A QUANTITATIVE ANALYSIS OF ENVIRONMENTAL INEQUALITY IN HOUSTON, TEXAS

by

Maggie Chamberlain, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Arts
with a Major in Sociology
August 2017

Committee Members:

Chad L. Smith, Chair

Matthew Clement

Susan Day

Craig Hanks

DEDICATION

To the members of O'Brien's Orchestra (Every. Single. One): The last 7 years have been a strange journey, and I've never regretted it…not once. Through this organization and with the love and support of its members I have experienced a tremendous amount of personal growth. Thank you for letting me learn to lead, laugh harder than I ever thought I could, and most importantly, for reminding me that I can "Be It".

To my Family, Mr. Isidoro Leon, Mrs. Julia Leon, Ms. Gavina Hobbs, Mrs. Julie Abraham, Mr. Wes Corwin and Ms. Magdalena Chamberlain: There is no way I can truly acknowledge the impact you all have had on me here. I love you. Thank you for always being with me.

ACKNOWLEDGEMENTS

To the members of my committee, Dr. Craig Hanks and Dr. Matthew Clement: Thank you for your assistance and feedback throughout this this project. I have immensely enjoyed having you both on my thesis committee and am confident the document and my future research endeavors have been positively impacted by your feedback and guidance.

To my mentor and committee member, Dr. Susan Day: Your mentorship and guidance throughout my time at Texas State University have had such a profound and positive impact on my development as a researcher and a person. I cannot possibly express the weight and sincerity with which I appreciate you.

To my thesis committee chair and mentor, Dr. Chad Smith: I cannot imagine a reality in which I would have been able to complete this project without your guidance, input, and patience. Your feedback and investment into this project and my own learning and development during my time as your student, both during my graduate and undergraduate career, are truly invaluable. Thank you, for everything.

To the Director and Staff of the Center for Diversity and Gender Studies, Dr. Audwin Anderson and Ms. Alyssa Garza: I am confident that I will never experience a work environment as supportive, understanding, and warm as the one I have had as your Graduate Assistant. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	**Description**
EPA	Environmental Protection Agency
U.S.	United States
EJ	Environmental Justice
PCB	Polychlorinated Biphenyl
UCC	United Church of Christ
TRI	Toxics Release Inventory
EPCRA	Emergency Planning and Community Right to Know Act
MSA	Metropolitan Statistical Area
MAUP	Modifiable Areal Unit Problem
GIS	Geographic Information Systems
IDW	Inverse Distance Weighted
TX	Texas
OLS	Ordinary Least Squares
GWR	Geographically Weighted Regression
VIF	Variance Inflation Factor

ABSTRACT

Drawing on concepts of disproportionality and privileged access to guide the research, this project explores the nature of environmental inequality in the Houston-Sugarland-Baytown Metropolitan Statistical Area. Using pollution data from the 2015 Toxics Release Inventory (TRI), sociodemographic data from 2015 American Communities Survey Estimates and distance based methods, this project addresses what groups are more likely to experience heightened levels of toxic releases from TRI sites. Specific variables examined include tract-level racial/ethnic composition, percent non-native, and percent below the poverty level. Segregation variables in this project include a tract-level Multigroup entropy index, Hispanic-white dissimilarity index, and black-white dissimilarity index. Moderating variables include tract level median home value and percent within the manufacturing industry, while population total per tract serves as a control variable in OLS and geographically weighted regression modelling. Findings highlight the impact spatial data have on analysis as well as methodological challenges caused by disproportionality in pollution data. Additionally, results regarding exposure to point-source pollution from TRI sites indicate that Hispanic and white groups were more likely to experience environmental inequality via residential proximity, suggesting that the burden of environmental inequality may be different in urban areas with a majority-minority population such as the Houston-Sugarland-Baytown MSA.

I. INTRODUCTION

The United States' Environmental Protection Agency (EPA) defines environmental justice as the "fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income, with respect to the development, implementation, and enforcement of environmental laws, regulations and policies" (U.S. Environmental Protection Agency 2016). The EPA goes on to clarify that environmental justice, sometimes referred to as environmental equity, will be achieved when all persons receive the same amount of protection from environmental health hazards and have "equal access to the decision making process" (U.S. Environmental Protection Agency 2016). Environmental *inequality* "focuses on broader dimensions of the intersection between environmental quality and social hierarchies" (Pellow 2000) with an emphasis on the unfair and unequal distribution of environmental burdens, or bads, along socially structured lines of class, race/ethnicity, national origin and other demographic and community characteristics. This thesis will focus on social structural forces that shape environmental inequality and will utilize quantitative methodology and publicly available data from the EPA and U.S. Census Bureau to weigh these outcomes.

Pellow (2000) reminds us that unequal distribution of environmental health hazards has historically been "disproportionally distributed around geographic areas with high concentrations of working poor, ethnic minorities, and/or politically disempowered groups" (2000: 591), specifically citing cases of the unequal distribution of human waste in ancient Rome, and he effectively demonstrates that instances of environmental inequality are embedded into our global history. Though instances of environmental injustice can be found across time and place, the environmental justice movement in the

U.S. is often cited as having origins in Warren County, North Carolina (Bullard, Johnson, and Lewis 2000). In 1978 Warren Country was a rural, predominantly African American, working class community. That year, residents of Warren County became aware that the state intended to place a landfill to hold roughly 40 thousand cubic yards of carcinogenic Polychlorinated Biphenyl (PCB) contaminated soil that was removed from along highways in North Carolina in Warren County. Residents and activists mounted a four-year resistance to the landfill placement, drawing national attention to issues related to environmental inequality.

Following the organized protests and national media coverage Warren County protesters received, the environmental justice (EJ) movement was born. Activists and community-based organizations worked with individuals in at-risk communities to champion environmental equity. In the late 1980s, the first scholarly and quantitative examination of environmental inequality in the United States was released by the United Church of Christ (UCC) Commission for Racial Justice. The UCC report (1987) found race/ethnicity was the strongest predictor of the location of toxic waste sites in the United States. This report marked the beginning of a flurry of empirical environmental justice (EJ) research released throughout the 1990s and early 2000s that investigated environmental inequality through residential proximity. Largely, EJ scholarship focuses on understanding and ameliorating environmental inequality.

Though this paper, and many studies across EJ literature, focus environmental inequality as measured by residential proximity to hazardous sites, including Superfund sites, landfills, and toxics release inventory (TRI) sites, the environmental justice scholarship is a broad and multidimensional scholarship that discusses and evaluates elements of environmental inequality as related to community access to greenspace, exposure to health hazards in the workplace, and the structural responses to environmental crises to name a few. Each of these scholarly endeavors strives to measure and understand environmental inequality in hopes of better addressing and ameliorating cases of environmental injustice. This scholarship, coupled with efforts by those in the EJ movement and public awareness of devastating environmental catastrophes across the globe, led to the 1986 Emergency Planning and Community Right to Know Act (EPCRA). This legislation was designed to assist local communities in their attempts to protect the environment and public health, while preserving public safety from environmental hazards. It is based largely on the formation of local emergency planning committees created by local fire fighters, health officials, community groups, industrial facilities, emergency managers and representatives from local government and media (U.S. Environmental Protection Agency 2016).

The explicit goal of the EPCRA is to make information on potential environmental health hazards known and readily available to the American public. In order to streamline the process of making data on environmental hazards available for public access, section 313 of the EPCRA created the EPA's Toxics Release Inventory (TRI) program, an annual inventory of sites across the nation that release and process chemicals both on and offsite. The TRI documents and measures roughly 600 chemicals released at TRI sites above acceptable thresholds that are known to pose environmental and health risks in a report available to anyone with internet access and interest through the EPA's Toxics Release Inventory website. Additionally, the EPA has created a user-friendly EJ screening tool that allows the public to view and assess the environmental quality of regions in which there is an interest.

Since its inception, data from the EPA's Toxics Release Inventory data have been used in scholarship employing quantitative methods to measure and understand environmental inequality. These studies commonly seek to measure environmental inequality through proximity and potential exposure to sites determined to negatively impact the environment that have negative impacts on human health. Examinations of the locations of Superfund sites, nuclear waste dumping, and Toxics Release Sites are common across the EJ literature. This thesis aims to add to that scholarship by examining environmental inequality as measured by geographic and residential proximity to toxics release sites while at the same time carefully considering dimensions of race/ethnicity, immigration status, socioeconomic status, segregation, and the path of least resistance, such as land value and available workforce. The basis for examining these variables specifically, as well as the motivation for selected methodology, stems from the following review of the literature.

II. LITERATURE REVIEW

An overview of the environmental justice literature reveals a few key themes. First, several scholars have quantified environmental inequality through an examination of residential proximity to TRI sites and also documented the ways in which geographic proximity to TRI sites has a negative impact on health and well-being (Agency for Toxic Substances and Disease Registry 2006; Atlas 2007; Bevc, Marchall and Picou 2005; Burningham and Thrush 2002; Downey and Willigen 2005; Johnson, Washington, Kind and Gomez 2014; Natural Resources Defense Council 2004). Second, the research reviewed indicates that vulnerable populations, particularly poor communities and communities of color, are at higher risk of experiencing environmental inequality than their affluent and white counterparts (Bullard, Mohai, Saha, and Wright 2008; Cushing, Gause, Meehan, Cendak, Wildean, and Aleeef 2015; Denq, Constance, and Joung 2000; Downey 2006; Fricker and Hengartner 2001; Grant, Traunter, Downey, and Thiebaud 2010; Hipp and Lakon 2010; Jones, Diez-Roux, Hajat, Kershaq, O'Neill, Guallar, Post, Kaufman, and Navas Acien 2014; Lejano and Iseki 2001; Lersch and Hart 2014; Mohai and Saha 2007, 2015; Pastor, Sadd, and Hipp 2001; Pellow 2001; Sicotte 2014; Smith 2007, 2009; Wu and Heberling 2013). Finally, the literature indicates that the inconsistent methodological approaches used to measure environmental inequality results in a wide variation of results. This portion of the literature is also characterized by a lively debate as to which methods are most appropriate for addressing environmental inequality, with many scholars across the discipline calling for a more uniform, comprehensive, and longitudinal approach to measuring environmental inequality. (Ash, Boyce, Chang, and Scharber 2013; Bullard, Mohai, Saha, and Wright 2008; Chakraborty, Maantay, Brender

2011; Conley 2011; Downey 2006; Mohai and Saha 2006, 2007, 2015; Mennis 2002).

Residential Proximity and Toxics Release Inventory Data

Because this thesis focuses upon Toxics Release Inventory (TRI) data, it is first necessary to highlight the literature on residential proximity to TRI data. Such data have been used as a way to measure environmental inequality, as many scholars suggest that geographic residential proximity to TRI sites increases the likelihood of exposure to environmental health hazards (Brender, Maantay, and Chakraborty 2011; Chakraborty et al. 2011). Scholars have found residential proximity, generally defined as being within a 1-, 3-, or 5- mile radius of a TRI site, increases the likelihood of exposure to poor air quality, contaminated sources of water, lead, polychlorinated biphenyl (PCBs), and pollutants that are known to have negative health impacts (Agency for Toxic Substances and Disease Registry 2006; Bevc, Marchall and Picou 2005; Johnson, Washington, Kind and Gomez 2014; Natural Resources Defense Council 2004; Paigen, Goldman, Magnant, Highland, and Steegmann 1987). These same studies have found higher rates of chronic and acute illness, including asthma, dermatitis, and even depression and anxiety, occur among persons who live within defined geographic distances of TRI sites.

