three sets of hypotheses concerning the prediction of recidivism. Each will be discussed separately below. Some hypotheses are based on previous research and are an attempt to replicate earlier studies. Others are hypotheses which have not been tested previously.

TABLE 4. INDEPENDENT VARIABLES FOR RECIDIVISM ANALYSIS

I. Personal characteristics of the defendant
   - Age
   - Race
   - Sex
   - Neighborhood of residence
   - Defendant employed at time of arrest
   - Length of time last job was held
   - History of chronic alcohol abuse
   - History of opiate use
   - Marital status*
   - Number of dependents*
   - Ability to read and write*
   - Education*
   - Occupation*
   - State of birth*

II. Characteristics of defendant's criminal history
   - Arrested in past five years
   - Whether defendant has used an alias
   - Number of previous arrests
   - Number of previous arrests for crimes against persons
   - First arrest for auto theft
   - Time since most recent arrest
   - Number of prior convictions in the District of Columbia
   - Characteristics of the current case:
     - Type of case
     - Victim-defendant relationship
     - Misdemeanor/felony
     - Case seriousness (Sellin-Wolfgang Index)
   - Number of codefendants
   - Number of witnesses

III. Actions taken by police
   - Case dismissed due to police malpractice
   - Time between offense and arrest

IV. Actions taken by prosecutor
   - Case papered
   - Case specially assigned to Major Violator's Unit
   - Case nollied
   - Successful diversion into special program
   - Acceptance of plea bargain
   - Number of continuances
   - Time between arrest and final disposition

V. Actions taken by defense counsel
   - Defense attorney type
   - Number of motions filed
   - Number of continuances

VI. Actions taken by judge
   - Bail decision
   - Dismissal of case
   - Finding of guilty
   - Sentence imposed

VII. Defendant's experience within the correctional system
   - Type of institution where incarcerated*
   - Time spent in each institution*
   - Number of transfers*
   - Participation in special programs (educational, vocational, therapy, counseling)*
   - Number of days "good time" gained*
   - Number of days "good time" lost*
   - Number of escapes*
   - Conduct record*

* Data available from the District of Columbia Correctional Data System (CRISYS).

not been tested previously.

(1) The first research objective is to determine the best prediction of recidivism from the information available when a defendant comes in for screening. The available information consists of personal characteristics of the defendant and characteristics of his criminal history. Below are listed the hypothesized directions of the relationship to recidivism for each variable. It should be noted that evenhandedness is as important a goal for criminal justice as the reduction of recidivism. Some variables, such as race and sex, are included in the analysis although they would not be used as items in a revised score.

	Direction of Hypothesized Relationship to Recidivism
I Personal Characteristics of the Defendant	
Age	-
Race (Black)	+
Sex (Male)	+
Defendant employed at time of arrest	-
Length of time last job held	-
History of chronic alcohol abuse	+
History of opiate use	+

An additional variable available on personal characteristics will be the neighborhood of residence of the defendant. It is expected that defendants living in neighborhoods with high concentration of offenders will be more likely to recidivate due to their opportunity to associate with persons committing crimes.

Characteristics of the defendant's criminal history are expected to be the most important determinants of recidivism:

	Direction of Hypothesized Relationship to Recidivism
II. Characteristics of the Defendant's Criminal History	
Arrested in past five years	+
Has used an alias	+
Number of previous arrests	+
Number of previous arrests for crimes against persons	+
First arrest for auto theft	+
Time since most recent arrest	-
Number of convictions	+
Characteristics of the current case:	
Misdemeanor rather than felony	-
Case seriousness (Sellin-Wolfgang Index)	-
Number of defendants	+
Number of witnesses	+

Two additional criminal history variables are the type of the current case and the relationship between the victim and the defendant in the current case. Using the results of previous studies, it is hypothesized that those arrested for personal victimizations involving violence, including homicides, assaults and forcible sex offenses, will be the least likely to recidivate. Those arrested for auto theft, robbery, and burglary will be most likely to recidivate. The hypothesis concerning the victim/offender relationship is that the closer the relationship between the victim and the defendant in the current case, the less likely the defendant will recidivate.

(2) The second research question involves the evaluation of particular actions taken during case processing by the police, prosecutor, defense counsel and judge. The personal characteristics of the defendant and the defendant's criminal history will be included as control variables. Two general hypotheses to be tested are: the further a person moves into the criminal justice system the more likely he will be to recidivate, and inefficiency in the handling of the defendant's case, as measured by time delays and continuances, will increase the likelihood of recidivism.

	Direction of Hypothesized Relationship to Recidivism
III. Actions taken by Police	
Case dismissed due to police malpractice	+
Time between offense and arrest	+
IV. Actions taken by Prosecutor	
Case papered	+
Case specially assigned to Major Violator's Unit	+
Case nollied	+
Successful diversion	-
Acceptance of plea bargain	+
Number of continuances	+
Time between arrest and final disposition	+
V. Actions taken by Defense Counsel	
Number of motions filed	-
VI. Actions taken by Judge	
Dismissal of case	+
Finding of guilty	+

The sentencing decision of the judge is also expected to have impact on recidivism. It is important to differentiate between short-term and long-term effects. In the short-term, incarceration may be effective in preventing a person from recidivating by keeping him or her off the streets, but may increase his or her chances of recidivating after confinement. Both effects can be examined. In the long-term, it is hypothesized that the sentences will range from most effective to least effective in the following order: special programs, fines, probations, short periods of incarceration, suspended sentence or unsupervised probation, and finally long periods of incarceration.

(3) The third research question concerns the effect of correctional programs and policies on recidivism. The population studied will be all persons sentenced to a period of incarceration, with the independent variables already mentioned included in the analysis as controls. Two of the most important independent variables will be the type of institution where incarcerated and the length of time incarcerated. It is hypothesized that persons incarcerated in high security prisons will recidivate more often when released, and that inmates from an institution with a lower ratio of professional to nonprofessional staff will recidivate more often than those from an institution with a higher ratio. Overall, the longer a person is incarcerated the more likely he or she will be to recidivate.

Within an institution, special correctional treatment programs can be evaluated. Special programs are hypothesized as having a beneficial effect, with persons who participate in them having lower recidivism rates than those who do not. Those who complete programs are predicted to recidivate less than those who drop out. A good conduct record, no escapes, and a high ratio of "good time" days gained over "good time" days lost are each hypothesized as having a negative effect on recidivism.

g. Panel design. The basic design of the recidivism analysis is to follow a panel of defendants in Washington, D.C. Criminal histories of defendants appearing in PROMIS will be gradually accumulated and analyzed. By following the defendants in PROMIS, one can continually monitor whether any actions taken during the processing of a case makes any difference in terms of predicting recidivism. One can focus on the differences between those who recidivate and those who do not. If the background of a panel group is analyzed, on the other hand, one does not know what happened to the group similar to the panel group who did not get involved in crime.

Some analysis will be completed on the prediction of recidivism using defendants who had at least one case in 1973. This panel of 1973 defendants can then be followed into 1974 and 1975. The advantage of beginning with 1973 data is that PROMIS has more complete information for this year than was collected in 1971 and 1972. However, a sample panel will also be formed of 1971 defendants who will then be followed into 1972, 1973, 1974 and 1975. The advantage of this smaller sample is that it will allow a five-year rather than a three-year follow-up period. After the three-year or five-year period, a standard procedure for constantly updating the score developed for predicting recidivism can be implemented.

The tasks to be completed in this area of research are:

- A literature search should be conducted on definitions of recidivism and ways of predicting recidivism.
- Definitions of recidivism should be formed which can be used with the PROMIS data base.

- The Gottfredson scale should be tested to see if it predicts recidivism.
- A new model based on variables within the four large hypothesized factors in recidivism should be developed and tested to see if it will predict recidivism better than the Gottfredson scale.

CHAPTER 5. DATA PURIFICATION FOR RESEARCH PURPOSES

A. Introduction

The process of "purifying" data should be important to the researcher who analyzes the data. This is particularly true when the end product of that research is a set of recommendations as to specific policy changes that should or should not be implemented. Quite often, such recommendations are based on "point estimates" of the amount of improvement in performance or output that a given policy change can be expected to produce. If these estimates are in error due to faulty data, implementation of the research recommendations can have a pernicious effect on the true performance of the system being analyzed.

In some cases the data errors will be randomly distributed over the observations that form the basis for policy recommendations. This type of error tends to obscure measurement of the true degree of association that exists between the factors under investigation. In other instances, the errors are systematically distributed in some way over the observations. This type of error often produces biases or distortion in the inferred relationships; hence, this latter type of error tends to be more harmful than the former. In either case, the determination must be made balancing the cost of data purification and the cost of implementing an erroneous research finding.

In light of the researcher's concern for reliable data, the research of the PROMIS data must include a thorough attempt to purify data and to estimate the reliability of each data element that will be used for analysis. However, the costs of data purification must be a significant factor when defining the scope of the effort.

B. PROMIS Data Base

PROMIS is designed primarily as an operational-based system to collect longitudinal data on cases and defendants as they flow through the prosecutorial system. The primary objective of this data gathering is the production of reports to aid the prosecutor in the preparation and processing of cases. Therefore, the PROMIS data base was developed more for efficient input/output processing of data which included textual information. Even though operations was the primary concern, the PROMIS data base does include a wealth of data for the researcher.

The PROMIS data base as it was formatted in 1973, the year chosen for the research data base, was not conducive to the application of standard statistical packages for statistical analysis. This was due to the following data base design characteristics:

- The duplication of data elements for multiple charges in a case, which is reflective of the case itself rather than each charge (e.g., continuances, case evaluation information, and witness information), makes it difficult to tabulate data accurately.
- The presence of alphanumeric data to describe certain variables creates inefficiency in using generalized statistical packages.
- The separation in certain situations of cases against one defendant that resulted from one criminal incident (e.g., indicted felonies, breakdowns).
- The lack of summary data to describe the status of a case requires additional processing in order to summarize dispositions and continuances for analysis.

The original PROMIS was recently redesigned and reprogrammed to improve upon the collection and storage of data. The new PROMIS design includes the following improvements:

- Duplication of case data has been eliminated
- Cases arising from the same criminal incident for the same defendant have been combined in the same physical location on the data base
- Numeric codes to summarize variables or data elements are stored rather than alphanumeric data, thus increasing the utility of the data for applying standard statistical packages to the data
- Summary fields have been designed to eliminate additional processing to determine the status of a case
- The editing and validity checking of the input data have been improved to enhance the reliability of data
- Error detection and correction procedures and updating of case data have been improved to facilitate the modification and purification of data.

Many of the improvements in the new system have been spawned by the witness cooperation study. This study provided an initial attempt in analyzing the PROMIS data and pointed out the necessity for improving PROMIS to enhance the research capability of the data.

The analysis of the data will be enhanced by reformatting the history data into the new PROMIS format. By doing this, the new software will be utilized for editing and updating research data during the data purification stage. In addition, the data base in the new format could be used more extensively than the old for conducting multivariate analysis.

There are approximately 50,000 case histories in the PROMIS data base collected since January 1, 1971. This volume alone precludes the use of the entire universe of cases for conducting purification and research due to the costs that would be involved. Only a sample is needed for research purposes. In addition, the early stages of data collection did not include the full set of data that will be required for analysis.

It has been determined that the cases referred to the Superior Court Division for prosecution in 1973 will constitute the sample frame for analysis. Not only does 1973 represent the most recent full year of data, but it was a year in which the data were most accurately and completely collected.

The use of 1973 cases would not only limit the volume of cases that would undergo data purification but would provide the researcher with the flexibility for modifying sampling strategies within the year's data. Within the 1973 cases, there are some data elements which are more important for research than others. In fact, there are several fields of information (e.g., defendant's name) which will not be used directly for analysis. Therefore, a priority list of elements was developed which rates each data element for research utility (Attachment V). The elements in this chart were assigned a priority number between I and IV. If a data element receives an "I," this means that the element is regarded as crucial for research; "II" implies that it is important; "III" implies that it might be useful; and "IV" suggests that it does not appear likely to be incorporated in the analysis.

Limiting the scope of the data purification task to 1973 cases and cleaning data which have a high priority will result in a purified data base for research becoming available sooner. At the same time, some data elements will be identified as "sufficiently reliable" (e.g., error rate less than 5%) to be used for research without further purification.

Limiting the sample to 1973 cases also enables the systems staff to utilize the new PROMIS software for reformatting and editing data. By extracting 1973 cases and processing these cases through the new software, which includes a more stringent edit, much of the data could be reedited by the system without a tedious attempt at manually identifying errors. In addition, the new software could then be used to easily update errors that are identified by the system. The end result of this data recycling would provide the researcher with a more purified set of data in order to conduct a manual comparison of data to source information, to apply various sampling techniques, and to begin generating frequency tabulations for describing the case workload of the Superior Court Division for the entire year of 1973.

The sampling of data from the universe of 1973 cases can be broken down into two stages for reliability analysis and preliminary research. Stage I will serve at least three purposes: first, it will provide information that will enable the researcher to determine on which variables the Stage II sample should be stratified; second, it will give the researcher an initial assessment of the reliability of the data elements on the PROMIS master file in order to determine the fields usable for research as they now exist on the files so that tabulations of total 1973 caseload statistics can be interpreted; third, once these data are cleaned, they will serve as a small, sample, purified data base for preliminary in-depth research.

To adequately meet these ends, it was decided that Stage I consist of a simple random sample of 100 cases. Each data element would be manually

checked with source documents and other available information in terms of omission rate, rates of discrepancy with source documents, internal inconsistency rate, and other appropriate reliability measures. It will be necessary also to have a not-applicable rate for each data element to reduce the incidents of frequency counts for blanks.

Stage II will be a stratified random sample, with the stratification plan to be determined on the basis of information obtained from both the Stage I sample and the summary frequency tabulation. The number of strata and the size of the sample within each stratum in Stage II can be determined after the results of Stage I.

The summary frequency tabulation in all 1973 cases will serve at least two useful functions: like Stage I, it will provide information on which to base the Stage II stratification scheme; it will also provide an overall picture of what is going on at the Superior Court, which is part of the first-year research requirement, and, of course, valuable in its own right as a means of understanding the system. This tabulation will be interpreted together with the Stage I sample, which will provide preliminary information about the extent to which the data are reliable. The items desired on this summary are listed in Attachment VI.

C. Specific Tasks

The purpose of the following tasks is threefold:

- Produce a file of all 1973 cases from which frequency counts can be generated. This file will be available upon completion of Task 2.
- Produce a purified file of 100 cases randomly selected from all 1973 cases by cleaning up all known errors in these cases through comparison to source documents. This file will

be available upon completion of Task 6.

- Produce a purified file of 500 cases selected from all 1973 cases based on frequency counts. This file will be available upon completion of Task 10.

1. Task 1. Convert 1973 cases into input transactions for the new PROMIS software.

a. General Description. The PROMIS master and history tapes will be run through two programs to create input transactions for the new PROMIS software. These two programs were designed and tested to provide test data for the new PROMIS. All transactions for the new system are generated except for the EVALUATION transaction which had not yet been defined when these programs were designed. A brief description of these two programs and the modifications which will be made to them follows:

b. PROM3. This program produces the CASE, GRAND JURY, BREAKDOWN, WITNESS and SPECIALLY ASSIGNED transactions. A minor modification will be made so that transactions will be generated only for those cases considered for prosecution in 1973 and those cases arising from 1973 cases. In addition, this program will be expanded to produce the EVALUATION transaction. In conjunction with this, a separate output file will be produced with PDID, COURT CASE key which will contain any fields which were on the EVALUATION form for the old PROMIS but were omitted from the EVALUATION form for the new PROMIS.

c. CONVERT1. This program produces the ARREST, ARRAIGNMENT, FINAL DISPOSITION, SENTENCING, and CONTINUANCE transactions. A minor modification will be made so that transactions will be

generated only for those cases considered for prosecution in 1973 and those cases arising from 1973 cases. Since it has been determined that the INSLAW Research Group requires sentencing information on indicted felonies (the only type of case for which it has been collected), a modification will be made to translate the sentencing information into transaction form.

2. Task 2. Run 1973 transactions through new PROMIS edit and update cycle and correct all errors detected by new PROMIS software.

a. General Description. The two transaction files produced in Task 1 will be run through the new PROMIS software to create a PROMIS master file of all 1973 cases in the new PROMIS format. This task will be divided into several subtasks.

b. Task 2A. Run 1973 transactions through the new PROMIS software. The output of Task 2A will be:

- A master file of cases which satisfied the editing criteria in the new PROMIS format.
- A list of cases which had editing errors. This list will distinguish between those transactions which had a major error that caused the entire case to be rejected (i.e., not retained on the master file) and those transactions which had a minor error (e.g., an invalid, nonrequired field) where the case was retained in the master file.
- A batch corrections error file to facilitate error transactions. When the field in error is corrected, the entire transaction will be reprocessed.

c. Task 2B. Manually correct all errors which caused a case to be rejected and as many of the other errors as possible. This will involve locating the source documents for these cases and coding corrections onto the error list. Corrections will then be keypunched from the error list which has been designed to be directly coded on and keypunched from. Based on a sample of 1973 cases run through the new software, it is estimated that 5% of 1973 cases will be totally rejected. Since the number of cases papered in 1973 was approximately 15,000, this will be a total of 750 cases. It is estimated that the total number of errors will be about 1500.

d. Task 2C. Run corrections through the new software until all error corrections have been accepted.

e. Task 2D. Run statistics on number of cases papered per month and compare to manual statistics of the U.S. Attorney's Office. If they differ greatly for any month, locate and enter missing cases.

3. Task 3. Select 100 cases randomly from all 1973 cases.

a. General Description. The 1973 master file in new PROMIS format created in Task 3 will be input to a program to randomly select 100 cases for the Stage I sample. All edit errors identified for this group of cases will be corrected using the new PROMIS software. This task will be divided into several subtasks.

b. Task 3A. Identify total number of 1973 cases so that a sample formula can be developed. This number will be a by-product of Task 2 in that one of the output totals in the update cycle is

the total number of case header records on the new master file. The sampling technique will be formulated as follows:

n = number of cases to be selected = 100

N = total number of cases

f = sampling rate = $\frac{n}{N}$

k = selection interval = 1/f, rounded to a whole number

To select the sample, the initial selection number will be a random number (r) selected from a random number table, i.e., between 1 and k. The selection numbers for sample cases will be as follows:

r, r + k, r + 2k, r + 3k, r = (m - 1) k

where each selection number is rounded up to the next integer.

c. Task 3B. Mark each of the 100 cases for the Stage I sample on the master file by inserting code "01" in the branch office field (otherwise unused). In addition, the PDID number and court case of these 100 cases will be printed out.

4. Task 4. Print the Stage I sample for error correction and reliability check.

a. General Description. The 100 cases in Stage I sample will be printed in a format so that a field-by-field comparison can be made between the data in PROMIS and the data on the source documents contained in the jackets filed in the U.S. Attorney's Office. This task will be divided into several subtasks.

b. Task 4A. Identify any crosschecks which can be made between fields in the data base which would allow some blank fields to be identified as not applicable. Examples might be allowing questions such as "was there provocation by victim" or "was there victim participation" to be blank if the charge were one that implied a victimless crime such as soliciting for prostitution or possession of marijuana. These crosschecks should be carefully selected so that the file can be processed without having to save entire records.

c. Task 4B. Design and write a program to print all fields in the PROMIS master file for each of the 100 cases. The order in which the fields are printed will be, in part, dictated by the record sequence of the master file but, whenever possible, will be printed in the same order in which they appear on the three source data entry forms. Some exceptions will be necessary. For example, prior to the inclusion of the witness information on the Case Processing and Trial Preparation Worksheet in September 1973, this information was taken from the PD163. However, in all cases in the printout, witness information will appear at the end of each case, while all other information on PD163 will appear at the beginning of the case. This will require those doing the manual comparison to look at the PD163 twice when checking cases papered prior to September 1973. Sufficient space will be left on the sheets to write in corrections.

In addition, this program will label each sheet with one of three jacket locations:

- Felony Trial Section - if the case was indicted or sent to be tried with a felony (Rule 105b).
- Grand Jury - if the case was bound over to the grand jury, then dismissed or ignored.

- File Control - all other cases.

One other output of this program will be five sets of blank forms containing only the labels of the fields (in the same format as the case printout but with no date) so that the reliability checkers can indicate how often each field was correct or incorrect when compared with the source documents. If it is determined that a field(s) has a consistency error that may be correctible by software, this should separately be noted.

d. Task 4C. Print the file created in Task 1 containing those fields on the old PROMIS evaluation form but omitted from the new form so that a reliability check can be made on these fields. If, in addition, it becomes imperative to correct known errors in these fields, software will have to be developed to update this file.

5. Task 5. Manually compare Stage I sample case to source documents.

a. General Description. Using the printouts from Task 4, a field-by-field check will be made of all data in PROMIS on a case to all data in the jackets filed in the U.S. Attorney's Office. This check has two purposes: to clean up all known errors in the Stage I sample; to manually record for each field in the data base the number of times it is correct, incorrect, and missing. This task will be divided into several subtasks.

b. Task 5A. Locate the jackets for all cases. As explained in Task 4, there are three separate places where the jacket may have been filed after final disposition: Felony Trial Section (second floor of the Pension Building), Grand Jury Section (basement of Building B), and File Control Section (first floor of Building B).

c. Task 5B. Manually conduct a field-by-field comparison of each field on the source document. As part of this task, a description or training class will be developed to detail where on the source form to check for each field. Alternate sources in the jacket of each field will be detailed. There will be two outputs of Task 5B. The first will be a notation for each field for each of the 100 cases as to whether the field agreed with the source form, disagreed with the source form, or was not entered into PROMIS although on the source form. The second will be the printout from Task 4, with corrections written in next to each field found to be in error.

d. Task 5C. Allow for addition of fields to the special file created in Task 1 of fields not on the new master file but on the old one. These fields will be defined prior to the start of Task 5, then compiled when going through the individual jackets. An example might be the defendant during the processing of a case.

6. Task 6. Correct all errors found in Stage I sample.

All errors noted on the printout in Task 5 will be coded into a format so that they will be accepted as transactions to the new PROMIS software. They will then be keypunched and run through the new software until all errors are corrected. At this point, the Stage I sample file is ready for statistical analysis through SPSS, Datatext, or other statistical packages.

7. Task 7. Produce 1973 frequency counts.

This task can be started any time after the completion of Task 2. The frequency counts requested by the INSLAW Research Group will be run using a statistical package.

8. Task 8. Create an extract file from the master file for that analysis which cannot be done directly by SPSS.

It is likely that some analysis will be requested that cannot be directly done from the PROMIS master file format. One example would be studies on police and prosecutor charges by charge priority. In order to define charge priority, an indicator would have to be inserted in each case. This task would be a program to develop such an indicator and any other than might be required.

9. Task 9. Determine which fields are "sufficiently" reliable to do an analysis on all 1973 cases.

Using the forms created in Task 5B, the percentage of time each field was reliable will be manually computed. Define an acceptable reliability, then identify fields which satisfy that reliability. It will then be possible to input all 1973 cases into a statistical package to do research on these fields.

10. Task 10. Design and execute sampling plan for Stage II sample.

This task can be started any time after the completion of Task 7. The frequency counts generated in Task 7 will be used to define the stratification for the Stage II sample. The sample will consist of 500 cases. The sample will be drawn based on a formula similar to that described in Task 3A. These cases will be marked on the master file with an "02" in the branch office field. Tasks 3C through 6 will then be repeated for the Stage II sample so that all errors are eliminated. The Stage II sample will then be available for analysis through statistical packages.

CHAPTER 6. APPLICATION OF MULTIVARIATE ANALYSIS TECHNIQUES

A. Introduction

While some of the statistical analysis of the PROMIS data will consist of compiling summary information (such as means, medians, standard deviations, and percentages) on single data elements, the preponderance of analysis will consist of the application of statistical techniques that examine relationships among variables. These techniques come under the heading of "multivariate analysis." This chapter provides an overview of some general principles of multivariate analysis, a review of the major techniques of multivariate analysis that appear to have potential for application within the PROMIS Research Project, and an outline of tasks to be performed in carrying out this application.

B. Basic Principles

The first principle of multivariate analysis is that the finding of a statistical relationship between two variables, in itself, says virtually nothing about the causal mechanism that lies behind that relationship. To know, for example, that the coefficient of correlation between two variables, say X and Y, is .90 does not suggest the extent to which changes in X produce changes in Y, or the extent to which changes in Y produce changes in X, or the extent to which changes in some other variable, such as Z, produce systematic changes in both X and Y. Various combinations of all three possible causal mechanisms may be at work. Worse yet, additional variables may, in fact, have an important role, within a highly intricate network of causal relationships that fully

explain how X and Y come to be associated statistically. (It is because of this that most models which are analyzed using multivariate techniques are said to be "naive models.")

How, then, can causality be studied--by constructing and comparing plausible sets of causal relationships among the variables that comprise the system under analysis. These sets of relationships may be depicted either schematically or algebraically. Each set of causal relationships is called a "model." Alternative models are typically constructed on an *a priori* (i.e., intuitive or theoretical) basis, rather than empirically. Empirical analysis comes to bear on the problem of selecting from these alternatives the model that describes the system most accurately.

The second principle of multivariate analysis deals with the mode of inference. There are two modes--point estimation and hypothesis testing. In point estimation, it is said, based upon both the empirically observed degree of statistical association and the presumed nature of the causal relationship, that the best estimate of the effect on the dependent variable (i.e., the variable that is being explained or predicted) of a given change in an independent variable (i.e., the variable that is postulated to affect the level of the dependent variable) is such and such. In hypothesis testing, if the two variables whose relationship is being investigated were, in fact, absolutely unrelated, the probability of obtaining a statistical association at least as close as has been obtained is such and such; if that probability is sufficiently small, the hypothesis that the two variables are unrelated is rejected.

The second principle follows: point estimation and hypothesis testing are useful in multivariate analysis primarily when they are combined;

each alone is seldom of value. The point estimate says nothing about the reliability of the estimate--the observed relationship may be large but, because of an insufficient number of observations, unstable. The hypothesis test, on the other hand, says nothing about the strength of the relationship (i.e., the degree to which a unit change in an independent variable can be expected to affect a dependent variable)--the observed relationship may be "statistically significant" at some arbitrary level, but the degree of impact of one variable on another may be negligible. The relationship between virtually any two variables within a system will be statistically significant at some arbitrary level if the number of observations is sufficiently large.

The third and, perhaps, most important principle of multivariate analysis is that the value of the analysis lies not in the correctness of the assumptions on which it is based, but in the ability of the analysis to accurately describe or predict the system under investigation. Virtually every multivariate technique makes assumptions about the properties of the system under analysis (e.g., that the dependent variable is normally distributed about every value of each of the independent variables). Some techniques make more restrictive assumptions than others. Too often it is said that a particular technique cannot be used in a particular instance because one or more of the assumptions on which the technique is grounded is violated. However, this reasoning loses sight of the fundamental objective of multivariate analysis--to accurately describe relationships between variables. If one technique consistently describes or predicts a phenomenon more accurately and efficiently than a second technique whose assumptions are less restrictive to that

application, then it would be difficult to justify using the second technique to anyone who places the understanding of the phenomenon above the formality of statistics. Accordingly, this principle will be used as the primary criterion for selection among alternative multivariate techniques, wherever conflicts may exist. It is recognized that the principle must be violated in those cases where scholarly mathematical statistics take precedence over scholarly social science.

The fourth principle discourages the tendency to apply a particular technique because it is familiar or available to the researcher, rather than appropriate to the problem. Several techniques should be considered for application in multivariate analysis. Sometimes modifications must be made to apply a technique (e.g., to apply multiple regression methods to nonlinear or discrete data). Each technique, generally, has been developed for application under somewhat different circumstances. The choice of the technique should be made based on the considerations given in the third principle. To paraphrase an old maxim: "If the only tool you possess is a hammer, you will be inclined to treat everything as if it were a nail." Curing the course of the project, the full range of multivariate techniques will be considered, and more than one technique will be applied to those problems that lend themselves to a set of alternative methods.

Other principles worthy of discussion in a more exhaustive treatment of multivariate analysis shall be ignored here to place greater emphasis on the importance of the four principles discussed above.

C. Major Techniques

The techniques discussed are those having potential for application within the PROMIS Research Project. These include multiple regression analysis, Goodman's technique for analyzing relationships among taxonomic variables, factor analysis and the related method of principal components, and a set of techniques that shall be treated under the heading "Other Multivariate Techniques."

1. Multiple Regression Analysis

The standard statistical solution to the problem of predicting the level of a dependent variable when given the levels of several independent variables is to apply multiple regression analysis. ("Multiple" refers to the existence of more than one independent variable, and "regression" will be described momentarily.) In its most basic form, this technique consists of "fitting" to a set of independent variables the straight line that minimizes the sum, across the observations, of the squared distances between the observed values of the dependent variable and the values of that variable that are "predicted" by the line. These distances are usually called "residuals," and the dependent variable is said to be "regressed" on the independent variables. References on regression analyses are given at the end of this section.

Regression analysis is widely used not only to predict unknown values of certain variables based on the known values of other variables; it is used also to describe relationships between variables, so that inferences about causality are possible. It is this latter use of regression analysis that makes it particularly well suited for application

within the PROMIS Research Project since it is the effect that given changes in policy factors have upon each of several different output or performance measures that is to be examined. For example, when everything else is held constant, what effect does an increase or decrease in the time between arrest and trial have upon the conviction rate? To what extent, in turn, is this time interval affected by the sex and level of experience of the prosecutor? Multiple regression analysis enables these types of questions to be addressed, even though all other factors may not, in reality, have been held absolutely constant. Assuming that the direction of causality is at least partly from independent to dependent variable, the analysis, when properly applied, measures the observed effect that each independent variable has upon the dependent variable, after taking account of the effects that all the other independent variables have upon the dependent variable.

When pairs of independent variables in a regression equation are suspected of moving together in a systematic way (this is referred to as "multicollinearity"), these subrelationships can also be analyzed using multiple regression. In such cases, the full model will consist of more than one regression equation. The two questions posed as examples in the above paragraph serve to demonstrate that the application of regression analysis in the PROMIS Research Project is likely to consist of estimating a system of related regression equations. Alternative systems or models can be evaluated in most cases by comparing each equation across the systems in terms of standard "goodness-of-fit" statistics, such as the F-ratio and the coefficient of multiple determination adjusted by the degrees of freedom. Such statistics measure the extent of total variation in the dependent variable that is explained by the independent variables, after accounting for the number of observations and the number of independent variables.

The predictive mode of regression analysis is also planned for use in the PROMIS Research Project. Of course, recommendations that are based on the application of the descriptive or explanatory mode of regression analysis will often carry the implication that performance will improve by some predicted amount if the policies are changed in a prescribed way. A more direct application of the predictive mode lies within the plan for simulation modeling within the project. This plan includes the use of regression analysis to determine branching ratios at selected nodes within the simulation model. For example, when a simulated case flows to a stage in the court system where the defendant is to plead either guilty or innocent to a charge, the model will look at pertinent characteristics of the case (such as the defendant's number of prior convictions, his relationship with the victim, and the elapsed time in the system); it will then feed these characteristics into a previously estimated regression equation that predicts the probability of a guilty plea at this stage, given these characteristics; and then it will combine the predicted probability of a guilty plea for this case with a random number to simulate whether he pleads guilty. The simulation model will evaluate prosecutorial decisions having to do with such issues as whether to accept or reject a given type of case brought from the police, how much of a reduction in formal charges to offer in return for a guilty plea at each stage, and what mix of prosecutors'

time to allocate to each prosecutorial activity. INSLAW is not aware of previous attempts to apply multiple regression analysis in this manner.

Regression analysis is widely applied in social research for reasons that go beyond its power to estimate and test alternative causal systems "parametrically" (i.e., based upon known relative frequency distributions, so that probability statements can be made about the reliability of the estimates). Much of its additional appeal is due to its flexibility. When first developed, regression analysis was applicable only to systems that were linear, or nearly so; also, it was used only to relate unbounded "scalar" (i.e., measurable) variables; further, it was suitable only when relationships were "homoscedastic" (i.e., the property that exists when the distribution of the dependent variable on each independent variable has a constant variance for all values of each independent variable). Now, due largely to extensions developed within the discipline of econometrics, regression analysis has been made capable of overcoming each of these earlier limitations. Nonlinearity can be dealt with adequately in most cases by transforming to the linear form those variables that are nonlinear with one another (e.g., by regressing the dependent variable on the logarithm or exponent of a nonlinear independent variable). "Taxonomic" (i.e., qualitative) or other discrete-valued variables, and combinations thereof, can be used as independent variables in a regression equation through the creation of "dummy" (i.e., binary or "zero-one") variables. A dichotomous dependent variable can be converted to a variable suitable for regression by aggregating individual observations into cells, expressing the dependent variable as a proportion, transforming the proportion to a suitable form (e.g., logit, probit, tobit transformations), and regressing this transformed variable on the independent variables. And homoscedasticity can usually be imposed through an "analysis of residuals," followed by the application of an appropriate scheme for weighting the observations.

Regression analysis is also useful to assist in formulating a model when the theory about the causal mechanisms of a system is weak. A popular version of multiple regression for assisting in model formulation is known as "stepwise regression." In stepwise regression, each principal dependent variable is regressed first, on the variable with which it is most highly correlated; then it is regressed on that variable plus the variable which produces the largest coefficient of multiple determination; next, it is regressed on those two variables plus the variable that produces the largest coefficient of multiple determination; this is repeated until all the variables have been entered, so that, in the end, there will be as many steps as there are explanatory variables. This process will identify how each entering variable affects the relationship between the dependent variable and all the previously entered variables. When these relationships turn out to be substantially altered by an entry, multicollinearity is revealed, and the variables most affected by the new entry can be viewed as likely principals in a separate causal relationship.

The basic elements of regression analysis are set forth in Applied Regression Analysis, by N.R. Draper and H. Smith;[1] the refinements noted

[1] N.R. Draper and H. Smith, Applied Regression Analysis (New York, John Wiley & Sons, Inc., 1967).

above can be found in H. Theil's Principles of Econometrics[2] and A.S. Goldberger's Econometric Theory.[3] An excellent example of the application of modern multiple regression analysis to the court system is W.M. Landes' "An Economic Analysis of the Courts,"[4] which investigates the determinants of pretrial settlement and the conviction rate.

2. Goodman's Technique

Leo Goodman has developed a methodology for analyzing relationships among taxonomic factors. This method has been widely cited in the sociology literature largely because so many of the factors that sociologists investigate are taxonomic or qualitative (e.g., race, region, attitude) rather than scalar (e.g., income, age, the Sellin-Wolfgang index).

As with multiple regression analysis, Goodman's technique attempts to infer causality out of a body of observations. The particular strength of his method is that it utilizes an unambiguous criterion for assessing alternative causal models that has advantages when the number of observations and the number of variables under investigation are small. This technique consists, first, of organizing the new data into a k-way cross-classification table (where k refers to the number of factors or variables involved in the study); then, the k-way cross-classification format is repeated, with raw observations replaced by expected frequencies that are estimated for each model of causality under investigation; next, the "main" effects and

[2] H. Theil, Principles of Econometrics (New York: John Wiley & Sons, 1971).

[3] A.S. Goldberger, Econometric Theory,(New York: John Wiley & Sons, 1964).

[4] W.M. Landes, "An Economic Analysis of the Courts," Journal of Law and Economics, Vol. 14, No. 1, April 1971, pp. 61-107.

"interaction" effects under each assumed causal model are estimated, based on observed odds ratios; finally, each hypothesized model of causality is scored for "goodness-of-fit" to the data, using a chi-square statistic.

An alternative mode of using this technique consists of an exhaustive comparison of all possible causal models that describe the main and interaction effects of the variables in the system. Goodman calls this the "saturated model." While this mode becomes rapidly unworkable as the number of variables grows beyond five or six, it does have some value in those instances where the number of variables in the system is small and the researcher has no intuition or theory about how the variables relate causally.

