

# UC Santa Barbara

## NCGIA Technical Reports

### Title

Environmental Equity in Los Angeles (93-6)

### Permalink

<https://escholarship.org/uc/item/6966d5t3>

### Author

Burke, Laretta M.

### Publication Date

1993-07-01

# **NCGIA**

## **National Center for Geographic Information and Analysis**

### **Environmental Equity in Los Angeles**

by  
Lauretta M. Burke  
University of California, Santa Barbara

Technical Report 93-6  
July 1993

**Simonett Center for Spatial Analysis  
University of California**  
35 10 Phelps Hall  
Santa Barbara, CA 93106-4060  
Office (805) 893-8224  
Fax (805) 893-8617  
ncgia@ncgia.ucsb.edu

**State University of New York**  
301 Wilkeson Quad, Box 610023  
Buffalo NY 14261-0001  
Office (716) 645-2545  
Fax (716) 645-5957  
ncgia@ubvms.cc.buffalo.edu

**University of Maine**  
348 Boardman Hall  
Orono ME 04469-5711  
Office (207) 581-2149  
Fax (207) 581-2206  
ncgia@spatial.maine.edu

## **ACKNOWLEDGEMENTS**

I am grateful for the support, guidance and criticism provided by Mike Goodchild, Luc Anselin and Joel Michaelsen over the course of this thesis. The research environment of the NCGIA was an invaluable asset to this analysis. I am indebted to Uwe Deichmann for his encouragement, suggestions, and technical assistance. Many thanks, also, to Diane Schweizer and Rusty Dodson, without whom this thesis would not have been as colorful or as snazzy. I am appreciative of the motivation offered by Fran McPoland, who enthusiastically publicized the analysis and the value of color maps.

## **ABSTRACT**

Environmental Equity in Los Angeles

by

Lauretta Marie Burke

The issue of environmental equity refers to whether the burden of environmental pollution is borne evenly across society, and more specifically, to whether racial minority and low-income communities bear a disproportionate share of exposure to pollution and environmental risk. Previous studies on the subject, conducted at a variety of scales, offer conflicting evidence regarding the importance of race and income in the relationship with environmental hazards. As race and income are highly correlated, the purpose of this analysis is to determine the significance of race in the relationship with environmental pollution when the effects of other important variables, such as income, have been removed.

In a case study for Los Angeles, the relationship between industrial facilities emitting toxic chemicals and demographic variables are examined at the census-tract-level of aggregation. Several exploratory data analysis and linear modeling techniques were implemented using the ARC/INFO GIS and S statistical software. Results suggest that race is a significant factor in this relationship. A detailed examination of the processes which have resulted in the current socioenvironmental landscape was conducted for three smaller study areas within Los Angeles. Policy issues related to the overall analysis are discussed.

## Table of Contents

### **Chapter 1. The Equity Issue**

Introduction

The Environmental Justice Movement

Environmental Discrimination Cited

Birth of a Movement

Summary and Comparison of Major Studies

1987 United Church of Christ Report

1992 EPA Report

National Law Journal

Summary

### **Chapter 2. Study Purpose and Methodology**

The Need for this Analysis

Study Purpose

Study Hypothesis and Assumptions

Data Sources

Data Accuracy Issues

Census Data

Toxic Release Inventory (TRI) Data

Census Tracts and TRI Facility Locations Combined

Methods of Analysis

1) Bivariate Mapping Analysis

2) Modeling the Number of TRI Facilities in a Tract:  
Generalized Linear Modeling to a Poisson Distribution

3) Modeling Presence or Absence of Facilities in a  
Census Tract: Logit Analysis

The Problem of Spatial Dependence

General Limitations of the Analysis

Time

Space

Scale Issues

Summary

### **Chapter 3. Background on Los Angeles County**

The Physical Landscape of Los Angeles County

The Industrial Landscape - TRI facilities within Los Angeles County

The Socioeconomic Landscape - Census Data for Los Angeles County

Racial / Ethnic Population Patterns

Visual Analysis

Minority Percentage and TRI Facility Locations

Per Capita Income and TRI Facility Locations

Population Density and TRI Facility Locations

Racial / Ethnic Composition and TRI Facility Locations  
Summary

Chapter 4. Analysis Results

Introduction

Method 1) Categorical Partitioning and Bivariate Mapping

Method 2) Modeling the Number of TRI Facilities in a Tract: Generalized Linear Modeling to a Poisson Distribution

Method 3) Modeling Presence or Absence of Facilities in a Census Tract: Logit Analysis  
Spatial Dependence

Summary

**Chapter 5. Neighborhood Analysis**

Detailed Study Purpose and Methodology

Method for Neighborhood Selection

Research Methods

Results of Neighborhood Analysis

**Chapter 6. Research Summary and Conclusions**

Policy Issues

Further Research

## **List of Tables**

### **Chapter 2**

Table 1 - Estimates of Census Undercount

### **Chapter 3**

Table 2 - Population by Race and Hispanic Origin

Table 3 - Summary Statistics for 1610 Census Tracts in L.A. County

Table 4 - Correlations Between Explanatory Variables

Table 5 - Correlations Between Dependent and Exploratory Variables

Table 6 - Summary Statistics by Racial/Ethnic Plurality

### **Chapter 4**

Table 7 - Model 1 Results

Table 8 - Summary of Three 2-Independent Variable Models

Table 9 - Sensitivity Analysis Summary

Table 10 - Model 3b Results

Table 11 - Model 4a Results

Table 12 - Summary of 4 Category Partitioning

Table 13 - Model 4b Results

Table 14 - Comparison of all Poisson Models

Table 15 - Logit Model Results

Table 16 - Prediction Success for Logit Model (0.50 Threshold)

Table 17 - Prediction Success for Logit Model (0.29 Threshold)

Table 18 - Comparison of Coefficients from the Full Logit Model with those Estimated through the Resampling

## List of Figures

### Chapter 2

Figure 1 - Study Hypothesis

### Chapter 3

Figure 2 - Physical Features of Los Angeles County

Figure 3 - TRI Facility Locations in Los Angeles

Figure 4 - Tracts Excluded from the Analysis

Figure 5 - Population Patterns in Los Angeles

Figure 6 - Minority Percentage and TRI Facility Locations

Figure 7 - Per Capita Income and TRI Facility Locations

Figure 8 - Population Density and TRI Facility Locations

Figure 9 - Race, Ethnicity and TRI Facilities in Los Angeles

### Chapter 4

Figure 10 - Bivariate Mapping of Minority Pct. and Per Capita Income

Figure 11 - Predicted TRI Facilities as a function of Per Capita Income (by Hispanic Grouping)

Figure 12 - Predicted TRI Facilities as a function of Per Capita Income (by Racial / Ethnic Grouping)

Figure 13 - Predicted Values from the Poisson Model

Figure 14 - Residuals from the Poisson Model

Figure 15 - Prediction Success for the Logit Model (@ 0.29)

Figure 16 - Deviance Residuals from Logit Model

Figure 17 - Logit Model Coefficients Estimated through Resampling



## **List of Appendices**

Appendix A - Map of Places in Los Angeles

Appendix B - The Distribution of the Main Variables

Appendix C - TRI Industrial Information for Los Angeles  
(Industrial Categories and Frequency of Occurrence)

Appendix D - Industrial Neighborhood Questionnaire

## Chapter 1. The Equity Issue

### Introduction

Environmental equity has recently emerged as an important issue both in the media and within the Federal government. The issue of environmental equity refers to whether people bear the burden of our technological advances - environmental pollution - evenly across society. In particular, the issue addresses whether or not racial minority and low-income communities bear a disproportionate share of exposure to pollution and environmental risk.

There are several terms in use in the literature which refer to the issue of the fairness of the distribution of environmental hazards - environmental equity, environmental justice and environmental racism. The three terms refer to the same basic issue, but differ somewhat in terms of tone and conviction.

### The Language of Equity

The term **environmental equity** refers to the issue and the actuality of the fairness of the distribution of environmental hazard with regard to the population. This is an appropriate starting point for scientific and policy inquiry. It implies no bias or presumptions. It is the point from which one can ask whether the distribution of environmental hazards is equitable across the population, or with regard to race, ethnicity, or income. The U.S. Environmental Protection Agency (EPA) chooses to use the term environmental equity to refer to this issue, and the broader issue of disproportionate risk on any population group, as defined by gender, age, income, or race (EPA, 1992).

**Environmental justice** is the political movement aimed at achieving environmental equity. The terms environmental equity and environmental justice are often used interchangeably, and environmental equity is sometimes referred to as a movement. A tenet of the environmental justice movement is that current inequities in the distribution of pollution are due, at least in part, to racial discrimination (Taylor, 1992). The environmental justice movement is largely a grassroots movement.

The term **environmental racism** was coined about a decade ago, when a series of protests failed to halt the siting of a hazardous waste landfill in predominantly Black Warren County, North Carolina (Satchell, 1992). The term is often used synonymously with environmental justice, in reference to the movement. But, the term itself contains the conclusion that inequities exist and a declaration of the cause of the inequities. The term is, of course, highly political, and is often used as a rallying cry to highlight inequities and motivate participation in the movement.

### The Value of Equity

Research on environmental quality in minority communities has been minimal and participation by minorities in the mainstream environmental movement has been sparse (Srnith, 1974, Bullard, 1990). The growing environmental justice movement has potential benefits in several areas. Firstly, participation by all segments of the population in the environmental movement is desirable, as this should effect a broader and better understanding of the nature of our environmental problems; how these problems are perceived; and the implications of proposed environmental policies. Diversity has great value, and this certainly applies to the area of environmental decision-making. Additionally, the movement is resulting in increased knowledge about environmental problems and solutions. This should ultimately result in better decisions supported by a greater consensus. The second benefit of the environmental justice movement is moral and legal in nature - environmental protection and the distribution of environmental hazards should be equitable across society. Federally funded programs (including environmental protection) are mandated to be non-discriminatory by the Civil Rights Act of 1964. The environmental justice movement presses the environmental decision making process to give issues of fairness consideration.

Although studies on the subject of the fairness of the distribution of environmental hazards were conducted as early as 1971 (CEQ, 1971), the issue of environmental equity and the related environmental justice movement have only gained attention and momentum within the past few years. The movement was slow to spawn, largely the result of history, timing, and limited resources, until a few key events in the 1980's provided the catalyst which generated the movement.

### The Environmental Justice Movement

Historically, participation by racial minorities in mainstream environmental organizations and the environmental movement has been slight. Within the U.S., the environmental movement gained momentum during the 1960's, at the same time that the civil rights movement was becoming a major force in national politics. Although there are some similarities between the two movements such as reliance on public participation, protest and legal action, the two movements have operated in relative isolation.

The environmental movement was closely aligned with, and largely grew out of, many old and well-established national conservation organizations. These membership organizations were historically White and middle to upper-middle class dominated (Smith, 1974). Although the environmental movement had the broad goals of pollution abatement and general improvement of environmental quality, there was also a large emphasis on the conservation of pristine areas, wildlife and scenic wonders (CEQ, 1971, Adams, 1992).

The civil rights movement, on the other hand, has deep roots in the Black church and other community organizations. The focus of this movement was social justice - political empowerment, equal education, fair employment practices and open housing for all people. Many social justice activists saw the environmental movement as a tactic to divert attention and resources away from the more important issue of the day - racism (Bullard, 1990).

In 1971, the President's Council on Environmental Quality issued its second annual report, which provided the first published information about inequities in the distribution of environmental hazards in the U.S. (Mohai & Bryant, 1992c). Within a chapter of the report dedicated to environmental problems of the inner-city, the report presented striking evidence of the correlation between poverty areas and high degrees of air pollution for three U.S. cities (CEQ, 1971). Although environmental inequities were documented by the CEQ report and nine other studies on the subject published during the 1970's (Mohai & Bryant, 1992a), the issue was not seriously adopted by either the mainstream environmental movement or the civil rights movement. That same year, in a hearing before the U.S. Civil Rights Commission, EPA Administrator Ruckelshaus testified that the EPA was a "technical and scientific agency, not equipped to judge disparate impacts on minority communities due to pollution" (Lavelle, 1992).

Some advocates of environmental justice suggest that concern about the environment cuts across all racial, ethnic and income strata, although participation in organizations dedicated to environmental protection has not (Bullard, 1990). Others suggest that due to the distractions caused by poverty and its accompanying ills, the traditional environmental objectives of clean air and water and the preservation of wilderness are not the central concerns of most inner city poor (CEQ, 1971). The type of issues adopted by environmental organizations has greatly influenced membership. As most mainstream environmental organizations have not focused on minority or inner-city pollution issues, minority participation in these organizations has been meager. Additionally, minorities are underrepresented in professional positions within national environmental organizations (Smith, 1974, Bullard, 1990).

### **Environmental Discrimination Cited**

Mainstream environmentalism and the vast majority of environmental policy enacted in this country over the past 25 years have ignored the issue of environmental equity, largely the result of focusing on more evident environmental problems. The mainstream environmental movement has focused on the improvement in overall environmental quality and the preservation of rare and pristine areas, at the expense of more consideration of the fairness of these policies. Many policies aimed at improving environmental quality have been regressive, meaning the policies place a disproportionate burden on low-income individuals (Smith, 1974).

A focus of environmental equity (or the environmental justice movement) has been the fairness of siting decisions for locally unwanted land uses (LULU's). LULU's are typically facilities which pollute or carry some sort of environmental risk, such as landfills. In his 1990 book, *Dumping in Dixie - Race, Class, and Environmental Quality*, Robert Bullard explores the thesis that "Black communities, because of their economic and political vulnerability have been routinely targeted for the siting of noxious facilities, locally unwanted land uses, and environmental hazards". Bullard proposes that the combination of limited residential options for Blacks; institutional barriers; and facility siting decisions have contributed to high concentrations of Blacks living in neighborhoods which are undesirable in terms of environmental quality. Facility siting often follows the "path of least resistance", meaning that Blacks and poor communities have often been targeted as facility locations because of their likely lack of political opposition (Bullard, 1990).

Many factors contribute to the siting of polluting industries in poor and minority neighborhoods. Land values and labor costs tend to be lower in poor neighborhoods, thus attracting industries seeking to reduce the cost of doing business (Mohai & Bryant, 1992c). Additionally, these communities often have the least ability to oppose the siting of an undesirable facility. In low-income communities, there are often more pressing problems related to poverty and there is sometimes the additional problem of language differences. Mohai & Bryant (1992c) make the argument that low-income and minority communities tend to be unaware of policy decisions affecting them; are not organized; lack the resources (time, money, contacts, and knowledge of the political system) for taking political actions; and tend to be underrepresented on governing bodies.

Beyond the limited ability of low-income and minority communities to resist the siting of unwanted facilities, several environmental equity scholars argue that some communities provide incentives which attract locally unwanted land uses. The trade-off is between jobs and a tax base versus environmental risk. Bullard (1990) argues that in order to improve the economic conditions of their constituents, many civil rights, business and political leaders provided incentives for industries by relaxing enforcement of

pollution standards and environmental regulations and by sometimes ignoring known violations. In many cases, facility siting occurs with little input from the local community, but is justified through the provision of jobs.

In some cases surrounding the siting of extremely undesirable facilities, such as landfills, the community is offered compensation for accepting the facility (Morell, 1987). This is an extreme example of the trade-off between economics and environmental risk. Under such conditions, the most economically disadvantaged communities would benefit the most from accepting the facility, as money would have higher utility in this context. Depending upon the evaluation criteria used, such deals would result in apparent environmental inequity.

An additional criticism by the environmental justice movement of mainstream environmentalism relates to the "not in my backyard" or NIMBY phenomenon. Some environmental justice supporters claim that environmentalists have been effective in defending their own backyards, without consideration of whose backyards the facilities then land in (Taylor, 1992). Opposition to facility siting in clean, pristine, middle-class and upper-class neighborhoods has sometimes pushed unwanted facilities into the neighborhoods of the less attentive or less politically powerful.

### **Birth of a Movement**

Beginning in the mid 1980's, minorities became increasingly active in the environmental justice movement. This activism resulted from the discovery of toxins in many minority communities; the publication of research linking race, class, and disproportionate exposure to toxins; the revelation that minority communities were targeted for the siting of hazardous facilities; and the organization of conferences and other meetings on the subject of environmental justice (Taylor, 1992).

There were several key events which provided momentum for the environmental justice movement. The following provides a brief synopsis of those events. Details of individual studies and reports are provided later, under *Summary and Comparison of Major Studies*.

The first major event to focus national attention on the issue of environmental pollution and race occurred in Warren County, NC in 1982. The predominantly Black community of Afton, NC was selected as a disposal site for PCB-contaminated soil, although the site was not physically well-suited as a disposal site (Bullard, 1990). Protests erupted, similar to the civil rights demonstrations of the 1960's, in which over 500 protesters were arrested (Coyle, 1992). Although the protests did not succeed in halting the project, they did succeed in focusing national attention on the issue.

The protests led to a 1983 study by the General Accounting Office (GAO) of hazardous waste landfill siting in the South. The GAO study examined the racial and socioeconomic characteristics of communities<sup>1</sup> surrounding four major off-site hazardous waste landfills. Though Blacks comprise about 20 percent of the overall population of the region, the communities with these landfills ranged from 38 to 90 percent Black (GAO, 1983)

Also in 1983, one of the first conferences addressing environmental equity was held - the Urban Environment Conference on Toxics and Minorities. The meeting served as a forum for dialogue among organized labor, minorities, and environmental groups (Bullard, 1990). In 1985, the National Council of Churches formed an Eco-Justice working group.

In 1987, the United Church of Christ's (UCC) Commission on Racial Justice (CRJ) released a landmark report on environmental equity - *Toxic Waste and Race in the United States*. The report represented the first systematic national analysis of the relationship between race, income, and the distribution of hazardous waste sites (Mohai & Bryant, 1992a). The study found that, on average, the higher the number and greater the severity of commercial hazardous waste facilities in a community, the higher the minority percentage of the population. The study also concluded that race is the single best predictor of where commercial hazardous waste facilities are located, even when the effect of income is taken into account (UCC, 1987).

The 1987 United Church of Christ report was a critical event for the environmental justice movement. It provided empirical evidence documenting environmental inequity and attributed it to race. At the time of the report's release, Dr. Benjamin Chavis Jr., Executive Director of the Commission on Racial Justice, termed the racial biases in the location of these facilities "environmental racism" (Mohai and Bryant, 1992c).

In 1990, two professors at the University of Michigan School of Natural Resources, Paul Mohai and Bunyan Bryant, convened the Michigan Conference on Race and the Incidence of Environmental Hazards. The meeting of about 20 people represented the first such meeting where a majority of the presenters were people of color. Several conference participants formed the Michigan

---

<sup>1</sup> Within the GAO study, a "community" was defined by all census areas (tracts) which had a border within 4 miles of the site.

Coalition, to continue work on this subject. In September of 1990, the Michigan Coalition met with EPA Administrator William Reilly (Mohai and Bryant, 1992b).

During 1990, EPA initiated several activities on environmental equity. EPA's Science Advisory Board published a report, *Reducing Risk: Setting Priorities and Strategies for Environmental Protection*. The report urged EPA to target the most promising opportunities for reducing the most serious risks to human health and the environment (Reilly, 1992), which should have effected a focus on urban problems. In July of 1990, Reilly formed the Environmental Equity Workgroup to review the evidence that racial minority and low-income populations bear a higher environmental risk burden than the general population. The final report of the Workgroup was released in July of 1992.

Reflecting the growing interest in environmental justice and minority participation in the environmental field, in April of 1990, a conference was convened at Howard University on environmental opportunities for minorities in the environmental field. Additionally, in 1990, Greenpeace released a report on environmental justice - *Playing with Fire*, and Bullard's *Dumping in Dixie* was published, documenting environmental discrimination in the South.

In October of 1991, the "First National People of Color Environmental Leadership Summit" was held on Capitol Hill and was attended by about 500 representatives of community groups and many representatives of mainstream environmental organizations (Coyle, 1992). Three months later, the University of Michigan's Law School convened a Symposium on Race, Poverty and the Environment.

In 1992, EPA released two publications on environmental equity and created an Office of Environmental Equity. The March/April issue of the EPA Journal was dedicated to the issue of environmental equity, entitled "Environmental Protection - Has It Been Fair?". The issue contains articles by outspoken researchers and activists for environmental justice, providing a broad background on the movement; on issues of race, poverty and the environment; on particular groups at risk including farm workers, Native Americans and inner-city children, and summarizes EPA's position on the issue.

In July of the same year, EPA released the report of its Workgroup on equity - *Environmental Equity - Reducing Risk for All Communities*. Although commended for addressing the issue at all, the EPA report has been widely criticized for being too cautious and conservative; exaggerating uncertainties surrounding the issue; not directly acknowledging inequities related to race; and for making recommendations which were too general and weak and which lacked any time-table for action (EPA, 1992).

In September of 1992, the National Law Journal published the results of its own comprehensive analysis of every environmental lawsuit concluded in the U.S. within the past seven years. The publication presented extensive evidence that federal environmental protection laws have not been implemented equally with regard to race (Lavelle & Coyle, 1992).

From the infrequent workshops and studies of the early 1970's, the environmental justice movement gained size, strength, organization and effectiveness to become an important influence in environmental decision-making today. The environmental justice movement has gained the attention and support of the minority community in a way the mainstream environmental movement never has. Through linking civil rights with environmental rights, re-introducing civil rights campaign strategies into environmental campaigns, and addressing issues of concern to the minority community, the movement has created a structure in which minorities feel comfortable participating (Taylor, 1992).

The main "players" in environmental justice have been established. The bulk of the environmental justice movement will be through grassroots organizations, more aligned with and structured like the social justice movement than the mainstream environmental movement. At the same time, many mainstream environmental organizations are attempting to include environmental justice concerns and forge alliances with these grassroots organizations (Suro, 1993, MacLachlan, 1992). With these two types of organizations will lie the vast majority of the environmental justice movement for the near future. While the EPA report recommends that the Agency "increase the priority that it gives to issues of environmental equity", it will take time for the bureaucracy to develop and introduce policies to include the consideration of equity within its programs.

As part of this movement, there has been a groundswell of grassroots organizations working on environmental justice. Within Los Angeles alone, several groups are having significant effects on environmental and land-use decision making. Concerned Citizens of South Central, one of the first African American environmental groups, prevented the siting of a toxic waste incinerator in South Central in 1985 (Kay, 1991). Concerned Citizens of South Central later aligned with a Hispanic group, the Mothers of East Los Angeles, to prevent the siting of the same incinerator in Vernon, next to the barrio of East Los Angeles. Together, the two groups filed a lawsuit which halted the project, and resulted in the introduction of legislation to require environmental impact statements to include the racial and ethnic makeup of the community (Kay, 1991).

Another group in Los Angeles, the Labor/Community Watchdog, is a multi-racial, grassroots, working-class organization focusing on air pollution in Los Angeles. The group uses a combination of education, scientific analysis and public protest to affect public policies, both aimed at improving air quality and avoiding policies which disproportionately affect low-income and minority individuals (Novick, 1992).

With the history of the environmental justice movement in context, the following section provides a comparison and more detail on several of the most important studies on environmental equity.

### **Summary and Comparison of Major Studies**

Over the past two decades there have been many studies and reports examining the issue of environmental equity. The studies have examined a variety of types of environmental hazards; assessed the hazard's relationship to an assortment of demographic indices; employed a wide range of analysis methods; and performed these analyses at a variety of scales.

There are many statistical tests and techniques available to the researcher performing such analysis, including significance testing of regression coefficients; discriminant analysis; difference of means tests; analysis of variance; matched-pairs tests; and several non-parametric tests. As many different statistical techniques have been applied within these studies, there is no common metric for comparison between or among studies. In most studies which evaluated the relationship between race or income and an environmental hazard, a test of significance of that variable was applied<sup>2</sup>. In studies which examined the relationship between both race and income and an environmental hazard, different techniques have been applied to evaluate the relative importance of those variables. The following overview of studies of environmental equity draws comparisons based on the studies' conclusions. In many cases, information was obtained from a secondary source and the particular statistical tests applied are not known. (A detailed review of three major studies follows this overview and comparison).

### **Overview**

Mohai and Bryant (1992a) conducted a review of 15 studies conducted since 1971 which provide "systematic information about the social distribution of environmental hazards." The studies examine the relationship of the hazards to income, race, or in some cases, both. The purpose of their review was to document the wide range of studies which have been performed; determine whether the evidence from these studies demonstrates a consistent pattern of environmental racism"; and assess the relative influence of income and race on the distribution of pollution. The review constitutes the best bibliography of equity studies to date.

The 15 studies varied considerably in terms of scope. About half the studies focused on a single urban area, while the rest focused on a region, a collection of urban areas, or were national in scope. (Several studies were performed both for the nation and for urban areas only.) Eleven of the studies examined the distribution of air pollution, four examined the distribution of solid or hazardous waste facilities, while one focused on toxic fish consumption. The studies were performed at a variety of scales and the scale and statistical methods cannot always be determined from the Mohai and Bryant review.

Of the four studies of the distribution of waste disposal sites, one focused on a single urban area (Houston); one on a set of urban areas; the 1983 GAO study focused on a region (the South); and the 1987 UCC study was national in scope. The three studies which evaluated the importance of income as a factor affecting the distribution of sites found it to be a significant factor. (The Houston study only looked at race.) All four studies evaluated the importance of race as a factor and found it to be significant. Only the 1987 UCC study evaluated the relative importance of income and race. This study, which was performed at the zip-code-level, concluded that race was the more significant factor in the relationship with occurrence of hazardous waste facilities. This study is discussed in detail later.

Of the ten studies of the distribution of air pollution, four focused on a single urban area; three on a set of urban areas; one was national in scope; and two performed studies both at the national level and upon urban areas only. Six of seven studies which evaluated whether air pollution is distributed equitably by race found race to be a significant variable in the relationship. (The study which found that race was not a significant factor focused on a single urban area.) All ten studies on air pollution evaluated the importance of income in this relationship. Every study which was conducted only on urban areas found income to be a significant factor, while those studies conducted on both rural and urban areas together found that income was not significant. This discrepancy is probably the result of both income and air pollution levels being higher in urban areas.

In six of the studies on the distribution of air pollution, the relative importance of both income and race was evaluated. Three of these studies found income to have the stronger relationship with air pollution, while three studies found race to have the stronger relationship.

---

<sup>2</sup> Significance tests typically used a 95% confidence level, though some used 90% or 99%.

A 1978 study by Asch & Seneca performed an analysis at the census-tract-level for a study of three urban areas, and at the level of city for a study of all urban areas in 23 states. The two analyses produced similar conclusions- that the relationship between median family-income levels and air pollution is stronger than the relationship between the minority percentage and air pollution. A study on the distribution of damage from air pollution by Gianessi et. al. (1979) presents a contrasting result. The national study found that minorities were much more likely than Whites to suffer greater damage from air pollution at all income levels.

Combined these studies present strong evidence of inequities in the distribution of environmental hazards, both by income-level and by race. It is not conclusive, though, whether race or income-level is the more important factor in this relationship.

The 1987 report by the United Church of Christ is regarded by many in the environmental justice movement as the conclusive empirical evidence required to raise public awareness about the disproportionate burden of environmental hazards on minorities. The results of this report will be examined in detail, along with two other important reports on environmental equity - the 1992 report by EPA's Environmental Equity Workgroup, and the 1992 study by the National Law Journal.

### **1987 United Church of Christ Report**

*Toxic Wastes and Race in the United States*, published in 1987 by the UCC Commission on Racial Justice was the first, and perhaps only, report to comprehensively document the presence of hazardous wastes in racial and ethnic communities throughout the U.S.. The report presents findings from two cross-sectional studies on demographic patterns associated with commercial hazardous waste facilities and uncontrolled toxic waste sites. A major objective of the study was to determine whether the racial composition of communities with and without commercial hazardous waste facilities was significantly different, in addition to determining which factors are important in this relationship.

One study focused on 415 operating, off-site commercial hazardous waste treatment, storage and disposal facilities, obtained from the EPA Hazardous Waste Data Management System. The methodological approach involved assigning all communities (as defined by 5-digit zip codes) to one of four categories, depending upon the number, type and size of facilities within the community.<sup>3</sup> Five variables were evaluated in this study - minority percentage of the population; mean household income; mean value of owner-occupied home; pounds of hazardous waste generated per person; and number of uncontrolled toxic waste sites per 1,000 persons in the community. The analysis used five different statistical tests in order to "derive findings that are independent of any single technique." The techniques used were: discriminant analysis; difference of means test; matched-pairs test; and non-parametric versions of the difference of means and matched-pairs tests.