Additionally, research found evidence that proximity to TRI sites may also account for an increased sense of neighborhood disorder and weak community ties (Bevc et al. 2005; Downey and Willigen 2005). Survey research conducted with individuals living near TRI sites finds residential proximity to TRI is correlated with poor mental health, including participants who describe feelings of personal powerlessness, perceived neighborhood disorder, increased concern about personal health, as well as higher levels of the physiological markers for stress associated with residential proximity to a TRI site

(Downey and Willigen 2005; Peek, Freeman, Stowe, and Goodwin 2009). Though a good deal of the literature examined for the purpose of this project, and in this theme in particular, focuses on TRI sites, it is important to note that similar results have been found in studies focusing on other point sources of pollution (Adeola 2000; Burningham and Thrush 2003; Fleming, O'Keefe, and Baum 1991), with results indicating that residential proximity to environmental hazards, whether they are TRI sites, Superfund sites, or landfills, generally has a negative impact on both physical and psychological well-being.

Race, Class, and Environmental Inequality

Findings across the literature indicate that poor communities and communities of color are more likely to experience environmental inequality (Bullard et al. 2008; Cushing et al. 2015; Downey 2006; Fricker and Hengartner 2001; Grant, Traunter, Downey, and Thiebaud 2010; Hipp and Lakon 2010; Lejano and Iseki 2001; Lersch and Hart 2014). Research indicates that a fairly strong relationship between socioeconomic status and environmental inequalities exists (Denq et al. 2000; Krieg 2005; Sicotte 2014; Smith 2007, 2009; Mennis 2002; Mohai and Saha 2006, 2015; Mohai, Lantz, Moernoff, House, and Mero 2009; Wu and Heberling 2013) with several studies finding socioeconomic factors such as economic deprivation and poverty to be the strongest indicators of residential proximity and exposure to hazardous facilities. These findings have been supported by other research that asserts that as distance from a TRI site increases, the density and percentage of persons living below the poverty line decreases (Mennis 2002); that poverty is the strongest predictor of landfill presence (Smith 2007, 2009; Pastor Sadd and Hipp 2001), exposure to chemical air releases (Wu & Heberling

2013) and that household income and level of education are strong predictors of proximity to toxic release sites (Mohai et al 2009).

In addition, findings across the environmental justice literature also suggest the relationship between race and exposure to environmental inequality is strong, with the bulk of the research suggesting race is the strongest predictor of environmental inequality (Cushing, Faust, August, Cendak, Wieland, and Alexeef 2015; Downey 2006; Frickner and Hengartner 2001; Grant et al 2010; Hipp and Lakon 2010; Jonston, Werder and Sebastian 2016; Lejano and Iseko 2001; Lersch and Hart 2014; Mohai and Saha 2015; Mohai et al 2009; Pastor, Sadd, and Hipp 2001; Pastor, Sadd, and Morello Frosch 2004; Pine, Marx, and Lakshmanan 2002; Sicotte 2014). When controlling for socioeconomic status, Fricker and Hengartner (2001) found that race/ethnicity is strongly associated with the presence of environmental bads in New York City, noting in particular that the relationship between percent Hispanic and the presence of environmental hazards is significant. Similarly, a few scholars noted the significant relationship between exposure to environmental toxins and the percent non-white in the region (Cushing et al. 2015; Hipp and Lakon 2010; Jones et al. 2014; Lejano and Iseki 2001; Pine, Marx, and Lakshmanan 2002; Ringquist 1997). Most recently, work by Mohai and Saha (2015) considered the relationship between race, class, and proximity to TRI sites through longitudinal analysis that accounted for sociopolitical and historical changes to the area surrounding TRI sites overtime. This analysis, possibly the first of its kind, suggested that though class matters, the most consistent determinant of proximity to TRI sites is race. However, their findings also indicate that over time the residential areas surrounding TRI sites only saw an increase in percent minority and percent poor composition. Noteworthy meta analyses that examine the environmental justice literature found that though a “race v. class” debate exist, with the majority of the literature examined found stronger evidence for race based environmental inequality and weaker evidence for environmental inequality based on class (Ringquist 2005).

An interesting yet smaller portion of the literature reviewed suggests that it is neither poverty nor race alone that are the best predictors of residential proximity to TRI sites; rather it is a combination of factors that creates a path of least resistance for corporations to place their facilities in certain areas (Anderton, Anderson, Oaks, and Fraser 1994; Brulle 2000; Hird and Reese 1998). The kind of available workforce, political participation, resistance, land value and proximity to primary channels of travel (such as major roads and seaports) are all variables that have been found by some scholars to be the best predictor of exposure to environmental inequality. However, this literature (Anderton, Anderson, Oaks, and Fraser 1994; Brulle 2000; Hird and Reese 1998) is a notably smaller and less developed portion of the environmental justice research as an overwhelming portion of the literature reviewed focuses on race and class in attempts to tease out which really determines experiences of environmental inequality. These conflicting findings across the EJ literature that focus on either race or class are known as the ‘race vs. class debate’ and have been attributed to differences in methods and approach (Downey 2006; Mohai and Saha 2006, 2015; Mohai et al 2009).

Though there has been considerable research that considers race and class separately, some scholars (Brulle and Pellow 2006; Lievanos 2015; Mohai and Saha 2015) call for an intersectional approach to environmental justice studies that considers the impact of both race and socioeconomic status simultaneously. This proposition is backed empirically by research that indicates strong interactions exist between race and class (Lievanos 2015), noting that when examining poor communities of color in addition to community characteristics such as immigration status and political engagement, the importance of examining the relationship between race and class is highlighted.

Brulle and Pellow (2006:315) effectively summarize the need to consider race and class together in addition to variables not largely considered by the extant literature in examinations of environmental inequality by asserting the following:

> The social production of environmental inequality cannot be understood through a singularly focused framework that emphasizes one form of inequality to the exclusion of others. Environmental injustices impact human beings unequally along lines of race, gender, class, and nation, so an overemphasis on any one of these factors will dilute the explanatory power of any analytical approach.

Different methods yield different results

The call for a change in the way environmental justice scholarship measures environmental inequality leads us to the final theme found across the literature: different methods yield different results. As previously mentioned, TRI data are widely used as a way to measure environmental inequality with several scholars using geographic proximity as a way to quantify exposure to environmental inequality. These studies employ a unit hazard method, also known as spatial coincidence method, to examine the demographic characteristics of areas surrounding TRI sites and compare them with the demographics of the larger surrounding area. The unit of measurement used in unit hazard analysis is the often census tract, a straight forward and sensible approach because census data are most commonly delivered as census tracts. Census tracts are one of the earliest quantitative methods used to examine environmental inequality and they use demographic data from the census in conjunction with data from the TRI. This method adequately provides insights into the geographic nature of environmental inequality, but has been criticized for failing to fully address exactly *where* people live in relation to TRI sites as there is great variation in census tract size. Additionally, using the census tract as a basis for analysis can be problematic depending on where the TRI is located, as a TRI site may sit within one tract but be geographically closer to persons living in a different tract (Bullard, Mohai, Saha, and Wright 2008; Chakraborty, Maantay, Brender 2011)

Following critiques of unit hazard analysis, distance-based methods were developed in an attempt to address limitations of the previously established method (Chakraborty, Maantay, and Brender 2011; Conley 2011; Downey 2006; Mohai and Saha 2006). Distance based methods, like unit hazard analysis, make use of TRI data and community demographics. However, rather than using the census tract as the unit of analysis, scholars that employ distance based methods develop circular geographic 'buffers' and only consider the demographics of those living within a specific geographic radius to TRI sites. This method uses geographic information systems (GIS) in analysis to develop these buffers and is able to control for variation in host-tract size while examining who specifically resides near TRI sites. Though this approach addresses the limitations of spatial-coincidence analysis, it has its own set of limitations in attempting to measure environmental inequality. Most notably, the selection of distance ranges that form the circular geographic buffers around sites of interest vary significantly without apparent justification with distances ranging anywhere from .5 to 3 km or 1 to 5 miles of a site of interest (Chakraborty, Maantay, and Brender 2011; Mohai and Saha 2006) often with little to no explanation given as to why these distances were chosen.

Pollution plume modeling is a recent development in environmental justice methods that has been hailed as the best and most sophisticated method available to measure environmental inequality. Like distance-based methods, pollution plume modeling sets geographic buffers surrounding TRI sites, but rather than being simple, circular buffers, the buffers generated in pollution plume modeling may vary in shape and size based on local meteorological conditions, composition of chemical released, and dispersion patterns of chemicals of interest (Ash, Boyce, Chang, and Scharber 2013; Chakraborty, Maantay, Brender 2011; Conley 2011). This method also makes use of GIS for analysis, but accounts for differences in chemical dispersion patterns by adding air dispersion modeling into the analysis. This method addresses the limitations of distance based-methods and unit-hazard analysis, but it faces severe challenges because of the requirement of significant amounts of data on chemical/pollutant properties and local meteorological conditions (Chakraborty, Maantay, Brender 2011), and it is the most time consuming and expensive of all methods available for measuring environmental inequality. These data requirements often make it an impractical and difficult option for environmental justice researchers (Chakraborty, Maantay, Brender 2011; Conley 2011; Mennis 2002; Mohai and Saha 2006, 2015)

Though a few options are available to researchers seeking to measure environmental inequality, across the environmental justice scholarship researchers are calling for careful consideration of methods employed and a more uniform approach be employed across future studies. However, the same scholars are careful to note that a change in methodical approach, much like a change in unit of analysis, will result in different estimations of risk and exposure to environmental inequality based on demographic characteristics (Downey 2006; Mohai and Saha 2006, 2015; Mohai et al

2009). Mohai and Saha (2006) demonstrated this when they applied both distance based methods and unit hazard analysis to the same set of data and found that results differed based on the method employed. Results from distance based methods indicated greater disparities existed between race, socioeconomic status, and proximity to toxic facilities than when applying unit hazard analysis. Additionally, Conley (2011) points out that though methods more sophisticated than unit hazard analysis exist, it does not mean that they will always be the most appropriate methods, nor do they ensure more reliable or valid results.

Gaps in the Literature

Examining Environmental Inequality in Relation to Segregation

A great deal of the quantitative Environmental Inequality literature measures environmental inequality in terms of TRI site proximity in urban areas. Using methods ranging from straight forward unit hazard analysis to more complex pollution plume modeling, significant research considers potential exposure to environmental toxins in relation to demographic composition of the area near or around a TRI site. However, with some exceptions (Smith 2007, 2009; Jones et al. 2014), most of the literature fails to consider the relationship between segregation, considered to be the geographic and measurable manifestation of a complex series of political, social, and economic inequalities, and environmental inequality. The underdeveloped nature of this portion of the literature is noteworthy and surprising, as segregation has been identified as “a major contributor to the creation and maintenance of environmental inequality” (Mohai and Saha 2015: 317) because sources of pollution have been found to be sited near neighborhoods that lack political influence and are socially isolated site locally unwanted land uses in such neighborhoods because they are socially isolated and relatively powerless politically (Bullard, Johnson and Torres 2000; Massey and Denton 1993).

Disaggregated Consideration for Race/Ethnicity

Most of the literature reviewed for the purpose of this project considers racial/ethnic variables to examine potential differences in exposure to pollution by race/ethnicity. However, a considerable portion of the literature aggregates the racial/ethnic categories or even fails to consider non-black racial/ethnic minorities. In cases where non-black racial/ethnic minorities are included in analysis, they are, at times, aggregated with black groups to create an overall non-white group to be compared to their white racial/ethnic counterparts. Consideration for differences between black versus white groups relative to experiences of environmental inequality are necessary and have provided a foundation upon which environmental justice work of all kinds, including researcher and activism, occurs. However, consideration for other racial/ethnic minorities alongside analyses that reviews differences between white and black groups' experience environmental inequality can only increase the richness of the data and such consideration has the potential to tell researchers of the similarities and differences between each groups exposure to pollution.