A chief distinction between this technique and that of multiple regression is that multiple regression isolates and estimates the extent of effect that each independent variable has on each dependent variable in the regression when all the other factors are held at their mean values; because all of Goodman's variables are taxonomic and, hence, have neither degrees of impact nor mean values, the notion of either isolating or estimating measurable effects, independent of other factors, is meaningless in his framework. Another distinction is that while multiple regression provides an estimate of the proportion of variation in the dependent variable that is attributable to (or "explained" by) the independent variables, Goodman's technique does not. He does, however, provide statistics that he regards as being "somewhat analogous" to the statistics of multiple regression analysis that measure the degree of impact between variables, as well as that which measures proportion of total variance explained by the independent variables.[5]

[5] L.A. Goodman, "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables," American Sociological Review, Vol. 37, February, 1972, pp. 42-44.

Goodman has applied his technique to an analysis of the determinants of a soldier's choice in region of assignment[6] and to an analysis of the factors affecting interracial sentiment.[7]

It may be appropriate to apply the technique to research issues within the PROMIS Research Project. In order to do so, aspects of the criminal justice system that involve a small number of dichotomous factors (up to four or five) will have to be identified. The problem of converting scalar variables into dichotomous ones can be overcome by reassigning the scalars the value zero if, for a given observation, the scalar takes on value less than its mean, and the value one otherwise. It must be noted, however, that considerable information about degree of relationship will be lost in this process; in particular, all information about the variance of the original variable will be lost.

One important issue in the area of prosecutor performance that may prove to be worthy of Goodman's technique is that having to do with the determinants of conviction: How do the most crucial factors pertaining to a case interrelate to cause the case to end up as either a conviction or nonconviction? Another issue is closer to the area of police-prosecution relationships: What factors determine whether the prosecutor decides to "paper" a case? The original determination of the factors that are "most crucial" in the analysis of each question may be identifiable as a result of the application of multiple regression techniques.

[6] Ibid.

[7] L.A. Goodman, "A General Model for the Analysis of Surveys," American Journal of Sociology, vol. 77, May 1972.

3. Factor Analysis and the Method of Principal Components

There is often much to be gained from grouping variables together to form new variables. At least three purposes can be served in clustering variables: first, when the number of original variables is large relative to the number of observations, grouping variables provides more reliable estimates of the most fundamental relationships of analysis; second, once formed, the clusters often reflect concepts that have useful interpretation in their own right; third, when the original set of explanatory variables contains members that are highly correlated with one another, it is often simpler and more economical to represent them as a smaller set.

Among the major techniques that lie within this domain and that appear to have potential within the PROMIS Research Project are factor analysis and the method of principal components. Both of these techniques are oriented more toward model formulation than toward model estimation or testing; neither requires a theory about how the explanatory variables are related to one another. Both techniques rearrange and reduce explanatory variables to a smaller set of factors. These rearrangements are determined by the patterns of association that are revealed from the correlation coefficients between the original variables. Accordingly, the first step of each of these techniques consists of the preparation of the correlation matrix, which reports the correlation coefficient for every pair of variables in the data set. The second step uses the correlation matrix to form new variables which are really linear combinations of the original variables.

It is at this second step that factor analysis and the method of principal components depart from one another. For the case of n original

independent variables, the principal components model constructs sequentially n uncorrelated components, such that each component consists of a "best" linear combination of all the original independent variables. These linear combinations are "best" in the sense that each combination consists of the unique set of weights that maximize the amount of residual variance that can be explained in the dependent variable after the effects of all the earlier components have been removed from the data. The crucial constraint in the formation of each component is that that component be absolutely uncorrelated with all previous components. The first component consists simply of the set of weights that maximizes the proportion of explained variance in the dependent variable. By the time the n^{th} component is formed, all of the variance in the dependent variable that can be explained by the full set of original variables will have been explained by the n components. The final expression will relate the dependent variable to the n independent components as a multiple regression equation. Consistent with the fundamental objectives of variables clustering, it is generally appropriate to regress the dependent variable only on the first few components, which will usually explain most of the variance in the dependent variable that can be explained with the data.

Factor analysis, on the other hand, clusters the original variable by using the correlation matrix in such a way that distinguishes between "common" factors (i.e., those factors that cause explanatory variables to be correlated) and "unique" factors (i.e., the effect that remains after the commonality is removed). Like the principal components model, the final expression in factor analysis will relate each dependent variable to a set of "factors" that is smaller than the original set of explanatory variables. Unlike principal components analysis, each new cluster need not explain the maximum amount of remaining variance in the dependent variable that is explainable using the original variables. Moreover, the clusters need not even be uncorrelated in factor analysis--when they are, the factors are called "orthogonal;" otherwise they are called "oblique."

There are several different ways of forming the clusters in factor analysis: they can be formed in a single process by using any of a number of different techniques (e.g., replacement versus nonreplacement of the main diagonal of the correlation matrix, Rao's canonical factoring method, the alpha-factoring technique, the image-factoring method). They can, alternatively, be reformed from any of these initial solutions through a process referred to as "rotation," of which there are several techniques (e.g., quartimax, varimax, equimax, oblique). When rotation is selected, it is customary to iterate the rotation process until an arbitrarily stable solution is produced.

Factor analysis and the method of principal components have often been applied to phenomena that are described empirically by variables whose meanings are not entirely clear. These techniques combine variables that are found to be associated in some way, and it is left to the researcher to assign meaning to each cluster. These techniques were developed and have been applied primarily by psychologists. Accordingly, it is natural to look for these techniques to have their greatest potential in the PROMIS Research Project within such areas as the patterns of criminal behavior and the determinants of victim proneness, which consist of behavioral variables that have strong elements of commonality among them.

Some prominent references in factor analysis and the method of principal components are H.H. Harman's Modern Factor Analysis,[8] D.N. Lawley and A.E. Maxwell's Factor Analysis as a Statistical Method,[9] and an article by C.R. Rao, "The Use and Interpretation of the Principal Component Analysis in Applied Research."[10]

4. Other Multivariate Techniques

While the three statistical techniques described above are expected to carry the primary load of the problems of multivariate analysis that fall within the PROMIS Research Project, other techniques will be used as well. Included among these are techniques that assist in establishing the population from which given observations arise (i.e., discriminant analysis and cluster analysis) and techniques for analyzing longitudinal data.

a. Cluster analysis and discriminant analysis. All the multivariate techniques discussed up to this point can predict outcomes based on given characteristics. Another set of techniques has been designed to enable inference in the reverse direction: Given the characteristics that describe the outcome, establish the population from which the observation came. In spite of the fact that this latter sort of problem is, on the

[8]D.N. Lawley, Modern Factor Analysis (Chicago: University of Chicago Press, 1967).

[9]D.N. Lawley and A.E. Maxwell, Factor Analysis as a Statistical Method (London: Butterworth & Co., Ltd., 1963).

[10]C.R. Rao, "The Use and Interpretation of Principal Component Analysis in Applied Research." Sankhya, v. 26, 1965, pp. 329-358.

surface, quite different from the former, it would seem nonetheless that techniques designed to deal with such a problem might be applicable to problems of prediction as well. Why, after all, should it matter whether the dependent variable is called an "outcome" or a "source?" As long as the technique is better at predicting the dependent variable than any other, whether or not the application of the technique is the "usual" one should be a secondary consideration. With this justification in mind, the following paragraphs describe cluster analysis and discriminant analysis.

While cluster analysis and discriminant analysis take on various meanings in the literature of multivariate analysis, they both refer generally to the problem of assigning observations to groups on the basis of a set of known characteristics. Both techniques employ a range of alternative algorithms from integer and dynamic programming to determine an efficient allocation of observations. Both rely upon "splitting" or "discriminant functions" to serve as the criterion for this allocation process. The chief distinction between the two is that cluster analysis assigns observations in such a way as to produce clusters that contain observations which look as much alike as possible, while discriminant analysis assigns observations in such a way as to minimize the expected total cost of misassignment. Accordingly, the splitting criteria for cluster analysis are referred to as coefficients of "similarity" (e.g., the fractional match coefficient, the Tanimoto match statistic, and the Mahalanobis D^2 statistic) and "stability," both of which measure the homogeneity of observations within each cluster, taking into account all the characteristics of each observation. The splitting functions in discriminant analysis, on the other hand, assign observations on the basis of "discriminant

scores," which are computed as linear combinations of the values of the characteristics that describe each observation, adjusted for misclassification costs.

Now, while the misclassification cost in discriminant analysis takes on a very clear meaning when the problem consists of a decision to assign each observation to a group, its meaning is less clear when the goal is the accurate description of a phenomenon. It is not apparent which of the criminal justice system processes to be studied in the PROMIS Research Project are of such a nature that a single observation error in one direction will be in any sense more harmful to an inference about causality than an error of the same magnitude in the opposite direction, given that the estimates are unbiased (i.e., given that across all the observations, the squared errors in either direction are offset by the squared errors in the opposite direction). If such issues are identified, then discriminant analysis should prove useful. Similarly, if the various splitting criteria of cluster analysis predict outcomes more accurately than other techniques to be used, then cluster analysis will be the preferred technique of multivariate analysis.

Cluster analysis is described more fully in a book by R.C. Tryon and D.E. Bailey entitled Cluster Analysis[11] and in a paper by Jerrold Rubin, "Optimal Classification into Groups: An Approach for Solving the Taxonomy

---

[11] R.C. Tryon and D.E. Bailey, Cluster Analysis (New York: McGraw-Hill, 1970).

Problems."[12] Discriminant analysis is presented in P.J. Dhrymes' Econometrics[13] and in D.F. Morrison's Multivariate Statistical Methods.[14]

b. Techniques for analyzing longitudinal data. Most of the data to be analyzed in the PROMIS Research Project will consist of cross-sectional observations of individual cases. However, under the topic "patterns of criminal behavior" will be an analysis of the factors associated with recidivism of defendants.[15] In order to investigate recidivistic behavior, defendants will have to be observed longitudinally as well as cross-sectionally, so that the number of times each defendant enters the system in separate case episodes can be observed. Thus, it will be possible not only to learn about the factors related to recidivism, but also to estimate the probabilities that each particular pattern of recidivism will occur, given a set of defendant characteristics. Furthermore, it also will be possible, given the defendant's personal characteristics, the pertinent characteristics of each crime he has committed, actions taken by the criminal justice system, and the sequence in which his crime episodes occurred, to predict the probabilities that he will not recidivate within a given period of time, that he will repeat his most recent offense, and that he will be charged next with a different offense.

[12]J. Rubin and H.P. Friedman, "A Cluster Analysis and Taxonomy System for Grouping and Classifying Data," IBM Program Library, August 1967.

[13]P.J. Dhrymes, Econometrics (New York: Springer-Verlag, 1974).

[14]D.F. Morrison, Multivariate Statistical Methods (New York: McGraw-Hill, 1967), pp. 130-133.

[15]The general approach discussed in this section also has potential for application to an analysis of factors associated with the tendency of victims to reappear within the criminal justice system.

These probabilities are most conveniently dealt with in a matrix of "transition probabilities," which depict the probabilities of going from each major offense category (e.g., misdemeanor against a family member) to another (e.g., no subsequent offense, repeat of same type of offense, felony against a stranger). The matrix of transition probabilities can be generated either separately for each major class of defendant (e.g., unmarried first offender) or at a more detailed level by taking all pertinent characteristics as predictor variables within a multiple regression framework. This general sort of analysis was carried out in a study of delinquency in Philadelphia, and reported by M.E. Wolfgang, R.M. Figlio and T. Sellin.[16] The INSLAW Research Group views this type of longitudinal analysis as being particularly crucial to a thorough understanding of the effect that each discretionary factor within the criminal justice system has upon whether and how a defendant recidivates.

D. Conclusion

The choice of which multivariate technique to apply depends on the problem and the ability of the technique to describe the elements of that problem, but the following simplified sequence of steps may serve to illuminate the procedure:

1. Step 1.

If one has too large a number of variables to analyze, factor analysis techniques may be used to reduce this number to a smaller set of "composite" variables by grouping variables that are correlated with each

[16]M.E. Wolfgang, R.M. Figlio, and T. Sellin, Delinquency in a Birth Cohort (Chicago: University of Chicago Press, 1972).

other. This is designed to simplify the analysis and interpretation of the data. This technique has been successfully used in structuring social indicators. Stepwise regression techniques have been used to reduce a set of independent variables to those most significant in explaining variation in a dependent variable, and have the advantage over factor analysis in producing a predictive model with readily defined explanatory variables.

2. Step 2.

If one has an unwieldy number of observations, the cluster analysis technique may be used to group observations into homogeneous groups or clusters. For example, this method has been successfully used to group data for small geographic areas into larger areas (say, groupings of census tracts) that are homogeneous with respect to the set of variables used for clustering. This technique can be useful for developing a simpler typology for characterizing the universe.

3. Step 3.

If one is able to predefine groups according to some criterion, particularly a qualitative one such as high, medium, and low crime areas, then discriminant analysis techniques can be used to assign observations to these groups based on the criterion. Discriminant analysis also allows one to come up with a single criterion function (decision rule) for assigning observations into groups based on minimizing the cost or error of misassignment. Cluster analysis can be used to group observations when the groups are not predefined, and discriminant analysis can be used when they are predefined. One way of defining groups is to use cluster analysis.

4. Step 4.

Having reduced the set of variables and data, one is now ready to develop descriptive or predictive models. Multiple regression techniques are useful in predicting values of a dependent variable on the basis of a set of explanatory or independent variables. Depending on the nature of the data, various types of data transformations may be necessary to improve the prediction capacity of the regression techniques, as explained above. Alternative multivariate models can be evaluated based on their predictive accuracy.

5. Step 5.

If the analyst has a data base (like PROMIS) that contains periodic measurements of a set of variables (time series), he may be interested in techniques for analyzing longitudinal data. Here, special attention must be paid to changes over time, which might for example change the nature of the groupings of observations and variables done for an earlier period. The PROMIS Research Project will attempt to test special techniques developed for longitudinal analysis in developing time trends. Thete techniques generally combine those used for cross-sectional analysis in Steps 1-4 above in dealing with the full set of data for all time periods. Also, the development and analysis of transitional probability matrices will be included as a task. In addition, the data from PROMIS may lend themselves to the application of statistical techniques designed specifically for the analyses of time series. It could prove useful to apply these techniques toward the aim of understanding the underlying trends, as well as any seasonal or cyclical components that exist in the data that describe crime and the criminal justice system.

E. Outline of Tasks

The actual process of carrying out the multivariate analysis will consist of a number of interrelated tasks. Some of these tasks intersect with tasks that will be performed under the three major areas of PROMIS Research scheduled for the first year of the project, and under such areas as analysis of data reliability and univariate analysis. This section presents an overview of the multivariate research process and a list of the tasks that comprise this process.

1. Overview of the analysis

First, a theoretical model of each subsystem under analysis will be developed. This model will consist of a set of presumed causal relationships among the factors that describe that subsystem. The subsystems themselves are likely also to be related to one another in several ways (e.g., arrests are both an immediate end product of police operations and an input to the prosecutorial process). The model of each subsystem will be developed as though data were available on every pertinent factor within that subsystem.

Then the full array of data elements contained in PROMIS will be reviewed and related to the factors identified in the theoretical models. Some of these data elements will measure the factors contained in the model more closely than others. Some of the factors modeled may have no counterparts at all in the set of data elements; in which case, a sample data base may be developed from source documents. In any event, each model will be reformulated by structuring relationships among the data elements in a manner that resembles each original model as closely as possible.

Next, each of these respecified models will be estimated and tested using the PROMIS data. Each of the applicable techniques described in Section B will be applied to this stage of the analysis.

At the same time, the full array of multivariate techniques will be applied to the entire set of PROMIS data elements, as though no theoretical model existed. This mode of analysis will complement the other by identifying relationships among data elements not hypothesized in the theoretically based mode. Thus, it will be possible to revise the theoretical models on the basis of this analysis, and compare the accuracy of the revised models with that of earlier versions.[17]

As previously noted, the comparison of alternative techniques will be based on the respective accuracy of prediction. The accuracy of prediction criterion will also apply to the comparison of alternative models. To assess this accuracy, a standard split sample validation procedure will be used. This will consist of using part of the observations to estimate the relationships among factors, and the remaining observations to measure how closely each model and technique predict the dependent variables among the latter set of observations. Those combinations of analytical model and

---

[17]It is appropriate to ask why theoretical models are even necessary if multivariate techniques are capable of inferring causal mechanisms in the absence of such models. This is a matter of sharp controversy among social scientists. The usual case is that neither theoretical nor empirical techniques alone can accurately explain the system under analysis. Theory alone suffers from lack of factual support, and empirical analysis alone suffers from measurement error. Either alone tends to suffer from failure to identify the full set of pertinent factors. By starting with a theory, however, then testing it empirically, and revising the theory in light of empirical findings, it is generally possible to more accurately explain social systems than by ignoring either the theory or the analysis of the data.

multivariate techniques that produce the most accurate predictions of each major dependent variable, as measured by the mean square error of prediction (i.e., the sum of the variance plus the squared bias of the predictor), will be reported as the findings of the multivariate analysis.

2. List of Tasks

- Formulation of hypothetical causal relationships.
- Matching of PROMIS data elements with factors identified in hypothetical models.
- Respecification of models in terms of PROMIS data elements.
- Empirical test of respecified models.
- Application of model-free mode of analysis.
- Revision of models.
- Test of accuracy of prediction.
- Report of first-year multivariate analysis results.

ATTACHMENT I

PROMIS DATA BASE DESCRIPTION

The PROMIS data base contains approximately 165 different types of information on each case including:

Information about the defendant: This includes name, alias, sex, race, date of birth, address, facts about prior arrests and convictions, employment status, and alcohol or drug abuse;

Information about the crime: The date, time and place of the crime, the number of persons involved in the crime, and information about the gravity of the crime in terms of the amount and degree of personal injury, property damage or loss, and intimidation using the seriousness scale developed by the criminologists Marvin Wolfgang and Thorsten Sellin;

Information about the arrest: The date, time and place of the arrest, the type of arrest, and the identity of the arresting officers;

Information about criminal charges: The charges originally placed by the police against the arrestee, the charges actually filed in court against the defendant and the reasons for changes in the charges by the prosecutor, the penal statute for the charge, the FBI Uniform Crime Report Code for the charge, and the project SEARCH Code for the charge;

Information about court events: The dates of every court event in a case from arraignment through motion hearing, continuances hearing and final disposition to sentencing, the names of the principals involved in each event including the defense and prosecution attorneys and judge, the outcomes of the events, and the reasons for the outcomes; and

Information about witnesses: The names and addresses of all witnesses, the prosecutors assessment of whether or not the witness is essential to the case, and any indications of reluctance to testify on the part of the witnesses.

Another key feature of PROMIS is the ability to track the workload of the criminal court process from three separate vantage points.

From the vantage point of the crime or criminal incident. This is accomplished by including in PROMIS the complaint number which the police department assigns to a reported crime. With this number, it is possible to follow the full history of the court actions arising from the crime even though those court actions may involve multiple cases, and multiple trials and dispositions;

From the vantage point of the accused person or defendant. This is accomplished by incorporating in PROMIS the fingerprint based number which the police department assigns to the individual following his arrest and which is used again by the department if the same individual is subsequently arrested. With this number, it is possible to accumulate criminal history files on offenders and to note incidents of recidivism; and

From the vantage point of the court proceedings: This is accomplished by including in PROMIS the docket number which the court assigns to the case pending before it. With this number, it is possible to trace the history of any formal criminal action from arraignment through final disposition and sentencing, and to account for the separate disposition of each count or charge.

The inclusion of these three numbers appears simple but it unique with PROMIS and extremely significant. The numbers provide an "instant replay" capability to track the criminal incident, the defendant, or the court actions, and they provide a basis for communication among the various constituent agencies of the criminal justice system. Without them, there is no comparability of information among the agencies.

ATTACHMENT II

SAMPLE STATISTICAL TABLES

TABLE 2.
ARRESTS IN 1973 BY OFFENSE TYPE OF
MOST SERIOUS CHARGE AGAINST THE DEFENDANT

				Number	Percent
I. Crimes Involving A Victim					
A. Personal Victimizations Involving Violence				5040	32.6%
1) Homicide			259		
a) Murder		200			
b) Manslaughter		49			
c) Negligent homicide		10			
2) Assault			2891		
a) Aggravated		2002			
b) Simple		684			
c) Assault on a police officer		205			
3) Forcible Sex Offenses			450		
a) Female victim 16 and over		357			
b) Victim under 16		72			
c) Male victim		21			
4) Robbery			1440		
a) Armed		726			
b) Other		714			
B. Personal Victimizations Without Violence				1898	12.3%
1) Larceny			1337		
2) Auto theft			372		
3) Fraud			189		
C. Crimes Against Residences Or Households				1370	8.9%
1) Burglary			1174		
2) Property destruction			164		
3) Arson			32		
D. Crimes Against Businesses Or Institutions				2099	13.6%
1) Robbery			217		
2) Burglary			372		
3) Larceny			1059		
4) Embezzlement and fraud			305		
5) Auto theft			74		
6) Arson			8		
7) Property destruction			64		
II. Crimes Without An Identifiable Victim				4757	30.8%
A. Weapons Offenses			1042		
1) Gun		827			
2) Other weapon		215			
B. Gambling			372		
C. Consensual Sex Offenses			836		
D. Drug Offenses			1872		
E. Bail Violations And Prison Breach			635		
III. Crimes Which Could Not Be Classified				296	1.9%
All Cases				15,460	100.0%

Reasons for Continuances

(Excluding continuances to initial trial date, to preliminary hearing, and to Superior Court Grand Jury)

Diversion programs	9.2%
Witness no show or appears unfit	4.4%
Police officer no show	5.3%
Counsel unavailable	13.9%
Counsel not prepared	2.7%
Defendant unavailable	4.7%
Defendant no show - bench warrant issued	14.6%
Court backlog	5.5%
System problems	5.3%

TOTAL: 8398 Reasons

PAPERED MISDEMEANORS RELEASE TYPE

	Personal Recognizance	Surety Bond	Cash Bond	Third-Party Custody	Other	Unknown	Total
Not Specially Assigned	46.5%	6.6%	1.9%	6.5%	2.1%	36.4%	5266 (81.6%)
Assigned to Major Violator's Unit	27.8%	14.3%	5.7%	13.2%	2.3%	36.5%	1186 (18.4%)
	43.1%	8.0%	2.6%	7.7%	2.1%	36.5%	

RELATIONSHIP BETWEEN CHARGES BROUGHT BY THE POLICE AND CHARGES PAPERED

	No. of Cases Brought by Police	No. of Cases Papered	Percentage Papered
Murder 1	101	89	88.1%
Rape	278	243	87.4%
Armed Robbery	678	651	96.0%
Burglary 1	291	235	80.8%

CASES IN WHICH THE VICTIM AND THE DEFENDANT WERE STRANGERS

	No. of Cases Brought by Police	No. of Cases Papered	Percentage Papered
Murder 1	16	16	100.0%
Rape	104	89	85.6%
Armed Robbery	379	368	97.1%
Burglary 1	114	93	81.6%

Distribution of Offenses, by Victim-Offender Relationship, Offender's Sex and Offense Type

Victim-Offender's Relationship / Offense Type	FEMALES				MALES			
	Family	Friend and Acquaintance	Stranger	Unknown	Family	Friend and Acquaintance	Stranger	Unknown
Violent	83.7	71.88	14.1	10	75	57.73	33.49	16.09
Forcible Sex				N = 1	5.8	8.9	3.4	1.1
Property	10.8	23.76	40.5	32	13.4	28.2	48.4	30.5
Victimless	2.7	2.8	43.9	55.4	3.3	3.2	13	49.58
Other	2.7	1.4	1.3	2.4	3.3	1.7	1.5	2.6
Total	100 (111)	100 (345)	100 (733)	100 (1288)	100 (514)	100 (2127)	100 (3995)	100 (5994)

Distribution of Offenses by Offense Type, Offender's Sex and Victim-Offender Relationship

Victim-Offender's Relationship / Offense Type	FEMALES					MALES				
	Family	Friend and Acquaintance	Stranger	Unknown	Total	Family	Friend and Acquaintance	Stranger	Unknown	Tota
Violent	16.2	43.2	18.1	22.47	100	9.7	31.39	34.2	24.66	100
Forcible Sex				N = 1		6.9	44.4	32.3	16.2	100
Property	1.4	10.1	36.9	51.3	100	1.5	13.5	43.6	41.2	100
Victimless	0	1.0	32.5	72	100	0	1.9	14.55	83	100
Other	5.8	9.8	19.6	60.7	100	6.2	13.9	21.9	57.5	100

Factors that Appear to Affect Whether a Misdemeanor is Accepted for Prosecution

Factor	Direction and Magnitude of Effect	Statistical Significance	Mean	Proportion
1. Male Defendant	--	High		.79
2. Number of Lay Witnesses	++	High	0.56	
3. Property or Evidence Recovered	++	High		.61
4. Apparent Property Motive	-	High		.38
5. Defendant Arrest Record	+	Low		.50
6. Defendant Age	+	Low	28.9	
7. Defendant Seriousness Score	+	Low	6.53	
8. Case Seriousness Index	+	Moderate	1.98	
9. Crime against Person	-	Moderate		.15
10. Crime against Institution	+	High		.16

ATTACHMENT III

NATIONAL ADVISORY COMMITTEE

Mr. Robert A. Shuker (Chairman)
Chief, Superior Court Division
United States Attorney's Office for
the District of Columbia
Fourth and F Streets, N.W.
Washington, DC 20001

Colonel Curtis Brostron
Secretary
St. Louis Board of Police Commissioners
1200 Clark Avenue
St. Louis, MO 63103

The Honorable William L. Cahalan
Wayne County Prosecuting Attorney
Frank Murphy Hall of Justice
1441 St. Antoine
Detroit, MI 48226

The Honorable William H. Erickson
Justice
Supreme Court of Colorado
201 State Capitol Building
Denver, CO 80203

Professor Edith E. Flynn
National Clearinghouse
505 East Green Street
Champaign, IL 61820

Mr. Paul L. Friedman
Assistant to the Solicitor General
Room 5136
Department of Justice
Washington, DC 20530

The Honorable Phillip H. Ginsberg
Public Defender
623 Second Avenue
Seattle, WA 98104

Mr. Lester C. Goodchild
Chief Executive Officer
Courts within the City of New York
80 Centre Street
New York, NY 10013

Mr. Willie King
Antioch School of Law
1624 Crescent Place, N.W.
Washington, DC 20009

Mr. Richard W. Kurrus
Attorney at Law
2000 K Street, N.W.
Washington, DC 20006

Dr. Albert J. Reiss, Jr.
Institution for Social and Policy Studies
Yale University
111 Prospect Street
New Haven, CT 06520

Professor Leslie T. Wilkins
State University of New York at Albany
1400 Washington Avenue
Albany, NY 12222

Professor Marvin E. Wolfgang
Director
Center for Studies in Criminology &
Criminal Law
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19174

Professor Hans Zeisel
University of Chicago Law School
1111 East 60th Street
Chicago, IL 60637

-and-

Vera Institute of Justice
30 East 39th Street
New York, NY 10016

ATTACHMENT IV

LOCAL ADVISORY COMMITTEE

Mr. Bruce D. Beaudin
Director
D.C. Bail Agency
601 Indiana Avenue, N.W.
Washington, DC 20004

Chief M.J. Cullinane
Metropolitan Police Headquarters
300 Indiana Avenue, N.W.
Washington, DC 20004

The Honorable Harold H. Greene
Chief Judge
Superior Court of the District
of Columbia
Washington, DC 20001

Mr. Norman A. Lefstein
Director
Public Defender Service
601 Indiana Avenue, N.W.
Washington, DC 20004

Mr. Benjamin H. Renshaw
Director
Office of Criminal Justice Plans
and Analysis
1329 E Street, N.W.
Washington, DC 20004

The Honorable Earl J. Silbert
United States Attorney
3600 District Court Building
Washington, DC 20001

ATTACHMENT V

RESEARCH PRIORITY OF EACH DATA ELEMENT

A. Defendant record:

PDID (I)
Court case (I)
Court case suffix (I)
Stated name (III)
Sex (II)
Race (II)
Date of birth (I)
Birth city (II)
Birth state (II)
Address code (II)
Urine date (III)
FBI no. IV
Complaint no. (II)

B. Charge record:

Co-def (I)
Search code (I)
MPD code (III)
Crt code (I)
Seq (III)
No paper (I)
NP reason (III)
Off. house no. (II)
Off. quad (II)
Off. date (II)
Off. time (III)

C. Case evaluation record:

Arrest date (II)
Arrest type (III)
PZ $ val prop. (II)
Evidence recov. (II)
Init. proc'g date (I)
Init. proc'g AUSA (I)
Init. rev'g ASA (II)
Arraignment atty (II)
Arr. atty type (II)
Arr. bail lockup (II)

Spec. assgn. (III)
FOT name (II)
Chemist report (III)
Opz. pow. (III)
Opz. vic. law ofcr. (III)
Opz. # vic. inj. (III)
Opz. # vic. tr. rl. (III)
Opz. # vic. hosp. (III)
Opz. # vic. kill (II)
Opz. def. status (III)
Opz. # vic. sex asslt. (II)
Opz. # vic. sex intim. (III)
Opz. # vic. intim. (II)
Opz. intim PDW (II)
Opz. # prem. ent'd (III)
Opz. # veh. stln. (III)
Stln. prop. (I)
Opz. narc. (II)
Opz. arr. in 5 yrs. (II)
Opz. # prev. arr. (II)
Opz. # prev. arr. per. (III)
Opz. yr. of last arr 1 (III)
Opz. yr. of last arr 2 (III)
Opz. yr. of last arr 3 (III)
Opz. 1st arr: auto (IV)
Opz. opiate (III)
Opz. alias (III)
Opz. alco. (II)
Opz. jobless 6 mo. (II)
Prb. sci. evid. (II)
Prb. presc. def. glt. (II)
Prb. corr. crimecom (II)
Prb. exculp. (II)
Prb. excul. evid. pre. (II)
Prb. time in D.C. (II)
Prb. spec. circum. (III)
Prb. def. disab. (III)
Prb. def. aid. abet. (III)
Prb. cond. rel. (II)
Prb. luck impeach (III)
Vic. corp. bus. (II)
Vic. rel. def. (I)
Vic. eye wit. rel. (I)
Vic. fical (II)
Vic. partic. (II)
Vic. crim. rec. (II)
Vic. narc. (II)
Vic. alco. (II)
Vic. sex (II)
Vic. age (II)
Vic. spec. circum. (II)

Vic. disab. (II)
Vic. reluc. (II)
Vic. empl. (II)
Vic. time in D.C. (II)
Wit. rel. to def. (I)
Wit. rel. to vic. (I)
Wit. obs. (II)
Wit. reluc. (II)
Wit. empl. (III)
Wit. crim. rec. (II)
Wit. spec. circ. (III)
Wit. age (III)
Wit. disab. (III)
Wit. eye wit. (II)
Wit. aider abettor (III)
Type crime consens (II)
Type crime injury (IV)
Opz. def. status (III)
Prob. win (III)
Lineup date (III)
Lineup time (IV)
Magist. compl. no. (III)
Defend. (I)
Opz. wts (I)
Polit. override (III)
Rank comp. (II)
Case closed (I)
Old ct. case (II)
Indict. date (I)
Pres. date (I)
Crime (I)

D. Case transaction record:

Init. plea (II)
JTD (I)
Tf. act. date (I)
Tf. AUSA (I)
Tf. type trial (I)
Tf. act. reason (I)
Tf. cont. date (I)
Tf. judge (II)
Tf. motion field (II)
Tf. motion date (II)
Fd. length trial (I)
Tf. fd. sent. min. (I)
Fd. sent. max. (II)
Fd. sent. $ (I)
Fd. sent. type (I)

DA recom. mos. (I)
DA recom. $ (I)
DA recom. con. conc. (I)
Trans. no. of. (II)
Last cont. date (II)
Contin. no. of. (I)

E. Witness record:

Wit. type (I)
Wit. ess. (II)
Wit. name (II)
Wit. address (II)
Wit. MPD rank (III)
Wit. MPD unit (III)
Wit. MPD badge (II)

ATTACHMENT VI

SUMMARY FREQUENCY TABULATION

For each of the following items, give both the absolute and relative (expressed as a percentage of total) frequency distribution, with total for each item, for calendar year 1973.

A. Felonies

1. Cases papered as felonies, by month
2. Cases brought to prosecutor as felonies, papered as felonies versus papered as misdemeanors versus not papered at all
3. Cases papered as felonies, by most serious police charge
4. Cases papered as felonies, by most serious prosecutor charge
5. Cases papered as felonies, by number of police charges at papering (0, 1, 2, 3, 4, 5, 6, 7, 8, 9 $\geq$ 10)
6. Cases papered as felonies, by number of total prosecutor charges at papering
7. Cases indicted as felonies, by number of charges at indictment
8. Cases papered as felonies, by number of defendants
9. Cases papered as felonies, by number of lay witnesses
10. Cases papered as felonies, by whether plea of guilt to one or more charges, verdict of guilt to one or more charges, verdict of innocent on all charges, other non-convictions, not yet closed, and other
11. Cases papered as felonies, by Sellin-Wolfgang index (single integers)
12. Defendants in cases papered as felonies, by Gottfredson score (nearest integer)
13. Defendants in cases papered as felonies, by number of previous arrests (0, 1, 2, 3, 4, 5, 6, 7, 8, 9 $\geq$ 10)
14. Defendants in cases papered as felonies, by age (< 20, 20-24, 25-29, 30-34, 35-39, 40-49, $\geq$ 50) -- give also mean and standard deviation

15. Defendants in cases papered as felonies, by race

16. Defendants in cases papered as felonies, by relationship to victim

17. Defendants in cases papered as felonies, by bail release status -- give also mean and standard deviation for amount of bail

    a. pre-indictment

    b. post-indictment

18. For all cases papered as felonies, give number of cases, and the mean and standard deviation for time in days between (if applicable) the following processes:

    a. offense and arrest

    b. arrest and papering

    c. papering and presentment

    d. presentment and preliminary hearing

    e. preliminary hearing and indictment

    f. indictment and arraignment

    g. arraignment and disposition

        (1) trial cases with jury

        (2) trial cases without jury

        (3) cases not brought to trial

    h. if guilty, time between disposition and sentencing

19. Repeat #18 separately for the following crime categories (most serious prosecutor charge at papering):

    a. murder

    b. rape

    c. robbery

    d. burglary

    e. assault

    f. other

B. Misdemeanors

(same as felonies, with following exceptions:)

2. cases brought to prosecutor as misdemeanors, papered versus no papered

18. for cases papered as misdemeanors, give total number of cases mean and standard deviation for time in days between (if applicable):

    a. offense and arrest

    b. arrest and papering

    c. papering and arraignment

    d. arraignment and disposition

        (1) trial cases with jury

        (2) trial cases without jury

        (3) cases not brought to trial

    e. if guilty, time between disposition and sentencing

19. omit

INSTITUTE FOR LAW AND SOCIAL RESEARCH
1125 15th St., N.W. Suite 625 Washington, D. C. 20005
(202) 872-9380 Cable INSLAW, WASHDC

William A. Hamilton, President

January 13, 1976

Ms. Voncile Gowdy
Project Monitor, Courts Branch
National Institute of
Law Enforcement and Criminal Justice
Law Enforcement Assistance Administration
633 Indiana Avenue, N.W.
Washington, DC 20531

Re: PROMIS Research Grant 75-NI-99-0111

Dear Ms. Gowdy:

We have prepared this letter of transmittal to accompany each copy of the PROMIS Research Design report, dated December 1975. This letter identifies the author(s) of each research design and will thus be useful for evaluation purposes and in responding to questions.