The second study in the report focused on the communities surrounding over 18,000 uncontrolled toxic waste sites from EPA's Comprehensive Environmental Response, Compensation, and Liability Act Information System (CERCLIS). This study was descriptive in nature, aimed at quantifying the number of persons from different racial and ethnic groups living in communities with uncontrolled toxic waste sites.

Regarding communities with commercial hazardous waste facilities, major findings of the first study included:

- 1) "Race proved to be the most significant among variables tested in association with the location of commercial hazardous waste facilities and this represented a consistent national pattern";
- 2) On average, the greater the number and the higher the severity of commercial hazardous waste facilities or landfills within a community, the higher the minority percentage of the population;
- 3) "Although socioeconomic status appeared to play an important role in the location of commercial hazardous waste facilities, race still proved to be more significant".

The findings of the first study seem well justified by the statistical results presented in the report. The results of one test, though, make it less indisputable that race is clearly the most significant factor. In an analysis of operating commercial hazardous waste facilities, a matched pairs test was applied comparing zip code to county-level facility occurrence within each of the 10 EPA Regions (UCC Table B-3). In this test, mean household income and mean value of owner occupied home were found to be significant in more of the ten EPA Regions than was the minority percentage.

---

<sup>3</sup> Group 1 comprised communities without any such facilities; communities in Group 2 had one facility which was not a landfill; communities in Group 3 have a hazardous waste landfill, but not one of the five largest; while communities in Group 4 have several facilities or one of the five largest landfills.

Some of the findings of the second study in the report regarding communities with uncontrolled toxic waste sites were presented in an alarming, yet inconclusive fashion, where the frame of reference for evaluation was absent. One can accurately derive from the report's findings that Blacks are heavily overrepresented in cities with the largest number of uncontrolled toxic waste sites. But, as the report states, "the presence of uncontrolled toxic waste sites was highly pervasive... with more than half of the U. S. population residing in communities with sites". Against this backdrop, the findings that "three of every five Black and Hispanic Americans live in communities with uncontrolled toxic waste sites" and "approximately half of all Asian/Pacific Islanders and American Indians" live in such communities are not outstanding. Further examination of this study's results (UCC Table C-1) shows that the percentage of each racial and ethnic group living in communities with uncontrolled toxic waste sites was roughly equal. Of the total population, 54.2% live in communities with sites, while 57.1% of all Blacks, 56.6% of all Hispanics, 53.6% of all Whites, 52.8% of all Asians and Pacific Islanders, and 46.4% of all Native Americans live in communities with sites.

In examining the results of the UCC report, a very different picture of the magnitude of environmental inequity emerges depending upon the environmental hazard assessed and the statistical technique applied. In an examination of communities (zip codes) with operating commercial hazardous waste facilities, one finds that the minority percentage of the population is twice as large in communities with facilities as those without. The statistical techniques applied in this study suggest that race is a more important factor than income in this relationship. When one examines the population living in communities with uncontrolled toxic waste sites, however, the disparity between the minority and White percentages of the population is slight -56.3 versus 53.6 percent. This difference could easily be attributed to the income effects.

The overall differences between these two studies might result from the facility types and time of siting. Uncontrolled toxic waste sites are much more pervasive, and the majority may have been sited prior to the NIMBY phenomenon.

### **1992 EPA Report**

EPA's Environmental Equity Workgroup released its report, *Environmental Equity -Reducing Risk for all Communities*, in June of 1992. One of the major findings of the report is:

"Racial minority and low-income populations experience higher than average exposures to selected air pollutants, hazardous waste facilities, contaminated fish, and agricultural pesticides in the workplace. Exposure does not always result in immediate or acute health effects. High exposure and the possibility of chronic effects, are nevertheless a clear cause for health concerns."

This rather significant finding was surrounded by other findings emphasizing our current lack of knowledge on the environmental contribution to racial differences in disease and death rates, and the general lack of data on environmental health effects by race and income. The report also emphasizes the uncertainties surrounding cumulative and synergistic effects and multiple pathways of exposure.

Although the report cites the disproportionate average exposure to pollutants of low-income and minority persons, it does little to explain the cause, other than by stating that, "these differences are deeply rooted in many aspects of society, such as historical residence, politics, commerce, geography, local land-use decisions, and other socioeconomic factors that affect where people live and work."

The EPA report provides a cautious, conservative balance to the more assured conclusions of discrimination provided by much of the environmental justice literature. The report cites several studies which suggest that many, if not all of the racially-based differences in cancer rates can be explained by the effects of poverty. The report's emphasis on health effects results in less attention being paid to the more conclusive literature on exposure. Although EPA cites the 1983 GAO study and the 1987 UCC report, the EPA report concludes that more study of this issue is required to fully understand the associations of race, income and facility location.

The EPA report and creation of the Office of Environmental Equity are important acknowledgement that environmental risk and exposure to pollution are not distributed evenly across society. The body of literature reviewed by the EPA Workgroup surely supported more aggressive conclusions and recommendations than those adopted in the final report. But, this is, nevertheless, an important starting point for the consideration of environmental equity in national policy.

### **National Law Journal**

The September 21, 1992 issue of the National Law Journal (NLJ) presented results of an analysis conducted by the NLJ of every environmental lawsuit concluded within the U.S. in the last seven years. Key findings of the report show that penalties imposed for environmental law violations in minority areas are lower than those imposed for violations in largely White areas. (Areas are defined by five-digit zip codes.) With regard to hazardous waste law enforcement under the Resource Conservation and Recovery Act



(RCRA), average penalties in White areas were six times those in minority areas (\$336,000 versus \$55,000), while there was little disparity in the average penalties assessed between high and low-income areas (\$110,000 versus \$113,000).

A slightly different picture emerges when the comparison is made of penalties assessed under all federal air, water and waste laws. In this case, the disparity between penalties assessed to high and low-income areas is larger than the disparity across race. Average penalties in high income areas were 53% higher than in low-income areas (\$147,000 versus \$96,000). Average penalties in White areas were 46% higher than in minority communities (\$153,000 versus \$105,000).

The NLJ study also found that within the Superfund program, abandoned hazardous waste sites in minority areas take 20% longer to be placed on the National Priority List (NPL)<sup>4</sup> than those in White areas, and the cleanup method selected tended to be less stringent in minority than in White communities. The EPA claims that penalties assessed are based on the particular conditions at the site and on the ability of the legally responsible party(ies) to pay.

Although Blacks and Hispanics dominate in urban areas where pollution levels are high enough to violate the standards of the Clean Air Act, Whites have disproportionately benefitted from litigation under the Act. Of 352 cases brought under the Act in the past seven years, the population which benefitted was 79% White, 14% Black, and 8% Hispanic. This disparity could result largely from a lack of financial resources in minority communities. Litigation is an effective, but expensive, tool for environmental remediation (Lavelle & Coyle, 1992).

This extensive analysis and expose by the NLJ provides evidence of unequal protection for minority and low-income individuals under current implementation of environmental statutes.

### **Summary**

The studies and reports described here provide validation for the environmental justice movement. There are many more case studies, articles, and newspaper clippings which provide further evidence of a system of environmental planning, protection, and remediation which has not been equitable across society. The anecdotes about injustice are abundant. The systematic analysis of the subject is uneven, but none-the-less provides significant evidence of inequity. The analyses described above were performed at a variety of scales, using a wide range of methods, examining a variety of types of pollution. There is ample evidence that environmental hazards are distributed inequitably with regard to income and inequitably with regard to race. But, given the knowledge accumulated through these studies, it is not conclusive whether race or income-level is the more significant factor in the relationship with environmental hazards. Nor has it been shown, in any analytical way, what processes contribute to this inequity. Is the inequity propagated through low-income minority individuals locating in environmentally poor neighborhoods, through a middle-class White exodus from those neighborhoods, or through the intentional targeting of such communities for the siting of unwanted facilities?

The next chapter describes in detail the purpose and methodology of this analysis, which is aimed at answering some of these questions.

## **Chapter 2. Study Purpose and Methodology**

This chapter provides a detailed description of the purpose of this study and the methodologies employed. It begins with a discussion of the study's purpose and is followed by a description of the study hypothesis and assumptions; a discussion of data sources and accuracy issues; a description of the analytical methods applied and limitations of these methods.

### **The Need for this Analysis**

Previous studies and reports offer conflicting evidence regarding the importance of race as a factor in the relationship with the distribution of pollution or environmental hazards. Several studies (Kruvant, 1975; Burch, 1976; Asch & Seneca, 1978) conclude that economics is the more significant factor in the relationship with environmental pollution, while three national studies examining how racial and socioeconomic factors relate to the location of environmental hazards (Gianessi et al., 1979; UCC, 1987; Gelobter, 1992) conclude that race is the more significant factor (Mohai & Bryant, 1992a).

In coarse-scale analysis, such as that employed in the national studies, one must be careful that the scale of analysis does not make results appear significant which are not, or create a relationship where there is none. This raises the issue of two related problems in geographical analysis - the modifiable areal unit problem (Johnson, 1984, Openshaw, 1984b) and the ecological fallacy (Robinson, 1950, Openshaw, 1984a). The ecological fallacy refers to the inappropriateness of drawing conclusions about individuals from an analysis conducted on an aggregated unit, such as a census tract. The modifiable areal unit problem refers to artifacts which

---

<sup>4</sup> A site must be placed on the NPL before cleanup can occur under the Superfund program.

result from the scale at which an analysis is conducted. If an analysis is repeatedly performed on the same base data set at increasingly aggregated levels (ie. census block-group, census tract, zip-code, or county), it will typically be found that the strength of the relationship increases. Although it is difficult to make comparisons across scales, the  $R^2$  associated with a linear relationship typically increases as the degree of aggregation is increased.

Beyond affecting the overall strength of the relationship, the degree of aggregation and the scope of a study can affect the significance of individual coefficients within a model. As discussed in the previous chapter, in national studies on air pollution, income does not typically appear to be a significant factor because both income and air pollution concentrations are higher in urban areas. But, when one examines in greater detail the relationship between income and air pollution within urban areas, it becomes apparent that there is a strong relationship (CEQ, 1971; Asch & Seneca, 1978; Gelobter, 1988).

In the national studies and most of the studies on urban areas, the analyses were performed with data typically aggregated to the zip code level. This scale of analysis might not be detailed enough to capture the relationship between the environmental hazard and the affected population, and the degree of aggregation might inflate the strength of the relationship. The average size of a zip code within Los Angeles County is roughly 37 square kilometers.<sup>5</sup> Zip codes are irregularly shaped and are sometimes discontinuous in space. The size of the zip codes might again be too coarse to pick up the actual relationship between the variables of interest, and a distorted picture of the true relationship could result (Openshaw, 1984). Thought needs to be given to the scale of analysis which is appropriate for the environmental hazard being examined, and the analysis should be conducted at that scale.

### **Study Purpose**

The main purpose of this case study of Los Angeles is to determine whether or not race is a significant factor in the relationship with proximity to environmental pollution when the effects of other important variables have been removed. More specifically, when controlling for income and population density, is race a significant variable in the relationship with environmental pollution?

To achieve this goal, several spatial and statistical techniques have been applied at the highest level of resolution that was both possible and appropriate. The analysis was performed at the census-tract-level for over 1600 census tracts in Los Angeles County. This choice of scale is justified later in this chapter in the section on *Scale Issues*, after the spatial accuracy of the data sets has been discussed.

### **Study Hypothesis and Assumptions**

Within a market-oriented economic system, it stands to reason that economically disadvantaged people would suffer a greater share of pollution than those better-off economically. Greater economic resources afford greater choice regarding where to live and how much can be spent on environmental factors and aesthetics. More money can be translated into a larger home; more property buffering, the home; a home closer to the shore and upwind from most traffic; or into additional consumer goods, as a matter of choice. A higher income does not guarantee that more money will be spent on one's living environment, but it affords greater opportunity to do so. The less economically well off one is, the less choice one has regarding where to live. The first premise of this analysis is that *money offers choice*. A corollary of this is that wealthy people will tend to live in cleaner neighborhoods.

It is unfortunately the case that minorities are typically more economically disadvantaged than Whites. The reasons for this are complex, but among others, include being new arrivals to an area; language differences; educational disadvantage; economic discrimination; and the restructuring and suburbanization of industry (Eggers & Massey, 1992). *Minorities are more likely than Whites to be economically disadvantaged, and therefore have less choice regarding where to live*. A corollary of this is that minorities will tend to live in more polluted neighborhoods.

The locations of industrial facilities in a landscape are the result of a complex set of physical, economic, and political factors working together, in addition to an element of chance. The location of transportation corridors and waterways; zoning regulations; local taxes; property values; land availability; and proximity to a market, labor force and input resources are among the plethora of factors that affect where a facility might be located.

Due in part to industrial zoning, industries are often located in the less densely populated parts of an urban landscape, and often along transportation arteries. Residential property values are typically lower in industrial areas than in areas without industry (Asabere & Huffman, 1991). For this reason, low income communities could develop in industrial areas, as the industry offers the dubious appeal of making the area a less desirable place to live and lowering property values and rents. This premise, that *areas with*

---

<sup>5</sup> The zip code area estimate is based on 280 5-digit codes in L.A. County (excluding P.O. box zip codes), and an area of 10,400 sq.km.

*polluting industries tend to be less expensive places to live* accounts for the potential circumstance of people with low incomes choosing to live in areas with industrial facilities.

The complement to this case involves the siting of new facilities. As mentioned previously, many factors affect locational choice. But, as environmental awareness has increased over the last twenty years, so has community resistance to the siting of new potentially polluting facilities and the NIMBY phenomenon. NIMBY implies an acceptance of the need for the given facility, as long as it is sited outside of the relevant area or "backyard" of those resisting the facility. (This differs from the more ideological "Not In Anybody's Back Yard" which encourages fundamental changes in industry and waste management, thereby eliminating the need for some facilities.) If a facility is perceived as undesirable, there is the chance that there will be resistance to its siting. Facility siting often follows the path of least resistance, coupled with considerations of minimizing costs and maximizing profits. The path of least resistance could include targeting areas where people are not politically active or well connected, have not indicated tendencies toward environmental activism, or perhaps do not have English as their first language (Dear, 1992). All other things being equal, *the siting of facilities which are perceived as undesirable often follows the path of least resistance.*

Largely due to zoning, industry will tend to be located in areas of low population density. Such areas would have fewer people with a given facility in their "backyard," and perhaps, therefore, provide less resistance. Also, industry tends to cluster in industrial zones, rather than being dispersed across the landscape. The premise here is that *industry is more likely to be located in areas of low population density.*

Given these assumptions, it is useful to examine the objective of this study - determining whether race is a significant factor in the relationship with environmental pollution when the effects of other important variables are removed. As stated earlier, minorities tend to be more economically disadvantaged; have less choice about where to live; and for reasons of affordability, might choose to live in areas close to industry. For these economic reasons alone, one would expect to see a disproportionately high number of minority residents in industrial areas. But, the question to be examined is whether or not there is an effect related to race which goes beyond what would be expected from economic factors alone.

The variables which will be used in this analysis represent the mean or median conditions for each of 1640 census tracts in Los Angeles County. The dependent variable of interest which is used as a surrogate for level of pollution is the number of "polluting facilities" within the census tracts. (More detail on this definition is provided in the next section, on Data Sources.) The explanatory variables of interest are:

- 1) population density for the tract (people per sq. km.);
- 2) an indicator of income in the tract - median per capita income will be used. In addition, the importance of median housing cost will be explored; and
- 3) some indicator of racial/ethnic composition in the tract - a single variable representing the minority percentage in the tract is evaluated, in addition to separate variables representing the Hispanic, Black, and Asian percentages for the tract.

These variables reflect the residential composition of the census tract. Although people are exposed to pollution in many different phases and locations during their daily and weekly routines, residence has been selected as the best single surrogate for exposure to pollution.

**\*\* A Note on Terminology:**

Although Black, White, Hispanic and Asian are a mix of *racial and ethnic categories*, the term *race* will be used when referring to these four categories.

To state the hypothesis of this study in statistical terms: If one were to control for the effect of income level and population density and then examine the number of polluting facilities in a census tract as a function of race, the null hypothesis states that the coefficients associated with the racial variable(s) are equal to zero. The alternative hypothesis states that for at least one of the racial variables, the coefficient is non-zero, with a statistically significant degree of confidence.

This statistical analysis examines the relationship between the explanatory variables and the dependent variable for a single point in time. As such, this analysis provides no information on causality, but rather, quantifies the current relationship.

Two cases will be examined - one where minorities are grouped into a single variable (Minority Percentage) and another where the White, Hispanic, Black and Asian percentages are treated separately. These hypotheses differ in terms of number of

variables depending upon whether or not the Hispanic, Black and Asian tract percentages are grouped into the single variable for Minority Percentage, or are examined separately.

A hypothesis testing framework will be used in this analysis, and individual coefficients will be evaluated at the .05 level of significance for a two-tailed test. This level of significance, admittedly the most commonly used in statistical tests (Gould, 1970), affords a high degree of confidence that the coefficient of interest is non-zero. Although this framework is technically appropriate for inferring characteristics about the whole population from a sample, it is also commonly used in modeling an entire population, such as in analyses including all U.S. counties. Within this analysis, all but 30 of the census tracts in Los Angeles will be included. The relative significance of the variables reflecting race and income will be evaluated through significance testing of model coefficients and pairwise model comparisons. For some models, 95% confidence intervals surrounding the coefficient for the racial variable will be evaluated, to establish a high degree of confidence that the coefficient for the population is not zero.

## **Figure 1 - Study Hypothesis**

---

**While controlling for the influence of Population Density and Per Capita Income level, Race / Ethnicity is significant in the relationship with the number of polluting facilities in an area.**

### **Case 1: Minorities Grouped**

**$H_0: B_W = B_M = 0$  (for White % and Minority %)**

**$H_1: B_M$  not equal 0**

### **Case 2: Racial / Ethnic Groups treated Separately**

**$H_0: B_W = B_B = B_H = B_A = 0$  (for White, Black, Hispanic and Asian %)**

**$H_1: B_i$  not equal 0 (for one or more coefficients)**

---

The analysis is made difficult by the degree of correlation between variables. In particular, the variable representing the minority percentage for the tract is highly negatively correlated with median per capita income for the tract (at -.71). As a result, the variable reflecting racial mix (Minority Percentage) includes a considerable amount of information about per capita income, while the Per Capita Income variable includes considerable information on racial makeup. The problem of multicollinearity results when two explanatory variables are highly correlated. In cases where the correlation approaches unity (linearity), the model breaks down computationally (Kennedy, 1985). A diagnostic test for multicollinearity will be applied to insure that the model assumption of independent variables is not violated.

### **Overview of Analysis**

This analysis is both spatial and statistical in nature. In order for any statistical techniques to be applied, the spatial relationship between sources and recipients of pollution needed to be determined. To do this:

- 1) a surrogate for pollution had to be established;
- 2) the relevant process scale needed to be determined;
- 3) a method for identifying the affected population needed to be selected; and finally,
- 4) methods for establishing the relationship between the source of pollution and the socioeconomic characteristics of the population needed to be developed.

## **Data Sources**

Two data sources were used in this analysis - one reflecting the socioeconomic characteristics of the county, the other reflecting the distribution of toxic emissions within the county. U.S Census TIGER files and Summary Tapes were used, providing data at the census tract-level for 1640 census tracts in LA county. The data included information on population size, race, ethnicity, median housing value and median per capita, household and family incomes for each tract.

The most comprehensive source of data on facilities releasing toxic substances to the environment is the U.S. Environmental Protection Agency's Toxic Release Inventory (TRI) system (U.S. EPA, 1989). The system was required by the Emergency Planning and Community Right to Know provisions (Title III) of the Superfund Amendments and Reauthorization Act (SARA). In addition to locational information, the TRI data base provides information on the type of facility by industrial category; type and amount of chemical(s) released; and pathway of emission to the environment for 722 facilities within Los Angeles County.

## **Data Assemblage**

The data sets were integrated using the ARC/INFO Geographic Information System (GIS), version 6.0. The Universal Transverse Mercator (UTM) projection was used. Within the GIS, a point in polygon overlay was performed in order to determine which census tract each TRI facility lies within.

## **Data Accuracy Issues**

### **Census Data**

Census tract boundaries were created in ARC/INFO from U.S. Census Bureau TIGER files by the Environmental Systems Research Institute. The data was provided in Lambert projection and was subsequently converted to the UTM projection. TIGER line files are based on 1: 100,000 scale maps from the U.S. Geological Survey, which meet national map accuracy standards at that scale. As such, the spatial accuracy of this data set is fairly high.

This study made use of 1990 Census data which was released prior to the 1990 Post-Enumeration Survey. The Post-Enumeration Survey provides an estimate of the error associated with the initial census estimates, and produces adjustments which the U.S. Census Bureau can use to correct the final census numbers. Whether or not the Bureau would use the adjustments in the final census count was the subject of a lawsuit filed by several states and counties, including Los Angeles (Exter, 1990). The Secretary of Commerce decided against using the adjustments, though aggregated adjustment coefficients are available (Hogan, 1992).

At the national level, the undercount of non-Hispanic Whites is relatively low (less than 1%), while the undercount of Hispanics and Blacks is around 5% and the undercount of Asians is around 2.4% (Hogan, 1992). Tenure (home-ownership) seems to be an important factor in the degree of undercounting. When the undercount estimates which apply to large urban areas in the western part of the U.S. are examined by housing tenure, a more striking pattern emerges. Table 1 provides the Revised Post-stratification Estimates of the undercount on a percentage basis. It is apparent from these estimates that the undercounting of renters is much greater than for homeowners, and the undercounting of minorities is much greater than for Whites.

The implications of these error estimates for this study of Los Angeles are that:

- 1) Overall population density is likely to be higher than the data suggests;
- 2) Population density is likely to be underestimated to a larger degree in tracts which have a large percentage of minorities (Hispanic, Black, or Asian) and where the percentage of homeowners is low;
- 3) For census tracts which have high Hispanic, Black, or Asian percentages, these percentages might be slightly underestimated.

---

**Table 1** Estimates of Census Undercount

Post-Stratification Estimates of Percent Undercount for the 1990 Census

| <u>Race / Ethnicity</u> | <u>Housing Tenure</u> |                  |
|-------------------------|-----------------------|------------------|
|                         | <u>Owner</u>          | <u>Non-owner</u> |
| White (Non-Hispanic)    | -0.3                  | 3.2              |
| Black                   | 6.1                   | 9.9              |
| Hispanic (Non-Black)    | 2.9                   | 5.9              |
| Asian                   | -1.5                  | 7.0              |

(Source: Hogan, 1992)

---

To concisely summarize the "relative error" in the estimates - the minority dominated tracts are likely to be more densely populated and might have slightly higher minority percentage than the data suggest. As will become apparent in the Chapter 3, most TRI facilities occur in low density tracts and in tracts which have a high minority percentage. As a result, not including these adjustments should not have a significant effect on the results of the analysis.

**Toxic Release Inventory (TRI) Data**

The locational information contained within EPA's TRI data base was obtained through several methods. For most facilities within the TRI data base, the facility operator or some other responsible party provided the facility locational information to EPA. Locational information includes the facility street address, city, county, and zip code in addition to latitude and longitude. The street address information, including zip code, tends to be fairly good, barring errors introduced during data entry. The accuracy of latitude and longitude coordinates tends to be lower, owing to the fact that not all information providers are accustomed to reading coordinates from a map, and might not have another means of obtaining this information. This is the primary means of obtaining latitude and longitude.

When the latitude and longitude are not provided by the facility, or when there is an obvious error in the coordinates, this information is obtained from EPA's ROADNET system if available, or the centroid (polygon center) for the zip code of the facility is used (EPA, 1991). An analysis on TRI locational data quality assurance performed for EPA compared coordinates obtained through these three sources and found that those provided by the facility to be the "Preferred" coordinates in 74% of the cases examined; ROADNET coordinates 14% of the time; and Zip code centroids 12% of the time (EPA, 1991).

Extensive quality assurance (QA) and quality control (QC) is required in the development and maintenance of a database like EPA's TRI. Resources have not permitted the necessary QA/QC for the all facilities in the TRI. EPA has concentrated its efforts on insuring the locational accuracy of the largest TRI facilities.

The shortcomings of locational accuracy of the TRI can be seen by displaying the raw TRI facility locations as a geographic coverage overlaid on a map of Los Angeles County. The number of facilities which lie well offshore reflects the problems with locational accuracy. Of the 722 TRI facilities in Los Angeles, 56 fell outside the county boundary.

Several methods exist to improve the locational accuracy of such a data set. These include visiting individual facilities with a Global Positioning System (GPS) which would provide a digital readout of geographic coordinates, accurate to within 6 meters (Long et. al., 1991). Another technique involves address matching. In an automated fashion, the facility street addresses are matched with the Census TIGER Files, where latitude and longitude of the street segments are known. This results in locational coordinates which are highly accurate. A third technique would involve manually looking up street addresses on maps and reading the latitude and longitude from the map, which is very time consuming and has a much lower accuracy.

The Teale Data Center in Sacramento has improved the quality of the TRI data set for LA County by doing an automated address matching for 379 of the largest of the 722 facilities. This provided a sound base on which further locational improvements

could be made. Of the 343 facilities whose coordinates had not been revised, 45 had locations which fell outside the LA County boundary and were definitely incorrect. These locations were improved through a point editing procedure.

The 379 facilities whose coordinates were improved through automated address matching are the most accurate, with an estimated accuracy of within 0.1 kilometer. A sample of 325 facilities was taken to assess the accuracy of the overall population. Of these 325, only 21 facilities (6.5%) were located outside their Zip code boundary (based on the Thomas Brother's Guide for Los Angeles - Zip Code Edition). Facilities whose locations fell outside their Zip Code boundaries were improved through the same method as used previously. This method is estimated to be accurate to within 0.5 kilometers of the actual location. The estimated average map accuracy of the whole data set (722 facilities) is 0.3 kilometer.

|                                     |       |   |            |  |
|-------------------------------------|-------|---|------------|--|
| <b>Average Locational Accuracy:</b> |       |   |            | <b>(These are conservative estimates.)</b> |
| Improved -                          | 379   | * | 0.1        |  |
| Corrected -                         | 61    | * | 0.5        |  |
| Unaltered -                         | 282   | * | 0.5        |  |
|                                     | ===== |   |            |  |
| <b>Average =</b>                    |       |   | <b>0.3</b> |  |

### **Census Tracts and TRI Facility Locations Combined**

The Census tract coverage which was created from Tiger files was obtained in Lambert projection and was converted to Universal Transverse Mercator (UTM). The UTM projection is appropriate for small geographic areas and is a projection in which areal distortion is minimal. The TRI was obtained as a tabular data file which included latitude and longitude. A geographic coverage (Lat/Lon) was created which was then converted to UTM for overlay with the Census coverage. As a TIGER file served as the basis for the address-matching for TRI facilities, errors due to misregistration should be minimal. Checking the facilities with "improved" coordinates using a Thomas Brothers Map of Los Angeles indicated good registration between the two coverages, as facilities did appear located in the correct positions within census tracts.

Further, as the Census data set does appear spatially correlated with regard to income and race (i.e. income groups and racial groups are clustered rather than random), this would serve to lessen the effect of errors where facilities are placed in adjacent tracts.

### **Methods of Analysis**

Several statistical techniques were applied in order to examine the influence of race on proximity to polluting facilities. Additionally, these techniques can be used to identify geographic areas which stand out as anomalies in terms of having many more or fewer polluting facilities than would be expected. A brief description of each method follows, along with a discussion of the limitations specific to each approach. Limitations which are relevant to all methods are discussed in the last section of this chapter.

Prior to performing the analysis, the relationships between all explanatory variables and the dependent variable were examined and were found to be approximately linear, and a diagnostic test for multicollinearity among the explanatory variables was performed. Two types of generalized linear models were applied - one for the case of modeling the number of TRI facilities in a census tract, and one for modeling presence or absence of a TRI facility in the tract. The statistical analysis is preceded by an exploratory analysis technique - bivariate mapping.

#### **Method 1) Bivariate Mapping Analysis**

This technique allows for a bivariate analysis of the landscape for the purpose of identifying tracts which are anomalies in terms of the value of their explanatory variables (i.e. highly-White, low-income tracts, or high-income, minority-dominated tracts), which then might be interesting tracts to examine in subsequent analysis.