III. THEORETICAL FRAMEWORK

Disproportionality is a concept conceived by Freudenburg that outlines the immense inequalities that exist among polluters in regard to the amount of pollution emitted per polluter. It is characterized by a double diversion that first involves an unequal pattern or disproportionality of "privileged access" (Freudenburg 2005: 89) to environmental resources. This includes not only access to green space or raw materials, but also to the environments' ability to absorb toxic wastes. Freudenburg (2005) notes that the expulsion of waste into the natural environment is done on a highly disproportional basis with profits from the disposal of waste benefiting a few individuals or entities while the costs of such disposal are borne by the entire society. The second diversion is considered a diversion of attention, in which disproportional access to environmental resource are normalized. That is, the unequal access to and use of environmental resources is assumed to be necessary and is rarely questioned.

This theory of privileged access was followed by empirical work identifying a small proportion of facilities that account for a large amount of total pollution. Analysis using a Gini Coefficient to measure inequalities in terms of toxicity rather than pounds of releases found that 8% of the two top polluting industries (SIC Code 33: Primary Metals Industry and SIC Code 28: Chemical and Allied Products) account for almost 80% of the toxicity risk from emissions (Freudenburg 2005: 98). This disproportionality only increased when Freudenburg examined sources of pollution within industry sectors, finding that a single facility accounted for more than 95% of the total pollution emitted within its industry sector (SIC Code 333: Nonferrous Metals).

Disproportionality, measured in either or both total pounds emitted and toxicity, provides a strong theoretical framework for empirical environmental justice research examining specific sources of pollution. Consideration for disproportionality in pollution has most recently been used by Jorgenson, Longhofer and Grant (2016) and has provided a clearer picture of what industries most directly and aggressively contribute to environmental degradation. Considering individual industries and facilities in examinations of pollution will allow for the consideration of disproportionality to be factored into analysis and will shed light on *who* is responsible for pollution and *how much* pollution they are responsible for. Additionally, this methodology has the potential to supplement our understanding of the developing environmental inequality literature.

IV. RESEARCH DESIGN

The research design for this project is outlined below and includes a discussion of the variables used in both the initial and second round of analyses, as well as how the variables are calculated. The gaps in the literature and the theoretical framework, along with the overarching themes across the literature, shape the research design. I will use GIS to employ distance-based methods to investigate the relationship between pollution and the demographic composition of the areas in Houston, Texas, that face exposure to pollution through geographic proximity. Houston has a historical significance to EJ scholarship, as it has been the focus of a number of classic and groundbreaking academic and legal cases including Bean v. Southwestern Management, Inc., one of the first lawsuits to charge environmental inequality as a civil rights violation and 'Solid Waste Sites and the Black Houston Community' (Bullard 1983), an influential quantitative analysis of environmental inequality that asserts that the siting or placing of solid waste sites in Houston were disproportionately placed near Black communities. Additionally, the Houston-Sugarland-Baytown Metropolitan Statistical Area (MSA) is one of the largest urban agglomerations in the state of Texas and hosts 495 TRI sites, more than the Austin and Dallas MSAs combined.

Unit of Analysis

The census tract is a widely accepted unit of analysis for environmental justice scholarship that focuses on proximity while it serves as a proxy for examining neighborhood exposure to environmental inequality. Additionally, the census tract is commonly the smallest, conventional, unit of measure across sources of data like the census and Toxics Release Inventory. Because of the spatial components of this research project, using the smallest available unit of measure is most appropriate in order to try to minimize the effects of the Modifiable Areal Unit Problem (MAUP) that occurs when spatial analysis of geographic data differs significantly on the basis of the selected geographic unit (ESRI 2016). That is, the larger the geographic unit of measure, the higher the chance that data and analysis will be skewed. For these reasons as well as the availability of the data, the unit of analysis for this project will be the census tract. In the Houston-Sugarland-Baytown MSA, there are 1070 census tracts spread across 9 counties. For a list of the counties and the census tracts that compose them, please refer to Appendix A.

Data Sources & Data Collection

For the purpose of this project, secondary sociodemographic, economic, and chemical release data were compiled and used in the analysis to investigate the nature of environmental inequality in the Houston-Sugarland-Baytown MSA.

TRI Data

At the time of analysis, the most recently available and completed version of the Toxics Release Inventory was collected in 2015. Chemicals listed on the annual TRI report fall into one of the following three categories: carcinogens or chemicals that cause other chronic human health issues, chemicals that cause acute human health issues, and chemicals that have a significant and negative impact on the environment (EPA 2016). Since each of the chemicals tracked by the TRI pose potential harm to human health and the environment, all individually listed chemicals (approximately 595) on the 2015 TRI will be considered to be part of the total releases.

Sociodemographic Data

Like the TRI data, the most recently available sociodemographic data at the tract level are from 2015 estimates available from the U.S. Census' American Communities Survey. In addition to providing sociodemographic variables, listed in detail below, the U.S. Census' American Communities Survey website map maker feature was used to create shapefiles, the most common geospatial data format, that form the basis for geospatial analysis in GIS.

Independent and Dependent Variables

Census tracts will be used as the geographic parameters from which data on a number of socio-demographic variables measuring Race/Ethnicity, Segregation, National Origin and Poverty will be reported.

<u>Race/ethnicity</u> will be reported as percent composition, measured continuously from 0 to 100 percent. Using racial/ethnic data reported by the census, the following variables were constructed. WHITE reports the percent total of non-Hispanic white persons residing in each tract and was calculated as follows: total non-Hispanic white population per tract/total population per tract*100. The same method is used for the calculation of subsequent racial/ethnic variables. BLACK reports the percent total of non-Hispanic Black persons residing in each tract while HISPANIC reports the percent total of Hispanic persons from any racial background living in each tract. The OTHER variable reports the cumulative percent total of persons residing in each tract that fall into one or more of the following racial/ethnic groups: American Cherokee, American Chippewa, American Navajo tribal, American Sioux tribal, Asian, Native Guamanian, Native Samoan, Native Other Pacific Islander, Some Other Race, and Two or More races. The abovementioned racial/ethnic groups were combined as a result of the low

population numbers of each of these groups. Additionally, the percent totals for all non-white persons living in each tract were aggregated to create the NWHITE variable.

The desire to examine residential segregation, or residential pattern measures (Iceland 2004), across multiple groups the Multigroup Entropy Index, or Thiel's H, was employed. The Multigroup Entropy Index, which will be referred to simply as the Entropy Index from here on, measures patterns of evenness among groups distributed across a geographic unit (Massey and Denton 1988). Though other commonly used measures of evenness exist, this measure allows for an examination of an unlimited number of groups, whereas other measures commonly examine evenness as a dichotomous measure (Reardon and Firebaugh 2002). The Entropy Index equation used for the purpose of this project will be measured using the following equation as discussed by White (1986):

$$h_i = -\sum_{j=1}^{k} p_{ij} \ln(p_{ij})$$

where k is the number of racial/ethnic groups of interest, p_{ij} is the proportion of the jth racial/ethnic group in tract I, n_j is the total population of the jth racial/ethnic group in tract I, and N_i is the total population in tract I (White 1986). Note that the maximum score for h_i is dependent on the number of racial ethnic groups examined, or ln(k), with tracts that hold higher values being more diverse, or less segregated, than tracts with lower values. For the purpose of this project, the maximum value of h_i is 1.386. So a tract with a score of 1.386 would have proportional amounts of persons from each racial/ethnic category examined whereas a tract with a score of 0 would have only a single racial/ethnic group represented in its population.

Though entropy is a previously established and acceptable proxy measure for segregation, the entropy scores measure a diversity of non-white groups within tracts rather than diversity as the researcher would define it. A score that increases along with the increasing non-white diversity could, and in this case does, result in a measure that reflects a smaller percentage of non-Hispanic whites than would be necessary for true and proportionate diversity. As a result, higher scores may not actually reflect higher levels of diversity, rather higher scores reflect higher levels of diversity within non-white communities. Thus, tracts that have higher entropy index scores often have smaller portions of non-Hispanic white persons but a greater range and distribution of Hispanic, black, and racial/ethnic other groups.

After consideration of the Entropy score and the implications of its consideration in initial analyses, the researcher elected to calculate tract level dissimilarity index scores as a means of comparison. Though a bulk of the segregation literature has done so at larger units of analysis such at the city or MSA level, researchers have previously employed the dissimilarity index at the tract-level to examine segregation that may be obscured if a larger unit of analysis, such as the city or MSA, is used (Akins 2009; Massey and Denton 1993; Smith 2009). Additionally, as the most commonly used measure of segregation, inclusion of the dissimilarity indices along side the entropy scores per tract would provide more robust insight.

The dissimilarity index to be used for the purpose of this project will be measured using the equation developed by Massey and Denton (1998) that has been previously used to measure segregation at the tract level where b represents the number of, in this example, black persons residing within the smaller geographic unit (in this case, the census block group), B represents the number of black persons living within the larger geographic unit (the census tract), while w represents the number of non-Hispanic white persons living in census block and W represents the number of non-Hispanic white persons living with the census tract.

$$\mathbf{D= [0.5\Sigma\ |b/B\text{-}w/W|]*100}$$

This equation was used to calculate dissimilarity indices between both black and non-Hispanic white groups, and non-Hispanic white and Hispanic groups. Since Hispanic is an ethnic rather than a racial category, the point was made to ensure that black Hispanics were not double counted in analysis. In order to do so, Black-Hispanics were considered as part of the black group rather than the Hispanic group in order to simplify analysis.

The dissimilarity index calculates "the percentage of a group's population that would have to change residence for each neighborhood to have the same percentage of that that group as the metropolitan area overall" (US Census Bureau 2002: 119) and will produce a score ranging continuously from 0.0 (total integration) to 1.0 (complete segregation). Scores ranging from 0 to .30 are indicative of low levels of segregation, measures between .30 and .60 indicate moderate levels of segregation, and measures above .60 indicate high levels of segregation (Massey & Denton 1993).

The variable measuring national origin (NNATIVE) reports a percent composition of all persons reported as being U.S. citizens by naturalization or not a U.S. citizen and is measured continuously from 0 to 100. This variable was calculated as follows: U.S. citizen by naturalization+not a U.S. citizen/total population per tract*100.

Poverty (POVERTY) is measured as a percent total per tract of persons for whom poverty status has been determined in the last 12 months. This variable is continuous, ranges from 0 to 100, and was calculated as follows: population for whom poverty status has been determined in the last 12 months/total population*100.

Moderating & Control Variables

Median home value will be measured continuously in U.S. Dollars based on 2015 estimates, and is reported per tract by the Census. The workforce composition variable is measured continuously from 0 to 100 and (PERCMAN) reports the percent composition of persons per tract over the age of 16 that are active in the workforce and that work a Manufacturing job in any industry. These two variables serve as proxy measures for the path of least resistance. The path of least resistance asserts that pollution producers will base the siting of their facilities on an ideal combination of land value and available workforce that will present the least amount of resistance to the building and continued operation of their facility. As a result, these variables will be used as moderating variables, in order to assess the strength of the relationship between the independent and dependent variables. Finally, Population Total, is a continuous measure that reports the total population per tract and will serve as a control variable in the analysis.

Dependent Variables

The TRI is a self-reported annual inventory collected by the EPA that provides valuable data on pollution. For the purpose of this project, the dependent variable will include the total amount of on/offsite chemical releases, measured continuously in millions of pounds, released both on and offsite into the surrounding air, water, and land reported on the 2015 TRI for the Houston-Sugarland-Baytown MSA.