The authors of each chapter are listed below:

1.	Introduction	Sidney H. Brounstein
2.	Analysis of Prosecutor Operations	Brian E. Forst
3.	Analysis of Police Operations From a Court's Perspective	Brian E. Forst
4.	Patterns of Criminal and Related Community Behavior	Kristen Williams
5.	Data Purification for Research Purposes	Joyce Deroy Dean Merrill
6.	Application of Multivariate Analysis Techniques	Brian E. Forst

Please call us if you have any questions.

Sincerely,

Sidney H. Brounstein
Sidney H. Brounstein
Vice President

SHB/kef

END

CENTER FOR URBAN ANALYSIS
CRIMINAL JUSTICE DEMONSTRATIONS

FINAL REPORT

73-DF-09-0020

COUNTY OF SANTA CLARA
OFFICE OF THE COUNTY EXECUTIVE
70 W. HEDDING ST.
SAN JOSE, CAL. 95110

OCJP # A1899-2-74

30 SEPTEMBER 1975

F. M. LOCKFELD
PROJECT DIRECTOR

for

REGION J, COUNTY OF SANTA CLARA
REGIONAL CRIMINAL JUSTICE PLANNING BOARD

The preparation of these materials was financially aided through Federal grant from the Law Enforcement Assistance Administration nd the California Office of Criminal Justice Planning under the mnibus Crime Control and Sate Streets Act of 1968, as amended. ıe opinions, findings, and conclusions in this publication are ıose of the author and are not necessarily those of the Law ıforcement Administration or the California Office of Criminal ıstice Planning."

Santa Clara County

CENTER FOR URBAN ANALYSIS

CRIMINAL JUSTICE DEMONSTRATIONS

FINAL REPORT

CENTER FOR URBAN ANALYSIS
CRIMINAL JUSTICE DEMONSTRATIONS

May 1973 - June 1975

Table of Contents

FINAL REPORT

CRIMINAL JUSTICE DEMONSTRATION PROJECTS

List of Figures and Exhibits

CENTER FOR URBAN ANALYSIS

CRIMINAL JUSTICE DEMONSTRATIONS

EXECUTIVE SUMMARY

On April 26, 1973, LEAA awarded a Pilot "0" Discretionary Grant to Santa Clara County to establish the Center for Urban Analysis, Criminal Justice Demonstrations. This provided $160,880 to support the development of management oriented problem-solving techniques using geographically related data. The mission of the Center was to provide new problem solving tools, serve as a resource for planning, and provide criminal justice agencies with information to increase their individual and collective abilites to:

- diagnose specific crime problems
- develop better baseline measures
- formulate better focused, more effective programs
- evaluate results.

Second year funding of $65,000 was awarded through the Regional Criminal Justice Planning Board to assist the institutionalization of the Center and to carry out demonstration projects which help criminal justice agencies make decisions and act on solutions to important problems. The Center was intended to be partially supported by non-criminal justice users so as to extend the base of potential financial support and enhance the probability of the Center to finance itself.

The projects conducted by the Center concerned the allocation of resources to prevent and control crime according to the geography and characteristics of the urban environment. The Center has the responsibility of maintaining a computerized map (called a Geographic Base File) of the County. The Geographic Base File employed is the DIME file originally established by the Bureau of the Census for the 1970 Federal Census. The DIME file was corrected, extended to countywide coverage, and maintained on a current basis. The Center provides the technology for matching address-coded records to the DIME file, so that events may be precisely located on a map and referenced to any size or shaped geographic area. The data may refer to a street address, such as "70 W. Hedding St.", a street intersection, such as "N. 1st St. and W. Hedding St.", or a place name, such as "Civic Center". Map coordinates referenced to the State Plane Coordinate System (similar to latitude-longitude), and other relevant geographic identification, are appended to the data concerning the event. The geographic referencing of event data enables criminal justice agencies to relate agency data to other relevant data of the urban environment, such as property data from the Assessor's records and demographic data from the Census.

The information may thus be displayed on maps produced by computer line printer, used for analysis that could not be performed without the geographic reference, and established in an interactive computer graphic system data base for examination and problem-solving. The interactive computer graphic system, GADS (Geographic Analysis and Display System) is a prototype system developed under a Joint Study Agreement between IBM-Research (San Jose) and Santa Clara County. The system uses a television-like computer terminal installed in agency offices. Agency personnel were able to explore and manipulate their data with reference to a map image they specified meaningful to the problems to be resolved.

The Center provided a general purpose statistical and tabulation system, SPSS (Statistical Package for the Social Sciences), as a method to summarize agency data in more traditional forms. The combination of user-oriented, flexible and versatile techniques provided criminal justice agencies with powerful tools to assist finding solutions to operational and management problems.

The projects chosen as demonstrations were selected on the basis of need and request for assistance by user agencies. Demonstration projects included the Adult Probation Department and five different law enforcement agencies of four generally different sizes, plus a crime analysis and evaluation project serving all police agencies in the County. Several one-time jobs unrelated to a project were completed for client agencies. Most of the demonstration projects focused on the use of calls-for-service (dispatch) data with police agencies. The projects assisted the management problems related to strategic deployment, agency organization, resource allocation and activity analysis. One project concerned crime-specific program planning, monitoring and evaluation.

Center staff provided agency assistance with problem formulation, data collection, data analysis, and assistance in the problem-solving process as required. The speed and flexibility with which alternative solutions could be tried and evaluated with the GADS system plus the minimal use of a programmer/analyst to which a problem definition had to be communicated, enabled agency personnel to use their special knowledge effectively in the problem-solving process. The direct involvement of administrators, managers, supervisors and line personnel enabled the development of solutions that all levels involved in the problem could understand and support.

Data collection for agencies lacking systematic data collection and data entry procedures was identified as a significant problem in the projects. Agencies were guided to increase their data collection and analysis capabilities as a result of the projects, and recognition was made that data collection costs are rapidly amortized if the information is timely, accessible and supports the allocation of criminal justice agency resources. A general conclusion supports placing flexible analysis tools as close to decision-makers as possible. It is recommeded that the interactive graphics system be made available to agencies throughout the County with enhanced capabilities for data entry and control at the agency terminal.

PROJECT ACTIVITIES

Problem Statement :

Agencies are concerned with allocating resources to prevent and control crime. To do this, criminal justice agencies need some method for efficiently allocating their resources according to the geography and characteristics of the urban environment. The allocation of patrol or service areas, identifying where events have occurred, targeting preventive programs for certain kinds of people, places or businesses - - all of these require knowledge of where something is located on a map. In other words, there is a need to relate criminal justice data to geography and to sources of information which can describe the urban environment of that geography.

A variety of agencies collect data potentially useful for criminal justice problems. However, there are several difficulties that inhibit effective use of such information:

A. One difficulty concerns the style the geographically related information is usually transmitted. Tabulations are not easily used in map-oriented problems, and data is frequently grouped in area units that do not conform to the needs of a particular agency.

B. Geographically related data is often available in a less than timely manner for purposes to which it could have been applied.

C. Geographic data often requires the intermediation of a specialist to translate, digest, extract and manipulate the information before it becomes useful to a criminal justice professional. This process submerges the special knowledge in criminal justice offical has to bring to the problem solving process. The technician is unable to acquire the perceptions and understanding of the professional and the full effectiveness of information in decision making is lessened.

Any place on a map can be described by an address, place name or an X-Y coordinate. If criminal justice agencies have access to a geographic base file, (computerized map), any address or place identification can be located into any size or shaped geographic area. This geo-coding capability (Criminal Justice Standard No. 4.8) provides a capacity to relate crime data to other information about the urban environment and significantly improves criminal justice planning on a region wide and individual agency level.

Goals and Objectives:

There are hierarchies of goals and objectives. The goal is to reduce crime. The first intermediate goal is to develop more effective projects and programs to reduce crime. The second intermediate goal is to provide more definitive information for project and program formulation and evaluation.

The objectives of the first year grant funding were to provide criminal justice agencies with information to increase their ability to:

1. Diagnose specific criminal justice problems;

2. Develop better baseline measures by geographic area;

3. Formulate better focused and more effective projects and programs; and

4. Evaluate the results of these projects and programs.

The objectives for the second year of the project funding were to:

1. Institutionalize the Center for Urban Analysis on an interprise basis, whereby user agencies budget and pay for services rendered by the Center. The Center was intended to be partially supported by non-criminal justice users so as to extend the base of potential financial support and enhance the probability of the Center to finance itself;

2. Carry out six candidate projects which help criminal justice agencies make decisions and act on solutions to important problems.

ACTIVITIES AND FUNCTIONS

The functions of the Criminal Justice Demonstrations, were to assist criminal justice agencies in the resolution of problems that had a geographic character and to provide new problem-solving tools. Assistance was provided in problem formulation, assembling appropriate and timely data, organizing and conveying the information in a manner that agencies could easily use, and in technical assistance in the problem solving process. The approach emphasized agency involvement in order to improve the abilities of agencies to resolve problems and to retain a focus on implementation.

A primary function of the Center has been the responsibility for establishing and maintaining the geographic base referencing system. The geographic base file GBF/DIME (Dual Independent Map Encoding), was originally developed by the Bureau of the Census for conducting the mail-back part of the 1970 census. The initial version of the DIME file for Santa Clara County covered the urbanized (north) portion of the County and consisted of approximate 47,000 records. The Center extended and updated to this file to a county-wide coverage. At the end of the project period the file consisted of approximately 70,000 records and reflected the current (June, 1975) street network and corporate jurisdiction boundaries. Ability to maintain the DIME file was also achieved during the project period. DIME file updates and correction processes originally requiring a three to four week period to accomplish, were successfully completed in less than a week.

The software enabling the matching of address coded records to the GBF/DIME was also developed by the Census Bureau. The initial program, ADMATCH, was replaced by a improved program, UNIMATCH, furnishing the ability to acommodate intersection coded events as well as place name and street address coded events. The Center established and maintains a version of UNIMATCH specifically adapted for use in Santa Clara County. The Census Bureau also developed the GRIDS program, which produces computer line-printer maps from geo-coded data. The Center refined the production of GRIDS maps to a high degree for maximum effectiveness.

Center staff aided users in problem formulation, in generating and organizing data, and in assistance through the analysis process. In addition to geo-coding and the production of GRIDS maps, the basic capabilities that the Center offered in its projects were: (1) interactive computer graphic system (GADS); (2) a general purpose statistical and tabulation system (SPSS); and (3) specialized analytic and data analysis capabilities.

Interactive computer graphic system.

A prototype interactive computer graphic system, GADS (Geodata Display and Analysis System), developed by IBM-Research, San Jose, under a Joint Study Agreement with Santa Clara County, is a user-oriented system that relates a structured data base to a user specific map. The system provides opportunity for exploratory examination of the data and capacity for sophisticated manipulation of data and map features. The terminal is a relatively inexpensive, remotable device with hard copy unit attachment. The system is economical in computer utilization. The control language has been successfully operated in-house by uniformed and non-uniformed personnel of different backgrounds in a variety of police agency problems.

General purpose statistical and tabulation system.

The program package SPSS, Statistical Package for the Social Sciences, developed by the National Opinion Research Center at the University of Chicago, now diseminated and maintained by SPSS, Inc., is a flexible, coherent set of procedures for data analysis. The system is characterized as user-oriented in that users can modify procedural controls established by Center staff to produce specific desired analysis.

Specialized analytic and data analysis capabilities.

Center staff undertook special processing of certain data files such as from the Assessor's records to enable the relating of crime data to property and Census data. Center staff also developed special analytic programs for particular projects. These latter abilities were used minimally, as a focus of the work of the Center was to retain the control and the responsibility for analysis with user agencies, utilizing the tools and procedures that were generally available rather than those that were required to be developed for a specific application.

The specific tasks undertaken under the project during the last year were:

1. Provide data handling services to users. This work consisted of three major categories. The first category encompassed the capabilities of preparing, processing and geo-coding user data into user defined areas. The second category consisted of establishing data bases and map definitions meaningful for the user-defined problems. The last category concerned maintenance of basic software to produce tabulations, summaries, maps and analysis of user data.

2. Maintain Reference files. This work encompassed the maintenance of the Geographic Base File, extraction capabilities of the Property System and other data sources, and related software maintenance.

3. Provide User Support. This work focused on making the resources of the Center available and accessible to Criminal Justice agencies and to support the user-specific projects.

4. Organization and Administration. This function provided management for the undertaking of the work.

Task 1 Provide Data Handling Services to Users

Most demonstration projects undertaken by the Center, as summerized below, dealt with the use of police agency calls-for-service data (CFS). Four of the five police agencies with whom the Center worked used dispatch data as source documents. The Sheriff's Department utilized patrol officer log sheets that had already undergone data entry at the time the Center starting working with the information. The Center worked with the agencies and dispatch functions in classification methods, data capture, conventions and personnel procedures regarding data entry requirements. Mechanics were established to transmit and return source documents between the Center and agencies. Personnel at the Center scanned dispatch tickets to clarify the documents, minimize data entry errors and evaluate data capture encoding techniques. Procedures were established with data entry staff at GSA/DP for clear, efficient and cost-effective data entry. Techniques were established to perform basic editing of the data, and to process the address information so that it could be utilized in the geo-coding process.

Two alternative methods for accomplishing the geo-coding of source records were undertaken. The first method involved identifying the correct study area or beat number to each record within a portion of the Geographic Base File appropriate to each agency. In this method, the correct assignment of study area or beat was simply transfered from the appropriately matched Geographic Base File record to the dispatch record. In the second method, the correct X-Y coordinate value was transferred from the matched DIME record. The X-Y coordinate values were then located to the correct study area or beat through a procedure called "Point-in-Polygon". A principal advantage of this latter procedure is that additional geographic identifiers of beat, study area, etc., are not required to be maintained directly on the GBF/DIME. The X-Y coordinates provides sufficient identification to locate that event to any described polygon representing a study area or beat.

Considerable effort was expended in determining efficient methods for preparing the base maps representing study areas or beats. The first procedure adopted was the interactive labelling of each side of a boundary segment of a study area polygon with the correct area number. Internal segments for these polygons were then automatically coded with the correct study area number. This procedure required the use of Operations Research Analyst staff, was time consuming in interactive computer operation and required extensive elasped time. It could only be done by one individual during the course of a map development. Technical problems with polygon closrue and numbering were significant. The revised procedure uses a special data entry form to identify each segment of a polygon. Each vertex is cross-referenced to a appropriate DIME file node; new nodes are created when a study area vertex does not have a DIME file node at that location. The product is a polygon description of the study area that is then automatically numbered, with sides of the polygon edges labelled with correct area numbers by a cross-reference between user and

computer-assigned area numbers. The resulting file is simplified in that it does not include each DIME file segment that may occur along the edge of a polygon. The work can done by technical personnel rather than professional personnel, and several people were able to simutaneously code polygon boundaries for project maps. The system is batch oriented, economical, and offers technical and operational benefits in terms of closure and direct verification of polygons.

Efforts in data base loading and establishment were closely integrated with work conducted under the Joint Study Program with IBM-Research. The interactive computer graphic system includes procedures, called "Extraction", that can selectively examine individual records and prepare a user specific data base for a particular problem. Alternatively, a staff prepared Fortran routine may be modified to prepare a data base. The Extraction procedure is extensive in computer utilization but requires minimal staff effort; adapting previously prepared Fortran loading programs requires more staff effort but is efficient in computer utilization.

Software for address matching (UNIMATCH) and computer line printer maps (GRIDS) are maintained by the Census Bureau. Updated versions of this software have been installed and documented in GSA-DP computer system procedures.

The SPSS system is maintained by SPSS, Inc., which issues periodic updates and modifications. SPSS is also installed, documented and maintained on a regular basis at GSA-DP. Center staff has documented the flow of procedures and programs from data entry of calls-for-service data through standardized SPSS control language to produce several standard summaries.

Task 2 Maintain Reference Files

The main work on GBF/DIME maintenance follows procedures established by the Bureau of Census. Staff was required to develop programs prior to formal issue by the Bureau to accomplish several functions required by the GBF/DIME activities. Staff was also required to prepare programs due to deficiences in Census Bureau prepared programs.

The level of maintenance required for the GBF in Santa Clara County is identified by the Census Bureau as exceptional in the nation. A large number of corporate boundary changes occur within Santa Clara County pursuant to County policy requiring annexation of development to cities within whose sphere of influence the development occurs. The house numbering system in the County is consistent only in the Southern portion of the County. In the remainder of the County each city has its own independent and frequently conflicting house numbering system; the County numbering system often conflicts with those independent systems. The Real Property System maintained by the Assessor provides a basic reference source for determination of addresses and street names. However, no single accurate source exists in the County for address ranges on street segments. A variety of sources were thus required to correct erroneous entries or to fill in gaps, and to extend the GBF countywide. Census Bureau procedures for the DIME file are oriented to a long range, systematic, comprehensive correction, update and extension process. The Center adopted an approach directed to providing an immediately usable reference file for project purposes. Center procedures emphasized firstly, inclusion of all segments in the file with correct node identification; secondly, addresses ranges along the segments; thirdly, accurate coordinates; fourth, corporate boundaries; and last the block number and zip code references. The Bureau has established a number of edit procedures in the CREATE and FIXDIME II programs used by the Center. Center staff also developed a number of edit programs and procedures to establish and verify the file. One of the most important files developed from the DIME file was an intersection reference file, enabling events coded to street intersections to be geocoded. The Bureau program ADDEDIT which chains along segments, identifying where numeric gaps exist along address ranges, or where address ranges are inconsistent, was not released for local operation until close to the end of the project year. An address edit performed by the Bureau on the file dated early spring, 1975, produced considerable information regarding address problems. Many real-world address range problems exist in Santa Clara County however; many of these had been previously verified, validated, and will continue to be noted as inconsistencies reflective of on-the-ground situations.

For each client agency, the Center derived from a sub-file, or Project DIME, from the master GBF/DIME. The Project DIME was examined and edited for accuracy through the interactive computer graphic system. This was particularly valuable when maintenance cycles on the DIME file were major efforts requiring months of preparation. As the maintenance cycle has become shorter, edits more comprehensive and the file more complete, this function has been reduced. The project file experience of interactively editing and correcting the DIME file continues to assist the generating of inputs to batch correction updates. A principal advantage of a Project DIME is that far fewer records are involved, providing efficient operation for geocoding and other related processing. The interactive computer graphic system was particularly helpful for displaying, validating and examining coordinate locations of the network.

The state plane coordinate system that requires 7 and 6 digits respectively for the X and Y coordinates is difficult to examine other than graphically. Plotting the file with a plotting device is time consuming, expensive and lacks direct correction capability. Interactively, the map may be economically displayed, examined for topological consistency, querried for attribute validity, and corrected as appropriate. The Center and IBM-Research are continuing work related to map and network techniques.

In addition to the GBF/DIME, the Center has made extensive use of the Assessor's files, both the Real Property system and the Unsecured Property system. Center staff participate in the Property Data Management committee that has responsibility for the system, development of additional software and output from the system. Work conducted by the Center in the process of establishing the DIME file assisted the cleaning up of multiple situs addresses in the Unsecured file and identified census tracts for parcels lacking such identification. In 1967, all parcels were identified with a state plane coordinate reference. As this process has not been maintained, only 75% of current parcels have X-Y coordinates. The Property Location Index, derived from the Property System, contains the coordinate values and was used as a reference file for matching prior to the establishment of the GBF/DIME.

The most important activity undertaken with the Assessor's Unsecured File system for project purposes dealt with selection from the 1972 Unsecured Roll

to provide baseline counts by type, by area, for the San Jose Police Department Burglary Analysis Unit. The catorgies required a crosswalk from the CAPER (Crime Analysis Program Evaluation and Research) definitions to the Unsecured Roll definitions in order to accomplish baseline measurements for the Burglaries Prevention project.

Task 3. Provide User Support

The primary function of this task was to maintain liaison with criminal justice agencies, making the Center's resources available and accessible to agencies, and assisting their analysis and problem solving. The Center was therefore required to work with agencies from the beginning of project formulation thru project conclusion. A significant tool the Center offered agencies was the interactive computer graphic system. Training in the operation and utilization of the system was provided for user agencies. A training manual based on the San Jose Police Departments Beat Analysis project was successfully used by police officers and non-uniformed personnel in the Gilroy, Mtn. View and San Jose Police Departments. The Adult Probation Dept. project required close working relationships with supervising Probation Officers to help them define potential service areas, composed of census tracts, prior to the availability of the interactive computer graphic system.

A typical sequence of user support began with the identification of the appropriate classifications, definitions for activities, facilitation of a coding system that encompassed the range of events an agency wished to work on, and identification of commonly used place names used by dispatchers and officers for familiar locations. Data collection, for a period of 2-3 months, was initiated with Center liaison for that process. Procedures for statistical summarization of the data thru SPSS and other programs were then established to produce desired analysis. Map areas were defined consistent with patrol personnel and management perceptions. Area units, or "atomic" zones, were designed to reflect the smallest area meaningful for the purposes of the project. After data entry, editing and geocoding, a data base was established that could respond to specific questions agencies wished to examine. Extraction capabilities provided opportunity to examine the full range of the data for crime-specific or management-specific events; area examination capabilities for beat development, etc. were available under the general GADS system. As agencies began to work on a specific problem or examine specific attributes, staff was available to provide assistance with the analysis. Terminals were installed in police agency offices for a period of two to four weeks, to facilitate data analysis and problem solving. Due to the extensive use of the system by the San Jose Police Department, a terminal was installed in the Research and Development Section for the period October 1974 - June 1975.

The processing of CAPER data focused on geocoding and production of computer line printer maps (GRIDS maps), for the client agencies of the CAPER program. As data turnaround was deemed very important by CAPER staff, overnight processing and expedited GRIDS map was accomplished for that project. Detailed cost analysis and estimates for CAPER processing were also developed.

Task 4. Organization and Management .

Research designs for the evaluation of the interactive computer graphic system with IBM Research and user agencies were developed. Technical aspects of installation of teleprocessing equipment including modems, terminals and telephone installation were also part of this task. The management of resources throughout the project, coordination with other governmental agencies, committee participation, arrangements for data processing and other services were also accomplished under this portion of the work program. Staff prepared and gave many demonstrations and presentations of the work of the Center to various local groups as well as visiting representatives of agencies elsewhere in the country.

Agencies Utilizing the Services of the Center

- San Jose Police Department
- Gilroy Police Department
- Campbell Police Department
- Mt. View Police Department
- Sheriff's Department
- Countywide CAPER
- Adult Probation Department
- Criminal Justice Pilot Program

Agencies Whose Services were Utilized by Project

- Bureau of the Census
- Census Use Study
- County Data Processing
- Assessors Department
- Communications Department
- IBM Research thru Joint Study Agreement
- RECAP (Regional Educational Center for Automated Processing), of the County Office of Education

The first year grant funding permitted the undertaking of one project with a non-criminal justice agency user. The Center began work with the East Side (San Jose) Union High School District data committee representing all the elementary districts tributary to that high school district, and did further detailed work in the first project year with the Mt. Pleasant Elementary school district.

EVALUATION OF REPORTS

The criminal justice demonstrations of the Center were not intended to generate reports authorized by the Center. Individual agencies were responsible for the analysis they were conducting and their work was internalized. Few reports were produced by user agencies. Lt. Robert Bradshaw and JoAnn Moore of the San Jose Police Department presented a paper at the 1974 URISA conference, "Man-Machine Interaction for Police Field Manpower Deployment", describing the use of the GADS system in the San Jose Beat alignment work. The Final Report of the San Jose Police Department Burglary Analysis Unit, where Center resources were

used extensively, was not completed at the time of this writing. Technical reports describing the interactive computer graphic system, GADS, produced by IBM-Research staff are listed in the Appendix. Methods and procedures adopted by the Census Bureau for DIME file processes are well-documented by Census publication. SPSS is documented in a handbook for that system (2nd Edition issued 1975). The projects generated insights for the participating agencies into their problem solving techniques. The direct involvement of administrators, supervisors and line personnel in the solution process produced if not an optimum solution, at least a solution that all levels involved in problem could understand and could support. The speed and flexibility with which alternative solutions could be generated and evaluated, plus the minimal use of a programmer/analyst to whom a problem definition had to be communicated, encouraged a re-evaluation of the factors and criteria used in the methodology. Users with prior ideas about which factors were important in determining how to redistribute service areas often found, during the course of solution development, that some factors were not as critical as hypothesized and that other factors needed to be examined more carefully.

PROJECT ACCOMPLISHMENT AND ANALYSIS

Project Results and Anticipated Results: The anticipated results of the project were to provide information to user agencies as a basis for decision-making and this result was achieved. There is, however, a difference between the decisions that were anticipated to flow from the information and those that, in fact, resulted. In the Countywide CAPER project, the decisions were administrative in nature (e.g. costs for certain services performed by the Center); operational decisions with CAPER data were to be made by the participating law enforcement agencies. Where specific agency problems were identified and worked on, decisions were assisted by information developed through the demonstration project; such as in the Adult Probation Department, San Jose Police Department Beat Development and Burglary Analysis projects. Virtually all agencies discovered serendipitous uses for the data to facilitate decisions on problems that had not been anticipated. Some police agencies began the work with an initial focus on possible beat re-alignment. During the course of the project, however, their interests became more involved with analysis of administrative activities and allocation of agency resources, of which the geographic assignment was only a portion. Project information was used as a variety of agency functions, especially in budget development.

There were several factors that led to results of different decision-making than anticipated. Some agencies wanted to use calls-for-service information as a basis for tactical resource deployment. This use of the CFS data was inappropriate in the project context as there was, necessarily, a significant time lag between the occurrence of the event and the establishment of the data base for the system. The delay was necessitated due to data collection and entry methods, and technical difficulties attendant to prototype development. An intersection reference file from the GBF/DIME was not available for use in the geocoding process until January, 1975, necessarily delaying the geocoding of intersection-coded events, some of which were dated July, 1974.

In several cases project management at the agency level changed during the course of the project , and the particular interests of the original understanding were modified by these personnel changes. During the course of the Gilroy Police Department project, the Deputy Chief accepted a position as Chief of a police agency out of the County. Responsibility for project cordination in the Campbell Police Department was shifted several times during the course of that project. Lack of identified day-to-day responsibility, discontinuity of interests and personnel shifts occured in the course of the Sheriff's Department project. Successful accomplishment of project objectives occured when agency continuity of personnel and interests existed throughout the project. A major project contemplated under second year funding was expedited processing of calls-for-service data. This was anticipated to have been accomplished with the San Jose Police Department using the daily log tape from the computer-assisted dispatch system anticipated to be operational during the project year. The computer-assisted dispatch system was delayed significantly and will not be operational until early 1976. It was thus not feasible to undertake a project with expedited data entry due to non-automated data collection systems.

Work contemplated with the proposed Juvenile Information System was not realized when that system was aborted.

The demonstration projects have had different levels of impact on different agencies. In most agencies there seems an increased committment to data collection activities and in the analysis of data to support a variety of applications. The San Jose Police Department Burglary Analysis Unit utilized the demonstration project to refine and identify high risk burglary target areas for crime-specific programs, and to develop baseline measures for pre- and post program evaluation. BAU staff believe that this assistance was of considerable utility for the work of that program.

It is not believed that the results could have obtained more efficiently by a different allocation of resources or of project activities. Given the state of technology and the technical resources of the Center, the project proceeded in an expeditious and efficient manner to deliver information to user agencies as rapidly as possible and in forms that made it convenient for agencies to utilize the information. The results of the project compare favorably with other projects using a similar approach and with projects using different approaches and methods. This project focused on removing an intermediary between the agency and the information base. It was distinguished from projects where a consulting group would externally manipulate agency data and provide a solution arrived at by non-criminal justice professionals, or at least non-agency personnel. Such solutions are frequently inappropriate for agency implementation because of the difficulty of transfering full information from the agency to the consulting group. This project also attempted to respond precisely, with limited but focused information, to questions raised by agency personnel rather than producing voluminous data hypothesized to be useful for agency purposes. The intent was to deliver only the desired data analysis to user agencies. Agencies were not overwhelmed with information, but received responses to specific questions. Information available through the GADS system was available interactively and was readily communicable because of the graphic medium.

The results which might have been expected in the absence of the demonstrations vary from project to project. In the burglary analysis project, previous work identified a rectangular area approximately 1 mile by 3 miles as the candidate target area for burglary prevention programs. The use of the detailed geo-coded information enabled considerably refined area identification, leading

to a much more efficient use of resources with greater control on the particular programs. The San Jose Police Department had manually prepared a revised beat pattern prior to working with the interactive computer graphic system. The solution generated with the interactive system was superior in several measures, was able to be comminicated to other city officials, and was able to be rapidly manipulated by the officers in accordance with different resource requirements. Agencies that undertook focused work on administrative activities were able to determine actual workloads, and improve their programming for allocating resources.

One of the major findings of the demonstration projects was the difficulty of data collection efforts with agencies with non-automated data collection procedures. Computer-assisted dispatch is programmed for implementation by the San Jose Police Department and has an objective of eventually becoming county-wide in its application. There are political and technical obstacles to the realization of this intent. At the present time many police agencies maintain their own communications and dispatch functions. For non-automated agencies, the project demonstrated the importance of establishing a data collection process. The manner in which the demonstrations undertook data collection, processing dispatch tickets through a batch data entry method, is not efficient for systematic data collection. This method was the least expensive and most efficient manner in which to handle that task during the demonstration projects. On a continuing basis however, the batch keypunch, (or key-to-tape or key-to-disk) process can not accomplish the objective of rapid data entry into a system. External batch data entry does not minimize editing requirements and has an indirect feedback to improve data recording. It is desireable to have agency personnel directly accomplish data entry in a system enabling easy and rapid retrieval in order to respond to questions that periodically occur with recent dispatch data. The system should have a capability of providing basic summaries of the data. As project experiences have encouraged agencies to undertake data collection efforts, the Center is currently examining the data entry question to develop alternative solutions to this project-identified problem area.

The demonstration projects have led to the development of processes, procedures and methods that are highly cost-effective for analysis of criminal justice agency data. The project benefited considerably from the resources of the Census Use Study personnel who worked closely with Center staff during the early part of the demonstrations. The demonstrations could not have been conducted without the Census Bureau's work on the DIME file, methodologies developed by the Bureau, and the procedures available in the SPSS programs. Major contributions were made by IBM-Research staff under the Joint Study Agreement. In addition to hosting and refining the interactive computer graphic system, IBM-Research staff also provided valuable assistance in evaluation of work performed by agencies and staff, and assisted in the accomplishment of several projects. The actual resources dedicated to the grant were thus significantly larger than the direct grant costs.

One result of the project is the capability of the Center to offer services to aid criminal justice agencies on an extremely economical basis. Geo-coding, is being accomplished for less than a penny per record on the County-wide file. The current DIME file is producing match rates exceeding 88% for County-wide CAPER. This minimizes the necessary manual geocoding that would otherwise have to be performed for non-matched records, and is timely in that the work can be accomplished on an overnight basis. GRIDS map production has also been refined for user agencies, is fully installed at County data processing, and is an economical producer of varying scale hard copy maps of geo-coded events.

GADS, the interactive computer graphic system, has been demonstrated as a highly efficient system for problem-solving purposes. The system requires low utilization of computer resources and has high effectiveness as a communications device. Terminal facilities are easily installed in agency offices. Police department personnel can be efficiently trained to use the system for operational purposes.

The GADS system has been benchmarked at three commerical installations using two lessons in the training manual developed in the San Jose Police Department Beat Analysis project. Commerical rates are in the range of $15 -$25 per connect hour for computer utilization, depending on the kind and pace of work. Hardware costs for modems, terminal and hard copy unit are approximately $3.50/hour. GADS operating costs are thus modest relative to system benefits. It would be appropriate for the existing version of GADS to be installed at County GSA-Data Processing and made available to criminal justice and other governmental agencies in Santa Clara County.

It is recognized that GADS is a prototype system, not a documented production product. Additional resources should be made available to develop and implement desirable modifications to the system. The interactive computer graphic system underwent several modifications during the course of the project. These changes were responsive to demonstration project experience with client agencies; the prototype system would benefit by additional tailoring to police agencies specifications and enhancement of user manipulation capabilities.

A forthcoming version of the SPSS will be interactive, rather than batch oriented, computer processing. It will be desirable to revise the batch procedures developed under the demonstration projects in a more convenient form for direct agency utilization on remote terminals.

The Center is undertaking preliminary work concerning the problem of data entry for smaller jurisdictions. Batch data entry of dispatch data costs approximately $.10 per record simply to establish machine readable card images. These costs should be able to be reduced and the information made more accessable to agencies through recent technology. Additional resources should be made available to support this effort. Many Census Bureau procedures for the maintenance and use of the GBF/DIME required modification and special development efforts by the Center. This work, as well as basic maintenance of the GBF/DIME, is necessary to be continued and supported.

In order to achieve the objective of the second project year, to institutionalize the Center on a self supporting basis, project expenditures were fundamentally necessary to develop establish techniques that would serve client agencies. The opportunity to serve one non-criminal justice client in the first project year has been fruitful in producing a number of users from the non-criminal justice area. These users improve the probability that the Center will be able to be self supporting in forthcoming years by spreading the costs of DIME file maintenance among different interest areas. The Center has experienced successful transferal of techniques and methods in a variety of projects with local government agencies, such as school and fire applications. To the extent that development costs can be underwritten by a variety of users, no single interest area is unfairly burdened and the benefits are distributed to the community of users.

FINDINGS AND RECOMMENDATIONS

The projects chosen as demonstrations were selected on the basis of need and request for assistance by the user agency. Agencies represented included Adult Probation and five different law enforcement agencies of four generally different sizes plus a crime analysis and evaluation project serving all police agencies in the County. In addition, several one-time jobs unrelated to a project were completed for client agencies. A total of eight (8) demonstration projects were undertaken and are described in the following section. One non-criminal justice agency project supported under the first year funding is not summarized.

Project objectives were achieved in terms of accomplishing the number of projects anticipated, and in institutionalizing the Center within County government. Project experience in terms of achieving the objectives of assisting agencies to make decisions and act on solutions to important criminal justice problems were uneven. One agency (Campbell Police Department) has made only slight use of the information developed in their project. The Sheriff's Department project was unable to accomplish project objectives for a variety of reasons, including place identification problems and lack of staff continuity. Other agencies utilized the information and tools provided to a greater extent and fulfilled the intent of the demonstrations. The San Jose Police Department Burglary Analysis Unit project exemplified the full range of the objective "to assist the diagnosis of specific problems, develop better baseline measures by geographic area, formulate more effective programs, and evaluate results."

In general, the demonstration projects focused on using calls-for-service (dispatch) data to assist in the mangement problem-solving tasks related to strategic deployment, agency organization, resource allocation, activity analysis; and crime-specific program planning, monitoring and evaluation. Two basic methods of organizing and presenting the data have enabled agencies to successfully operate in a mode that places the decision-maker in direct contact with data relevent to a current problem. They are:

1. An interactive computer graphic system.

   A prototype interactive computer system, GADS (Geo-data Analysis and Display System), developed by IBM-Research, San Jose, under a joint study agreement with Santa Clara County, is a user-oriented system that relates a structured data base to a user-specific map. The system provides opportunity for exploratory examination of the data and capacity for sophisticated manipulation of data and map features. The terminal is a relatively inexpensive, remotable device with hard copy attachment. The system is economical in computer utilization. The control language has been successfully operated in-house by uniformed and non-uniformed personnel of diverse backgrounds in a variety of police agency problems.

2. A general purpose statistical and tabulation system.

   The program package SPSS (Statistical Package for the Social Sciences), developed by the National Opinion Research Center at the University of Chicago, now disseminated and maintained by SPSS, Inc., is a flexible, coherent set of procedures for data analysis. The system is characterized as user-oriented in that users can readily modify previously developed procedural controls to produce the desired analysis.

