131243

ARSON MEASUREMENT, ANALYSIS, AND PREVENTION

Final Report

James Alan Fox
College of Criminal Justice
Northeastern University

Prepared for the National Institute of Justice
under Grant #86-IJ-CX-0071

The views expressed in this report do not necessarily reflect the official position of the National Institute of Justice. I wish to thank John Pinette, Julie Bonsteel, and Kimberly Gatto of Northeastern University; Jennifer Mieth of the Massachusetts Department of Public Safety, Division of Fire Prevention; Kristina Rose of the National Criminal Justice Reference Service; and Michael Karter of the National Fire Protection Association for their assistance during various phases of this project. I am particularly grateful for the support and patience of my project monitor, Lois Felson Mock of the National Institute of Justice.

TABLE OF CONTENTS

1. The Problem of Arson

Arson, the crime of intentionally setting fire to property, has long been a major problem in the United States. In fact, with the exception of burglary, no other crime accounts for more direct property loss than arson. According to the National Fire Protection Association (NFPA) (see Hall, 1990), in 1989 the direct property loss from fires of incendiary or suspicious origin was 1.7 billion dollars nationally. Not included in this figure, moreover, are indirect losses (e.g., cost of relocation, loss of tax revenue, escalation of insurance premiums) which constitute an estimated four dollars for every one dollar of reported direct fire loss (Insurance Bureau of Canada, 1981).

Even more tragic than the economic loss is the major threat to human life that this crime represents. The NFPA also reported that in 1989, 615 civilians died as a result of arson fires (Hall, 1990). Added to this, of course, are the physical injuries and psychological scars suffered by victims of this crime.

Over time, the growth of arson has been staggering, at least until the most recent years. From 1955 to 1975, the incidence of reported arson rose 1,500%; and from 1975 to 1981, it rose an additional 250%. Also, whereas in 1964 arson represented only 3% of fire losses, by 1981 it represented as much as 40% of losses due to fire (Karter, 1982).

In the 1980s, the incidence of arson declined somewhat. For example, from 1981 to 1989, the estimated number of incendiary or suspicious structural fires nationally dropped precipitously, from 154,500 to 97,000 (Hall, 1990). In part, this downward trend may be due to the successful efforts of community activists who have forced policy-makers to prioritize the arson problem. Additionally, improvements in the economy during the 1980s may also have contributed to the reduction in the incidence of arson. Finally, the downturn in arson, may be partially related to changes in the age structure of the U.S. population. Adolescents are particularly prone to setting deliberate fires; in fact, half the arrests for arson in this country involve persons under the age of 20 (FBI, 1990). The decline in the size of this age group during the 1980s would have logically produced a reduction in the number of arson incidents.[1]

Unfortunately, this trend may also be short-lived. Expected resurgence in the relative size of this age group in the 1990s, the so-called "baby-boomerang effect" for the offspring of the original post-Word War II baby-boomers, may mean more crime, arson being one of them. Moreover, the economic recession that appears to be overtaking the nation is an ominous indicator of trouble in the years ahead.

---

[1]For a similar demographic argument in relation to other crimes, see Fox (1978, 1981, 1990).

2. Types of Arson

Much of the literature on the subject of arson has focused narrowly on the psychological characteristics of the arsonist. In their recent book, for example, Wooden and Berkey (1984) classify juvenile arsonists into four types--the "playing-with-matches firesetter," the "crying-for-help firesetter," the "delinquent firesetter," and the "severely disturbed firesetter."

Psychological explanations are also frequently applied to adult firesetting (e.g., see Rider 1980). For example, fire is often used as an instrument of revenge against an estranged lover or a former employer, as a means or hurting someone either directly through injury or indirectly by destroying the victim's property or business. Additionally, arson may be used instrumentally as a form of intimidation or in order to cover-up evidence of some criminal enterprise.

Quite different from psychological theories and profiles of vandals, pyromaniacs, and revenge-seekers are the economic explanations concerning arson-for-profit (see, for example, Brady, 1983; and Moore, 1981). Through a wide variety of clever schemes, the destruction of property by fire can produce a substantial payoff, either immediately in the form of insurance proceeds, or in the long run by removing impediments to real estate development.

In neighborhoods that are on the decline economically, particularly those in large urban centers, arson is in essence the "sale" of the property to the insurance company. Whenever a building has a greater paper value than market value, the likelihood of an arson fire tends to run high.

A number of arson convictions in recent years have uncovered unscrupulous speculators who purchased run-down tenements and "improved" them on paper alone. For example, an investor may buy and sell the same property several times within a short time span (perhaps between holding companies owned by the same individual), each time artificially increasing the purchase price and, as a consequence, boosting the insurance coverage. After several conveyances of the property, it can be worth on paper several times the actual dollar investment; a fire will then yield an enormous profit.

Under similar logic, arson buildings are often found to have had several smaller fires leading up to the large blaze that razes the property. Rather than simply an indication of fire proneness, a fire history may indicate a deliberate ploy to soak the property of as much money as possible. After each of the smaller fires, the owner uses part of the modest insurance settlement to pay corrupt contractors to "repair" the damage. The final large fire not only pays out the full value of the property, but covers up evidence that previous damage had never been repaired. Overall, the owner can collect a sum far exceeding that which was paid to acquire the property.

In certain "redlined" areas, furthermore, multiple-unit property owners may find it difficult to secure improvement loans from banks.[2] Unable financially to maintain his building, the owner may decide to abandon the property and take a tax loss, leaving it fair game for vandalism (e.g., juvenile arson). Alternatively, he may attempt to recover whatever value remains in the building through arson (i.e., insurance fraud).

So-called "upscale" or gentrifying neighborhoods are not immune to arson, but its form is quite different from "downscale" fires. Rather than being a final element in a sequence of maneuvers to exact profit, the arson fire in a gentrifying neighborhood removes an obstacle to profit-gathering.

In a fast developing real estate market, property values can spiral upward, increasing faster than rent payments, particularly in rent controlled buildings. An owner, seeing the opportunity for a quick turnaround of profit may decide to have a well-

---

[2]The practice of redlining is when a bank or lending institution illegally discriminates against loan applications from property owners in certain neighborhoods. They fear that a continued decline in property values will depreciate the loan's collateral, making it difficult to recover the loan even through foreclosure.

controlled fire set in the property in order to scare out tenants whose rent is costing him money.

Similarly, condominium conversion fires are designed to accomplish several purposes simultaneously. An apartment owner in a neighborhood where property values are rising and into which well-to-do professionals are moving, could profit considerably by converting his building to sale units. A ten rental-unit apartment building may be worth double its value if sold as ten individual condominiums. However, the landlord is not permitted, by ordinance, to evict tenants or to increase their rents significantly. A well-engineered fire will, however, not only force out the low-paying renters, but provide funds for rehabilitating the building and converting it to condominiums.

Still another form of arson in gentrifying neighborhoods is directly tied to the development of property. In a so-called "parcel-formation" fire, a building owner will burn his property and collect the insurance just prior to selling the land to a developer who is planning to construct a shopping mall or an office building on the site.

While structural arson may receive its fair share of attention, vehicular arson, although limited in the extent of destruction, is also a significant problem in some cities. Few cars burn accidentally or spontaneously; in contrast, there are a number of compelling motivations for vehicular arson.

In addition to cars that are stolen and later torched by teenaged thieves when the joyride is done, there are some fundamental economic factors underlying vehicular arson. Given the ease at which one may obtain an auto loan with little or no money down and with only a superficial credit check, many car owners soon find themselves unable to meet their payments. By torching the car, not only do they not lose whatever equity they may have in the car, but their credit rating is not affected.

Also, given the peculiarities of some auto insurance regulations, it may become profitable to "give away" a car to fire. An accumulation of body damage not covered by collision insurance can reduce a car's value below the blue book price. In such cases, a fire that destroys the car and covers the body damage can pay off (see Fox, 1985).

3. Studies of Arson Correlates

A large volume of literature is devoted to the topic of arson. Most that is written, however, concerns the technical aspects of arson investigation, the psychological dynamics of pyromania, or even the economic aspects of this crime (e.g., see New York State Academy of Fire Science, n.d.). There are only a handful of quantitative studies that have focused on arson rates and their correlates.

Karter and Donner (1978) conducted one of the earliest and most simplistic examinations of fire and arson correlates. They analyzed the effects of demographic and socio-economic characteristics on fire rates (not specifically arson rates) in five major metropolitan areas: Kansas City; Syracuse; Newark; Phoenix; and Toledo. Fire rates--expressed as the number of fires per 1,000 population--were measured at the census tract level and analyzed in relation to a variety of census population and housing characteristics. The housing characteristics included home ownership, age of structures, crowdedness within structures, vacancy, and size of structures, while population measures included racial composition, poverty, affluence, unemployment, poverty by race, undereducation, undereducation by race, high-school education, transiency, family stability, and the elderly population.

Karter and Donner examined the five communities separately and selected the single housing and single population characteristics most highly correlated with the fire rate in each city. This was determined by the percentage of housing units per tract which possessed a given variable or attribute.

In Syracuse, the population characteristic found to be most frequently related to a high rate of fire was family stability (the percent of children living with both parents) and the housing characteristic determined to be most strongly correlated to fire was crowdedness (the percent of year-round units with at least 1.01 persons per room). Newark fires were found to be most strongly correlated with poverty (the percent of persons below the poverty level) and ownership (the percent of owner-occupied buildings). The remaining three cities had similar fire rate correlates: poverty and crowdedness for Phoenix; poverty and ownership for Toledo; and family stability and vacancy (percent of year-round housing units that were vacant) for Kansas City.

Karter and Donner (1977) conducted a more elaborate study of fire rates and census characteristics using data from Kansas City, Syracuse, Newark, Phoenix, Toledo, Seattle, Fairfax County (Virginia), Charlotte, and St.Petersburg. Aggregating by census tract, fire data for 1970 and 1975 were correlated with the same eleven population characteristics and five housing characteristics from the 1970 Census (specifically, racial composition, poverty, affluence, unemployment, poverty status by race, undereducation, undereducation by race, high-school education, transiency, family stability, elderly population, home ownership, age of structure, crowdedness within the structure, vacancy, and size of the structure). Unlike their simpler study, in this effort Karter and Donner considered more precise assessments of statistical association as well as alternative measures of fire rates, including the number of fires per 1,000 people and the number of fires per 1,000 housing units.[3]

The first phase of the analysis involved intra-community comparisons of housing and population characteristics on fire rate activity. For each dependent variable (per capita and per housing units fire rates), three regressions were computed: a "housing model" which contained the five housing characteristics, a "population model" comprised of the eleven population characteristics, and a "combined model" which contained all sixteen population and housing variables. For all communities and both fire rates, each set of characteristics was found to be highly significant. For seven of the communities, the percent of

---

[3]Karter and Donner (1977) had compared four measures of fire rate: fires per 1,000 population; fires per 1,000 housing units; fires per the square root of the product of 1,000 population and 1,000 housing units; and the square root of fires per the square root of the product of 1,000 population and 1,000 housing units. They found that the first three rates yielded very similar results and, without clear reason, decided only to focus on the first two. The differences created by the fourth measure are understandable, as there is no justification for taking the square root of the fire count.

variation explained in the two fire rates differed by no more than 10% in any of the three models. Phoenix and St. Petersburg were the two exceptions. For both of these cities, the variation explained in the population-based rate was considerably less than that explained in housing-based rate. This finding was consistent across all three models.

Next, the effect of the population characteristics on the fire rates after adjusting for the effect of housing characteristics (and vice versa) were calculated and tested for significance. For the most part, it was found that both sets of variables explained a significantly greater share of the variance in fire rates than either set alone.[4]

The final phase of analysis involved only Seattle, in which block group data were compared to the tract level data, in order to determine whether the census characteristics would remain significant at a smaller geographical level. Other geographical levels, such as the block-face and the dwelling levels, were considered but were ruled out because the data proved to be difficult to obtain and to be unavailable. The analysis of fire rates and census characteristics at tract and block group levels was completed in three ways: 1) descriptive comparisons, 2) fire rate correlations at tract and block group levels, and 3) an

[4]Homogeneity of the effects of individual factors across cities were examined, yet the results were not overly encouraging.

analysis of variance (ANOVA). Again, the Karter and Donner chose to use two dependent variables: fires per 1,000 people and fires per 1,000 housing units.

Mean fire rates were found to be similar at the tract and block group levels; however, as expected, more variation existed at the block group level. For fires per 1,000 population, the correlations between census characteristics and fire rates were 50% higher at the tract level than the block group level. For fires per 1,000 housing units, in 13 out of the 15 census characteristics correlations were at least 50% higher at the tract level than the block group level. Using ANOVA and an intra-class coefficient analysis, it was found that the block groups <u>within</u> a tract were more homogeneous with respect to fire rate and census characteristics than were block groups overall.

In a similar study, Schaenman, Hall, Schainblatt, Swartz, and Karter (1977) examined the effects of census variables on the fire rates of five U.S. cities: Fairfax County (Virginia), St. Petersburg, Charlotte, San Diego, and Seattle. Specifically, Schaenman at al. focused on the extent to which inter-tract variation in residential fire rates could be explained by the same set of sixteen socioeconomic and housing characteristics taken from the 1970 census.

Schaenman et al. performed a series of regression analyses in which fire rates were expressed in terms of residential fires per 1,000 population. The amount of data for each community varied; for example, four years of data were available for Fairfax County, while a mere six months of data for San Diego were available.

Across the regressions, three variables emerged as strong predictors of residential fire rates: parental presence, poverty, and undereducation (the percentage of persons over age 25 with less than eight years of education). On average, parental presence accounted for 40% of the variation in the six analyses, while undereducation, on average, accounted for 39% of total variation. Poverty explained approximately 40% of the total variation across the regression runs.

Another regression analysis combined or pooled the data from all of the cities except San Diego. In this analysis, parental presence accounted for 52% of the variation, undereducation explained 45%, and poverty explained 50% of the variation in fire rate. It is important to note, however, that these percentages reflect variation explained by only one variable at a time, based on simple bivariate correlations.

Two multiple regressions were then conducted in order to test the strength of the form of fire rate. Two fire rates were compared: one based on fires per 1,000 people, and one based upon fires per 1,000 households. Parental presence, undereducation, housing vacancy, and poverty (in that order) accounted for a combined 60% of the total variation for fires per 1,000 population, with parental presence accounting for 52% of this figure. Fifty-eight percent of the variation in fires per 1,000 households were accounted for by undereducation, housing crowdedness, housing vacancy, elderly population, and poverty; of these, undereducation explained 51% of the variation. Due to the similarities between these two sets of results, Schaenman et al. concluded that the form of the fire rate does not an appreciable difference.[5]

Like Schaenman et al., Gunther (1981) conducted a study of the socioeconomic correlates of residential fires. His analysis focused upon neighborhood differences in the causation of fire in inner-city Toledo, Ohio, using fire data for 1976-1979 along with 1970 Census data. Since previous studies had found socioeconomic variables such as education, single parent households, crowdedness, and family stability to be confounded by family income and race variables, Gunther used only the latter two as the socioeconomic factors in neighborhood definitions.

---

[5]Schaenman et al. acknowledged the existence of substantial multicollinearity: "The strong and moderately strong variables were generally highly correlated, so any conclusions as to which ones explain the most variation in fire rate must be considered very tentative" (p. 62). They made no effort, however, to correct or adjust for its effects.

Census tracts were subjectively (and unscientifically) grouped by a long-time Toledo resident and employee of the U.S. Fire Administration. Cause of fire, the dependent variable, was divided into seven categories: incendiary-suspicious, smoking, children playing, cooking, heating, electrical distribution. and appliances. These were classified as either "non-equipment" or "equipment" causes.

Analysis of census tracts revealed a strong correlation between income and the frequency of fires per 1,000 population within a census tract. Fire rates were substantially greater in tracts with median incomes less than $4,000, especially in inner-city tracts. When compared at equal levels of income, racial composition did not reveal any significant effects on tract fire rates. Race, Gunther concluded, has no significant influence over fire rates, "except insofar as the differences show up as differences in income" (p. 56).

The Institute of Puget Sound Needs (1976) analyzed arson data taken from the Seattle Fire Department for the years 1965-75 along with 1970 Census data in an attempt to identify relationships between arson activity and demographic variables among census tracts. Several different types of arson ,including juvenile arson and adult arson, were considered separately, and each was analyzed in both per capita and per housing unit forms.

On the basis of regression analysis, several variables were found to be significant in their association with juvenile and adult arson. For adult arson, significant correlates included: proximity to the central business district, presence of subsidized housing projects, prevalence of low rent units, proximity to hospitals, and employment and new business starts The last two variables, employment and new businesses, both economic in nature, explained 57% of the total variation in adult arson activity among census tracts. All of the economic variables included in the regression (15 in all) together accounted for 73% of the variance.

Bennett, Merlo, and Leiker (1987) analyzed patterns of incendiary and accidental fires in Springfield, Massachusetts census tracts, combining fire data for 1980-1984 taken from the Springfield Fire Department along with 1980 Census data. The dependent variable was identified as fire frequencies, by type, by census tract. Fire rates (incendiary and accidental) were calculated both per square mile and per 1,000 population.

A stepwise regression was produced using a set of thirteen independent variables representing social, economic and housing characteristics. Using conventional thresholds for variable selection, only two variables entered into the equation, namely, Vacancy (the percentage of housing units reported vacant in the 1980 Census) and Tenements (the percentage of housing structures with 5 or more units). Vacancy alone accounted for 61.9% of the variation in reported arson rates. When tenements was added to the regression, both variables together explained 70.4% of the variation. Bennett et al. concluded that vacancy may be a significant factor in the distribution of incendiary fires in Springfield and that this variable has not been given sufficient attention in previous research.

The program of research by Karter and his associates may have been pioneering, but, more recently, Pettiway has produced the most complex analyses of arson patterns. In his first approach to arson correlates, Pettiway (1983) focused on 727 cities with populations of 25,000 or over. Despite coverage, consistency and completeness problems in this data source, Pettiway used arson figures from the 1980 Uniform Crime Reports, in addition to 58 measures drawn from the 1977 County and City Data Book.

Rather than using population-based rates of arson, Pettiway developed "arson risk rates" with a denominator reflecting the targets of arson. Following a variable-based clustering, five dimensions were found to be significantly correlated with arson activity: 1) economic base 2) SES, 3) crime, 4) taxes-expenditures, and 5) age-crowding. Economic base was defined as the number of retail, manufacturing, and wholesale establishments per capita; total retail sales per capita; population density;

and the change in population density. The SES dimension was comprised of traditional indices of status, including income, education, housing values, and rent. Crime was a measure of index offenses; however, arson did not load on this dimension. Taxes-expenditures measured such items as per capita expenditures, percent expenditures for education, and the average property taxes per capita. Age-crowding included the percent of population over eighteen years of age, death rates per 1,000 persons, and crowding.

In addition to variable clustering, the cities were clustered or grouped into 15 sub-types. Arson rates were determined to be highest in two city clusters: a cluster defined as a low tax and low age city type, and a cluster described as low SES and high crime. While high arson rates in this study seemed to occur only under these two environmental conditions, low arson rates occurred under a variety of conditions. Low arson rates were found in city types that were average on all dimensions, and the lowest arson rates were located in cities with older, high status populations and low general crime rates.

Pettiway concluded that "certain mixtures of the environmental-structure dimensions of urbanism are generally good predictors of arson-risk rates. The association between the level of arson and the structural attributes of cities within types should not be considered causal relations but...as 'contingent control' relations....Low SES has been considered the root cause of expressive and instrumental violence, but its influence on arson may be conditioned by SES's interaction with other dimensions of urbanism" (1983, p. 173).

Pettiway (1985a) continued his study of arson with an examination of ways to standardize arson rates. Utilizing data from the 1980 Census and 1978-1979 fire data from Houston, both aggregated to census tract levels, two arson rates were constructed and contrasted: an indirect rate which utilized indirect measures of opportunity as its divisor, and a direct rate whose divisor was equal to the product of population times opportunity.

Through factor analysis, Pettiway reduced a 27 x 27 correlation matrix of census variables to six significant factors; these factors accounted for 56% of the total variance. The most significant factor was family dissolution, which accounted for 24.2% of the variance, followed by population density (7.8%), race-resources (6.8%), single-family residences (5.5%), vacancy (5.4%), and old housing (5.2%). By using factor-based scales for independent variables, Pettiway conducted a multiple regression, focusing on differences in results under varied operationalizations of the arson rate as the dependent variable.

Due to the change in the degree of correlation from one operationalization to another, arson rate distributions were examined for skewness. One scale was found to be particularly skewed, creating a tremendous effect on the fit of the regression model. A log transformation was utilized in order to reduce the skewness; when this transformed measure was used as a dependent variable, results were somewhat different. However, five of the six variables which had been associated with the untransformed rate measure remained correlated after the transformation. Across rate measures, four variables yielded consistently positive results: 1) the percentage of structures with thirteen or more stories, 2) the percentage of the population employed as handlers, equipment cleaners, or laborers, 3) the percentage of structures with seven to twelve stories, and 4) median family income.

Using the same data, Pettiway (1985b) chose next to compare neighborhood and environmental factors in Houston. The tract population was used as the base for arson rates in this study. Non-populated and sparsely-populated tracts were excluded from the analysis in order to avoid spatial skewness. Any skewness which did appear in the arson rate was transformed using a log binomial transformation.

Pettiway began the analysis by grouping the census tracts into homogenous clusters. Differences in the arson rates between groups were measured with an analysis of variance, and a stepwise regression assessed the impact of neighborhood membership on arson rates.

Variable cluster analysis revealed eight dimensions of urban structure which accounted for 88% of the variance in the original matrix of ninety census variables. The highest arson rate occurred in a dimension characterized by a high percentage of rental units, low-income families, female-headed households, old housing, and structures lacking in some facilities.

Although significant differences in arson rate were found among the eight neighborhood types, an analysis of covariance determined that neighborhood-type membership did not have a significant effect on arson rates beyond the individual effect of a dimension's attendant variables. Based on these results, Pettiway concluded that neighborhoods were not unique in the nature their arson problems, and thus arson prevention strategies need not be tailored for different parts of the city.

Pettiway (1988) next investigated whether arson patterns differed between ghetto and non-ghetto neighborhoods. Using the same fire and census data for Houston census tracts, tracts were classified in terms of "ghetto zones" and "non-ghetto zones" based on the percentage of blacks within each census tract. Ghetto zones consisted of those areas with at least a 30% black population. The arson rate measure (the dependent variable) was defined according to opportunity rates and was calculated using the ratio of single-family residential land use to other types of land use and the ratio of the population under 25 years to the population over 25 years.

The dependent variable was then regressed against a set of independent variables for each residential cluster. It was determined that 105 census tracts constituted the Houston black residential community and that 66.6% of the tracts were ghetto core areas (75% of the population was black). Not surprisingly, great differences were found through a discriminant analysis between the ghetto and non-ghetto areas. High-loading variables which characterized the black residential community were: female-headed households, separated males, level of unemployment, percent of population between the ages of 14 and 18, and stable population. Discriminating variables indicated that the indices of age, family dissolution, residential stability, economic position, land-use pattern, structural density, and structural inadequacy mark the differences between ghetto and non-ghetto clusters.