Shapefile Creation

As mentioned previously, shapefiles based on the tracts of interest were created using the ‘Map Maker’ feature on the U.S. American Communities Survey website. These shapefiles provide the geographic basis for analysis in GIS, but had no sociodemographic or TRI data associated with them. Rather, the shapefiles generated held only spatial data and outlined the individual tracts of interest. For a visual representation of what the original shapefiles downloaded look like, please refer to Appendix B. Following the download of the sociodemographic data in Excel format, I used the Join feature in GIS to combine data on the variables of interest with their respective tracts. This join resulted in a shapefile that had both geographic and sociodemographic data for the area of interest.

The raw TRI data including information on the total on/offsite releases was downloaded in Excel format and contained a variety of geographic data per site. The geographic data provided by the TRI included a SITE ID assigned by the EPA per TRI site, the longitude/latitude per site and, most relevant to this project, the physical street address per site. This information allowed for the use of a Geocoding tool in GIS to create a point on the map document that corresponds with the geographic location of each TRI site creating a visual representation of the physical locations.

Due to the structure of the raw data, which reports data from a single site via multiple data point entries, the layer resulting from the initial round of geocoding included multiple points per site. This structure prevents proper statistical analysis from being conducted, as each site has the potential to be represented more than once. In order to amend this issue, I used the Dissolve feature in GIS to combine all points that share the same geographic location. This created a layer in GIS that that combined all the data previously represented with multiple points with a single point. For a visual representation of the resulting shapefile, see Appendix C. At this stage in data preparation, the amount of toxic releases is demonstrated by individual static points. However, this does not represent the nature of pollution. As defined by the TRI itself, the reported chemicals are released into the air, land, and water surrounding the TRI site.

Interpolation

In attempts to accurately represent the nature of TRI releases with the tools and skills I am currently equipped with, I employed the use of the interpolation tool in GIS. Interpolation estimates the value of the input variable, in this case Total TRI Releases, across cells on my shapefile by using a "linearly weighted combination" (ESRI 2017) of a sample of points. By using the Interpolation tool estimates of the total toxic releases per tract were generated, even in tracts that did not have TRI sites within their geographic boundary. Generally speaking, chemical releases, especially releases to air and water, do not respect constructed geographic boundaries such as tract boundaries. To classify area as being exposed to pollution and one as not on the basis of whether or not it hosts or is within a specific geographic proximity to a TRI site does not adequately shed light onto the dynamic and often pervasive nature of pollutants. Thus, Interpolation provides more robust analysis, as it provides a more accurate representation of what exposure to chemical emissions from TRI sites may look like.

IDW Interpolation

Though there are multiple methods of Interpolation available, the locally dependent nature of the Toxic Release Data made the Inverse Distance Weighted (IDW) method the most appropriate method, as it is weighted based on, as the name implies, an inverse distance function. This means that the farther away from the input variable source, in this case the TRI site, the less impact that TRI site will have on the estimated Total Releases per tract. IDW interpolation will consider and quantify the estimated releases for tracts near multiple TRI sites by taking into account the total releases across multiple sites simultaneously and weighting their estimated impact on a target area accordingly. In addition to being the most appropriate method for interpolation due to the nature of the data, IDW interpolation has been found to provide the best estimation for releases such as TRI releases and addresses Tobler's first law of geography (Vorapracha, Phonpransert, Khanaruksombat, and Pijarn 2015). For a visual representation of the resulting map, see Appendix D.

The review of the data following initial analysis revealed intense disproportionality to be discussed in the following sections of this paper, prompting concern that the largest outliers in the MSA may be driving the results from initial modelling. In order to investigate whether the disproportionality in total TRI releases per site was skewing the model, estimation of the total releases per tract was re-run using IDW interpolation without the largest two outliers. The re-interpolation generated estimates per tract that did, in fact, deviate from the initial interpolation estimates. For a visual representation of the new IDW results please see Appendix G**.**

V. ANALYSIS/RESULTS

Analysis for this project are discussed below as follows. First, an overview of the descriptive statistics are provided to give insight into the test area. Second, discussions of the exploratory regression models used to develop subsequent models and vet the variables of interest are discussed. Finally, ordinary least squares models and subsequent geographically weighted regression models are discussed. Note that there are two sets of all analyses, with exception of geographically weighted regression. The first set of analyses presented in each case are run with the entirety of the TRI data that are available, while the second set of analyses for each test are run with the largest two outliers excluded.

According to 2015 estimates, the Houston-Sugarland-Baytown MSA (here forth just referred to as Houston) is home to 6,346,652 people that live across 1,070 census tracts. There are nine counties that comprise the MSA, including Austin, Brazoria, Chambers, Fort Bend, Galveston, Harris, Liberty, Montgomery and Waller county. Notable municipalities include Houston, Galveston, Lake Jackson, Sugarland and The Woodlands to name a few. Its population is 38.20% White, 36.11% Hispanic. 16.75% Black, and 8.95% Other, making it a majority minority MSA as just over 61% of the population fall into racial/ethnic minority groups. 2.26% of the total population for the MSA are non-native, having either being naturalized citizens or non-citizens, while 15.64% of the total population was classified as below the poverty line in 2015.

In 2015, the TRI reports that Houston was home to 495 TRI sites that released a combined 91.3 million pounds of on/offsite releases into the air, land, and water surrounding each TRI site ranking it among the top 20 urban areas in regard to total releases per square mile. According to the TRI report, the majority of the releases in the Houston area were to land (56.9 million lbs), followed by releases to air (18.1 million lbs.) and then water (6.1 million lbs) with only 10.1 million pounds reported as being Off-Site.

An examination into the amount of total releases per site quickly revealed grave disproportionality in the amount of releases per site. A handful of TRI sites located in the densest part of the MSA reported 0 releases, an outcome that warranted further investigation leading to the identification of a potential issue with the raw data. In the TRI dataset 0 actually means 0, and is not indicative of releases below a certain threshold established by the TRI. However, by their very nature, sites listed on the TRI are responsible for toxic releases into their surrounding communities. Due to the self-reported nature of the TRI, and the lack of incentive to report releases, or to report them accurately, it is quite possible that any or all of the reported releases are incomplete or inaccurate.

In addition to several sites reporting 0 releases, two sites account for roughly half of the total releases in the Houston MSA. The largest single site in terms of total releases accounts for roughly 36% of the total releases in the MSA. Ascend's Performance Materials Chocolate Bayou plant, located in Alvin, TX, manufactures "and supplies chemicals, fibers, and plastic products" (ASCEND 2017) for commercial and industrial use and reports that 99.37% of its releases were to Land. Meanwhile, the second largest site in the MSA, TM Deer Park Services, accounts for 14.1% of the total releases in the Houston MSA. TM Deer Park Services is a Hazardous Waste Treatment and Disposal Facility and, in 2015, reported that 99.95% of its releases were to surrounding land. It is noteworthy that though both Ascend and TM Deer Park report an overwhelming majority of their releases are land releases, both sites are closer to the coast of the Gulf of Mexico than most of the TRI sites in the MSA.

The statistical analysis and results reported below were all conducted using the Statistical Analyst tools built in to ArcGIS 10.4, including Exploratory Regression, Ordinary Least Squares Regression, and Geographically Weighted Regression when appropriate. These tests were run in the order that they are mentioned above as a way to test the strength and validity of the variables and dataset, and to ensure that the model was properly developed.

Exploratory Regression Results with Largest Two Outliers

Exploratory Regression was run in order to help determine whether the independent/explanatory variables selected would yield “any properly specified OLS models” (ESRI 2017). Though the resulting report generated by the Exploratory Regression on the variables of interest generated a great deal of information, including the strength of a possible OLS models and the variable stability, the information utilized was provided by the Summary of Variable Significance and the Summary of Multicollinearity portions of the generated report.

The Summary of Variable Significance table indicates how consistent variable relationships are along with the proportion of times the variable was found to be statistically significant during testing. As mentioned, the Summary of Variable Significant table also provides information on how consistent variable relationships were found to be during testing. Strong independent or explanatory variables will be found to be consistently and primarily either positive or negative as well as consistently significant.

The Summary of Multicollinearity table is used alongside the Summary of Variable Significance to determine which explanatory variables should be removed, if any, in order to strengthen subsequent analysis. Specifically, the Summary of Multicollinearity table indicates how many times each variable ran into issues of Multicollinearity when run in the exploratory regression models and which variables were also included in those models. In essence, this table indicates whether or not the model will suffer from multicollinearity and will indicate what variables are problematic by reporting the VIF (<10 within acceptable range, with investigation warranted if VIF>4), the number of violations (how many times the variable resulted in multicollinearity), and its covariates.

The first exploratory regression run on the independent variables measuring home value, poverty, segregation (entropy), nativity, non-white population, percent manufacturing, and total population size provided the most promising results. Results for this model are reported in Table 1A below. Note that, for the exploratory models, variables are ordered from largest to smallest in terms of significance. All variables were found to have consistent explanatory power, with 4 variables (Population, Nativity, Percent Manufacturing, and Home Value) found to be significant 100% of the time. More specifically, Population, Nativity, and Home Value were found to have a negative relationship with total releases per tract 100% percent of the time while Percent Manufacturing was found to have a positive relationship with total releases per tract 100% of the time.

The dichotomous race variable, Non-white, was significant 78.95% of the time and indicated a mostly negative relationship with total releases per tract. In this exploratory model, poverty was found to be significant 70.18% of the time and had a mostly positive relationship with the dependent variable. Entropy, the measure of segregation in this model, was the least statistically significant of all variables and was only found to be significant 26.32% of the time. Of the times it was found to be statistically significant, Entropy exhibited a primarily positive relationship with the dependent variable. Due to the way Entropy is measured, this means that as diversity within a tract increased, so did the number of total releases per tract.

TABLE 1A: SUMMARY OF VARIABLE SIGNFICANCE | EXPLORATORY REGRESSION MODEL I

	%Significant	%Negative	%Positive
POPULATION	100.00	100.00	0.00
NATIVITY	100.00	100.00	0.00
MANUFACTURING	100.00	0.00	100.00
HOME VALUE	100.00	100.00	0.00
NON-WHITE	78.95	89.47	10.53
POVERTY	70.18	36.84	63.16
ENTROPY	26.32	8.77	91.23

The Summary of Multicollinearity report for the first exploratory regression run are reported in Table 1B below. None of the variables measured revealed any violations of multicollinearity, and the VIF for each variable (reported below) were well within the acceptable range. As a result of both the summary of Variable Significance and the Summary of Multicollinearity, this model moves on to the next phase of analysis and is fit for both OLS and GWR modelling.

TABLE 1B: SUMMARY OF MULTICOLLINEARITY | EXPLORATORY REGRESSION MODEL I			
	VIF	Violations	Covariates
POPULATION	1.10	0	-----
POVERTY	2.15	0	-----
NON-WHITE	2.70	0	-----
NATIVITY	1.58	0	-----
MANUFACTURING	1.07	0	-----
HOME VALUE	1.45	0	-----
ENTROPY	1.18	0	-----

The second exploratory regression was run on the independent variables measuring Home Value, Poverty, Segregation (Entropy), Nativity, Percent Manufacturing, and the disaggregated race variables on the White, Hispanic, Black, and Other populations per tract. Like the first exploratory regression, all variables were found to have consistent and explanatory power, with variables (Home Value and Nativity) found to be significant 100% of the time. More specifically, both Home Value and Nativity were found to have a negative relationship with total releases per tract 100% percent of the time.

The Other race variable was found to be significant 96.32% of the time and indicated a mostly negative (60.74%) relationship with the dependent variable. The 'Hispanic' variable was statistically significant 78.53% of the time, and also had a mostly negative relationship with the dependent variable (62.58%). The 'Black' variable was found to be significant 63.19% of the time, and indicated a mostly negative relationship (98.77%) with the dependent variable while the 'White' variable was just barely mostly significant (55.21%) and indicated a mostly positive relationship with the dependent variable (69.33%).