This method of analysis involves partitioning the whole set of observations into separate groupings based upon income level and minority percentage. First, the number of categories (n and m) for each of the two variables is selected. Next, thresholds for partitioning the data into quantiles are calculated, such that the same number of observations falls into each single variable (1-dimensional grouping). This partitioning is done for both variables (Per Capita Income and Minority Percentage). This results in an n by m matrix, where the number of observations in each row or column are roughly equal, but the number of observations in each cell would not be equal due to the negative relationship between Per Capita Income and Minority Percentage. For example, cells reflecting

a high Per Capita Income and low Minority Percentage are likely to have more observations than those with a high Per Capita Income and high Minority Percentage. A bivariate mapping is achieved through representing one variable through changes in hue and the other through changes in lightness.

### **Limitations**

This analysis method should be useful for visualization of the relationship between race, income and facility occurrence. One limitation of this method is that it will only be applied to two variables at a time. Therefore, when income-level and minority percentage are examined, the influence of population density is not considered. Another limitation of this approach is that the results could be an artifact of the thresholds selected and resultant groupings. This limitation could be overcome through evaluation at several different levels of resolution.

### **Method 2) Modeling the Number of TRI Facilities in a Tract: Generalized Linear Modeling to a Poisson Distribution**

The intent of this analysis is to determine the relationship between the explanatory variables (population density, income, and one or more variables representing racial composition) and the dependent variable (number of TRI facilities in a census tract). This dependent variable is a count variable with a very skewed distribution. The variable ranges between 0 and 25, but 83% of the tracts have no facilities and only 4% of the tracts have more than 2 facilities. (The distributions of all variables are provided in Appendix B.) Owing to the skewed distribution, using Ordinary Least Squares estimation and modeling to a Normal (Gaussian) distribution is not appropriate, as the necessary assumptions of a normally distributed dependent variable and normally distributed error term would be violated. (Kennedy, 1985). A non-linear model, fitted to a Poisson distribution will be used.

A test for heteroskedasticity within the classical linear model revealed that the variance of the error term in the model is not constant over the range of the explanatory variables. This violates another assumption of classical linear regression - that the error term has uniform variance (Kennedy, 1985). Fitting the model to a Poisson distribution greatly reduces the degree of heteroskedasticity.

Several combinations of explanatory variables will be evaluated within this model structure. The influence of Population Density and median Per Capita Income will be consistently included in the model.<sup>6</sup> The explanatory variables used as an indication of race / ethnicity will vary. A variable reflecting the total Minority Percentage will be used in the basic model, while three separate variables reflecting the Hispanic, Black and Asian percentages of each census tract will be evaluated in a second model. Additionally, binary indicator variables (dummy variables) reflecting, racial / ethnic categories will be used to isolate the influence of race. The importance of the variable representing race will be evaluated through significance testing.

### **Limitations of the model**

As discussed earlier, the dependent variable - number of TRI facilities in a tract is highly skewed, where 96 percent of all tracts have 2 or fewer facilities, and only 1.3 percent of all tracts having more than 5 facilities. The mean number of TRI facilities in a tract is 0.41. A generalized linear model of the form implemented will predict a mean fitted value which is equivalent to the mean value of the dependent variable. But, the fitted values will have a narrower range (and variance) than the actual data. The model will predict some low values, but not zero. As a result, it is expected that many tracts where there are no facilities present will have small negative residuals. Also, the model is unlikely to predict facility occurrence as high as some of those which occur in reality, such as those with more than five facilities. So, in cases where the actual number of facilities is quite high, even if the model predicts a high value in terms of the model range, a positive residual could be expected. In general, it can be said that such a model will tend to overpredict in the cases of 0 occurrence, and underpredict in the case of high facility occurrence.

### **Method 3) Modeling Presence or Absence of Facilities in a Census Tract: Logit Analysis**

Logit analysis can be employed to model presence or absence of TRI facilities in a census tract. This second type of model is being implemented in order to insure that results are not dependent upon a few highly industrial tracts with an extreme number of TRI facilities. Logit models are generalized linear models fitted to a binomial distribution. The model assesses the likelihood of facility presence for each tract, given the values of the dependent variables for that tract. In general, the same explanatory variables will be assessed as were used in the modeling of number of facilities per tract - Population Density, Per Capita Income, and several variables

---

<sup>6</sup> Median Housing Cost was evaluated for its value as a variable representing economic status (replacing Per Capita Income). Per Capita Income proved to be a more significant variable when included with Population Density and Minority Percentage as explanatory variables.



which indicate racial makeup. The dependent variable in this case will be a binary variable reflecting whether or not any TRI facilities are present in a census tract.

### **Advantages and limitations of the model**

Using logit analysis to model the presence / absence of facilities in a tract is, in many regards, simpler than modeling the number of facilities in a tract. By modeling presence absence, the problem of trying to predict a widely distributed count variable has been eliminated. This method also eliminates the problem of highly influential observations (outliers) in terms of the dependent variable. But, one of the limitations of logit modeling is the associated loss of information.

Although all facilities do not have the same type or level of emissions, it can still be assumed that on average, close proximity to several facilities proves more of a hazard than close proximity to only one (EPA, 1989). A limitation of the logit model is that presence of a single facility in a tract is treated the same as the presence of 20.

A logit model produces fitted values between 0 and 1 which reflect the predicted probability of a TRI facility being present in a given tract. It is standard practice to use a threshold of 0.5 for interpretation of the model prediction (Pindyck and Rubinfeld, 1991). Fitted values below 0.5 are interpreted as a prediction that no facility is present, while fitted values greater than 0.5 are interpreted as a prediction of facility presence.

Roughly 17% of all tracts have facilities. The logit model will produce a set of fitted predictions (one per observation) which have a mean equal to the actual occurrence of the dependent variable (0.17). When the prediction threshold is applied, it is likely that only a small percentage of observations will exceed the threshold of 0.5. The logit model is therefore likely to severely underpredict the number of tracts with facilities. It should also be noted that given the disproportionate absence of facilities vs. presence in tracts (83% vs. 17%), the success rate for predicting absence should be much higher than the success rate for predicting presence.

### **The Problem of Spatial Dependence**

As discussed previously, industrial facility siting is not a random process and facilities tend to cluster in industrial zones. Additionally, as an initial visual analysis of the explanatory variables reveals (described in Chapter 3), there is a high degree of spatial autocorrelation within each of the explanatory variables. Low-income census tracts are concentrated in the center of the city, as is the majority of the minority population.

Both the modeling of the number of facilities per tract and modelling of presence absence of facilities within a tract used variable-sized census tracts as the unit of analysis. Over 1600 tracts were used in the analysis, hence, there were over 1600 observations used in the generalized linear models. Each observation was treated as an independent observation, regardless of its location in space. This ignores the potential problem of spatial dependence (Anselin, 1989).

Spatial autocorrelation or spatial dependence is the presence of spatial pattern in a mapped variable due to geographic proximity. The presence of spatial dependence violates a basic assumption of many standard statistical tests - that observations are independent (Haining, 1990). Ignoring spatial dependence can result in a biased analysis with inaccurate results, both in terms of variable significance and coefficient value.

There are methods for dealing with spatial dependence. First, there are tests for spatial autocorrelation, such as the Moran, Geary and Join-count statistics (Goodchild, 1986), which help to determine whether further consideration of spatial dependence is warranted. In the case of census tracts in Los Angeles, the simple visual analysis of the distribution of income, and racial mix of the population shows that the population is not randomly distributed, and spatial autocorrelation exists. This is also true of the distribution of the TRI facilities across the landscape.

Once the determination has been made that observations are spatially autocorrelated, several techniques are available to address the problem. Spatial regression analysis, available in spatial statistical software such as SPACESTAT (Anselin, 1992), can perform the linear regression analysis, while taking into account proximal and adjacent effects. As modeling of discrete data (such as logit analysis) is not available in SPACESTAT, this was not an alternative.

Another option for dealing with spatial dependence is through repetitively evaluating coefficients of models which include a randomly-sampled, spatially-independent, subset of all observations. This resampling technique is similar to bootstrapping (Diaconis, et. al, 1983), except sampling is not done with replacement. Bootstrapping is a robust estimation technique which exploits randomness in resampled data sets, in order to infer parameter estimates, measures of bias and variance, and pseudo-significance levels (Anselin, 1990). The resampling procedure involves, first, determining a minimum distance threshold for spatial dependence, beyond which observations are likely to be spatially independent. Next, a single observation is randomly chosen, and a set of observations which are

beyond the distance threshold from all other observations are selected, and the model is run on this "resampled" subset of observations. The resampling and the estimation of model coefficients is then repeated many times (hundreds to thousands), and from this set of coefficient estimates, a final set of coefficient estimates can be empirically estimated. This final set of coefficients provides a very good representation of the relationship between the explanatory and dependent variables, while circumventing the problem of spatial dependence.

The resampling technique will be applied to a logit analysis.<sup>7</sup> Standard logit analysis (of all 1600+ observations) and the resampling technique will be performed on the same set of variables, so that the importance of spatial dependence can be evaluated.

### **General Limitations of the Analysis**

This analysis is based on the evaluation of a very limited set of explanatory variables reflecting socioeconomic characteristics of the population as these influence the occurrence of facilities which deal with or emit toxic substances. But, many other factors which are quite relevant to the siting of industrial facilities were not included, such as proximity to transportation corridors and waterways or the effect of zoning, local taxes, etc. The only type of environmental hazard considered in this analysis is industrial facilities in the TRI. Additionally, all TRI facilities were treated uniformly, regardless of the type of facility, or types and amounts of chemicals emitted.

Regression analysis assumes that explanatory variables are exogenous (originating from external causes). Factors within an urban landscape are inextricably related. Within the proposed models, the explanatory variables are closely related to the dependent variable. It is possible that the explanatory variables influence TRI facility occurrence. It is also possible that the occurrence of TRI facilities influences the characteristics of the surrounding community - its population density, economic class, and racial / ethnic composition. This analysis will quantify the relationship between these variables, but as stated earlier, will not establish causality.

### **Time**

This analysis was performed for a single time period, and in essence reflects a single "snapshot in time". As such the models implemented reflect the relationship among race, income, population density and TRI facilities at the end of the 1980's, but provide no insight into the processes which have resulted in this landscape and these relationships. As such, conclusions about facility siting decisions "targeting" certain types of communities would be inappropriate.

Within this type of cross-sectional analysis, the racial/ethnic composition of a community at the time of facility siting is not considered. Further analysis is therefore required of individual neighborhoods, in order to understand the process over time. Several neighborhoods will be selected for further study, and the socioeconomic dynamics of the neighborhood will be examined in conjunction with the inception of facilities in the neighborhood. Results of this analysis will be presented in Chapter 5.

### **Space**

A limitation to modeling at the census tract level is the effect of aggregation of the population and polluting facilities to the variable-sized unit of the census tract. This results in one's proximity to a facility being averaged across the tract. So, whether one is adjacent to a facility, within 1000 meters; or at the opposite side of the tract, proximity is treated as uniform.

The variable size of tracts introduces additional disuniformity. As tracts in the northern portion of the county are much larger than in southern LA county, one can be 30 km from a facility and still in the same tract in the rural north, while some urban tracts are only 1 kilometer across. This problem was reduced in model implementation through the exclusion of tracts with a population density of less than 100 persons per square kilometer. Subsequent to the exclusion of these 30 tracts, the mean tract area was reduced to 2.5 sq. km from 6.3 sq. km., and the number of observations was reduced from 1640 to 1610.

In some densely populated areas in the southern portion of the county, one could live within a tract with no TRI facilities, while being in close proximity (less than 1 kilometer) to several facilities. Linear modeling aggregated to the census tract level will not pick up these proximal effects or the effects of adjacency. This shortcoming will serve to reduce the significance of the modeling results, and reduce the predictive capabilities of the model.

---

<sup>7</sup> This examination of spatial dependence is being performed for the logit modeling only, as a resampling program was available for logit modeling, but not for poisson modeling.

## Scale Issues

The relevant scale of analysis for examining proximity to toxic facilities is the subject of extensive ongoing research (EPA, 1989; Wilson, 1991). How big is the "backyard" associated with the NIMBY (Not In My Back Yard) syndrome? This relates to one's subjective perception of relevant distance. Research by Dear (1992) and others address this issue of perceived zone of effect. In physical terms, the relevant scale of effect varies with facility type; type of chemical emitted; medium of emission; and physical factors such as surface water flow; groundwater depth and flow; climate; precipitation and humidity; stack height and predominant winds.

In research on comparative risk underway at EPA concerning population proximity to TRI facilities, 2 mile and 4 mile (3.2 and 6.4 km.) distances are being used as the relevant thresholds for proximity (Personal communication, Lauren Hall, EPA). Owing to the spatial accuracy problems of the raw TRI data, much more refined analysis would seem inappropriate.

From an examination of risk assessments performed on individual TRI facilities in Los Angeles for the South Coast Air Quality Management District, the "zone of effect"<sup>8</sup> was estimated to extend as little as 0.5 km. from a facility to as much as 20-30 km. Most census tracts within L.A. County are between 1.0 and 2.0 kilometers on a side. In most cases, analysis at the census tract level would result in only a subset of the "affected population" being included in the observation. It is therefore true that for the smallest census tracts, only the most affected population will be included, while for the largest census tracts, more than the affected population might be included.

## Summary

Three varied methods of analysis have been described. These methods will be applied with the intention of identifying whether or not race is an important factor in the relationship with proximity to TRI facilities; and to identify neighborhoods which stand out as having an unexpectedly high number of TRI facilities, given the characteristics of the neighborhood, other than race.

Chapter Three provides an introduction to the landscape of L.A. County. The chapter provides an overview of the physical landscape; the socioeconomic and industrial landscapes; and describes the inter-relationship among variables representing these landscapes.

---

<sup>8</sup> "Zone of effect" is the area within which this facility contributes a  $1 \times 10^{-6}$  aggregate cancer risk.

### **Chapter 3. Background on Los Angeles County**

This chapter provides a description of the physical setting of Los Angeles, the distribution of industrial facilities, and the socioeconomic landscape that comprise this multi-cultural city. A statistical summary of important variables is provided. The chapter concludes with an exploratory spatial analysis.

#### **The Physical Landscape of Los Angeles County**

Los Angeles County, a land area of approximately 10,400 sq. kilometers (4,060 sq. miles) is striking for its wide variety of topography and climate. The area is comprised of high mountains in the north; low, rugged intervening hills; broad, flat valley floors; and a tremendous coastal plain.

The high, rugged San Gabriel Mountains comprise most of the northern half of Los Angeles County, stretching from the northwest corner of the county across to the center of the county's eastern border. This section of the county is predominantly public land - the Angeles National Forest, and is sparsely populated. The only major communities to the north of the mountains are Lancaster and Palmdale, skirting the Mojave Desert. (Figure 2 provides an orientation to the area.) In the western panhandle of the county, near the coast, lie the much lower Santa Monica Mountains. These are also sparsely populated, predominantly public lands, held as national forest, national recreation area, and state parks. The Hollywood Hills lie at the eastern edge of these mountains.

Between the San Gabriel and Santa Monica Mountains lies the wide, flat San Fernando Valley. Burbank lies at the eastern edge of the valley. Further east, still bordered by the San Gabriels on the north, is the San Gabriel Valley, stretching from Pasadena, nearly to Pomona. Both valleys have been important agricultural areas.

South of these valleys lies the southern third of the county, the vast Los Angeles coastal plain. It is on the plain that the vast majority of Los Angelinos live. Downtown Los Angeles lies in the center of the northern portion of the plain. Just to the east, the Los Angeles river runs from Vernon, nearly due south to Long Beach Harbor. The cities of Inglewood, Torrance, Compton and Santa Fe Springs are among the many which lie on the plain. Palos Verdes is the promontory which divides the more industrial south-facing coast, dominated by Wilmington, Terminal Island and Lono, Beach, from the western-facing coast, which is comprised of beach communities such as Manhattan Beach, Marina Del Rey, and Santa Monica, in addition to LAX airport.

Figure 2

## Features and Places in Los Angeles County



### The Industrial Landscape - TRI facilities within Los Angeles County

EPA's Toxic Release Inventory (TRI) database includes facilities within Standard Industrial Classification (SIC) numbers 2000 - 3999, whose emissions of any of set of toxic chemicals exceeds prescribed thresholds. These SIC categories include food and tobacco processing; textile mill and apparel production; lumber, paper and furniture processing; printing; chemical processing; petroleum refining; rubber and plastics manufacture; leather, stone, clay and glass industries; primary and fabricated metal industries; commercial machine and computer manufacturing; transportation equipment; analytical and optical goods; and miscellaneous manufacturing. (For a detailed list of industrial categories and their frequency of occurrence within Los Angeles County, see Appendix C.) One hundred thirty nine different chemicals are emitted by facilities within Los Angeles County.

The 1989 TRI data base includes 722 facilities for L.A. County. The pattern of TRI facilities in L.A. County is far from random. Most facilities are in the southernmost quarter of the county. Large concentrations of facilities can be identified around Vernon<sup>9</sup>, south of Compton, and near La Puente and along the Interstate 5 corridor. Figure 3 reflects the distribution of TRI facilities within LA County.

### **The Socioeconomic Landscape - Census Data for Los Angeles County**

The US Census data on the 1640 tracts which make up LA county provide a wealth of information on its estimated 8,863,164 inhabitants, and the socioeconomic landscape of the county. Census-tract-level data used in this study includes the total population for the tract; population percentage by race and ethnicity; median household and per capita income and median home value. By combining the total population for the tract with the tract area (which is available in the GIS coverage), the population density of the tract has been calculated.

The US Census Bureau treats race separately from Hispanic ethnicity. At the most aggregate level, the population is categorized into five racial categories and two categories regarding Hispanic Origin (of Hispanic Origin, or not of Hispanic Origin). The five racial categories are White, Black, Indian/Eskimo, Asian and Other. The whole population, therefore, can be categorized by race, Hispanic-origin, or both. Summary statistics for LA County are presented in Table 2.

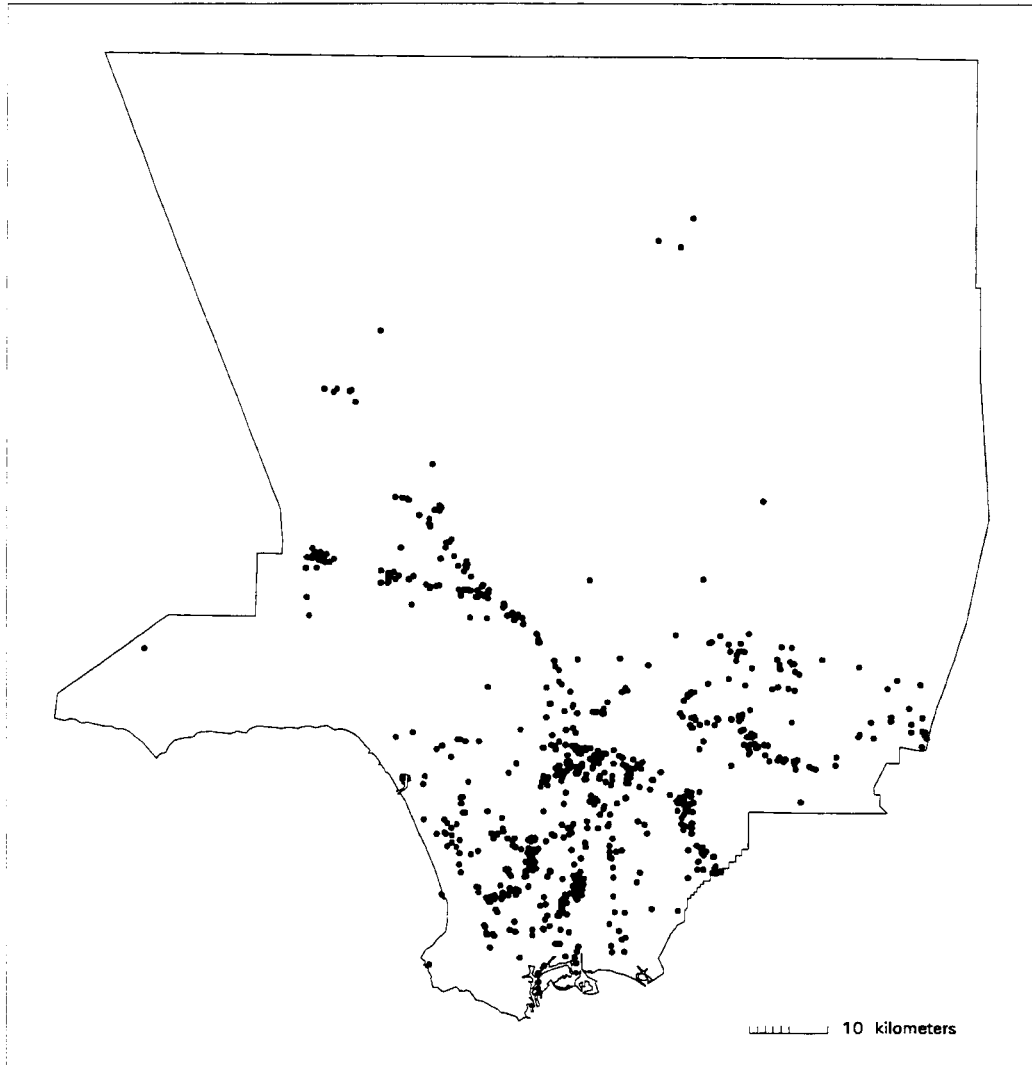
The vast majority of persons claiming Hispanic Origin also claimed White or Other as their race. Less than 4% of the persons of Hispanic Origin claimed Black, Indian, Eskimo, or Asian as their race. Due to this distribution and in order to be consistent with other work in this area, this two-dimensional matrix has been recategorized into the five ethnic/racial categories which will be used in this analysis:

---

<sup>9</sup> Vernon was identified in a 1991 analysis by the San Francisco Examiner as the "dirtiest zip code" in the State. (See Kay, 1991).

Figure 3

### TRI Facility Locations in Los Angeles County



• Indicates TRI Facility Locations

Data Source:  
1989 U.S. EPA Toxic Release Inventory

**Table 2 - Population by Race and Hispanic Origin**

Los Angeles County  
Total Population: 8,863,164

Percentage Breakdown by Race and Hispanic Origin

| <u>Race</u>    | <u>Of Hispanic Origin (%)</u> | <u>Not of Hispanic Origin (%)</u> |
|----------------|-------------------------------|-----------------------------------|
| White          | 16                            | 41                                |
| Black          | 1                             | 11                                |
| Indian, Eskimo | 0                             | 0                                 |
| Asian          | 1                             | 10                                |
| Other          | 20                            | 0                                 |

Source: 1990 U.S. Census

- 
- |   |        |     |
|---|--------|-----|
| 1) Hispanic of any Race                         | (POPH) | 38% |
| 2) White (non-Hispanic)                         | (POPW) | 41% |
| 3) Black (non-Hispanic)                         | (POPB) | 11% |
| 4) Asian (non-Hispanic)                         | (POPA) | 10% |
| 5) Other (non-Hispanic Other and Indian/Eskimo) | (POPO) | 0%  |

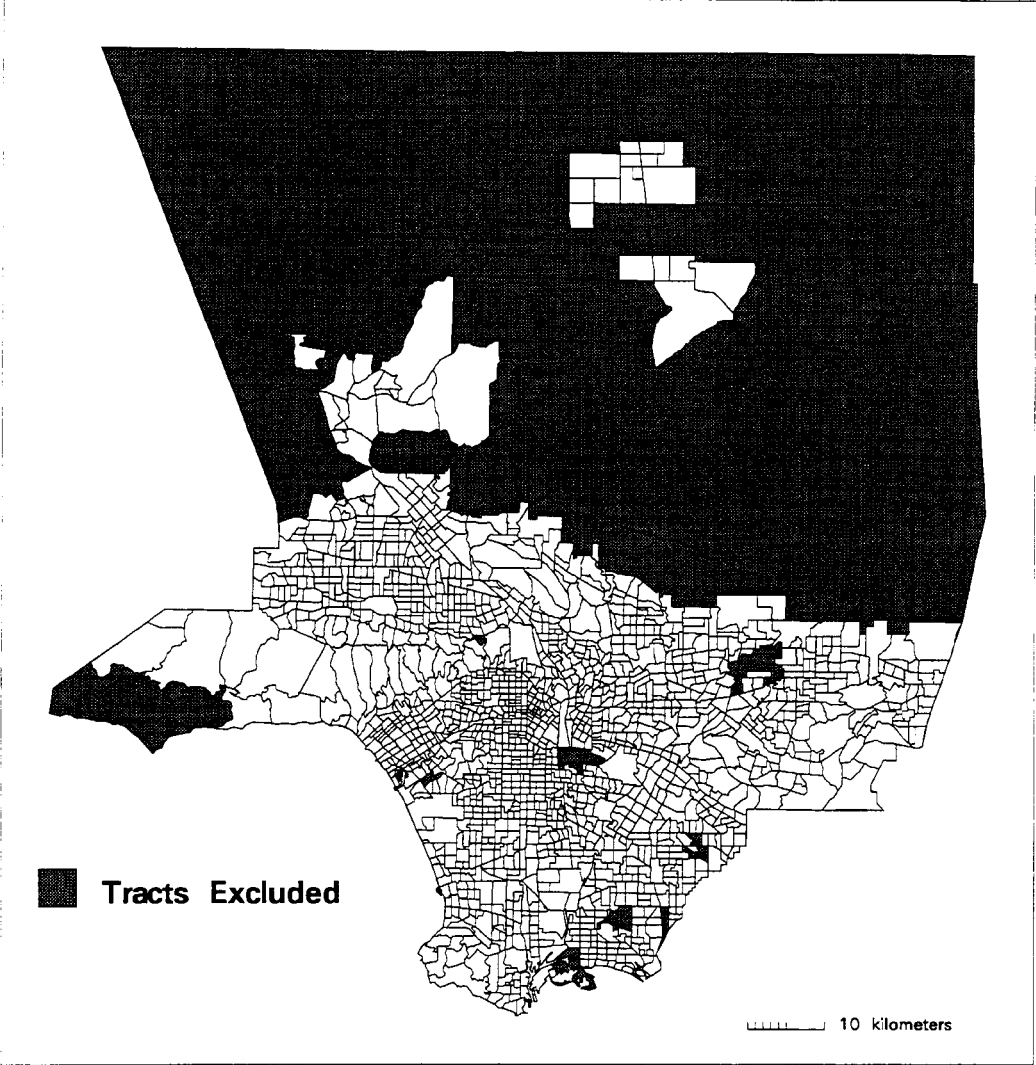
The 1640 census tracts in LA county had a median population of 5029. Five of these tracts had a population of less than 100 persons. These tracts were unusual, non-residential tracts, including, Universal City and Long Beach Harbor. These tracts were deleted from the data set as outliers. The mean population density within the remaining 1635 census tracts is 4246 persons per square kilometer. Twenty-five of these tracts have a population density of less than 100 persons per square kilometer. The majority of these tracts are rural tracts within the national forest in the northern half of the county. A few of these twenty-five tracts are sparsely-populated, highly-industrial tracts in the southern half of the county. These 25 tracts were also excluded from the analysis as outliers, resulting in a data set with 1610 observations (census tracts). This exclusion reduced the population under study by 4066 (0.05%); the number of TRI facilities by 39 (5.4%); but reduces the land area under study by 4084 sq. km. (39.8%). This allows a focus on the more densely populated southern half of the county. Figure 4 reflects which tracts have been excluded from the analysis.

Of these 1610 tracts, 278 contain one or more TRI facilities and 1332 contain none. Table 3 presents some summary statistics for these 1610 tracts.



Figure 4

**Census Tracts Excluded from the Analysis**  
**Population Density less than 100 pp./ sq. km.**



**Table 3 - Summary Statistics**

Summary Statistics for 1610 Census Tracts in LA County (with population > 100 and Density > 100 persons / sq.km.)

|                               | <u>Mean</u> | <u>Median</u> | <u>Min.</u> | <u>Max.</u> | <u>Variance</u>   |
|-------------------------------|-------------|---------------|-------------|-------------|-------------------|
| Population (persons)          | 5435        | 5048          | 119         | 28,792      | $6.0 \times 10^6$ |
| Tract Area (sq.km.)           | 2.52        | 1.36          | 0.06        | 87.64       | 26.73             |
| Pop. Dens. (pp/sq.km.)        | 4324        | 3490          | 119         | 30,000      | $1.2 \times 10^7$ |
| Median Per Capita Income (\$) | 17,353      | 14,436        | 2331        | 107,694     | $1.4 \times 10^5$ |
| Housing Value (\$)            | 242,955     | 211,450       | 0           | 500,001     | $1.3 \times 10^7$ |
| Minority %                    | 56.8        | 56.7          | 6.0         | 99.9        | .096              |
| White %                       | 43.2        | 43.3          | 0.0         | 94.0        | .096              |
| Hispanic %                    | 34.9        | 27.1          | 2.5         | 98.6        | .073              |
| Black %                       | 11.1        | 2.7           | 0.0         | 93.4        | .038              |
| Asian %                       | 10.2        | 6.4           | 0.0         | 80.9        | .012              |
| TRISITES                      | 0.41        | 0             | 0           | 25          | 2.4               |

### The Distribution of Relevant Variables

In general, none of the variables used in this analysis are normally distributed, but rather have varying degrees of positive skewedness. Appendix B provides histograms reflecting the distributions of the major variables which give an indication of how the distributions deviate from normal.