Finally, a regression analysis was conducted. Due to the highly skewed arson rate measure, a log transformation was used to normalize the rate, and separate regressions were performed on 103 ghetto and 232 non-ghetto tracts.

In the ghetto neighborhoods, the median family income, the percentage of males aged 25-34 years old, and the percentage of buildings built prior to 1939 accounted for 37% of the variation in arson. In non-ghetto areas, the following variables accounted for 42% of the arson variation: percent of buildings constructed prior to 1939, percent of buildings with 13 or more floors, percent of buildings with 7 to 12 floors, and percent of units lacking kitchen facilities. Pettiway concluded that, "with the exception of [the percent of owner-occupied housing built 1939 or earlier], the models differ not only in the variables associated with the arson rate structure but also in the degree of explanation associated with each model. For the non-ghetto model the degree of association...is higher and appears to be conditioned by structural characteristics....In ghetto areas, the explanation appears to be mixed: the arson rate structure is explained best by age of housing and characteristics of the population. Income is an explanatory variable that is correlated more highly with the rate measure for ghetto tracts" (1988, pp. 124-125).

Pettiway surmised that the arson rate structure in ghetto areas, related to income and age of housing stock, is a product of economic and racial segregation. He postulated that the presence of old housing stock in ghetto areas provides more opportunity in terms of a targets and may be attracting arsonists from outside of the ghetto tracts as well. In non-ghetto areas, the structural density indicates competition for centrally located spaces, and arson in these areas represents the pressures of urban growth.

In sum, the literature is quite varied in scope (from single-city analyses to multi-city analysis), in unit of analysis (from city-wide rates to block group rates), in complexity (from two variable models to models encompassing dozens of correlates), in technical approach (from simple correlation to analyses of clusters), in methodological soundness (from straight-forward data analysis to adjustments for skewness), and in arson measurement (from population and housing based rates to indirect measures). Despite this diversity, researchers appear to reach similar conclusions about environmental, housing and demographic correlates of arson. Chiefly, the literature reviewed above indicates that indices of poverty, family dissolution, housing quality and building vacancy are consistently correlated with measures of arson incidence.

4. Arson Data Sources

The purpose of this project was to replicate the statistical work of Karter, Pettiway, and others, while advancing the level of technical soundness. Regardless of technique, of course, the most fundamental issue in relation to the quality of results involves the data source itself.

Of all serious crimes, arson has received the least attention from a quantitative standpoint. Surely, the potential for research on arson has been limited by a severe lack of reliable data on arson incidents.

4.1 UCR Arson Data

In 1979, following a congressional mandate, the FBI elevated arson to the status of a Part I or "Index" offense and thus began collecting and disseminating reports of arson incidents (see FBI, 1980). The results of this endeavor, however, have been most disappointing and incomplete.

For 1981, for example, the Massachusetts communities for which reports of arson data were absent included Boston and Lynn, two cities whose arson problems have been noteworthy and newsworthy (FBI, 1982). For 1982, moreover, the FBI noted that law enforcement agencies representing 87 percent of the U.S. population furnished at least <u>six</u> monthly arson reports (FBI, 1983). This is hardly the level of coverage and consistency achieved for other portions of the Part I return.

Although nearly a decade has past since arson was made a Part I offense, the coverage still remains relatively poor. In 1989, for example, only 72 percent of agencies provided full reports of arson (FBI, 1990).[6] More important, the pattern of missing agencies is not random. In a comparison of official UCR reports with sample survey data from 683 fire departments, Jackson (1988) observed that the tendency to report arson data to the UCR was inversely related to the prevalence of arson, thereby constituting a serious bias to the official data.

In addition to the coverage problem is the superficial nature of these data. They are limited to the number of arson incidents, number of clearances for arson offenses, type of structures, and estimated monetary value from arson fire loss. Besides excluding suspicious fires altogether (which will grossly understate the enormity of the crime), these reports lack most of the critical pieces of information necessary to understand the nature of arson. Finally, and most important, UCR data on arson

---

[6]Undoubtedly, the non-compliance problem is a result of the fact that arson is fundamentally different from all other Part I offenses. A report of a robbery or auto theft, for example, is generally determined to be founded or unfounded within a brief period of time, often 24 hours. In contrast, it may take a week or longer following a fire for the determination of cause to be made by the fire department or fire marshal. This time lag may be largely responsible for the data problems inherent in the UCR counts of arson.

provide only monthly aggregates for entire reporting jurisdictions on all items. A data source to study the nature and correlates of arson would need to be incident-based.

4.2 National Fire Incident Reporting System

Fortunately, an alternative data resource for studying arson has emerged in recent years. While the FBI and its network of police agencies may not have given arson data collection high priority, the Federal Emergency Management Agency (FEMA) and its network of state fire prevention and safety agencies have. In the late 1970s, the National Fire Data Center was established which, in conjunction with participating states, began to design a uniform fire reporting system. As this National Fire Incident Reporting System (NFIRS) developed and was experimented with by several states, participating states established a National Association of NFIRS States (see NFIRS News, 1981), and targeted January, 1982 as the start of the uniform, national implementation of this computerized fire data system. Two uniform reporting schedules have been developed: one for fire incidents and one for casualties (both fatalities and injuries to both fire fighters and civilians).

The NFIRS forms solicit a wide variety of data on each fire incident, be it structural (e.g., building), vehicular, or outdoors (e.g., brush and trash fires). Included are:

- Time, day, and date
- Time "in service" and response time
- Location, including zip code and census tract
- Name, address and telephone of occupant
- Name, address and telephone of owner
- Method of alarm (telephone, public and private alarms, etc.)
- Type of situation found (structural fire, vehicular fire, hazardous condition, service call, false alarm, etc.)
- Action taken (extinguishment, rescue, remove hazard, etc.)
- Number of alarms
- Number fire fighters and equipment (engines, etc.) at scene
- Number of injuries and fatalities
- Type of complex (single family or two-family dwelling, apartment, office, or educational complex, etc.)
- Property use (residential, storage, institutional, educational, commercial, industrial, restaurant, etc.)
- Area of fire origin (kitchen, basement, closet, etc.)
- Level of origin compared to ground level
- Termination stage (pre-smolder, pre-flame, open flame, etc.)
- Equipment used in ignition (fryer, heater, wiring, etc.)
- Form of ignition (match, lightening, etc.)
- Type of material ignited (gasoline, oily rags, etc.)
- Ignition factor (incendiary, suspicious, accidental, natural, undetermined)
- Construction type (brick, wood, etc.)
- Extent of flame and smoke damage (i.e., reach of fire/smoke)
- Existence and performance of smoke detectors and sprinklers
- Material carrying flame and avenue of spread
- Method of extinguishment (hydrant, extinguisher, etc.)
- Estimated total dollar loss
- Time from alarm to application of extinguishment agent

Next, the NFIRS Form solicits information on every victim of injury or death due to the fire incident. Included are:

- Time, day, date of incident
- Name, address and telephone of victim
- Severity (injury or death)
- Affiliation (fire fighter, emergency personnel, civilian)
- Familiarity with structure
- Location of ignition (same room, floor, building, etc.)
- Precondition (asleep, bedridden, intoxicated, too young or old to act, awake and alert, etc.)
- Conditions preventing escape (locked door, path blocked by fire, etc.)

- Activity at time of injury (escaping, rescue attempt, fire control, etc.)
- Cause of injury (trapped, overcome by smoke or flames, trampled, etc.)
- Nature of injury (Burns, asphyxia, wound, shock, etc.)

The Massachusetts Fire Incident Reporting System (MFIRS), operational since the beginning of 1982, is managed by the Massachusetts Department of Public Safety, Division of Fire Prevention. This computerized data base of fire incident and casualty reports completed by fire departments statewide for every fire is based on data collection schedules consistent with the NFIRS, with the notable addition of insurance data--name of carrier, level of insurance, and settlement (see Appendix B).

5. Data Definition

Two primary data sources were used to assemble the data files on arson and its socio-economic and demographic correlates: a) the 1980 Census of Massachusetts, Summary Count 3A and 3B (STF3A for census tracts and STF3B for zip codes) and b) the Massachusetts Fire Incident Reporting System (MFIRS) tapes for the years, 1983-1985.

The census file was used to construct a wide range of demographic and socio-economic variables for the analysis of arson rates. Census tapes provide data in the form of counts (e.g., counts of persons or of housing units) displayed in cross-tabulations for various sub-classifications. For example, a typical count might be the number of 17-year-old black males without jobs, given in a table of age by sex by employment status with a base of black residents; or the number of vacant year-round housing units with complete plumbing in a table of occupancy status by availability of plumbing for year-round units.

Although straight-forward, calculation of rates requires careful matching of counts with their appropriate bases or denominators. In addition, since the records in the census files are hierarchically structured, ceratin census tracts are split across records. For these few instances, count data were aggregated before calculation of rates and summary statistics.

Finally, while the census data are not plagued with missing data (leaving aside the issue of census undercouting), data are suppressed in tables having a base of less than 30 persons or less than 10 housing units, in order to maintain the confidentiality of the data (i.e., for disclosure control). For most of our calculations, suppression was not a problem because the level of aggregation was sufficiently large. However, for some race-specific calculations, counts for non-whites were suppressed. Nevertheless, since race is not a major focus of this research,[7] we were able to avoid much of the complexities that arise from suppression.

Tapes of fire incident data (MFIRS) for the years 1983-85 were obtained from the Office of the State Fire Marshal.[8] The MFIRS file was aggregated by zip code and by census tract in order to obtain counts of both residential and vehicular fire and arson (fires determined incendiary or suspicious in origin).

---

[7]The percentage non-white is used as a control variable in the later analysis. Because its partial effect, over and above socio-economic factors like poverty, is not particularly great, there seemed little need to use race-specific variables in the analysis.

[8]The tape containing incidents for the year 1982 was also obtained. Because this was the first year of the program, the data were suspect for many jurisdictions. For this reason, the 1982 MFIRS reports were not used.

Other types of fire, mostly small ones such as brush fires, were not within the focus of this analysis.

By far the most significant problem faced in this research involved missing information on zip code and census tract on the fire incident reports. Since these are the very codes on which the data were to be aggregated, missing identifiers would render the incident as if the fire had never occurred.

Zip codes were unavailable for only 2395 records, a bit more than 3 percent of the 60,450 structural and vehicular incident reports in the three-year file. On the face of it, this level of missing data would pose little problem, assuming that the pattern of omissions was random rather than systematic--that is, if the incomplete records were randomly distributed geographically, owing perhaps to clerical oversights.

On closer examination, however, it was determined that of the 2395 records with missing zip codes, 2035 were from the Boston Fire Department. Unlike smaller departments which may serve but one or two zip code areas, Boston spans dozens of zip codes. Conceivably, for some of the incident reports, the person filling out the MFIRS form did not expend the effort necessary to determine the precise zip code of the fire location.

Thus, while a missing identifier rate of 3% is generally acceptable, the fact that these are not nearly random in distribution would seriously bias the calculated rates. Since the aggregate measures are constructed from observed counts, the direction of error is necessarily negative.

Because of this problem, efforts were made to determine the missing zip codes. It was intended that the ADDMATCH program, distributed but not supported by the Census Bureau, would be used to search for missing zip codes and missing census tract codes (see below) from the GBF/DIME files.[9]

Unfortunately, the ADDMATCH program (written in COBOL) was particularly machine-dependent on the IBM mainframe, and was not easily transferable to the VAX. As a result, we prepared from scratch a FORTRAN program to match and fill-in missing identifiers from the GBF/DIME files for Massachusetts SMSAs.

The first version of our matching program was applied to a "street-only" version of the GBF/DIME file for Boston. Of the 2035 missing zip codes for Boston incidents, 1515 were located and inserted in the data file. A second version of the matching

---

[9]The remaining 360 missing zip codes outside of Boston would remain missing; Although it would be possible to seek out the corresponding FDID codes and then transfer these to census place codes, the yield would be hardly worth the extensive labor required.

program was applied to the standard GBF/DIME files for the Massachusetts SMSAs to fill-in missing census tract codes.

Several problems related to the form of the incident reports created difficult hurdles for matching tasks. Street names were written (and misspelled) in a wide variety of ways. Additionally, descriptive addresses, like "Behind Shaw's Market" made the matching process less successful than been had hoped.

Following the matching effort, a number of oddities appeared in the early manipulations of the data files supplemented with the newly matched tract codes. Many of the original tract codes (those not filled in by the matching program) seemed to be unusual. When analyses were stratified by city, a large number of out-of-range tract codes began to emerge. By generating frequency distributions of tract codes for separate cities, the source of the problem was uncovered. Apparently, the fire departments, perhaps unfamiliar with census tract codes, failed to fill out the data collection forms properly.

Census tract codes consist of a four digit prefix and, for a few areas, an additional two digit suffix (e.g., 01 and 02). The vast majority of the time the tract in which the fire incident occurred had only four digits, e.g., 2501. In many instances, rather than filling the four digits into the prefix field and leaving the suffix blank, the fire departments would center or right justify the number. Later the leading blanks were zero-filled upon machine inputing of data. Thus, a code properly written as 2501bb (bb denotes two blanks) could appear as 025010 or 002501. In addition, even within the same fire department, the type of mistaken-entry (centering or right justification) was not consistent.

Another type of error (occurring in almost every case from one fire department) was the omission of the first two digits of the tract code. For example, one city's tract codes range from 7301bb to 733102. Perhaps because all of their tracts begin with 73, the constant 73 was dropped. A fire in 7301 would be coded as just 1 (and this 1 could be anywhere in the six column field).

Once the problem was discovered to be pervasive throughout the data set, its solution was relatively straight-forward yet extremely tedious and time-consuming. A lengthy program was written to realign the codes within each of the nine most fire-plagued cities. (We would have liked to have used the tenth--Lynn, but the codes were almost always missing and could not be recovered from the GBF/DIME files since Lynn is not an SMSA central city.) As shown in the table below, in each of the nine major cities, at least 88 percent of the census tract codes were recovered.

City	# Fires	% Valid Tract Codes
Boston	21,096	91%
Springfield	2,568	98%
Worcester	2,481	93%
Fall River	2,326	88%
New Bedford	2,213	98%
Brockton	1,931	97%
Lawrence	1,747	94%
Lowell	1,581	95%
Cambridge	1,459	95%

Our focus in the tract level analysis was thus restricted to these nine cities. Not only do these nine represent a large portion (45%) of the fire incidents in Massachusetts, but including smaller, more rural areas in this analysis (if it were even possible), would not be terribly fruitful. That is, the statewide patterns of arson are well-represented by the zip code file. For the larger cities, in which arson is more prevalent, the tracts provide the greater level of geographic detail needed.[10]

In the end, three files were used in the analysis. The first file contained data on 592 Massachusetts zip codes of which 306 were retained in the analysis; the second on the 389 census

[10]Using the matching program, we were also able to identify 75% of the three digit block codes for the City of Boston (these identifiers are not part of the MFIRS system). We were wary of analyzing a file in which a quarter of the cases had to be ignored. Also, we are inclined to believe that the tract may be a small enough level of aggregation for our analytic purposes. Also, further disaggregation would not have come without a price. Not only do the "rates" become less reliable as one disaggregates, but the extent of suppression in census data grows.

tracts in the nine largest cities of which 289 were analyzed; and finally, the 161 census tracts in the City of Boston of which 129 were retained.

6. Methodological Approach

The overall objective of this research was to assess the demographic and socio-economic correlates of residential and vehicular arson rates in Massachusetts. More specifically, we hoped to disentangle the effects of racial composition, poverty, urbanness and other characteristics on the incidence of intentional firesetting.

As the review of recent literature indicates, this objective was far from a novel. Given the generally consistent findings from earlier studies concerning the strongest correlates of arson, moreover, we did not expect to uncover startling and unanticipated results. It was our modest expectation to improve upon earlier research in terms of methodological rigor and scientific soundness.

In contrast to the consistency of substantive results, the existing studies are quite uneven in their degree of methodological sophistication and strength. A variety of technical flaws, some minor and some significant, appear in these studies, specifically:

- failure to utilize appropriate denominators in the calculation of arson rates;
- failure to consider or adjust for irregularities in arson rates;

- failure to weight observations in cross-sectional analyses;
- failure to employ multivariate statistical methods and to correct for ill-conditioned or multi-collinear data.

Using the MFIRS data for 1983-85, described above, we hoped to resolve each of these problem areas.

6.1 Calculation of Arson Rate

Perhaps it is due to tradition emanating from the FBI's Uniform Crime Reports or perhaps it is due to the inaccessibility of alternatives bases, many investigators (e.g., Bennett et al, 1987) have used population figures to turn arson counts into arson rates. Despite custom, people do not burn, buildings and cars do. Thus, it is clearly more appropriate to cast arson rates in terms of the availability of targets, for example, the number of buildings or housing units for residential arson, and the number of cars for vehicular arson.

The choice between using the number of housing units or the number of buildings in the calculation of residential arson rates is a matter of debate. We feel that, while the stock of housing units may be preferable for calculating fire rates, the count of buildings would be more defensible for arson risk measurement.

It is reasonable to suggest that, in terms of fire incidence generally, risk tends to increase with the number of exposures (e.g., housing units). That is, the more housing units, the more potential ignition sites. Therefore, the count of housing units would be most appropriate for a general measure of fire incidence.

In terms of deliberate firesetting, on the other hand, the entire building would constitute the target at risk. While one could argue that a large, multi-unit dwelling would be the pyromaniacs dream, for most forms of arson the desirability of targeted structures may depend on other factors than just the number of units.

Following the lead of Sparks (1980), Pettiway (1985a) recommended that both measures of population and of opportunity (property) should be included in the calculation of arson rates. Specifically, he calculated the (direct) arson rate as

$$\text{Arson} = \frac{A+0.5}{PO}$$

where A is the number of arson offenses, P is the population, and O is the measure of opportunity or targets (e.g., housing units).

While it may make sense to adjust for both types of density, doing so by diving by the product of both overadjusts arson counts. This overadjustment can be illustrated using the following hypothetical data for two areas:

		Area A	Area B
a.	Arson Fires	8	12
b.	Population in 1000	20.2	30.3
c.	Buildings in 1000	1.2	1.8
d.	Arson per 1000 Pop	0.396	0.396
e.	Arson per 1000 Bldgs	6.667	6.667
f.	Arson per 1 million (Pop*Bldgs)	0.330	0.220
g.	Arson per 1000 (Pop + Bldgs)	0.374	0.374
h.	Arson per 1000 sqrt(Pop*Bldgs)	1.625	1.625

Area B is proportionally fifty percent larger than Area A in both population and stock of housing. Area B also has experienced fifty percent more arson fires. The arson rate per 1000 population (row d) and that per 1000 buildings (row e) are properly adjusted. The rate which includes both denominators multiplied together (row f) as Pettiway (1985a) did, "penalizes" the larger area by double-adjusting.

Alternative formulations encompassing both adjustment factors can be fashioned, which avoid this overadjustment problem. For example, the denominator can be defined as the sum of the two "stock" measures (row g) or even the product of the square root of the two factors (row h). While these two alternative approaches work just fine in our hypothetical case, with real data, they would adjust more for the characteristic having the larger standard deviation (i.e., scale).

To correct for this deficiency, one could, of course, adjust the count data for sum of the two factors or the square root of the product of two factors after the factors are transformed into standard score form. More generally, count data could be adjusted for n factors by adding them of by dividing by the n-th root of the product of the factors all in standard form.

In the case at hand, we choose not to invoke the very elaborate steps needed to adopt a multiple adjustment approach. As shown in the correlations of Table 1 and 2, all the formulations except that suffering from the double adjustment problem (RARATE3 and CARATE3) are highly inter-correlated across all three data sets.[11] We shall settle, therefore, on the more substantively meaningful singly-adjusted rate, Residential Arson per 1000 Residential Buildings (RARATE1) and Car Arson per 1000 Automobiles (CARATE1). Furthermore, RARATE1 on average correlates .95 with the other alternatives except RARATE3, and CARATE1 averages a .96 correlation with the other measures with the exception of CARATE3. Henceforth, these rates will be labeled simply RARATE and CARATE, respectively.

6.2 Transformation for Skewness

Early examinations of the arson rate distributions uncovered certain outliers apparently resulting from recording peculiarities. For example, we found one zip code (01901--Lynn

---

[11]Correlations were computed only for observations having non-zero values for arson counts. This was done for the sake of comparability with later analyses of log transformed rates, for which zero values are undefined.

Table 1: Comparison of Residential Arson Rate Measures

RARATE1 = Residential arson per 1000 residential bldgs
RARATE2 = Residential arson per 1000 population
RARATE3 = Residential arson per 1000 residential bldgs and per 1000 population
RARATE4 = Residential arson per square root 1000 residential bldgs and per square root 1000 population
RARATE5 = Residential arson per 1000 year-round units

a. Zip Codes (N=289)

Variable	Mean	SD	RARATE1	RARATE2	RARATE3	RARATE4	RARATE5
RARATE1	4.168	9.460	1.000	0.943	0.709	0.989	0.890
RARATE2	0.792	1.261	0.943	1.000	0.622	0.981	0.976
RARATE3	1.041	5.696	0.709	0.622	1.000	0.674	0.528
RARATE4	1.772	3.378	0.989	0.981	0.674	1.000	0.940
RARATE5	1.928	2.491	0.890	0.976	0.528	0.940	1.000

b. Urban Census Tracts (N=306)

Variable	Mean	SD	RARATE1	RARATE2	RARATE3	RARATE4	RARATE5
RARATE1	13.483	16.618	1.000	0.955	0.885	0.989	0.925
RARATE2	1.944	2.131	0.955	1.000	0.896	0.988	0.966
RARATE3	5.186	8.773	0.885	0.896	1.000	0.900	0.838
RARATE4	5.069	5.859	0.989	0.988	0.900	1.000	0.955
RARATE5	4.616	4.918	0.925	0.966	0.838	0.955	1.000

c. Boston Census Tracts (N=129)

Variable	Mean	SD	RARATE1	RARATE2	RARATE3	RARATE4	RARATE5
RARATE1	16.786	17.370	1.000	0.938	0.836	0.984	0.927
RARATE2	2.313	2.221	0.938	1.000	0.892	0.984	0.973
RARATE3	6.650	10.125	0.836	0.892	1.000	0.877	0.842
RARATE4	6.175	6.073	0.984	0.984	0.877	1.000	0.966
RARATE5	5.522	5.543	0.927	0.973	0.842	0.966	1.000

Table 2: Comparison of Automobile Arson Rate Measures

CARATE1 = Automobile arson per 1000 automobiles
CARATE2 = Automobile arson per 1000 population
CARATE3 = Automobile arson per 1000 automobiles and per 1000 population
CARATE4 = Automobile arson per square root 1000 automobiles and per square root 1000 population

a. Zip Codes (N=289)

Variable	Mean	SD	CARATE1	CARATE2	CARATE3	CARATE4
CARATE1	9.669	28.075	1.000	0.973	0.552	0.994
CARATE2	3.026	6.455	0.973	1.000	0.542	0.992
CARATE3	1.060	3.465	0.552	0.542	1.000	0.548
CARATE4	5.304	13.320	0.994	0.992	0.548	1.000

b. Census Tracts (N=306)

Variable	Mean	SD	CARATE1	CARATE2	CARATE3	CARATE4
CARATE1	47.720	88.568	1.000	0.924	0.935	0.981
CARATE2	11.065	16.137	0.924	1.000	0.878	0.980
CARATE3	20.702	50.890	0.935	0.878	1.000	0.925
CARATE4	22.548	36.599	0.981	0.980	0.925	1.000

c. Boston Census Tracts (N=129)

Variable	Mean	SD	CARATE1	CARATE2	CARATE3	CARATE4
CARATE1	96.869	119.138	1.000	0.905	0.928	0.977
CARATE2	20.883	20.611	0.905	1.000	0.872	0.975
CARATE3	43.518	72.218	0.928	0.872	1.000	0.923
CARATE4	44.253	47.825	0.977	0.975	0.923	1.000

Central) which had a residential arson rate of over 200 per 1000 residences, and five zip codes (including 01901) which had excessive automobile arson rates (over 100 per 1000 cars). For both these variables, outliers were trimmed to the 99th percentiles.