Manufacturing was found to be significant 76.69% of the time and had a mostly positive relationship with the dependent variable (95.09%). Poverty was found to be significant 73.62% of the time and had a mostly negative relationship with the dependent variable. Again, segregation as measured by Entropy was found to be significant the least amount of the time only being found significant 26.38% of the time. Of the times it was found to be significant, it was found to have a mostly positive relationship with the dependent variable.

TABLE 2A: SUMMARY OF VARIABLE SIGNFICANCE | EXPLORATORY REGRESSION MODEL II

	%Significant	%Negative	%Positive
POPULATION	100.00	100.00	0.00
HOME VALUE	100.00	100.00	0.00
NATIVITY	100.00	100.00	0.00
OTHER	96.31	60.74	39.26
HISPANIC	78.53	62.58	37.42
MANUFACTURING	76.69	4.91	95.09
POVERTY	73.62	87.12	12.88
BLACK	63.19	98.77	1.23
WHITE	55.21	30.67	69.33
ENTROPY	26.38	15.34	84.66

The Summary of Multicollinearity report for the second exploratory regression run yielded mostly positive results. Only one variable, Nativity, revealed potential issues with multicollinearity and had a VIF outside of the acceptable range. An interesting finding surrounds the failure of the Summary of Multicollinearity to report any co-variates. However, an examination of the VIFs reported per variable led the researcher to deduce that there were potential multicollinearity issues with the Non-Native variable and the Hispanic and Other race variables, as the VIF for both the Hispanic and Other variables are higher than the remaining race variables and within the VIF range that warrants investigation. After considering both the summary of variable significance and the summary of multicollinearity, it became apparent that for this model, the non-native variable should not be included in the subsequent OLS and GWR modelling.

TABLE 3A: SUMMARY OF MULTICOLLINEARITY | EXPLORATORY REGRESSION MODEL II

	VIF	Violations	Covariates
POPULATION	1.50	0	
HOME VALUE	1.26	0	-----
HISPANIC	6.52	0	-----
WHITE	2.90	0	-----
BLACK	1.41	0	-----
OTHER	6.52	0	-----
POVERTY	2.47	0	-----
NATIVITY	11.06	22	-----
MANUFACTURING	3.85	0	-----
ENTROPY	1.30	0	-----

Following guidance provided by the summary of multicollinearity and variable significance for Model I, exploratory regression was re-run with the revised model in order to assess how the removal of the nativity variable influenced the relationship between variables. With the removal of the nativity variable from the model, significance for the percent manufacturing variable increased sharply, from 76.69% to 100.00% significant, though the relationship with total releases per tract remained the same (as expected). The resulting exploratory regression generated the following summary of variable significance and resulted in further refining of the model. Additionally, variable significance among the disaggregated racial/ethnic variables shifted across all three groups, with a decrease in percent significance for the Hispanic and White racial/ethnic groups and increase in significance for the Black group. Again, the nature of the relationship between these variables and the total releases per tract remained the same directionally, but was clarified. Poverty lost some significance across the test models run, whereas entropy gained significance.

TABLE 3B: SUMMARY OF VARIABLE SIGNFICANCE | REVISED EXPLORATORY REGRESSION MODEL II			
	%Significant	%Negative	%Positive
POPULATION	100.00	100.00	0.00
HOME VALUE	100.00	100.00	0.00
OTHER	100.00	100.00	0.00
MANUFACTURING	100.00	0.00	100.00
WHITE	96.93	3.68	96.32
HISPANIC	73.01	77.30	22.70
BLACK	70.55	67.48	32.52
POVERTY	46.63	62.58	37.42
ENTROPY	39.88	15.95	84.05

The Summary of Multicollinearity report for the revised exploratory regression run on Model II yielded results that indicated further refinement needed to occur once the nativity variable was removed. The racial ethnic groups for Hispanic, Black, and White all indicated severe multicollinearity that would present problems for further analysis. However, the racial/ethnic group 'Other' was not collinear with the rest of the racial/ethnic groups, but underwent an increase in VIF indicating its inclusion in the model may be the source of some of the collinearity seen across the Hispanic, Black, and White groups. After careful consideration for the substantive impacts of excluding the racial/ethnic 'Other' group, this group was removed from subsequent analysis and the exploratory regression was run for a third and final time.

TABLE 3C: SUMMARY OF MULTICOLLINEARITY | REVISED EXPLORATORY REGRESSION MODEL II			
	VIF	Violations	Covariates
POPULATION	1.10	0	-----
POVERTY	2.19	0	-----
MANUFACTURING	1.13	0	-----
HOME VALUE	1.53	0	-----
ENTROPY	1.57	0	-----
HISPANIC	33.00	13	WHITE, BLACK
WHITE	41.18	22	HIPSANIC, BLACK
BLACK	23.08	6	WHITE, HISPANIC
OTHER	6.01	0	-----

The final exploratory regression model tested the relationship between the variables measuring population total, percent manufacturing, median home value, percent white, percent Hispanic, percent black, poverty rates, and entropy per tract. This model generated satisfactory results that indicated subsequent analysis with these variables, including OLS and GWR modelling, could be conducted. The Summary of variable significance clarified variable significance for all of the variables selected. Just as in previously exploratory regression models population, percent manufacturing, and home value per tract were found to be strong variables that were significance 100% of the time. Population and home value both exhibited a negative relationship with the dependent variable whereas percent manufacturing indicated a positive relationship with the dependent variable. In this model, the significance for percent white per tract increased to 100%, with a strong positive relationship with the dependent variable. Percent black and percent Hispanic both indicated mostly negative relationships with the dependent variable, and entropy was found to have a mostly positive relationship with the variable but decreased drastically in the number of times it was found to be significant in testing.

TABLE 4A: SUMMARY OF VARIABLE SIGNIFICANCE | FINAL EXPLORATORY REGRESSION MODEL II			
	%Significant	%Negative	%Positive
POPULATION	100.00	100.00	0.00
HOME VALUE	100.00	100.00	0.00
MANUFACTURING	100.00	0.00	100.00
WHITE	100.00	0.00	100.00
BLACK	69.70	57.58	42.42
HISPANIC	67.68	63.64	36.36
POVERTY	42.42	50.51	49.49
ENTROPY	4.04	26.26	73.74

The resulting summary of multicollinearity for the final exploratory regression run for this model indicated that the removal of the racial/ethnic 'Other' group did in fact alleviate some of the issues revealed in the previous model, but did not completely address collinearity entirely. This model, though indicative of some collinearity between the Hispanic and white racial/ethnic groups, is strong enough for additional analysis. Additionally, further removal of the racial/ethnic groups for the sake of addressing collinearity would weaken the substantive implications of any analysis as the removal of additional racial/ethnic groups, especially ones that comprise a substantial portion of the population in the MSA, would fail to accurately represent the test area.

TABLE 4B: SUMMARY OF MULTICOLLINEARITY | FINAL EXPLORATORY REGRESSION MODEL II			
	VIF	Violations	Covariates
POPULATION	1.10	0	-----
POVERTY	2.19	0	-----
MANUFACTURING	1.13	0	-----
HOME VALUE	1.47	0	-----
ENTROPY	1.34	0	-----
HISPANIC	8.59	7	WHITE
WHITE	10.96	16	HISPANIC
BLACK	6.13	0	-----

Exploratory Regression Results without Top Two Sites

The first set of exploratory regressions tested the same variables as the initial models, but included the new IDW estimates. For the first exploratory regression run, variables measuring population, poverty, percent non-white, percent manufacturing, home value, and entropy were tested in order to test if the relationships that were evident in the first round of testing would hold in the absence of the extreme outliers.

For this model, 4 variables were found to have consistent explanatory power, with three variables (non-native, percent manufacturing, and home value) indicating 100% variable significance with the variable for non-white indicating significance most of the time (68.42% of the time). Of the variables that were found to have 100% variable significance, percent manufacturing held a positive relationship with the dependent variable, and percent non-native and home value held a negative relationship with the dependent variable. The non-white variable was significant 68.42% of the time and was found to have a mostly negative relationship with the dependent variable (73.68%). Population, which had been significant 100% of the time in the previous model, lost a substantial portion of its significance during this exploratory regression and was found to be significant only 19.30% of time. Of the times it was fond to be significant, it held a negative relationship with the variable as it had in previous analysis. Finally, entropy in this test lost all significance.

TABLE 5A: SUMMARY OF VARIABLE SIGNIFICANCE | EXPLORATORY REGRESSION MODEL I

	%Significant	%Negative	%Positive
NATIVITY	100.00	100.00	0.00
MANUFACTURING	100.00	0.00	100.00
HOME VALUE	100.00	100.00	0.00
NON-WHITE	68.42	73.68	26.32
POVERTY	24.56	40.35	59.65

TABLE 5A CONTINUTED: SUMMARY OF VARIABLE SIGNIFICANCE | EXPLORATORY REGRESSION MODEL I			
	%Significant	%Negative	%Positive
POPULATION	19.30	100.00	0.00
ENTROPY	0.00	28.07	71.93

No issues of multicollinearity were made apparent by the summary of multicollinearity generated for this exploratory regression, as it was identical to the previous summary of multicollinearity generated for the independent variables examined.

TABLE 5B: SUMMARY OF MULTICOLLINEARITY | EXPLORATORY REGRESSION MODEL I			
	VIF	Violations	Covariates
POPULATION	1.10	0	-----
POVERTY	2.15	0	-----
NON-WHITE	2.70	0	-----
NATIVITY	1.58	0	-----
MANUFACTURING	1.07	0	-----
HOME VALUE	1.45	0	-----
ENTROPY	1.18	0	-----

The second set of exploratory regressions were run with the new IDW estimates on the independent variables measuring population, poverty, percent manufacturing and the disaggregated racial/ethnic variables (percent white, percent Hispanic, and percent black) yielded the following results. Just as with the first set of exploratory regressions, as a precaution, the black-white and Hispanic-white dissimilarity indices were run separately. The first exploratory regression was run with the Hispanic-white dissimilarity index and the previously mentioned variables.

For this model, 5 variables were found to have consistent explanatory power, with two variables (percent manufacturing and home) indicating 100% variable significance, while the remaining variables indicated healthy levels of variable significance more than half of the time, indicating sufficient variable stability. Of the variables that were found to have 100% variable significance, percent manufacturing held a positive relationship with the dependent variable while home value held a negative relationship with the dependent variable. The percent white variable was significant 81.82% of the time and was found to have a positive relationship with the dependent variable 100% of the time. Percent black was found to be significant 56.57% of the time and held a mostly negative relationship with the dependent variable (83.84% of the time).

The percent Hispanic variable was mostly found to have a positive relationship with the dependent variable and was significant 79.80% of the time. Unlike in previous exploratory models, poverty was found to have a mostly negative relationship with the dependent variable (61.62% of the time) and was only significant in 20.20% of the models run. Similar deviations from previous models were found with the population variable, which only held significance 34.34% of the time but, like previous models, was found to have a negative relationship with the dependent variable. Finally, Hispanic-white dissimilarity was found to be significant only 4.04% of the time and held a negative relationship 100% of the time.

TABLE 6A: SUMMARY OF VARIABLE SIGNIFICANCE | EXPLORATORY REGRESSION MODEL IIA

	%Significant	%Negative	%Positive
MANUFACTURING	100.00	0.00	100.00
HOME VALUE	100.00	100.00	0.00
WHITE	81.82	0.00	100.00
BLACK	56.57	83.84	16.16
HISPANIC	51.52	20.20	79.80
POPULATION	34.34	100.00	0.00
POVERTY	20.20	61.62	38.38
HW DISSIMILARITY	4.04	100.00	0.00

The Summary of Multicollinearity generated for this model revealed some multicollinearity between the variables measuring the percent Hispanic and percent white. Though some multicollinearity is present between the two aforementioned variables, the VIFs for both are within a technically acceptable range. After weighing the potential impacts of removing one of these groups for the sake of decreasing collinearity, I decided to keep both the percent Hispanic and percent white variables as removing them could potentially weaken both the statistical and substantive significance of any further analysis. The remaining variables all held VIFs that fell well within the accepted parameters. As a result, OLS modelling was conducted to generate coefficients for each of the variables in question.