### Relationships Between Variables

It is well known that socioeconomic factors are very related. Among others, race; years of education; literacy; and income are very interdependent. Within Los Angeles County, the typical White resident has had more years of education; has a higher income; owns more property; lives in a less densely populated area; and has fewer children than the typical Black or Hispanic resident of Los Angeles County.

Within this study, these interdependencies are reflected in the correlations between the explanatory variables which reflect the racial and ethnic percentages; per capita income; housing value; and population density at the census tract level. Table 4 presents the correlation coefficients relating racial percentages, income and population density.

**Table 4 - Correlation between Explanatory Variables**

|               | <u>Per Cap.<br/>Income</u> | <u>Housing<br/>Value</u> | <u>Pop.<br/>Density</u> | <u>Minority<br/>Percentage</u> |
|---------------|----------------------------|--------------------------|-------------------------|--------------------------------|
| % White       | .72                        | .72                      | -.43                    | -.1                            |
| % Hispanic    | -.64                       | -.62                     | .42                     | .77                            |
| % Black       | -.24                       | -.35                     | .09                     | .48                            |
| % Asian       | -.01                       | .12                      | .01                     | .06                            |
| Per Cap Inc.  | 1                          | .82                      | -.39                    | -.72                           |
| Housing Value | .82                        | 1                        | -.25                    | -.72                           |
| Pop. Density  | -.39                       | -.25                     | 1                       | .43                            |

Aside from the obvious correlation between per capita income and housing value (at .82), per capita income is highly correlated with percentage White (at .72) and highly negatively correlated with the Hispanic percentage (at -.64). There is less of a negative correlation between per capita income and percentage Black (at -.24) and no significant relationship between per capita income and Asian percentage (at -.01). Housing Value is similarly correlated with the racial percentages. Summary statistics at the County level confirm the inference of these relationships - average per capita income by race within Los Angeles is \$20,531 for Whites, \$12,018 for Blacks, \$14,584 for Asians, and only \$8,100 for Hispanics (U.S. Census, STF-3, 1990).

Population density has an inverse of the same general relationship, with racial percentage, though not as strong. Population density is negatively correlated with percentage White (at -.43), positively correlated with Hispanic percentage (at .42) and shows no correlation with percentage Black or Asian. This suggests that tracts with a high percentage of Whites have the lowest population density, while those dominated by Hispanics have the highest. It follows that Per Capita Income would be negatively correlated with Population Density (at -.397), reflecting that the higher one's income, the more space one can afford.

As the correlations between these variables do not approach unity, they can be included as explanatory variables within the same model. A diagnostic test for multicollinearity was performed to insure that the correlation between variables was not so high as to invalidate model results. The "condition number" was calculated, which is the ratio of the largest to smallest of the Eigen values associated with the matrix resulting from the multiplication of X-transpose by X matrixes. As the condition number calculated, 4.2, is considerable smaller than the guideline of 100 (Montgomery & Peck, 1982), multicollinearity does not appear to be a problem.

### **Racial / Ethnic Population Patterns**

Figure 5 provides an indication of both the distribution of the population and the racial mix within each census tract in southern Los Angeles County<sup>10</sup>. The pattern is striking. Near the center of map, the barrio of East Los Angeles stands out as a Hispanic enclave. From north of East L.A., south almost to Compton and east to Santa Fe Springs, Hispanics are clearly the predominant group. To the south and west of this area lie Watts, Compton and Inglewood, areas which are predominantly Black and Hispanic. There are a few neighborhoods, such as Monterey Park and the eastern edge of the county, where Asians predominate, though these areas also include many Hispanics and some Whites. Peripheral to this dense, urban core, particularly along the coast and in parts of the San Fernando Valley, Whites predominate.

### **Relationship Between Explanatory and Dependent Variables**

The dependent variable of interest for initial modeling is the number of TRI facilities in each tract (TRISITES). The variable is highly skewed. Of the 1610 tracts, 1332 (83%) have no TRISITES, while 278 (17%) have one or more TRISITES. The mean TRISITES is 0.41, while the maximum is 25. The variable TRI01 reflects facility presence or absence within the tract. Table 5 provides the correlations between TRISITES and the major explanatory variables. Of the four explanatory variables examined, Population Density has the strongest correlation with TRISITES.

<sup>10</sup> A map of neighborhood names has been provided for reference in Appendix A.

**Table 5 - Correlations Between Dependent and Exploratory Variables**

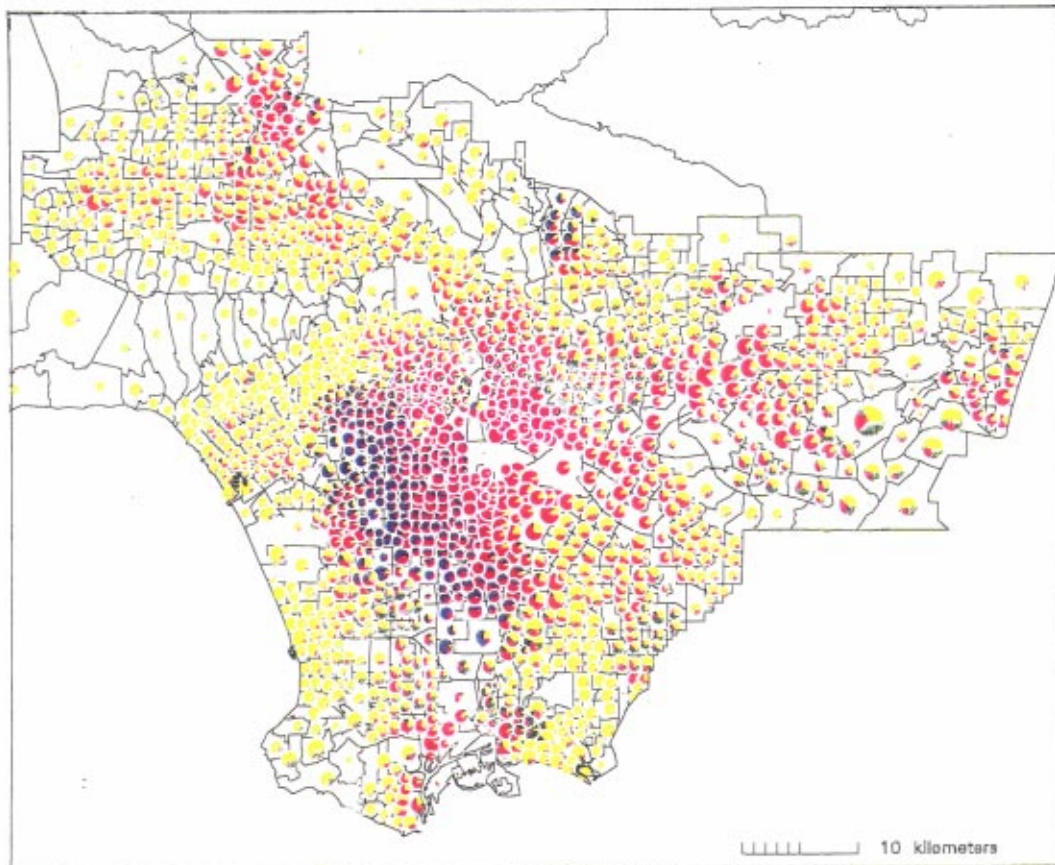
---

|                     | <u>TRISITES</u> |
|---------------------|-----------------|
| Per Capita Income   | -.101           |
| Housing Cost        | -.118           |
| Population Density  | -.146           |
| Minority Percentage | .123            |

---

Figure 5

## Population Patterns in Los Angeles The Racial / Ethnic Landscape



Proportion by Racial Group  
Disc-size Proportional to Total Population

-  Asian
-  Black
-  Hispanic
-  White (non-Hispanic)

Data Source:  
1990 U.S. Census  
1989 U.S. EPA Toxic Release Inventory

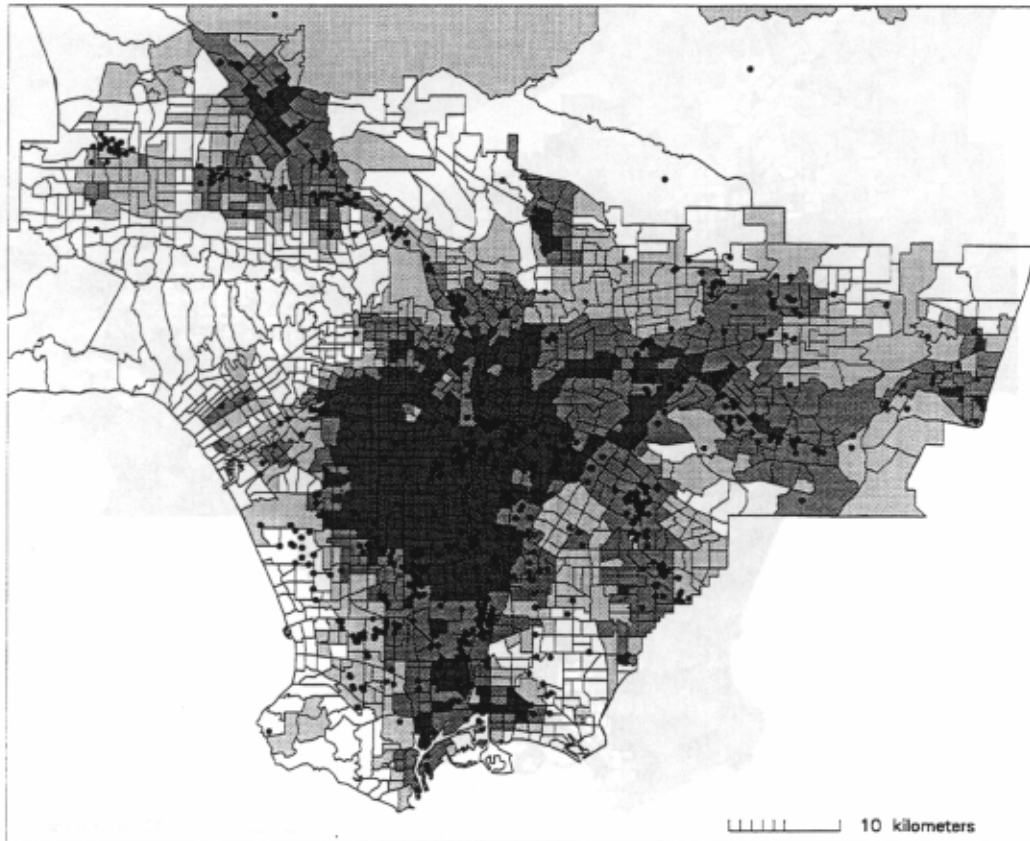
Map by Laretta Burke  
National Center for Geographic  
Information and Analysis,  
Santa Barbara, California

### Visual Analysis

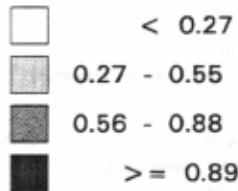
Having examined the nature of the landscape in Los Angeles County; the distribution of the TRI facilities in the county; and the demographic relationships at the census-tract-level, it is useful to visually examine the spatial interrelationships between these demographic variables and the TRI facilities. As discussed earlier, the three explanatory variables of greatest interest are the Minority Percentage (equivalent to 1 - the White Percentage); the Per Capita Income; and the Population Density.

Figure 6

## Minority Percentage and TRI Facilities by Census Tract in Los Angeles County



### Minority Percentage



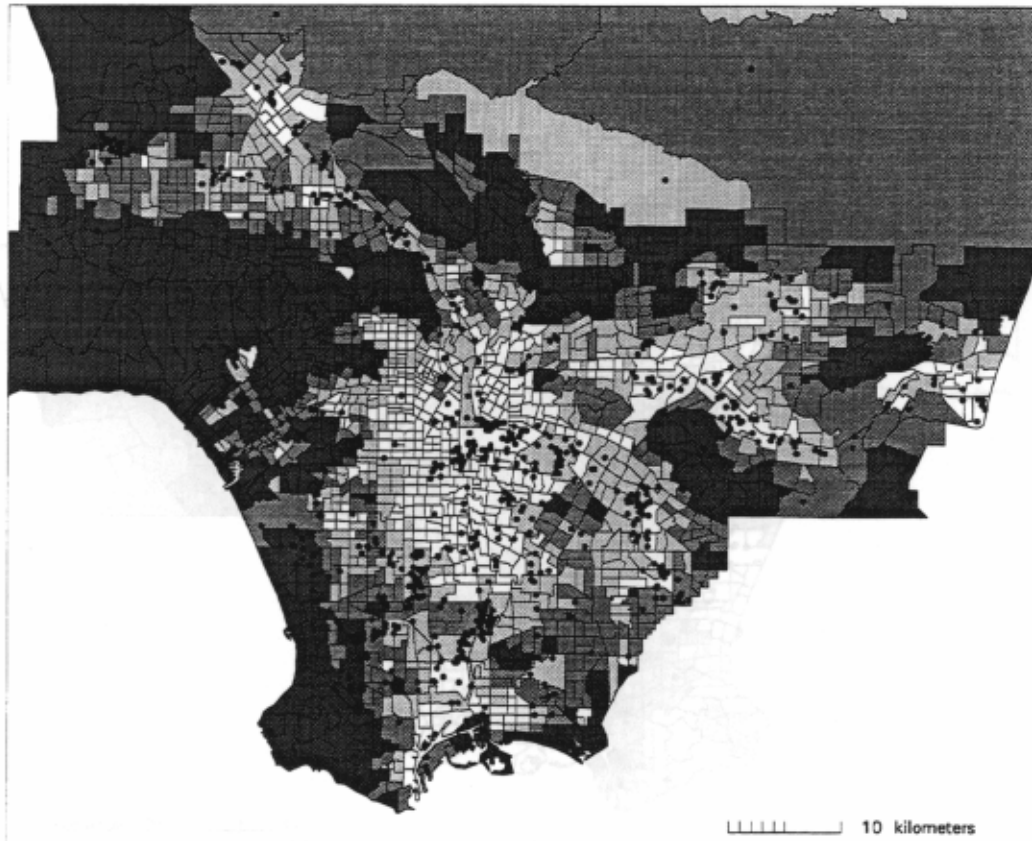
Data Source:  
1990 U.S. Census  
1989 U.S. EPA Toxic Release Inventory

Map by Laretta Burke  
National Center for Geographic  
Information and Analysis,  
Santa Barbara, California

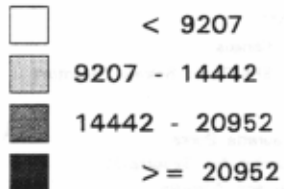
• Indicates TRI Facility Locations

Figure 7

## Per Capita Income and TRI Facilities by Census Tract in Los Angeles County



Per Capita Income (\$)



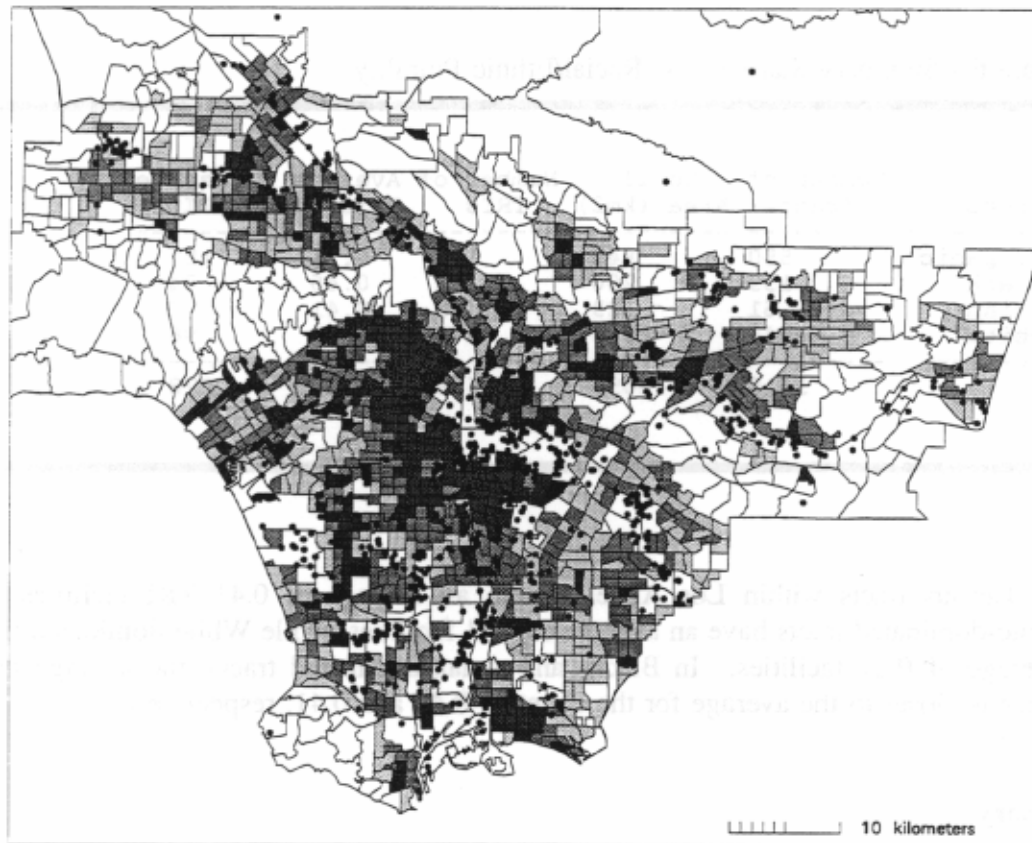
Data Source:  
1990 U.S. Census  
1989 U.S. EPA Toxic Release Inventory

Map by Laretta Burke  
National Center for Geographic  
Information and Analysis,  
Santa Barbara, California





• Indicates TRI Facility Locations

Figure 8

### Population Density and TRI Facilities by Census Tract in Los Angeles County



People Per Square Kilometer

-  < 2158
-  2158 - 3452
-  3452 - 5322
-  >= 5322

Date Source:  
1990 U.S. Census  
1989 U.S. EPA Toxic Release Inventory

Map by Laurretta Burke  
National Center for Geographic  
Information and Analysis,  
Santa Barbara, California

• Indicates TRI Facility Locations

#### Minority Percentage and TRI Facility Locations

A quantile mapping of the Minority Percentage in each census tract and TRI facility locations is presented in Figure 6. Though the correlation between these two variables is only .123, visual analysis of this map reveals a strong positive correlation



between these two variables. This anomaly probably results from the visual analysis focusing on the TRI facility locations, though there are many minority-dominated tracts which have no facilities. There are few facilities in census tracts where the Minority Percentage is less than 27%.

### Per Capita Income and TRI Facility Locations

Figure 7 presents a quantile mapping of Per Capita Income and TRI facility locations. Again, the relationship is striking. Many facilities lie within the low-income urban core, and many others follow low-income corridors which emanate from this core. The pattern is very similar to that observed in Figure 6. The El Segundo area stands out as a high-income area (also having a low minority percentage) which has many TRI facilities.

### Population Density and TRI Facility Locations

A quantile mapping of Population Density by census tract and TRI facility locations is presented in Figure 8. At this scale of analysis, Population Density appears much more heterogeneous across the landscape than either Per Capita Income or Minority Percentage. Low-density tracts carve into high-density areas. Noteworthy in this regard is the area between Vernon and Huntington Park, where several low-density, highly-industrial tracts are adjacent to high-density tracts.<sup>11</sup> Modeling at the census-tract level will, unfortunately, not capture these proximal effects. Overall, TRI facilities appear clustered in low-density tracts.

### Racial / Ethnic Composition and TRI Facility Locations

The visual analysis of the Minority Percentage and TRI facility location (Figure 6) and the correlation between these two variables (at .12) have shown that there is a positive relationship between these two variables. The relationship between individual racial/ethnic groups and TRI facilities can also be examined. By categorizing each census tract according to which group comprises the largest percentage (plurality) and overlaying this with TRI facility locations, a striking pattern is revealed. (See Figure 9). The majority of the facilities appear to lie in Hispanic-dominated tracts. Table 6 presents summary statistics by the four racial / ethnic groups for the 1610 tracts used in this analysis which confirm this impression.

**Table 6 - Summary Statistics by Racial/Ethnic Plurality**

| Group    | Number of Tracts | Total Area (km <sup>2</sup> ) | Number of TRIs | Average TRIs/tct. | Presence (TRIO1) |
|----------|------------------|-------------------------------|----------------|-------------------|------------------|
| Hispanic | 550              | 849                           | 393            | 0.71              | 151              |
| Black    | 155              | 200                           | 71             | 0.46              | 24               |
| Asian    | 61               | 123                           | 25             | 0.41              | 7                |
| White    | 844              | 2885                          | 179            | 0.21              | 96               |
|          | 1610             | 4057                          | 668            | 0.41              | 278              |

Census tracts within Los Angeles have an average of 0.41 TRI facilities per tract. Hispanic-dominated tracts have an average of 0.71 facilities, while White-dominated tracts have an average of 0.21 facilities. In Black- and Asian-dominated tracts, the average number of facilities is closer to the average for the county - 0.46 and 0.41, respectively.

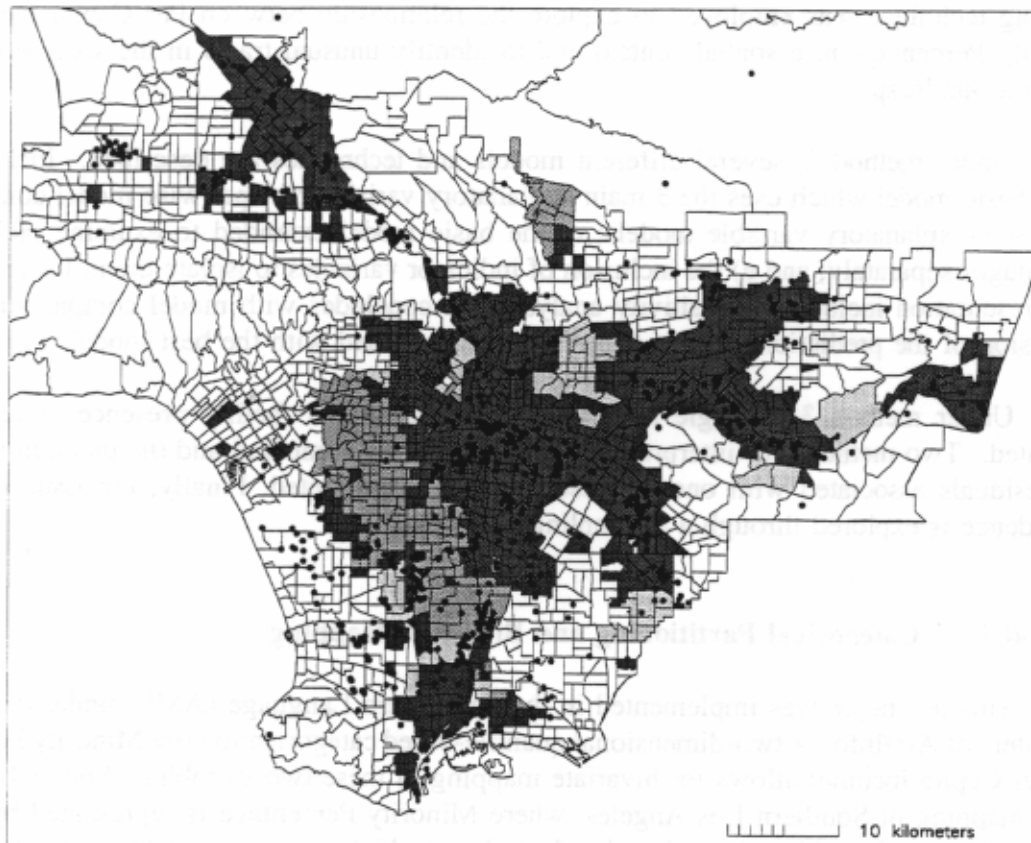
### Summary

The evidence presented in this chapter suggests that there is a relationship between TRI facility occurrence within census tracts and each of the explanatory variables (Minority Percentage, Per Capita Income, and Population Density) and that these variables are sufficiently unique that they can simultaneously be included in a generalized linear regression analysis. Additionally, it has been shown TRI facility occurrence is higher than average in Hispanic-dominated census tracts. This conclusion, though noteworthy, does not address the topic of this thesis - whether or not race is an important factor when the influence of other important variables have been controlled for. This issue is the subject of the next chapter.





<sup>11</sup> Note: One of these very low-density tracts, just south of Vernon, has been excluded from the analysis, with a population density of less than 100 persons per sq. km..

Figure 9

## Race, Ethnicity and TRI Facilities In Los Angeles County



Dominant Racial or Ethnic Group  
(Plurality in Each Census Tract)

-  Asian
-  Black
-  Hispanic
-  White (non-Hispanic)

Data Source:  
1990 U.S. Census  
1989 U.S. EPA Toxic Release Inventory

Map by Laretta Burke  
National Center for Geographic  
Information and Analysis,  
Santa Barbara, California

- Indicates TRI Facility Locations

## **Chapter 4. Analysis Results**

### **Introduction**

The results of three methods of analysis will be presented in this chapter: 1) bivariate mapping of minority proportion and income; 2) generalized linear modeling to a Poisson distribution (modeling the number of TRI facilities in a census tract); and 3) logit analysis (modeling presence or absence of TRI facilities in a census tract).

Under the first method, results of an exploratory analysis are presented. A bivariate mapping technique was employed to explore the relationship between Per Capita Income and Minority Percentage in a spatial context and to identify unusual tracts in the socio-economic industrial landscape.

Under method 2, several different models and techniques are described. Included are: a) the basic model which uses the 3 main explanatory variables; b) pairwise model comparisons of three 2-explanatory variable models; c) the basic model extended to examine racial/ethnic percentages separately; and e) the inclusion of indicator variables to isolate the influence of race. The presentation includes a sensitivity analysis and concludes with model comparisons, and a discussion of the prediction success and residuals associated with the best model.

Under method 3, a single logit model predicting TRI facility presence or absence is presented. Two methods for interpreting model results are compared and the prediction success and residuals associated with one of these methods is presented. Finally, the issue of spatial dependence is explored through a resampling method.

### **Method 1) Categorical Partitioning and Bivariate Mapping**

This technique was implemented in the Arc Macro Language (AML) under the Arcplot subsystem of Arc/Info. A two-dimensional quantile-based categorization (by Minority Percentage and Per Capita Income) allows for bivariate mapping of these two variables. Figure 10 reflects a 4x4 mapping of Southern Los Angeles, where Minority Percentage is represented by change in hue. Change from blue through red reflects low to high categories of Minority Percentage. Per Capita Income is represented through change in lightness (or strength) of the color. Pale shades represent low Per Capita Income categories, while more intense colors represent high income categories.

The bivariate mapping technique reveals the high correlation between income and minority composition in a spatial context. Of the 1608 tracts categorized in the 4 x 4 (16 category) partitioning, 67% were assigned to the four cells (categories) along the diagonal. (The key in Figure 10 reflects categorical assignments). As expected, occurrence decreases as one moves further from the diagonal. Only three census tracts were assigned to the highest income, highest Minority Percentage category, while only one tract is assigned to the lowest income, lowest Minority Percentage category.

The technique was useful for identifying census tracts that diverge from the dominant relationship between income and race. Tracts which have a high Minority Percentage and high Per Capita Income can be identified, mostly west of the poor urban core, including tracts near Inglewood, south of Norwalk and south and west of Compton. Similarly, tracts which have a low Minority Percentage and low Per Capita Income can be identified south of Burbank, north of San Fernando, and north of Covina. These tracts tend to have far fewer facilities than those in the high Minority Percentage, high Per capita Income groupings. Both categories are areas which are rare in the landscape, and could be interesting areas for subsequent analysis.