Additionally, we observed some inflated and unreliable rates calculated for a number smaller areas. Ultimately, these unreliable observations would have insignificant influence on the analytic results because of the population weighting scheme used (see discussion below). That is, these small areas would have appropriately small weights in the data analysis. Still, we preferred to purge the data sets of these cases. Specifically, only those areas having at least 200 persons and 200 cars were retained for analysis.

Even after trimming the outliers and after eliminating rates considered unreliable due to small bases, a further data problem exists--skewness. As often occurs when examining rates of rare events in heterogenous areas, most observations are low in incidence while a few high-incidence areas severely skew the distribution. Such skewness, if uncorrected, can dramatically affect the analysis of the rates. In a regression framework (to be used here), for example, skewness, particularly in the dependent variable, not only affects the validity of significance tests based on normality assumptions, but, more importantly, makes the regression coefficients unduly sensitive to small changes at the tail of the distribution. That is, the contribution of an observation varies directly in proportion to the distance of an observation from the group centroid (i.e., the means). Therefore, the tail of a skewed distribution carries excessive weight in the analysis.

Fortunately, skewness is often easy to correct by transforming by logarithms. As shown in Figs 1-3, a log transformation not only eliminates skewness in all the arson rates, but even causes the distributions to approach normality in shape.

The log transformation does have its drawbacks. First, although minor, the transformation alters the interpretation of regression results. Specifically, changes in the independent variables relate to percentage changes in rates, not natural changes. More importantly, the log transformation is problematic for areas having zero rates, since the log of zero is undefined.

Some researchers add a small number (say .01) before taking logs to avoid this problem. Pettiway (1985a) added 0.5 to his arson counts before taking logs. Of course, there are no guidelines to determine exactly what constant to add, and, furthermore, the results can be quite sensitive to small changes in this constant. Also, if there are a number of such

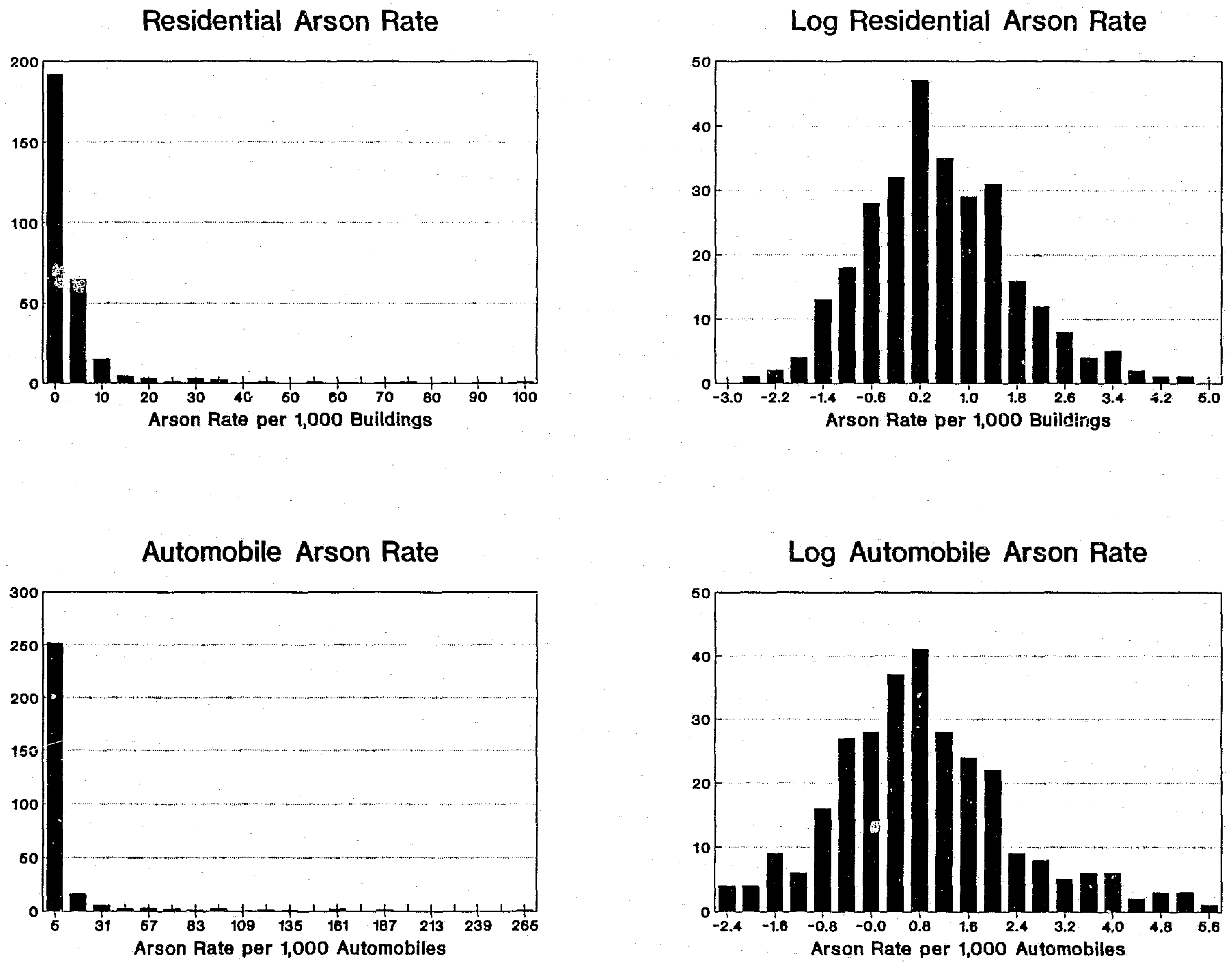


Figure 1: Rates for Massachusetts Zip Codes

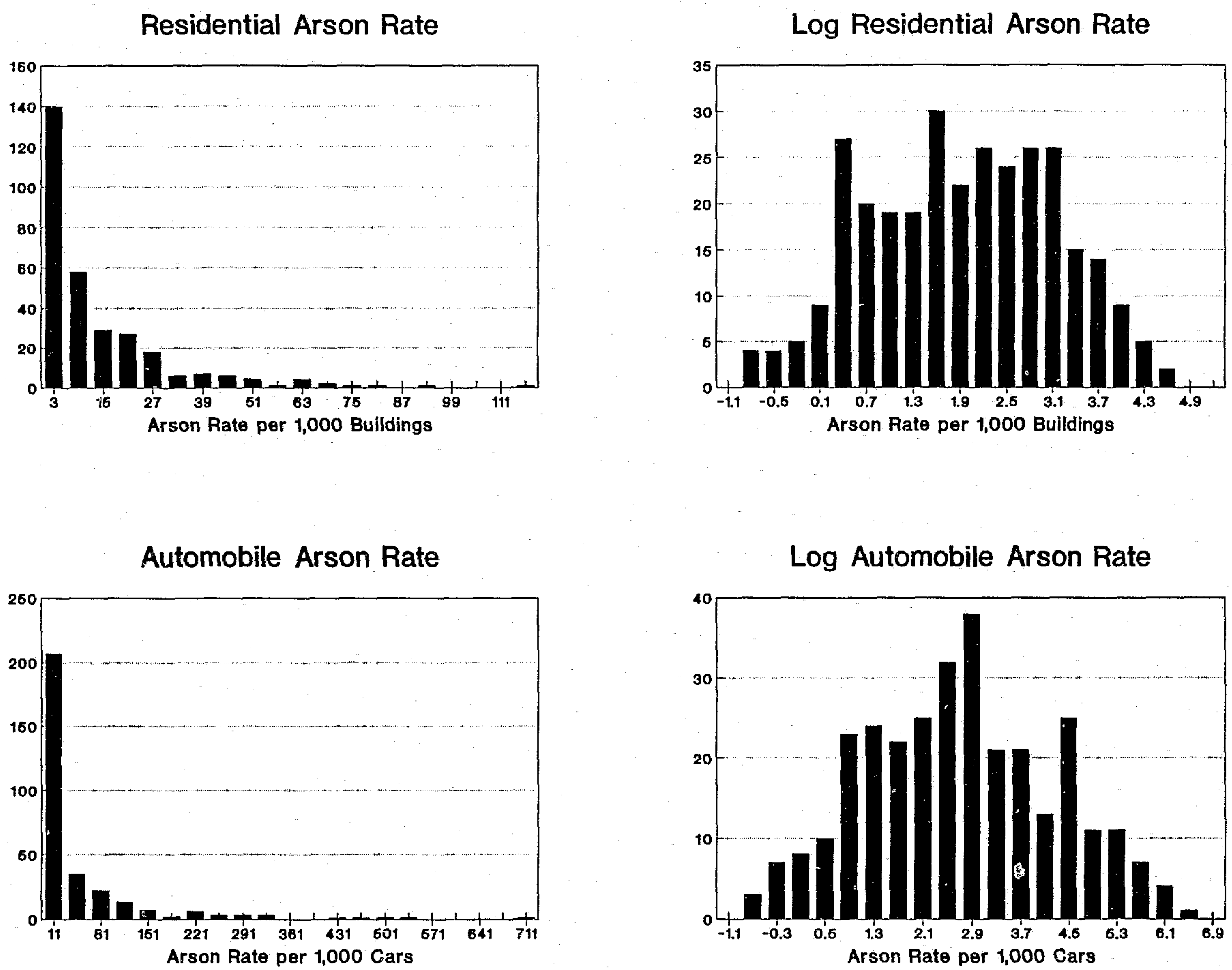


Figure 2: Rates for Census Tracts

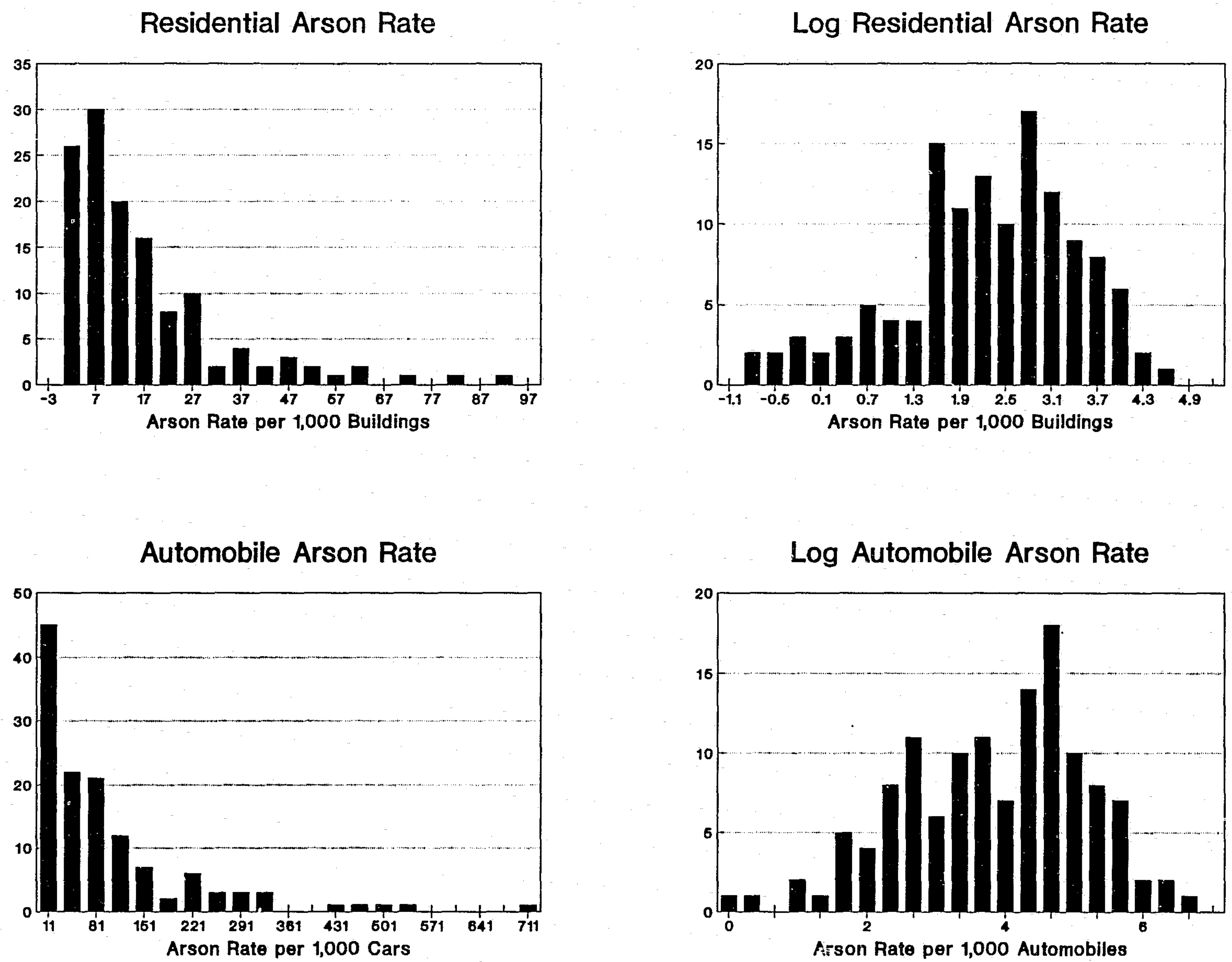


Figure 3: Rates for Boston Tracts

Table 3: Descriptive Statistics for Arson Measures

a. Zip Codes

Variable	N	Min	Max	Sum	Mean	SD	Skew
NPERSONS	592	0	65158.00	5737037.00	9690.94	11963.18	1.75
NUMCARS	582	0	36615.00	2816733.00	4839.75	5680.45	1.79
RTN	585	0	505	26088.00	44.59	68.29	2.71
RAN	584	0	180	3707.00	6.35	15.27	5.07
CTN	585	0	1548	42951.00	73.42	166.25	4.74
CAN	584	0	1279	18500.00	31.68	116.23	6.91
RARATE	289	.08	99.48	1204.49	4.17	9.46	6.21
LRARATE	289	-2.52	4.60	153.39	.53	1.23	.47
CARATE	289	.09	270.44	2794.36	9.67	28.08	5.70
LCARATE	289	-2.44	5.60	248.93	.86	1.50	.58

b. Boston Census Tracts

Variable	N	Min	Max	Sum	Mean	SD	Skew
NPERSONS	161	300.00	11072.00	561745.00	3489.10	2049.12	.99
NUMCARS	161	0	3749.50	162259.00	1007.82	823.52	1.33
RTN	160	2	93	3596.00	22.48	16.13	1.31
RAN	161	0	34	1029.00	6.39	6.38	1.63
CTN	161	2	270	13515.00	83.94	57.55	1.02
CAN	161	1	248	9482.00	58.89	50.05	1.11
RARATE	129	.48	92.96	2165.43	16.79	17.37	1.93
LRARATE	129	-.73	4.53	292.51	2.27	1.18	-.53
CARATE	129	1.07	721.15	12496.09	96.87	119.14	2.42
LCARATE	129	.06	6.58	497.30	3.86	1.33	-.39

b. Census Tracts

Variable	N	Min	Max	Sum	Mean	SD	Skew
NPERSONS	389	0	11072.00	1411670.00	3628.97	1822.21	.78
NUMCARS	388	0	4609.75	509052.75	1311.99	887.45	.89
RTN	387	0	98	8651.00	22.35	15.81	1.50
RAN	388	0	34	2040.00	5.26	5.90	1.92
CTN	389	0	270	21424.00	55.07	48.36	1.75
CAN	388	0	248	12208.00	31.46	41.13	2.09
RARATE	306	.43	114.50	4125.86	13.48	16.62	2.50
LRARATE	306	-.84	4.74	593.60	1.94	1.21	-.07
CARATE	306	.44	721.15	14602.22	47.72	88.57	3.75
LCARATE	306	-.81	6.58	830.67	2.71	1.56	.12

NPERSONS: Number persons
NUMCARS: Number of automobiles (estimated)
RTN: Residential-building fires
RAN: Residential-building arson fires
CTN: Automobile fires
CAN: Automobile arson fires
RARATE: Residential arson per 1000 residential building
LRARATE: Log residential arson per 1000 residential bldgs
CARATE: Automobile arson per 1000 automobiles
LCARATE: Log automobile arson per 1000 automobiles

Table 4a: Descriptive Statistics for Census Variables for Zip Codes

Variable	N	Min	Max	Mean	SD	Skew
UR	306	2.09	2.32	7.33	3.51	1.17
ZUR	306	0.74	.15	1.88	0.48	-0.13
LFP	306	31.70	7.94	58.97	7.89	-0.69
ZLFP	306	3.46	.37	4.07	0.15	-1.35
PTLPOVL	306	1.42	6.36	19.07	10.64	1.08
ZPTLPOVL	306	0.35	.20	2.78	0.62	-0.71
MDNINC	306	4792	2147	12543.0	3762.54	0.34
ZMDNINC	306	8.47	1.09	9.39	0.32	-0.49
PCTFAM1P	306	0.00	91.94	34.30	17.74	0.62
ZPTFAM1P	301	1.05	4.52	3.42	0.55	-0.62
PGT1PPRM	306	0.00	2.96	4.31	3.91	2.61
ZPGT1PRM	297	-1.60	.33	1.19	0.79	-0.33
DENSITY1	306	2.48	1.96	6.27	2.31	1.77
ZDENS1	306	0.91	.94	1.78	0.34	0.28
DENSITY2	306	1.00	.00	2.65	1.11	0.90
ZDENS2	306	0.00	.79	0.89	0.41	0.01
PTEN75	306	17.48	8.65	51.05	13.01	0.53
ZPTEN75	306	2.86	.47	3.90	0.26	-0.28
AVEOCC	306	3.01	1.05	10.05	2.96	-1.50
ZAVEOCC	306	1.10	.89	2.26	0.34	-0.93
PCTVAC	306	0.00	5.34	7.98	6.19	2.29
ZPCTVAC	305	-0.46	.92	1.81	0.77	-0.34
PCTVBLDG	306	0.00	5.62	6.97	5.76	2.31
ZPTVBLDG	305	-0.72	.92	1.62	0.85	-0.36
PCTRENTU	306	3.12	10.00	67.64	20.51	-0.81
ZPCTRENT	306	1.14	.61	4.14	0.46	-0.03
PCNTNPLG	306	0.02	2.34	2.72	3.12	3.22
ZPCTNPLG	306	-3.91	.27	0.34	1.49	-1.54
PCNTNBAT	302	0.13	2.44	4.32	4.27	2.55
ZPCTNBAT	302	-2.04	.38	0.97	1.17	-1.07
PCNTNKIT	306	0.05	2.71	2.27	2.90	3.37
ZPCTNKIT	306	-3.10	.08	0.16	1.33	-0.91
PCTBL	306	0.02	9.24	14.71	25.79	2.10
ZPCTBL	306	-4.20	.58	0.74	2.57	-0.58
PCT1019M	306	1.34	2.29	8.29	2.85	0.81
ZPT1019M	306	0.29	.15	2.05	0.39	-1.28

UR	Unemployment rate
LFP	Labor force participation
PTLPOVL	Pct persons below poverty
MDNINC	Median income
PCTFAM1P	Pct of families with 1 parent
PGT1PPRM	Pct units with > 1 person per room
DENSITY1	Persons per building
DENSITY2	Units per building
PTEN75	Pct occ tenure since 1975
AVEOCC	Average occupancy tenure
PCTVAC	Pct vacant housing units
PCTVBLDG	Pct vacant buildings
PCTRENTU	Pct occupied units rented
PCNTNBAT	Pct housing units without compl bath
PCNTNPLG	Pct housing units without compl plumbing
PCNTNKIT	Pct housing units without compl kitchen
PCTBL	Pct black
PCT1019M	Pct age 10-19 male

Table 4b: Descriptive Statistics for Census Variables for Tracts

	N	Min.	Max.	Mean	SD	Skew
UR	289	0.00	19.13	5.17	2.41	2.00
ZUR	287	0.08	2.95	1.56	0.42	-0.02
LFP	289	0.40	81.13	63.43	6.12	-0.82
ZLFP	289	3.71	4.40	4.14	0.10	-1.23
PTLPOVL	289	0.44	44.98	9.38	6.80	1.80
ZPTLPOVL	289	-0.81	3.81	2.02	0.67	-0.02
MDNINC	289	6650	47542	18799.7	6049.47	1.02
ZMDNINC	289	8.80	10.77	9.79	0.32	-0.14
PCTFAM1P	289	0.00	69.56	19.05	11.12	1.70
ZPTFAM1P	287	1.24	4.24	2.82	0.51	0.20
PGT1PPRM	289	0.00	17.52	2.37	1.99	3.77
ZPGT1PRM	284	-2.08	2.86	0.64	0.73	-0.61
DENSITY1	289	1.82	14.22	4.15	1.63	2.66
ZDENS1	289	0.60	2.65	1.37	0.31	1.10
DENSITY2	289	1.01	6.00	1.60	0.86	2.83
ZDENS2	289	0.01	1.79	0.38	0.38	1.70
PTEN75	289	26.28	84.29	44.86	9.38	1.21
ZPTEN75	289	3.27	4.43	3.78	0.20	0.38
AVEOCC	289	3.78	15.68	11.11	2.08	-0.94
ZAVEOCC	289	1.33	2.75	2.39	0.22	-1.90
PCTVAC	289	0.89	29.00	5.39	4.55	2.02
ZPCTVAC	289	-0.11	3.37	1.41	0.72	0.39
PCTVBLDG	289	0.52	28.49	4.58	4.34	2.25
ZPTVBLDG	289	-0.65	3.35	1.20	0.76	0.50
PCTRENTU	289	5.04	100.00	37.33	22.27	0.89
ZPCTRENT	289	1.62	4.61	3.44	0.63	-0.21
PCNTNPLG	289	0.02	14.41	1.38	1.69	3.79
ZPCTNPLG	289	-3.91	2.67	-0.39	1.50	-1.12
PCNTNBAT	289	0.13	25.58	3.39	3.20	3.15
ZPCTNBAT	289	-2.04	3.24	0.72	1.23	-1.12
PCNTNKIT	289	0.05	33.54	1.37	2.40	9.47
ZPCTNKIT	289	-3.10	3.51	-0.20	1.04	-0.53
PCTBL	289	0.02	90.52	3.18	10.13	6.07
ZPCTBL	289	-4.20	4.51	-0.46	1.81	-0.16
PCT1019M	289	0.00	15.23	8.94	1.82	-0.70
ZPT1019M	288	0.76	2.72	2.17	0.23	-1.98