TABLE 6B: SUMMARY OF MULTICOLLINEARITY | EXPLORATORY REGRESSION MODEL IIA

	VIF	Violations	Covariates
POPULATION	1.06	0	-----
POVERTY	2.25	0	-----
MANUFACTURING	1.13	0	-----
HOME VALUE	1.45	0	-----
HISPANIC	7.70	2	WHITE
WHITE	9.51	16	HIPSANIC
BLACK	5.90	0	-----
HW DISSIMILARITY	1.12	0	-----

Findings from the exploratory regression model run with the black-white dissimilarity index mostly mirrored the findings of the exploratory regression model with the Hispanic-white dissimilarity index. Six variables were found to have consistent explanatory power, with two variables (percent manufacturing and home) indicating 100% variable significance, while the remaining variables indicated healthy levels of variable significance more than half of the time, indicating variable stability. Like the previous version of this model run, out of the variables that were found to have 100% variable significance, percent manufacturing held a positive relationship with the dependent variable while home value held a negative relationship with the dependent variable. The percent white variable was significant 82.83% of the time and was found to have a positive relationship with the dependent variable 100% of the time. The black-white dissimilarity index held significance 64.65% of the time and held a negative relationship with the dependent variable 100% of the time. The percent black variable held the same significance and directional relationship as in the previous version of this model, and was found to be significant 56.57% of the time and held a mostly negative relationship with the dependent variable (83.84% of the time).

The percent Hispanic variable was mostly found to have a positive relationship with the dependent variable and was significant 53.54% of the time. Population, again, was found to have a negative relationship with the dependent variable (100% of the time) and was only significant in 36.36% of the models run. Similar deviations from previous models were found with the poverty variable, which only held significance 22.22% of the time and held a negative relationship with the dependent variable 100% of the time.

TABLE 7A: SUMMARY OF VARIABLE SIGNIFICANCE | EXPLORATORY REGRESSION MODEL IIA

	%Significant	%Negative	%Positive
MANUFACTURING	100.00	0.00	100.00
HOME VALUE	100.00	100.00	0.00
PERCENT WHITE	82.83	0.00	100.00
BW DISSIMILARITY	64.65	100.00	0.00
PERCENT BLACK	56.57	83.84	16.16
PERCENT HISPANIC	53.54	19.19	80.81
POPULATION	36.36	100.00	0.00
POVERTY	22.22	64.65	35.35

The Summary of Multicollinearity generated for second version of this model revealed the same multicollinearity between the variables measuring the percent Hispanic and percent white. Just as in the previous model, though multicollinearity is present between the two aforementioned variables, the VIFs for both are within an acceptable and both variables were included in subsequent OLS modelling.

TABLE 7B: SUMMARY OF MULTICOLLINEARITY | EXPLORATORY REGRESSION MODEL IIB

	VIF	Violations	Covariates
POPULATION	1.06	0	-----
POVERTY	2.17	0	-----
MANUFACTURING	1.13	0	-----
HOME VALUE	1.45	0	-----
HISPANIC	7.70	2	WHITE
WHITE	9.51	16	HIPSANIC
BLACK	5.90	0	-----
BW DISSIMILARITY	1.00	0	-----

OLS Results

Following the vetting of the variables during exploratory regression testing, two sets of Ordinary Least Squares regression models were developed to provide a global set of parameter estimates for the area of interest. The first set of models includes the aggregated race variable, non-white along with variables measuring poverty, nativity, manufacturing, entropy as well as home value and total population. The adjusted R^2 for this model is .16. and all variables except for the variable measuring poverty rates per tract were found to be statistically significant at α=0.01. However, though the measure for poverty was not statistically significant, a close look at the directional nature of the result is worth consideration as it's p=0.07 value falls just outside of significance at α=0.05.

TABLE 8A: OLS MODEL I RESULTS					
	Coefficient	*Std. Error*	*T-Statistic*	*Prob.*	*VIF*
POPULATION	*-4.728*	*1.28*	*-3.67*	*0.00**	*1.099*
MANUFACTURING	*3913.77*	*824.65*	*4.74*	*0.00**	*1.07*
HOME VALUE	*-0.23*	*0.03*	*-6.37*	*0.00**	*1.47*
POVERTY	*995.30*	*558.19*	*1.78*	*0.07*	*2.18*
NON-WHITE	*-1074.57*	*287.87*	*-3.73*	*0.00**	*2.72*
NON-NATIVE	*-3345.11*	*448.88*	*-7.47*	*0.00**	*1.58*
ENTROPY	*52146.41*	*19225.27*	*2.71*	*0.00**	*1.20*

*Significant at 0.01
R^2: .17
Adjusted R^2: .16
AICc: 28689.84
Moran's I: .21

The second model includes the disaggregated racial/ethnic variables that passed the exploratory regression testing for variables significance and fell within reasonably acceptable range for multicollinearity. This includes variables measuring the percent white, Hispanic, and black categories along with variables measuring poverty, manufacturing, home value and entropy. The adjusted R^2 for this model is .11. Variables measuring population total, manufacturing, median home value, and the racial/ethnic variables for percent white and percent black were all found to be statistically significant at α=0.01. Non-significant variables for this model include variables measuring the percent Hispanic, segregation (entropy), and poverty rates per tract. However, though these variables were not significant, they fell just out of range for statistical significance (α=0.07) so their relationship with the dependent variable is worth consideration for their substantive implications. For a visual representation of the OLS models, see Appendix E.

TABLE 8B: OLS MODEL II RESULTS					
	Coefficient	*Std. Error*	*T-Statistic*	*Prob.*	*VIF*
POPULATION	*-5.27*	*1.32*	*-3.99*	*0.00**	*1.10*
MANUFACTURING	*3655.57*	*872.74*	*4.18*	*0.00**	*1.13*
HOME VALUE	*-0.28*	*0.03*	*-7.34*	*0.00**	*1.49*
POVERTY	*42.24*	*582.22*	*0.07*	*0.94*	*2.24*
ENTROPY	*38358.16*	*21264.00*	*1.80*	*0.07*	*1.39*
HISPANIC	*1105.71*	*612.53*	*1.80*	*0.07*	*9.25*
WHITE	*3009.33*	*596.334*	*5.04*	*0.00**	*10.99*
BLACK	*1339.73*	*618.51*	*2.16*	*0.03**	*6.39*

*Significant at 0.01
R^2: .12
Adjusted R^2: .11
AICc: 28695.01
Moran's I: 0.30

OLS Model Results without Top Two Sites

Ordinary Least Squares models were developed for each version of the exploratory regression models outlines in the previous section. The first model includes the aggregated racial/ethnic non-white variable, alongside variables measuring poverty, nativity, population, home value, and entropy. With the exclusion of the two largest outliers there was a loss of significance across the variables measuring population, percent non-white, and entropy. In this model only variables measuring nativity, percent manufacturing, and home value retained their significance with percent manufacturing being the only statistically significant variable to hold a positive relationship with the dependent variable. The adjusted R^2 for this model is .07 and results can be seen in the table below.

TABLE 9: REVISED OLS MODEL I RESULTS					
	Coefficient	Std. Error	T-Statistic	Prob.	VIF
POPULATION	-1.70	1.07	-1.58	0.11	1.09
MANUFACTURING	3301.36	690.01	4.78	0.00*	1.07
HOME VALUE	-0.14	0.03	-4.55	0.00*	1.47
POVERTY	323.59	4567.05	0.69	0.48	2.18
NON-WHITE	-311.71	240.87	-1.29	0.19	2.72
NON-NATIVE	-1676.05	375.59	-4.46	0.00*	1.58
ENTROPY	17189.98	16086.26	1.06	0.28	1.20

*Significant at 0.01
R^2: .08
Adjusted R^2: .07
AICc: 28252.716337
Moran's I: 0.16

The second OLS model run includes the disaggregated racial/ethnic variables along with the remaining independent variables examined previously in Model II. Most notably, population and percent black lost significance with the new IDW estimates for the dependent variable and percent Hispanic gained significance. However, significance was retained for percent white, home value, and percent manufacturing as was the direction of these relationships.

TABLE 10: REVISED OLS MODEL II RESULTS					
	Coefficient	Std. Error	T-Statistic	Prob.	VIF
POPULATION	-2.09	1.08	-1.93	0.05	1.10
MANUFACTURING	2802.95	714.82	3.92	0.00*	1.13
HOME VALUE	-0.15	0.03	-4.98	0.00*	1.49
POVERTY	-308.43	476.87	-0.64	0.51	2.24
ENTROPY	25504.86	17416.39	1.46	0.14	1.39
HISPANIC	1325.60	501.70	2.64	0.00*	9.24
WHITE	1799.99	488.43	3.68	0.00*	10.99
BLACK	909.52	506.59	1.79	0.07	6.39

*Significant at 0.01
R^2: .07
Adjusted R^2: .06
AICc: 28267.865202
Moran's I: 0.17

The following model includes the aggregated race variable, non-white race variable along with the variables measuring poverty, percent manufacturing, home value, and the Hispanic-white dissimilarity index calculations. Results for this OLS model can be seen in Table 11 and are as follows. Variables measuring percent manufacturing, median home value, and the aggregated racial/ethnic non-white variable were all found to be statistically significant at α=0.01. Non-significant variables for this model include variables measuring population, poverty, and the Hispanic-white dissimilarity index variable. However, though these variables were not significant, they fell just out of range for statistical significance (α=0.07) so their relationship with the dependent variable is worth consideration.

TABLE 11: REVISED OLS MODEL I RESULTS

	Coefficient	Std. Error	T-Statistic	Prob.	VIF
POPULATION	-1.79	1.06	-1.68	0.09	1.06
MANUFACTURING	3172.24	689.94	4.59	0.00*	1.05
HOME VALUE	-0.16	0.03	-5.59	0.00*	1.40
POVERTY	140.55	456.64	0.30	0.75	2.05
NON-WHITE	-791.40	210.97	-.375	0.00*	2.05
HW DISSIMILARITY	-177.62	237.48	-0.74	0.45	1.08

*Significant at 0.05
** Significant at 0.01
R^2: .26
Adjusted R^2: .06
AICc: 28269.51
Moran's I: 0.17

The second set of OLS models correspond with the set of exploratory regression models discusses previously that include variables that measure the disaggregated racial/ethnic variables percent white, percent black, percent Hispanic, and variables measuring home value, percent manufacturing, population and poverty per tract. The first model includes the aforementioned variables and the Hispanic-white dissimilarity index calculations. For this model, variables measuring percent manufacturing, home value, and the percent white variable were statistically significant at α=0.01, while the percent Hispanic variable was significant at α=0.05. Non-significant variables for this model include population, poverty, percent black and the Hispanic-white dissimilarity index. Results for this model can be seen in table 12 below.

TABLE 12: OLS MODEL IIA RESULTS

	Coefficient	Std. Error	T-Statistic	Prob.	VIF
POPULATION	-1.76	1.06	-1.65	0.09	1.06
MANUFACTURING	2742.63	714.59	3.83	0.00**	1.13
HOME VALUE	-0.16	0.03	-5.18	0.00**	1.46
POVERTY	-342.87	482.18	-0.71	0.47	2.29
HISPANIC	1065.66	465.22	2.29	0.02*	7.93
WHITE	1567.90	456.90	3.43	0.00**	9.60
BLACK	760.67	493.39	1.54	0.12	6.05
HW DISSIMILARITY	-172.82	241.73	-0.71	0.47	1.12

*Significant at 0.05
** Significant at 0.01
R^2: .06
Adjusted R^2: .06
AICc: 28269.51
Moran's I: 0.17

The second model includes the above listed variables and the black-white dissimilarity calculations. For this model, variables measuring percent manufacturing, home value, and percent white were all statistically significant at α=0.01, with the percent Hispanic variable holding significance at α=0.05. Non-significant variables for this model include percent black, population, poverty, and the black-white dissimilarity index and the adjusted R^2 value was .06.