The necessary conditions for either of these systems to be employed are:

1. Systematic data collection and data entry;
2. Geographic coding of data; (The Center performs the "geo-coding" functions of identifying X-Y coordinates and user zones (beat, study areas, etc.) with event data, using Census developed methods (DIME file and the program UNIMATCH), and additional procedures developed by IBM-Research and the Center)
3. Data base loading, maintenance and assistance in problem solving. (Functions also performed by Center staff.)

Summary and Conclusions

The police agency projects undertaken under the previous limited funding have been restricted in duration and scope; e.g. data collection of two or three months, 1/12 sample data, etc. These demonstration projects have generated valuable insights and modifications to the GADS system, initialized the DIME file and refined its use in police applications, and demonstrated the versatility and transferrability of SPSS procedures in the calls-for-service analysis. In addition, the projects generated insights for the participating agencies into their problem solving techniques. The direct involvement of administrators, managers, supervisors, and line personnel in the solution process produced - if not the optimum solution - at least a solution that all levels involved in the problem could understand and could support. The speed and flexibility with which alternative solutions could be tried and evaluated, plus the minimal use of a programmer/analyst to which a problem definition had to be communicated, encouraged a re-evaluation of the factors and criteria used in the methodology. Users with preconceived ideas about which factors were important in deciding how to redistribute service areas often found, during the course of solution development, that some factors were not as critical as hypothesized and that other factors needed to be examined more carefully.

The projects accomplished under the LEAA grants have demonstrated the combination of a user-oriented, flexible and versatile interactive graphics computer system plus traditional statistical analysis techniques provide law enforcement administrators with powerful tools to assist in finding solutions to operational and management problems.

Recommendations

1. It is recommended that the interactive computer graphic system, GADS, be developed beyond the current prototype stage into a documented production version, installed at the Santa Clara County Data Processing Center and made available to criminal justice and other governmental agencies within the County. Several important modifications to the prototype should be accomplished. The most significant change concerns an adaptation of the system architecture to a configuration less dependent on a mainframe computer, with greater emphasis on establishing capabilities at the distributed end of the system. The distributed end of the current prototype is a terminal, connected by telephone line to a host computer where the data management occurs and the program is executed. The revised system should utilize either minicomputers or intelligent terminals at the distributed end, to achieve economies in operation, reduce mainframe contention, improve analysis capabilities through supplementary software, and provide opportunity for less costly and more efficient data entry. The distributed architecture should utilize mainframe capabilities for data management and certain program economies, but should provide users with maximum cost-effective capabilities at the remote end. The intelligent terminal/minicomputer offers smaller agencies a means to retain control over data entry while providing the necessary capability to review and summarize the data in a timely manner. The connection to a mainframe would enable large scale data base establishment and manipulation capabilities in an appropriate environment.

2. It is recommended that the interactive version of SPSS, when available, be installed at County Data Processing and made available to criminal justice agency users throughout the County. This recommendation is consistent with the findings from the projects undertaken by the Center supportive of placing flexible analysis tools as close to decision-makers as possible.

3. It is recognized that user agencies will be required to maintain systematic data collection efforts as the basis for analysis. Project experience supports a conclusion that data collection costs and efforts are rapidly amortized if the information is used in decisions regarding the allocation of criminal justice agency resources. The use of information in such decision-making is dependent on the ability of the system to respond in a timely, appropriate and communicable manner to selective inquiries.

SUMMARY PROJECT DESCRIPTIONS

1. Adult Probation Department Case Assignment Project

A. Objectives: Assist a determination for geographic assignment of supervising probation officers.

B. Methodology: In order to analyze the mobility of probationers over a two year period, the residences of a sample of approximately 630 probationers (10% of males and 20% of females), previously drawn for a Probationer Needs Study by the Criminal Justice Pilot Program, was obtained from Departmental records. If probationers move a great deal, case assignment by geographic area could require a large number of transfers among probation officers. Probationer moves were analyzed to shed some light on the potential magnitude of this problem. It was determined that almost 51% of the cases did not move in this period, 26% moved only once, and 23% moved 2 to 4 times. A second sample of 1973 probationers was drawn to determine if the then-current distribution of probationers was similar to the 1972 pattern. Both samples listed sex and type of offense for each probationer. Offense types were classified as felony, misdemanor-drunk driving, misdemeanor-drugs, or misdemeanor-other; the most serious offense was identified when the probationer was paroled for more than one offense. The proportion of males and females was almost the same in the two samples, however the proportion of probationers by offense type showed increases in misdemeanor-drunk driving (from 7 to 11%) and misdemeanor-drugs (from 3 to 10%), with a decline in felony (from 54 to 42%). From 1972 to 1973, the number of cases appeared to have increased somewhat in the north county and south San Jose areas, and decreased slightly in the mid-county areas.

Supervising Probation Officers identified alternative groupings of census tracts as potential service areas by indicating boundaries on plastic overlays over a census tract map base. Computer line printer (GRIDS) maps at the same scale as the census tract base maps, indicating the residences of probationers, by type, for the samples, graphically represented the distributions. A computer program developed by Center staff analyzed the number of moves made by probationers (1972 sample) into, out from, and within each grouping of census tracts.

C. Results:

Probationer moves over the two year period prior to May 1972 were tested against configurations of 5,6, and 9 service areas. The numbers of transfers increased, as expected, with the number of service areas. Up to 5 areas, each additional service area caused about 200 additional transfers per year. When the county was divided into five test service areas, about 47% of the moves occured within the test areas, while 26% of the moves which crossed service area boundaries were into or out of the county and would require a case transfer whether or not service areas were established. Analysis of the number of actual cases involved in the moves indicated that about 7.7% of the caseload would be involved in interzonal moves every year if the system of 5 service areas were instituted.

This analysis formed a significant part of Departmental consideration regarding the establishment of service area assignments. Qualitative factors were also an integral part of the decision-making by departmental personnel. It was reported that misdemeanor-drunk driving cases were to be assigned on a service area basis as a test of the geographic case assignment concept.

Exhibits A-C illustrate portions of the analysis. Exhibit A illustrates a grouping of 6 service areas with the case distribution for both samples. Exhibit B illustrates residence change summaries for the 6 service area example. Exhibit C summarizes the moves by probationers for a 5 service example.

2. San Jose Police Department "Beat Design Project"

A. Objectives: To redesign the service area (beat) structure in order to equalize workload while maintaining a prescribed level of available patrol time.

B. Methodology: The first phase of this project was implemented and executed by means of a joint study between IBM-Research and SJPD, during which time 12500 call-for-service (CFS) records were collected, categorized and aggregated to a computer-based map of San Jose divided into 273 beat building blocks (BBB's) (Exhibit 1). Teams of SJPD personnel were trained and used IBM's Geo-data Analysis and Display System (GADS) to aggregate BBB's into larger geographic service areas (beats).

The data was comprised of a sample of 30 days of data selected from throughout 1972 and balanced for proportional weekends/week days, holidays/non-holidays and seasons of the year. Variables measured in the BBB and beat aggregation process were Total CFS, Multiple Unit CFS, In-Progress CFS (priority) and Consumed Time in minutes (from call received to event cleared). Each variable was aggregated to 4-hour spans of the day as well as totaled for each day and totaled for all days. No attempt was made to capture the type of event because, for San Jose's resource allocation problem, any priority call would be treated the same regardless of type or nature.

The second phase of this project was executed through a joint study agreement between IBM-Research, SJPD and Santa Clara County. The Center assumed responsibility for providing all hardware exclusive of the host computer, all data preparation services and technical assistance as required. For this phase a second set of data of 13,677 CFS records was sampled from 1974 activity and loaded into GADS. This data was aggregated similar to the 1972 data except that Multiple Unit CFS were not used and variables were aggregated by 1-hour time span instead of 4-hour spans. The 1972 data was reloaded into GADS by 1-hour time span to be compatible with the 1974 data.

Again, SJPD personnel used GADS to aggregate BBB's into beats based on the variables available and criteria as defined by the agency. The primary difference between the two phases was that, during the first phase all GADS usage was performed at IBM's Research facility while during phase 2, SJPD was provided with an on-site terminal, located in the office of Research and Statistics.

C. Results: Phase 1 actually had 2 ending points. Upon reaching the first, SJPD personnel determined that the existing structure of 34 beats (Exhibit 2) should be expanded to 43 beats. City budgetary limitations then imposed a constraint of 40 beats and the problem was re-executed. A 40 beat solution (Exhibit 3) was rapidly developed which conformed to all constraints and objectives for number of beats, equal workload between beats and preservation of minimum preventive patrol time. Exhibit 4 is a 2-dimensional scattergram displaying the value of total consumed time for each of the old 34 beats, and Exhibit 5 is a scattergram displaying the same data for each of the new 40 beats. The new beat structure provided a significantly more equitable distribution of workload between beats in terms of consumed time.

The phase 2 work involved re-examining the old beat structure with regard to a finer breakdown of the data and to validate the new beat structure with data sampled subsequent to its implementation. With very few exceptions the new beat structure proved to be as workable as the data indicated it would be. The exceptions were easily accounted for in terms of known shifts in the make-up of the city and adjustments to the beat plan were readily formulated.

3. San Jose Police Department "Burglary Prevention" Project

A. Objectives: To assist the Burglary Analysis Unit (BAU) of the San Jose Police Department in its Tactical Research Approach Concept by providing specialized data processing services; specifically, by providing data bases of selected offense report data, geo-coding the data, producing graphic and tabular displays of the data and providing technical assistance as required.

B. Methodology: From SJPD's offense reporting system all burglaries for 1973 and 1974 were extracted, corresponding to the periods before and after several of the BAU's programs had been implemented in selected target areas. These records were then geo-coded to x-y coordinates, Census Tract and Beat Building Blocks (BBB's) as defined by SJPD in their calls-for-service study. The data was then used to prepare statistical tables of burglaries by Census Tract, Month and Year, and "GRIDS" maps (Exhibit 7) of Total Burglaries and Private Residence Burglaries by year.

The County Assessor's Unsecured Master File was geo-coded, categorized and tabulated to produce a listing of the number of businesses by category by Census Tract by year (Exhibit 8). This set of tables, in conjunction with the count of burglaries by Census Tract, provided the BAU with the key factors needed to produce a "risk factor" for commercial burglary. It should be noted here that this use of the Assessor's data was the first attempt by a local law enforcement agency to incorporate another agency's non-law enforcement type data into their analysis through the Center's capabilities.

In addition to the GRIDS maps and statistical tables, the data was loaded into GADS and BAU personnel were trained in the use of the system. BAU personnel were able to use the unique area definition and aggregation capabilities of GADS on the terminal installed at SJPD for the calls-for-service study, in their own environment at their own schedule and pace. It should be noted that while the calls-for-service study was concerned with data as aggregated by BBB, the BAU study was concerned with viewing and listing the data by event and aggregating the data by target areas unique to their work; the capabilities were directly available through the GADS system (Exhibit 6).

C. Results: Formal evaluation of the Burglary Prevention Study has not yet been completed, consequently the final results and conclusions of this project have not yet been determined. However, BAU personnel have found GADS extremely useful for defining target areas and extracting pre-and post-program implementation control data. The ability to integrate demographic, census and land use data into the problem analysis phase enabled BAU personnel to develop new and useful measurement functions.

4. Santa Clara County Sheriff "Patrol Analysis" Project

A. Objectives: To determine the feasibility of providing geo-coding capability for the Sheriff's patrol activity report data, to reference Census and property use data to the Sheriff' reporting areas and to demonstrate graphical and tabular output of the data.

B. Methodology: A variety of GRIDS maps of Fiscal 1973/74 data from the Sheriffs patrol activity reporting system (COPANA) were produced. Because of the volume of data collected in the Sheriff's patrol activity reporting system (over 131,000 events in fiscal 1973-74) a sample of data was selected for further detailed analysis. This sample consisted of all events during the last 2 weeks of June, 1974 and totaled 5,701 events.

One of the unique features of the Sheriff's reporting system was the assignment by each reporting officer of an event location in terms of a grid code. This grid code provides for a maximum resolution of 6,250,000 square feet (each grid is a square 2,500 on a side); therefore, one of the first problems was to provide geo-coding capability to a finer resolution so that the data could later be aggregated into more meaningful analysis areas (e.g. census tract, beat, traffic zone, grids of smaller dimensions, etc.). As a baseline for this effort, an example set of GRIDS maps displaying Total Events, Total Part 1 Events, Total Felony Events, Total Minutes Consumed on Felony Events, Felony Events as a Per Centage of Total Events and Time Consumed on Felony Events as a Per Centage of Total Consumed Time was produced (Exhibits 9-11) by assigning the location of each event to the centroid of the grid cell it was reported to be in. The plan was then to geo-code the data to x-y coordinates and to re-execute the GRIDS maps to the finer resolution thus provided.

Unfortunately, while several powerful programs and many techniques are available for unscrambling free-field address data, the quality of address data in the Sheriff's reporting system had degraded so severely as to almost preclude the mechanical geo-coding process (Exhibit 12). Many attempts were made to convert street and place names to consistent values and to standardize the address part of each record into the identifiable components of direction, name, type, city and house number. After expending over 200 man-hours on the geo-coding process, the best match of data to reference files was less than 50%.

The additional amount of resources required to complete the geo-coding process was deemed too extensive due to the quality of source data. Because of pressing requirements of other agencies, this project was administratively assigned a relatively low priority and little effort was subsequently expended.

C. Results: The geo-coding process is critical to the analysis of geographically related data. The difficulties encountered in geo-coding the Sheriff's data precluded the accomplishment of most project objectives. The feasibility of producing hard copy GRIDS maps of selected data was proven and demonstrated. It was determined that, until such time as the Sheriff's Office can implement increased discipline in their location data collection, geographic analysis of their data will be limited in resolution to that provided by their manual grid coding system.

5. Mt. View Police Department "Calls-for-Service" Project

A. Objectives: To determine and demonstrate the feasibility and usefulness of displaying MVPD CFS data in graphical and tabular forms as aids to management and operational information systems.

B. Methodology: A 100% sample of MVPD dispatch data from July, August and September, 1974 (22,808 events) was selected, edited and used as input to produce SPSS reports and input to GADS. The source documents of dispatch data were edited manually by an intern provided by the Regional Criminal Justice Planning Board (RCJPB). Missing data elements and consistency of addresses and activity codes were resolved in this process. The data was then keypunched and machine edited for missing data elements and consistency.

Statistical tables produced included a count of the number of events by emergency/routine code, time of day (2-hour spans), day of the week and by month; a count of the response time and total time expended by event category and month; and several different combinations (crosstabulations) of these factors (Exhibits 13 and 14).

The data was also geo-coded to x-y coordinates and loaded into a GADS system, using a map defined by MVPD personnel of 251 BBB's (Exhibits 15 and 16). A terminal was installed at MVPD headquarters for the analysis phase of the project and MVPD personnel were trained to use the system. The powerful selection, extraction and aggregation capabilities of GADS were then used to analyze the data for a wide variety of factors (Exhibits 16 and 17).

C. Results: MVPD did not pursue a clearly defined management problem through GADS, however, they found GADS to be an easy and versatile system to use in responding to a variety of management and operational questions. Of particular interest to Mt. View management was the availability of data on administrative calls and the time spent on them. For the first time, MVPD administrators were able to count and comprehend the time spent on meals, personal breaks, training, court appearances, vehicle maintenance, public relations, and similar activities.

One of the key factors in the relative success of this project was the manual screening of the source data prior to data entry. Inconsistencies in source data, similar to those described with the Sheriff's patrol analysis report data neccessitate a form of screening or manual intervention.

6. Campbell Police Department "Call-for-Service" Project

A. Objectives: To demonstrate the feasibility and usefulness of displaying CPD dispatch data in graphical and tabular forms as aids to management and operational information systems.

B. Methodology: A 100% sample of CPD dispatch data from September and October, 1974 (13,350 events) was collected and edited for input to produce SPSS reports and input to GADS. Source documents were screened manually by an intern provided by the RCJPB for missing data elements and consistency of codes. The data was then keypunched and machine edited for consistency and validity.

A total of 19 statistical tables were produced using SPSS (Exhibit 18). As for Mt. View, CPD management was particularly interested in analysing data relevant to consumed time for administrative functions and low priority calls (e.g. "barking dog" calls, motorist assist calls, etc.).

The data was also geo-coded to x-y coordinates with the intention of loading it into a GADS system. With this in mind, CPD defined a set of 42 RBB's which were translated into a GADS map (Exhibit 19).

Subsequent to geo-coding the data but prior to actually installing a terminal or training CPD personnel to use GADS, CPD management decided that the age of data precluded its use for tactical resource deployment problems. CPD also concluded that relevant management policy data had been produced through the earlier tabulations and the project was terminated.

C. Results: Tactical resource deployment depends on the use of essentially current data in a continuous flow. The costs of generating this kind of data from non-automated data systems are relatively high and time lags are difficult to overcome. CPD objectives to use project collected data for operational deployment were inappropriate given the nature of data capture, data entry and data validation. Traditional forms of data collection are feasible for providing input to long range strategic and management problem solving analysis programs.

The concept of using calls-for-service data to capture management information relative to the administrative chores of the police function remains a viable approach. Even without the geographical capabilities of GADS, CPD management was able to document, using SPSS, the amount of field officer time lost to non-patrol activities.

7. Gilroy Police Department "Calls-for'Service " Project

A. Objectives: To demonstrate the feasibility and usefulness of displaying GPD dispatch data in graphical and tabular forms as aids to management and operational information systems.

B. Methodology: A 100% sample of GPD dispatch data from July, August and September, 1974 (14,655 events) was collected, screened, edited and input to SPSS and GADS. The source documents were screened manually by a RCJPB intern for missing data elements and consistency of codes. The data was keypunched and machine edited.

A set of statistical tables, similar to those produced for CPD and MVPD, were produced. GPD was particularly interested in analysing the data relevant to consumed time for administrative functions (e.g. vehicle service, court appearances, transportation of prisoners/evidence, etc.). Summary graphs of total events (less administrative calls) by day of week and time of day are presented as Exhibits 20 and 21.

In addition to the SPSS tables, the data was geo-coded to x-y coordinates and loaded into a GADS system. GPD personnel were trained to use the system, a terminal was installed at GPD headquarters, and GPD personnel then used GADS to analyse their data. Exhibits 22-25 are representative of the type of analysis presentations available to GPD management in GADS.

C. Results: GPD management, staff and line personnel reported considerable utility and efficiency with the types of analysis provided through the combination of hard copy statistical reports and interactive graphical inquiry/retrieval systems. Without being buried by volumes of paper they were quickly and easily able to extract precisely the data needed to answer specific questions.

8. Countywide CAPER Project

A. Objectives: To demonstrate and subsequently to provide geo-coding and GRIDS map services.

B. Methodology: Countywide CAPER (Crime Analysis - Project Evaluation-Research) is an LEAA grant funded project designed to collect and process offense report data for all law enforcement agencies in the county. The Center processed CAPER data previous to and including calendar year 1974 as projects under grant funds. CAPER has subsequently sub-contracted with the Center to geo-code the data and then to produce GRIDS maps from the geo-coded data.

In practice, CAPER personnel regularly visit each agency and transcribe offense data onto coding sheets from which two files on magnetic tape are produced. One file is used as input to CAPER's own SPSS routines and the other file is hand carried to the Center for geo-coding. After the file has been geo-coded it is returned to CAPER where the rejects from the geo-coding process are coded by hand. The completed file is then returned to the Center where a variety of different GRIDS maps are produced depending on the requests from each agency.

C. Results: From January 1974 to June 1975 the Center has processed 175,769 CAPER records and geo-coded a total of 131,340, or 74.7% overall. The Center's target for geo-coding is a 90% match rate and is currently running with better than an 88% rate, with the target rate expected to be achieved before the end of this calendar year. Current geo-code processing consumes less than 3/100 second per record (on the County's IBM S370/158 computer) and about 1 man-hour per batch of data - regardless of the number of records. As an example, at current rates geo-coding a batch of 30,000 records would cost about $110 or about $0.0036 per record. The current agreement with CAPER provides for a turnaround of their data in one working day, and no problems have been encountered meeting this commitment.

The GRIDS processing is a much less "production" type job than the geo-coding because of the varied requests for maps. However, one of the most powerful aspects of GRIDS is its ability to produce an almost endless variety of maps without limiting the user to a set number of maps produced on a regular basis which may not be of interest at any particular point-in-time. Currently, a typical GRIDS job consists of one map displaying total events and total burglaries for each of the thirteen county agencies, plus maps showing the same data for two of the Sheriff's contract cities, plus three special interest maps (one for each of three of the 13 agencies)- each calendar quarter.

Because of the custom nature of each GRIDS job the maps are produced in two steps, with the output of the first step being verified before the maps are printed on vellum paper in step 2. For this reason a maximum of three working days are required for completion of each job, assuming that no significant changes have been requested. If significant changes have been requested then the job turnaround will depend on the complexity of the changes and staff resources available to implement them; no jobs have been delayed more than 1-2 days for this reason.

9. Miscellaneous Short Projects

A. Land Use Changes and Reported Offenses

The San Jose Police Department in conjunction with the Stanford Research Institute requested the Center's assistance in determining whether the change from fraternity/sorority residences to day care centers and half-way houses, in the vacinity of San Jose State College, had any effect on crime rates in that area.

The predominant change in land use occurred during 1972-73 and, consequently, two samples of data were selected from SJPD's offense reporting system. Each sample consisted of all records for the first half of each year used in the study. The data was then geo-coded and used to prepare GRIDS maps (one for each year) displaying events in the campus area (Exhibits 26 and 27).

Analysis of the data and the results, conclusions and recommendations of the study have not been made available to the Center as of the date of this report.

B. Campbell Police Department Burglary Study

In 1973, prior to the advent of Countywide CAPER, the Campbell Police Department sponsored a study of burglaries in their city. As part of this study, they collected data for all burglaries from 1971 thru 1973, manually geo-coded this data and commissioned the Center to produce 11 GRIDS maps. The maps displayed data values as follows:

1971 Burglaries
1972 Burglaries
1973 Burglaries
Burglaries of Public Places
Burglaries of Commerical Places
Burglaries of Residential Places
Burglaries of Autos
Burglaries from 8 A.M. - 4 P.M.
Burglaries from 4 P.M. - 12 A.M.
Burglaries from 12 A.M. - 8 A.M.
Total Burglaries

Exhibits 28 and 29 are two of the GRIDS maps that were produced.

Analysis of the data and the results, conclusions and recommendations of the study have not been made available to the Center as of the date of this report.

C. Criminal Justice Pilot Program

The Criminal Justice Pilot Program (CJPP) has been involved in many aspects of law enforcement information analysis. One of the in-house studies performed by CJPP involved an in-depth analysis of robberies in San Jose. As part of this study CJPP sampled and prepared offense report data for robberies in San Jose during 1971-1973 and, among other things, requested the Center to prepare the following GRIDS maps from the data:

Total Commercial Robberies	Total Non-Commerical Robberies
Cleared Commercial Robberies	Cleared Non-Commerical Robberies

Residences of Adult Robbery Offenders
Residences of Juvenile Robbery Offenders
Residences of All Robbery Offenders
Average Travel Distance Between Event and Suspect Residences and Number of Events

Some of these maps were produced several times to display the data at different scales and resolutions. Exhibit 30 is an example of one of these maps.

APPENDIX A

G A D S
TECHNICAL CAPABILITIES AND ASSUMPTIONS

SYSTEM SUMMARY

The Geo-Data Analysis and Display System (GADS) is an on-line interactive graphics system which provides users with a set of generalized capabilities to select, display, manipulate, and interpret data in a sequence which they determine during interaction with the system. GADS consists of two subsystems:

1. Extraction

   This subsystem provides a set of interactive functions for selecting, aggregating and subsetting geo-coded data from multiple "event" files, to form an on-line "extracted data base".

2. Analysis and Display

   This subsystem provides a set of interactive functions for conversational problem exploration and solution through the display and manipulation of maps and map-related data.

APPENDIX A

GADS

SYSTEM DESCRIPTION

GADS, as previously mentioned, consists of two subsystems: (1) Extraction and (2) Analysis and Display. The components of each of these subsystems are shown in Figure 1. Note that the "build and maintain" functions which are not part of GADS are shown as being separate from the GADS components.

Each of the GADS subsystems shown in Figure 4 are now described:

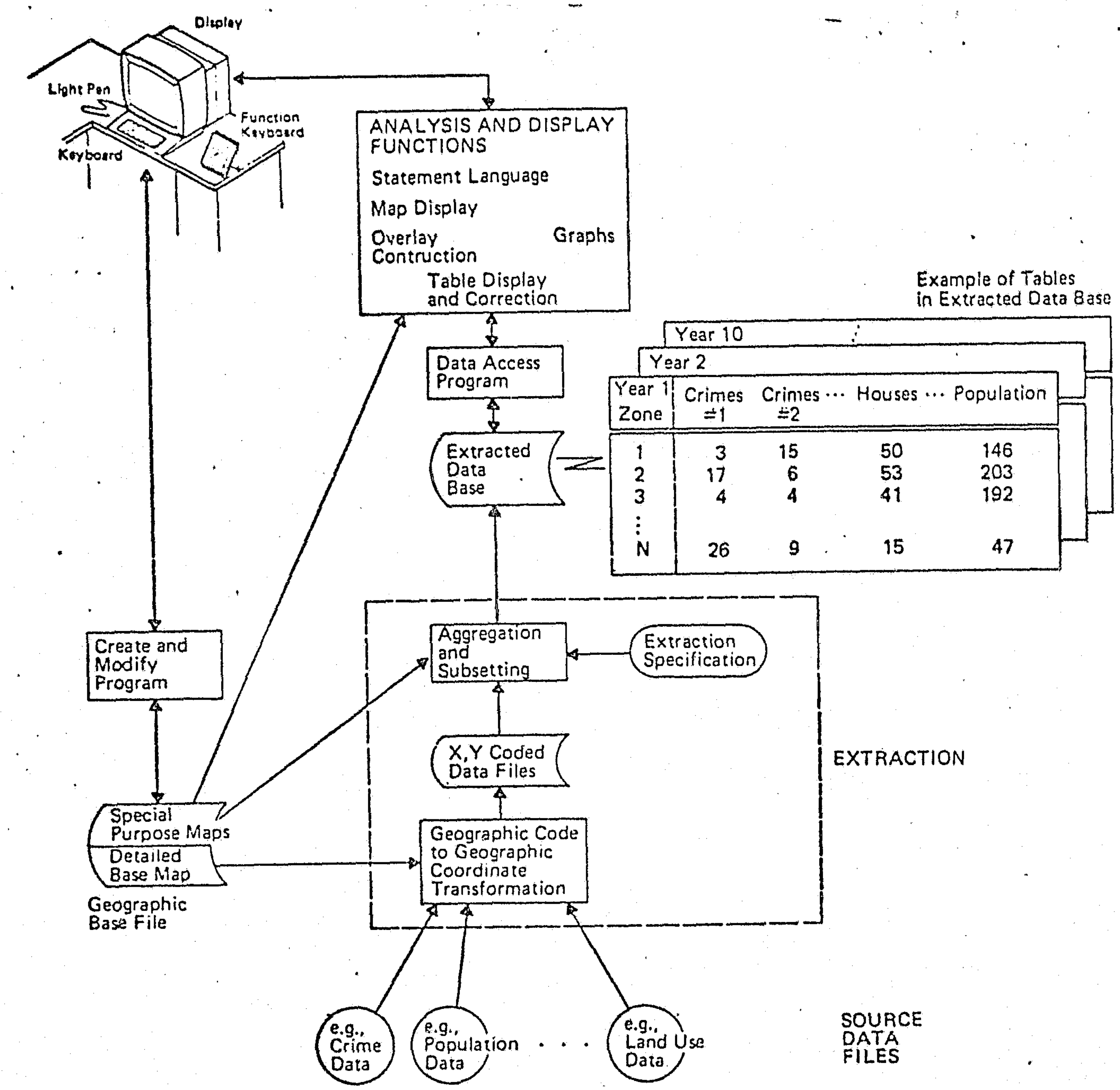
1. Extraction

The GADS extraction subsystem performs the extra-action of data from multiple event files. The inputs to extraction are the event files; a base map, and a selection, aggregation and subsetting specification; the output is an on-line extracted data base. These components are described below:

a. Event File

The event files (created using the Census Bureau's UNIMATCH or ADMATCH programs) contain an x-y coordinate in each of its records, so that extracted data can be related to points, lines, or polygons on the computerized base map.

There must be record definitions for each event file; these are supplied by the user to GADS and are stored in the computer.



~~GADS architecture~~

GADS ARCHITECTURE

FIGURE 1

~~FIGURE 4~~

APPENDIX A

b. Map Display and Overlay Construction

The map display functions support the display of the symbols created with GADS statements. The user can display one or more maps, with one or more sets of symbols, either simultaneously or consecutively. Maps may be expanded around any zone, and lines between zones containing the same symbol may be eliminated to form a sub-map of areas with similar data values. The user may point to a symbol in a zone to get a display of data values for that zone. A typical display would contain 1000 lines and over 250 symbols. During use of these functions, several displays might be requested (e.g., change scale, change symbols, change map).

The overlay construction functions are used to create, alter, retrieve and save maps. These overlay maps must be formed from combinations of the basic zones in the base map. To identify zones, one can display statement-created symbols, dots, or numbers in each zone. By pointing to the identifiers, existing zones can be extended or new zones created from combinations of basic zones. All maps can be enlarged to facilitate overlay construction. Because overlay maps can be saved in a map library, they can be referred to in the statement language to form map displays or new variables based on data values aggregrated according to an overlay map. Use of the overlay construction functions will usually require frequent map redrawing to reflect changes.

c. Graph Display

Scatter diagrams are a familiar mode of data presentation and are the second data display mode in GADS. With these functions the user may create one, two, or three dimensional scatter diagrams. Each point (or line in 3-D) on the scatter represents one zone. The zone number for any point is displayed if the point is pointed to, and any zone number can be entered and the appropriate point on the graph will be flagged. The scatter diagrams are automatically scaled, unless the user enters specific scales on each axis. The user may request a cumulative summation of the y-axis variable.

d. Table Display and Manipulation

The third data display mode in GADS is Tables. The Table based functions can be used to: display any Table in the extracted data base (either any n-tuple in a Table or the values of all variables for any basic zone), alter values in a Table and log the changes, and print hard copies of the Tables. This dictionary can be used as a reference in constructing statements. As a data protection mechanism, the user may create one or two working Tables in the extracted data base and move data between these Tables and other Tables in the extracted data base.

APPENIDX A

GADS

Reference to Exhibits Displaying GADS Functions

Exhibit 1	Maps and Symbols	San Jose CFS Study
Exhibit 2	" "	" " " "
Exhibit 3	" "	" " " "
Exhibit 4	2-Dimensional Histogram	" " " "
Exhibit 5	" " "	" " " "
Exhibit 15	Maps,	Mt View CFS Study
Exhibit 16	Maps and Symbols	" " " "
Exhibit 17	3-Dimensional Histogram	" " " "
Exhibit 19	Maps and Symbols	Campbell CFS Study
Exhibit 22	3-Dimensional Histogram	Gilroy CFS Study
Exhibit 23	Maps and Symbols	" " "
Exhibit 24	" "	" " "
Exhibit 25	Maps and Extracted Data	" " "

County of Santa Clara General Services Agency: Data Processing Center	BY P. Watanabe		SECTION 1.0	
	APPROVAL		REVISION 00	PAGE 1
	APPROVAL		DATE 3/9/74	

Advanced Analysis/Mathematical Library
SPSS
Introduction

1.0 Introduction

SPSS, Statistical Package for the Social Sciences, is installed and maintained by GSA DPC. The primary purpose of this Report is to document the Procedures adopted by GSA DPC for the use of SPSS. This Report is written to be used as a supplement to the regular SPSS manual by documenting Installation dependent operational conventions.

It is the Users' responsiblity to procure SPSS manuals. However, GSA DPC will make available a limited number of SPSS manuals on a loan basis. A SPSS manual will so be available at GSA DPC to be used as a reference manual when using the facilities at the Center.

2.0 General Program Description

The Statistical Package for the Social Sciences (SPSS) is an integrated system of computer programs for the analysis of social science data. The system has been designed to provide the social scientist with a unified and comprehensive package enabling him to perform many different types of data analysis in a simple and convenient manner. SPSS allows a great deal of flexibility in the format of data. It provides the user with a comprehensive set of procedures for data transformation and file manipulation, and it offers the researcher a large number of statistical routines commonly used in the social sciences.

In addition to the usual descriptive statistics, simple frequency distributions, and crosstabulations, SPSS contains procedures for simple correlation (for both ordinal and interval data), partial correlation, multiple regression, factor analysis, and Guttman scaling. The data-management facilities can be used to modify a file of data permanently and can also be used in conjunction with any of the statistical procedures. These facilities enable the user to generate variable transformations, to recode variables, sample, select, or weight specified cases, and to add to or alter the data or the file-defining information. SPSS enables the social scientist to perform his analysis through the use of natural-language control statements and requires no programming experience on the part of the user.

3.0 Program Reliability/Maintainability

3.1 Overview

SPSS is distributed by NORC (National Opinion Research Center) located at the University of Chicago. This group is responsible for upgrading and documenting the capabilities of SPSS. We have made a formal request to NORC for information about their general Reliability/Maintainability policy for SPSS. Their response will be documented as a revision to this section. Presently there is no formal commitment by NORC to guarantee the performance of SPSS.

3.2 Program Reliability

SPSS has been an operational system since 1967. Studies conducted at Stanford University have shown that the SPSS procedures are very reliable. SPSS has been successfully implemented in over 200 facilities. These facilities include major Universities and government agencies. Various Santa Clara County departments and agencies have successfully used SPSS.

3.3 Program Maintainability

A Support Agreement can be purchased from NORC for the maintenance of SPSS. The Agreement entitles GSA DPC to all new releases, program fixes and subscription to the SPSS Newsletter. The newsletter documents known errors and bypasses available.

3.4 Problem Handling

Should problems occur during the execution of SPSS, the problem description and associated run documentation should be submitted to GSA DPC for resolution.

4.0 General Operational Description

SPSS is an extremely flexible Statistical Package. It not only has the ability to perform various statistical routines for data analysis, it also has fairly extensive input/output capabilities. Unfortunately such capabilities require the user to have not only knowledge of SPSS Procedures but also a very extensive knowledge of IBM JCL (Job Control Language).

Additionally there are GSA DPC constraints that must be considered when using SPSS. The most significant GSA DPC constraint is the limited availability of permanent on-line disk storage areas for SPSS generated System Files.

GSA DPC has designed several SPSS Catalogued Procedures which will minimize the amount of JCL knowledge required by the User. If the User cannot conform to the constraints specified for the use of these Procedures, he must code his own JCL. All JCL must conform to GSA DPC Standards documented in the <u>Facility User Guide for GSA DPC</u>.

5.0 Detailed Operational Description

5.1 <u>Overview</u> - This section will describe the SPSS Catalogued Procedures installed at GSA DPC. The use of these Procedures will minimize the amount of JCL that the User is required to code in order to make an SPSS computer run. However, the User is cautioned to observe the limitations imposed by the Catalogued Procedures.

5.2 <u>Catalogued Procedure Limitations</u> - The Catalogued Procedures are extremely useful when the User is creating and/or retreiving SPSS generated System Files. The Catalogued Procedures will automatically store User's SPSS System Files on a permanently resident Disk File(TEST.JAA.PRMUS). Subsequent processing can be accomplished without any additional Operater intervention. Hence, these subsequent runs should experience better response times(turn-around).
The amount of disk space presently allocated for TEST.JAA.PRMUS is twenty(20) cylinders. Assuming an average sample size(# of cases) of 1,000 observations, there is sufficient space available for approximately sixty(60) System Files. Should the capacity of this disk area be exceeded, SPSS runs will terminate abnormally. SPSS System Files must be purged from TEST.JAA.PRMUS before any subsequent processing commences.