UR	Unemployment rate
LFP	Labor force participation
PTLPOVL	Pct persons below poverty
MDNINC	Median income
PCTFAM1P	Pct of families with 1 parent
PGT1PPRM	Pct units with > 1 person per room
DENSITY1	Persons per building
DENSITY2	Units per building
PTEN75	Pct occ tenure since 1975
AVEOCC	Average occupancy tenure
PCTVAC	Pct vacant housing units
PCTVBLDG	Pct vacant buildings
PCTRENTU	Pct occupied units rented
PCNTNBAT	Pct housing units without compl bath
PCNTNPLG	Pct housing units without compl plumbing
PCNTNKIT	Pct housing units without compl kitchen
PCTBL	Pct black
PCT1019M	Pct age 10-19 male

Table 4c: Descriptive Statistics for Census Variables for Boston Census Tracts

Variable	N	Min	Max	Mean	SD	Skew
UR	129	2.24	1 .27	7.00	3.20	0.86
ZUR	129	0.81	.79	1.84	0.46	-0.11
LFP	129	35.17	7 .94	59.28	7.41	-0.25
ZLFP	129	3.56	.37	4.07	0.13	-0.85
PTLPOVL	129	2.89	49.29	20.78	10.55	0.48
ZPTLPOVL	129	1.06	3.90	2.88	0.60	-0.78
MDNINC	129	4814	22307	12554.4	3711.12	0.46
ZMDNINC	129	8.48	10.01	9.39	0.31	-0.40
PCTFAM1P	129	0.00	91.94	37.27	20.43	0.30
ZPTFAM1P	124	1.70	4.52	3.51	0.58	-0.70
PGT1PPRM	129	0.00	27.96	5.34	4.83	2.25
ZPGT1PRM	126	-0.29	3.33	1.40	0.77	0.15
DENSITY1	129	3.39	14.69	6.78	2.12	1.05
ZDENS1	129	1.22	2.69	1.87	0.30	0.12
DENSITY2	129	1.25	5.83	2.98	1.16	0.86
ZDENS2	129	0.22	1.76	1.02	0.38	0.90
PTEN75	129	23.73	87.65	52.46	13.59	0.72
ZPTEN75	129	3.17	4.47	3.93	0.25	0.08
AVEOCC	129	3.01	18.05	9.53	3.14	-0.04
ZAVEOCC	129	1.10	2.89	2.19	0.37	-0.78
PCTVAC	129	1.20	50.34	10.31	7.42	2.12
ZPCTVAC	129	0.18	3.92	2.09	0.74	-0.48
PCTVBLDG	129	0.73	50.62	9.04	6.79	2.30
ZPTVBLDG	129	-0.32	3.92	1.91	0.84	-0.72
PCTRENTU	129	23.36	100.00	72.45	16.79	-0.53
ZPCTRENT	129	3.15	4.61	4.25	0.27	-1.41
PCNTNPLG	129	0.02	17.78	2.80	3.01	2.50
ZPCTNPLG	129	-3.91	2.88	0.48	1.30	-1.59
PCNTNBAT	126	0.13	25.42	4.59	4.79	2.37
ZPCTNBAT	126	-2.04	3.24	1.04	1.11	-0.86
PCNTNKIT	129	0.05	12.72	2.19	2.29	2.30
ZPCTNKIT	129	-3.10	2.54	0.23	1.26	-1.14
PCTBL	129	0.02	97.24	25.94	34.02	1.06
ZPCTBL	129	-4.20	4.58	1.04	3.14	-0.62
PCT1019M	129	1.34	16.62	8.37	3.01	0.04
ZPT1019M	129	0.29	2.81	2.04	0.45	-1.44

UR	Unemployment rate
LFP	Labor force participation
PTLPOVL	Pct persons below poverty
MDNINC	Median income
PCTFAM1P	Pct of families with 1 parent
PGT1PPRM	Pct units with > 1 person per room
DENSITY1	Persons per building
DENSITY2	Units per building
PTEN75	Pct occ tenure since 1975
AVEOCC	Average occupancy tenure
PCTVAC	Pct vacant housing units
PCTVBLDG	Pct vacant buildings
PCTRENTU	Pct occupied units rented
PCNTNBAT	Pct housing units without compl bath
PCNTNPLG	Pct housing units without compl plumbing
PCNTNKIT	Pct housing units without compl kitchen
PCTBL	Pct black
PCT1019M	Pct age 10-19 male

vast majority of the variables, especially PGT1PPRM, PCTVAC, both density measures DENSITY1 and DENSITY2, the three facilities measures PCNTNPLG, PCNTNBAT, and PCNTNKIT, and PCTBL. Some of the variables (the arson variables in particular) are skewed by the "rare event" phenomenon described above. Other variables, those that are percentages, tend to be skewed because of constraints of bounding at zero and one. Apparently, only the percentage measures which are moderate in central tendency are not subject to extensive skewness.

For the variables that are skewed, log transformations will be used in the analysis. As expected, the skewness is less severe as the observation set becomes more homogeneous (as in the Boston data set). For the sake of comparability, however, the same pattern of transformation will be applied to all three data sets.

LRARATE and LCARATE will represent the logged arson rates, while a "Z" prefix will be used to indicate a log transformed census variable. Based on the results of Tables 4a-4c, all the census variables except LFP (labor force participation), AVEOCC (average occupancy tenure), PCTRENTU (percent rental units), and PCT1019M (Percent age 10-19 male) will be log-transformed.

6.3 Weighting Scheme

Statistical methods designed for analyzing variation across micro-units, such as people, cannot necessarily be transferred without modification to aggregates. This is particularly true of analyses of inter-state, inter-city, or inter-tract cross-sectional variation. The more variable are the units in size, the more problematic.

First, geographic units, such as states, zip codes or even census tracts, are not the same size in terms of population. Application of statistical techniques without weighting tacitly and unjustifiably gives equal contribution to all observations regardless of size. Many researchers, for example, focus on inter-city variation by applying statistical methods to city-level aggregate data. In the process, data for a medium-sized city like Cleveland would count equally to those for large-sized city like Los Angeles. It makes sense, therefore, to weight observations proportional to their size in order to achieve results reflective of the whole.

Second, data from smaller units are more variable or less reliable that those from larger units. To avoid heteroscedasticity common with aggregate cross-sections, observations should be weighted inversely proportional to their variance. This is essentially equivalent to a population-based weighting scheme, since the variance is inversely proportional to the number of units that make up the aggregate. That is, for example, the homicide rate, say, for Cleveland is less stable or reliable than that for Los Angeles.

In the analyses below, all observations are weighted proportional to their population counts. Shown in Table 5, weights (labeled WTPERS) are calculated as the resident population of a unit (zip code or census tract) divided by the average population within the respective data set.[13] By this approach, the average weight is set to unity, and, therefore, the analytic sample sizes are maintained at their proper levels.

The result of this weighting approach is that places contribute to the analysis proportional to their populations. For example, the well-populated zip code having a weight of 3.78 counts in the analysis as if it were almost four distinct data points, whereas the zip code with a .03 weight contributes very little to the overall results. In addition to its intuitive appeal, observations are weighted in such a way as to discount unreliable rates from smaller units.

Finally, this approach has the additional advantage of minimizing the problem of missing and suppressed data. Undefined log-transformation for zero arson rates and Census Bureau

---

[13]In addition to being a meaningful weighting approach, this scheme avoids confounding directly an adjustment factor (e.g., number of buildings or number of cars) with the weighting factor.

Table 5: Population Weights

a. Zip Codes

	Mean	Std Dev	Minimum	Maximum	N
Unweighted:					
NPERSONS	17247.56	12940.42	464.00	65158.00	289
WTPERS	1.00	.75	.03	3.78	289
Weighted:					
NPERSONS	26922.85	14578.21	464.00	65158.00	289
WTPERS	1.56	.85	.03	3.78	289

b. Urban Census Tracts

	Mean	Std Dev	Minimum	Maximum	N
Unweighted:					
NPERSONS	3863.15	1765.92	739.00	11072.00	306
WTPERS	1.00	.46	.19	2.87	306
Weighted:					
NPERSONS	4667.75	1966.88	739.00	11072.00	306
WTPERS	1.21	.51	.19	2.87	306

c. Boston Census Tracts

	Mean	Std Dev	Minimum	Maximum	N
Unweighted:					
NPERSONS	3797.93	2002.27	739.00	11072.00	129
WTPERS	1.00	.53	.19	2.92	129
Weighted:					
NPERSONS	4845.35	2304.24	739.00	11072.00	129
WTPERS	1.28	.61	.19	2.92	129

NPERSONS = Resident Population
WTPERS = Population Weight

suppression of potentially revealing information are problematic only for smaller or sparsely-populated areas. Had these observations not been eliminated because of data limitations, they would be greatly minimized in their importance when weighted down by WTPERS.

6.4 Multicollinearity

Table 6 displays correlations of the census variables with the arson rates.[14] Most of the socio-economic variables posses moderate to strong associations with the arson measures, particularly ZPTLPOVL, ZMDMINC, ZDENS1, ZDENS2, ZPCTVAC, and PCTRENTU. The two demographic variables, ZPCTBL and PCT1019M, are weaker or inconsistent in their association with the arson rates.

It would seem to be an obvious point that the analysis should consider arson correlates in combination, rather than individually. That is, it is likely that a good deal of overlap exists among the variables in their explanatory potentials. Not all investigators have accommodated this consideration in their analyses, but we assume that the failure of many of them to do so is a function of era of their work.

---

[14]As is discussed below, a few of the variables (LFP, AVEOCC, and ZPTVBLDG) had to be eliminated for statistical reasons.

Table 6: Correlations of Census Variables and Arson Rates

	Zip Codes		Tracts		Boston	
	LRARATE	LCARATE	LRARATE	LCARATE	LRARATE	LCARATE
ZUR	0.548	0.495	0.384	0.382	0.380	0.615
LFP	-0.525	-0.443	-0.389	-0.304	-0.291	-0.460
ZPTLPOVL	0.787	0.649	0.691	0.448	0.708	0.351
ZMDNINC	-0.762	-0.631	-0.621	-0.376	-0.640	-0.432
ZPTFAM1P	0.748	0.625	0.587	0.422	0.646	0.526
ZPGT1PRM	0.603	0.588	0.435	0.524	0.527	0.498
ZDENS1	0.674	0.621	0.639	0.301	0.656	0.013
ZDENS2	0.708	0.624	0.606	0.257	0.554	-0.130
ZPTEN75	0.534	0.315	0.463	0.082	0.344	-0.336
AVEOCC	-0.446	-0.241	-0.500	-0.128	-0.467	0.214
ZPCTVAC	0.691	0.520	0.652	0.523	0.646	0.519
ZPTVBLDG	0.664	0.501	0.684	0.462	0.709	0.444
PCTRENTU	0.715	0.630	0.630	0.332	0.602	0.038
ZPCTNPLG	0.545	0.449	0.402	0.255	0.294	0.107
ZPCTNBAT	0.386	0.321	0.386	0.223	0.332	0.199
ZPCTNKIT	0.587	0.438	0.375	0.189	0.328	0.112
ZPCTBL	0.478	0.391	0.405	0.194	0.504	0.199
PCT1019M	-0.143	-0.109	0.246	0.230	0.429	0.536

ZUR	Log unemployment rate
LFP	Labor force participation
ZPTLPOVL	Log pct persons below poverty
ZMDNINC	Log median income
ZPTFAM1P	Log pct families with 1 parent
ZPGT1PRM	Log pct units with > 1 person per room
ZDENS1	Log persons per building
ZDENS2	Log units per building
ZPTEN75	Log pct occ tenure since 1975
AVEOCC	Average occupancy tenure
ZPCTVAC	Log pct vacant housing units
ZPTVBLDG	Log pct vacant buildings
PCTRENTU	Log pct occupied units rented
ZPCTNPLG	Log pct housing units without compl bath
ZPCTNBAT	Log pct housing units without compl plumbing
ZPCTNKIT	Log pct housing units without compl kitchen
ZPCTBL	Log pct black
PCT1019M	Pct age 10-19 male

In all the analyses to follow we are interested in the partial effects of various components of the socio-economic and demographic profile. This will be accomplished through a multiple regression approach.

Unfortunately, multicollinearity tends to be substantial with highly aggregated data such as the census data used here. This was a problem that consistently plagued analyses by Karter and his associates and which caused Gunther (1981) to consider only two factors. The correlation matrices shown in Tables 7a through 7c reveal a large number of the intercorrelations which exceed .70 (a customary benchmark or rule of thumb for assessing the presence of multi-collinearity; a few of the association even exceed a prohibitive level of .90, as in the relationship between the extent of rental units (PCTRENTU) and residential density (ZDENS2), between population density (ZDENS1) and housing density (ZDENS2), and among zip codes between persons below the poverty level (ZPTLPOVL) and both median income (ZMDNINC) and the prevalence of families with one parent (ZPTFAM1P). Clearly multicollinearity is a problem which must be resolved before estimation can be done in an efficient manner.

The multivariate focus is designed specifically to identify overlap between such factors as race, poverty and urbanness. However, some researchers have naively forged ahead with regression, asking that the algorithm accomplish something

Table 7a: Correlation Matrix of Socio-Economic and Demographic Variables
Zip Codes (N = 288)

	UR	POVL	MDINC	FAM1P	DENS1	DENS2	TEN75	VAC	RENTU	GT1PRM	NBAT	NPLG	NKIT	PCTBL	1019M
ZUR	1.00	0.67	-0.76	0.62	0.28	0.35	0.18	0.61	0.41	0.59	0.40	0.48	0.50	0.31	-0.07
ZPTLPOVL	0.67	1.00	-0.92	0.91	0.72	0.80	0.60	0.74	0.85	0.68	0.41	0.64	0.65	0.54	-0.32
ZMDNINC	-0.76	-0.92	1.00	-0.85	-0.65	-0.76	-0.52	-0.75	-0.82	-0.70	-0.47	-0.67	-0.66	-0.41	0.32
ZPTFAM1P	0.62	0.91	-0.85	1.00	0.67	0.76	0.56	0.73	0.83	0.62	0.38	0.58	0.61	0.61	-0.30
ZDENS1	0.28	0.72	-0.65	0.67	1.00	0.94	0.66	0.46	0.88	0.52	0.37	0.58	0.52	0.52	-0.20
ZDENS2	0.35	0.80	-0.76	0.76	0.94	1.00	0.72	0.59	0.96	0.51	0.38	0.62	0.57	0.49	-0.42
ZPTEN75	0.18	0.60	-0.52	0.56	0.66	0.72	1.00	0.57	0.72	0.33	0.35	0.51	0.44	0.47	-0.30
ZPCTVAC	0.61	0.74	-0.75	0.73	0.46	0.59	0.57	1.00	0.62	0.51	0.39	0.50	0.57	0.40	-0.17
PCTRENTU	0.41	0.85	-0.82	0.83	0.88	0.96	0.72	0.62	1.00	0.53	0.41	0.67	0.61	0.50	-0.47
ZPGT1PRM	0.59	0.68	-0.70	0.62	0.52	0.51	0.33	0.51	0.53	1.00	0.34	0.48	0.51	0.41	-0.04
ZPCTNBAT	0.40	0.41	-0.47	0.38	0.37	0.38	0.35	0.39	0.41	0.34	1.00	0.76	0.52	0.08	-0.08
ZPCTNPLG	0.48	0.64	-0.67	0.58	0.58	0.62	0.51	0.50	0.67	0.48	0.76	1.00	0.64	0.23	-0.24
ZPCTNKIT	0.50	0.65	-0.66	0.61	0.52	0.57	0.44	0.57	0.61	0.51	0.52	0.64	1.00	0.32	-0.21
ZPCTBL	0.31	0.54	-0.41	0.61	0.52	0.49	0.47	0.40	0.50	0.41	0.08	0.23	0.32	1.00	-0.06
PCT1019M	-0.07	-0.32	0.32	-0.30	-0.20	-0.42	-0.30	-0.17	-0.47	-0.04	-0.08	-0.24	-0.21	-0.06	1.00

ZUR	Log unemployment rate
LFP	Labor force participation
ZPTLPOVL	Log pct persons below poverty
ZMDNINC	Log median income
ZPTFAM1P	Log pct families with 1 parent
ZPGT1PRM	Log pct units with > 1 person per room
ZDENS1	Log persons per building
ZDENS2	Log units per building
ZPTEN75	Log pct occ tenure since 1975
AVEOCC	Average occupancy tenure
ZPCTVAC	Log pct vacant housing units
ZPTVBLDG	Log pct vacant buildings
PCTRENTU	Log pct occupied units rented
ZPCTNPLG	Log pct housing units without compl bath
ZPCTNBAT	Log pct housing units without compl plumbing
ZPCTNKIT	Log pct housing units without compl kitchen
ZPCTBL	Log pct black
PCT1019M	Pct age 10-19 male

Table 7b: Correlation Matrix of Socio-Economic and Demographic Variables
Census Tracts (N = 293)

	UR	POVL	MDINC	FAM1P	DENS1	DENS2	TEN75	VAC	RENTU	GT1PRM	NBAT	NPLG	NKIT	PCTBL	1019M
ZUR	1.00	0.54	-0.59	0.54	0.16	0.11	0.03	0.38	0.24	0.43	0.26	0.31	0.20	0.20	0.26
ZPTLPOVL	0.54	1.00	-0.81	0.73	0.67	0.66	0.59	0.68	0.77	0.48	0.40	0.51	0.43	0.44	0.23
ZMDNINC	-0.59	-0.81	1.00	-0.64	-0.58	-0.60	-0.42	-0.53	-0.73	-0.39	-0.41	-0.49	-0.41	-0.23	-0.19
ZPTFAM1P	0.54	0.73	-0.64	1.00	0.40	0.41	0.39	0.61	0.55	0.36	0.26	0.36	0.33	0.54	0.29
ZDENS1	0.16	0.67	-0.58	0.40	1.00	0.91	0.71	0.50	0.86	0.33	0.42	0.51	0.39	0.24	0.16
ZDENS2	0.11	0.66	-0.60	0.41	0.91	1.00	0.75	0.56	0.92	0.17	0.48	0.57	0.42	0.18	-0.13
ZPTEN75	0.03	0.59	-0.42	0.39	0.71	0.75	1.00	0.43	0.75	0.11	0.34	0.43	0.35	0.37	-0.05
ZPCTVAC	0.38	0.68	-0.53	0.61	0.50	0.56	0.43	1.00	0.59	0.38	0.48	0.52	0.48	0.29	0.14
PCTRENTU	0.24	0.77	-0.73	0.55	0.86	0.92	0.75	0.59	1.00	0.27	0.51	0.63	0.45	0.25	-0.04
ZPGT1PRM	0.43	0.48	-0.39	0.36	0.33	0.17	0.11	0.38	0.27	1.00	0.18	0.21	0.20	0.36	0.36
ZPCTNBAT	0.26	0.40	-0.41	0.26	0.42	0.48	0.34	0.48	0.51	0.18	1.00	0.74	0.40	-0.03	-0.04
ZPCTNPLG	0.31	0.51	-0.49	0.36	0.51	0.57	0.43	0.52	0.63	0.21	0.74	1.00	0.43	0.05	-0.05
ZPCTNKIT	0.20	0.43	-0.41	0.33	0.39	0.42	0.35	0.48	0.45	0.20	0.40	0.43	1.00	0.13	0.05
ZPCTBL	0.20	0.44	-0.23	0.54	0.24	0.18	0.37	0.29	0.25	0.36	-0.03	0.05	0.13	1.00	0.27
PCT1019M	0.26	0.23	-0.19	0.29	0.16	-0.13	-0.05	0.14	-0.04	0.36	-0.04	-0.05	0.05	0.27	1.00

ZUR	Log unemployment rate
LFP	Labor force participation
ZPTLPOVL	Log pct persons below poverty
ZMDNINC	Log median income
ZPTFAM1P	Log pct families with 1 parent
ZPGT1PRM	Log pct units with > 1 person per room
ZDENS1	Log persons per building
ZDENS2	Log units per building
ZPTEN75	Log pct occ tenure since 1975
AVEOCC	Average occupancy tenure
ZPCTVAC	Log pct vacant housing units
ZPTVBLDG	Log pct vacant buildings
PCTRENTU	Log pct occupied units rented
ZPCTNPLG	Log pct housing units without compl bath
ZPCTNBAT	Log pct housing units without compl plumbing
ZPCTNKIT	Log pct housing units without compl kitchen
ZPCTBL	Log pct black
PCT1019M	Pct age 10-19 male

Table 7c: Correlation Matrix of Socio-Economic and Demographic Variables
Boston Census Tracts (N = 123)

	UR	POVL	MDINC	FAM1P	DENS1	DENS2	TEN75	VAC	RENTU	GT1PRM	NBAT	NPLG	NKIT	PCTBL	1019M
ZUR	1.00	0.46	-0.49	0.59	0.13	-0.04	-0.17	0.36	0.13	0.49	0.15	0.13	0.18	0.30	0.51
ZPTLPOVL	0.46	1.00	-0.80	0.69	0.73	0.65	0.50	0.63	0.78	0.53	0.29	0.34	0.41	0.54	0.34
ZMDNINC	-0.49	-0.80	1.00	-0.60	-0.58	-0.53	-0.22	-0.50	-0.71	-0.47	-0.30	-0.30	-0.35	-0.31	-0.41
ZPTFAM1P	0.59	0.69	-0.60	1.00	0.37	0.31	0.16	0.66	0.44	0.43	0.20	0.22	0.36	0.56	0.47
ZDENS1	0.13	0.73	-0.58	0.37	1.00	0.91	0.70	0.45	0.85	0.34	0.35	0.39	0.29	0.36	0.18
ZDENS2	-0.04	0.65	-0.53	0.31	0.91	1.00	0.75	0.44	0.90	0.12	0.36	0.39	0.28	0.23	-0.10
ZPTEN75	-0.17	0.50	-0.22	0.16	0.70	0.75	1.00	0.23	0.64	0.07	0.17	0.24	0.27	0.36	-0.15
ZPCTVAC	0.36	0.63	-0.50	0.66	0.45	0.44	0.23	1.00	0.46	0.43	0.36	0.37	0.40	0.35	0.35
PCTRENTU	0.13	0.78	-0.71	0.44	0.85	0.90	0.64	0.46	1.00	0.29	0.34	0.42	0.31	0.33	0.00
ZPGT1PRM	0.49	0.53	-0.47	0.43	0.34	0.12	0.07	0.43	0.29	1.00	0.15	0.19	0.20	0.58	0.54
ZPCTNBAT	0.15	0.29	-0.30	0.20	0.35	0.36	0.17	0.36	0.34	0.15	1.00	0.72	0.28	-0.05	0.04
ZPCTNPLG	0.13	0.34	-0.30	0.22	0.39	0.39	0.24	0.37	0.42	0.19	0.72	1.00	0.33	0.13	0.01
ZPCTNKIT	0.18	0.41	-0.35	0.36	0.29	0.28	0.27	0.40	0.31	0.20	0.28	0.33	1.00	0.19	0.19
ZPCTBL	0.30	0.54	-0.31	0.56	0.36	0.23	0.36	0.35	0.33	0.58	-0.05	0.13	0.19	1.00	0.33
PCT1019M	0.51	0.34	-0.41	0.47	0.18	-0.10	-0.15	0.35	0.00	0.54	0.04	0.01	0.19	0.33	1.00

ZUR	Log unemployment rate
LFP	Labor force participation
ZPTLPOVL	Log pct persons below poverty
ZMDNINC	Log median income
ZPTFAM1P	Log pct families with 1 parent
ZPGT1PRM	Log pct units with > 1 person per room
ZDENS1	Log persons per building
ZDENS2	Log units per building
ZPTEN75	Log pct occ tenure since 1975
AVEOCC	Average occupancy tenure
ZPCTVAC	Log pct vacant housing units
ZPTVBLDG	Log pct vacant buildings
PCTRENTU	Log pct occupied units rented
ZPCTNPLG	Log pct housing units without compl bath
ZPCTNBAT	Log pct housing units without compl plumbing
ZPCTNKIT	Log pct housing units without compl kitchen
ZPCTBL	Log pct black
PCT1019M	Pct age 10-19 male

empirically that theoretically we cannot. The algorithms tries, but hedges its bet through inflated standard errors and extended confidence intervals around the coefficients.[15]

As an alternative, we employed a factor analytic approach to extract from the socio-economic profiles the most prominent underlying dimensions. Next, factor scores on the major dimensions were generated for use as regressors in place of the large mix of census variables.[16]

---

[15]Multicollinearity occurs when two or more predictors are so highly intercorrelated that the regression procedure has difficulty separating the unique effects of each variable. Whatever coefficients that are obtained are, as a result, unstable or unreliable because they can change substantially with minor modifications in specification (choice of predictors). The stability of a regression hyper-plane depends on the extent to which the predictor variables are orthogonal (uncorrelated) or oblique (correlated). As an analogy, consider a table top mounted on a pedestal base which is comprised of a single vertical pole with two crossed legs. If the legs are perpendicular ("uncorrelated," in a sense), the table is stable; if the legs are oblique ("correlated"), the table can wobble. The wobble is caused by the fact that the table top cannot determine for certain how much weight to put on one leg versus the other. The more oblique the legs, the greater the instability of the table. In the extreme, if the legs go in the same exact direction, the table collapses. Similarly, in regression analysis, high correlations among predictors cause the algorithm to have difficulty determining how much weight to assign to each regressor. Perfect collinearity breaks down the algorithm completely.