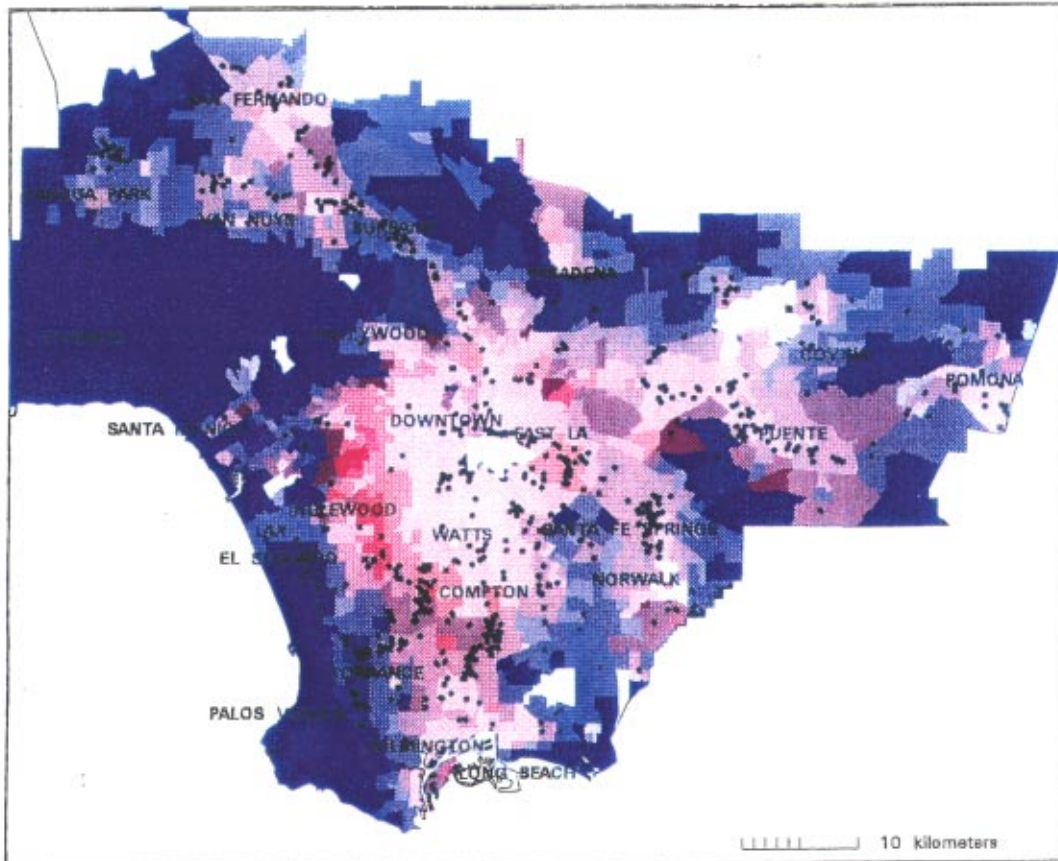
### **Method 2) Modeling the Number of TRI Facilities in a Tract: Generalized Linear Modeling to a Poisson Distribution**

As discussed earlier, owing to the highly skewed distribution of the dependent variable (number of TRI facilities in a census tract) the model was fitted to a Poisson distribution.

The primary variables of interest for their relationship with the number of TRI facilities in a tract (TRISITES) are median Per Capita Income (PCINC), Minority Percentage of the population (MINORITY), and Population Density (POPDENS). Minority Percentage of the population is the sum of the individual Hispanic, Black, Asian and Other categories. Also of interest is how the percentage of the population of each of these ethnic or racial groups relates to the dependent variable and the other explanatory variables.

Figure 10

**Bivariate Mapping of Minority Percentage and Per Capita Income  
by Census Tract in Los Angeles County**



| Minority Percentage |       |      |     | Per Cap Income |
|---------------------|-------|------|-----|----------------|
| low                 | ----- | high |     |                |
| 12                  | 86    | 12   | 3   | high           |
| 96                  | 224   | 67   | 17  |                |
| 4                   | 85    | 245  | 69  |                |
| 1                   | 8     | 80   | 312 | low            |

Data Source:  
1990 U.S. Census  
1989 U.S. EPA Toxic Release Inventory

Map by Laretta Burks  
National Center for Geographic  
Information and Analysis,  
Santa Barbara, California

• Indicates TRI Facility Locations

A set of multiple-linear regression models was examined in order to evaluate the relative relevance of the three primary explanatory variables Per Capita Income, Minority Percentage and Population Density. These are:

- 1) using all 3 variables to predict TRISITES; and
- 2) doing 3 pairings of 2 variables each to predict TRISITES.

(Table 14 summarizes results of all the Poisson models.)

**a) 3 Explanatory Variable Model:  
TRISITES <- Minority %, Population Density, Per Capita Income**

The three-explanatory variables used in this model:

- 1) **minority** - the percentage of the tract population who are Black, Asian, or Hispanic
- 2) **std.popdens** - a standardized version of the number of persons per square kilometer in the tract
- 3) **std.pcinc** - a standardized version of the median per capita income for the tract

In this model, all coefficients were significant at the .05 level.  $Rho^2$  for the model is 0.345. Population Density proved to be the most significant variable, with a t-value of 24.9. Minority Percentage and Per Capita Income were significant variables, with t-values of 10.2 and 7.9, respectively. The signs of the coefficients reflect that an increase in predicted TRISITES would result from either an increase in the Minority Percentage of the population; a decrease in Population Density; or a decrease in Per Capita Income. The 95% confidence interval for the coefficient associated with Minority Percentage is 1.8 to 2.6. A Chi-Square test was used to confirm that the model is significant at the .05 level. Table 7 provides a summary of the model results.

**Table 7 - Model 1 Results**

| <b>TRISITES &lt;- Minority %, Population Density, Per Capita Income</b> |              |                   |                |
|---|--------------|-------------------|----------------|
| Coefficients:   |              |                   |                |
|   | <u>Value</u> | <u>Std. Error</u> | <u>t-value</u> |
| (Intercept)   | -3.20        | 0.15              | -20.8          |
| minority  | 2.19         | 0.21              | 10.2           |
| std.popdens   | -2.20        | 0.08              | -24.9          |
| std.pcinc   | -0.82        | 0.10              | -7.9           |
| Null Deviance: 2934.981 on 1609 df                                      |              |                   |                |
| Residual Deviance: 1922.392 on 1606 df                                  |              |                   |                |
| $Rho^2 = 1 - (1922.392 / 2934.981) = 0.345$                             |              |                   |                |

**b) Comparison of Three 2-Explanatory Variable Models**

Running pairwise models including only two of Minority Percentage, Per Capita Income, and Population Density yielded similar results. In all three models both variables were significant, with t-values well over 2.0. The best of these three models used Minority Percentage and Population Density to predict TRISITES, with  $Rho^2$  of 0.318. The second best 2-explanatoryvariable model used Per Capita Income and Population Density ( $Rho^2$  of 0.307), while the weakest model included Minority Percentage and Per Capita Income ( $Rho^2$  of 0.054). Table 8 provides summaries of the three models.

A comparison of t-values from these models elucidates a relationship among the variables. In the first of these models, Population Density and Minority Percentage had t-values of 23.7 and 22.9, respectively. In the second model, Population Density and Per Capita Income had t-values of 23.6 and 19.4, respectively. In the third model, both t-values were much lower, at 4.3 for Minority Percentage and 3.2 for Per Capita Income. This indication of significance reflects the fact that Per Capita Income and Minority Percentage contain some redundant information, while Population Density contains much more unique information.

**Table 8 - Summary of Three 2-Independent Variable Models**

|   |
|---|
| <p><u>Model 2a</u><br/><u>TRISITES &lt;- Minority %, Population Density</u></p> <p>TRISITES &lt;- -3.9 + 3.5 * minority - 2.1 std.podens<br/>(22.9) (23.7)</p> <p>Rho<sub>2</sub> = 1-(2001.003/2934.981) = 0.318</p> <p>=====</p> <p><u>Model 2b</u><br/><u>TRISITES &lt;- Per Capita Income, Population Density</u></p> <p>TRISITES &lt;- -1.9 -1.6 * std.pcinc -2.0 std.podens<br/>(19.4) (23.6)</p> <p>Rho<sub>2</sub> = 1-(2034.65/2934.981) = 0.307</p> <p>=====</p> <p><u>Model 2c</u><br/><u>TRISITES &lt;- Minority %, Per Capita Income</u></p> <p>TRISITES &lt;- -1.6 + 1.0 minority -0.3 std.pcinc<br/>(4.3) (3.2)</p> <p>Rho<sup>2</sup> = 1-(2776.729/2934.981) = 0.054</p> |
|---|

### Sensitivity Analysis

A sensitivity analysis was performed to evaluate the sensitivity of Per Capita Income and Minority Percentage across different ranges of Population Density. The 1610 observation data set was stratified into five groupings based on Population Density, such that there were roughly equivalent numbers of observations in each grouping. The basic three-explanatory-variable model was applied to each of these five stratum. Table 9 provides a summary of this analysis.

As expected, the total number of TRI facilities in each stratum diminishes with increasing Population Density. Rho<sup>2</sup> also decreases in higher Population Density strata. Of the three variables, Minority Percentage appears the most sensitive to Population Density. The coefficient associated with Minority Percentage ranges from 2.9 in the lowest-density grouping to -2.1 in one of the higher-density groupings. The t-values associated with this coefficient drop from 10.4 to 0.37 with increasing Population Density. This coefficient is not significant or is borderline significant in four of the five strata.

The Per Capita Income variable is also sensitive to the Population Density partitioning. The magnitude of the coefficient associated with this variable generally increases with increasing Population Density, ranging from -0.6 in the lowest-density stratum to -3.4 in one of the higher-density stratum. The t-values associated with this variable are significant for all strata.

**Table 9- Sensitivity Analysis Summary**

| Density Threshold | # Tracts | Total TRIs | ---Coefficients and t-values--- |                |                | RHO <sup>2</sup> |
|-------------------|----------|------------|---------------------------------|----------------|----------------|------------------|
|                   |          |            | POPDENS                         | PCINC          | MINORITY%      |                  |
| < 1782            | 300      | 333        | -2.77<br>(6.6)                  | -0.59<br>(5.4) | 2.91<br>(10.4) | 0.41             |
| 1782-2959         | 326      | 117        | -3.79<br>(3.8)                  | -1.11<br>(3.6) | 1.09<br>(2.1)  | 0.21             |
| 2959-4055         | 329      | 96         | -1.37<br>(1.2)                  | -1.79<br>(3.9) | 1.35<br>(1.9)  | 0.22             |
| 4055-5991         | 328      | 82         | -1.16<br>(1.6)                  | -3.37<br>(4.6) | -2.06<br>(1.9) | 0.16             |
| > 5991            | 327      | 40         | -1.03<br>(3.0)                  | -2.54<br>(2.3) | -0.58<br>(0.4) | 0.15             |
| -----             |          |            |                                 |                |                |                  |
| Full Model:       |          |            |                                 |                |                |                  |
| < 100             | 1610     | 668        | -2.20<br>(24.9)                 | -0.83<br>(7.9) | 2.20<br>(10.2) | 0.35             |

Minority Percentage appears to have its greatest significance relative to Per Capita Income in the lowest-density grouping. At higher Population Density strata, Minority Percentage is not significant.

These results show that within the main 1610 observation model, much of the strength of Minority Percentage relative to Per Capita Income comes from observations which have a low Population Density. These tracts tend to be either low-income, minority-dominated and fairly industrial, or sparsely-populated, high-income and White-dominated. Census tracts which have ten or more TRI facilities are typically more extreme in degree of minority proportion than in degree of poverty. (This evaluation is based on a comparison of standardized values.) These tracts are highly influential observations.

### Interpretation

In the full model, all three explanatory variables are relevant in the relationship with TRISITES. Population Density appears to be the most significant of the three variables. The best two-explanatory-variable model (using Minority Percentage and Population Density) is a slightly better model than the one pairing Per Capita Income and Population Density. This suggests that, when all census tracts are included in the model, the Minority Percentage of a census tract is a slightly more significant factor in the relationship with TRISITES than is the median Per Capita Income for the tract. In the full model, Per Capita Income is an important variable, but the Minority Percentage appears to include much of the same information, and some additional relevant information which makes Minority Percentage slightly more valuable in the relationship with TRISITES.

In an analysis of the sensitivity of these coefficients to Population Density, it was revealed that much of the strength of the racial variable relative to the income variable comes from observations with a low Population Density and a large number of TRI facilities. Additionally, it was shown that the Per Capita Income variable is more stable across different Population Density strata than is the Minority Percentage variable.

### c) Influence of Individual Ethnic and Racial Percentages on TRISITES

The first set of models examined have confirmed the significance of Minority Percentage in the relationship with TRI facility occurrence. This section explores the importance of individual ethnic / racial categories in the relationship with TRISITES.

The Minority Percentage of the population examined earlier, is the sum of the Hispanic, Black, Asian and Other categories. As the population classified as "Other" is so small as to be insignificant, it was not considered further. The Hispanic, Black and Asian percentages of the census tracts were examined using generalized-linear-regression analysis, as was used previously. (The White

Percentage of the population could not also be included in this model, as multicollinearity would result and the model would break down computationally (Kennedy, 1985)).

Three new explanatory variables were included in the model:

- 1) **poph** - the proportion of the tract population classified as Hispanic
- 2) **popb** - the proportion of the tract population classified as Black
- 3) **popa** - the proportion of the tract population classified as Asian

When these three ethnic/ racial variables were used as the only explanatory variables for predicting TRISITES, Hispanic Percentage and Black Percentage were both significant at the .05 level (t-values of 12.6 and 3.8, respectively). The Asian Percentage of the population did not prove to be a significant variable (t-value 0.4). (Model summary is provided in the Table 14.)

The two racial/ethnic variables which proved significant (poph and popb) were subsequently included in a model with Population Density and Per Capita Income. In the resulting model all variables are significant with t-values of 25.1 for Population Density; 12.3 for Hispanic Percentage; 7.0 for Black Percentage; and 5.5 for Per Capita Income. With  $Rho^2$  of .361, this model appears slightly stronger than the model which used the single variable Minority Percentage as an indication of racial mix. Table 10 presents a summary of this model.

### Interpretation

Variables representing both the Hispanic and Black population percentages were highly in the relationship with TRISITES. The Asian Percentage of the population was not a significant variable. Of the four racial groups, Asians are the most uniformly distributed across the county, and this could account for the lack of significance of this variable.

**Table 10 - Model 3b Results**

| <b>TRISITES &lt;- Hispanic Pct., Black Pct., Per Capita Income, Population Density</b> |              |                   |                |
|--|--------------|-------------------|----------------|
| Coefficients:  |              |                   |                |
|  | <u>Value</u> | <u>Std. Error</u> | <u>t-value</u> |
| (Intercept)  | -3.07        | 0.13              | -23.8          |
| poph   | 2.79         | 0.23              | 12.3           |
| popb   | 1.59         | 0.23              | 7.0            |
| std.pcinc  | -0.57        | 0.10              | -5.5           |
| std.popdens  | -2.19        | 0.09              | -25.1          |
| Null Deviance: 2934.981 on 1609 d.f.   |              |                   |                |
| Residual Deviance: 1875.664 on 1605 d.f.   |              |                   |                |
| $Rho^2 = 1 - (1875.664 / 2934.981) = .361$   |              |                   |                |

### d) Inclusion of Indicator Variables

Indicator (or binary dummy) variables can be used to group observations and to make continuous variables discrete. In this analysis, indicator variables were used to examine the mean characteristics of different ethnic and racial groups at different thresholds. In particular, Hispanic vs. non-Hispanic groupings were examined, along with variables representing which racial group has the plurality (largest percentage) in each tract.

### Indicator Variable Model 1 - Hispanic Groupings

In previous analysis, the Hispanic Percentage proved to be the most valuable of the racial ethnic variables examined. But, inclusion of the Hispanic Percentage greatly reduces the significance of the Per Capita Income variable because the two are highly negatively correlated (at -.64) and hence, Hispanic Percentage already reflects much of the income information.



Through the use of an indicator variable which distinguishes highly Hispanic from notvery-Hispanic tracts, much of the detail is lost regarding the precise Hispanic Percentage of the tract, but average characteristics of these groups can be estimated. Relating to this, the high negative correlation between Hispanic Percentage and Per Capita Income is lost, so Per Capita Income becomes a more significant variable in the model.

2:)

Two variables were established to reflect the degree of Hispanic domination of the census tract. For moderately Hispanic tracts, HMID was set to 1 if the Hispanic Percentage was between 0.3 and 0.6 and set to 0 otherwise. For Highly Hispanic tracts, HHIGH was set to 1 if the Hispanic Percentage was 0.6 or greater. Of the 1610 tracts, the Hispanic Percentage was less than 0.3 859 times (53%); between 0.3 and 0.6 413 times (26%) and 0.6 or greater 338 times (21%).

The indicator variables HMID and HHIGH were combined in a model with Per Capita Income and Population Density to predict TRISITES. All variables in this model proved significant.  $Rho^2$  for the model is .335. Table 11 provides a model summary.

**Table 11 - Model 4a Results**

| <b>TRISITES &lt;- HMID, HHIGH, Per Capita Income, Population Density</b> |              |                   |                |
|--|--------------|-------------------|----------------|
| Coefficients:  |              |                   |                |
|  | <u>Value</u> | <u>Std. Error</u> | <u>t-value</u> |
| (Intercept)  | 1.81e+00     | 1.76e-01          | 10.29          |
| hmid   | 5.12e-01     | 1.14e-01          | 4.49           |
| hhigh  | 1.06e+00     | 1.21e-01          | 8.77           |
| pcinc  | -8.46e-05    | 8.03e-06          | -10.53         |
| popdens  | -5.97e-04    | 2.49e-05          | -23.93         |
| Null Deviance: 2934.981 on 1609 d.f.                                     |              |                   |                |
| Residual Deviance: 1952.366 on 1605 d.f.                                 |              |                   |                |
| $Rho^2 = 1 - (1952.366 / 2934.981) = .335$                               |              |                   |                |

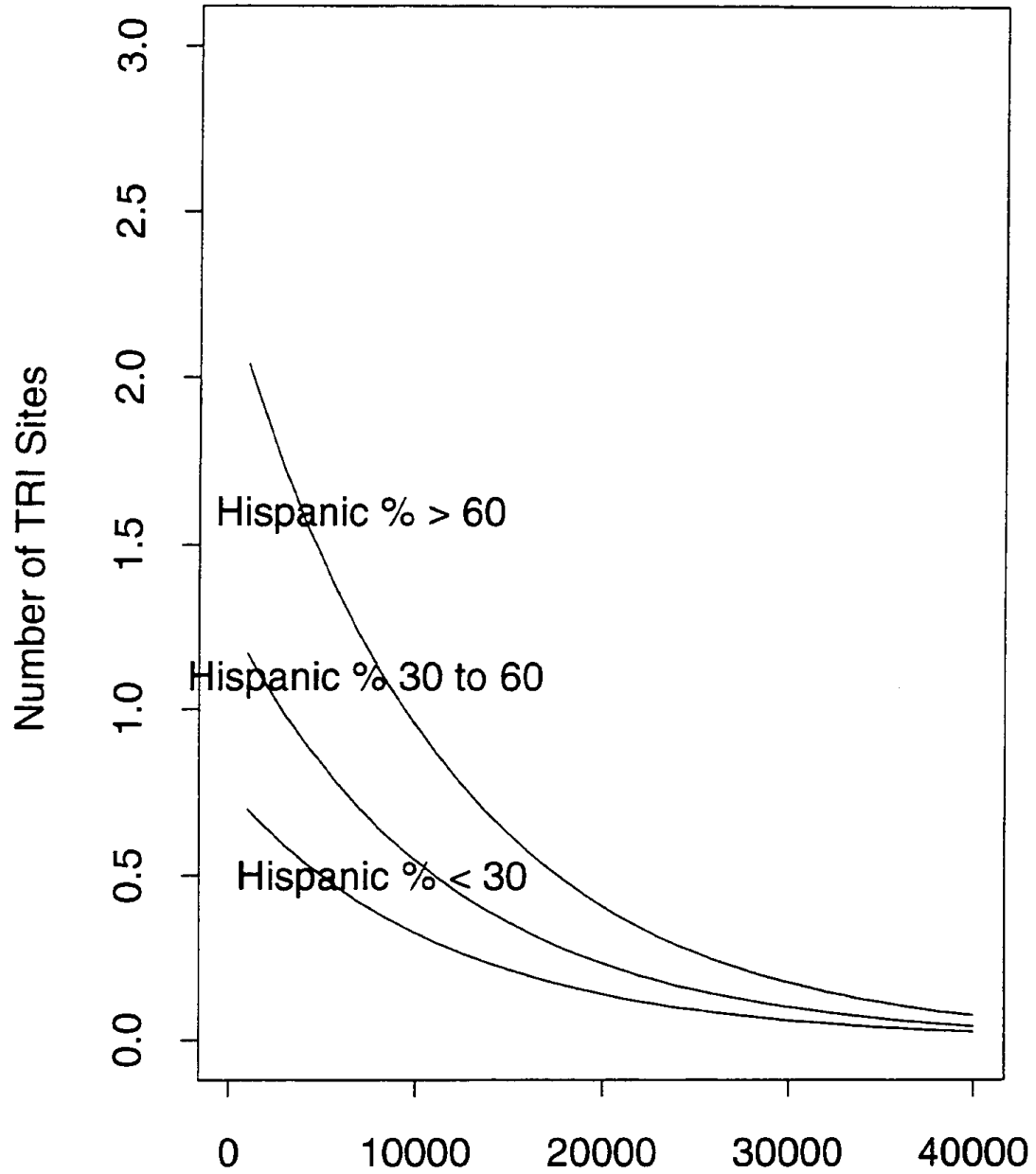
**Results**

A plot of the predicted number of TRISITES for the three classes of Hispanic dominance versus Per Capita Income reflects the influence of this variable on the model. (See Figure 11.) This plot represents the model prediction for a constant Population Density of 3490 - the median observed value for Population Density. It can be seen that, particularly at lower income levels, the predicted number of TRI facilities in a tract is much higher for tracts which are greater than 60% Hispanic than for those which are less than 30 % Hispanic. For a median Per Capita Income of \$8,000 and a Population Density of 3490 persons per km<sup>2</sup>, the predicted number of TRI facilities is 0.39 for tracts which are less than 30% Hispanic, 0.65 for tracts which are between 30 and 60% Hispanic, and 1. 13 for tracts which are greater than 60% Hispanic. This difference decreases at higher income levels. For example, for a median Per Capita income of \$20,000, the predicted number of TRISITES drops to 0.14, 0.23 and 0.41 for Hispanic Percentages of less than 30, between 30 and 60, and greater than 60, respectively.

**Indicator variable 2- Dominant Racial / Ethnic Group**

Four indicator variables were created reflecting which racial/ethnic group had the highest percentage in the tract. Of 1610 tracts, the largest percentage was White 844 times (52%); Hispanic 550 times (34%); Black 155 times (10%); and Asian 61 times (4%). Summary statistics were taken on this partitioning of data (See Table 12). The characteristics which are noteworthy from these statistics are that Hispanic dominated tracts are the poorest; are the densest; and have the highest mean number of TRISITES by a considerable margin. Also noteworthy is that the median Minority Percentage varies widely - is highest in the Black dominated tracts (at .98) followed by the Hispanic dominated tracts (at .88); the Asian dominated tracts (at .78) and the White dominated tracts( at .28).

**Figure 11 -** Predicted TRI Facilities as a function of Per Capita Income (by Hispanic Grouping)



Per Capita Income

Predictions assume median population density (3490 pp/sq.km.)

**Table 12 - Summary of 4 Category Partitioning**

|                   | ----Racial/Ethnic Plurality--- |              |                 |              | ALL           |
|-------------------|--------------------------------|--------------|-----------------|--------------|---------------|
|                   | <u>White</u>                   | <u>Black</u> | <u>Hispanic</u> | <u>Asian</u> | <u>Tracts</u> |
| number of tracts  | 844                            | 155          | 550             | 61           | 1610          |
| median PCINC      | 20375                          | 10325        | 8524            | 14591        | 14436         |
| median POPDENS    | 2740                           | 4533         | 4976            | 3805         | 3490          |
| median % White    | .715                           | .024         | .121            | .222         | .433          |
| median % Black    | .021                           | .613         | .027            | .035         | .027          |
| median % Hispanic | .135                           | .258         | .665            | .245         | .271          |
| median % Asian    | .074                           | .013         | .055            | .447         | .064          |
| median % Minority | .285                           | .976         | .880            | .778         | .567          |
| mean # TRI sites  | 0.21                           | 0.46         | 0.71            | 0.41         | 0.41          |

Of these indicator variables, only Hispanic Majority and Black Majority proved significant. In a model using these plus Population Density and Per Capita Income to predict TRISITES, all variables were significant. The model was significant with a Rho<sup>2</sup> of .347. A model summary is provided in Table 13.

**Table 13 - Model 4b Results**

| <b>TRISITES &lt;-</b>                               | <b>Hispanic Majority,</b>  | <b>Black Majority,</b> | <b>PC</b>      |
|---|----------------------------|------------------------|----------------|
|   | <b>Income, Pop Density</b> |                        |                |
| Coefficients:                                       |                            |                        |                |
|   | <u>Value</u>               | <u>Std. Error</u>      | <u>t-value</u> |
| (Intercept)   | 1.50e+00                   | 1.79e-01               | 8.36           |
| imh   | 1.19e+00                   | 1.14e-01               | 10.40          |
| imb   | 8.41e-01                   | 1.48e-01               | 5.67           |
| pcinc   | -7.39e-05                  | 7.92e-06               | -9.33          |
| popdens   | -6.01e-04                  | 2.50e-05               | -24.06         |
| Null Deviance: 2934.981 on 1609 df                  |                            |                        |                |
| Residual Deviance: 1915.637 on 1605 df              |                            |                        |                |
| Rho <sup>2</sup> = 1 - (1915.637 - 2934.981) = .347 |                            |                        |                |

The model predictions for these three groupings can be plotted against Per Capita Income (while holding Population Density constant). See Figure 12. The separate lines reflect predictions of number of TRISITES in a tract for tracts with a Hispanic plurality; Black plurality; and White or Asian plurality.

The plots reflect that for a given Per Capita Income, Hispanic dominated tracts have the highest number of TRISITES, followed by the Black dominated tracts, followed finally by White or Asian dominated tracts. At a median Per Capita Income of \$10,000 and a Population Density of 3490 persons per km<sup>2</sup>, the predicted number of TRISITES is 0.26 for tracts which have a White or Asian plurality; 0.61 for tracts with a Black plurality; and 0.86 for tracts which have a Hispanic plurality. Once again this difference decreases at higher income levels.

This model supports the conclusion that the Hispanic Percentage of the population is a significant variable, and illuminates the value of this variable, as distinct from the influence of income effects. The Black Percentage of the population is also a significant variable.