5.3 <u>SPSS System File Management</u>

5.3.1 TEST.JAA.PRMUS Usage - The following rules governing the use of this disk area should be observed:

(1) Individual System Files should not contain more than 1,000 observations. However, should space be available and approval is granted, the User may execeed this limit. Requests should be directed to Mary Wiggers at ext. 4161.

(2) Individual System Files should not be kept longer than a period of two weeks. Requests for a longer period will require approval from Mary Wiggers.

County of Santa Clara General Services Agency: Data Processing Center	BY F. Watanabe		SECTION 5.0	
	APPROVAL		REVISION 01	PAGE 2
	APPROVAL		DATE 9/3/74	

Advanced Analysis/Mathematical Library
SPSS
Detailed Operational Description

5.3.2 TEST.JAA.PRMUS Purging Policy - GSA DPC will be responsible for the purging of SPSS System Files.

(1) The purging of individual SPSS System Files should be directed to Mary Wiggers.

(2) Periodically, GSA DPC will purge all SPSS System Files residing on TEST.JAA.PRMUS. GSA DPC will create a magnetic tape back-up copy of all the SPSS System Files before purging. SPSS System Files on this back-up tape will be available upon request for a period of four weeks.

Presently, the purging of all SPSS System Files will on the first and third Monday of each month.

5.3.3 TEST.JAA.PRMUS Back-up Capability - GSA DPC has installed several Catalogued Procedures to aid the User in "backing-up" their System Files onto Magnetic Tape. This capability allows the User to save his SPSS System File beyond the two week period. The Users are encouraged to use this capabilty for SPSS System File of significant size. The Procedures are documented in Section 5.5.

APPENDIX B

SPSS

REFERENCE TO EXHIBITS

Exhibit 8	Crosstabulation	San Jose Burglary Study
Exhibit 13	One-way frequency histogram	Mt View CFS Study
Exhibit 14	Aggregation of Variables	" " " "

1.0 General Program Description

1.1 Background

The Census Use Study, a small-area research study sponsered by the Bureau of the Census, was established in New Haven, Conn., in September 1966. It was established to explore the current uses and future needs of small-area data and data handling and display techniques in local, State and Federal agencies.

The Grid-Related Information Display System (GRIDS) was designed and written by the Census Use Study staff to provide a flexible, easy-to-use computer mapping system.

1.2 General function

GRIDS can map files whose data characteristics are unknown to provide a quick view of the data for analysis and can also map files where complicated data manipulation is required.

GRIDS produces three types of maps: (1) shaded mapes where the printed symbols vary with the value level, (2) density maps where the number of printed symbols varies with the value level (e.g. one symbol for every 10,000 units), and (3) value maps where the mapping values themselves are printed.

The input to GRIDS consists of one or more data values to be mapped and a pair of coordinates associated with each set of data values. The coordinates determine the location of the data values on the maps and must be supplied to GRIDS by an external routine (i.e. manually, UNIMATCH or ADMATCH).

GRIDS reads the data file to be mapped and manipulates the data values or coordinates if desired. A special programming language may be executed by GRIDS to perform most data processing tasks. This language is very simple and the user need not be a programmer to operate the system.

There is no limit to the number of data records in the file, and a grid cell (from 1 to 55X55 characters) may cover almost any mapping area desired. Up to five(5) completely independent maps may be produced for each run, and the system will only need to read the data file once.

A complete diagnostic editing system is included in the program logic. The program will scan all control cards for errors and does not stop after the first error is encountered.

2.0 Program Reliability/Maintainability

2.1 General Policy

The Census Use Study Group does not assume the responsibility or updataing the GRIDS programs. The local chapter(San Francisco) has assured us that they will aid Users in the successful implementation of the programs as they have done in the past. However, there is no binding agreement between the User and the local chapter of the Census Use Study Group which obligates them to provide this service. The Census Use Study Group does make available program documentation material to aid in the use and maintenance of the programs.

2.2 Program Reliability

The Census Use Study Group makes the GRIDS programs available for use by local, State and Federal agencies. Santa Clara County's Planning Department has used GRIDS successfully for past three years. Other Agencies outside the County has had similar success with GRIDS.

2.3 Program Maintainability

GSA DPC has the high-level language version(Fortran) of the GRIDS programs. Thus it is possible for GSA DPC to maintain the programs. GSA DPC will maintain GRIDS on its Computer Supported Mathematical Library.

2.4 Problem Handling

Should problems occur during the execution of GRIDS, the problem description and associated run documentation should be submitted to GSA DPC for resolution.

County of Santa Clara General Services Agency: Data Processing Center	BY F. Watanabe		SECTION 3.0	
	APPROVAL		REVISION	PAGE 1
	APPROVAL		DATE July 30, 1974	

Advanced Analysis/Mathematical Library
GRIDS
General Operational Description

3.0 General Operational Description
GRIDS, as installed at GSA DPC, is normally executed in two phases.

3.1 Phase I
During the execution of this phase, GRIDS will produce a magnetic tape copy of the resultant maps, the maps themselves printed on stack paper and the routine GRIDS control card listings and diagnostics. The listings, diagnostics, hard-copy maps and the job set-up will be returned to the user. The job set-up will be annotated with the number of the tape containing the copy of the maps. The user has ten days beginning upon completion of Phase I in which to inspect the results of the run and advise GSA DPC to initiate Phase II. If, at the end of ten days, GSA DPC has not been so informed, the tape containing the maps will be released for use by other jobs.

3.2 Phase II
The execution of this phase does not actually use GRIDS, but rather, causes the magnetic tape copy of the maps produced in Phase I to be printed on unlined vellum paper(U012). The maps thus produced can be combined with a mylar overlay of the analysis area showing streets, blocks, zones or areas and this resulting sandwich can be used to make inexpensive "blue-line" prints or higher quality photographic negatives and prints.

The procurement of the vellum paper, mylar overlays and finished prints are the user's responsibility. However, technical support may be obtained from the Center for Urban Analysis. They may contacted at (408) 299-3285.

Advanced Analysis/Mathematical Library
GRIDS
PHASE I - Detailed Operational Description

4.1 Phase-I - Detailed Operational Description

4.1.1 Input Requirements

A. Grids control cards - These controls are used by GRIDS to determine the input format of the incoming data, to specify any special processing required, and to specify the type of map desired.

B. Grids Data File - This file contains the value and location data. It is normally the output of an external geo-coding routine, UNIMATCH, but can be generated manually. The medium for this file can be punched cards, magnetic tape or disk. If the medium is other than punched cards, a control card specifying the location of the Data File must be included.. See DATAUNIT keyword parameter for control cards. It is suggested that the DATAUNIT value be set to 10 when using this keyword parameter.

4.1.2 Outputs Generated

A. GRIDS Map Pool Tape - This tape contains the necessary data to print the required maps. It is automatically pooled to magnetic tape so that maps can be reprinted without re-executing Phase I. The user can reprint his maps (Phase II) for a duration of 10 days. After that time period the pool tape will not be available.

B. GRIDS Maps - Phase I will automatically print a copy of the required map on standard computer output forms. i.e. 11 x 14 one-part paper six liner per inch.

County of Santa Clara
General Services Agency, Data Processing Center

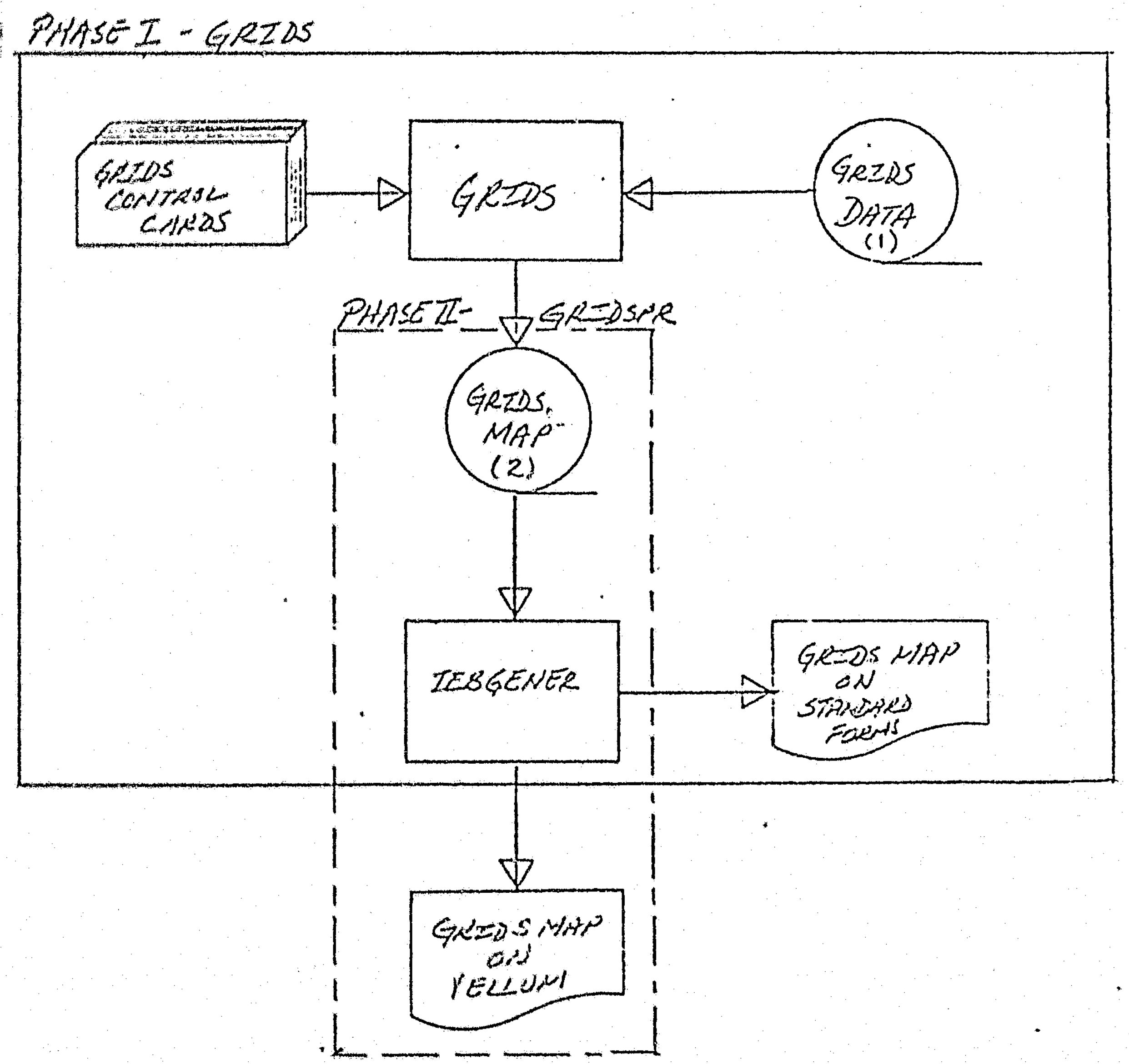
F. Watanabe

1

July 30, 1974

Advanced Analysis/Mathematical Library
GRIDS

SYSTEM FLOW CHART



(1) GRIDS DATA - CAN BE CARDS, TAPES, DISK ETC
(2) GRIDS MAP POOL TAPE - AUTOMATIC RETENTION PERIOD OF 10 DAYS

APPENDIX C

GRIDS

Reference to Exhibits Using GRIDS Functions

Exhibit 7	Double Value Map	San Jose Burglary Study
Exhibit 9	Single Value Map	Sheriff's Patrol Analysis Study
Exhibit 10	Shaded Map	" " " "
Exhibit 11	Single Value Map	" " " "
Exhibit 26	" " "	San Jose Land Use Study
Exhibit 27	" " "	" " " " "
Exhibit 28	" " "	Campbell Burglary Study
Exhibit 29	" " "	" " "
Exhibit 30	Shaded Map	CJPP Robbery Study

APPENDIX D

GEO-CODING

The process by which an event whose location is known by address or place name is referenced to service area (e.g. school district, census tract, police beat, etc.) is known as geo-coding. The Center performs geo-coding by means of its DIME file and the program UNIMATCH. Figure 5 depicts the problem of overlapping service areas; and each of the elements in the geo-coding process are described below.

A. DIME File (Dual Independent Map Encoding):

The Census Bureau's DIME File concept, which is required in the creation of event files, is a geographic base file which describes an area in terms of line segments, nodes and enclosed areas (blocks). Each street, river, canal, railroad track, municipal boundary, or other map feature can be considered as one or more straight line segments. Curved lines can be divided into a series of straight line segments. When features intersect or when straight line segments change direction, nodes are formed. The area enclosed by a set of line segments are called blocks. Using street names and address ranges, the state plane coordinate system and other identifier; each segment, node, and block is uniquely identified along with its geographic characteristics (Figures 6,7, and 8). The DIME file is built and maintained using programs provided by the Census Bureau and programs developed and maintained by the Center for Urban Analysis.

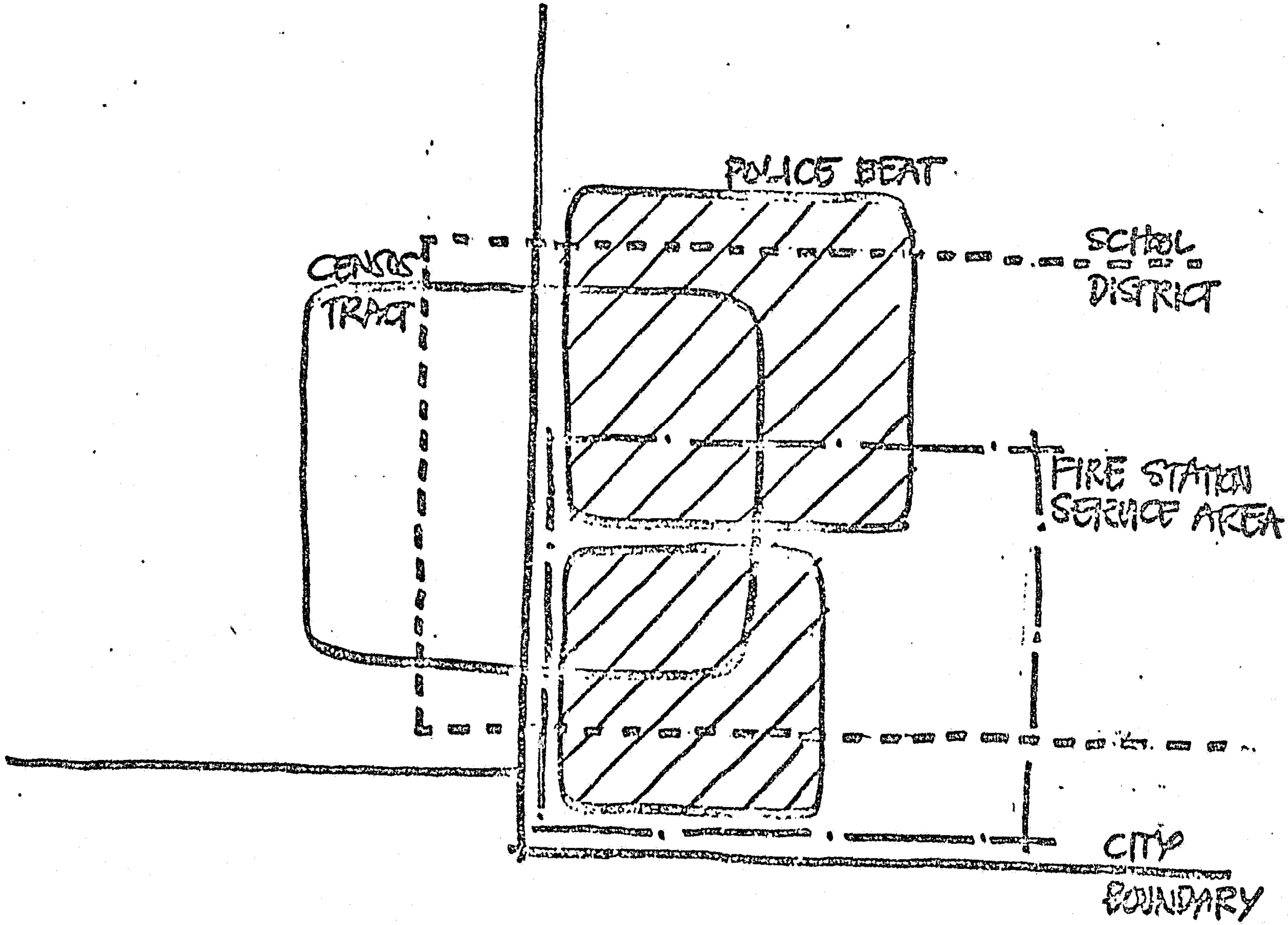
B. UNIMATCH:

UNIMATCH is a generalized record linkage system. The linking process involves two input files. The first, the data file, is the file to which information is to be attached. The second file, the reference file (usually extracted from the DIME File), is the file that supplies the information to be linked. A data file record is said to be matched if a unique reference file record exists that satisfies the criteria of the matching application.

Geo-coding then is a specialized form of record linkage application whereby house numbers, directions, street names and types, and city jurisdiction codes are the matching criteria and x-y coordinates, census tracts and user service area codes are the transferred or linked data.

As was GRIDS, UNIMATCH was designed and written by the Census Use Study staff to provide a flexible, easy to use computerized record linkage system.

Figure 5



OVERLAPPING SERVICE AREAS

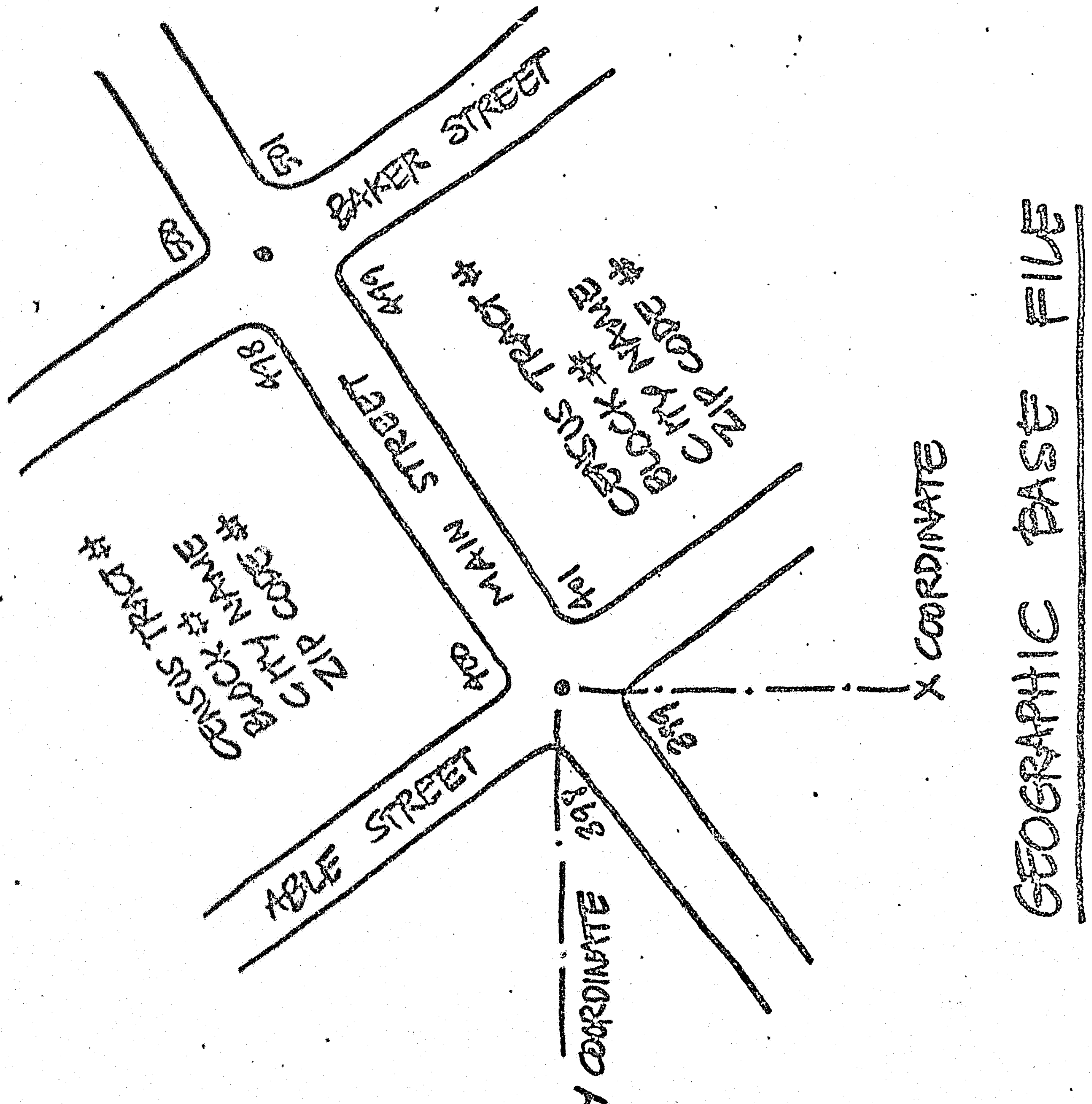


Figure 6

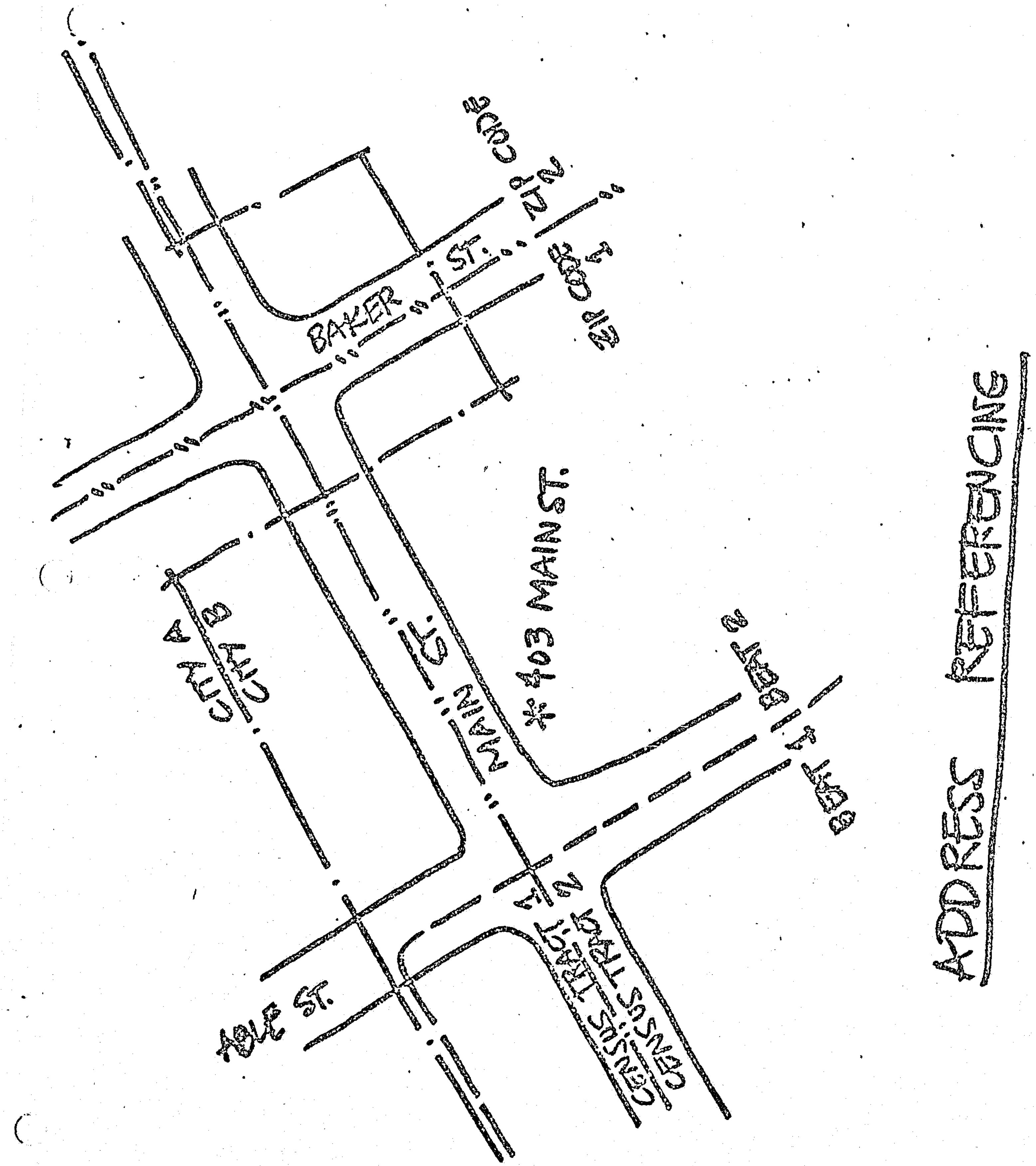


Figure 7

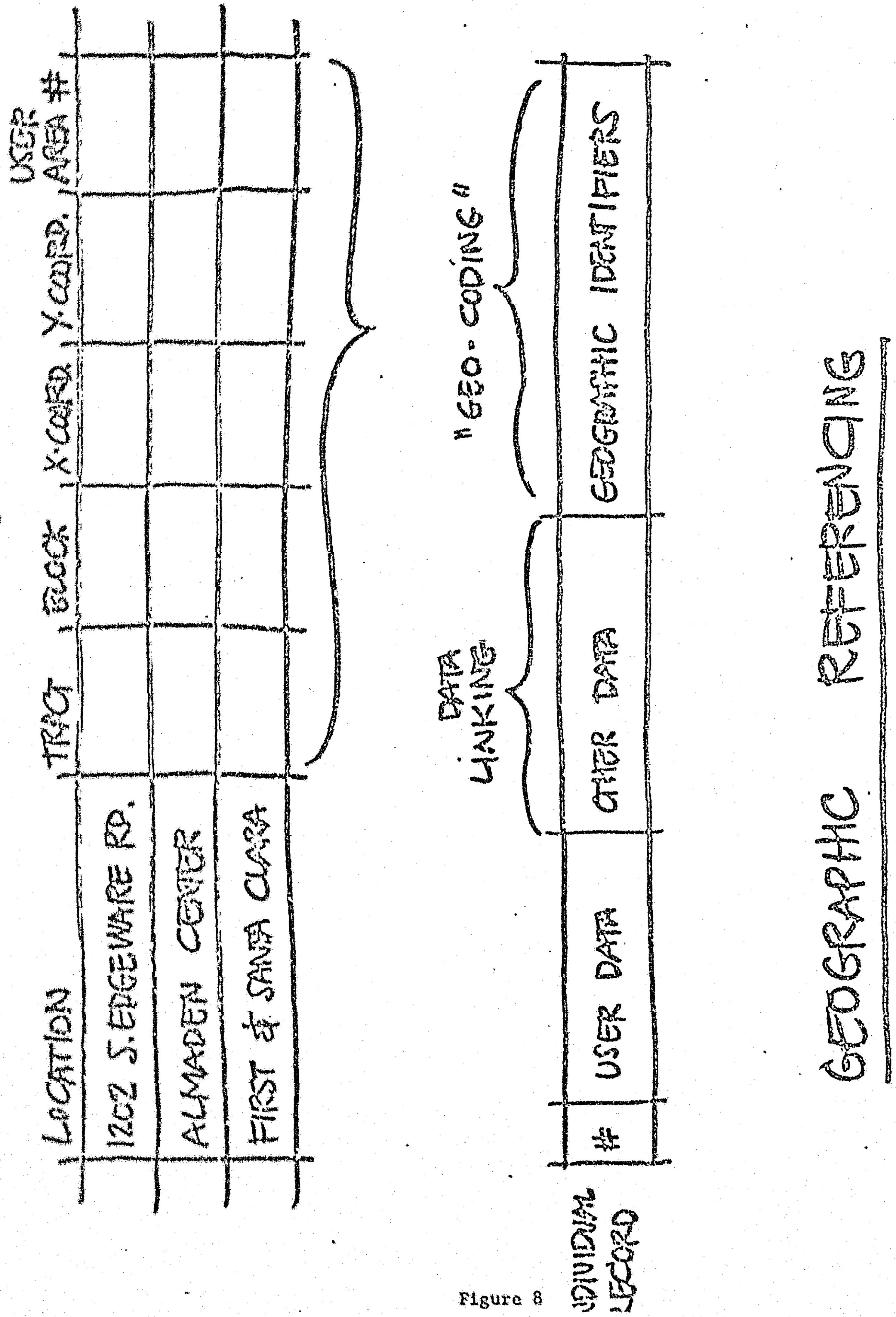
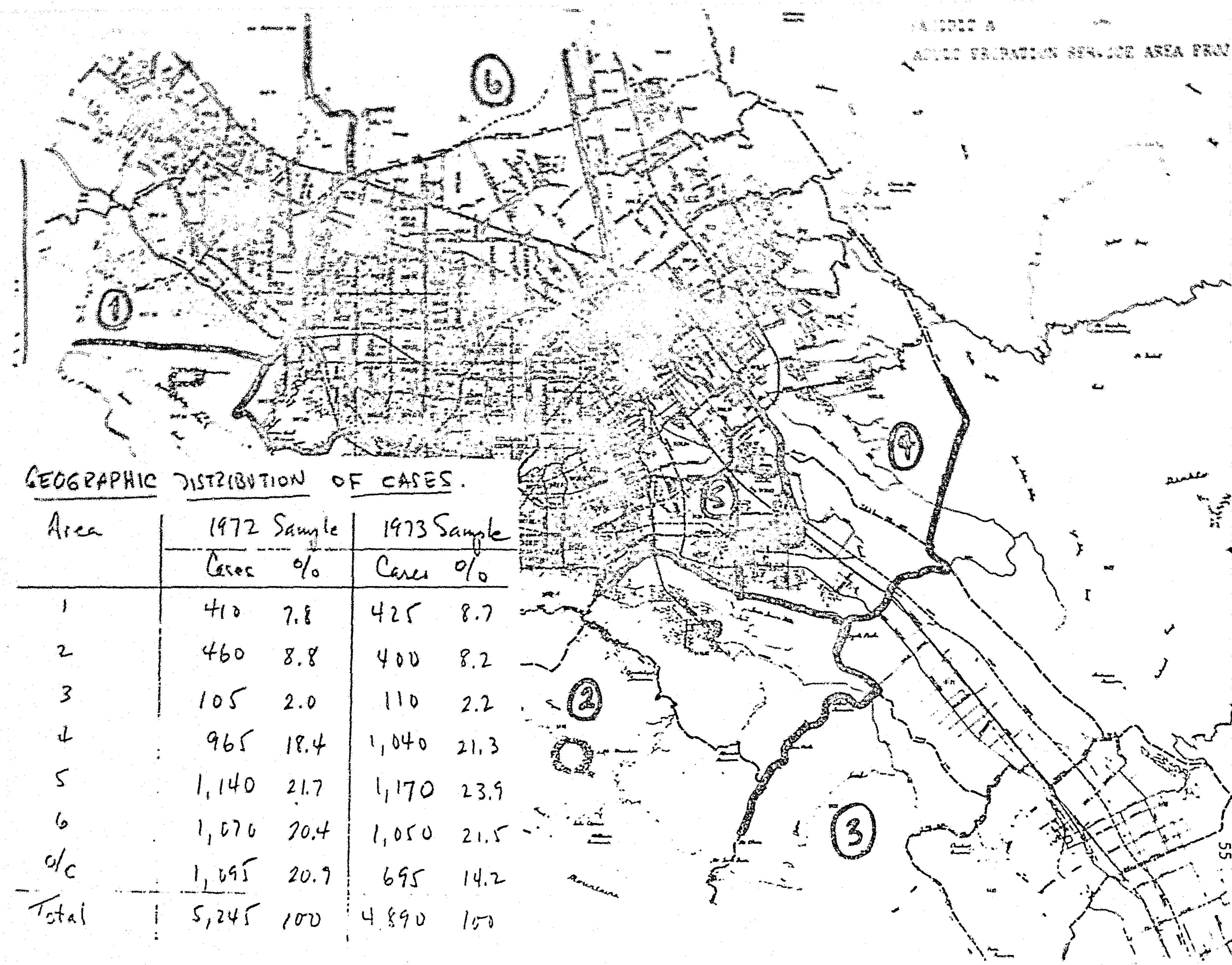


Figure 8

APPENDIX E

EXHIBITS



GEOGRAPHIC DISTRIBUTION OF CASES.