[16]There are several biased regression methods available for handling multicollinear or ill-conditioned data (see Fisher, 1981). We considered the ridge approach, but encountered some troubling computational problems in the process. Overall, the factor approach was computationally manageable and at the same time maintained a sufficient degree of substantive meaningfulness.

Thirteen socio-economic and housing variables were factor analyzed to extract the more prominent dimensions to the profile. Three variables (LFP, AVEOCC, and ZPTVBLDG) had to be removed to ensure that the correlation matrix was positive-definite and thus could be factor analyzed. Additionally, the two demographic variables (ZPCTBL and PCT1019M) were kept separate since, as control variables, they have very different meaning from the rest of the census measures.

As is customary, factors with eigenvalues greater than one were retained. By this criterion, retained factors were those which contribute more than their share to the total variance represented by the thirteen observable variables. Overall, the three factors account for 82.4% of the total variance among the zip codes, 74.6% of the variance among the urban census tracts, and 74.0% among of the variance for the Boston tracts. Thus a seemingly acceptable degree of variance in the total set of thirteen variables can be represented by the three underlying dimensions.

Tables 8a through 8c also show the factor matrices after rigid orthogonal rotation (Varimax). This is designed to produce a so-called "simple structure" in that, to the extent possible,

Table 8a: Varimax Rotated Factor Solution for Zip Codes

	Factor 1 (Urbanness)	Factor 2 (Poverty)	Factor 3 (Hous Qual)	Communality
ZUR	-0.0281	0.8886	0.2591	0.858
ZPTLPOVL	0.5922	0.7312	0.2006	0.926
ZMDNINC	-0.4799	-0.7835	-0.2736	0.919
ZPTFAM1P	0.5776	0.7040	0.1490	0.851
ZPGT1PRM	0.2622	0.7256	0.1626	0.622
ZDENS1	0.8729	0.2455	0.2014	0.863
ZDENS2	0.8914	0.3363	0.2035	0.949
ZPTEN75	0.8054	0.1224	0.2243	0.714
ZPCTVAC	0.3920	0.6834	0.2049	0.663
PCTRENTU	0.8486	0.4116	0.2366	0.945
ZPCTNBAT	0.1433	0.1846	0.9209	0.903
ZPCTNPLG	0.4034	0.3245	0.7723	0.864
ZPCTNKIT	0.3502	0.4944	0.5172	0.635
Eigenvalue	8.3181	1.3479	1.0451	
Pct of Var	64.0%	10.4%	8.0%	
Cum. Pct	64.0%	74.4%	82.4%	

ZUR	Log unemployment rate
ZPTLPOVL	Log pct persons below poverty
ZMDNINC	Log median income
ZPTFAM1P	Log pct families with 1 parent
ZPGT1PRM	Log pct units with > 1 person per room
ZDENS1	Log persons per building
ZDENS2	Log units per building
ZPTEN75	Log pct occ tenure since 1975
ZPCTVAC	Log pct vacant housing units
PCTRENTU	Log pct occupied units rented
ZPCTNPLG	Log pct housing units without compl bath
ZPCTNBAT	Log pct housing units without compl plumbing
ZPCTNKIT	Log pct housing units without compl kitchen

Table 8b: Varimax Rotated Factor Solution for Census Tracts

	Factor 1 (Urbanness)	Factor 2 (Poverty)	Factor 3 (Hous Qual)	Communality
ZUR	-0.1227	0.8386	0.2162	0.765
ZPTLPOVL	0.5834	0.6982	0.2322	0.882
ZMDNINC	-0.4609	-0.6769	-0.2789	0.748
ZPTFAM1P	0.3477	0.7443	0.1250	0.691
ZPGT1PRM	0.0684	0.6868	0.0507	0.479
ZDENS1	0.8584	0.2079	0.2476	0.841
ZDENS2	0.8902	0.1280	0.3397	0.924
ZPTEN75	0.8588	0.0526	0.1727	0.770
ZPCTVAC	0.3844	0.5092	0.4556	0.615
PCTRENTU	0.8406	0.3036	0.3561	0.926
ZPCTNBAT	0.1844	0.1104	0.8850	0.829
ZPCTNPLG	0.3121	0.1884	0.8201	0.805
ZPCTNKIT	0.2703	0.2141	0.5550	0.427
Eigenvalue	6.8923	1.7157	1.0943	
Pct of Var	53.0%	13.2%	8.4%	
Cum. Pct	53.0%	66.2%	74.6%	

ZUR	Log unemployment rate
ZPTLPOVL	Log pct persons below poverty
ZMDNINC	Log median income
ZPTFAM1P	Log pct families with 1 parent
ZPGT1PRM	Log pct units with > 1 person per rocm
ZDENS1	Log persons per building
ZDENS2	Log units per building
ZPTEN75	Log pct occ tenure since 1975
ZPCTVAC	Log pct vacant housing units
PCTRENTU	Log pct occupied units rented
ZPCTNPLG	Log pct housing units without compl bath
ZPCTNBAT	Log pct housing units without compl plumbing
ZPCTNKIT	Log pct housing units without compl kitchen

Table 8c: Varimax Rotated Factor Solution for Boston Census Tracts

	Factor 1 (Urbanness)	Factor 2 (Poverty)	Factor 3 (Hous Qual)	Communality
ZUR	-0.1880	0.8330	0.0689	0.734
ZPTLPOVL	0.6194	0.6964	0.1564	0.893
ZMDNINC	-0.4473	-0.7021	-0.1613	0.719
ZPTFAM1P	0.2028	0.8138	0.1205	0.718
ZPGT1PRM	0.0758	0.7168	0.0594	0.523
ZDENS1	0.8693	0.2729	0.2045	0.872
ZDENS2	0.9295	0.1057	0.2441	0.935
ZPTEN75	0.8612	-0.0773	0.0798	0.754
ZPCTVAC	0.2848	0.6216	0.3543	0.593
PCTRENTU	0.8592	0.3186	0.2188	0.888
ZPCTNBAT	0.1291	0.0943	0.8939	0.825
ZPCTNPLG	0.1990	0.0997	0.8818	0.827
ZPCTNKIT	0.2060	0.3187	0.4256	0.325
Eigenvalue	6.1201	2.1128	1.3721	
Pct of Var	47.1%	16.3%	10.6%	
Cum. Pct	47.1%	63.4%	74.0%	

ZUR	Log unemployment rate
ZPTLPOVL	Log pct persons below poverty
ZMDNINC	Log median income
ZPTFAM1P	Log pct families with 1 parent
ZPGT1PRM	Log pct units with > 1 person per room
ZDENS1	Log persons per building
ZDENS2	Log units per building
ZPTEN75	Log pct occ tenure since 1975
ZPCTVAC	Log pct vacant housing units
PCTRENTU	Log pct occupied units rented
ZPCTNPLG	Log pct housing units without compl bath
ZPCTNBAT	Log pct housing units without compl plumbing
ZPCTNKIT	Log pct housing units without compl kitchen

observable variables load strongly on only one underlying dimension or factor.[17]

The first factor, labeled "Urbanness," has strong loadings for population and housing density (ZDENS1 and ZDENS2), on housing tenure (ZPTEN75), and percent rental units (PCTRENTU), and moderately on certain economic measures involving income. The second factor, labeled "Poverty," has high positive loadings for the unemployment rate (ZUR), percent of persons below the poverty line (ZPTPOVL), percent of one parent families (ZPTFAM1P), percent of units with more than one person per room (ZPGT1PRM), and percent of vacant housing units (ZPCTVAC), as well as a strong negative loading on the median income (ZMDNINC). The final factor, labeled "Housing Quality," has strong loadings for percent of housing units without complete pluming and percent without complete bath facilities (ZPCTNPLG and ZPCTNBAT) and modestly on the percent of without complete kitchen (ZPCTNKIT).

In addition to the fairly good level of fit, as measured by the cumulative percent of variance explained, most of the observable variables, with the exception of ZPCTNKIT, have reasonably high communalities.[18] That is, for the most part,

---

[17]A factor loading is the correlation of an observable variable with an underlying factor.

[18]A communality represents the percent of variance in an observable that is accounted for by the underlying factors in combination.

the three factors adequately substitute for the thirteen separate census variables.

7. Regression Results

Tables 9a, 9b, and 9c provide descriptive statistics (means, variances, and correlations) of the two arson variables and the five independent variables, FS1 (Urbanness factor), FS2 (Poverty factor), FS3 (Housing Quality factor), ZPCTBL (Percent black), and PCT1019M (Percent age 10-19 male) to be included in the regression, as well as a set of location dummies for the nine largest cities in Massachusetts. The three factors enjoy the benefits of being mutually uncorrelated, which permit their independent effects of the arson rates to be estimated reliably. The demographic control variables (ZPCTBL and PCT1019M), because they were not included in the factor analysis, are, of course, not uncorrelated with the three socio-economic factors. With only a few exceptions, the intercorrelations of the demographic controls and the socio-economic factors are modest. Thus, the methodological problems created by using correlated regressors are far outweighed by the advantages of keeping the socio-economic and demographic effects separate.

In Tables 10a and 10b the arson rates are regressed on these three factors along with the two demographic control variables, followed by dummy variables for each of the nine largest cities in the state. These dummy variables control for global city differences in arson experience combined with fire department differences in classification of fires as arson. Specifically,

Table 9a: Means, Standard Deviations and Correlations for Regression Variables
Zip Codes (N = 288)

Variable	Mean	SD	LRARATE	LCARATE	FS1	FS2	FS3	ZPCTBL	PCT1019M
LRARATE	0.604	1.222	1.000	0.778	0.528	0.591	0.204	0.479	-0.143
LCARATE	1.096	1.550	0.778	1.000	0.423	0.528	0.143	0.391	-0.109
FS1	0.000	1.000	0.528	0.423	1.000	0.000	0.000	0.483	-0.380
FS2	0.000	1.000	0.591	0.528	0.000	1.000	0.000	0.366	-0.079
FS3	0.000	1.000	0.204	0.143	0.000	0.000	1.000	-0.082	-0.042
ZPCTBL	-0.041	1.644	0.479	0.391	0.483	0.366	-0.082	1.000	-0.063
PCT1019M	8.830	1.557	-0.143	-0.109	-0.380	-0.079	-0.042	-0.063	1.000
BOST	0.105	0.307	0.422	0.575	0.430	0.259	-0.027	0.376	-0.092
BROC	0.019	0.137	0.129	0.089	-0.034	0.108	0.075	0.134	0.066
CAMB	0.025	0.157	0.058	0.038	0.231	-0.092	0.063	0.143	-0.048
FALL	0.019	0.135	0.154	0.150	0.073	0.102	0.107	-0.081	-0.084
LAWR	0.020	0.141	0.129	0.137	0.017	0.107	0.062	-0.008	-0.059
LOWE	0.019	0.135	0.130	0.057	0.061	0.052	0.077	0.023	0.020
NEWB	0.022	0.148	0.052	0.045	-0.047	0.184	0.040	0.043	-0.085
SPRI	0.031	0.174	0.083	0.011	-0.031	0.217	-0.177	0.190	0.036
WORC	0.033	0.180	0.088	0.036	0.104	0.061	-0.028	0.077	-0.022

LARARTE	Log residential arson per 1000 buildings
LCARATE	Log automobile arson per 1000 automobiles
FS1	Urbanness factor score
FS2	Poverty factor score
FS3	Housing quality factor score
ZPCTBL	Log percent population black
PCT1019M	Population age 10-19 and male
BOST to WORC	Dummy variables for Boston, Brockton, Cambridge, Fall River, Lawrence, Lowell, New Bedford, Springfield, and Worcester

Table 9b: Means, Standard Deviations and Correlations for Regression Variables
Urban Census Tracts (N = 293)

Variable	Mean	SD	LRARATE	LCARATE	FS1	FS2	FS3	ZPCTBL	PCT1019M
LRARATE	1.768	1.220	1.000	0.517	0.500	0.515	0.241	0.405	0.246
LCARATE	2.495	1.444	0.517	1.000	0.122	0.530	0.156	0.194	0.230
FS1	0.000	1.000	0.500	0.122	1.000	0.000	0.000	0.280	-0.059
FS2	0.000	1.000	0.515	0.530	0.000	1.000	0.000	0.425	0.416
FS3	0.000	1.000	0.241	0.156	0.000	0.000	1.000	-0.160	-0.115
ZPCTBL	0.821	2.410	0.405	0.194	0.280	0.425	-0.160	1.000	0.270
PCT1019M	8.437	2.826	0.246	0.230	-0.059	0.416	-0.115	0.270	1.000
BOST	0.411	0.493	0.205	0.568	0.268	0.016	-0.019	0.155	-0.060
BROC	0.076	0.265	-0.101	-0.122	-0.190	-0.031	-0.082	0.049	0.103
CAMB	0.064	0.245	0.010	-0.182	0.258	-0.191	0.035	0.139	-0.011
FALL	0.077	0.266	-0.018	-0.029	-0.060	-0.037	0.166	-0.361	-0.055
LAWR	0.049	0.216	0.111	0.069	-0.027	0.072	0.125	-0.013	-0.050
LOWE	0.061	0.240	-0.096	-0.171	-0.098	-0.095	0.022	-0.121	0.086
NEWB	0.068	0.252	-0.132	-0.203	-0.197	0.069	0.104	-0.101	-0.064
SPRI	0.101	0.301	-0.073	-0.222	-0.221	0.197	-0.243	0.133	0.107
WORC	0.095	0.293	-0.059	-0.174	0.056	-0.042	-0.024	-0.034	-0.022

LARARTE	Log residential arson per 1000 buildings
LCARATE	Log automobile arson per 1000 automobiles
FS1	Urbanness factor score
FS2	Poverty factor score
FS3	Housing quality factor score
ZPCTBL	Log percent population black
PCT1019M	Population age 10-19 and male
BOST to WORC	Dummy variables for Boston, Brockton, Cambridge, Fall River, Lawrence, Lowell, New Bedford, Springfield, and Worcester

Table 9c: Means, Standard Deviations and Correlations for Regression Variables
Boston Census Tracts (N = 123)

Variable	Mean	SD	LRARATE	LCARATE	FS1	FS2	FS3	ZPCTBL	PCT1019M
LRARATE	2.066	1.278	1.000	0.487	0.461	0.610	0.201	0.504	0.429
LCARATE	3.475	1.319	0.487	1.000	-0.268	0.718	0.135	0.199	0.536
FS1	0.000	1.000	0.461	-0.268	1.000	0.000	0.000	0.286	-0.128
FS2	0.000	1.000	0.610	0.718	0.000	1.000	0.000	0.516	0.643
FS3	0.000	1.000	0.201	0.135	0.000	0.000	1.000	-0.083	-0.017
ZPCTBL	1.267	2.810	0.504	0.199	0.286	0.516	-0.083	1.000	0.330
PCT1019M	8.234	3.045	0.429	0.536	-0.128	0.643	-0.017	0.330	1.000
BOST	1.000	0.000	---	---	---	---	---	---	---
BROC	0.000	0.000	---	---	---	---	---	---	---
CAMB	0.000	0.000	---	---	---	---	---	---	---
FALL	0.000	0.000	---	---	---	---	---	---	---
LAWR	0.000	0.000	---	---	---	---	---	---	---
LOWE	0.000	0.000	---	---	---	---	---	---	---
NEWB	0.000	0.000	---	---	---	---	---	---	---
SPRI	0.000	0.000	---	---	---	---	---	---	---
WORC	0.000	0.000	---	---	---	---	---	---	---

LARARTE	Log residential arson per 1000 buildings
LCARATE	Log automobile arson per 1000 automobiles
FS1	Urbanness factor score
FS2	Poverty factor score
FS3	Housing quality factor score
ZPCTBL	Log percent population black
PCT1019M	Population age 10-19 and male
BOST to WORC	Dummy variables for Boston, Brockton, Cambridge, Fall River, Lawrence, Lowell, New Bedford, Springfield, and Worcester

Table 10a: Regression Results for Residential Arson (Varimax Solution)

Variable	Zip Codes (1)	(2)	(3)	(4)	Census Tracts (5)	(6)	(7)	(8)	Boston Tracts (9)	(10)	(11)
Urbanness	0.702	0.648	0.608	0.639	0.581	0.568	0.550	0.564	0.578	0.541	0.557
	13.292	10.891	11.866	12.488	11.918	10.461	10.935	12.018	7.638	8.226	8.826
Poverty	0.731	0.671	0.641	0.648	0.523	0.503	0.477	0.517	0.571	0.538	0.562
	16.017	13.250	14.637	15.718	9.542	8.536	8.813	10.094	5.646	6.242	6.685
Hous Qual	0.257	0.241	0.275	0.236	0.328	0.324	0.323	0.335	0.271	0.250	0.308
	6.250	5.739	7.393	6.627	6.954	6.608	7.180	7.914	3.848	4.176	5.010
Pct Bl	0.006	0.004	0.019	-0.002	0.053	0.059	0.071	0.070	0.048	0.061	0.082
	0.177	0.126	0.663	-0.082	2.342	2.343	3.074	3.267	1.557	2.289	3.138
Pct 10-19m	0.104	0.093	0.097	0.110	0.043	0.045	0.042	0.044	0.070	0.060	0.051
	3.564	3.190	3.844	4.386	2.333	2.424	2.458	2.741	2.300	2.302	2.009
Boston		0.317	0.403	0.408							
		1.897	2.768	3.060							
Brockton		0.646	0.573	0.614		-0.072	-0.161	-0.135			
		2.115	2.218	2.667		-0.382	-0.933	-0.843			
Cambridge		-0.082	-0.118	-0.109		-0.357	-0.445	-0.453			
		-0.298	-0.507	-0.522		-1.744	-2.388	-2.553			
Fall River		0.518	0.510	0.506		0.026	-0.088	0.079			
		1.651	1.919	2.133		0.130	-0.476	0.433			
Lawrence		0.555	0.528	0.476		0.267	0.460	0.620			
		1.881	2.117	2.110		1.203	2.175	2.947			
Lowell		0.537	0.500	0.504		-0.155	-0.262	-0.232			
		1.755	1.933	2.182		-0.759	-1.383	-1.322			
New Bedford		-0.099	-0.130	-0.102		-0.448	-0.518	-0.463			
		-0.347	-0.542	-0.473		-2.241	-2.847	-2.708			
Springfield		0.150	0.130	0.218		-0.146	-0.252	-0.178			
		0.597	0.605	1.065		-0.832	-1.533	-1.157			
Worcester		0.132	0.074	0.064		-0.306	-0.375	-0.261			
		0.569	0.369	0.359		-1.854	-2.501	-1.820			
Constant	-0.312	-0.301	-0.287	-0.406	1.365	1.436	1.526	1.500	1.425	1.571	1.588
Multiple R	0.827	0.837	0.874	0.893	0.768	0.781	0.814	0.845	0.806	0.842	0.860
R^2	0.685	0.701	0.765	0.797	0.590	0.610	0.663	0.715	0.649	0.708	0.739
Adjusted R^2	0.679	0.685	0.752	0.785	0.583	0.591	0.646	0.700	0.634	0.695	0.727
S.E.	0.692	0.686	0.579	0.514	0.788	0.780	0.706	0.652	0.774	0.656	0.625
N	288	288	274	256	293	293	283	270	123	118	114

Table 10b: Regression Results for Automobile Arson (Varimax Solution)

Variable	Zip Codes (1)	(2)	(3)	(4)	Census Tracts (5)	(6)	(7)	(8)	Boston Tracts (9)	(10)	(11)
Urbanness	0.752	0.444	0.418	0.441	0.217	−0.105	−0.106	−0.114	−0.284	−0.270	−0.294
	8.668	5.010	5.196	5.724	2.886	−1.826	−1.959	−2.209	−3.440	−3.436	−3.954
Poverty	0.855	0.657	0.658	0.621	0.804	0.862	0.865	0.898	0.982	0.945	0.918
	11.401	8.711	9.565	9.998	9.570	13.961	14.993	16.066	8.908	9.183	9.265
Hous Qual											
Pct Bl	−0.036	−0.046	−0.040	−0.077	−0.056	−0.075	−0.083	−0.089	−0.070	−0.065	−0.070
	−0.679	−0.919	−0.888	−1.897	−1.630	−2.864	−3.397	−3.798	−2.078	−2.055	−2.295
Pct 10-19m	0.116	0.082	0.088	0.175	0.017	0.035	0.033	0.030	0.034	0.025	0.030
	2.430	1.874	2.221	4.613	0.609	1.803	1.804	1.727	1.029	0.819	1.004
Boston		1.955	1.972	2.217							
		7.868	8.632	11.051							
Brockton		0.899	0.802	0.858		−1.632	−1.679	−1.649			
		1.985	1.982	2.481		−8.179	−9.083	−9.427			
Cambridge		0.501	0.455	0.524		−1.210	−1.249	−1.246			
		1.226	1.250	1.668		−5.612	−6.269	−6.407			
Fall River		1.302	1.249	1.339		−1.303	−1.368	−1.400			
		2.799	3.009	3.765		−6.284	−6.972	−7.033			
Lawrence		1.278	1.207	1.249		−0.870	−0.977	−0.873			
		2.915	3.088	3.677		−3.726	−4.323	−3.793			
Lowell		0.482	0.413	0.435		−1.844	−1.806	−1.774			
		1.061	1.019	1.255		−8.532	−8.888	−9.228			
New Bedford		0.194	0.110	0.283		−2.462	−2.523	−2.431			
		0.458	0.291	0.873		−11.691	−12.955	−12.986			
Springfield		−0.276	−0.386	0.004		−2.531	−2.654	−2.664			
		−0.744	−1.157	0.012		−13.967	−15.476	−16.143			
Worcester		0.182	0.130	0.200		−1.697	−1.751	−1.661			
		0.524	0.415	0.743		−9.745	−10.882	−10.559			
Constant	0.069	0.078	0.103	−0.698	2.397	3.298	3.376	3.378	3.281	3.395	3.343
Multiple R	0.685	0.765	0.801	0.845	0.550	0.830	0.854	0.869	0.778	0.784	0.801
R^2	0.469	0.585	0.642	0.715	0.303	0.688	0.729	0.755	0.605	0.614	0.642
Adjusted R^2	0.461	0.566	0.624	0.699	0.293	0.675	0.717	0.744	0.591	0.600	0.628
S.E.	1.138	1.021	0.910	0.774	1.214	0.824	0.758	0.715	0.843	0.783	0.737
N	288	288	274	256	293	293	283	270	123	118	114

Cols. (1), (5), and (9) involve only the five economic and demographic regressors for the zip codes, the census tracts, and the Boston tracts, respectively. Cols. (2) and (6) then introduce dummies for the zip codes nd the tracts. Of course, since all the Boston tracts observations are within the same urban area, the city dummies are not used for this data set.