### **Model Prediction and Results**

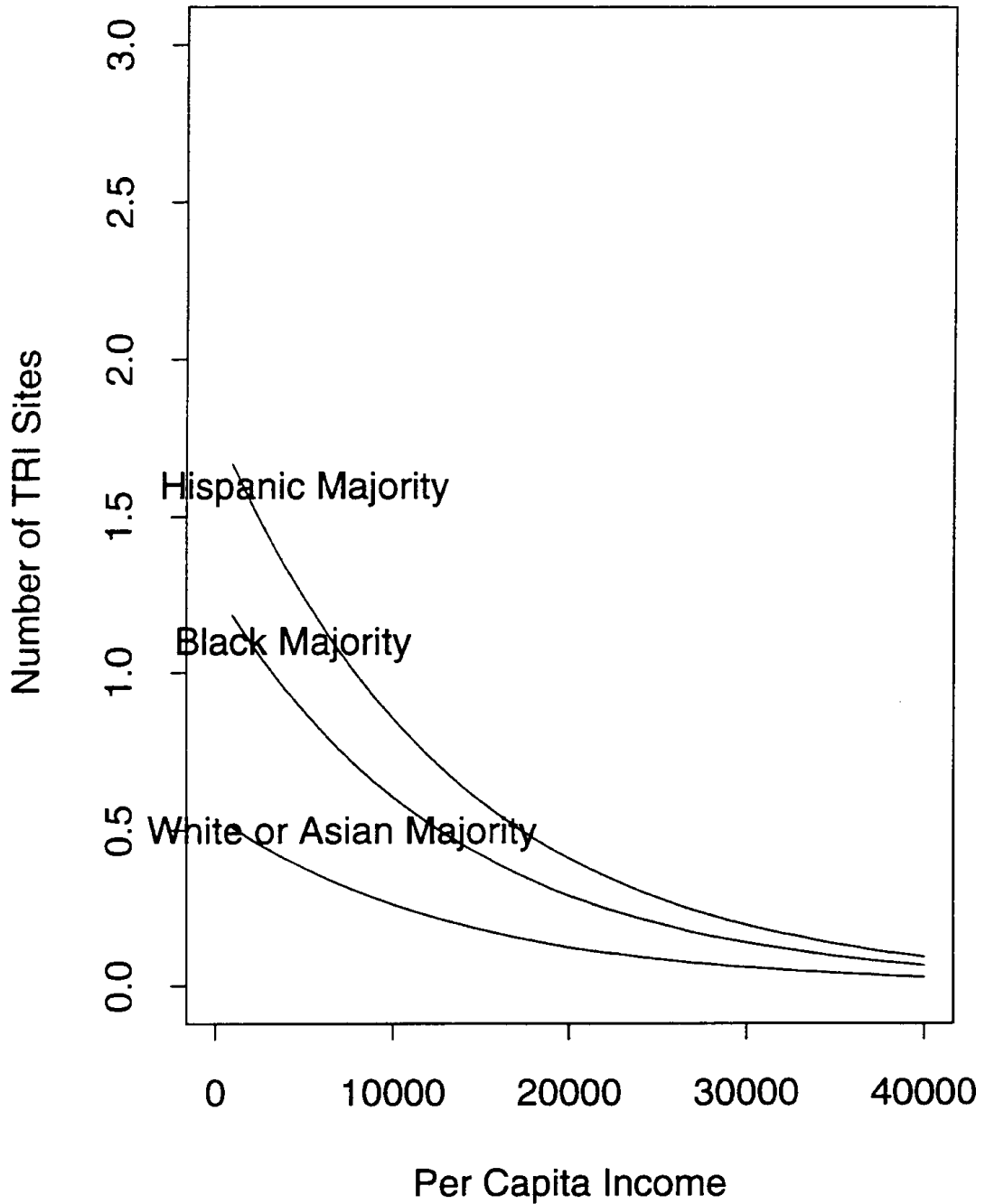
Table 14 provides a comparison of all the 1610 observation Poisson models used to predict TRISITES.

Of the models for predicting TRISITES which have been described, the four-variable model using Hispanic Percentage, Black Percentage, Per Capita Income and Population Density appears to have the best predictive capability, as indicated by its low residual deviance and high  $Rho^2$  (0.361). The actual values of TRISITES ranged between 0 and 25, with a mean of 0.41. The fitted values from this model ranged from near 0 to 9.5, with a mean of 0.41. The deviance residuals associated with these predictions ranged from -3.5 to 9.3, with a mean of -0.3.

A plot of the predicted values from this model combined with the actual location of the TRI facilities indicates a good relationship in most areas. (Figure 13). It can be seen, though, that there are many areas where the model predicted a relatively high number of TRISITES where none occur and areas where the predicted value is low and facilities do occur. These errors can be examined more closely through an examination of the residual deviance for each census tract.

In tracts where there is no TRI facility, the model overpredicted to a varying degree. The mean fitted value was 0.29 for tracts where no facility was present, and 1.00 for tracts which had one or more facilities. In tracts where there is a large number of TRI facilities, the model always severely underpredicts. Figure 14 reflects the spatial distribution of large overprediction and underprediction by the model. Dark grey reflects underprediction, while light grey reflects overprediction. The discs in each census tract are scaled to the magnitude of the deviance residual. Deviance residuals range between -3.5 and 9.3 for this model. Only tracts where the deviance residual is greater than 1 or less than -1 are displayed.

**Figure 12 -** Predicted TRI Facilities as a function of Per Capita Income (by Racial / Ethnic Grouping)



Predictions assume median population density (3490 pp/sq.km.)

**Table 14 - Comparison of all Poisson Models**

| Dependent Variable TRISITES |   | (t-value in parenthesis) |                 |              |                   |                        |
|-----------------------------|---|--------------------------|-----------------|--------------|-------------------|------------------------|
| <u>Model #</u>              | <u>Independent Variables</u>                      | <u>#</u>                 | <u>Deviance</u> |              | <u>df</u>         | <u>Rho<sup>2</sup></u> |
|                             |   |                          | <u>Null</u>     | <u>Resid</u> | <u>Null,Resid</u> |                        |
| 1)                          | minority, DENS, PCI<br>(10.2) (24.9) (7.9)        | 3                        | 2935            | 1922         | 1609,1606         | 0.345                  |
| 2a)                         | minority, DENS<br>(22.9) (23.7)                   | 2                        | 2935            | 2001         | 1609,1607         | 0.318                  |
| 2b)                         | PCI, DENS<br>(19.4) (23.6)                        | 2                        | 2935            | 2035         | 1609,1607         | 0.307                  |
| 2c)                         | minority, PCI<br>(4.4) (3.2)                      | 2                        | 2935            | 2777         | 1609,1607         | 0.054                  |
| 3a)                         | popb, popa<br>(12.7) (3.8) (0.4)                  | 3                        | 2935            | 2752         | 1609,1606         | 0.062                  |
| 3b)                         | popb, PCI, DENS<br>(12.3) (7.0) (5.4) (25)        | 4                        | 2935            | 1876         | 1609,1605         | 0.361                  |
| 4a)                         | hmid, hhigh, PCI, DENS<br>(4.5) (8.8) (10.5) (24) | 4                        | 2935            | 1952         | 1609,1605         | 0.335                  |
| 4b)                         | imb, PCI, DENS<br>(10.4) (5.7) (9.3) (24.1)       | 4                        | 2935            | 1915         | 1609, 1605        | 0.347                  |

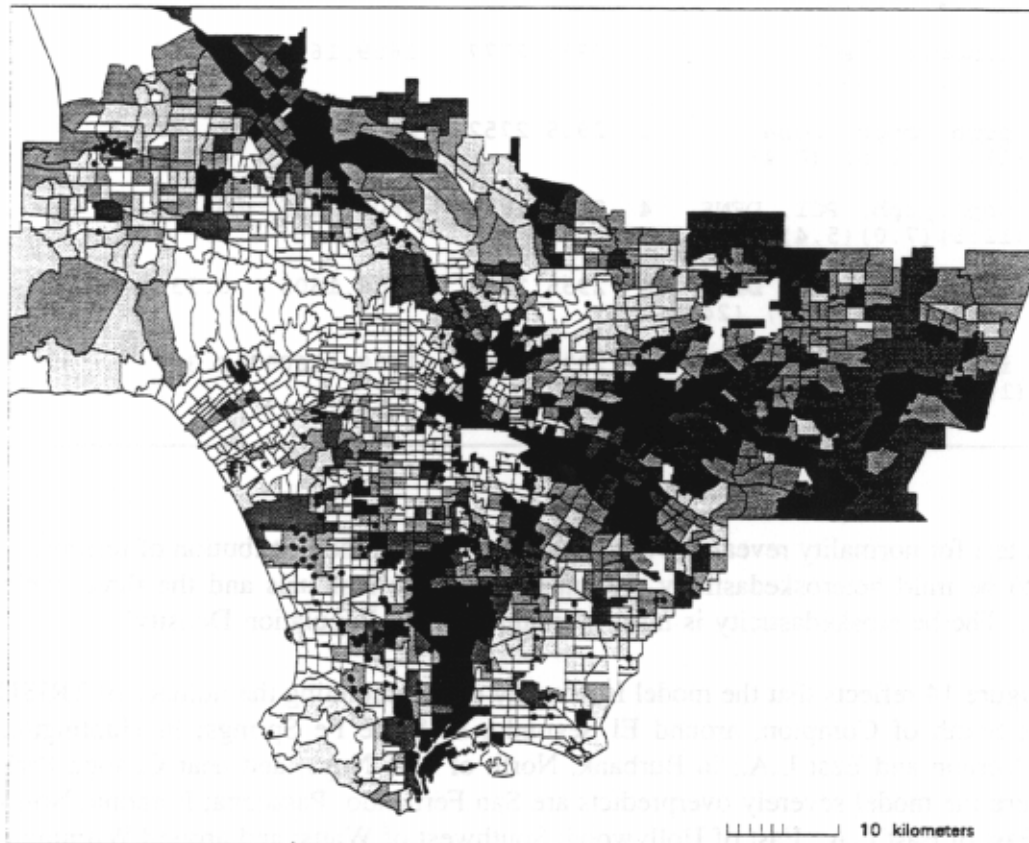
A test for normality revealed a slight positive skew in the distribution of residuals. There appears to be mild heteroskedasticity present between the residuals and the three explanatory variables. The heteroskedasticity is most pronounced with Population Density.

Figure 14 reflects that the model is severely underpredicting the number of TRISITES in Torrance; South of Compton; around El Segundo; in Sante Fe Springs; in Huntington Park; between Vernon and East L.A.; in Burbank; North of Van Nuys; and near Canoga Park. The areas where the model severely overpredicts are San Fernando; Pasadena; Pomona; North of Le Puente; East of East L.A.; East of Hollywood; Southwest of Watts; and around Wilmington. In some areas, such as Sante Fe Springs, East L.A. and San Fernando, we see locations of high overprediction and high underprediction in close proximity to one another. It can be observed in Figure 14 that the model often predicts a relatively high value for a number of contiguous census tracts, as adjacent tracts are often similar demographically, but the TRI facilities occur clustered in only some of these tracts. This reality results in the case of high overprediction adjacent to high underprediction.






The spatial distribution of residuals suggests that there is spatial dependence which has not been accounted for in the Poisson model. This reduces the reliability of model results. Spatial dependence is examined under the next method - logit analysis.

Figure 13

**Predicted Values from Poisson Model (facilities per tract)  
and Actual TRI Facility Locations**



**Predicted TRI Facilities per Tract**

-  < 0.12
-  0.12 - 0.24
-  0.25 - 0.49
-  0.50 - 0.99
-   $\geq 1.00$

Data Source:  
1990 U.S. Census  
1989 U.S. EPA Toxic Release Inventory

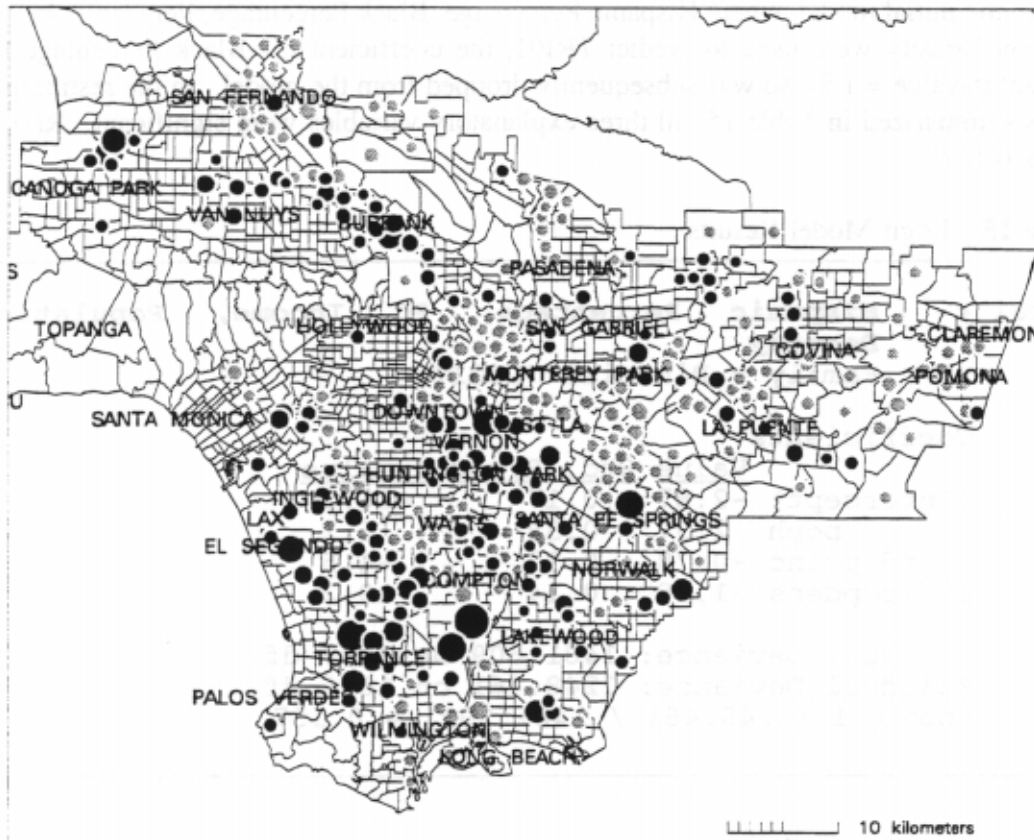
Map by Laretta Burke  
National Center for Geographic  
Information and Analysis,  
Santa Barbara, California

- Indicates TRI Facility Locations

Figure 14

## Residuals from Poisson Model

(Deviance Residuals greater than 1)



Large Residuals  
Disc size Scaled to Residual Deviance

- Underprediction
- Overprediction

Data Source:  
1990 U.S. Census  
1989 U.S. EPA Toxic Release Inventory

Map by Loretta Burke  
National Center for Geographic  
Information and Analysis,  
Santa Barbara, California



### Method 3) Modeling Presence or Absence of Facilities in a Census Tract: Logit Analysis

A logit model was employed to examine presence or absence of TRI facilities in a census tract. The same explanatory variables as were used in the "preferred" Poisson model were used here (Hispanic Percentage, Black Percentage, Per Capita Income and Population Density), and the same 1610 observations. The dependent variable of interest is TRI01, a binary variable for which a value of 1 indicates the presence of at least one facility in a census tract.

#### Results

In the initial model, where Hispanic Percentage, Black Percentage, Per Capita Income and Population Density were used to predict TRI01, the coefficient for Black Percentage was not significant (t-value = 1.8), so was subsequently dropped from the model. In the resulting model, which is summarized in Table 15, all three explanatory variables were significant.  $Rho^2$  for the model is 0.157.

**Table 15 - Logit Model Results**

```
TR01 <- Hispanic Percentage, PC Income, Population
Density
Family = Binomial (LOGIT)

Coefficients:
      Value Std. Error t-value
(Intercept) -2.87    0.17   -16.5
      poph    2.37    0.38     6.1
      std.pcinc -0.78    0.17    -4.5
      std.popdens -1.48    0.14   -10.2

Null Deviance: 1481.509 on 1609 df
Residual Deviance: 1248.461 on 1606 df
Rho2 = 1 - (1248.461 / 1481.509) = 0.157
```

The predicted or fitted values from this model, which reflect the predicted probability of a TRI facility being present in a given tract, ranged from near 0 to 0.81. The mean fitted value is 0.173, which is equivalent to the actual occurrence of tracts with TRI facilities present (17.3%). There are several methods for interpreting these fitted values, though the most common practice is to use a threshold of 0.5 for interpretation of the model prediction. Fitted values below 0.5 are interpreted as a prediction that no facility is present, while fitted values of 0.5 and above are interpreted as a prediction of facility presence. Table 16 summarizes model prediction success using the threshold of 0.5. Given this threshold, the model predicts only 71 tracts (4.4%) as having facilities. Of these, 41 are correct, so the success rate for predicting presence is 57.7%. Similarly, of the 1539 tracts where the model predicts absence, 1302 are correct, so the success rate for predicting facility absence is 84.6%. Overall raw prediction success using the 0.5 threshold is 83.4%.

**Table 16 - Prediction Success for Logit Model (0.5 Threshold)**

| Actual<br>TRI01                  | Predicted Value |               | Actual<br>Total | Actual<br>Share | %<br>Correct |
|----------------------------------|-----------------|---------------|-----------------|-----------------|--------------|
|                                  | < 0.5<br>(0)    | >= 0.5<br>(1) |                 |                 |              |
| 0                                | 1302            | 30            | 1332            | 82.7%           | 97.7%        |
| 1                                | 237             | 41            | 278             | 17.3%           | 14.7%        |
| Predicted Total                  | 1539            | 71            | 1610            |                 |              |
| Predicted Share                  | 95.6%           | 4.4%          | 100%            |                 |              |
| % Correct                        | 84.6%           | 57.7%         | 83.4%           |                 |              |
| Normalized<br>Prediction Success | -0.1            | 3.1           |                 |                 |              |

An alternative way of looking at prediction success is to examine the success given the actual condition rather than the predicted condition. Here, we find that 98% of tracts without facilities are correctly predicted, while only 15% of tracts with facilities are predicted correctly. Prediction success is uneven as a result of the differential difficulty of predicting facility presence vs. absence. As 82.7% of all tracts have no facility, it is much easier to correctly predict facility absence than presence. As a result, a normalized measure of prediction success can be calculated by subtracting the predicted share from the percent correct for both predicted absence and presence, and then dividing by the actual share for each of those categories (Wrigley, 1985). This measure indicates that prediction success is only slightly better than chance for predicting facility absence, but is fairly successful in predicting the more difficult facility presence.

Although the prediction interpretation threshold of 0.5 is the most common in practice, other methods can be considered. Of the 1610 tracts, 278 have facilities present. The threshold could be adjusted downward to a level where the proportion of tracts predicted to have a facility equals the proportion in reality. Using a threshold of 0.29 achieves this objective and results in the model prediction presented in Table 17. Using the threshold of 0.29, prediction success across the actual categories becomes somewhat more even (88% and 43%), and the normalized prediction success has become slightly more even (0.1 and 1.5). The overall raw prediction success of the model is reduced slightly, to 80.4.

Figure 15 shows spatially the results of classification using the 0.29 threshold. Areas which are lightly shaded were classified correctly. Areas which are dark grey have facilities, but have predicted probabilities of having a facility of less than 0.29. Areas which are light grey do not have any facilities in the tract, but have a predicted probability of greater than 0.29 of having one. Areas northeast of Monterey Park, north and east of East L.A., north of Sante Fe Springs, north of La Puente and around San Fernando stand out as areas where the model "overpredicted", when the 0.29 threshold was used. Far fewer tracts would be classified in this category if the 0.5 threshold were used.

**Table 17 - Prediction Success for Logit Model (0.29 Threshold)**

| Actual<br>TRI01                  | Predicted Value |                | Actual<br>Total | Actual<br>Share | %<br>Correct |
|----------------------------------|-----------------|----------------|-----------------|-----------------|--------------|
|                                  | < 0.29<br>(0)   | >= 0.29<br>(1) |                 |                 |              |
| 0                                | 1174            | 158            | 1332            | 82.7%           | 88%          |
| 1                                | 158             | 120            | 278             | 17.3%           | 43%          |
| Predicted Total                  | 1332            | 278            | 1610            |                 |              |
| Predicted Share                  | 82.7%           | 17.3%          | 100%            |                 |              |
| % Correct                        | 88.1%           | 43.2%          | 80.4%           |                 |              |
| Normalized<br>Prediction Success | 0.1             | 1.5            |                 |                 |              |

Using the 0.29 threshold reduces the number of tracts where "underprediction" has occurred. These tracts, shaded in dark grey, have facilities, but have a predicted probability of having a facility which is less than 0.29. The areas which stand out are similar to those areas identified through an examination of residuals from the Poisson model. They are LAX and El Segundo, Torrance, around Compton, Huntington Park, Burbank, Van Nuys, and Canoga Park. As the use of a 0.29 threshold is generous in classifying tracts as having a facility, these "underpredictions" can be considered severe.

An analysis of the deviance residuals associated with this model supports this conclusion. The deviance residuals associated with this model ranged from -1.52 to 3.39, with a mean of -0.15. A plot of the deviance residuals squared reveals the areas of extreme under or overprediction, and a broader pattern of misclassification emerges. Figure 16 presents negative deviance residuals (overpredictions) in light grey and positive deviance residuals (underpredictions) in dark grey. In this case, severe underprediction reflects tracts where the probability is quite low that a facility would be present, given the Population Density, Per Capita Income and Hispanic Percentage of the tract, but at least one facility is present. Unlike the examination of residuals from the Poisson model, where severe underprediction tended to identify tracts with a large number of TRI facilities, these residuals show the least likely tracts for a facility to occur where one does occur.

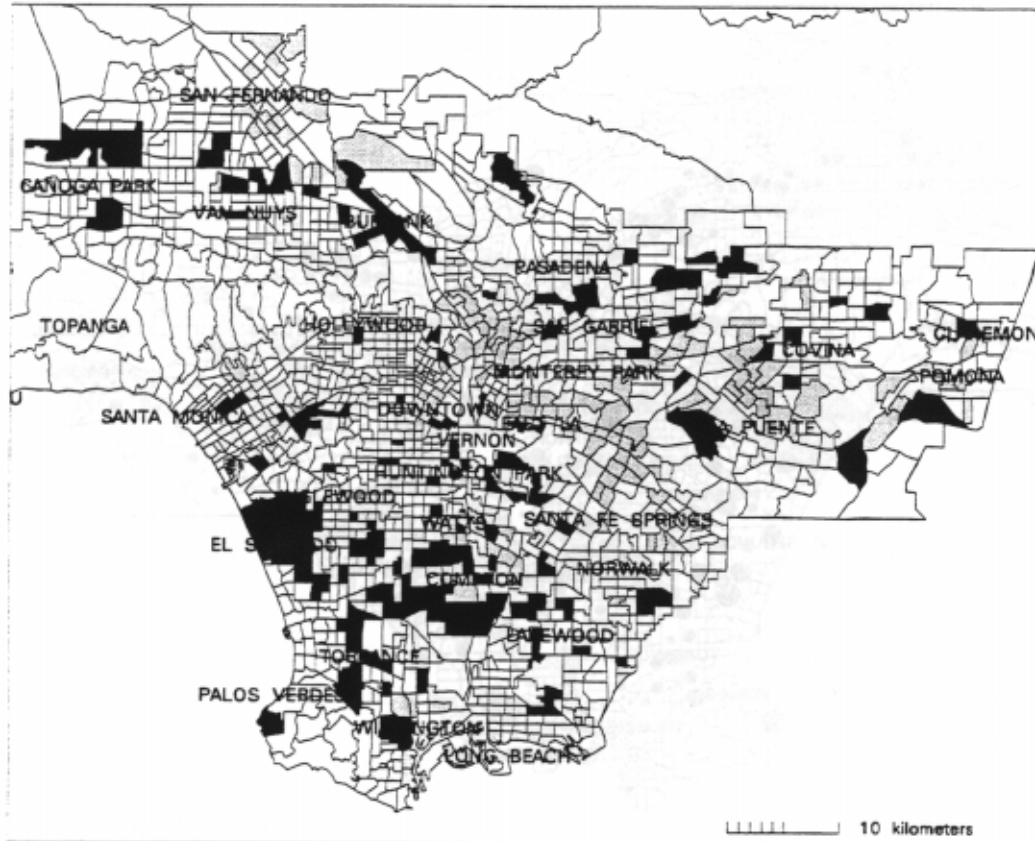
### **Spatial Dependence**

The importance of spatial dependence within the context of the logit model was evaluated through a sensitivity analysis involving the resampling technique described in Chapter 2. The procedure involves first selecting a minimum distance threshold representing the neighborhood surrounding a census tract. Next, a census tract is randomly selected as the first observation in the "resampled data set", and all tracts whose polygon centroid lies within the threshold are excluded as neighbors. This procedure is then repeated until the original set of observations is exhausted. A logit model is then run on the randomly selected resampled data set. This process was repeated 2000 times for each distance threshold, resulting in a sample of 2000 estimates of each coefficient, associated t-values and residual deviance. From these, a single set of coefficients can be estimated. The larger the distance threshold selected, the larger the average number of neighbors each tract will have; and the smaller the average number of observations in each model run.

This technique was applied to the same variables as were used in the original logit model. (See Table 15 for summary of original logit model). The explanatory variables Population Density, Per Capita Income and Hispanic Percentage were used to predict presence or absence of facilities within the tract (TRI01). The method was applied for distance thresholds of one, two and three kilometers and 2000 iterations. Beyond the three kilometer threshold the average number of observations in the resampled subsets become too low for the results to be reliable.

Figure 15

## Prediction Success for the Logit Model Prediction Threshold of 0.29



### TRI Facility Presence

|  |             |                |
|--|-------------|----------------|
|  | ACTUAL: YES | PREDICTED: YES |
|  | ACTUAL: YES | PREDICTED: NO  |
|  | ACTUAL: NO  | PREDICTED: NO  |
|  | ACTUAL: NO  | PREDICTED: YES |

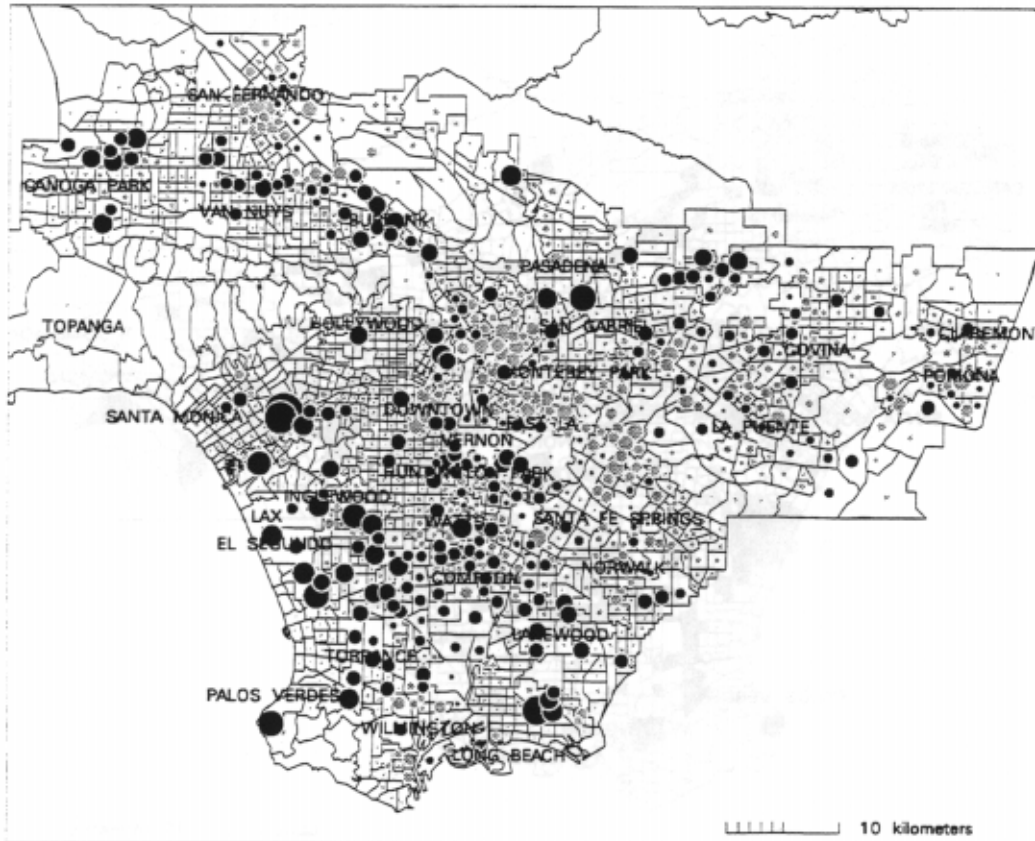
Data Source:  
1990 U.S. Census  
1989 U.S. EPA Toxic Release Inventory

Map by Laretta Burke  
National Center for Geographic  
Information and Analysis,  
Santa Barbara, California

Figure 16

## Residuals from Logit Model

(Deviance Residuals Squared)



### Logit Residuals

Disc-size scaled to square of Residual Deviance

- Underprediction
- Overprediction

Data Source:

1990 U.S. Census

1989 U.S. EPA Toxic Release Inventory

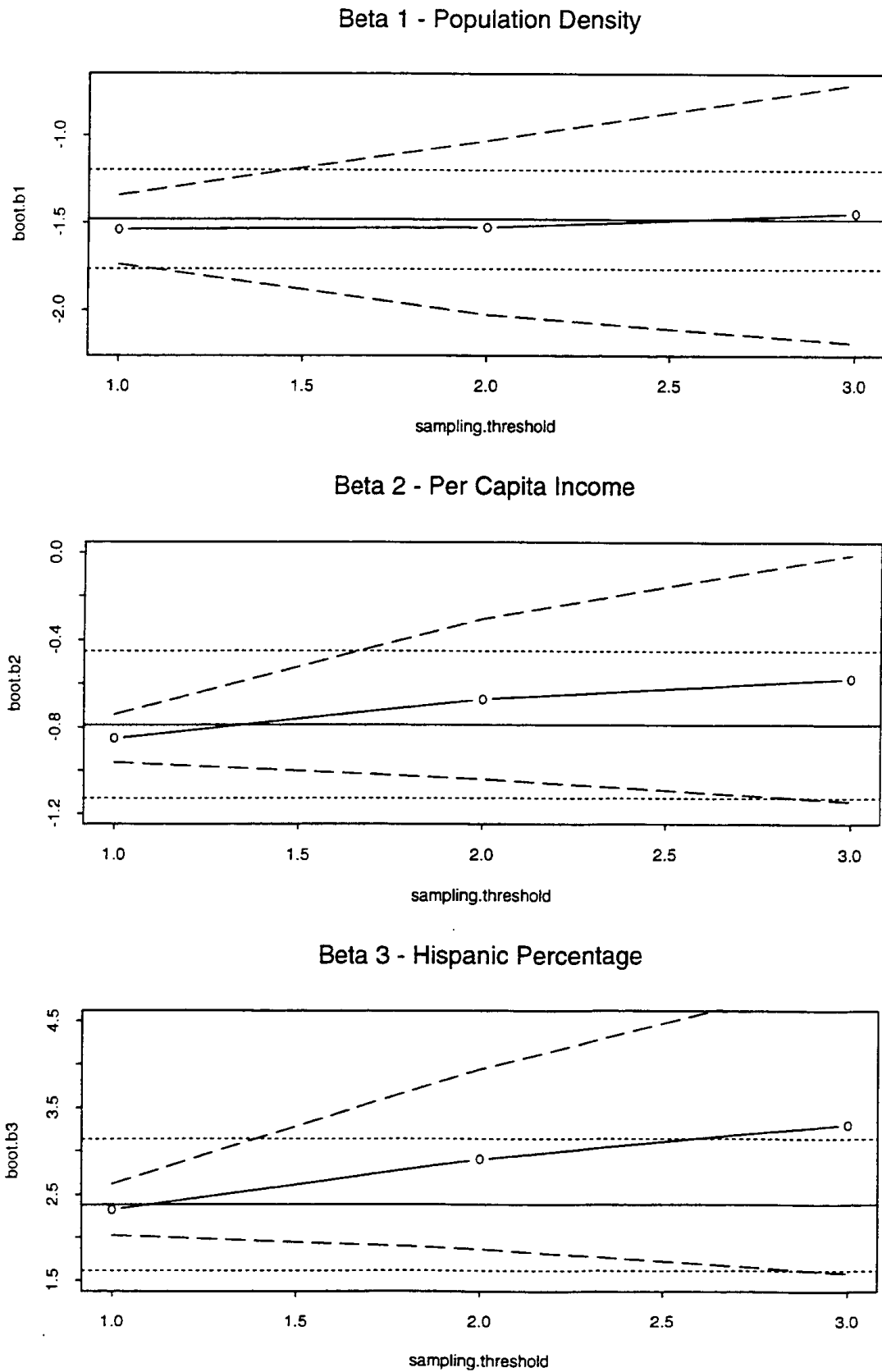
Map by Laretta Burke  
National Center for Geographic  
Information and Analysis,  
Santa Barbara, California

## Results

As the minimum distance threshold was increased from one to three kilometers, the magnitude of the coefficients associated with Population density and Per Capita Income decreased slightly (7% and 32%, respectively), while the magnitude of the coefficient associated with Minority Percentage increased slightly (41%). The t-values associated with these coefficients decreased considerably with the increased threshold, though all coefficients were still significant at the three kilometer threshold. Table 18 and Figure 17 compare the estimated values for all coefficients at the three distance thresholds with the coefficients estimated in the full logit model. Ninety-five percent confidence intervals were calculated for both the full model coefficients and the coefficients estimated through resampling and are presented in Figure 17. The straight solid line represents the coefficient estimated in the full model, while the fine dashed line reflects the 95% confidence interval for this estimate. The three points along a solid line represent the coefficient estimates for the three distance thresholds, while the coarse dashed lines reflect 95% confidence intervals for these estimates. As the distance threshold increases, the standard error associated with the estimate increases, resulting in wider 95% confidence intervals about the estimates.