Area	1972 Sample		1973 Sample	
	Cases	%	Cases	%
1	410	7.8	425	8.7
2	460	8.8	400	8.2
3	105	2.0	110	2.2
4	965	18.4	1,040	21.3
5	1,140	21.7	1,170	23.9
6	1,070	20.4	1,050	21.5
O/C	1,095	20.9	695	14.2
Total	5,245	100	4,890	100

EXHIBIT B

ADULT PROBATION SERVICE AREA PROJECT

PROBATION CASE ADDRESS BY SERVICE AREA / 1972 Sample 6 Areas

NUMBER OF CASES BY SEX

AREA	TOTAL CASES LAST ADDRESS	CHANGE OF ADDRESS IN	OUT	WITHIN	MALES LAST ADDRESS	CHANGE OF ADDRESS IN	OUT	WITHIN	FEMALES LAST ADDRESS	CHANGE OF ADDRESS IN	OUT	WITHIN
1	49	24	30	12	33	12	17	7	16	12	13	5
2	54	25	30	12	38	14	19	8	16	11	11	4
3	12	5	4	4	9	2	2	3	3	3	2	1
4	114	54	42	31	79	36	24	28	35	18	18	3
5	131	55	63	65	97	38	40	45	34	17	23	20
6	130	57	59	37	84	31	35	27	46	26	24	10
O/C	139	76	68	58	80	43	39	24	59	33	29	34
TOTAL	629	296	296	219	420	176	176	142	209	120	120	77

NUMBER OF CASES BY TYPE OF OFFENSE (MOST SERIOUS CONVICTION)

AREA	FELONY LAST ADDRESS	CHANGE OF ADDRESS IN	OUT	WITHIN	MISD/DRUNK DRIVING LAST ADDRESS	CHANGE OF ADDRESS IN	OUT	WITHIN	MISD/DRUGS LAST ADDRESS	CHANGE OF ADDRESS IN	OUT	WITHIN	MISD/OTHER LAST ADDRESS	CHANGE OF ADDRESS IN	OUT	WITHIN
1	24	16	15	7	2	2	4	0	4	3	0	2	19	3	11	3
2	25	20	22	8	2	1	1	0	3	0	0	0	24	4	7	4
3	4	3	1	0	1	1	1	1	1	1	1	0	6	0	1	3
4	65	33	28	19	12	4	1	2	0	0	0	0	37	17	13	10
5	72	34	38	46	11	3	3	4	3	0	0	2	45	18	22	13
6	77	40	38	20	9	2	5	1	4	1	3	1	40	14	13	15
O/C	73	42	46	37	8	4	2	1	2	2	3	3	56	28	17	17
TOTAL	340	188	188	137	45	17	17	9	17	7	7	8	227	84	84	65

Moves by Probationers
1972 Sample

	Moves per Case					
	0	1	2	3	4	Total
Number of Cases	330	160	83	42	14	629
Number of Moves *						
-within zones		55	62	50	19	136
-between zones		40	34	35	15	124
-out of county		65	70	41	22	198
Total		160	166	126	56	508
Case Involved in each type of move *						
-within zone		55	45	27	9	136
-between zones		40	26	22	9	97
-out of county		65	44	20	8	137

* Computed for case of five service areas

Inter zone Transfers per year $= \frac{1}{2}\left(\frac{97}{629}\right) = 7.7\%$

Exhibit 1

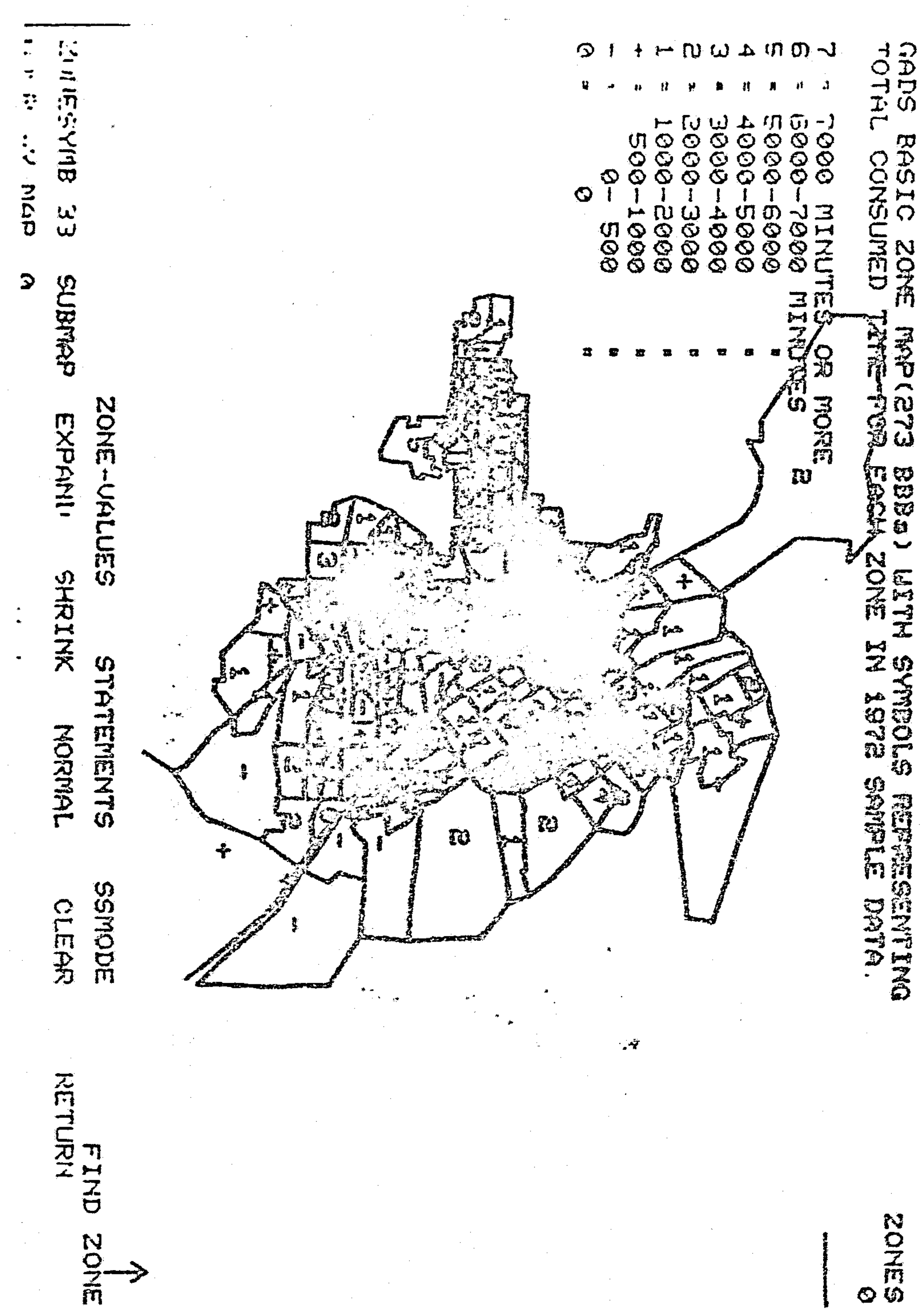


Exhibit 2

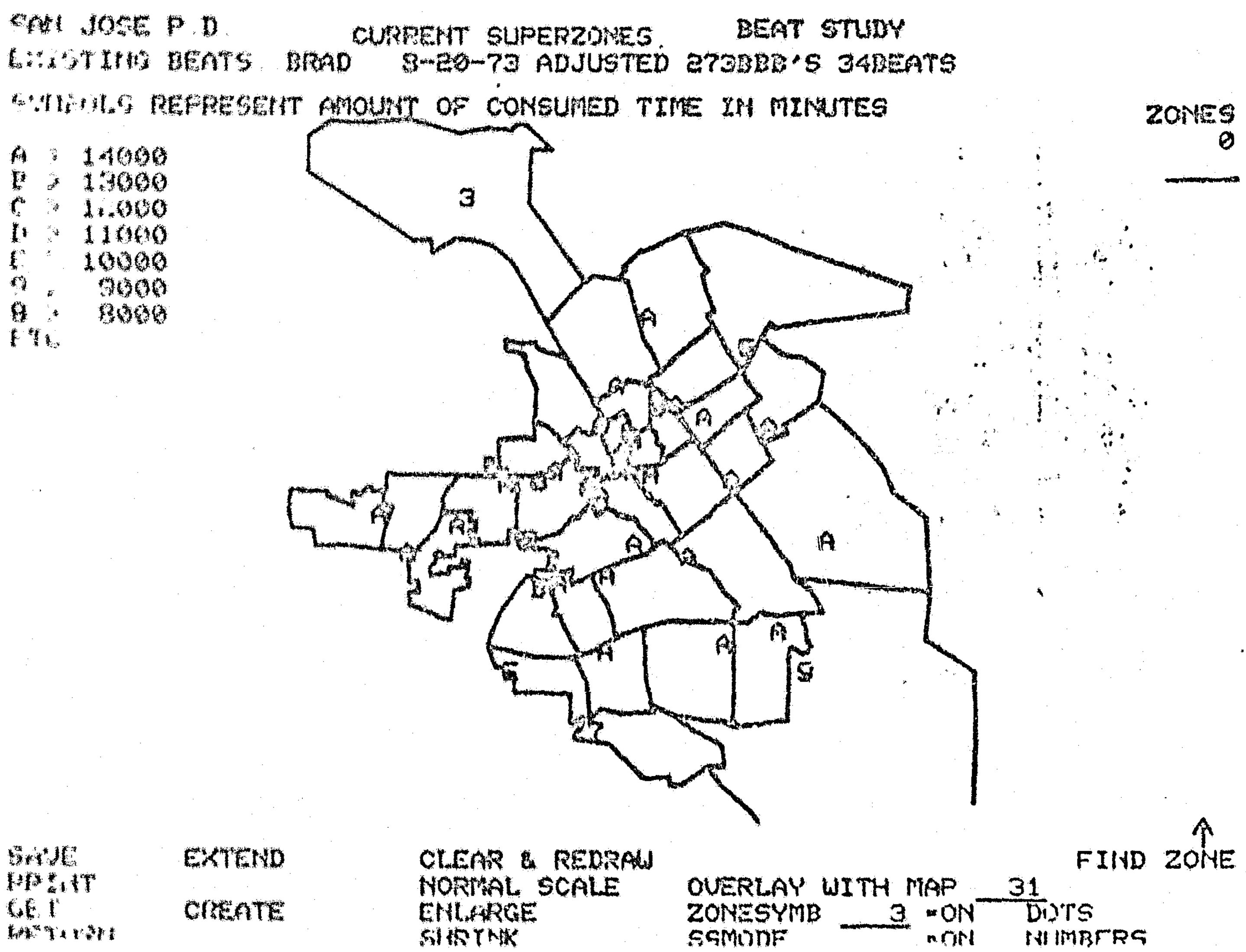


Exhibit 3

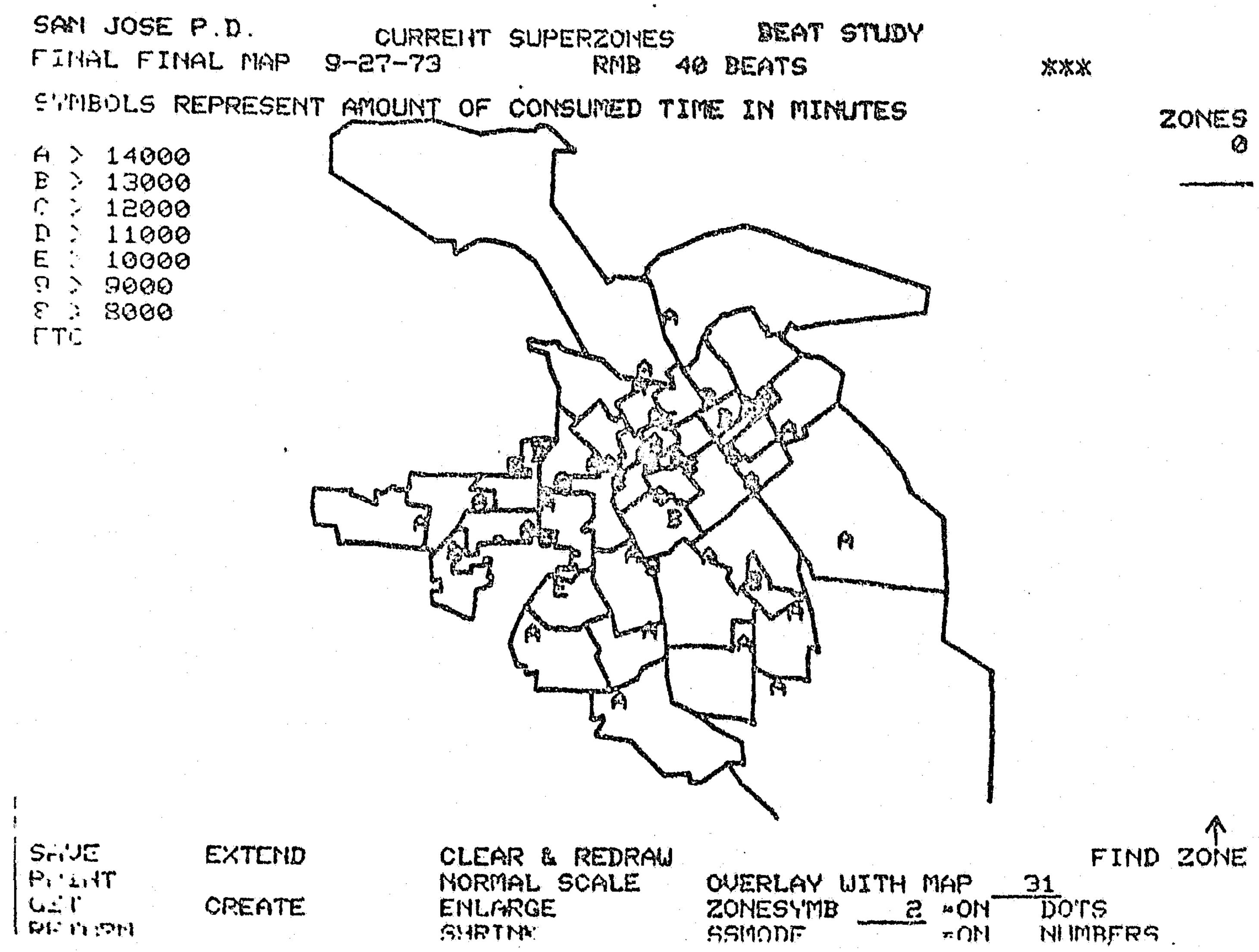
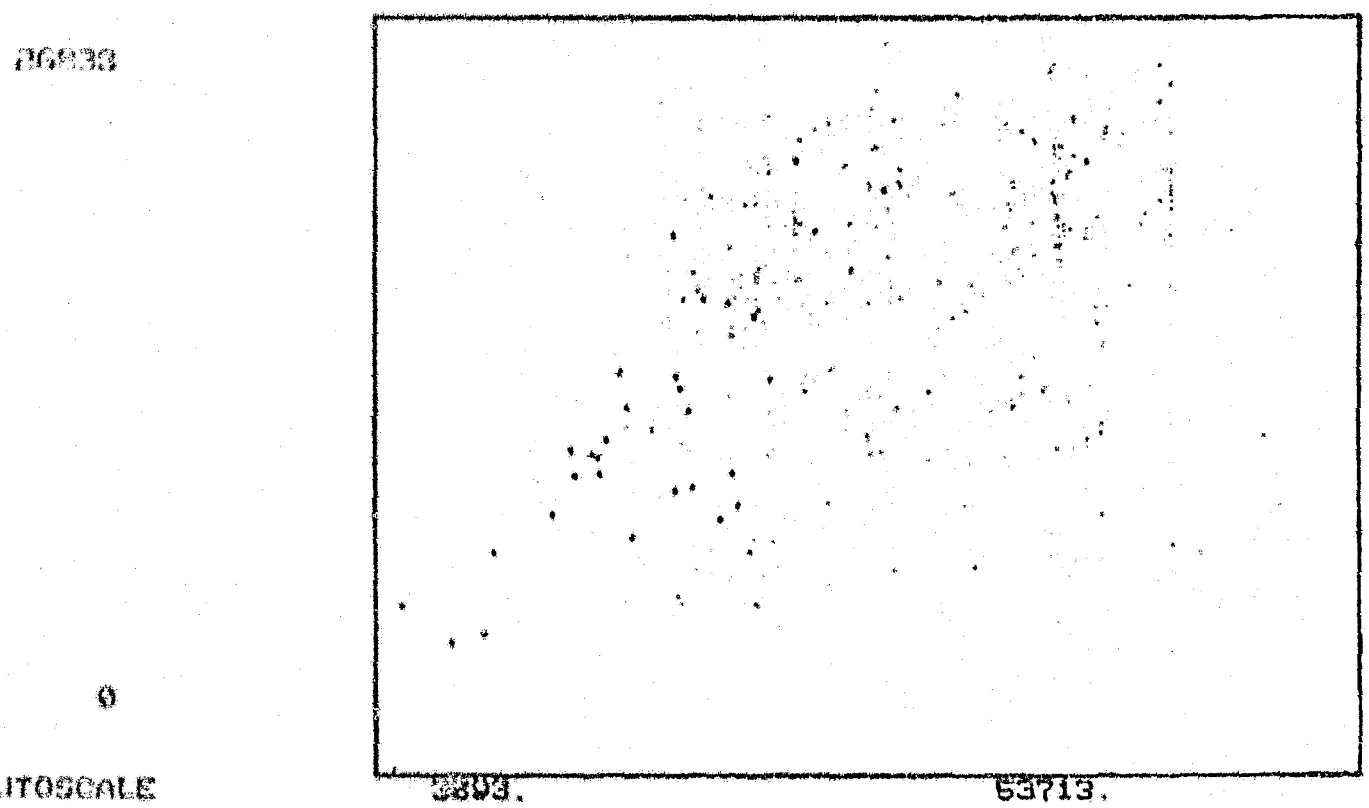


Exhibit 4

SAN JOSE POLICE DEPARTMENT BEAT STUDY
EACH DOT REPRESENTS THE VALUE OF TOTAL CONSUMED TIME IN MINUTES FOR A BEAT
USING THE EXISTING BEAT STRUCTURE BEFORE THE STUDY.
THE X AXIS IS 1974 DATA AND THE Y AXIS IS 1972 DATA.



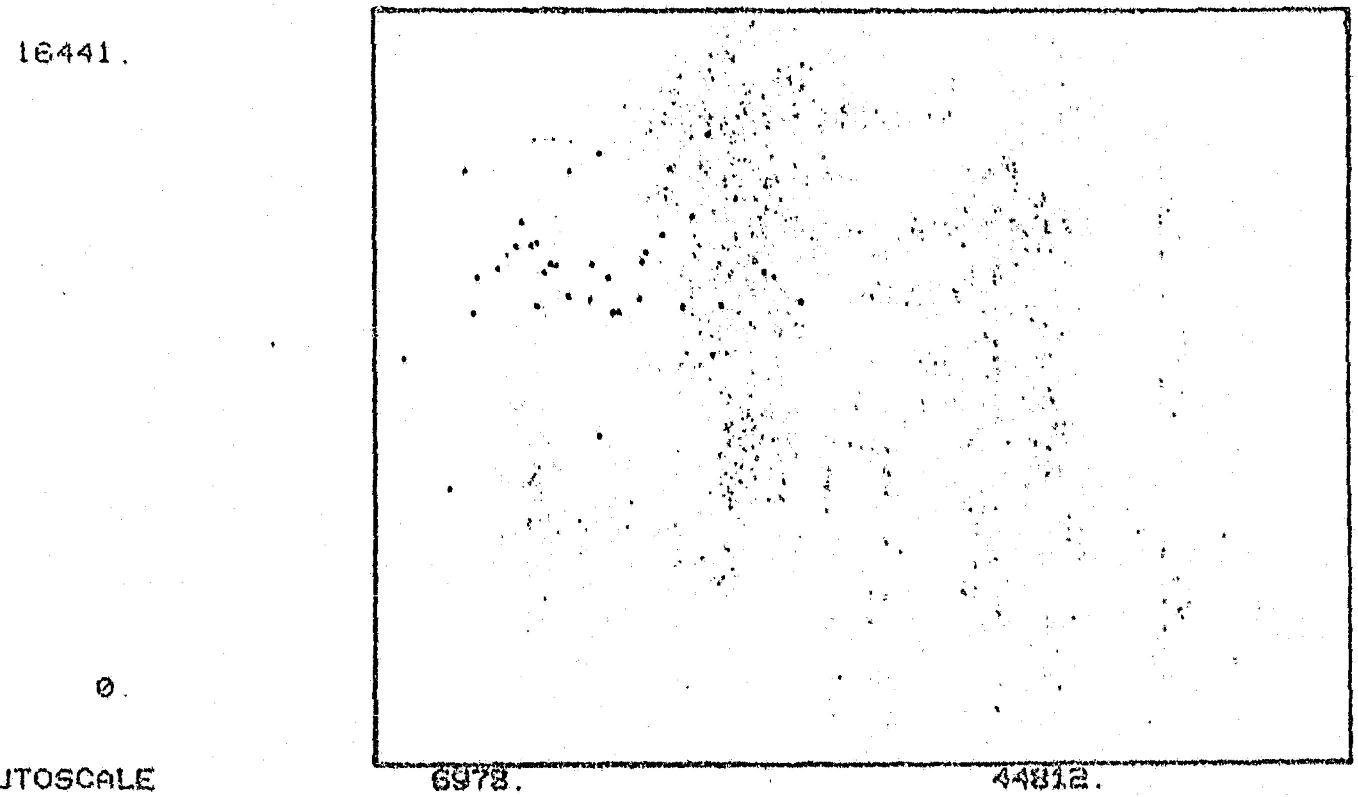
AUTOSCALE

REDRAW
RETURN

VARNAME	UNIT	USEMAP		
			ZONE	0
TOTCTIM	91	15		
TOTCTIM	61	15	ZONESYMB	1

Exhibit 5

SAN JOSE POLICE DEPARTMENT BEAT STUDY
EACH DOT REPRESENTS THE VALUE OF TOTAL CONSUMED TIME IN MINUTES FOR A BEAT
THE X AXIS IS 1974 AND THE Y AXIS IS 1972.
(DATA SHOWS A MUCH MORE HOMOGEBIOUS GROUPING THAN BEFORE STUDY).

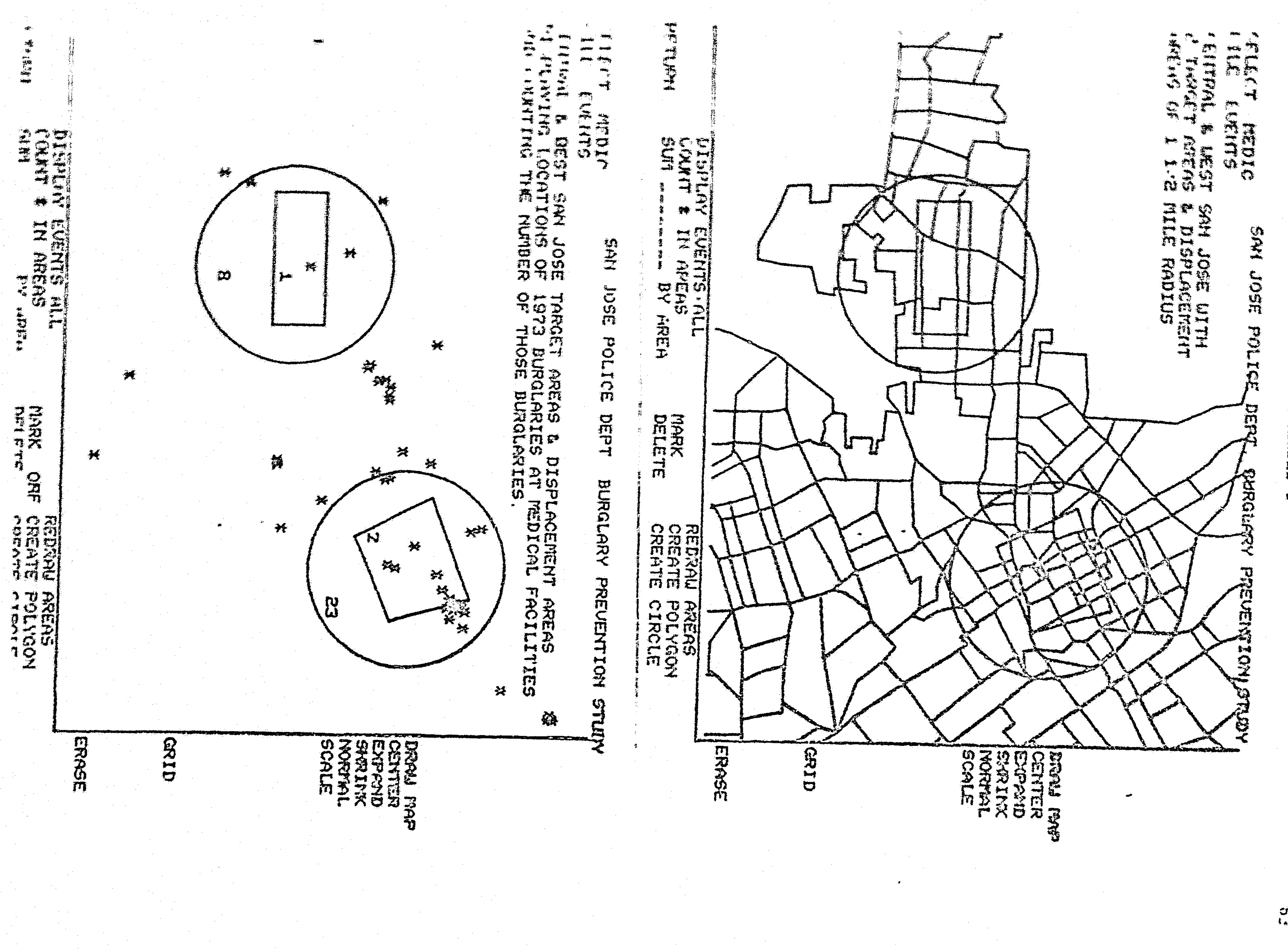


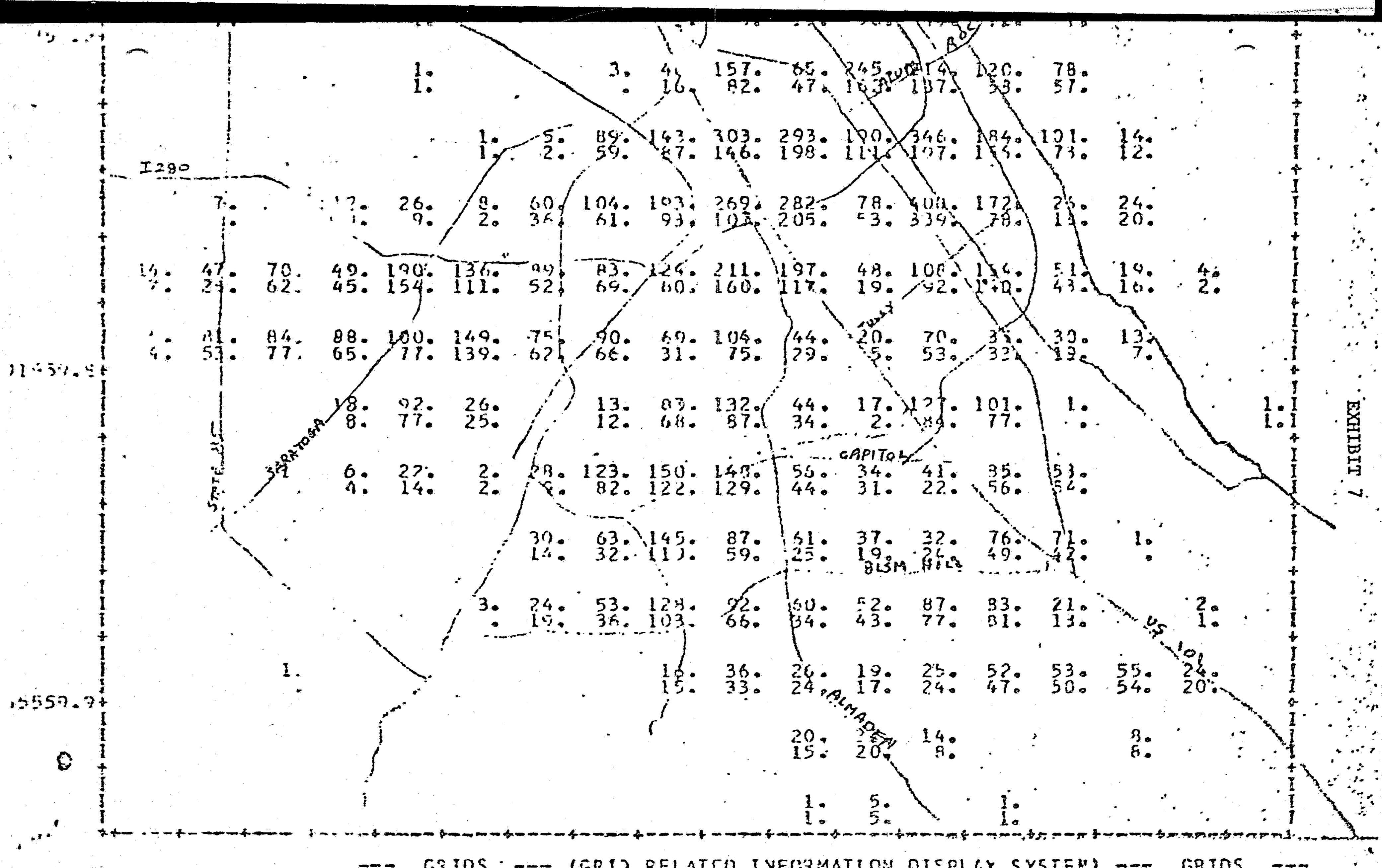
AUTOSCALE

REDRAW

RETURN

VARNAME	UNIT	USEMAP	ZONE	0
TOTCTIM	91	31		
TOTCTIM	61	31	ZONESYMB	1.





TOTAL BURGLARIES & PRIVATE RESID. BURGS, SAN JOSE - 1973

SJPD BURGLARY PREVENTION GRANT STUDY
DATA SOURCE - 1973 C.A.P.E.R.
MAP PREPARED BY THE CENTER FOR URBAN ANALYSIS, SANTA CLARA COUNTY
DATE PREPARED: FEB. 21, 1975

FIGURES REPRESENT TOTAL CASES ON TOP LINE AND
CASES WITH PREMISE CODE OF 10,11,12,13,14 OR 15 ON BOTTOM LINE.

MAP SCALE IS 1 INCH = 2 MILES.
EACH GRID CELL IS 1 MILE SQUARE

INESS CATEGORIES BY CENSUS TRACT - AS OF MARCH 1, 197   06/19/74   PAGE 3

FILE   (CREATION DATE = 06/19/74)

* * * * * * * * * * * * * * * * * C R O S S T A B U L A T I O N   O F * * * * * * * * * * * * * * * * * *
CT   CENSUS TRACT   BY CODE   BUSINESS CATEGORY
* PAGE 1 OF 45

CT	COUNT / ROW PCT / COL PCT / TOT PCT	CODE: REP - INDUST. 1.	MEDICAL 2.	GAS STA. 3.	DEPART. STORE 4.	GROCERY STORE 5.	MANUFAC-TURING 6.	PERSONAL SERVICES 7.	RETAIL 8.	WARE-HOUSE 9.	LIQUOR STORE 10.	ROW TOTAL
500100.	COUNT	25	0	6	0	3	90	80	35	20	1	271
	ROW PCT	9.2	0.0	2.2	0.0	1.1	33.2	29.5	12.9	7.4	0.4	2.2
	COL PCT	4.0	0.0	1.9	0.0	1.4	5.5	1.9	1.1	20.8	1.0	
	TOT PCT	0.2	0.0	0.0	0.0	0.0	0.7	0.6	0.3	0.2	0.0	
500200.	COUNT	14	42	6	1	5	28	316	24	8	1	438
	ROW PCT	3.2	9.7	1.4	0.2	1.1	6.4	72.1	5.5	1.8	0.2	3.5
	COL PCT	2.2	5.4	1.9	1.5	2.3	1.7	7.4	0.8	8.3	1.0	
	TOT PCT	0.1	0.1	0.0	0.0	0.0	0.2	2.5	0.2	0.1	0.0	
500300.	COUNT	6	2	8	0	4	114	135	47	5	1	346
	ROW PCT	1.7	0.6	2.3	0.0	1.2	32.9	39.0	13.6	1.4	0.3	2.8
	COL PCT	1.0	0.2	2.5	0.0	1.9	6.9	3.1	1.5	5.2	1.0	
	TOT PCT	0.0	0.0	0.1	0.0	0.0	0.9	1.1	0.4	0.0	0.0	
500400.	COUNT	6	14	0	0	3	9	113	14	3	0	171
	ROW PCT	3.5	8.2	0.0	0.0	1.8	5.3	66.1	8.2	1.8	0.0	1.4
	COL PCT	1.0	1.7	0.0	0.0	1.4	0.5	2.6	0.4	3.1	0.0	
	TOT PCT	0.0	0.1	0.0	0.0	0.0	0.1	0.9	0.1	0.0	0.0	
500500.	COUNT	6	64	5	0	3	8	103	39	0	1	237
	ROW PCT	2.5	27.0	2.1	0.0	1.3	3.4	43.5	16.5	0.0	0.4	1.9
	COL PCT	1.0	7.7	1.6	0.0	1.4	0.5	2.4	1.2	0.0	1.0	
	TOT PCT	0.0	0.5	0.0	0.0	0.0	0.1	0.8	0.3	0.0	0.0	
500600.	COUNT	15	18	5	0	0	18	163	37	1	0	268
	ROW PCT	5.6	6.7	1.9	0.0	0.0	6.7	60.8	13.8	0.4	0.0	2.1
	COL PCT	2.4	2.2	1.6	0.0	0.0	1.1	3.8	1.2	1.0	0.0	
	TOT PCT	0.1	0.1	0.0	0.0	0.0	0.1	1.3	0.3	0.0	0.0	
500700.	COUNT	3	2	1	0	2	28	46	35	0	1	125
	ROW PCT	2.4	1.6	0.8	0.0	1.6	22.4	36.8	28.0	0.0	0.8	1.0
	COL PCT	0.5	0.2	0.3	0.0	0.9	1.7	1.1	1.1	0.0	1.0	
	TOT PCT	0.0	0.0	0.0	0.0	0.0	0.2	0.4	0.3	0.0	0.0	
COLUMN TOTAL		623	833	320	67	214	1650	4287	3150	96	102	12566
		5.0	6.6	2.5	0.5	1.7	13.1	34.1	25.1	0.8	0.8	100.0

(CONTINUED)