Overall, both arson rates are most strongly influenced by the Urbanness and Poverty factors. The Housing Quality factor is also significant in the residential arson equations, but is excluded from the vehicular arson equation because of its substantive irrelevance.

The effect of the Urbanness factor weakens substantially when the urban dummies are added and when the data are limited to Boston cases. This is because most of the explanatory power of the Urbanness factor is overtaken by direct urban indicators.

The effect of the percentage of population that is black, although significant for some specifications, is relatively weak overall. Interestingly, the race effect in arson statistics is far weaker than in other types of crime.[19] Moreover, most of

[19]The predominance of blacks in arrest statistics does not extend to arson. In 1989, for example, only 24.5% of all persons arrested for arson and 16.0% of all juveniles arrested for arson were black (FBI, 1990). These are far below comparable numbers for the other Index offenses.

whatever race effect does exists is subsumed by the socio-economic variables in the equations.

The weak effect of the age/sex variable is also noteworthy. The age effect for arson has frequently been discussed in the literature.[20] However, the age factor is strongest for smaller fires, such as brush fires or trash fires, which were not included in the arson measures used here. In addition, whatever age effects that do exist, they tend not to show up in cross-sectional analyses. That is, cross-sectionally, there does not tend to be much variation among the units, in terms of the percent 10-19 male (see Tables 4a-4c). What does not vary cannot exert much predictive power. Thus, while the drop in the rates of arson over the 1980s may be traced to the decline in the adolescent population, this effect does not emerge in a cross-sectional design, of the sort used here.

As suggested by the multiple correlations, car arson rates are generally more difficult to explain across all specifications. This is a result of cross-unit displacement in car arson reports. That is, car fires are counted corresponding to the location of the fire not to the address of the owner. Cars that are abandoned and torched in an area far away from their garagings constitute a significant source of measurement

---

[20]In 1989, for example, the percentage of arson arrests involving persons under the age of eighteen was 43.4%, greater than that for any other Index offense (FBI, 1990.)

error, which cannot be examined in a regression model and which attenuate the multiple correlations.

The fit for the car arson data among the census tracts is noticeably poor, that is, until the city dummies are entered. There is substantial inter-city variation in the car arson problem, not related to urbanness or poverty. Car arson is acutely problematic in Boston and Lawrence, as reflected in the dummy coefficients. Boston is the base contrast in the census tract data; all eight city dummies are negative, and all but that for Lawrence are significantly negative.

As is customary with regression methods, the residuals were examined for any peculiarities which may have distorted the results. Inspection of residual plots (not shown) showed a general normality to the shape of the residuals, owing largely to the transformations performed on all the variables to eliminate skewness. Despite the reduction in skewness, the analysis did reveal a few (but not an excessive number) of larger residuals, that is, cases in which the arson rates were significantly lower or higher than one would predict based on their socio-economic, demographic, and locational profile.

To ensure that the results are robust to the effects of outliers, observations with standardized residuals larger than 2.5 and 2.0 in absolute magnitude were successively removed and the results recomputed in Cols (3) and (4) for the zip codes, cols (7) and (8) for the tracts, and cols. (10) and (11) for the Boston tracts.

The coefficients do not change markedly when outliers are eliminated. One consistent impact of eliminating observations with large residuals is that the multiple correlations increase. This is to be expected, because the removal of outliers homogenizes the distribution and removes the very cases which reduce the multiple correlation.

While the factor analysis performed on the socio-economic census variables achieved a number of methodological advantages, these do not come without the imposition of several rigid assumptions, the most stringent of which is that the factors are uncorrelated. Obviously, we might wish to relax the restriction that Urbanness, Poverty, and Housing Quality are necessarily independent dimensions, by permitting an oblique rotation of the factors.

The oblique factor solution is shown in Tables 11a-11c. Note that while the overall fit of the factor solution, as measured by the percent of variance explained and by the set of communalities is unchanged by an oblique rather than orthogonal rotation, modest correlations among the factors emerge. While this is reasonable in the real wold, there are certain drawbacks.

Table 11a: Oblique Rotated Factor Solution for Zip Codes

	Factor 1 (Poverty)	Factor 2 (Urbanness)	Factor 3 (Hous Qual)	Communality
ZUR	0.8833	0.2183	0.4641	0.8576
ZPTLPOVL	0.8852	0.7684	0.5060	0.9255
ZMDNINC	-0.9196	-0.6842	-0.5632	0.9190
ZPTFAM1P	0.8456	0.7399	0.4472	0.8515
ZDENS1	0.4998	0.9266	0.4450	0.8629
ZDENS2	0.5908	0.9655	0.4742	0.9491
ZPTEN75	0.3702	0.8372	0.4203	0.7140
ZPCTVAC	0.7885	0.5659	0.4536	0.6626
PCTRENTU	0.6572	0.9469	0.5151	0.9455
ZPGT1PRM	0.7862	0.4442	0.3958	0.6217
ZPCTNBAT	0.3925	0.3266	0.9457	0.9027
ZPCTNPLG	0.5635	0.5848	0.8988	0.8644
ZPCTNKIT	0.6602	0.5321	0.6903	0.6346
Factor 1	1.0000			
Factor 2	0.4992	1.0000		
Factor 3	0.4831	0.4205	1.0000	
Eigenvalue	8.3181	1.3479	1.0451	
Pct of Var	64.0%	10.4%	8.0%	
Cum. Pct	64.0%	74.4%	82.4%	

ZUR	Log unemployment rate
ZPTLPOVL	Log pct persons below poverty
ZMDNINC	Log median income
ZPTFAM1P	Log pct families with 1 parent
ZPGT1PRM	Log pct units with > 1 person per room
ZDENS1	Log persons per building
ZDENS2	Log units per building
ZPTEN75	Log pct occ tenure since 1975
ZPCTVAC	Log pct vacant housing units
PCTRENTU	Log pct occupied units rented
ZPCTNPLG	Log pct housing units without compl bath
ZPCTNBAT	Log pct housing units without compl plumbing
ZPCTNKIT	Log pct housing units without compl kitchen

Table 11b: Oblique Rotated Factor Solution for Urban Census Tracts

	Factor 1 (Urbanness)	Factor 2 (Poverty)	Factor 3 (Hous Qual)	Communality
ZUR	0.0572	0.8323	0.3452	0.7650
ZPTLPOVL	0.7199	0.8094	0.5245	0.8818
ZMDNINC	-0.6086	-0.7767	-0.5302	0.7484
ZPTFAM1P	0.4765	0.8002	0.3690	0.6905
ZDENS1	0.9131	0.3773	0.5115	0.8413
ZDENS2	0.9515	0.3190	0.5899	0.9242
ZPTEN75	0.8735	0.2142	0.4087	0.7701
ZPCTVAC	0.5474	0.6291	0.6399	0.6146
PCTRENTU	0.9342	0.4850	0.6286	0.9256
ZPGT1PRM	0.1824	0.6883	0.2106	0.4790
ZPCTNBAT	0.3864	0.2770	0.9043	0.8294
ZPCTNPLG	0.5074	0.3628	0.8948	0.8054
ZPCTNKIT	0.4137	0.3392	0.6399	0.4269
Factor 1	1.0000			
Factor 2	0.3360	1.0000		
Factor 3	0.5000	0.3966	1.0000	
Eigenvalue	6.8923	1.7157	1.0943	
Pct of Var	53.0%	13.2%	8.4%	
Cum. Pct	53.0%	66.2%	74.6%	

ZUR	Log unemployment rate
ZPTLPOVL	Log pct persons below poverty
ZMDNINC	Log median income
ZPTFAM1P	Log pct families with 1 parent
ZPGT1PRM	Log pct units with > 1 person per room
ZDENS1	Log persons per building
ZDENS2	Log units per building
ZPTEN75	Log pct occ tenure since 1975
ZPCTVAC	Log pct vacant housing units
PCTRENTU	Log pct occupied units rented
ZPCTNPLG	Log pct housing units without compl bath
ZPCTNBAT	Log pct housing units without compl plumbing
ZPCTNKIT	Log pct housing units without compl kitchen

Table 11c: Oblique Rotated Factor Solution for Boston Census Tracts

	Factor 1 (Urbanness)	Factor 2 (Poverty)	Factor 3 (Hous Qual)	Communality
ZUR	-0.0582	0.8042	0.1742	0.7340
ZPTLPOVL	0.7263	0.7838	0.4012	0.8930
ZMDNINC	-0.5595	-0.7679	-0.3713	0.7190
ZPTFAM1P	0.3295	0.8416	0.3013	0.7179
ZDENS1	0.9197	0.4048	0.4246	0.8720
ZDENS2	0.9615	0.2527	0.4456	0.9347
ZPTEN75	0.8442	0.0440	0.2411	0.7540
ZPCTVAC	0.4188	0.6912	0.5094	0.5930
PCTRENTU	0.9182	0.4502	0.4443	0.8876
ZPGT1PRM	0.1825	0.7225	0.1992	0.5230
ZPCTNBAT	0.2760	0.2176	0.9036	0.8247
ZPCTNPLG	0.3432	0.2304	0.9073	0.8270
ZPCTNKIT	0.3107	0.3917	0.5083	0.3252
Factor 1	1.0000			
Factor 2	0.2801	1.0000		
Factor 3	0.3737	0.3165	1.0000	
Eigenvalue	6.1201	2.1128	1.3721	
Pct of Var	47.1%	16.3%	10.6%	
Cum. Pct	47.1%	63.4%	74.0%	

ZUR	Log unemployment rate
ZPTLPOVL	Log pct persons below poverty
ZMDNINC	Log median income
ZPTFAM1P	Log pct families with 1 parent
ZPGT1PRM	Log pct units with > 1 person per room
ZDENS1	Log persons per building
ZDENS2	Log units per building
ZPTEN75	Log pct occ tenure since 1975
ZPCTVAC	Log pct vacant housing units
PCTRENTU	Log pct occupied units rented
ZPCTNPLG	Log pct housing units without compl bath
ZPCTNBAT	Log pct housing units without compl plumbing
ZPCTNKIT	Log pct housing units without compl kitchen

First, the "simple structure" in which observed variables tend to align themselves with one and only one dimension (as in Varimax rotation) no longer holds. Indeed, since the solution allows Urbanness and Poverty, for example, to be correlated, many variables load strongly on both. (Note also, that among the zip codes, the Urbanness and Poverty factors reverse in order.) Thus, while the solution is no longer as "clear-cut" as before, we can still safely assign the same substantive meanings to the three factors.

Second, one of the two justifications for employing factor analysis in the first place (multicollinearity among predictors, with the other being data dimension reduction) is negated by allowing the factors to be correlated. This concern notwithstanding, the regression results along with the introduction of dummy controls and the elimination of observations with large standardized residuals are displayed in Tables 12a-12c and Tables 13a-13b.

The results are similar to the earlier findings using the orthogonal factors. Because only the rotation is different, not the overall fit of the factor solution, many of the regression statistics are unchanged by substituting oblique factors for orthogonal. In particular, the effects of the variables outside of the factor solution, specifically ZPCTBL, PCT1019M as well as

Table 12a: Means, Standard Deviations and Correlations for Regression Variables Zip Codes (N = 288) Oblique Rotation

Variable	Mean	SD	LRARATE	LCARATE	FS1	FS2	FS3	ZPCTBL	PCT1019M
LRARATE	0.604	1.222	1.000	0.778	0.420	0.337	0.035	0.479	-0.143
LCARATE	1.096	1.550	0.778	1.000	0.394	0.260	-0.001	0.391	-0.109
FS1	0.000	1.231	0.420	0.394	1.000	-0.373	-0.348	0.275	0.007
FS2	0.000	1.188	0.337	0.260	-0.373	1.000	-0.236	0.392	-0.340
FS3	0.000	1.175	0.035	-0.001	-0.348	-0.236	1.000	-0.201	0.019
ZPCTBL	-0.041	1.644	0.479	0.391	0.275	0.392	-0.201	1.000	-0.063
PCT1019M	8.830	1.557	-0.143	-0.109	0.007	-0.340	0.019	-0.063	1.000
BOST	0.105	0.307	0.422	0.575	0.171	0.357	-0.122	0.376	-0.092
BROC	0.019	0.137	0.129	0.089	0.094	-0.069	0.059	0.134	0.066
CAMB	0.025	0.157	0.058	0.038	-0.146	0.232	0.049	0.143	-0.048
FALL	0.019	0.135	0.154	0.150	0.061	0.029	0.079	-0.081	-0.084
LAWR	0.020	0.141	0.129	0.137	0.086	-0.019	0.040	-0.008	-0.059
LOWE	0.019	0.135	0.130	0.057	0.022	0.034	0.059	0.023	0.020
NEWB	0.022	0.148	0.052	0.045	0.176	-0.094	0.014	0.043	-0.085
SPRI	0.031	0.174	0.083	0.011	0.251	-0.051	-0.206	0.190	0.036
WORC	0.033	0.180	0.088	0.036	0.044	0.090	-0.051	0.077	-0.022

LARARTE	Log residential arson per 1000 buildings
LCARATE	Log automobile arson per 1000 automobiles
FS1	Urbanness factor score
FS2	Poverty factor score
FS3	Housing quality factor score
ZPCTBL	Log percent population black
PCT1019M	Population age 10-19 and male
BOST to WORC	Dummy variables for Boston, Brockton, Cambridge, Fall River, Lawrence, Lowell, New Bedford, Springfield, and Worcester

Table 12b: Means, Standard Deviations and Correlations for Regression Variables
Urban Census Tracts (N = 293) Oblique Rotation

Variable	Mean	SD	LRARATE	LCARATE	FS1	FS2	FS3	ZPCTBL	PCT1019M
LRARATE	1.768	1.220	1.000	0.517	0.360	0.401	0.070	0.405	0.246
LCARATE	2.495	1.444	0.517	1.000	0.017	0.475	0.060	0.194	0.230
FS1	0.000	1.172	0.360	0.017	1.000	-0.173	-0.424	0.261	-0.075
FS2	0.000	1.106	0.401	0.475	-0.173	1.000	-0.280	0.413	0.434
FS3	0.000	1.203	0.070	0.060	-0.424	-0.280	1.000	-0.265	-0.153
ZPCTBL	0.821	2.410	0.405	0.194	0.261	0.413	-0.265	1.000	0.270
PCT1019M	8.437	2.826	0.246	0.230	-0.075	0.434	-0.153	0.270	1.000
BOST	0.411	0.493	0.205	0.568	0.260	-0.012	-0.073	0.155	-0.060
BROC	0.076	0.265	-0.101	-0.122	-0.158	0.007	-0.038	0.049	0.103
CAMB	0.064	0.245	0.010	-0.182	0.260	-0.222	0.007	0.139	-0.011
FALL	0.077	0.266	-0.018	-0.029	-0.096	-0.060	0.178	-0.361	-0.055
LAWR	0.049	0.216	0.111	0.069	-0.066	0.050	0.117	-0.013	-0.050
LOWE	0.061	0.240	-0.096	-0.171	-0.089	-0.086	0.053	-0.121	0.086
NEWB	0.068	0.252	-0.132	-0.203	-0.223	0.070	0.131	-0.101	-0.064
SPRI	0.101	0.301	-0.073	-0.222	-0.172	0.262	-0.217	0.133	0.107
WORC	0.095	0.293	-0.059	-0.174	0.065	-0.043	-0.029	-0.034	-0.022

LARARTE	Log residential arson per 1000 buildings
LCARATE	Log automobile arson per 1000 automobiles
FS1	Urbanness factor score
FS2	Poverty factor score
FS3	Housing quality factor score
ZPCTBL	Log percent population black
PCT1019M	Population age 10-19 and male
BOST to WORC	Dummy variables for Boston, Brockton, Cambridge, Fall River, Lawrence, Lowell, New Bedford, Springfield, and Worcester

Table 12c: Means, Standard Deviations and Correlations for Regression Variables
Boston Census Tracts (N = 123) Oblique Rotation

Variable	Mean	SD	LRARATE	LCARATE	FS1	FS2	FS3	ZPCTBL	PCT1019M
LRARATE	2.066	1.278	1.000	0.487	0.349	0.514	0.071	0.504	0.429
LCARATE	3.475	1.319	0.487	1.000	-0.362	0.714	0.097	0.199	0.536
FS1	0.000	1.097	0.349	-0.362	1.000	-0.184	-0.313	0.241	-0.188
FS2	0.000	1.073	0.514	0.714	-0.184	1.000	-0.238	0.487	0.648
FS3	0.000	1.110	0.071	0.097	-0.313	-0.238	1.000	-0.175	-0.065
ZPCTBL	1.267	2.810	0.504	0.199	0.241	0.487	-0.175	1.000	0.330
PCT1019M	8.234	3.045	0.429	0.536	-0.188	0.648	-0.065	0.330	1.000
BOST	1.000	0.000	---	---	---	---	---	---	---
BROC	0.000	0.000	---	---	---	---	---	---	---
CAMB	0.000	0.000	---	---	---	---	---	---	---
FALL	0.000	0.000	---	---	---	---	---	---	---
LAWR	0.000	0.000	---	---	---	---	---	---	---
LOWE	0.000	0.000	---	---	---	---	---	---	---
NEWB	0.000	0.000	---	---	---	---	---	---	---
SPRI	0.000	0.000	---	---	---	---	---	---	---
WORC	0.000	0.000	---	---	---	---	---	---	---

LARARTE	Log residential arson per 1000 buildings
LCARATE	Log automobile arson per 1000 automobiles
FS1	Urbanness factor score
FS2	Poverty factor score
FS3	Housing quality factor score
ZPCTBL	Log percent population black
PCT1019M	Population age 10-19 and male
BOST to WORC	Dummy variables for Boston, Brockton, Cambridge, Fall River, Lawrence, Lowell, New Bedford, Springfield, and Worcester

Table 13a: Regression Results for Residential Arson (Oblique Solution)

	Zip Codes				Census Tracts				Boston Tracts		
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Urbanness	0.883	0.815	0.768	0.791	0.712	0.695	0.674	0.693	0.686	0.642	0.670
	16.185	13.075	14.164	15.622	14.532	12.819	13.389	14.546	8.988	9.694	10.441
Poverty	0.925	0.851	0.821	0.823	0.654	0.631	0.605	0.644	0.669	0.630	0.661
	18.893	15.346	16.994	17.873	12.023	10.947	11.429	12.715	6.665	7.359	7.898
Hous Qual	0.583	0.541	0.534	0.525	0.577	0.565	0.555	0.572	0.481	0.447	0.510
	13.859	12.132	13.137	13.834	12.510	11.560	12.271	13.132	6.807	7.433	8.181
Pct Bl	0.006	0.004	0.021	0.002	0.053	0.059	0.071	0.069	0.048	0.061	0.082
	0.177	0.126	0.720	0.083	2.342	2.343	3.074	3.162	1.557	2.289	3.138
Pct 10-19m	0.104	0.093	0.102	0.103	0.043	0.045	0.042	0.045	0.070	0.060	0.051
	3.564	3.190	3.993	4.442	2.333	2.424	2.460	2.833	2.300	2.302	2.009
Boston		0.317	0.388	0.390							
		1.897	2.649	2.879							
Brockton		0.646	0.563	0.607		-0.072	-0.161	-0.143			
		2.115	2.168	2.618		-0.382	-0.930	-0.893			
Cambridge		-0.082	-0.105	-0.110		-0.357	-0.443	-0.453			
		-0.298	-0.448	-0.526		-1.744	-2.377	-2.550			
Fall River		0.518	0.523	0.505		0.026	-0.088	0.073			
		1.651	1.959	2.115		0.130	-0.477	0.399			
Lawrence		0.555	0.527	0.470		0.267	0.459	0.613			
		1.881	2.103	2.067		1.203	2.169	3.024			
Lowell		0.537	0.503	0.505		-0.155	-0.226	-0.240			
		1.755	1.933	2.173		-0.759	-1.216	-1.368			
New Bedford		-0.099	-0.142	-0.117		-0.448	-0.519	-0.475			
		-0.347	-0.585	-0.540		-2.241	-2.850	-2.744			
Springfield		0.150	0.078	0.205		-0.146	-0.253	-0.188			
		0.597	0.357	0.990		-0.832	-1.538	-1.205			
Worcester		0.132	0.062	0.060		-0.306	-0.375	-0.288			
		0.569	0.311	0.335		-1.854	-2.497	-1.985			
Constant	-0.312	-0.301	-0.323	-0.335	1.365	1.436	1.525	1.491	1.425	1.571	1.588
Multiple R	0.827	0.837	0.872	0.894	0.768	0.781	0.814	0.845	0.806	0.842	0.860
R^2	0.685	0.701	0.761	0.798	0.590	0.610	0.662	0.713	0.649	0.708	0.739
Adjusted R^2	0.679	0.685	0.748	0.787	0.583	0.591	0.646	0.699	0.634	0.695	0.727
S.E.	0.692	0.686	0.582	0.517	0.788	0.780	0.707	0.652	0.774	0.656	0.625
N	288	288	275	253	283	293	284	268	123	118	114

Table 13b: Regression Results for Automobile Arson (Oblique Solution)

Variable	Zip Codes (1)	(2)	(3)	(4)	Census Tracts (5)	(6)	(7)	(8)	Boston Tracts (9)	(10)	(11)
Urbanness	0.587	0.282	0.289	0.231	0.149	-0.106	-0.102	-0.114	-0.231	-0.221	-0.225
	6.686	3.358	3.708	3.437	2.154	-2.083	-2.120	-2.511	-2.894	-2.905	-3.164
Poverty	0.683	0.462	0.509	0.420	0.657	0.717	0.720	0.743	0.818	0.790	0.789
	8.958	6.330	7.398	7.008	7.950	11.909	12.684	13.755	7.610	7.853	8.300
Hous Qual											
Pct Bl	0.065	0.023	0.034	-0.007	0.018	-0.053	-0.064	-0.065	-0.053	-0.048	-0.062
	1.130	0.447	0.707	-0.164	0.628	-1.945	-2.457	-2.651	-1.518	-1.475	-1.973
Pct 10-19m	0.044	0.028	0.064	0.074	-0.033	0.038	0.036	0.038	0.046	0.036	0.038
	0.852	0.623	1.534	2.066	-0.909	1.870	1.896	2.085	1.337	1.122	1.267
Boston		2.309	2.222	2.496							
		8.955	9.266	12.133							
Brockton		1.093	0.892	1.046		-1.778	-1.824	-1.793			
		2.272	2.068	2.921		-8.642	-9.466	-10.032			
Cambridge		0.749	0.667	0.789		-1.196	-1.240	-1.230			
		1.734	1.725	2.443		-5.316	-5.884	-6.111			
Fall River		1.798	1.691	1.782		-1.223	-1.278	-1.275			
		3.697	3.878	4.915		-5.670	-6.186	-6.195			
Lawrence		1.575	1.441	1.481		-0.814	-0.922	-0.937			
		3.402	3.474	4.226		-3.345	-3.863	-4.098			
Lowell		0.828	0.682	0.784		-1.902	-1.964	-1.831			
		1.728	1.589	2.201		-8.448	-9.328	-9.210			
New Bedford		0.399	0.240	0.410		-2.476	-2.539	-2.479			
		0.890	0.597	1.225		-11.267	-12.326	-12.624			
Springfield		-0.284	-0.537	0.068		-2.687	-2.824	-2.761			
		-0.714	-1.492	0.212		-14.401	-15.839	-16.200			
Worcester		0.369	0.264	0.411		-1.714	-1.772	-1.710			
		1.004	0.792	1.478		-9.445	-10.436	-10.428			
Constant	0.709	0.473	0.278	0.138	2.366	3.278	3.355	3.305	3.165	3.290	3.268
Multiple R	0.594	0.729	0.771	0.823	0.489	0.813	0.836	0.856	0.761	0.766	0.791
R^2	0.352	0.532	0.595	0.678	0.239	0.661	0.700	0.732	0.579	0.587	0.625
Adjusted R^2	0.343	0.510	0.575	0.660	0.228	0.647	0.686	0.719	0.565	0.572	0.611
S.E.	1.256	1.085	0.971	0.803	1.269	0.858	0.800	0.739	0.870	0.810	0.754
N	288	288	275	253	293	293	284	268	123	118	114

the constant term, are unchanged. Additionally, the measures of overall fit of the regression equations (e.g., multiple correlations and standard errors) are also invariant to the type of factor rotation.