The coefficients estimated for Population Density appear the least sensitive to the distance threshold selected. There is little difference between the coefficient estimated for Population Density in the full model as compared with the three estimated through the resampling method. There is slightly more change in the parameter estimates for Per Capita Income, though the coefficients estimated through resampling always lie within the 95% confidence interval for the full model coefficient. The parameter estimates for Hispanic Percentage appear the most sensitive to the distance threshold selected. When a one kilometer threshold is used, the estimated coefficient is quite similar to that of the full model. In the case of the three kilometer threshold, the estimated coefficient slightly exceeds the 95% Confidence Interval for the full model coefficient.

Figure 17 - Logit Model Coefficients Estimated through Resampling



**Table 18 - Comparison of Coefficients from the Full Logit Model with those Estimated through the Resampling Technique at 3 Distance Thresholds**

| <u>Distance Cutoff (m)</u> | <u>Avg. # Neighb.</u> | <u>Avg. # Obs.</u> | <u>Intercept B0</u> | <u>POPdens B1</u> | <u>PCINC B2</u> | <u>POPH B3</u> |
|----------------------------|-----------------------|--------------------|---------------------|-------------------|-----------------|----------------|
| 1000                       | 1.3                   | 1111               | -2.79               | -1.53<br>(15.2)   | -0.85<br>(15.2) | 2.31<br>(14.2) |
| 2000                       | 7.7                   | 429                | -3.02               | -1.52<br>(5.7)    | -0.66<br>(3.5)  | 2.91<br>(5.5)  |
| 3000                       | 17.4                  | 235                | -3.24               | -1.42<br>(3.7)    | -0.58<br>(2.0)  | 3.26<br>(3.8)  |
| -----                      |                       |                    |                     |                   |                 |                |
| 0 (Full Model)             | -                     | 1610               | -2.87               | -1.48<br>(10.3)   | -0.79<br>(4.6)  | 2.38<br>(6.1)  |

This analysis has allowed for an examination of the influence of spatial dependence on the error term of the model. The results suggest that spatial dependence has not had much of an effect on the model's coefficients. As increasing the distance threshold to three kilometers results in increasing the magnitude of the coefficient associated with Hispanic Percentage and decreasing the magnitude of the coefficients associated with Population Density and Per Capita Income, it is safe to assume that the Hispanic Percentage variable is at least as important as it appears in the full model.

**Summary**

Several conclusions can be drawn from these three very different methods of analysis. Consistently, race (as represented by Minority Percentage) proved to be an important variable in the relationship with TRI facility occurrence in a census tract. Both in the case of modeling the number of TRI facilities in a tract (Poisson model) and in the case of modeling the presence or absence of TRI facilities (Logit model), Minority Percentage proved to be an important variable, while the influence of Per Capita Income and Population Density were controlled. Within the full model, Minority Percentage appears to have a slightly stronger relationship with the number of TRI facilities in a census tract than does Per Capita Income. A sensitivity analysis suggests that much of the strength of the Minority Percentage variable comes from observations with low Population Density and many TRI facilities.

Of the individual racial / ethnic groups examined for Los Angeles, Hispanics appear to be most disproportionately exposed to TRI facilities. The Black population also appears to have a greater exposure to TRI facilities than would be expected based upon income and population density.

The results of the logit analysis strengthened the conclusions drawn from the Poisson analysis by showing that results were not dependent upon a few highly influential industrial tracts. Also, it was shown that spatial dependence does not significantly affect the results of the logit analysis.

Residuals from the Poisson and logit models will be combined with knowledge gained through the bivariate mapping in order to identify neighborhoods for more detailed analysis of the process which has resulted in the current socio-environmental landscape. The detailed neighborhood analysis will be the subject of the next chapter.



## **Chapter 5. Neighborhood Analysis**

This chapter presents the method and results of a more detailed analysis of three areas or neighborhoods". The chapter begins with a brief description of the purpose of the detailed analysis, and is followed by descriptions of the method of neighborhood selection; research methods employed; and study results.

### **Detailed Study Purpose and Methodology**

The statistical analyses described in the previous chapter provide evidence of environmental inequity in Los Angeles. It is apparent that low-income and minority individuals generally live in closer proximity to TRI facilities than do higher-income and White individuals. These statistical analyses describe a static condition - a "snapshot in time", but do not provide any insight into the processes which have resulted in this socio-environmental landscape. The purpose of this neighborhood analysis is to provide more detail regarding the development history of a few study areas within the County. A closer examination of the physical features of each area, the zoning for the area, and the timing of industrial and residential development was conducted with the goal of providing insight regarding the factors which have resulted in the current socio-environmental landscape.

### **Method for Neighborhood Selection**

Information gained from the analysis of residuals from the Poisson and logit models and the bivariate mapping were jointly used in the selection of three neighborhoods for further study. The selection method was somewhat ad-hoc, employing some constraints and flexible criteria regarding the study areas. It was desirable to select study areas with at least moderate population density ( greater than 1000 pp/sq.km.), as to exclude the highly industrial, non-residential tracts. At the same time, it was desirable to select areas with at least a moderate number of TRI facilities. Additionally, rather than examining three study areas of similar socio-economic characteristics, it would be preferable to examine three neighborhoods of varied ethnic composition and economic class.

In a demographic landscape as varied as that of Los Angeles, it was not difficult to identify neighborhoods which met these criteria. Utilizing the large positive residuals from the Poisson model tended to distinguish tracts with many more facilities than would be expected given the population density, minority composition and per capita income of the tract. Overall, these residuals tended to point toward tracts with a large number of TRI facilities or moderate to high population density. The residuals from the logit model tended to distinguish tracts with a moderate to high population density, or a high per capita income. The bivariate mapping was then used to identify neighborhoods which diverged from the strong negative relationship between income and minority percentage, such as high income areas with a high minority percentage.

Three neighborhoods or study areas were selected for further analysis:

- 1) Huntington Park - This area was selected because of the large concentration of census tracts which had high residuals in both the Poisson and logit models. This area is at the heart of the poor inner city, and is located just south of Vernon, regarded by many as the most polluted place in Los Angeles. The Huntington Park area is very densely populated and is almost exclusively Hispanic.
- 2) Burbank - This area was also selected for study due to the large residuals in both the Poisson and logit models. The area is densely populated, is of moderate income, and has a moderate percentage of minority residents. The area is ethnically diverse, with a mix of White, Hispanic and Asian residents, though TRI facilities tend to be in tracts which have a higher Hispanic percentage.
- 3) Compton / Gardena (west of Compton) - This area stood out in the bivariate mapping as an area with a moderate population density, moderate per capita income and a moderate minority percentage. This area also had large residuals in both the Poisson and logit models. This study area is ethnically mixed, with a significant percentage of all four ethnic groups.

### **Research Methods**

Several sources of information and collection techniques were used in the neighborhood analysis. These included site visits, telephone interviews, and use of existing materials such as zoning maps. This information was synthesized in a descriptive fashion.

Information on the history of TRI facilities within the neighborhoods was obtained through telephone interviews with 50 TRI facility operators. Questionnaires asking for the same information were sent to several others which could not be contacted by phone. (Appendix D contains the questionnaire which was used for the phone interviews and mail survey.) The survey focused on when the facility began operation at the current location; whether it was the original location for the business; what factors influenced the

decision to site the facility in that location; and on the nature of the surrounding area- whether it is purely industrial, highly residential or mixed. Some facilities were also asked about the number of employees; what percentage came from the surrounding community; and about the nature of the pollution associated with their industrial process and the pollution control technologies they were using. Unlike the mailed questionnaires, the response rate on the phone survey was very high. Only two facilities declined to provide information over the phone.

Each neighborhood was also visited to obtain a better sense of the nature and character of the study sites, the aura of the TRI facilities, and to verify the actual proximity of the industrial sites to residential areas. The site visits were made on a Sunday, while few facilities were operating.

Both historic and current maps of the area were used to provide additional information on important features, such as roads, railroads and airports; on zoning for each area; and to identify changes in the community over time. Information on when residential structures were built was obtained from the 1990 Census. Interviews with government officials and local environmental activists provided additional perspective on local history and environmental perception within the community.

### **Results of Neighborhood Analysis**

The seeds of the current residential and industrial landscape of Los Angeles were planted long ago. Most of the industrial zones within the County were well defined by 1930, though much industrial growth occurred after 1940. Population growth, and residential expansion have been fundamental to the history of Los Angeles. Although much of Los Angeles was developed by 1940, most residential construction occurred after 1955 (Census, 1990.)

Most TRI facilities within the study areas opened before 1952, prior to widespread environmental awareness and the rise of the NIMBY phenomenon. TRI facilities do not present the same hazard as toxic waste dumps. They are industrial facilities, which often offer jobs for the surrounding communities, while only presenting an indefinite and largely unknown risk. It appears that there has been little need for targeting of communities by TRI facilities because these facilities tend to be located in long-term industrial areas, and because there is not much community resistance to these facilities. To a great extent, historic zoning has resulted in the landscape which is apparent today.

Industrial facilities are businesses focused on making a profit. Most of the facilities have been in place for a long time, and chose current locations for reasons of access, land availability, low land prices, and because of the industrial nature of the area. Industries need to be in industrial areas for reasons of zoning; pollution; convenience; and to fit in. In this sense, it is a path of least resistance - industries can most easily locate in industrial areas. Many of those industries have had to change processes and implement control technologies over time in order to comply with new regulations.

Prior to 1940, all three of the study areas had both residential and industrial land uses. Housing typically densified in these areas during the 1950's and 1960's, at the same time that industry expanded within industrial zones. The growth of both industry and residential use consumed open space and reduced the buffering between these conflicting land uses. This development largely followed the zoning and planning which had been established decades earlier.

## Chapter 6. Research Summary and Conclusions

Los Angeles is a city of contrasts. It is a diverse, multi-cultural, industrial city. It is a city of large open spaces and world renowned congestion. It is a city of year-round leisure pursuit; a place for agriculture, industry, and the premier city of the automobile. Though a city of tremendous wealth and extravagance, its center is one of extreme poverty, frustration and neglect.

The history of ethnic migration to this area is long and colorful, resulting in a city of paramount ethnic diversity. But, segregation in the city is historic and complex. Many parts of the city are extremely integrated, while South Central remains an area dominated by a Black and Hispanic underclass. The expanse surrounding East Los Angeles has become the largest concentration of Hispanics north of Mexico City.

Los Angeles is a far more industrial city than many realize. It is a major center for oil refineries, the petro-chemical industry, metal finishing and plating, textiles and aerospace. There are over 700 manufacturing facilities in Los Angeles which are required by law to report their routine processing, emission and disposal of toxic chemicals for inclusion in EPA's TRI database. These facilities range from seemingly benign industries, such as fruit canning and ice cream manufacture, to oil refineries and metal foundries. One-hundred-thirty-nine different TRI-listed toxic chemicals are emitted by TRI facilities in Los Angeles County. Many facilities are in the TRI due to the emission of a single chemical, while some oil refineries emit as many as thirty. TRI facilities have very different types and rates of emission, employ different control technologies and pose widely different risks to the surrounding community.

The analysis presented in this report applies only to TRI facilities, and only to Los Angeles, and is not generalizable beyond this case study. Many other environmental hazards exist within Los Angeles, which were not included in this analysis. Environmental hazards such as landfills and conventional air pollutants have not been considered. By volume, toxic emissions from TRI facilities pale in comparison to emissions from automobiles in Los Angeles.

The statistical analyses described here have confirmed that there is a significant relationship between income, minority percentage and TRI facility occurrence in a census tract. In general, the poorer an area, and the higher the minority percentage of the population, the greater the number of TRI facilities which occur in that area. When all census tracts were included in the analysis, the minority percentage of the tract appeared to have a slightly stronger relationship with TRI facility occurrence than does the median per capita income of the tract. This relationship was found to be very sensitive to the minimum population density threshold used, suggesting that the presence of minority individuals in low density, highly industrial tracts adds strength to the racial variable relative to the income variable. Overall, it is not possible to conclude whether income or race is more important in this relationship with TRI facility occurrence. Both are significant variables.

The statistical analyses assessed static conditions in 1989. The more detailed study of three neighborhoods added a temporal dimension to the analysis. Zoning has had a major influence on the socio-environmental landscape which exists today. Many sections of greater Los Angeles had industrially-zoned areas adjacent to residentially-zoned areas in the 1920's. By the 1940's this blueprint or scheme for control of space and form had come to fruition.

Industries, acting as profit seeking enterprises, make largely rational decisions. They operate to minimize cost, while maximizing revenue. The objectives include acquiring land at low cost; acquiring facilities well suited to intended use; having good, efficient access to transportation, customers and suppliers; locating in a safe area where workers with appropriate skills are available; and locating in an area where the facility can operate legally and at minimum cost. Of course, zoning can be changed to accommodate new uses, but it is easier to locate in an area with compatible zoning from the outset. Also, environmental regulations can vary from locality to locality, but this does not appear to have been an important factor for TRI facilities in Los Angeles County.

Given that most TRI facilities within the study areas began operation at current locations before 1952, and few have been sited within the last twenty years, industrial siting decisions do not appear to be a major factor controlling the socio-environmental landscape. Most of the current housing stock was built since 1955. A large influx of foreign-born persons beginning in the 1970's has significantly elevated the demand for inexpensive housing. Residential and, to some extent, industrial expansion have increased the proximity between housing and industry.

Given the more pressing motivations guiding industrial decision making, facility siting does not appear to have resulted in the high number of minority individuals living in close proximity to industrial facilities. The current socio-environmental landscape, therefore, must be a product of residential choice, or perhaps lack of choice.

Residential segregation is apparent in many parts of Los Angeles. The urban core of South Central and East Los Angeles is dominated by Blacks, and increasingly, by Hispanics, while the outlying areas are dominated by middle and upper-class Whites. This segregation results, in part, from the effect of aggregation, particularly for foreign speaking residents. A Spanish speaking population must reach some size in order to support many desirable services, such as Spanish-speaking banks, shops and theaters. East Los

Angeles offers the appeal of these services for newly arriving Spanish-speakers. Such effects alone, however, do not account for the degree of segregation present in Los Angeles.

The statistical analysis showed that at a given income level, Hispanic and Black individuals were more likely to be living in close proximity to a TRI facility than White or Asian individuals. This result could be attributed to two factors - perception and value of the environment or limited residential choice. If individuals' perception of the environment differs across racial and ethnic groups, perhaps as a result of culture, experience, differing degrees of environmental awareness, etc., this could result in a differing valuing of proximity to industry. Hence, people might be willing to pay differently for not living near industries which pose an unknown environmental risk. Another explanation of the model results lies in residential segregation. If differing residential options are available to people of the same economic class, but of different race or ethnicity, this could account for the importance of race as a factor within the model.

### **Policy Issues**

The bivariate mapping employed in this analysis confirmed the strong correlation between income and race. The greater the Minority Percentage in a tract, the poorer the tract tends to be, with very few tracts diverging from this pattern. Due to the high correlation between income and race, which of the two factors is more important is nearly a moot point. To examine low-income neighborhoods is to examine minority-dominated neighborhoods, and the inverse is true. And both hold true most of the time.

This finding has relevance to potential policies for addressing the high minority proximity to TRI facilities. If one wanted to address the issue of a high percentage of minority and low-income individuals living near TRI facilities, one would not need to focus on ethnic composition or income-level. Rather, if one simply focuses on areas with high concentrations of TRI facilities (or the most polluted areas, in general) one would by default be examining the areas with, on average, a high minority percentage and low income. Hence, a risk-based approach, which focuses on areas of high pollution and the greatest risk to the population, which treats people equally, would generally identify areas of low income and high minority percentage. Such a strategy can only be effective if environmental protection and priority setting are kept separate from political interests.

The solutions to the problems of environmental equity are closely related to broader problems in American society today. Firstly, the strong relationship between race and poverty is at the root of the high minority exposure to environmental pollution, as poverty limits choice. For it ever to be the case that the exposure of minority individuals to pollution is no higher than that of the average resident, the economic disparity between minorities and non-Hispanic Whites must be rectified.

Secondly, if residential segregation has resulted in the importance of race as a factor beyond what would be expected from income-level alone, then, once again the issue of environmental equity is closely tied to broader social problems. Residential segregation must be remedied.

There are some policies and actions available, however, which are more closely related to environmental policy, which could help to lessen environmental inequities. Insuring that environmental regulations are uniformly enforced, regardless of the race, ethnicity or class of the affected population, would go a long way toward addressing the equity issue. Small industrial businesses offer a great benefit to the surrounding community, particularly in low income areas. It is crucial that these facilities employ all required and appropriate pollution control methods, to minimize risk to the surrounding communities.

Within regional planning agencies, greater attention needs to be paid to environmental concerns and the co-location of conflicting land-uses. Particularly in the cases of most limited choice, such as low-income and public housing projects, it seems inappropriate, at best, to site these in areas of high environmental risk. Although it did not appear to be the case with TRI facilities, more noxious facilities such as landfills, incinerators and pipelines are often proposed for the most disadvantaged areas. People should be treated equally within the environmental landscape, and such targeting should not persist. To this end, environmental reports and planning documents should be published and distributed in the languages which are appropriate for all affected neighborhoods.

And, finally, the quality of environmental education is poor in general, and even worse in inner-city schools. Environmental education should be improved in both primary and secondary schools as a means of improving environmental awareness and of improving people's ability to address environmental concerns within their communities.

### **Further Research**

This research examined the relationship between the population and TRI facility occurrence within census tracts. The analysis could be expanded in three ways:

- 1) Within this analysis, all TRI facilities were treated as equivalent. The analysis could be enhanced by evaluating the relative risk posed by individual facilities in terms of type and volume (or overall toxicity) of emissions.
- 2) Additionally, other sources of pollution could be included in order to get a broader picture of environmental equity in Los Angeles. Sources such as incinerators, municipal and hazardous waste landfills, and Superfund sites could be included, in addition to information on concentrations of conventional pollutants.
- 3) Finally, a similar, frame-independent analysis could be performed which is less constrained by census tract boundaries. Such an analysis could be implemented by buffering facilities, and then analyzing the characteristics of the "affected population" within the buffer. Alternatively, a population potential function could be used which would allow for a distance-based evaluation of exposure.

Such enhancements would be interesting extensions of this thesis.

## References

- Adams, John, 1992. "The Mainstream Environmental Movement", EPA Journal, v18, n1: 25-27.
- Anselin, Luc, 1989. "What is special about spatial data: alternative perspectives on spatial data analysis", National Center for Geographic Information and Analysis (NCGIA) technical paper (89-4), Santa Barbara, CA.
- Anselin, Luc, 1990. "Some robust approaches to testing and estimation in spatial econometrics", Regional Science and Urban Economics, v4, n2:141-163.
- Anselin, Luc, 1992. SPACESTAT, software developed by Luc Anselin, National Center for Geographic Information and Analysis (NCGIA) software (92-1), Santa Barbara, CA.
- Asabere, Paul K. & F. E. Huffman, 1991. "Zoning and industrial land values: the case of Philadelphia", AREUEA Journal, v19, n2:154-160.
- Asch, Peter & J. Seneca, 1978. "Some evidence on the Distribution of Air Quality," Land Economics, 54(3): 278-297
- Barber, Gerald M., 1988. Elementary statistics for geographers, Guilford Press, New York.
- Brewster Maps, 1963. Property zoning atlas, Brewster Maps, Los Angeles.
- Bullard, Robert D., 1990. Dumping in Dixie: race, class, and environmental quality, Westview, Press, Boulder, CO.
- Burch, William , R, 1976. "The Peregrin Falcon and the Urban Poor: Some Sociological Interrelations", (pgs. 308-316) in Human Ecology, An Environmental Approach, P. Richardson and J. McEvoy (eds.), Duxbury Press, Belmont, CA.
- Chambers, John M. & T. J. Hastie (eds.), 1992. Statistical models in S, Wadsworth & Brooks / Cole Advanced Books, Pacific Grove, CA.
- Council on Environmental Quality (CEQ), 1971. The Second Annual Report of the Council on Environmental Quality, U.S. Government Printing Office, Washington, DC.
- Coyle, Marcia, 1992. "When Movements Coalesce", The National Law Journal, v15, n3:SIO.
- Davis, Mike, 1990. City of quartz: excavating the future in Los Angeles, Verso, London and New York.
- Dear, Michael J., 1991. Gaining community acceptance, Robert Wood Johnson Foundation, Princeton, NJ.
- Dear, Michael J., 1992. "Understanding and overcoming the NIMBY syndrome", Journal of the American Planning Association, v58, n3:288-299.
- Diaconis, Persi, B. Efrain, 1983. "Computer Intensive Methods in Statistics", Scientific American, v248, n5:116-130.

- Eggers, Mitchell L. & D. S. Massey, 1991. "The Structural Determinants of Urban Poverty: a comparison of Whites, Blacks, and Hispanics", *Social Science Research*, v20, n3:217-255.
- Eggers, Mitchell L. & D. S. Massey, 1992. "A Longitudinal Analysis of Urban Poverty: Blacks in U.S. Metropolitan Areas between 1970 and 1980", *Social Science Research*, v21, n2:175-203.
- Exter, Thomas, 1990. "The Survey that Measures the Census", *American Demographics*, November, 1990.
- Gelobter, Michael, 1992. "Toward a Model of Environmental Discrimination", in Bryant & Mohai, (eds.), *Race and the Incidence of Environmental Hazards: A time for Discourse*, Westview Press, Boulder, CO
- Gianessi, Leonard, H. Peskin, E. Wolff, 1979. "The Distributional Effects of Uniform Air Pollution Policy in the U.S.," *Quarterly Journal of Economics*, (May): 281-301.
- Godsil, Rachel D., 1991. "Remedying environmental racism", *Michigan Law Review*, v90, n2:394-427.
- Goldman, Benjamin A., 1991. *The truth about where you live: an atlas on toxins and mortality*, Times Books / Random House, New York.
- Goodchild, Michael, 1986. *Spatial Autocorrelation*, CATMOG series: 47, Norwich.
- Goodchild, Michael & S. Gopal (eds.), 1989. *The Accuracy of Spatial Databases*, Taylor & Francis, London & New York.
- Gould, Peter, 1970. "Is statistix inferens the geographical name for a wild goose?", *Economic Geography*, v.46, 439:448.
- Gould, Peter & R. White, 1986. *Mental Maps*, Allen & Unwin, Boston.
- Haining, Robert, P., 1990. *Spatial Data Analysis in the Social and Environmental Sciences* Cambridge University Press, New York.
- Hogan, Howard, 1992. "New Estimates from the 1990 Post-enumeration Survey", paper presented to the 1992 Annual Meeting of the Population Association of America, Denver, CO.
- Johnson, R.J., 1984. "Quantitative ecological analysis in human geography: an evaluation of four problem areas", in *Recent Developments in Spatial Data Analysis*, G. Bahrenberg, M. Fischer & P. Nijkamp (eds.), Aldershot, Gower.
- Kay, Jane, 1991. "Fighting Toxic Racism", in *San Francisco Examiner*, April 7, 1991 po's. A1, A 12.
- Kennedy, Peter, 1985. *A Guide to Econometrics*, MIT Press, Cambridge, MA.
- Kruvant, William J., 1975. "People, Energy and Pollution", (pgs. 125-167) in *The American Energy Consume*, D.K. Newman and D. Day (eds.), Ballinger, Cambridge, Mass.
- Lavelle, Marianne, & M. Coyle, M, 1992. "Unequal Protection - the Racial Divide in Environmental Law" *The National Law Journal*, v15, n3:S2-S4.
- Lavelle, Marianne, 1992. "Residents Want Justice, the EPA Offers Equity", *The National Law Journal*, v15, n3:SI2.
- Logan, John R. & H. L. Molotch, 1987. *Urban fortunes: the political economy of place*, University of California Press, Berkeley, CA.
- Long, D. S., S.D. DeGloria, J.M. Galbraith, 1991. "Use of Global Positioning System in soil survey", in *Journal of Soil and Water Conservation*, v46, n4:293-297.
- Los Angeles City Council, 1925. *Los Angeles municipal atlas: official -Zoning- maps of the city of Los Angeles*, as authorized by the City Council, Los Angeles.
- Los Angeles County, Regional Planning Commission, 1973. *General Plan of Los Angeles County*, Los Angeles County Regional Planning Commission, Los Angeles.
- MacLachlam, Claudia, 1992. "Tension Underlies Rapport with Grassroots Groups", *The National Law Journal*, v. 15,n3:S 10.