TABLE 13: OLS MODEL IIB RESULTS

	Coefficient	Std. Error	T-Statistic	Prob.	VIF
POPULATION	-1.81	1.06	-1.69	0.08	1.06
MANUFACTURING	2756.35	714.70	3.85	0.00**	1.13
HOME VALUE	-0.16	0.03	-5.26	0.00**	1.45
POVERTY	-412.61	471.81	-0.87	0.38	2.19
HISPANIC	1054.30	464.89	2.26	0.02*	7.92
WHITE	1538.55	455.19	3.38	0.00**	9.52
BLACK	712.36	488.92	1.45	0.14	5.94
BW DISSIMILARITY	-2.39	5.49	-0.43	0.66	1.00

*Significant at 0.05
** Significant at 0.01
R^2: .06
Adjusted R^2: .06
AICc: 28269.83
Moran's I: 0.17

Geographically Weighted Regression Results

The models reported above both generate a single set of parameter estimates across the entire test area. As discussed previously, the data and variables in question include spatial components that are not thoroughly addressed by the OLS parameter estimates. In order to better address and understand the relationship between variables with consideration to the spatial nature of the data, localized model should be developed that. As a result, the same models tested and reported above were used to conduct Geographically Weighted Regression (GWR). GWR runs thousands of regression models across the entirety of the test area in bandwidths based on manually selected "kernel type, bandwidth method, distance and number of neighbors" (ESRI 2017) so that the spatial nature of the data may be accounted for.

Prior to shifting from a global OLS model to a local and geographically weighted model, an investigation into whether the variables are spatially autocorrelated is necessary. Using the Global Moran's I tool in ArcGIS, I tested for spatial autocorrelation on the residuals of OLS models. This tests whether geographic patterns in the residuals of the variables are clustered, dispersed, or random (ESRI 2017), and will generate a score that indicates whether spatial autocorrelation will provide challenges during analysis. The Moran's I values are reported below AICc values in the tables above and indicate that, for both models, spatial autocorrelation is not an issue. As a result, GWR testing is possible and will face no challenges stemming from spatial autocorrelation.

Geographically Weighted Regression

The GWR model I corresponds to OLS model I and includes variables on Population total, Poverty, the dichotomous Non-White variable, Nativity, Percent Manufacturing, Home Value, and Entropy. The resulting output generated an adjusted R^2 value of .81. The GWR model II corresponds to the OLS model II and includes the disaggregated race variables for White, Hispanic, Black along with variables for Poverty, Manufacturing, Home Value, and Entropy. The model generated an adjusted R^2 value of .80.

Since both GWR models generated a similar adjusted R^2 value, a closer look at the AICc was taken to determine the most appropriate model and gauge the explanatory power of each. The AICc is "measure of model performance" (ESRI 2017) and is most helpful in situations like this one where comparing different GWR models is necessary. Generally speaking, the model with the lower AICc is a better fit for the data. If the AICc for models in question varies by less than three, then the explanatory power of each model is considered equivalent. However, if the AICc differs by more than three, then the model with the lower value is considered a better fit. Additionally, comparing AICc values for a GWR model to the OLS AICc value provides insight into the benefits and motivation of moving from a global model (OLS) to a local model (GWR).

TABLE 14: ADJUSTED R^2 AND AICC VALUES FROM OLS AND GWR MODELS

	OLS R^2	AICc OLS	GWR R^2	AICc GWR
MODEL I	.16	28634.19	.81	27057.19
MODEL II	.11	28695.01	.80	27116.60

Even though both GWR models generated the similar adjusted R^2 values, the results of the AICc comparison allows for the strongest model, Model I, to be identified. However, the AICc comparison of both models suggests that though they both hold statistical significance and considerable explanatory power the model with the aggregated racial/ethnic variable as well as the dichotomous variable on nativity in analysis is a more appropriate fit for the research question. For the map corresponding to the GWR analysis, see Appendix F.

VI. DISCUSSION

Results from the exploratory regression Model I (Table 5A) suggest that as Population per tract increases (β=-4.728), the total releases per tract decreases. This is an interesting finding considering population has been previously found to be positively associated with other measures of pollution. This finding could be the result of disproportional releases from TRI sites located in less densely populated areas within the MSA, a suggestion supported by the map resulting from IDW interpolation (see Appendix D) that indicates the largest sources of pollution from the TRI are located in rural and suburban areas near the coast.

Median home value was found to have a slightly negative relationship with the total releases per tract (β=-0.23), meaning that as home values increase the total releases per tract decrease (and vice versa). This finding lends support to the notion that a path of least resistance, composed in part by land value, is a factor in identifying where TRI facilities are located.

As the percentage of persons working a manufacturing job increases, so does the total number releases per tract (β=3913.77). This finding, like the finding for home value, lends support to the 'path of least resistance' explanation for exposure to pollution and is indicative of the tendency for persons to live and work within the same area.

Finally, as entropy increased within a tract so did the total pollution per tract (β=52146.41). This finding was considered with great caution, as it could be misleading. Entropy, as discussed previously, considers residential patterns across multiple groups and measures evenness with higher scores indicating higher levels of diversity with the maximum score (1.386 in this case) indicating an evenness that reflects the distribution of racial/ethnic groups within the total test area. Taking the racial/ethnic composition of the MSA into consideration, this means that a tract with a score of 1.386, for example, would have a racial/ethnic composition that mirrors the MSA. Specifically, a tract with a score of 1.386 would have 38.20% White, 36.11% Hispanic. 16.75% Black, and 8.95% Other.

As the number of naturalized or non-citizens' increases, the total releases per tract decreases (β=-2245.11). This could be indicative of a demographic trend for immigrant communities to be located in more densely populated parts of the test area. Similarly, when looking at race as an aggregated variable, as the non-white population per tract increased the total releases per tract decreased (β=-1074.57). However, this finding contradicts not only the existing environmental inequality literature, but also contradicts the second model. In the second OLS model, disaggregated racial/ethnic variables were included rather than an aggregated non-white variable. Results from the OLS model indicate that a *positive and statistically significant* relationship exists between the total amount of TRI releases per tract and the percent black and white in a tract. This contradiction is further evident when considering the positive directional, yet non-statistically significant, findings in regard to the percent Hispanic and overall diversity/segregation as measured by the Entropy index. Though the results of the first OLS model could be a function of population density and residential patterns for non-white communities, the inconsistency with the literature and even the second model in this project warrant further investigation.

The OLS Model I indicates a statistically significant relationship exists between all of the variables in Model I with the exception of the variable that measures poverty per tract. Though the relationship between total TRI releases per tract and poverty rates is not statistically significant, its positive relationship (β=995.30) it falls just outside of the range of significance (α=0.07) so it's implications and the directional findings are considered here. This finding suggests that as there is a unit increase in the population that falls below the poverty line, there is an increase in the total releases per tract, and falls in line with previous research that indicates the same relationship exists between the concentration of poverty per tract and exposure to pollution.

The OLS model coefficients reported above are for the general global model applied to the entire test area, without the consideration or weighting of spatial data. This OLS model generated an adjusted R^2 of .16, indicating that there is a missing component while addressing the relationship between the total releases per tract and the variables selected. However, when accounting for the spatial nature of the TRI release data and weighting variables across the test area accordingly, the GWR for this model generates an adjusted R^2 of .81, increasing the explanatory power the variables have relative to the total TRI releases per tract.

VII. IMPLICATION OF RESULTS

First and perhaps most striking is the considerable amount of support the descriptive statistics have for Freudenberg's theory of Disproportionality. Two sites out of more than 400 account for roughly half of the total releases in the Houston-Sugarland-Baytown MSA, with a single site accounting for over 32% of the total releases in the MSA. Furthermore, the single largest source of TRI releases in the test area falls into the SIC Code 28: Chemicals and Allied Products industry that was found to be one of the two industries that account for almost 80% of toxicity risk (Freudenburg 2005: 98).

That the findings highlight severe disproportionality provided support to Freudenburg's theory and aligns with the corresponding literature, but required additional analysis be conducted as the disproportionality provided challenges to statistical investigation. Subsequent analysis on the test area with the exclusion of the two largest outliers confirmed the suspicion that these outliers likely had an undue influence upon the results and were driving the results. That the disproportionate nature of TRI releases seems to have driven some of the results for the first model touches upon a methodological conundrum that should be explored in future research. Though outliers present challenges to statistical analysis. For comparison, the two largest outliers in the TRI were excluded during the secondary analysis, removing them from the dataset fails to provide a complete picture of what pollution in Houston looks like and who is potentially exposed via residential proximity. Though taking the results from both the initial and secondary models congruently allows for some insight, I am unsure if either model addresses the research question completely as each captures a potentially incomplete picture of what environmental inequality looks like in the test area. Any subsequent analysis should be considered carefully and additional methods to address the disproportionate releases should be investigated.

Results from the first model (both OLS and GWR) indicate that total releases, or pollution, per tract is a function of geographic location and displays the power that the addition of spatial data have on determining the relationship between the independent variables of interest and total releases per tract. This model also indicates that TRI releases in the Houston MSA are highest in tracts that are more proportionately diverse, less densely populated and are adequately described as suburban/rural. Additionally, this model supports the notion that tracts that have those within the manufacturing industry and experiencing poverty are most likely to experience increased exposure to TRI releases.

These findings lend support to previous research and would be enhanced by additional, time-series based analysis that could indicate how the relationship between the variables of interest and proximity to TRI sites/exposure to total TRI releases may or may not changes with demographic trends in the area. Though the implications of the findings fit within the boundaries of previously established research, each of the resulting coefficients are small relative to the estimated total releases emitted per tract should be considered as the estimated total releases per tract is very throughout the entire test area. Even tracts with low total releases were still estimated to have an exposure to 8 million pounds TRI releases when estimates were calculated with the inclusion of the largest two TRI sites. Though initially unintentional, the inclusion of two measures for segregation in analysis, the dissimilarity and entropy indices, allowed for a more robust and in depth quantitative examination of the relationship between the residential patterns of different racial/ethnic groups and their potential exposure to pollution from TRI sites. The initial models suggest that as diversity increases, or segregation decreases, within a tract, the total pollution per tract decreases. This is counter-intuitive to previous findings in the literature and is what prompted further investigation and the inclusion of dissimilarity indices for Hispanic-white and black-white groups. Results from those analyses confirmed what the initial models were suggesting that as segregation (as measured by dissimilarity indices) increased, total estimated pollution per tract decreases. These findings held for models run with and without the largest two TRI sites included in estimates.

In addition to the findings relative to segregation and diversity contradicting the literature, the findings on potential exposure to pollution by race and ethnicity do not align with previous environmental inequality literature in general, and even environmental inequality literature specific to the area. Specially, findings from analyses that removed the largest outliers while using the disaggregated racial/ethnic variables suggest that Hispanic and white populations were the more likely to reside in tracts that have higher TRI releases than their black counterparts. Though there is a directional finding that suggests that black groups also experience some increased exposure to TRI releases, it lacks statistical significance. Additionally, the coefficients generated by the analysis indicate that though some increase in estimated TRI releases may occur, the estimated increase in TRI releases for black groups is not nearly as high as the estimated increases seen for Hispanic and white groups. Though counter and contradictory to the existing literature, these findings prompt future investigation into potential social-structural factors.