One notable difference among the factors themselves is that Housing Quality appears to have stronger effects under the oblique rotation. This is likely a consequence of the fact that the factors are no longer simple or "clear-cut" in meaning. That is, the Housing Quality factor contains some moderate loadings for variables representative of Poverty and Urbanness.

Examination of the entire set of regression results reveals nothing that is terribly surprising. Overall the fits are quite good, but this, of course, can be expected of aggregated data such as these. It is quite clear that the levels of urbanness and of poverty have the closet association with residential arson, and these relationships remain largely unchanged after inclusion of the city dummies. Neither of the demographic controls contribute much to the regression fits, and, in fact, they are often non-significant in their effects.

8. Conclusion

The objectives of this research were for the most part successfully achieved. Although a number of unexpected data problems were encountered, the statistical results confirmed the important role of poverty and urbanness and the relatively unimportant role of race in the production of arson rates.

It is unfortunately the case that findings of this kind of analysis do not directly suggest strategies for arson prevention. The effects of poverty are far-reaching; the problem of arson is but a small part of the devastation of neighborhoods brought on by economic disorganization, joblessness and despair.

Although the rate of arson has subsided somewhat in recent years, the future does not look encouraging. Not only will the impending increase in the juvenile population mean more juvenile arson, but, more important, the recession overtaking the nation, and particularly the slow real estate market, would suggest that the recent downward trend in arson may soon reverse itself.

In the absence of major social change in American cities, the most effective strategies for attacking the arson problem combine enforcement and legislative initiatives. For years, arson has been one of the most profitable and low risk form of criminal activity. Efforts to minimize the incentives for arson and increase the likelihood of apprehension will likely have a positive impact. A number of programs and initiatives undertaken in Massachusetts in recent years demonstrate the inroads that can be made toward alleviating the arson problem.

In 1987, for example, the Massachusetts legislature passed a bill that directly attacked the problem of vehicular arson in the state. By this law, an insured motorist claiming a fire loss to his/her vehicle is required to file a sworn written report with the fire department having jurisdiction.

Previously, when a claimant had only to deal with his/her own insurance company, generally the worst that would result from a fraudulent claim was that the claim would be denied. Even when the company's own Special Investigative Unit (SIU) uncovered evidence of insurance fraud, there was very little chance that the case would be referred to law enforcement authorities for possible prosecution (see Fox and Tracy, 1988).

The added involvement of an official government agency stipulated under the new legislation has apparently been a deterrent. Early data on fire frequency have shown a substantial downturn in the incidence of vehicular arson in Massachusetts. Following a stable rate of fire claim frequency through most of the 1980s, the claim frequency per exposure (insured vehicle) declined 6.5% from 1986 to 1987, 10.9% from 1987 to 1988, and

14.0% from 1988 to 1989 (Automobile Insurance Bureau of Massachusetts, 1990). While there are, of course, other confounding factors, it is hard to dismiss these post-law changes in vehicular arson rates.

In addition to increasing the risk or perceived risk of detection and/or apprehension, strategies for reducing the opportunity for profit from arson can also be effective. In Boston, this has been accomplished by empowering those who have most to lose by arson--the tenants. By a law enacted in the mid-1970s and revised in 1986, any tenant is entitled to learn the name of the insurance company covering his/her rental unit, the insured value of the property, and the person or persons who would receive loss payments under the policy.

A property that is overinsured, that has excessive tax liens against it, that has multiple code violations, that has several vacant apartments, that has been sold frequently and recently, or that has experienced a number of small fires is a prime candidate for arson. In Boston, tenants have the right to research the economic well-being of their building and can gain assistance from the Boston Arson Prevention Commission in doing so.[21]

---

[21]A superb handbook and guide for tenants called "Does Someone Want to Burn Your Building Down?" has been prepared by the Jamaica Plain Arson Prevention Action Council (Bolger, 1988).

Because of some outspoken tenants rights groups, the Commonwealth of Massachusetts, and the City of Boston in particular, has been a model for arson prevention (see Hammett, 1987). Sadly, it is difficult to motivate most people to take arson as seriously as other forms of crime, even though arson often results in the loss of human life. Most Americans cannot identify with the victims of arson, in the same way that they can empathize with victims of street crime. Most crimes cut across all classes of society, as least to some degree; arson plagues only the underclasses--the poor and minorities.

Almost anyone might on occasion lie awake worrying that an intruder could break into his home and terrorize his family. Almost anyone might on occasion worry that her teenaged son will be shot down on the street by a drug-crazed criminal. But few of us ever worry that an arsonist will burn down our home, that is, of course, unless you are poor, powerless, and living in a building that you do not own and that someone else would prefer go up in smoke.

9. References

Automobile Insurance Bureau of Massachusetts (1990). "1991 Private Automobile Insurance Rates--Part Two: Cost Containment--Fraudulent Claims Payment." Automobile Insurance Rating Filing Docket No. G90-15.

Bennett, W.D., A.V. Merlo, and K.K. Leiker (1987). "Geographical Patterns of Incendiary and Accidental Fires in Springfield, Massachusetts, 1980-1984." Journal of Quantitative Criminology. 3 (1): 47-64.

Bolger, D. (1988). "Does Someone want to Burn Your Building Down?" Jamaica Plain Arson Prevention Action Council, Jamaica Plain, MA.

Brady, J. (1983). "Arson, Urban Economy, and Organized Crime: The Case of Boston." Social Problems 30 (April).

Fahy, R.F. and A.L. Norton (1989). How Being Poor Affects Fire Risk. Fire Journal (January/February): 28-36.

Federal Bureau of Investigation (1980, 1982, 1983, and 1990). Uniform Crime Reports. U.S. Government Printing Office, Washington, D.C.

Fisher, J.C. and R.L. Mason (1981). "The Analysis of Multicollinear Data in Criminology," in J.A. Fox (ed.), Methods in Quantitative Criminology. Academic Press, New York, pp. 99-125.

Fox, J.A. (1978). Forecasting Crime Data: An Econometric Analysis. Lexington Books, Lexington, MA.

Fox, J.A. (1981). "Forecasts of Crime Rates and Their Relation to Demographic and Economic Indicators." Congressional Record, United States House of Representatives Subcommittee on Crime, June 10, 1981.

Fox, J.A. (1985) "Insurance Rates Spur Auto Thefts," The Boston Herald, January 23, 1985.

Fox, J.A. (1990) "The Demographic Cycle of Violence," Congressional Record, Senate Judiciary Committee, July 31, 1990.

Fox, J.A. and P.E. Tracy (1988). "Cost Containment Pertaining to Insurance Fraud," Report to the Office of the Attorney General of Massachusetts, Insurance Division, Boston, MA.

Gunther, P. (1981). Fire-Cause Patterns for Different Socioeconomic Neighborhoods in Toledo, Ohio. Fire Journal May: 52-58.

Hall, J.R. (1990). "U.S. Arson Trends and Patterns--1989." Fire Analysis Division, National Fire Protection Association.

Hammett, T.M. (1987). "Toward Comprehensive Anti-Arson Programs." Issues and Practices in Criminal Justice series, National Institute of Justice, Washington, D.C.

Institute for Puget Sound Needs (1976). An Analysis of Arson in a Socio-Economic Framework, Unpublished report prepared for the Seattle Mayor's Task Force on Arson.

Insurance Bureau of Canada (1981). "A Viewpoint of Insurance." Insurance Bureau of Canada.

Jackson, P.G. (1988). "Assessing the Validity of Official Data on Arson." Criminology 26 (1):181-195.

Karter, M.J. (1982). "Fire Loss in the Unites States during 1981." Fire Journal 76 (September).

Karter, M.J., and A. Donner (1978). The Effects of Demographics on Fire Rates. Fire Journal. January: 53-65.

Karter, M.J., and A. Donner (1977). Fire Rates and Census Characteristics--An Analytical Approach (Final Report II), NFPA Grant No. 76043.

Moore, M.N. (1981). The Socio-Economics of Arson in Boston, Unpublished report, Boston Arson Prevention Commission, Boston, Massachusetts.

National Fire Data Center (1980). "National Fire Incident Reporting System News." Federal Emergency Management Agency, United States Fire Administration, Washington, D.C.

New York State Academy of Fire Science (n.d.). "Selected Articles on Arson and Related Subjects."

Pettiway, L.E. (1983-84). Arson and American City Types. Journal of Environmental Systems. 13 (2): 157-176.

Pettiway, L.E. (1985a). Measures of Opportunity and the Calculation of the Arson Rate: The Connection Between Operationalization and Association. Journal of Quantitative Criminology. 1 (3): 241-268.

Pettiway, L.E. (1985b). The Context Of Arson Incidence: Neighborhood and Environmental Factors. Journal of Environmental Systems 15 (2): 127-158.

Pettiway, L.E. (1987). Arson for Revenge: The Role of Environmental Situation, Age, Sex, and Race. Journal of Quantitative Criminology. 3 (2): 169-184.

Pettiway, L.E. (1988). Urban Spatial Structure and Incidence of Arson: Differences Between Ghetto and Nonghetto Environments. Justice Quarterly, 5 (1): 113-129.

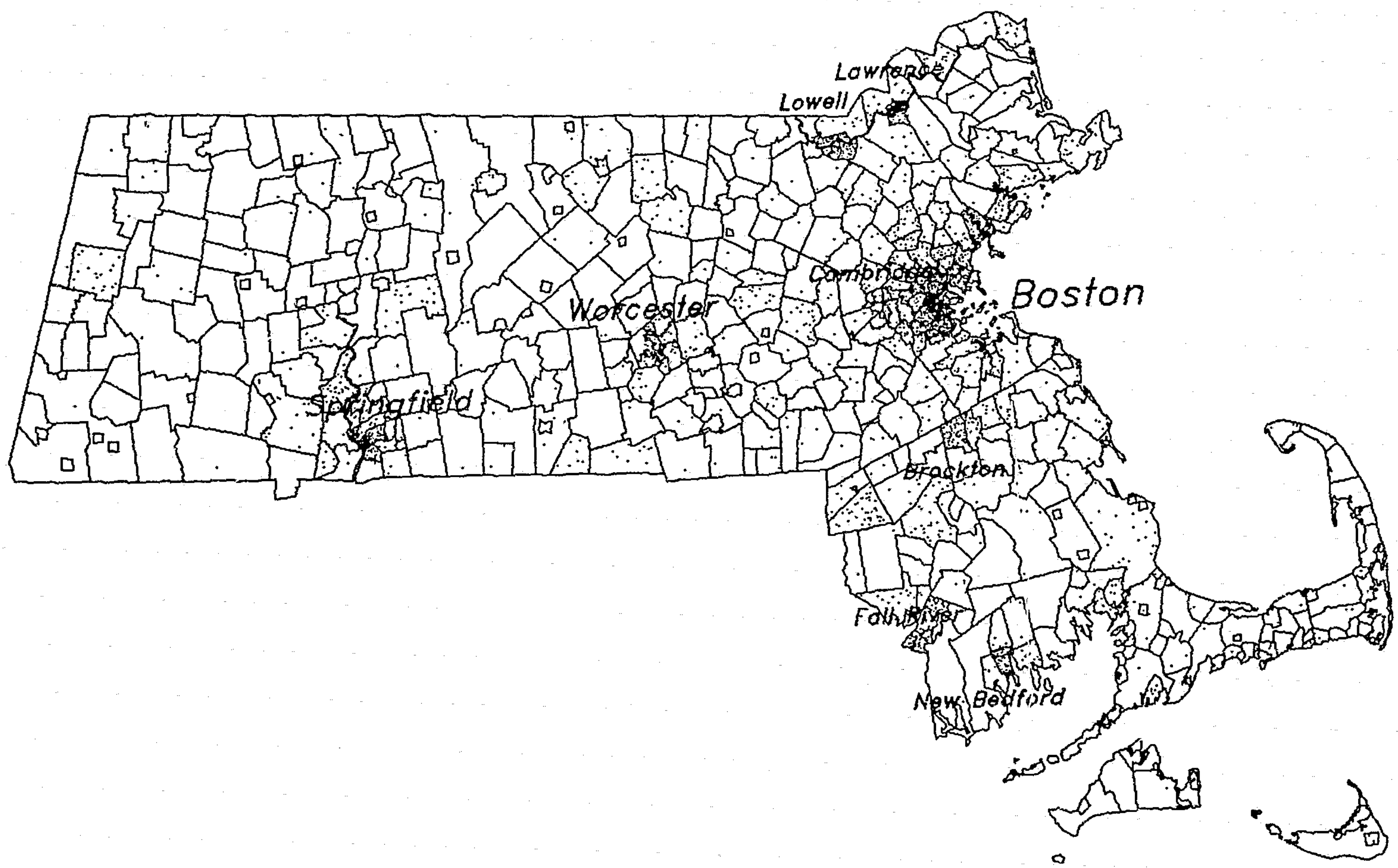
Rider, A.O. (1980). "The Firesetter: A Psychological Profile." FBI Law Enforcement Bulletin (June, July and August).

Schaenman, P.S., J.R. Hall, A.H. Schainblatt, J.A. Swartz, and M.J. Karter (1977). Procedures for Improving the Measurement of Local Fire Protection Effectiveness. The Urban Institute, Washington, D.C.

Sparks, R.F. (1980), "Criminal Opportunities and Crime Rates." In S.E. Fienberg, and A. Reiss (eds.), Indicators of Crime and Criminal Justice: Quantitative Studies. U.S. Government Printing Office, Washington D.C., pp. 18-18.

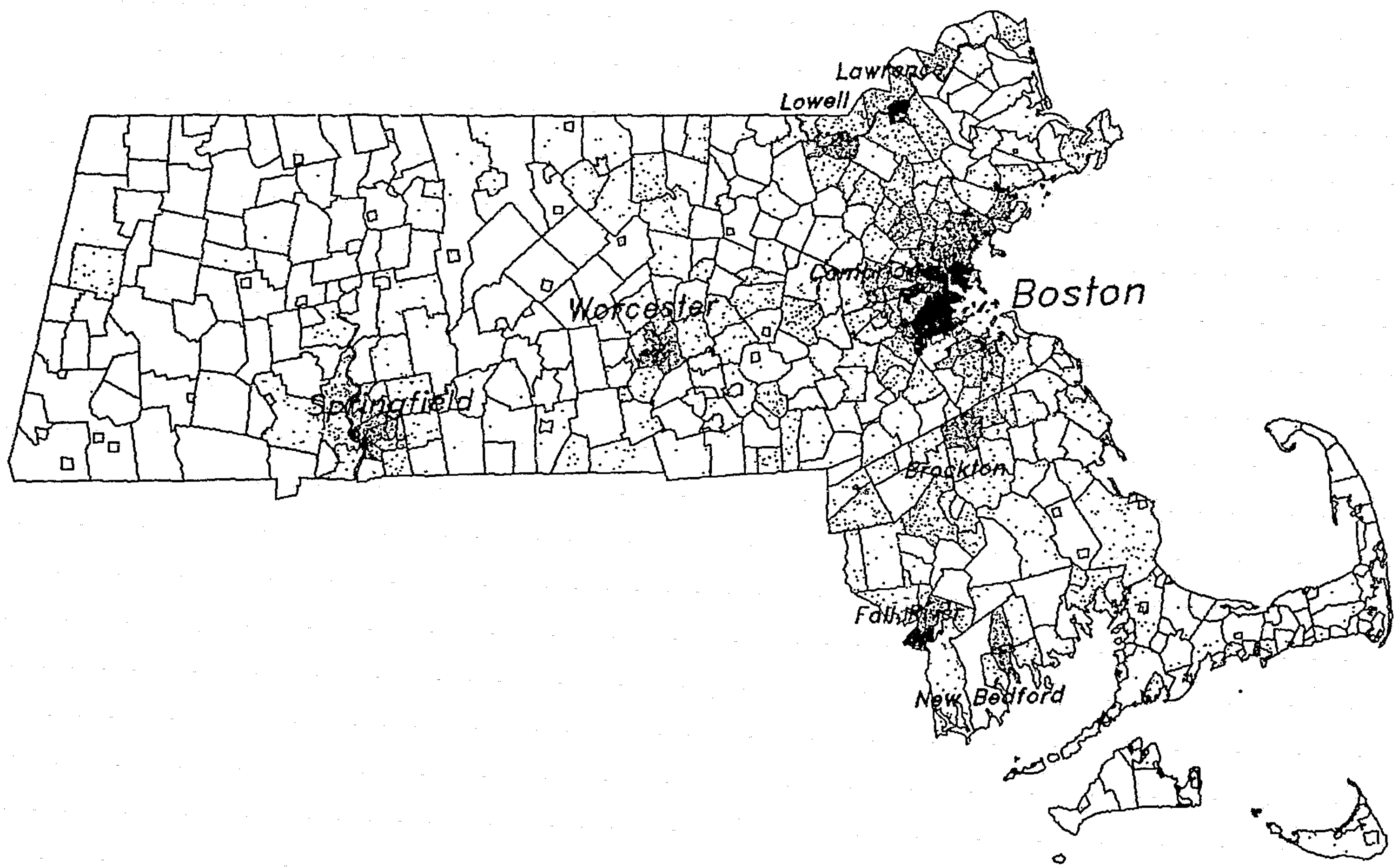
Appendix A: Dot-Density and Hatch Pattern Maps