Exhibit 8

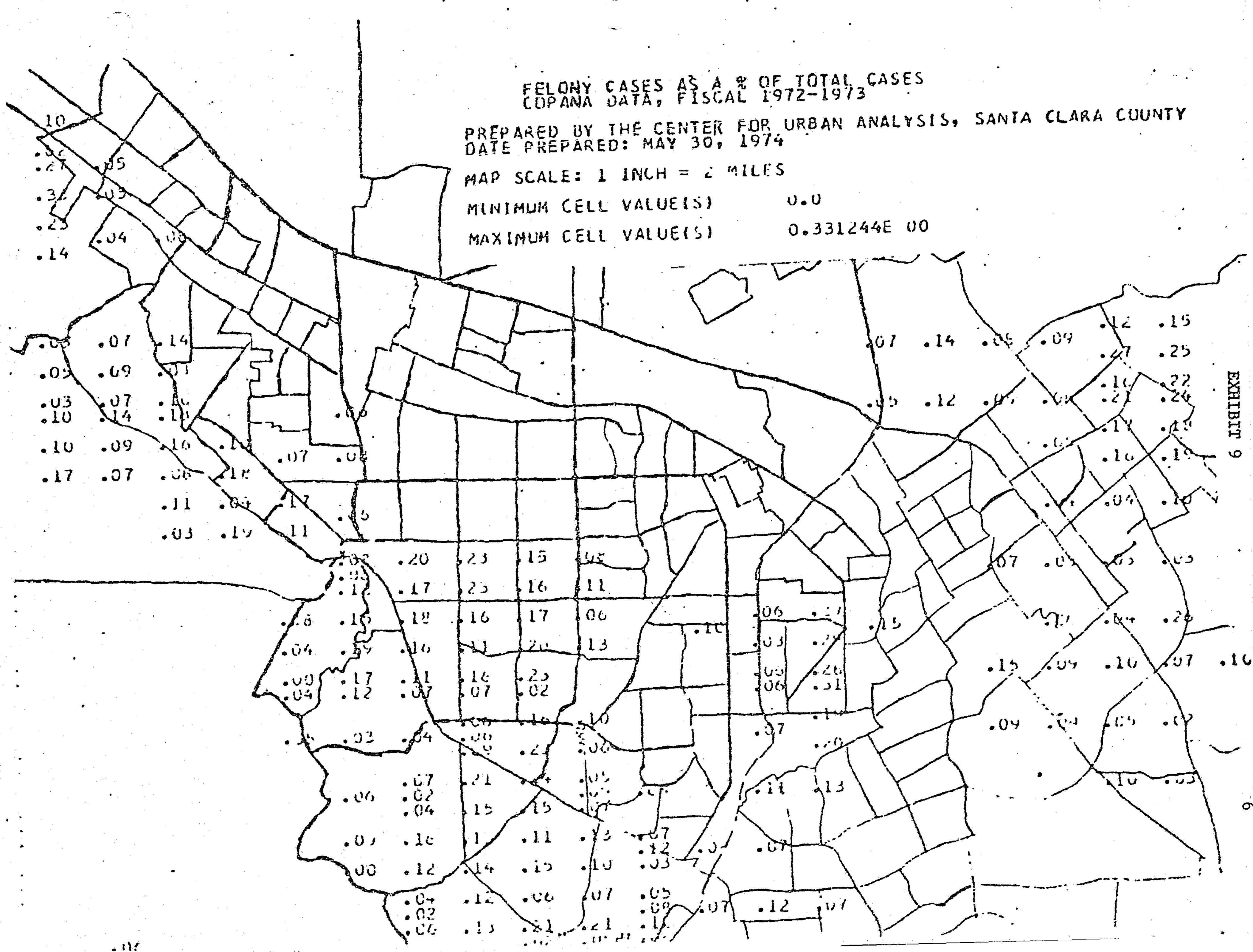


EXHIBIT 9

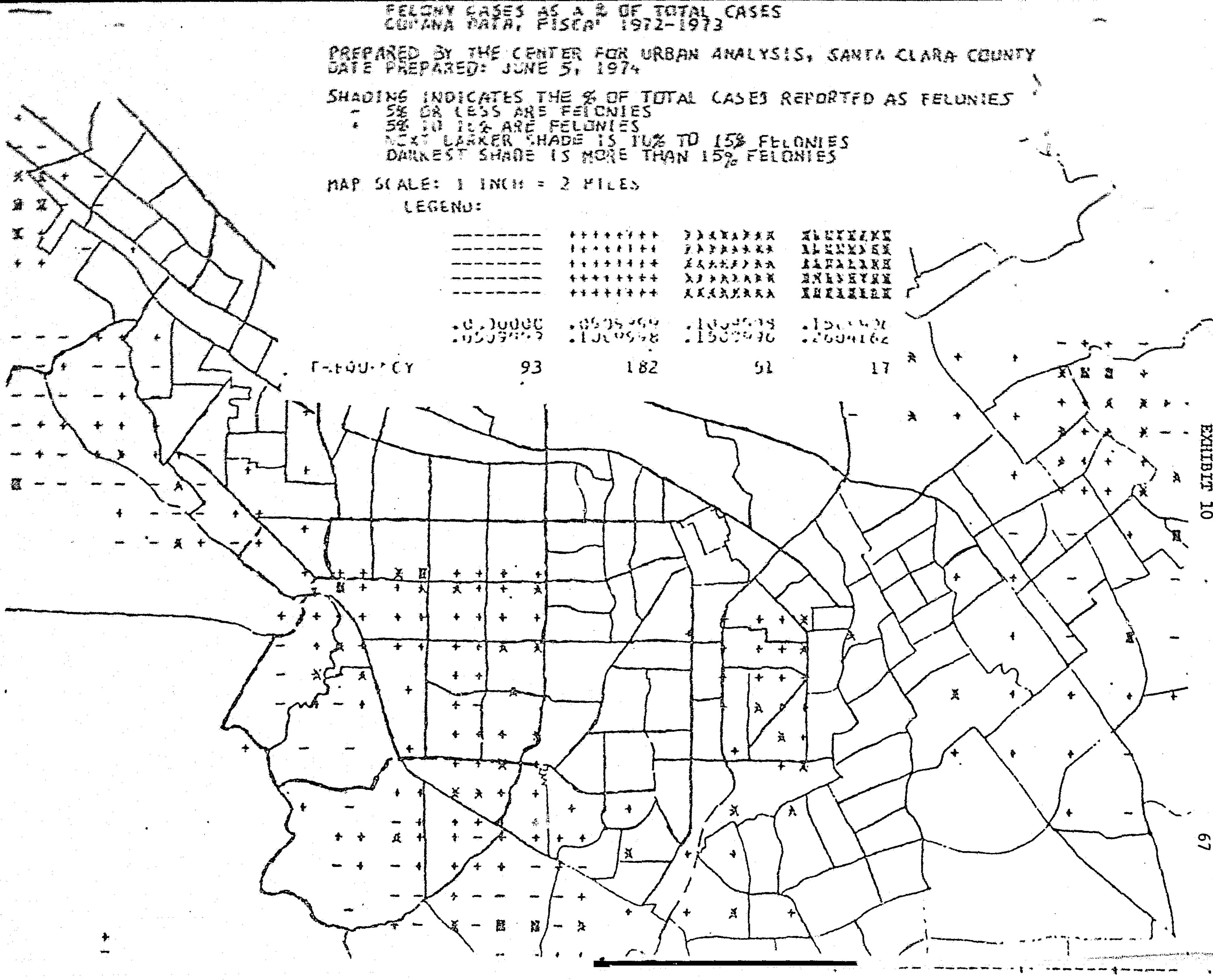


EXHIBIT 10

67

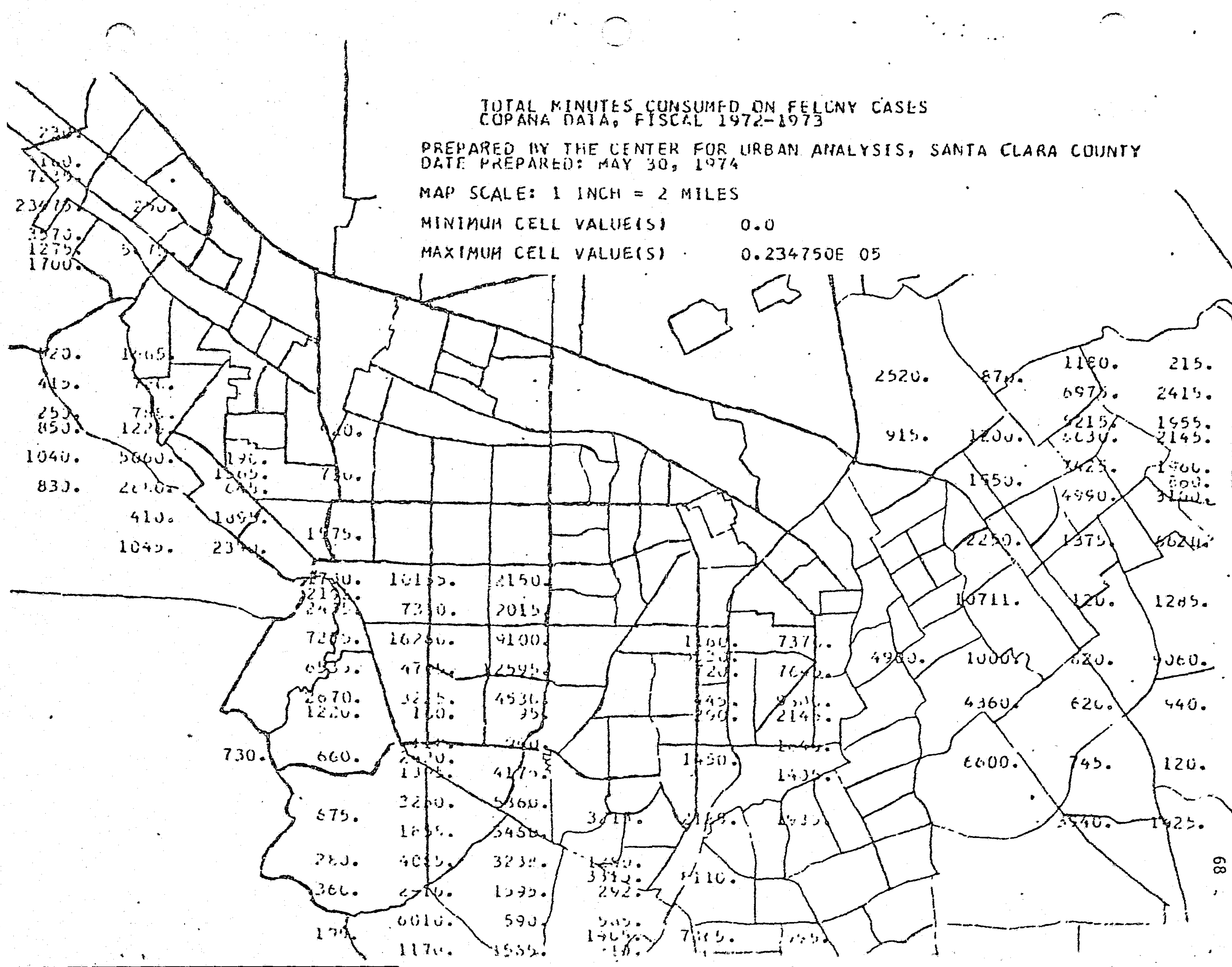


EXHIBIT 11

68

ZHR MIN HSE # STREET NAME		CITY	CORRECTED ADDRESS, X Y
1020 015	WHITE FIR & RED FIR	CU	
1509 015	BURB & MCCLELLEN	CU	
1017 020	MCCLELLAN & STELLING R	CU	R: INCONSISTANT STREET TYPE
1006 015	MCCLELLAN & JOHN	CU	} NON EXISTANT STREET TYPE
1007 015	MCCLELLAN & JOHN	CU	
1020 015	MCCLELLAN & JOHN	CU	
1019 020	BLUEHILLS DR & SR 85	SJ	SR 85: INCONSISTANT STREET NAME
1011 020	N/B SVALE & BOLLINGER	CU	N/B: DIRECTION OF TRAVEL - NOT AN ADDRESS; &: INCONSISTANT INTERSECTION DELIMITER
1019 015 001601	HWY 9	SJ	
1019 015 001601	SARATOGA/SVLE RD	RD SJ	/: INCONSISTANT INTERSECTION DELIMITER
1001 005	85 AND SEA GULL	SA	85: INCONSISTANT STREET NAME
1001 045	85 AND SEA GULL	SA	
1009 020	SR 85 & BLAUER	SA	
1017 055 012946	ARROYO DE ARGUELLO	SA	
1020 060	SR#85 & COX AREA	SA	SR#85: INCONSISTANT STREET NAME; AREA: NOT PRECISE ENOUGH
1009 015	SCULLY NR GOLETA	SA	
1022 040	SAR HI GYM	SA	INCONSISTANT PLACE NAMES
1023 010	OAK ST SCHL	SA	
1023 015	OAK ST SCHL	SA	
1018 165	HWY 85 & BIG BASIN	SA	
1014 010	HWY 280 NR SAR-SVALE	CU	NOT PRECISE ENOUGH

69

51

```
# .. ACCI. TYPE MIN NAME AND ALL                                02/10/75        PAGE   17
FILE    TVIEW   (CREATION DATE = 02/10/75)
VARIABLE   DOW          DAY OF WEEK

     CODE
           I
       1.  ********************* (  1873)   8.2 PCT
           I     MON
           I
           I
       2.  ****************** (  1562)   6.8 PCT
           I     TUE
           I
           I
       3   ********************* (  1530)   6.5 PCT
           I     WED
           I
           I
       4.  ********************** (  2142)   9.4 PCT
           I     THUR
           I
           I
       5.  ************************* (  2526)  11.1 PCT
           I     FRI
           I
           I
       6.  ************************* (  2482)  10.9 PCT
           I     SAT
           I
           I
       7.  ********************* (  2046)   9.0 PCT
           I     SUN
           I
           I
       0.  ********************************************************************************* (  8247)  36.2 PCT
 (MISSING) I     UNKNOWN
           I
           I.........I.........I.........I.........I.........I.........I.........I.........I.........I.........I
           0       1000      2000      3000      4000      5000      6000      7000      8000      9000     10000
           FREQUENCY

STATISTICS..

MEAN         4.203         STD ERROR     0.016         MEDIAN       4.394
MODE         5.000         STD DEV       1.949         VARIANCE     3.798
KURTOSIS    -1.143         SKEWNESS     -0.196         RANGE        6.000
MINIMUM      1.000         MAXIMUM       7.000

VALID   OBSERVATIONS -     14561
MISSING OBSERVATIONS -      8247
```

Exhibit 13

70

Exhibit 14

A G G R E G A T E

GROUPING VARIABLES... EVENT INCIDENT CODE - NUMERIC

GR-SC	GROUP-CODES	N	NS	VARNAME	SUM	MEAN	SD	VALUE	PCTGT	PCTLT
1	0.	960								
			960	RESPCN	121.99	0.13	0.92		0.10	99.79
			960	TOTAL	1310.54	1.37	3.70		2.60	93.65
2	2.	2								
			2	RESPCN	0.40	0.20	0.12		0.0	100.00
			2	TCTAL	2.87	1.43	1.16		C.0	100.00
3	3.	84								
			84	RESPCN	4.07	0.05	0.06		0.0	100.00
			84	TCTAL	191.08	2.27	5.23		4.76	86.90
4	4.	45								
			45	RESPCN	2.93	0.07	0.21		0.0	100.00
			45	TOTAL	59.07	1.31	4.36		4.44	95.56
5	5.	159								

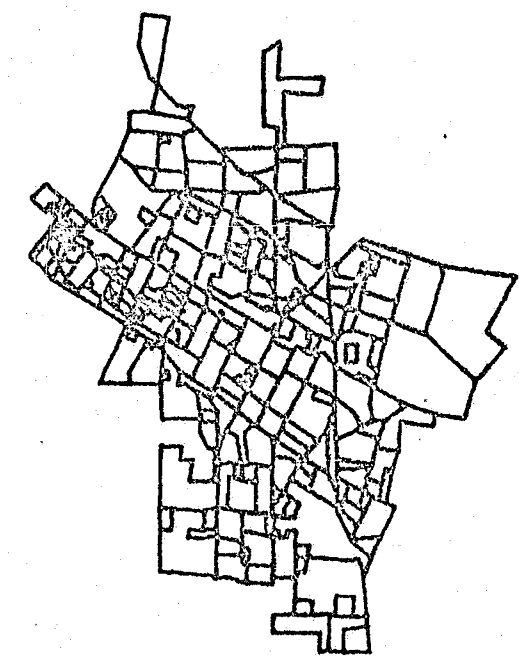
Exhibit 15

MT VIEW P D
CO BASIC ZONES

CURRENT SUPERZONES

CALLS FOR SERVICE STUDY

ZONES



SAVE
PRINT
EXTEND
CREATE
CLEAR & REDRAW
NORMAL SCALE
ENLARGE
SHRINK
OVERLAY WITH MAP 1
ZONESYMB 1
SSMODE
FIND ZONE
DOTS
NUMBERS

Exhibit 16

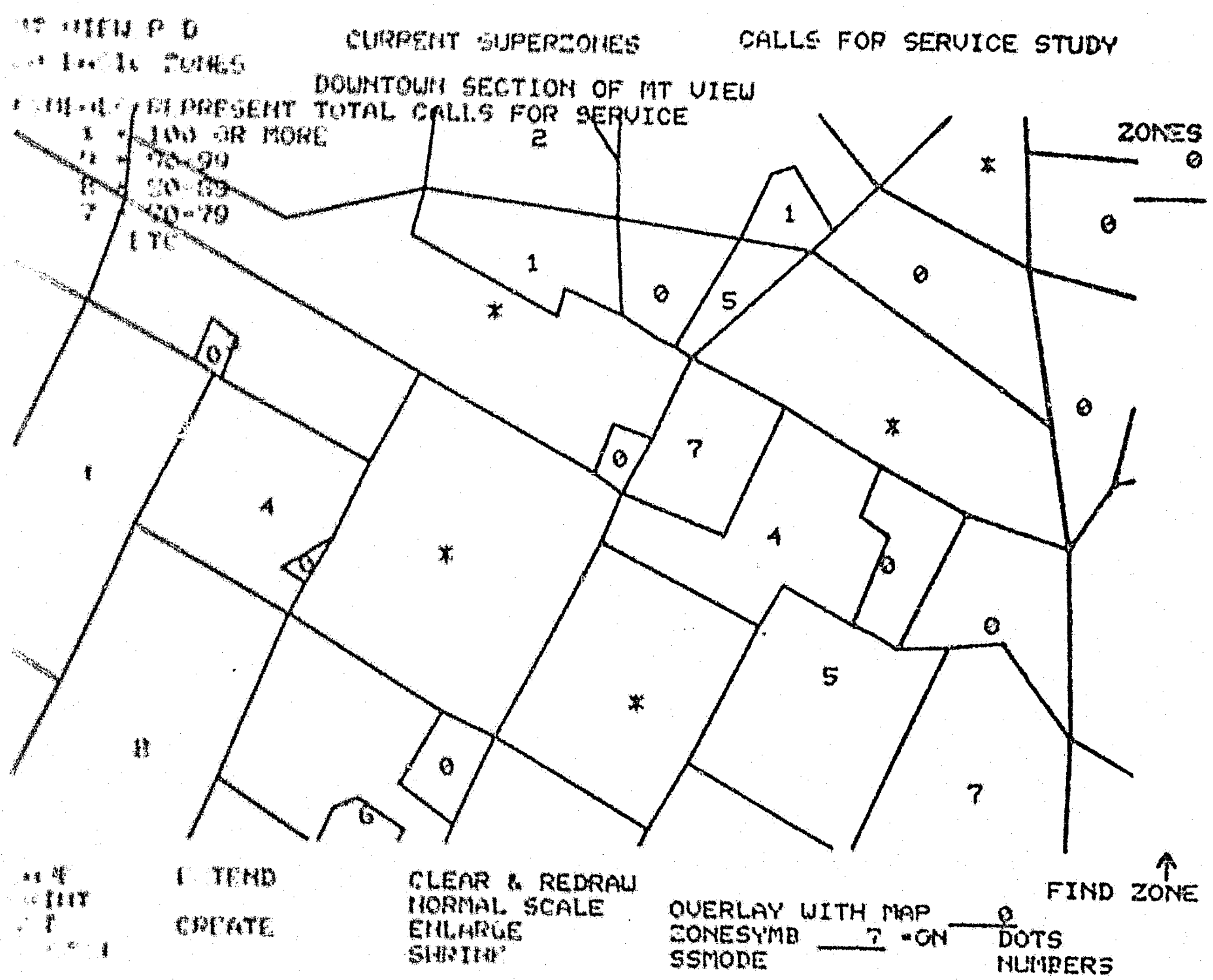


Exhibit 17

Mt View Police Department Calls for Service Study

Each stick represents the values for a "BBB" for Total Calls (on the x-axis), Total Consumed Time (on the y-axis) and Average Response Time (on the z-axis).

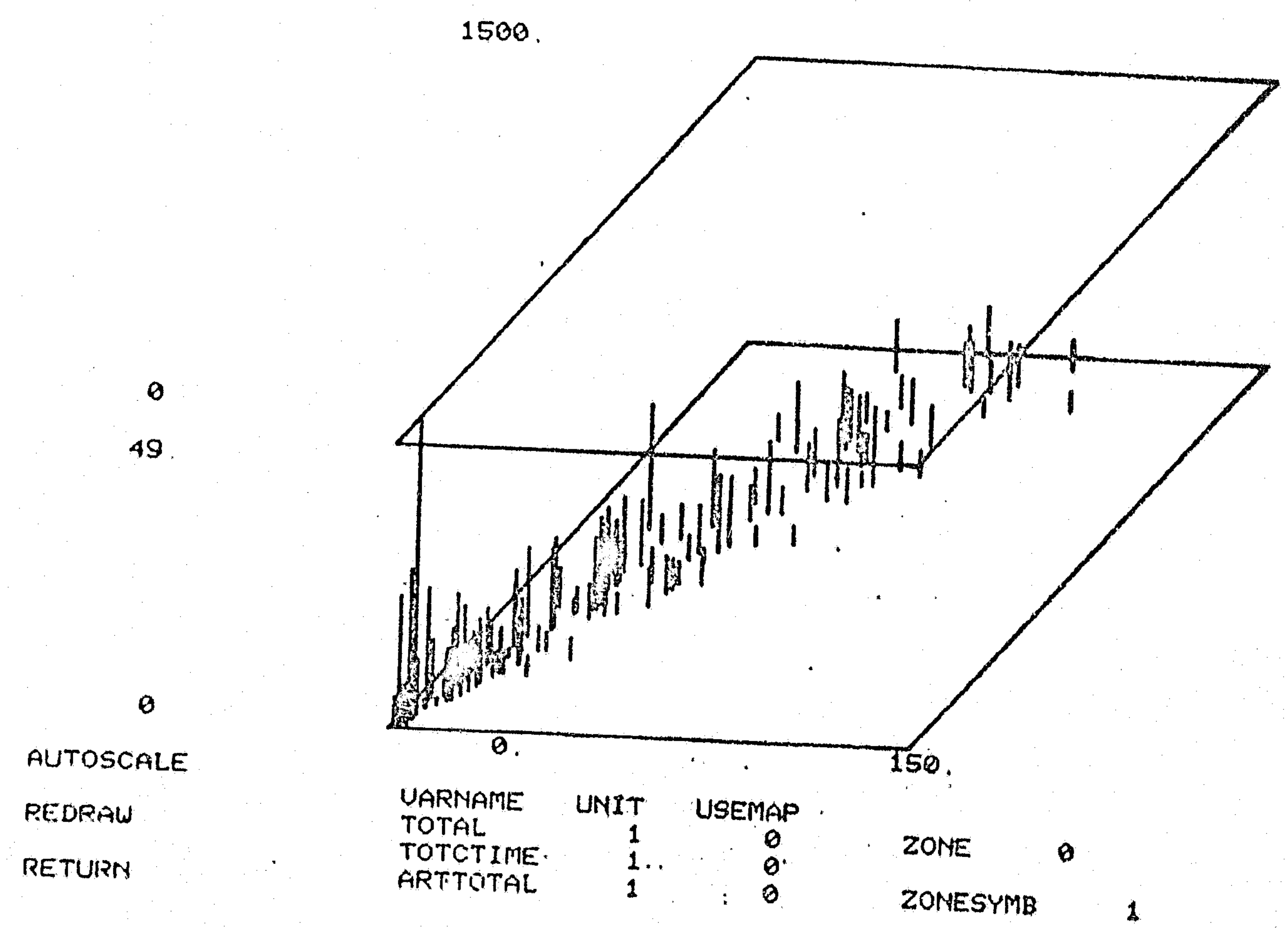


EXHIBIT 18

Criminal Justice
Demonstration Project

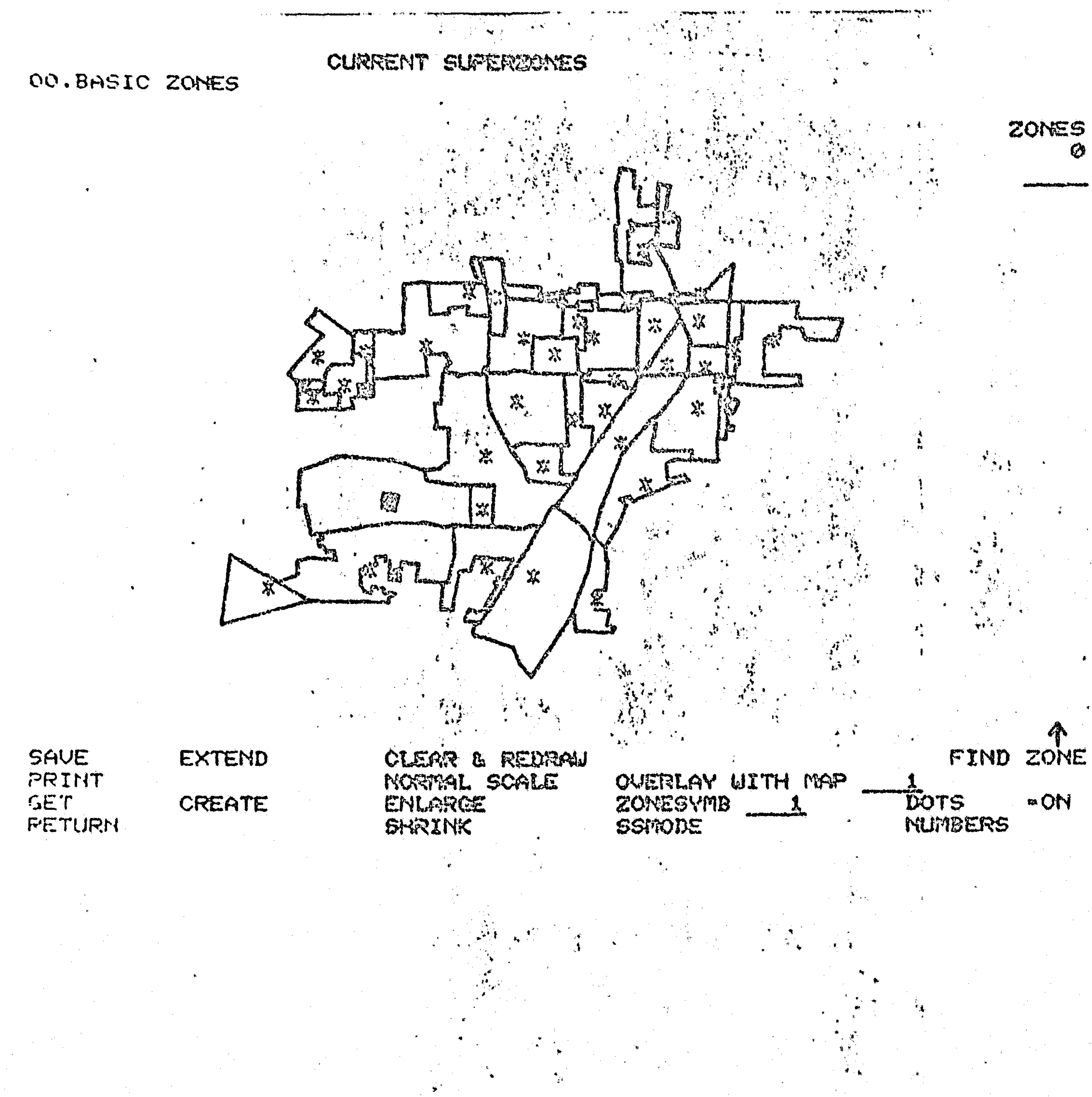
Campbell Police Department
March 5, 1975

Table of Contents of Reports

Exhibit 19

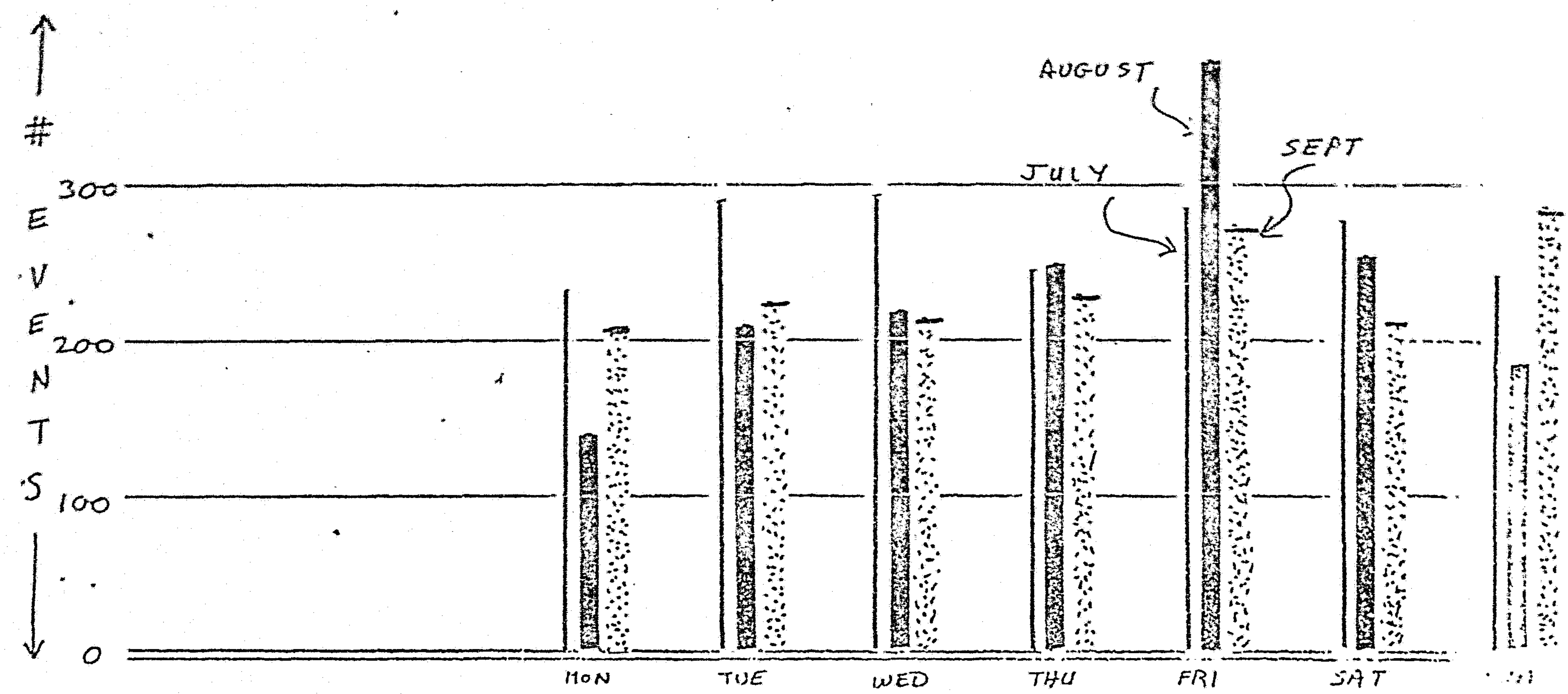
Campbell P.D. Calls-for-Service Study

Basic Zone Map of Campbell with * at the Centroid of Each Zone.



EVENTS (LESS 900 [illegible]) BY DAY

EXHIBIT 20

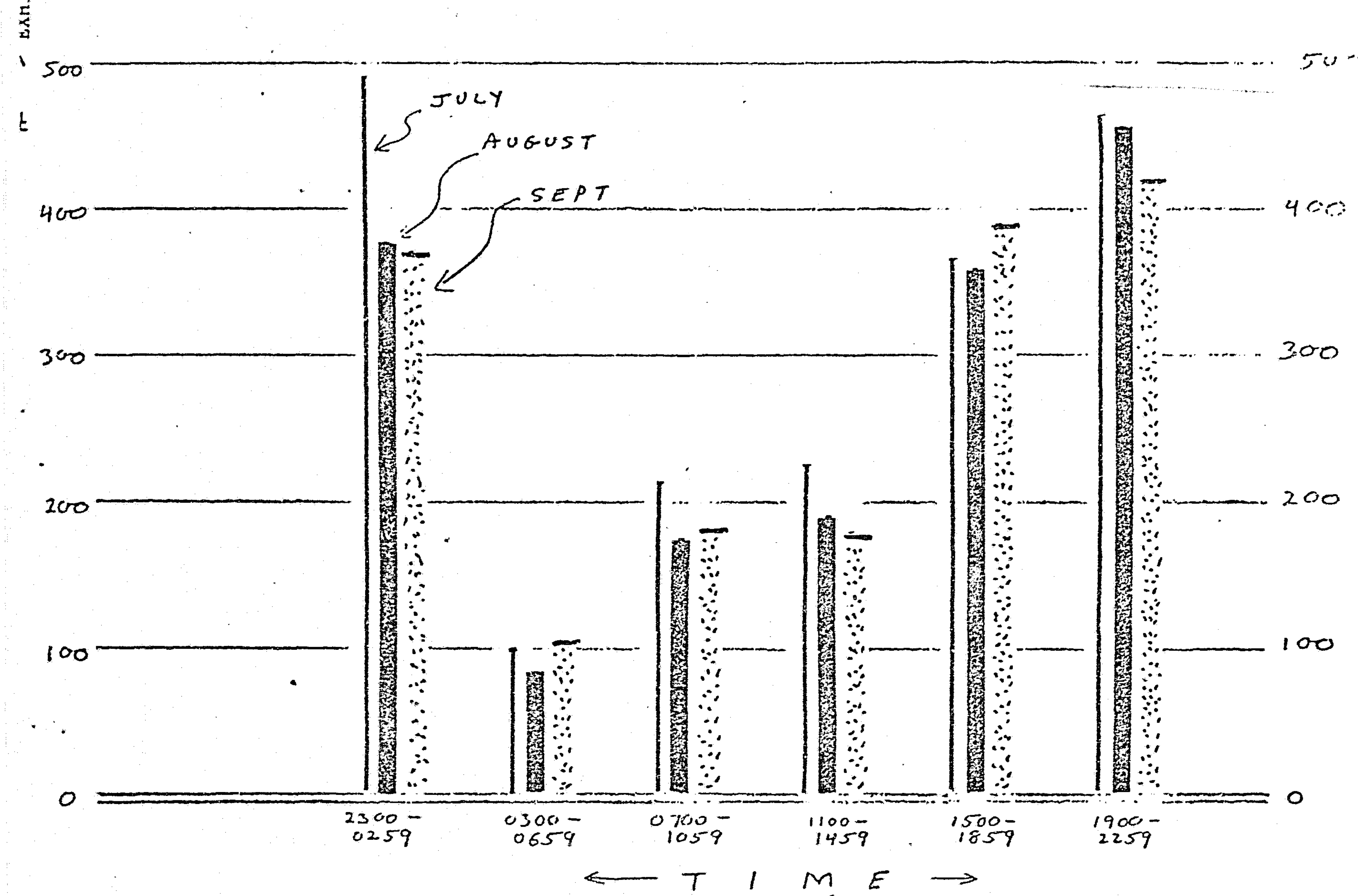


SOURCE: GILROY P.D. DISPATCH RECORDS JULY - SEPT 1974

PREPARED BY: CENTER FOR URBAN ANALYSIS
DEC 5, 1974
S.M.D.

EVENTS (LESS 900 CODES) BY HOUR

EXHIBIT 21

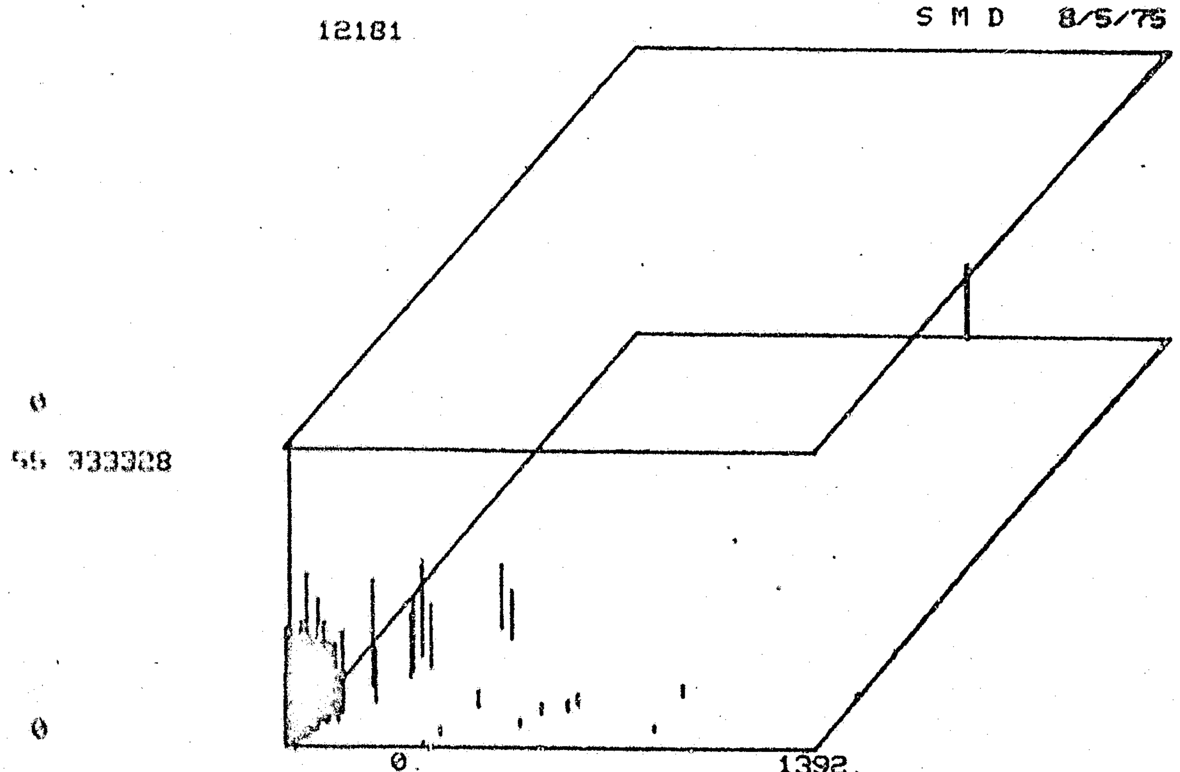


SOURCE: GILROY P.D. DISPATCH RECORDS JULY - SEPT 1974

PREPARED BY: CENTER FOR URBAN ANALYSIS
DEC 5, 1974
S.M.D.