- Maddala, G.S., 1983. Limited-dependent and qualitative variables in econometrics, Cambridge University Press, New York.
- Mann, Eric & the WATCHDOG, 1991. L.A.'s lethal air: new strategies for policy, organizing, and action, Labor / Community Strateg Center, Los Angeles.
- Massey, Douglas S. & N. A. Denton, 1992. "Racial Identity and the spatial assimilation of Mexicans in the United States", Social Science Research, v21, n3:235-260.
- Mohai, Paul and B. Bryant, 1992a. "Environmental Racism: Reviewing the Evidence", paper delivered at the University of Michigan Law School Symposium on Race, Poverty and the Environment, January 23, 1992. (Also forthcoming in Bryant & Mohai, Eds., Race and the Incidence of Environmental Hazards: A time for Discourse, Westview Press, Boulder, CO).
- Mohai, Paul and B. Bryant, 1992b. "The Michigan Conference: A turning Point", EPA Journal, v18, nl:9:10.
- Mohai, Paul and B. Bryant, 1992c. "Race, Poverty and the Environment", EPA Journal, v18, nl:6-8.
- Montgomery, Douglas C. and Elizabeth Peck, 1982. Introduction to linear regression analysis, Wiley, New York.
- Morell, Robert, 1987. "Siting and the Politics of Equity" in Resolving Locational Conflict, Robert Lake, ed., Center for Urban Policy Research, New Brunswick, NJ.
- Nelson, Howard J. & W. A. Clark, 1976. Los Angeles - the metropolitan experience, Ballinger, Cambridge, MA.
- Novick, Michael, 1992. "The Watchdog", EPA Journal, v18, nl:51-52.
- Openshaw, Stan, 1984a. "Ecological fallacies and the analysis of areal census data", Environmental and Planning- A, v. 16:17-3 1.
- Openshaw, Stan, 1984b. The modifiable areal unit problem. CATMOG series: 38, Norwich: Geo Books.
- Pindyck, Robert S. and D. L. Rubinfeld, 1991. Econometric Models and Economic Forecasts, McGraw-Hill, New York.
- Reilly, William K., 1992. "Environmental Equity: EPA's Position", EPA Journal, v 18, n 1: 18--22.
- Robinson, W.S., 1950. "Ecological correlations and the behavior of individuals", American Sociological Review, v.15: 351-357.
- Satchell, Michael, 1992. "A whiff of discrimination?", U.S. News & World Report, v 112, n 17: 34-35.
- Smith, James, N. (ed.), 1974. Environmental Quality and Social Justice in Urban America, Conservation Foundation, Washington, DC.
- Suro, Robert, 1993. "Pollution-weary Minorities Try Civil Rights Tack", New York Times, Jan. 11, 1993 A I & B7.
- Taylor, Dorceta, 1992. "The Environmental Justice Movement", EPA Journal, v18, nl:23-25.
- Tierra Concepts, 1993. Regional economic analysis and business climate assessment -Huntington Park, California, Tierra Concepts. Los Angeles.
- Turner, Eugene & J. P. Allen, 1991. A human mosaic: an atlas of ethnicity in Los Angeles County, 1980-1986, California State University, Northridge.
- United Church of Christ (UCC), 1987. Toxic waste and race in the United States: A national report on the racial and-socio-economic characteristics of communities with hazardous waste sites. Commission for Racial Justice, United Church of Christ, New York, NY.
- United States Environmental Protection Agency (EPA), 1989. Toxic Chemical Release Inventory, Risk Screening Guide, Vol. L, Washington, DC.
- United States Environmental Protection Agency (EPA/OTS), 1991. TRI Location Data Quality Assurance for Geographic Information Systems, U.S. EPA, Office of Toxic Substances, Washington, DC.

United States Environmental Protection Agency (EPA), 1992. Environmental Equity - Reducing Risk for All Communities, Washington, DC.

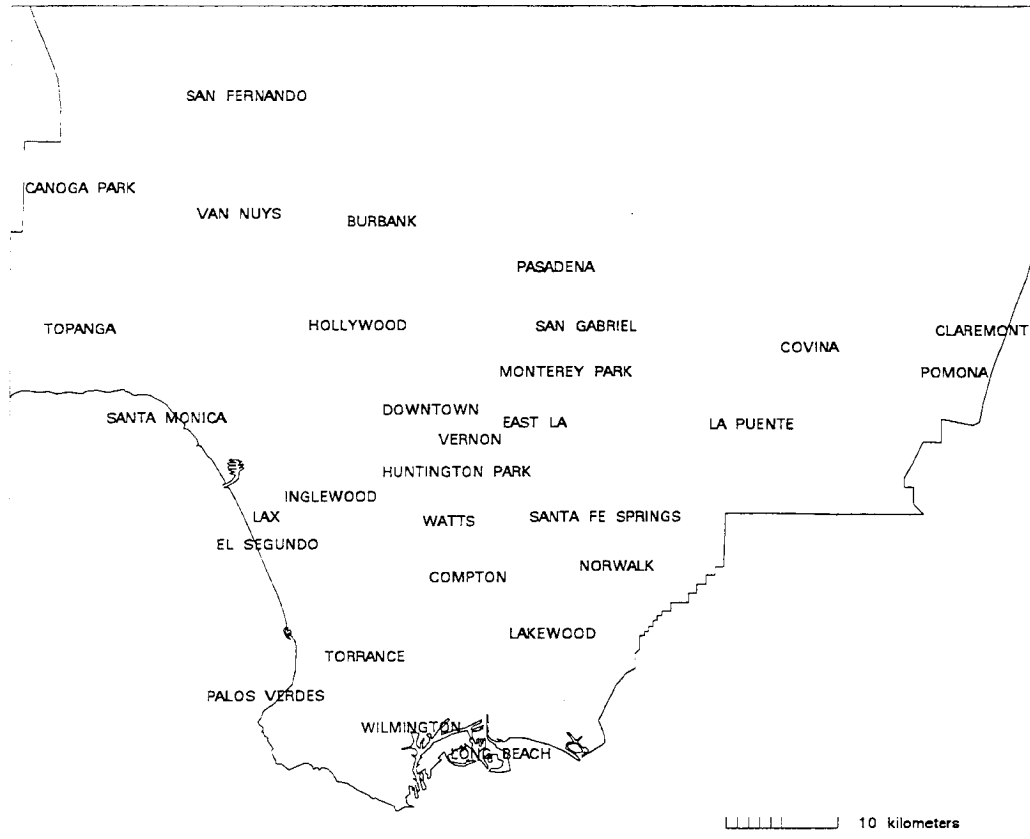
United States General Accounting Office (GAO), 1983. Siting of hazardous waste landfills and their correlation with racial and economic status of surrounding communities. U.S. GAO, Washington, DC.

Wilson, Albert, R., 1991. Environmental Risk: Identification and Management, Lewis Publishers, Chelsea, MI.

Wrigley, Neil, 1985. Categorical Data Analysis for geographers and environmental scientists. Longman, London & New York.

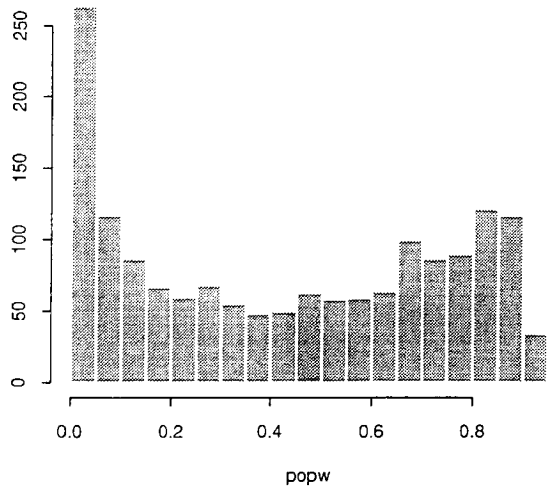


# Appendix A - Map of Places in Los Angeles

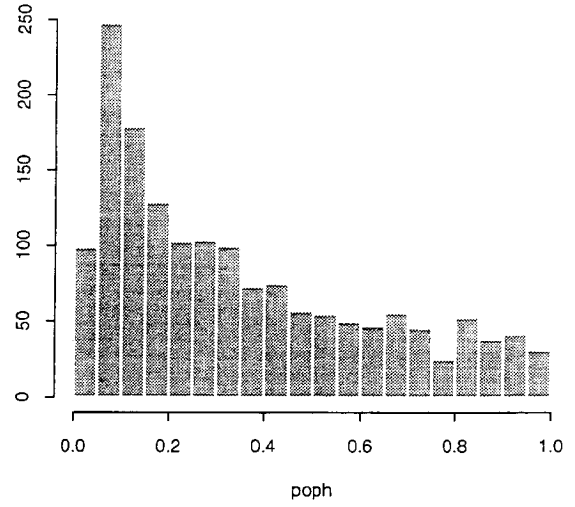


# Appendix B - Distribution of Main Variables

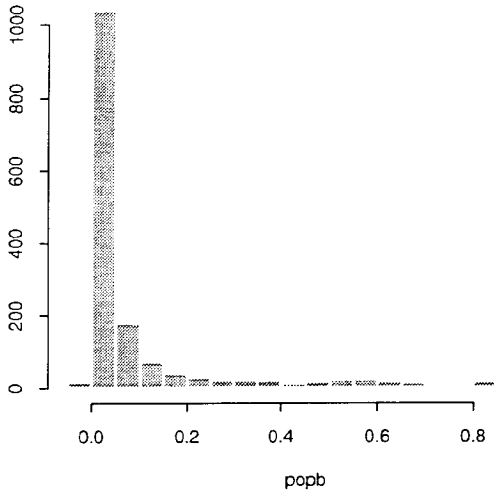
## Percentage White



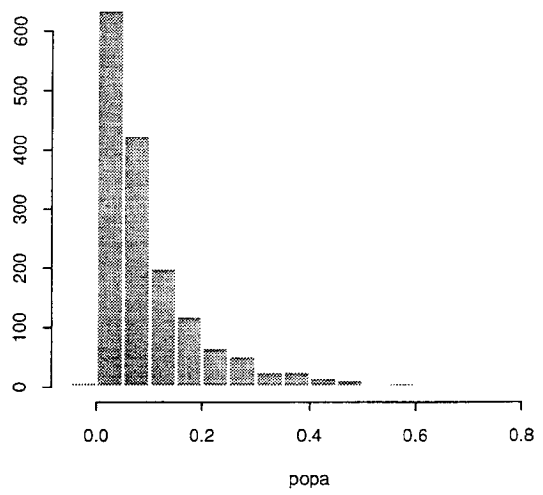
## Percentage Hispanic



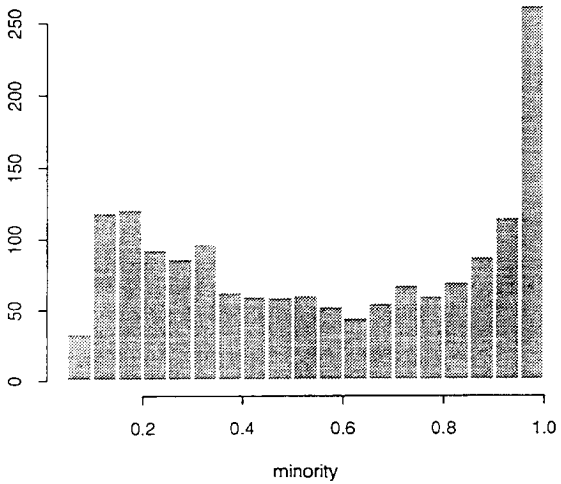
## Percentage Black



## Percentage Asian

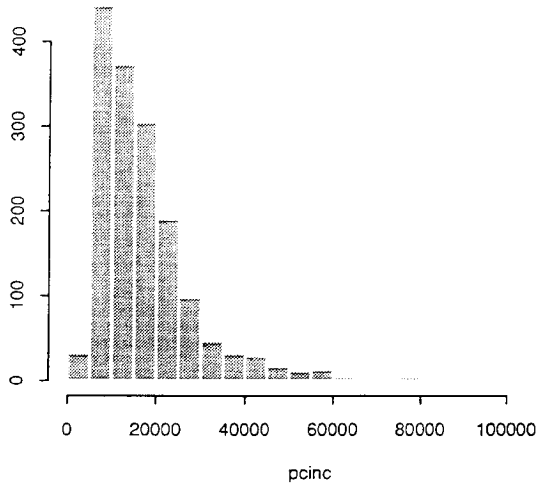


## Minority Percentage

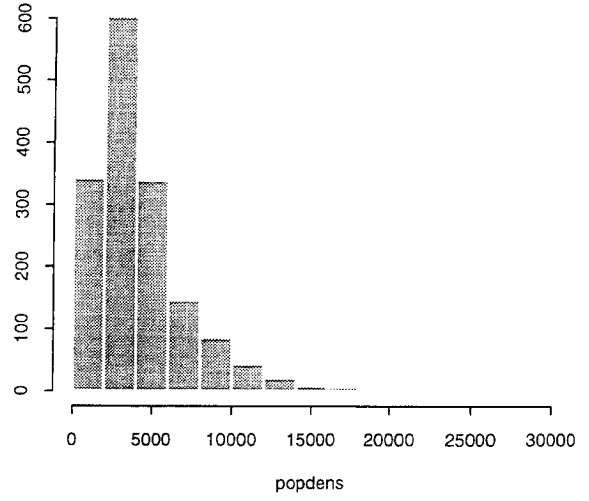


Distribution of Main Variables

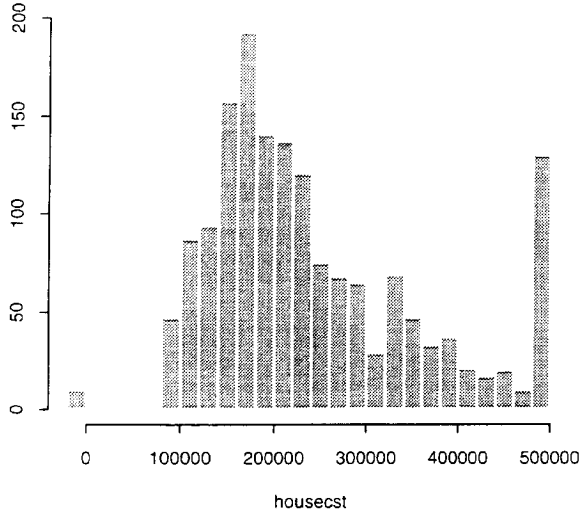
Per Capita Income (\$)



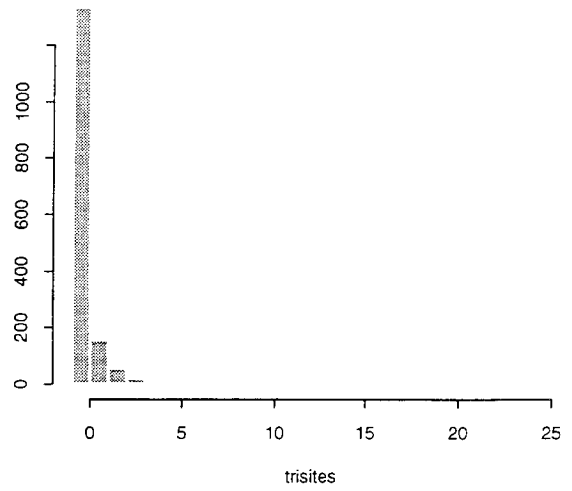
Population Density (pp/sq.km.)



Housing Value (\$)



TRI Facilities



## Appendix C - Toxic Release Inventory (TRI) Industrial Information:

### 1) Standard Industrial Categories (SIC) and Frequency of Occurrence within Los Angeles County

| SIC  | FRQ | INDUSTRY                             |
|------|-----|--------------------------------------|
| 2000 | 1   | FOOD AND KINDRED PRODUCTS            |
| 2023 | 1   | CONDENSED AND EVAPORATED MILK        |
| 2024 | 2   | ICE CREAM AND FROZEN DESSERTS        |
| 2026 | 1   | FLUID MILK                           |
| 2033 | 1   | CANNED FRUITS AND VEGETABLES         |
| 2035 | 2   | PICKLES, SAUCES, AND SALAD DRESSINGS |
| 2037 | 1   | FROZEN FRUITS AND VEGETABLES         |
| 2041 | 2   | FLOUR AND OTHER GRAIN MILL PRODUCTS  |
| 2047 | 2   | DOG, CAT, AND OTHER PET FOOD         |
| 2075 | 1   | SOYBEAN OIL MILLS                    |
| 2077 | 2   | ANIMAL AND MARINE FATS AND OILS      |
| 2082 | 3   | MALT BEVERAGES                       |
| 2086 | 4   | BOTTLED AND CANNED SOFT DRINKS       |
| 2087 | 1   | FLAVORING EXTRACTS AND SYRUPS, NEC   |
| 2089 | 1   | FOOD AND KINDRED PRODUCTS            |
| 2091 | 1   | CANNED AND CURED SEAFOODS            |
| 2257 | 2   | CIRCULAR KNIT FABRIC MILLS           |
| 2273 | 1   | CARPETS AND RUGS                     |
| 2295 | 3   | COATED FABRICS, NOT RUBBERIZED       |
| 2299 | 2   | TEXTILE GOODS, NEC                   |
| 2431 | 1   | MILLWORK                             |
| 2434 | 2   | WOOD KITCHEN CABINETS                |
| 2490 | 1   | MISCELLANEOUS WOOD PRODUCTS          |
| 2491 | 1   | WOOD PRESERVING                      |
| 2499 | 2   | WOOD PRODUCTS, NEC                   |
| 2511 | 5   | WOOD HOUSEHOLD FURNITURE             |
| 2512 | 2   | UPHOLSTERED HOUSEHOLD FURNITURE      |
| 2514 | 1   | METAL HOUSEHOLD FURNITURE            |
| 2517 | 1   | WOOD TV AND RADIO CABINETS           |
| 2521 | 5   | WOOD OFFICE FURNITURE                |
| 2522 | 5   | METAL OFFICE FURNITURE               |
| 2599 | 3   | FURNITURE AND FIXTURES, NEC          |
| 2621 | 2   | PAPER MILLS                          |
| 2631 | 1   | PAPERBOARD MILLS                     |
| 2657 | 3   | FOLDING PAPERBOARD BOXES             |
| 2671 | 1   | PACKAGING PAPER                      |
| 2672 | 4   | COATED PAPER                         |
| 2673 | 2   | PLASTIC, FOIL AND PAPER BAGS         |
| 2674 | 1   | UNCOATED PAPER AND MULTI-BAGS        |
| 2677 | 2   | ENVELOPES                            |
| 2700 | 1   | PRINTING AND PUBLISHING              |
| 2752 | 5   | COMMERCIAL PRINTING, LITHOGRAPHIC    |
| 2759 | 1   | COMMERCIAL PRINTING, NEC             |
| 2782 | 3   | BLANKBOOKS AND LOOSELEAF BINDERS     |
| 2796 | 1   | PLATEMAKING                          |

2812 2 ALKALIES AND CHLORINE  
2813 7 INDUSTRIAL GASES  
2816 2 INORGANIC PIGMENTS  
2819 19 INDUSTRIAL INORGANIC CHEMICALS, NEC  
2821 20 PLASTICS MATERIALS AND RESINS  
2822 2 SYNTHETIC RUBBER  
2834 5 PHARMACEUTICAL PREPARATIONS  
2836 1 BIOLOGICAL PRODUCTS  
2841 16 SOAP AND OTHER DETERGENTS  
2842 9 POLISHES AND SANITATION GOODS  
2843 5 SURFACE ACTIVE AGENTS  
2844 4 TOILET PREPARATIONS  
2851 34 PAINTS, VARNISH, LACQUERS  
2865 1 CYCLIC CRUDES AND INTERMEDIATES  
2869 7 INDUSTRIAL ORGANIC CHEMICALS, NEC  
2875 1 FERTILIZERS, MIXING ONLY  
2879 3 AGRICULTURAL CHEMICALS, NEC  
2891 19 ADHESIVES AND SEALANTS  
2893 5 PRINTING INK  
2899 14 CHEMICAL PREPARATIONS, NEC  
2911 16 PETROLEUM REFINING  
2952 2 ASPHALT FELTS AND COATINGS  
2992 10 LUBRICATING OILS AND GREASES  
2999 2 PETROLEUM AND COAL PRODUCTS, NEC  
3052 2 RUBBER AND PLASTIC HOSES  
3053 2 GASKETS  
3060 2 FABRICATED RUBBER PRODUCTS, NEC  
3061 2 PLASTICS, FILM AND SHEET  
3069 2 PLASTICS PRODUCTS, NEC  
3079 4 RUBBER OR PLASTIC PRODUCTS  
3080 4 RUBBER OR PLASTIC PRODUCTS  
3081 1 RUBBER OR PLASTIC PRODUCTS  
3082 1 RUBBER OR PLASTIC PRODUCTS  
3083 2 RUBBER OR PLASTIC PRODUCTS  
3086 5 RUBBER OR PLASTIC PRODUCTS  
3087 4 RUBBER OR PLASTIC PRODUCTS  
3089 11 RUBBER OR PLASTIC PRODUCTS  
3221 2 GLASS CONTAINERS  
3231 1 GLASS PRODUCTS  
3272 1 CONCRETE PRODUCTS, NEC  
3275 2 GYPSUM PRODUCTS  
3315 3 STEEL WIRE AND RELATED PRODUCTS  
3316 1 COLD FINISHING OF STEEL SHAPES  
3317 3 STEEL PIPE AND TUBES  
3321 1 GRAY IRON FOUNDRIES  
3324 6 STEEL INVESTMENT FOUNDRIES  
3325 1 STEEL FOUNDRIES, NEC  
3341 9 SECONDARY SMELTING  
3351 1 COPPER ROLLING AND DRAWING  
3353 1 ALUMINUM SHEET, PLATE, AND FOIL  
3354 6 ALUMINUM EXTRUDED PRODUCTS

3357 1 NONFERROUS WIRE DRAWING INSULATING  
3362 2 BRASS, BRONZE, AND COPPER FOUNDRIES  
3363 3 ALUMINUM DIE-CASTINGS  
3364 1 NONFERROUS DIE-CASTINGS  
3369 10 NONFERROUS FOUNDRIES, NEC  
3398 9 METAL HEAT TREATING  
3399 1 PRIMARY METAL PRODUCTS, NEC  
3411 11 METAL CANS  
3412 1 METAL BARRELS, DRUMS, AND PAILS  
3417 1 FABRICATED METAL PRODUCTS  
3423 2 HAND AND EDGE TOOLS, NEC  
3425 1 HAND SAWS AND SAW BLADES  
3429 3 HARDWARE, NEC  
3431 2 METAL SANITARY WARE  
3432 5 PLUMBING FITTINGS AND BRASS GOODS  
3441 1 FABRICATED STRUCTURAL METAL  
3442 1 METAL DOORS, SASH, AND TRIM  
3443 2 FABRICATED PLATE WORK %BOILER SHOPS<  
3446 1 ARCHITECTURAL METAL WORK  
3449 1 MISCELLANEOUS METAL WORK  
3451 1 SCREW MACHINE PRODUCTS  
3452 7 BOLTS, NUTS, RIVETS, AND WASHERS  
3462 1 IRON AND STEEL FORGINGS  
3469 2 METAL STAMPINGS, NEC  
3470 2 METAL SERVICES, NEC  
3471 78 PLATING AND POLISHING  
3478 1 FABRICATED METAL PRODUCTS  
3479 17 METAL COATING AND ALLIED SERVICES  
3489 1 ORDNANCE AND ACCESSORIES, NEC  
3491 2 INDUSTRIAL VALVES  
3494 2 VALVES AND PIPE FITTINGS  
3496 5 MISC. FABRICATED WIRE PRODUCTS  
3498 1 FABRICATED PIPE AND FITTINGS  
3499 11 FABRICATED METAL PRODUCTS, NEC  
3531 1 CONSTRUCTION MACHINERY  
3533 1 OIL FIELD MACHINERY  
3546 1 POWER DRIVEN HAND TOOLS  
3561 2 PUMPS AND PUMPING EQUIPMENT  
3562 1 BALL AND ROLLER BEARINGS  
3573 1 ELECTRONIC COMPUTING EQUIPMENT  
3585 3 REFRIGERATION AND HEATING EQUIPMENT  
3589 3 SERVICE INDUSTRY MACHINERY, NEC  
3600 1 ELECTRIC AND ELECTRONIC EQUIPMENT  
3621 1 MOTORS AND GENERATORS  
3624 1 CARBON AND GRAPHITE PRODUCTS  
3643 2 CURRENT-CARRYING WIRING DEVICES  
3644 1 NONCURRENT-CARRYING WIRING DEVICES  
3645 5 RESIDENTIAL LIGHTING FIXTURES  
3646 1 COMMERCIAL LIGHTING FIXTURES  
3662 1 RADIO AND TV COMMUNICATION EQUIPMENT  
3663 1 RADIO AND T.V BROADCAST EQUIPMENT

3669 2 COMMUNICATIONS EQUIPMENT, NEC  
3671 1 ELECTRON TUBES, RECEIVING TYPE  
3672 7 CATHODE RAY TELEVISION PICTURE TUBES  
3674 4 SEMICONDUCTORS AND RELATED DEVICES  
3675 4 ELECTRONIC CAPACITORS  
3678 1 ELECTRONIC CONNECTORS  
3679 13 ELECTRONIC COMPONENTS, NEC  
3691 7 STORAGE BATTERIES  
3699 1 ELECTRICAL EQUIPMENT SUPPLIES, NEC  
3700 1 TRANSPORTATION EQUIPMENT  
3711 2 MOTOR VEHICLES AND CAR BODIES  
3713 1 TRUCK AND BUS BODIES  
3714 11 MOTOR VEHICLE PARTS AND ACCESSORIES  
3715 1 TRUCK TRAILERS  
3721 10 AIRCRAFT  
3722 1 AIRCRAFT PARTS  
3724 2 AIRCRAFT ENGINES AND ENGINE PARTS  
3728 23 AIRCRAFT EQUIPMENT, NEC  
3741 1 TRANSPORTATION EQUIPMENT  
3761 2 GUIDED MISSILES AND SPACE VEHICLES  
3764 2 SPACE PROPULSION UNITS AND PARTS  
3769 1 SPACE VEHICLE EQUIPMENT, NEC  
3792 1 TRAVEL TRAILERS AND CAMPERS  
3800 1 INSTRUMENTS AND RELATED PRODUCTS  
3812 3 NAVIGATION INSTRUMENTS  
3822 1 ENVIRONMENTAL CONTROLS  
3823 1 PROCESS CONTROL INSTRUMENTS  
3827 1 OPTICAL INSTRUMENTS AND LENSES  
3830 3 OPTICAL INSTRUMENTS AND LENSES  
3841 1 SURGICAL AND MEDICAL INSTRUMENTS  
3843 1 DENTAL EQUIPMENT AND SUPPLIES  
3851 1 OPHTHALMIC GOODS  
3861 4 PHOTOGRAPHIC EQUIPMENT AND SUPPLIES  
3944 1 GAMES, TOYS, AND CHILDREN  
3949 2 SPORTING AND ATHLETIC GOODS, NEC  
3951 1 PENS AND MECHANICAL PENCILS  
3996 1 HARD SURFACE FLOOR COVERINGS  
3999 8 MANUFACTURING INDUSTRIES, NEC

## **Appendix D - Industrial Neighborhood Questionnaire**

I am a graduate student at UCSB doing research on decisionmaking surrounding the siting, of industrial facilities within Los Angeles. I am looking at several hundred facilities within L.A., and am trying to gather some information about why different businesses chose their current locations.

I have a few questions for the site manager, or some other person knowledgeable of the history of the facility. The discussion should only take about 5 minutes.

- 1) When did the facility begin operation at the current location? (year).
- 2) Did your company open here or did you move to this location?
- 3) What is the nature of your business?
- 4) What is the nature of the neighborhood surrounding your facility? (ie. industrial, commercial, residential, mixed.)
- 5) What is the name of the neighborhood / area? Is it an incorporated area?
- 6) Is the area "run down"?
- 7) What factors influenced the decision to site the facility in this area?
- 8) How many employees are there at this site?
- 9) Does the facility employ many people from the surrounding community, or do most people commute from more than 5 miles?
- 10) Is there much pollution associated with your industrial process? (either toxic or conventional).

If so, what pollution control technologies are being employed to reduce the pollution?