Figure 4: Residential Arson by Zip Code



1 Dot = 1 Fire

Figure 5: Automobile Arson by Zip Code



1 Dot = 1 Fire

Figure 6: Log Residential Arson Rate by Zip Code

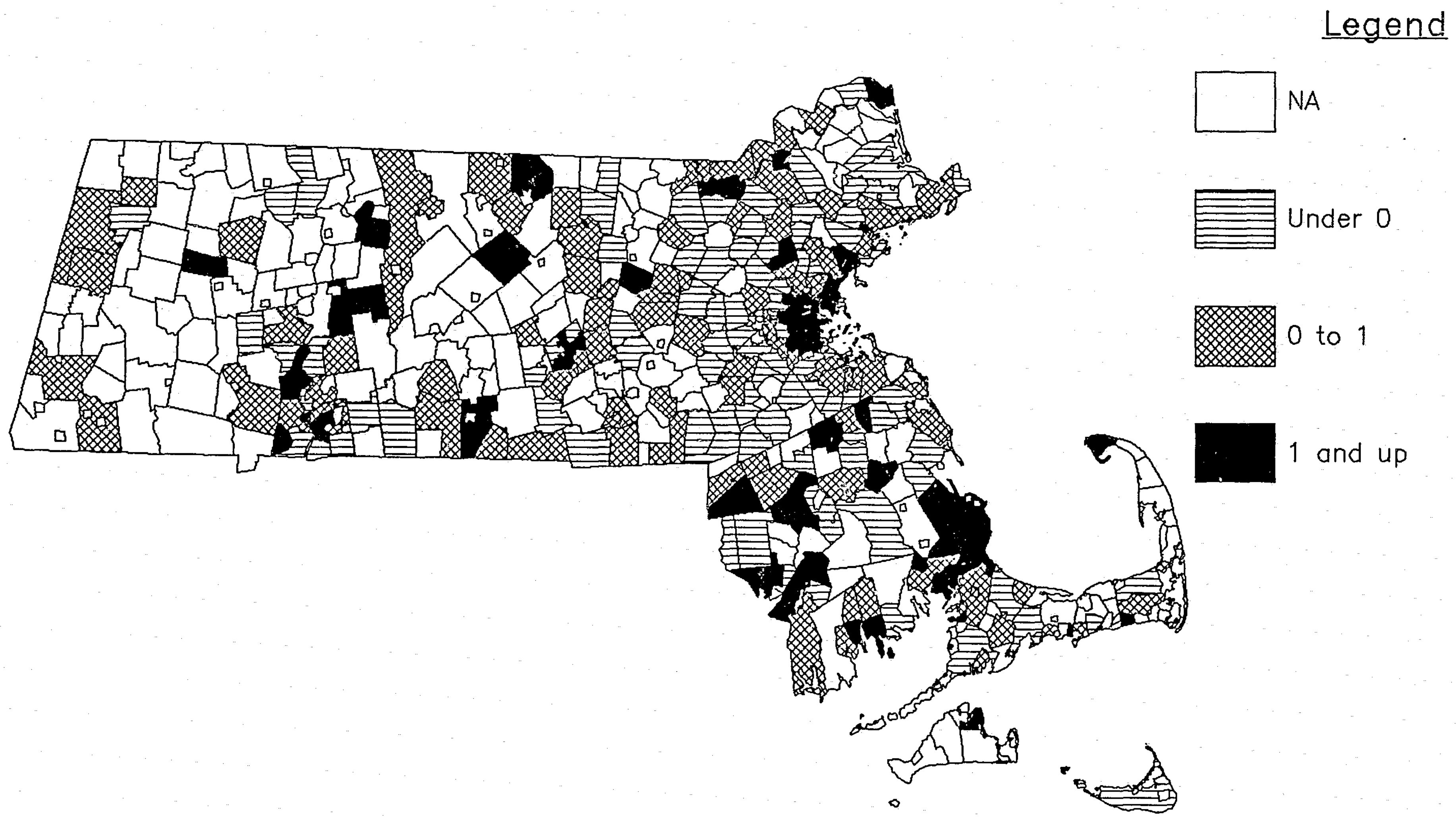


Figure 7: Log Automobile Arson Rate by Zip Code

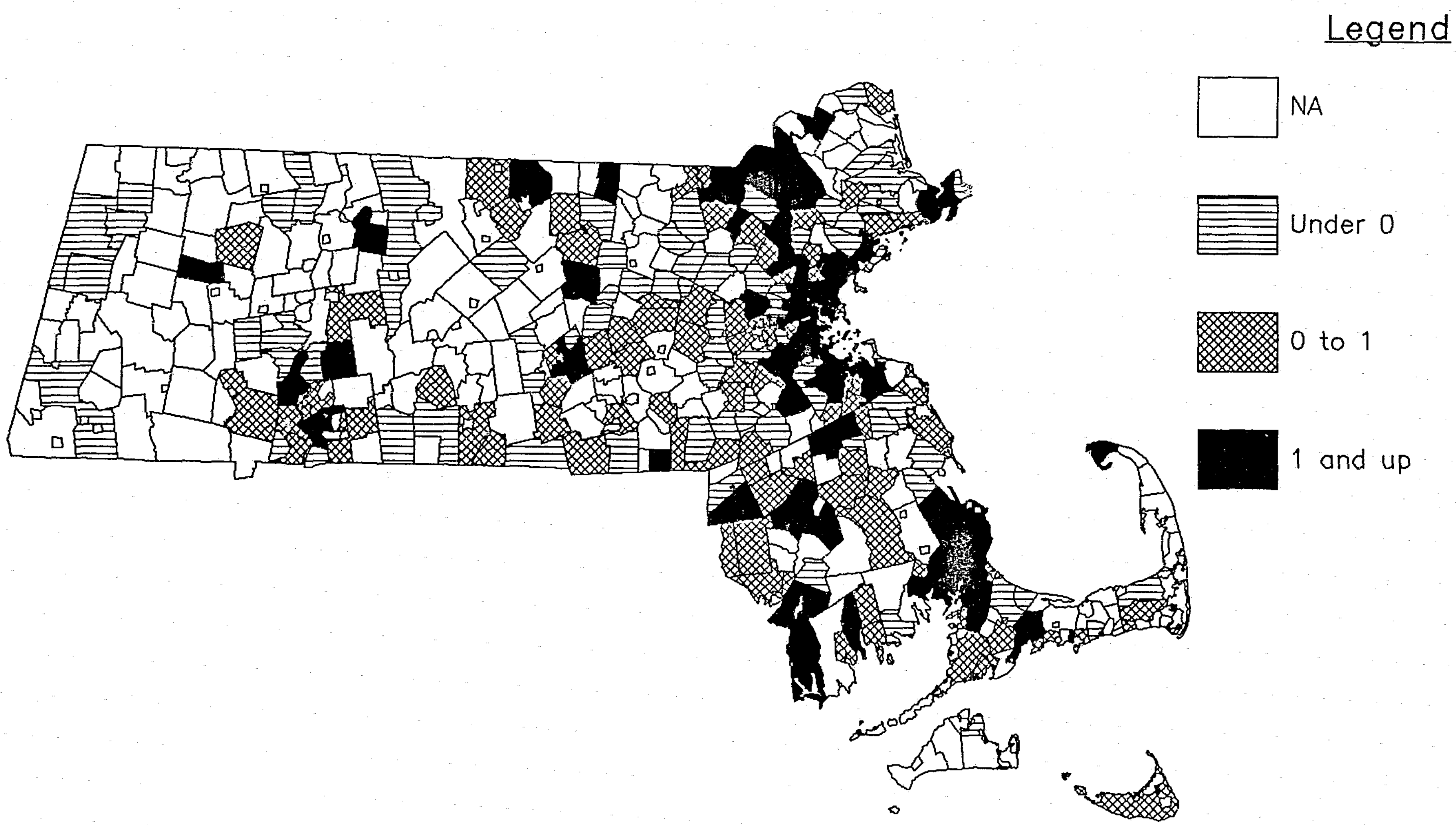


Figure 8: Residential Arson by Boston Tract

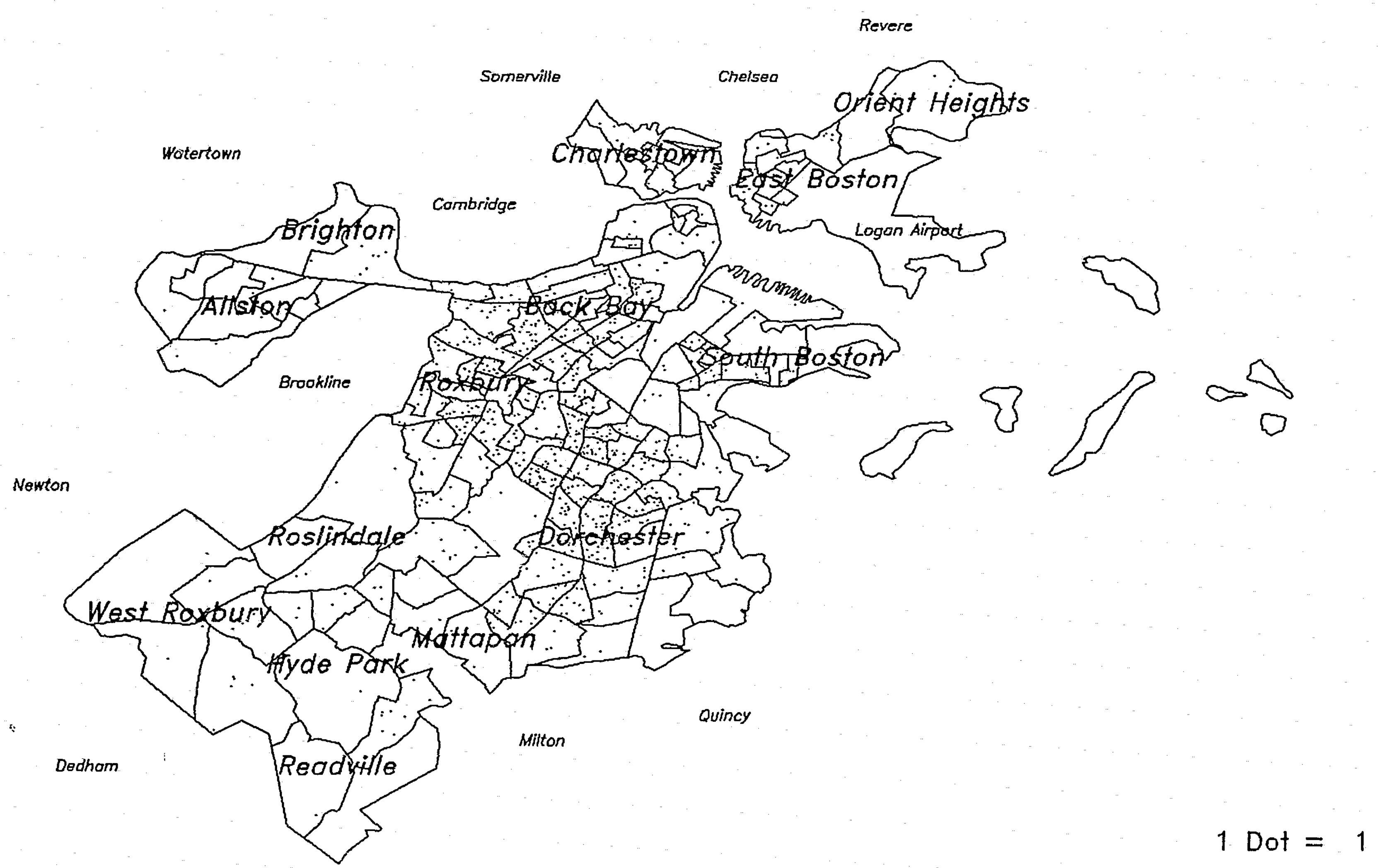


Figure 9: Automobile Arson by Boston Tract

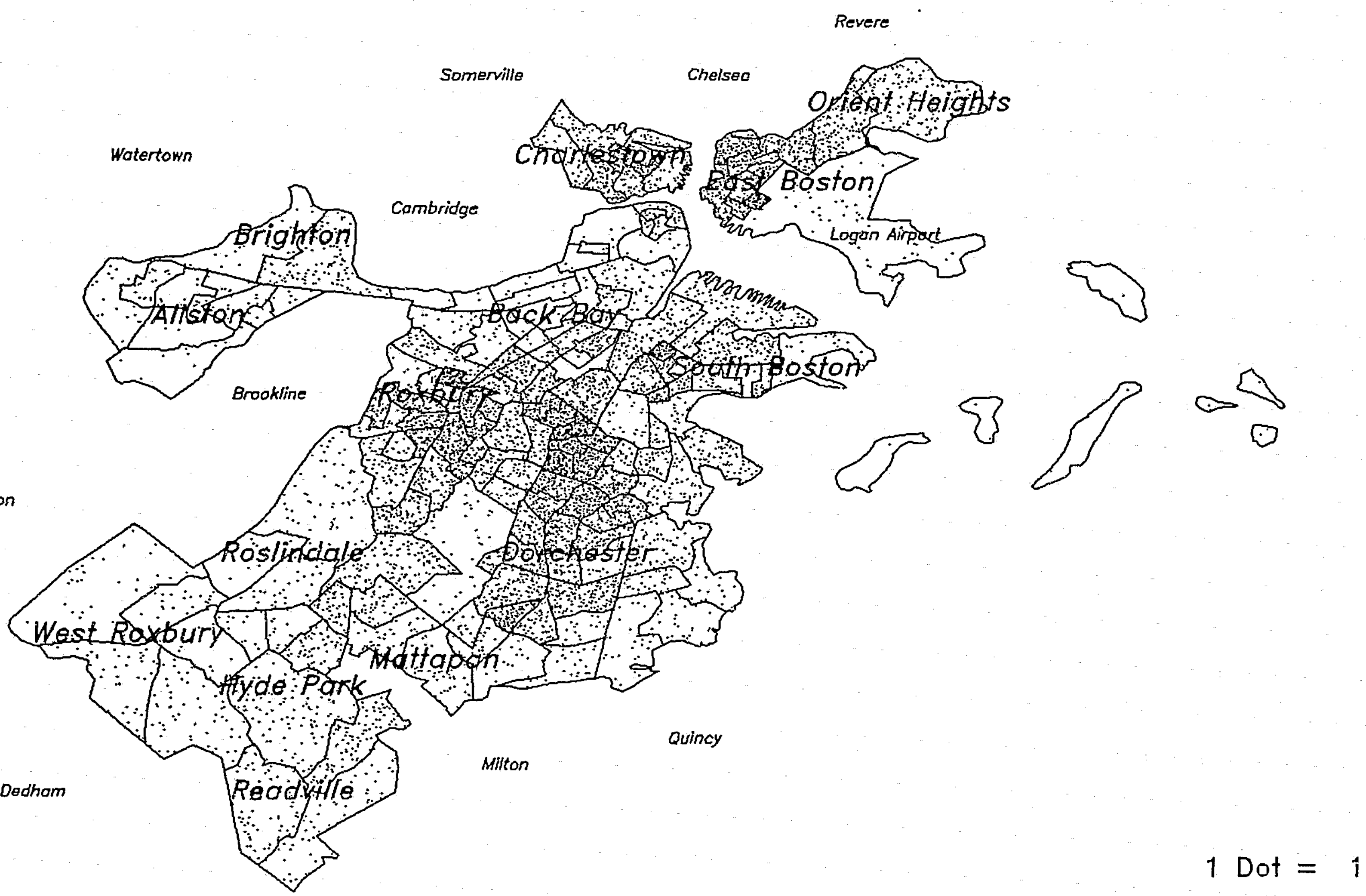


Figure 10: Log Residential Arson Rate by Boston Tract

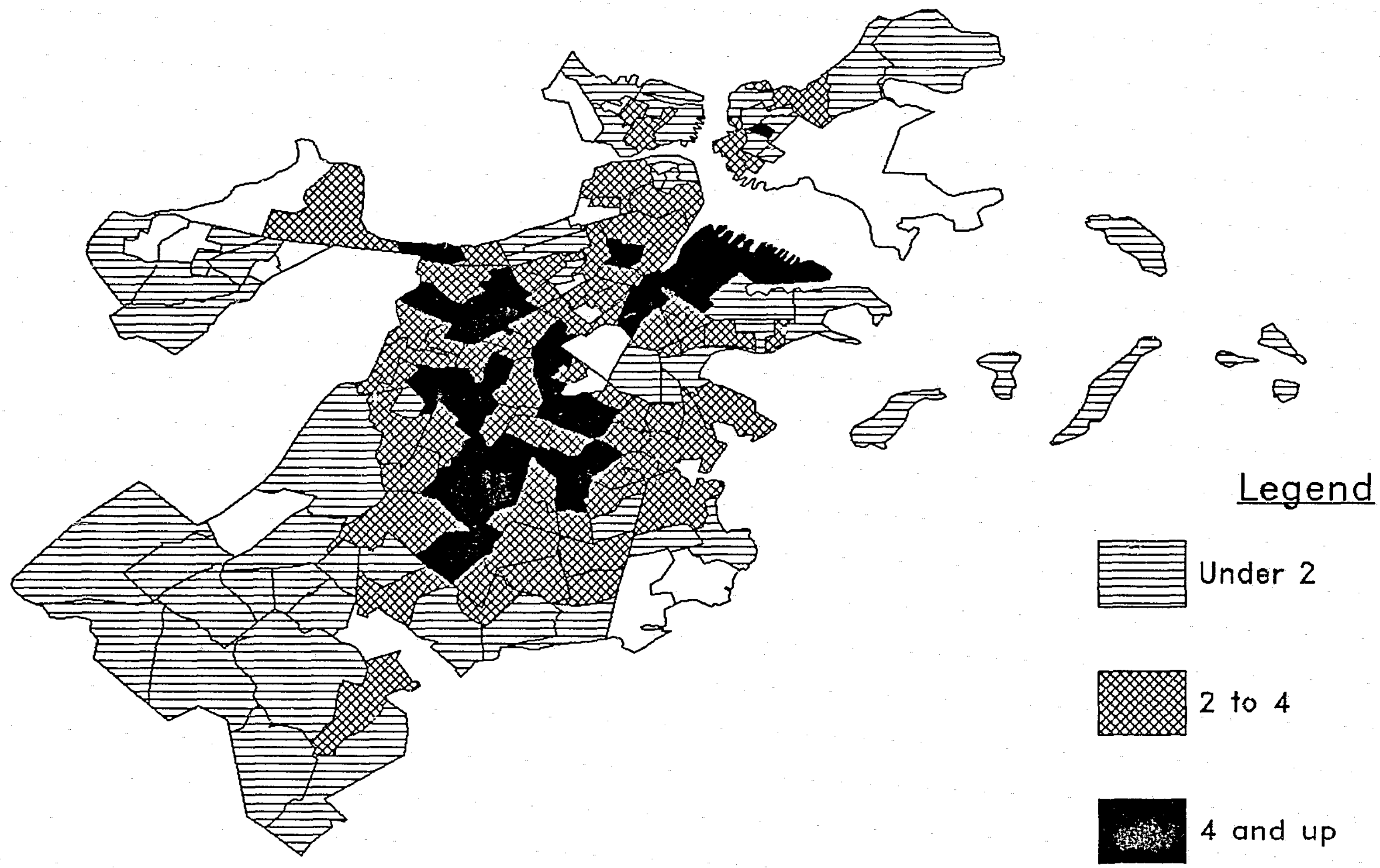
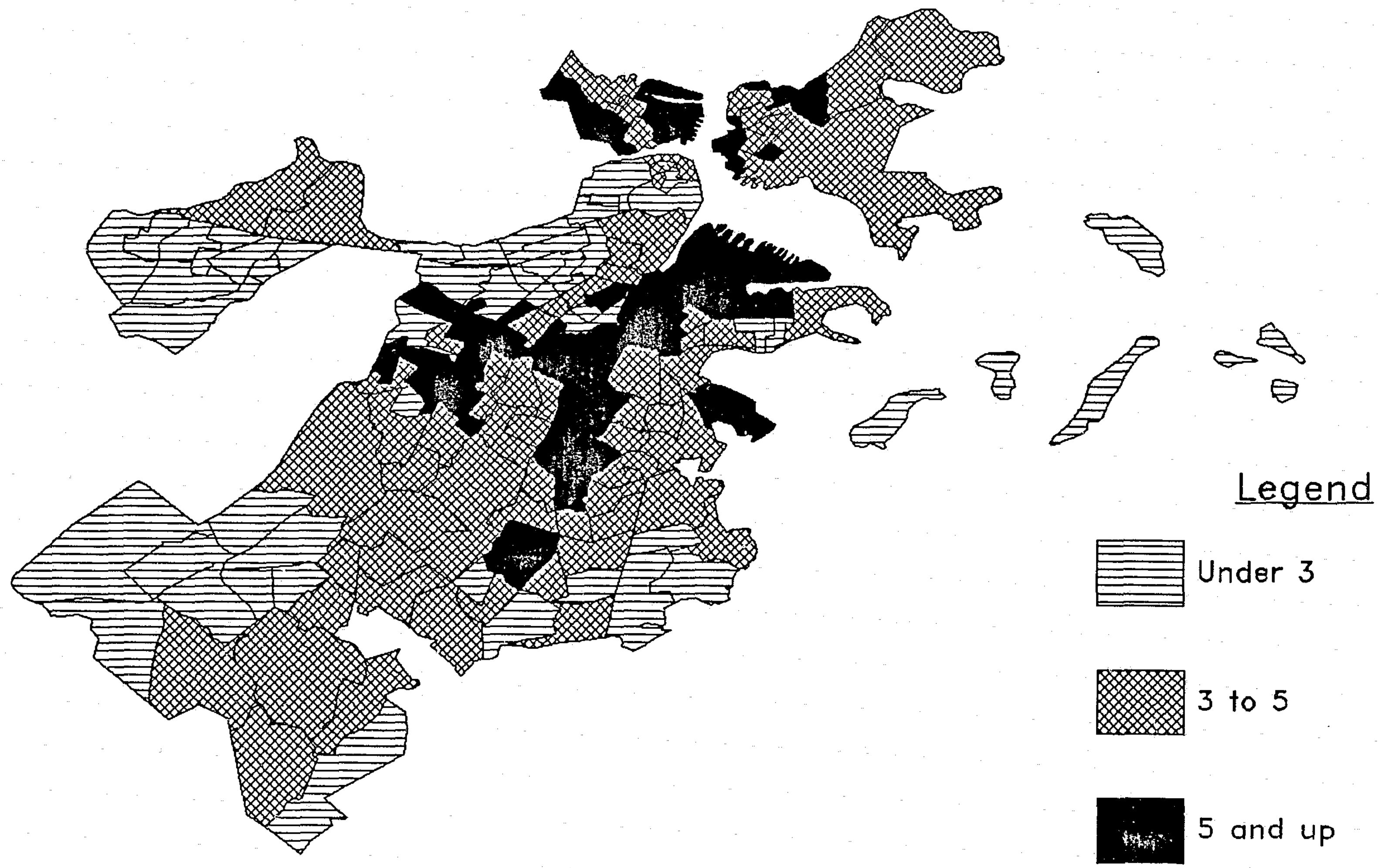


Figure 11: Log Automobile Arson Rate by Boston Tract



Appendix B: MFIRS Coding Form

DEPARTMENT OF PUBLIC SAFETY—DIVISION OF FIRE PREVENTION
1010 COMMONWEALTH AVENUE, BOSTON 02215

FP32

Fill In This Report In Your Own Words

Fire Department

MASSACHUSETTS INCIDENT REPORT

LAYOUT 2
1 ☐ Delete
2 ☐ Change

WHITE COPY - FIRE DEPARTMENT RECORD

COMPLETE ON ALL INCIDENTS

A 1	FDID	Incident No.	Exp. No.	Mo.	Day	Year	Day of the Week	Time Alarm	"On Scene"	"In Service"

B	CORRECT ADDRESS	No.	Dir.	Name	Type	Zip Code	Census Tract

C	Occupant Name	Telephone	Room or Apt.
D	Owner Name	Address	Telephone
E	Method of Alarm from Public	Type of Situation Found	

F	Type of Action Taken	Co. Inspection District	Shift	No. Alarms	Mutual Aid 1 ☐ Rec'd 2 ☐ Given
G	No. Fire Service Personnel Used at Scene	No. Engines Used at Scene	No. Aerial Apparatus Used at Scene	No. Other Vehicles Used at Scene	

COMPLETE IF CASUALTY OR FIRE

H 2	No. Incident-related Injuries* — Fire Service / Others	No. Incident-related Fatalities* — Fire Service / Others	Complex
I	Fixed Property Use	Mobile Property Type**	

COMPLETE IF FIRE

ALL IGNITIONS

J	Area of Fire Origin	Level of Fire Origin	Termination Stage
K	Equipment Involved in Ignition (if any)**	Form of Heat of Ignition	
L	Type of Material Ignited	Form of Material Ignited	Ignition Factor

FOR STRUCTURE FIRE ONLY

M	Structure Type	Construction Type	Construction Method
N	Extent of Flame Damage	Extent of Smoke Damage	Extent of Water Damage
O	Extent of Fire Control Damage	Detector Performance	Sprinkler Performance
P	IF FLAME SPREAD BEYOND ROOM OF ORIGIN — Type of Material Generating Most Flame	Avenue of Flame Travel	
Q	IF SMOKE SPREAD BEYOND ROOM OF ORIGIN — Type of Material Generating Most Smoke	Avenue of Smoke Travel	

ALL FIRES

R	Method of Extinguishment		
S	Estimated Total Dollar Loss	Property Damage Classification	Time from Alarm to Agent Application
S1	Total Insurance on Building-Vehicle	Damage to Building-Vehicle	
S2	Damage to contents-Vehicle	Insurance Paid	
S3	Public Adjuster	Insurance Agency	Insurance Company

CANARY COPY - FIRE MARSHAL'S OFFICE

Entries contained in this report are intended for the sole use of the State Fire Marshal. Estimations and evaluations made herein represent "most likely" and "most probable" cause and effect. Any representation as to the validity or accuracy of reported conditions outside the State Fire Marshal's office, is neither intended nor implied.

COMPLETE ON ALL INCIDENTS

Officer in Charge (Name, Position, Assignment) Date

Member Making Report (If Different from Above) Date

FIRE MARSHAL
F M ___ 1 ☐ Yes 2 ☐ No

** Complete Below

☐ Check box if remarks are made on reverse side.

U 3	If Mobile Property	Year	Make	Model	Serial No.	License No. (if any)
V 4	If Equipment Involved In Ignition	Year	Make	Model	Serial No.	Voltage (if any)

All Information Above Must Be Reported Within 48 Hours After Fire.

List name, age, sex, and description of injury for each casualty on form FP33