EXHIBIT 22

GILROY POLICE DEPARTMENT CALLS FOR SERVICE
EACH STICK REPRESENTS THE VALUES FOR A ZONE   THE X AXIS IS TOTAL CALLS.
THE Y AXIS IS TOTAL CONSUMED TIME IN MINUTES AND THE Z AXIS IS AVERAGE
CONSUMED TIME IN MINUTES DURING THE PERIOD JULY - SEPT 1974
S M D   8/5/75



AUTOSCALE
REDRAW
RETURN

VARNAME	UNIT	USEMAP
TOTAL	1	0
TOTCTIME	1	0
TEMP	1	0

ZONE 0

ZONESYMB 1

EXHIBIT 23

CURRENT SUPERZONES

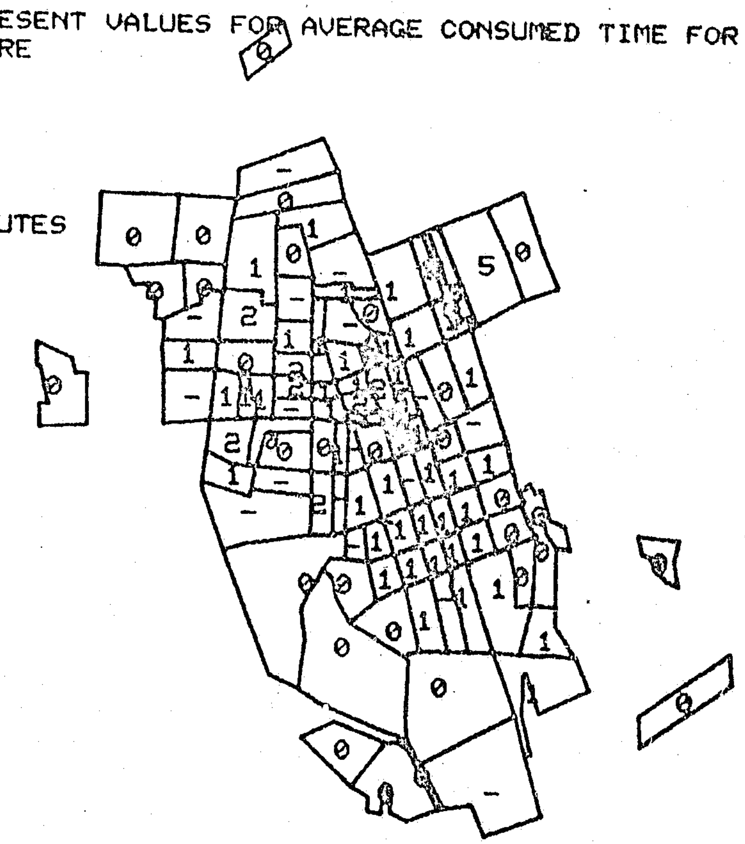
90 BASIC ZONES

SYMBOLS REPRESENT VALUES FOR AVERAGE CONSUMED TIME FOR EACH ZONE.

5 = 50 OR MORE
4 = 40-49
3 = 30-39
2 = 20-29
1 = 10-19
- = 5-9
0 = 0-4 MINUTES

S M D
8/5/75

ZONES
0



SAVE
PRINT
GET
RETURN

EXTEND

CREATE

CLEAR & REDRAW
NORMAL SCALE
ENLARGE
SHRINK

OVERLAY WITH MAP 1
ZONESYMB 2 =ON
SSMODE

FIND ZONE

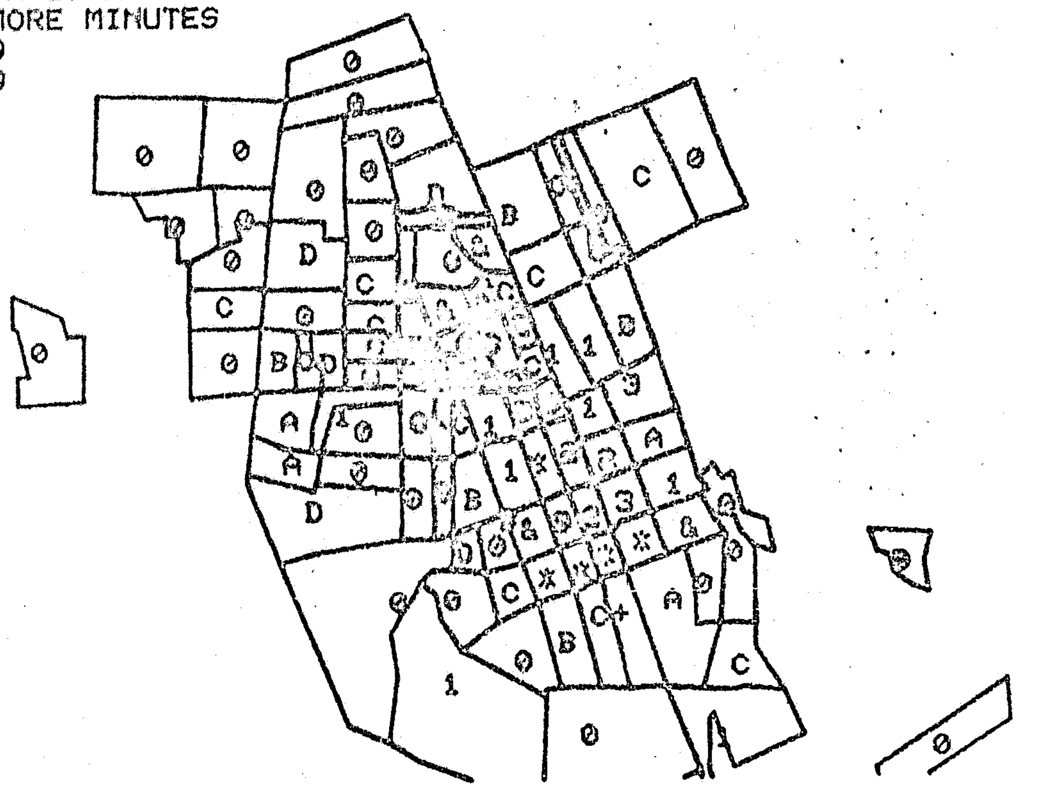
DOTS
NUMBERS

EXHIBIT 24

GILROY P .D CURRENT SUPERZONES CFS STUDY
00 BASIC ZONES

BASIC ZONE MAP OF GILROY WITH SYMBOLS REPRESENTING TOTAL CONSUMED TIME IN EACH ZONE. ZONES Ø

*=3000 OR MORE MINUTES
&=2000-2999
1=1000-1999
+= 850-999
= 700-849
= 550-699
= 400-549
= 250-399
= 100-249
= 50-99
Ø= 0-49

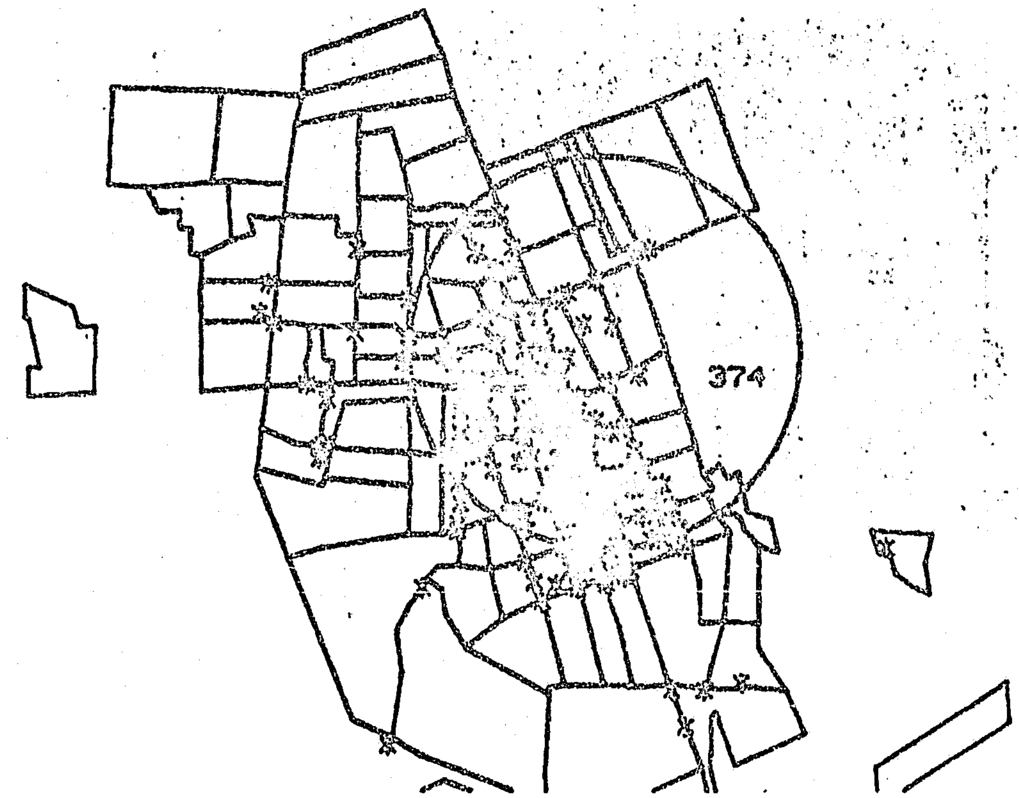


SAVE EXTEND CLEAR & REDRAW FIND ZONE
NORMAL SCALE OVERLAY WITH MAP 1
CREATE ENLARGE ZONESYMB 1 -ON DOTS
SHRINK SSMODE NUMBERS

EXHIBIT 25

SELECT CODE636
FILE EVENTS

GILROY POLICE DEPARTMENT CFS STUDY
BASIC ZONE MAP OF GILROY WITH * WHERE-EVER AN EVENT WITH ACTIVITY CODE = 636 OCCURRED AND COUNTING THE NUMBER OF THOSE EVENTS WITHIN A 1 MILE RADIUS CIRCLE.

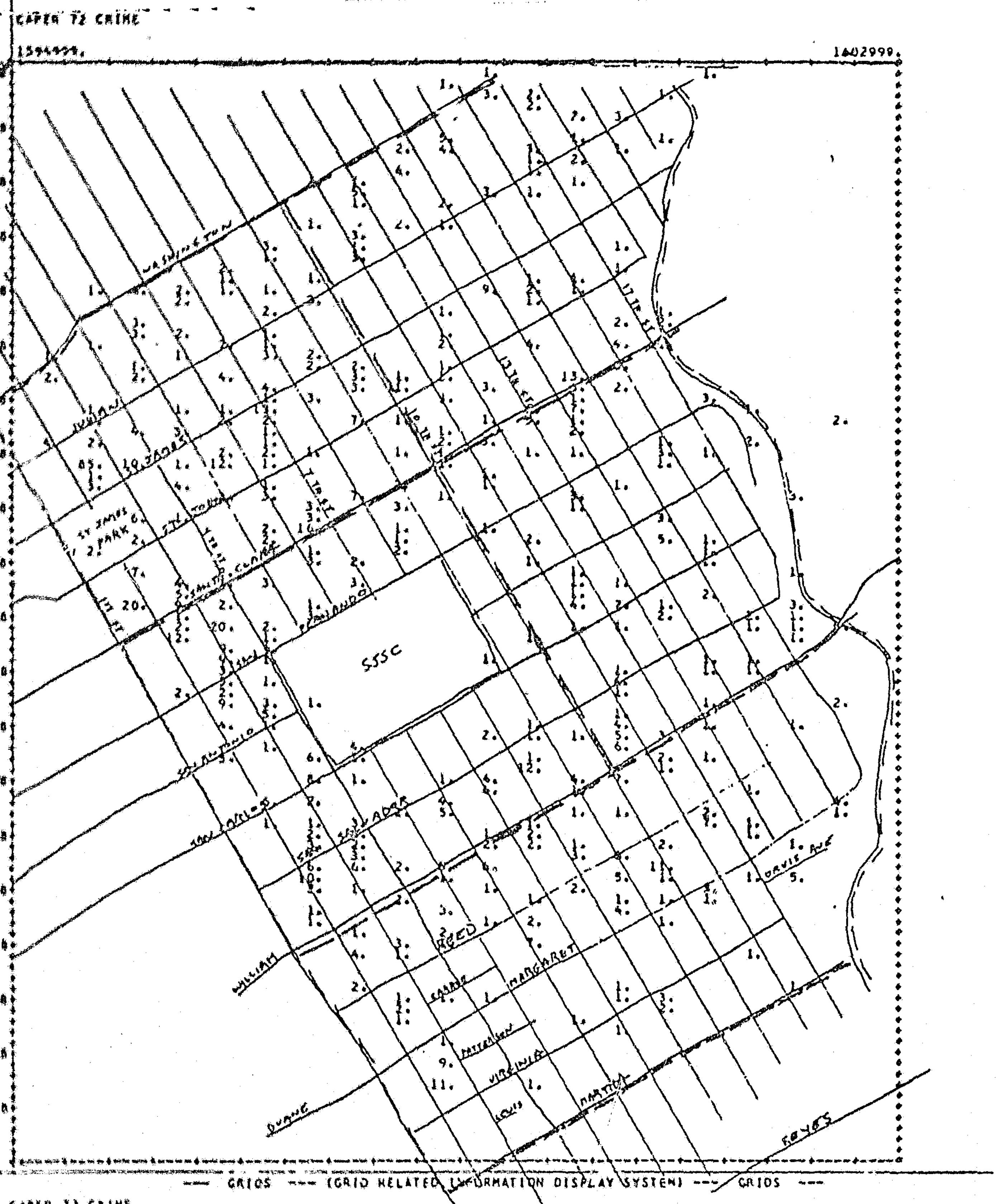


DRAW MAP
CENTER
EXPAND
SHRINK
NORMAL
SCALE

GRID

ERASE

RETURN DISPLAY EVENTS ALL
COUNT # IN AREAS
SUM BY AREA
MARK
DELETE
REDRAW AREAS
CREATE POLYGON
CREATE CIRCLE
END OF FILE



— GRIDS — (GRID RELATED INFORMATION DISPLAY SYSTEM) — GRIDS —

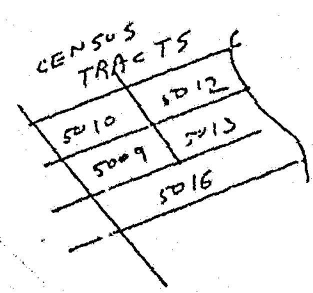
CAPER 72 CRIME
DATA SHOWS TOTAL CAPER EVENTS DURING FIRST HALF 1972
IN CENSUS TRACTS 5009, 5010, 5012, 5013 AND 5016
SOURCE - CAPER 72

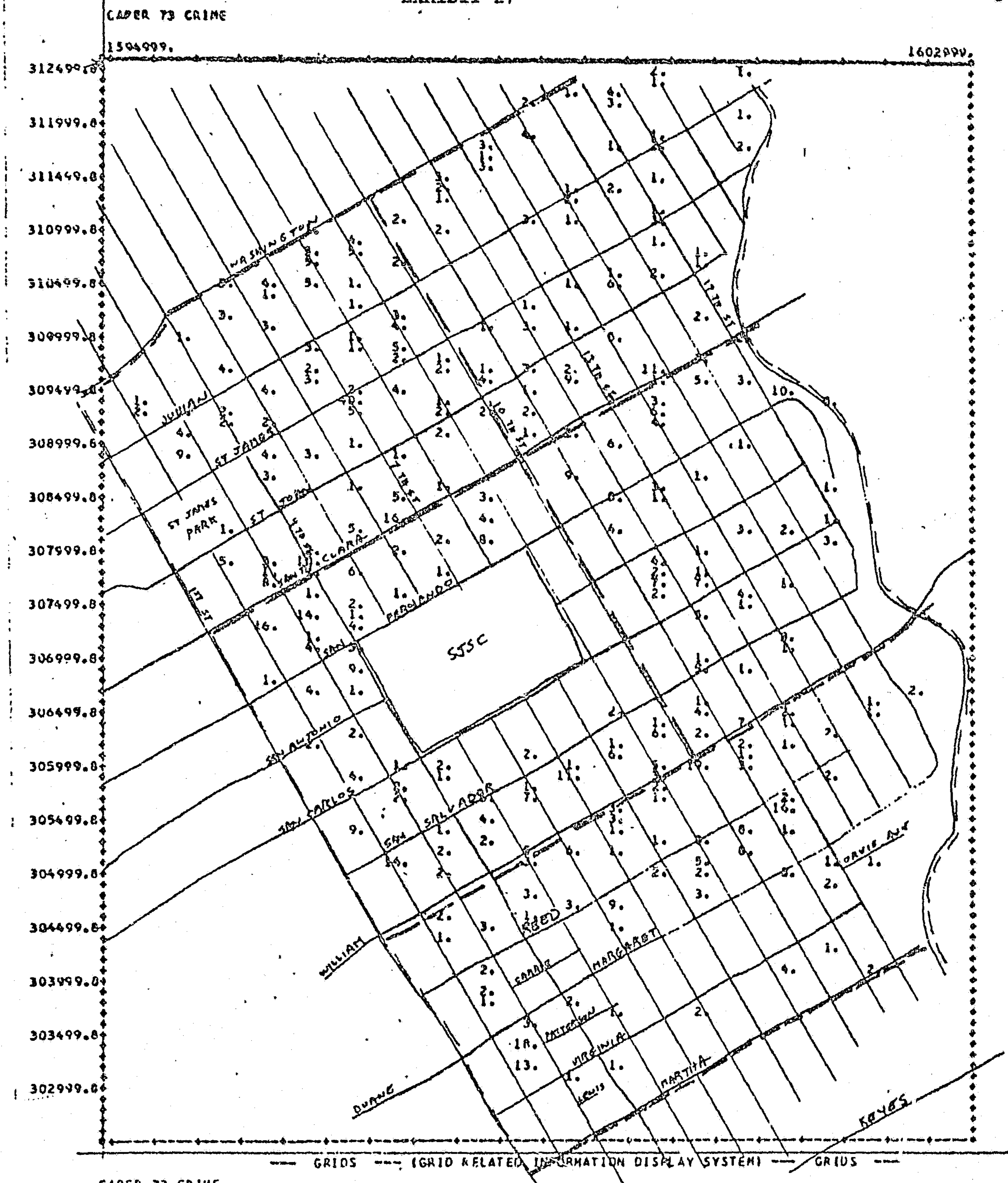
PREPARED BY THE CENTER FOR URBAN ANALYSIS, COUNTY OF SANTA CLARA
DATE PREPARED: APRIL 8, 1974

MAP SCALE - 1 INCH EQUALS 400 FEET

MINIMUM CELL VALUE(S) 0.100000E 01

MAXIMUM CELL VALUE(S) 0.850000E 02





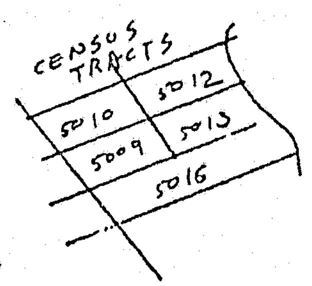
CAPER 73 CRIME
DATA SHOWS TOTAL CAPER EVENTS DURING FIRST HALF 1973
IN CENSUS TRACTS 5009, 5010, 5012, 5013 AND 5016

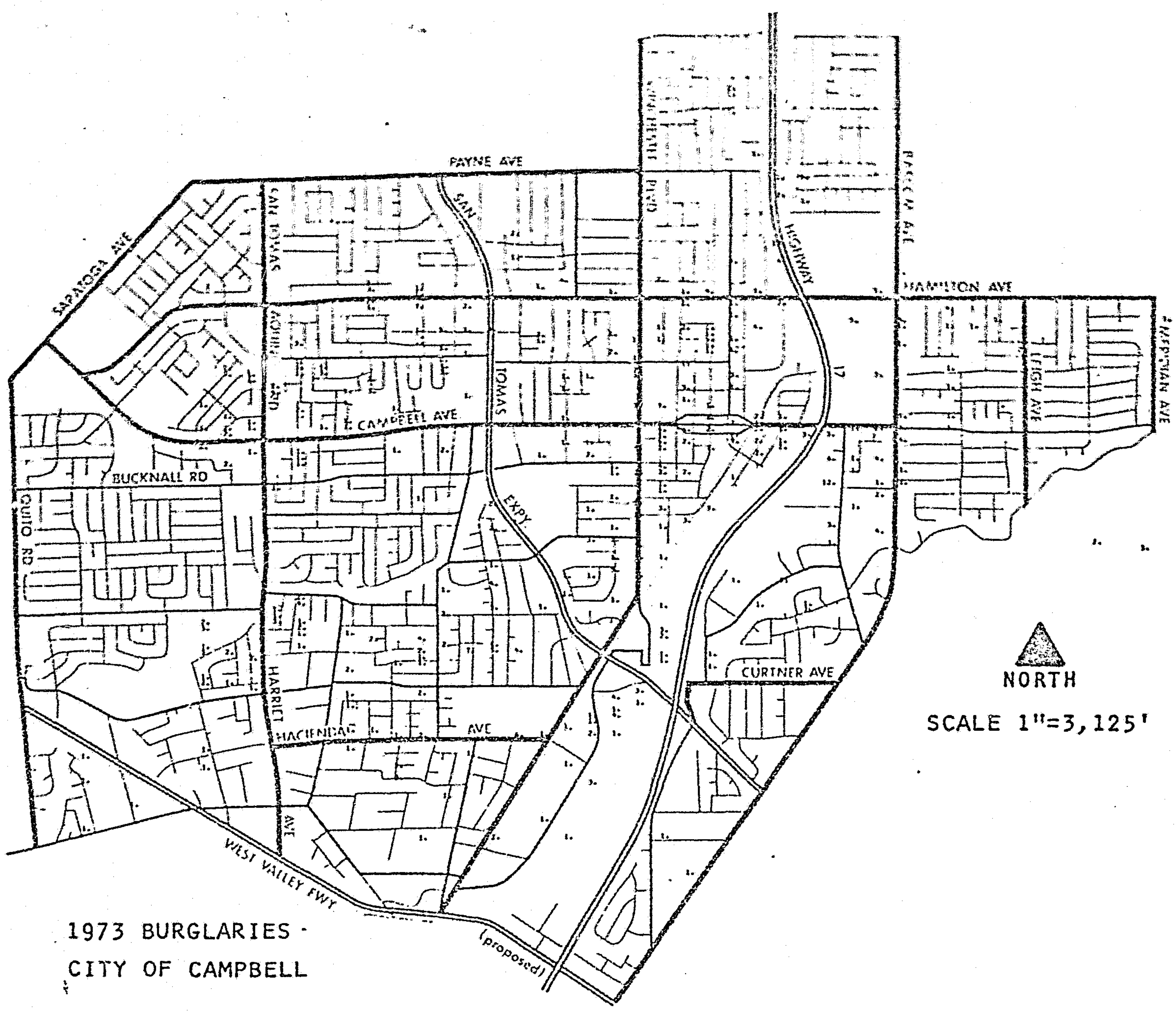
PREPARED BY THE CENTER FOR URBAN ANALYSIS, COUNTY OF SANTA CLARA
DATE PREPARED: APRIL 8, 1974

MAP SCALE - 1 INCH EQUALS 800 FEET

MINIMUM CELL VALUE(S) 0.100000E 01

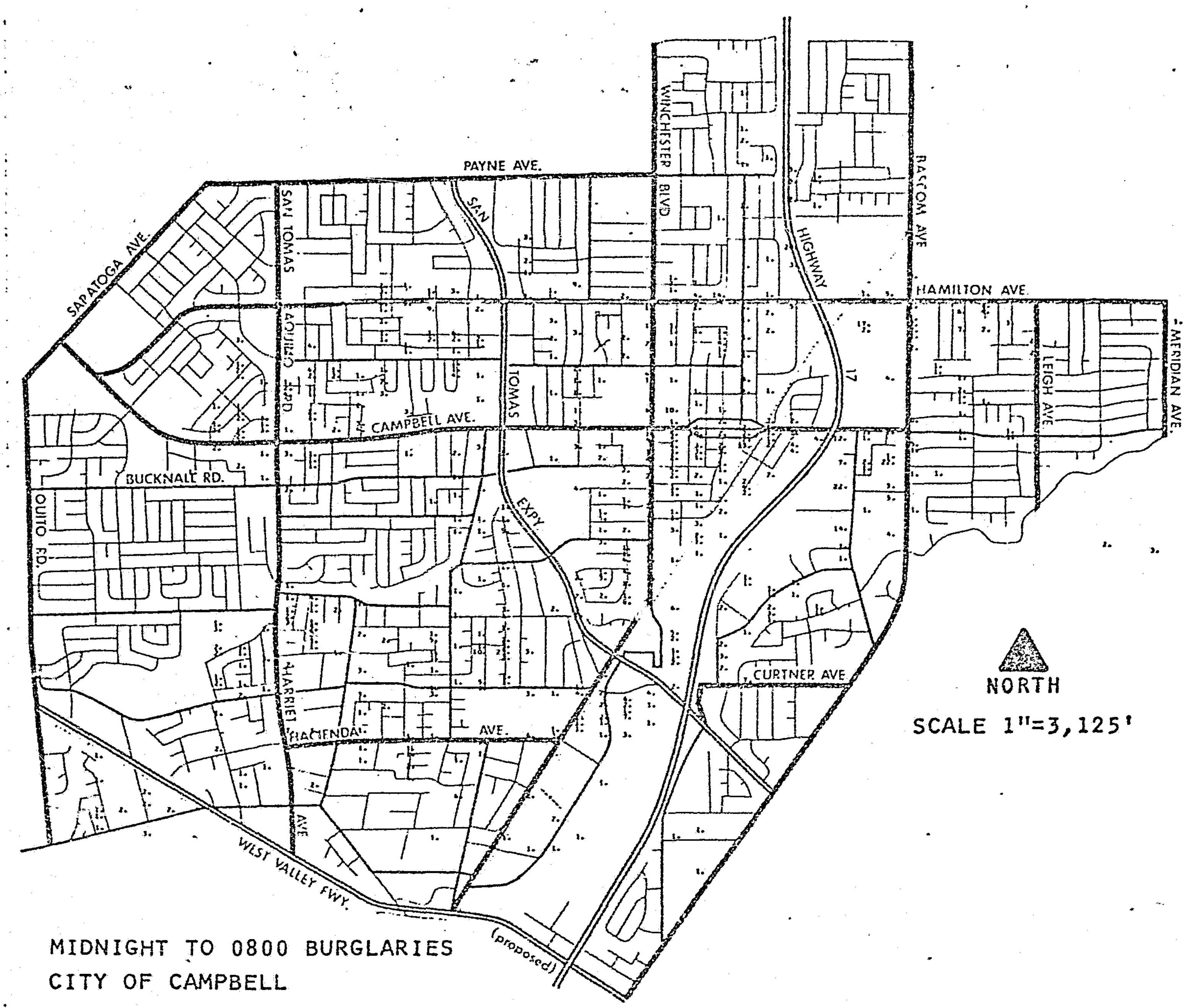
MAXIMUM CELL VALUE(S) 0.400000E 02





1973 BURGLARIES
CITY OF CAMPBELL

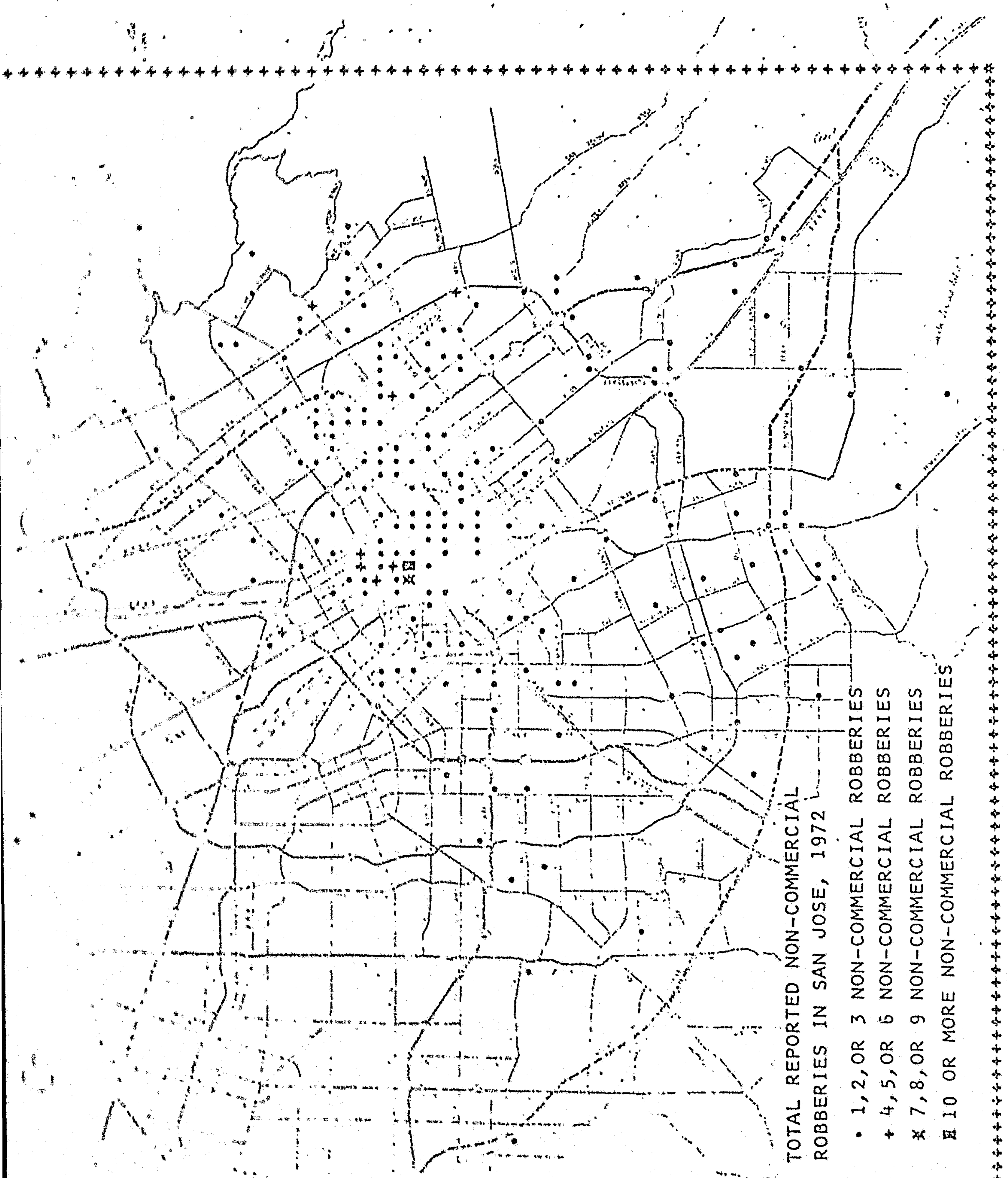
EXHIBIT 28



MIDNIGHT TO 0800 BURGLARIES
CITY OF CAMPBELL

EXHIBIT 29

EXHIBIT 30



REFERENCES

Bibliography of IBM Papers on Interactive Graphics

1. Mantey, P. E., Carlson, E. D.; "Integrated Data Bases for Municipal Decision-Making"; AFIPS Conference Proceeding volume 44, AFIPS Press, Montvale, N. J. 07645

2. P. E. Mantey, J. L. Bennett, E. D. Carlson "Information for Problem-Solving: The Development of An Interactive Geographic Information System," IEEE Int. Conf. on Communications, Volume II. Seattle, Washington, June, 1973

3. E. J. Cristiani, R. J. Evey, R. E. Goldman, P. E. Mantey, "An Interactive System for Aiding Evaluation of Local Government Polices", IEEE Transactions on Systems, Man and Cybernetics, Volume SMC-3, No. 2, March, 1973

4. E. D. Carlson, J. L. Bennett, G. M. Giddings, P. E. Mantey, "The Design and Evaluation of an Interactive Geo-Data Analysis and Display System", Proceedings of the IFIP Congress 74, International Federation for Information Processing, Stockholm, August, 1974. North Holland Publishing Company, Amsterdam, 1974

5. E. D. Carlson and J. A. Sutton, "A Case Study of Non-Programmer Interactive Problem-Solving", IBM Research Report RJ1382, IBM Research Laboratory, San Jose, CA, April, 1974

6. G. M. Giddings and E. D. Carlson, "An Interactive System for Creating, Editing and Displaying a Geographic Base File," IBM Research Report RJ1288, IBM Research Laboratory, San Jose, CA, 1973

7. B. F. Grace, "A Case Study of Man/Computer Problem-Solving: Observations on Interactive Formation of School Attendance Boundaries", IBM Research Report RJ1483, IBM Research Division, San Jose, CA, February, 1975

8. E. D. Carlson, B. F. Grace, J. A. Sutton, "Observations on End User Requirements for Interactive Problem-Solving Systems", to appear in MANAGEMENT DATAMATICS

(IBM confidential at this time)

9. D. M. Wytock, B. F. Grace, "Computer Assistance for Planning School Attendance Boundaries", paper prepared for presentation at 1975 URISA Conference

Budget Category	D-3238 Grant Funds	D-3238 Match Funds	A1899-2 Grant Funds	A1899-2 State Match	A1899-2 Local Match
Personal Services	110,097.	25,079.	36,262.	3,611.	-0-
Travel	1,197.	-0-	788.	-0-	-0-
Consulting Services	32,678.	10,000.	1,500.	-0-	-0-
Operating Expenses	16,908.	22,040.	26,450.	-0-	3,611.
Equipment	-0-	2,412.	-0-	-0-	-0-
TOTALS	160,880.	59,531.	65,000.	3,611.	3,611.

Appendix F = Cost Information

APPENDIX F: COSTS

Countywide CAPER Geocoding

Geocoding the Countywide CAPER data currently requires several passes thru the computer. However, the various phases can be processed as one job and the Center's costs to process 30,000 data records can be estimated as follows:

Set up Costs (2 hours x $25.00)	$ 50.00
Machine Costs*	$ 75.00
Total/Record	$ 0.004
Total	$125.00

Based on the Center's current matching experience of 62%, the cost per geocoded record is now $0.007. Upon implementation of the D.I.M.E. file as our primary geographic base file this cost should be reduced by 10-15% as the number of matches increase. However, at the same time, the ability to match intersection data will result in an increased cost. We estimate the additional run necessary to match intersection data will cost about the same as the non-intersection job and will result in a match rate of 90 - 95% of the intersection records. Therefore, upon implementation of the D.I.M.E. and D.I.M.E. Intersection files, the costs to process 30,000 records and successfully geocode 90% of them will be:

Set up Costs (4 hours x $25.00)	$100.00
Machine Costs*	$150.00
Total/Record	$ 0.008
Total/Geocoded Record	$ 0.010
Total	$250.00

*Machine Costs are those costs billed to the Center by GSA-Data Processing Center. A detail breakdown by item is available upon request.

Dec. 18, 1974

Dec. 18, 1974

APPENDIX F: COSTS

Countywide CAPER GRIDS Maps

Producing a set of GRIDS maps requires three separate steps after all geo-coding is completed. The first step is to convert the agency code to a numeric value and is required by the GRIDS program regardless of the number or types of maps requested. Based on previous experience, the costs for this operation, for a batch of 30,000 records are:

Set up cost (1 hour x $25.00)	$25.00
Machine Cost*	$13.53
Total/Record	$ 0.00128
Total	$38.53

The costs associated with the production of the maps, step 2, vary depending on the number and types of maps requested. If the requested maps differ from previously created maps only in the source data used, the following costs for a run of 16 maps displaying 30,000 records can be anticipated:

Set up cost (2 Hours x $25.00)	$ 50.00
Machine Cost*	$ 95.28
Total/Map	$ 9.08
Total	$145.28

If the requested maps differ from previously created maps in any way other than using different data, the set up costs will increase depending on the complexity of the change. However, a reasonable cost for designing, developing and implementing a totally new map should not exceed $25.00.

The third step in the process is the final printing and assembly of the maps produced in step 2. Again, based on past experience, the following costs for a run of 16 maps can be anticipated:

Set up cost (1.5 hour x $7.00)**	$ 10.50
Machine Cost*	$ 23.11
Assembly cost (9 Min/Map x 16 Maps x $15.00)	$ 36.00
Paper Cost (6 Pages/Map x $0.055/Page x 16 Maps)	$ 5.33
Total/Map	$ 4.68
Total	$ 74.94

Summary (assuming 16 previously set up maps displaying 30,000 records):

Set up Cost	$ 85.50
Machine Cost*	$131.92
Assembly Cost	$ 36.00
Paper Cost	$ 5.33
Total/Map	$ 16.17
Total	$258.75

**Data Processing Center's Data Control Clerk

*Machine Costs are those costs billed to the Center by GSA-Data Processing Center. A detail breakdown by item is available upon request.

CONTINUED

1 OF 2