First, these findings could be a function of the test area. The test area for this project is much larger than the test area in previous research that examines environmental inequality. Specifically, previous research on the test area focuses primarily on Houston city proper or Harris county alone (Bullard 1983; Johnson et al. 2014). This project expands the test area to include eight additional counties that encompass the entire MSA, as the researcher feels that the sprawling nature of the urban area makes considering *only* the city or single county alone an arbitrary decision considering the distribution of the population. The larger test area, though arguably more appropriate than consideration of a single city given the urban sprawl throughout the MSA, may be the reason for some of the deviations between the findings of this project and the existing literature.

Second, the Houston-Sugarland-Baytown MSA is a majority-minority urban agglomerate. The racial/ethnic composition of the MSA alone could result in environmental inequalities that may not be found in other areas. It may be that, for areas in which the racial/ethnic minority populations compose a substantial or majority of the total population, that environmental burdens look differently than they would in areas in which whites are the majority or in which Hispanics are a much smaller portion of the population. That is, in areas with a majority or substantial Hispanic population, or any racial/ethnic minority, environmental inequality may fall along different and unexpected cleavages of race/ethnicity than the literature has previously found.

Though the MSA is a majority-minority area, this is a relatively recent development. A great deal of previous environmental inequality research was conducted in the 1980's and early 1990's, when minority groups may not have accounted for a majority of the total population (Anderton et al. 1994; Bullard 1983; United Church of

Christ 1987). Though consideration and comparison of environmental inequality via proximity to TRI sites via longitudinal analysis would be the only way to make any definitive commentary on the nature of this relationship over time, brief consideration of racial/ethnic demographic trends reveals that the Hispanic population in the MSA has grown substantially. That the Hispanic population accounts for one third of the total population in the MSA, along with other changes in the demographic composition of the MSA, could mean that the environmental burden of proximity to pollution via TRI sites has shifted along with these changing trends.

VII. LIMITATIONS AND FUTURE RESEARCH

This project's limitations, of which there are many, provide an outline from which future research can be informed and strengthened. The research project discussed here is cross sectional in nature, considering data that are from a single year. The inclusion of time-series analysis, particularly with a test area that bears historical and cultural significance for environmental justice like Houston, would allow researchers to investigate how the relationship between exposure to pollution/chemical releases changes or remains constant over time. It would also allow for results derived from 2015 data to be contextualized and could confirm whether or not the results here are an anomaly or if they are indicative of relationships that hold constant over time.

The inclusion of longitudinal analysis on the test area and selected variable could also provide more insight into the finding between home value and total releases per tract. Currently, the results suggest that as total releases per tract increase, home values decrease. This finding *could* provide support to the path of least resistance hypothesis, lending any support to this claim would require consideration for whether locations for these TRI sites were based on home value (as a proxy for land value) or whether the home values decreases following the presence of a TRI site. As this is a cross-sectional project the data are only able to shed light on what was occurred in 2015.

Consideration for on-site releases separate from total releases in future research could provide a stronger basis from which interpolation could occur, and might more accurately represent the way chemical releases are distributed across the test area. Pollution plume modelling that includes data for elevation, topology, climate and wind direction could also strengthen future analysis, so long as the researcher remained focused on the research question and was able to derive substantive answers from the data and analysis. Additionally, this project only includes a single source of pollution and does not include measures for other sources of pollution. The inclusion of multiple sources of pollution into the analysis would enhance the analysis and such inclusion could address some of the questions that have been raised by the results of this research.

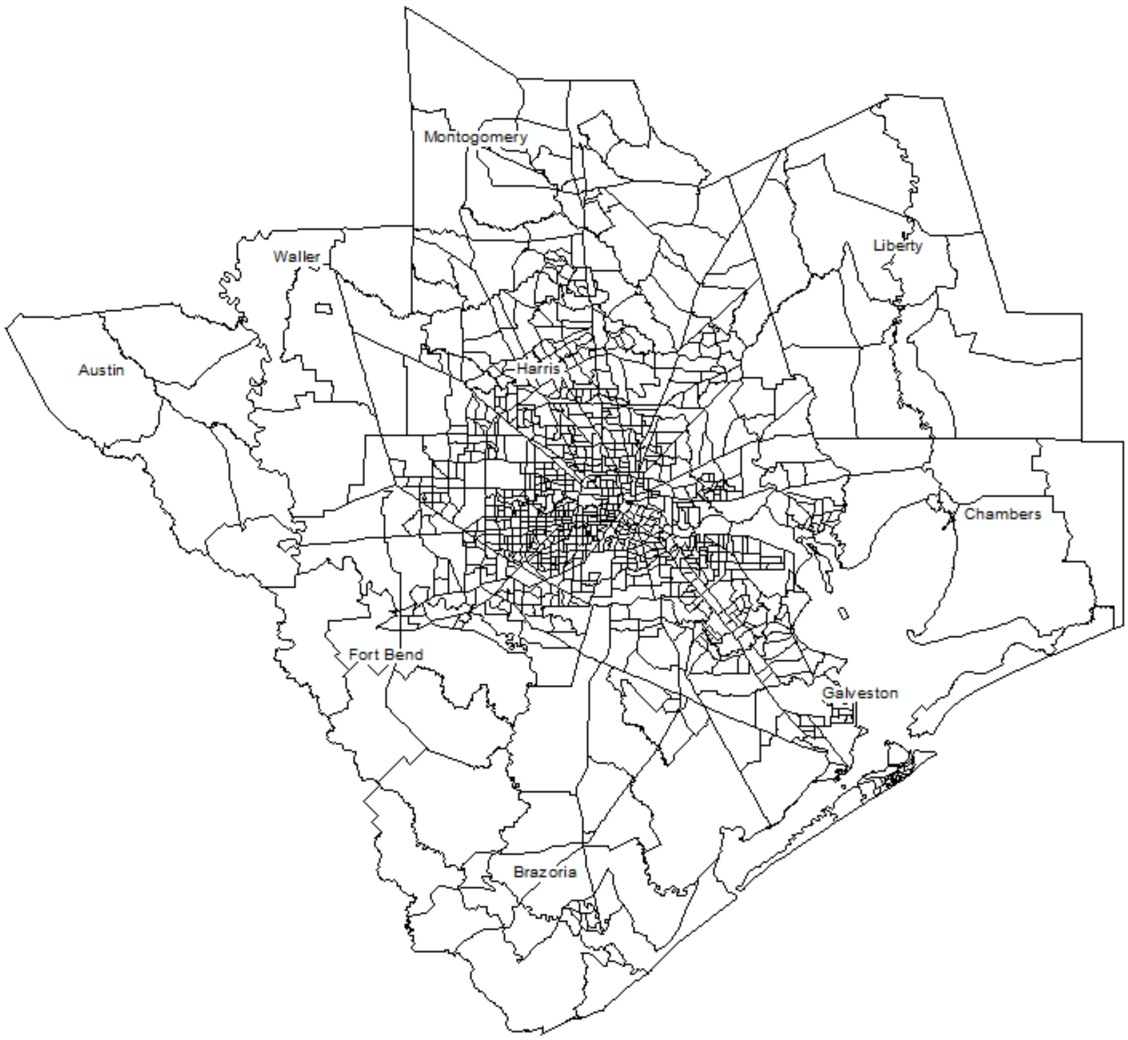
Though it has been well established in many ways by the literature that proximity to TRI sites and sources of pollution in general is detrimental, there was no inclusion of public health data or any quality of life measure in this study. Adding these kinds of data in future research would allow for a more robust discussion of the implications of proximity to pollution for those living closest to these sits. It, like the additional analysis of on-site releases with a time series approach, would strengthen the potential explanatory power of the models used and could provide more meaningful answers to the research question. Finally, this research project employs a strictly quantitative approach. Though appropriate for the specific research question, the addition of qualitative data allowing for persons living within the test area to be represented by more than secondary data estimates would more personally tie the research and findings to the communities being studied.

APPENDIX SECTION

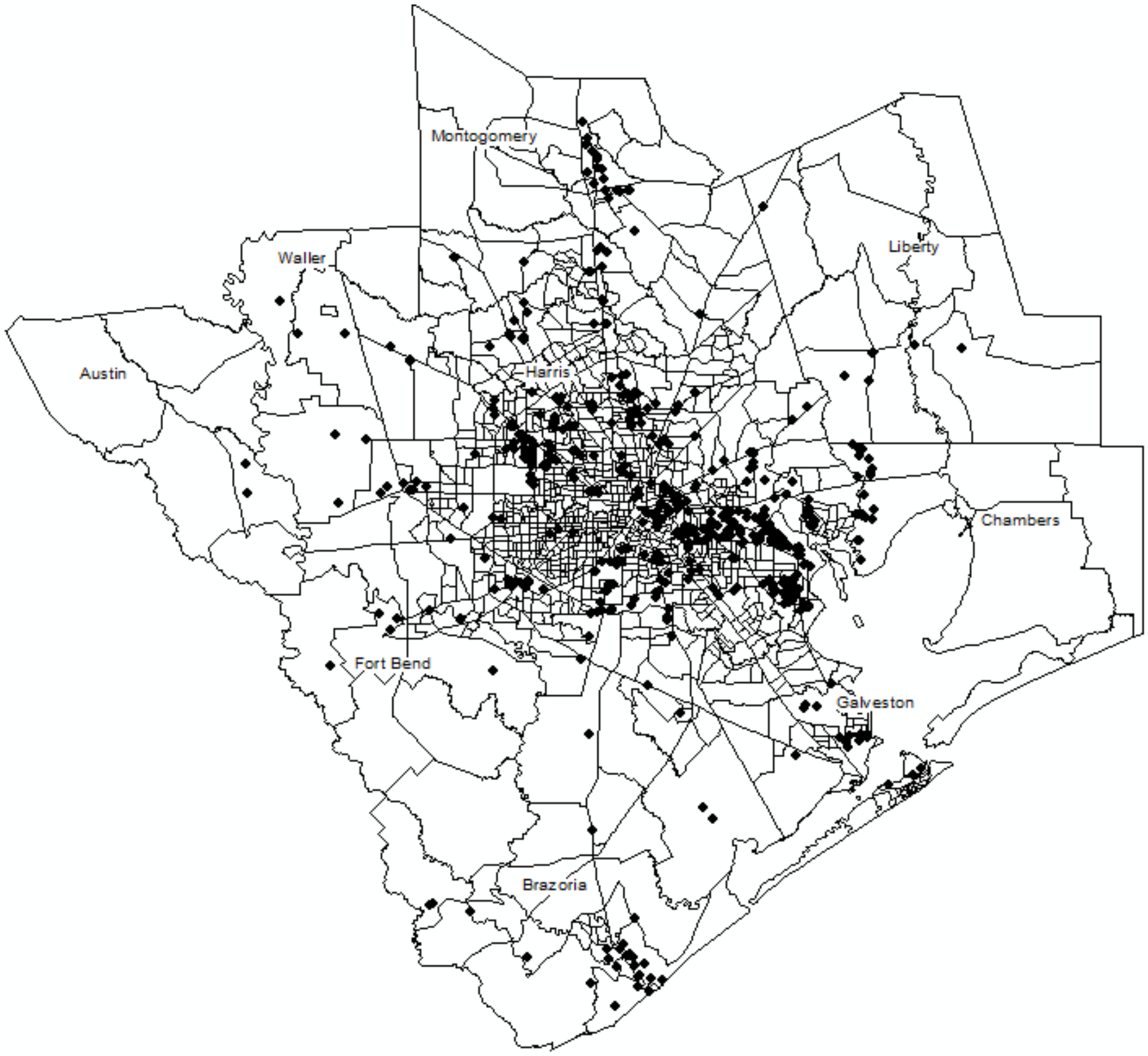
Appendix A: List of Counties in Test Area

Austin
Brazoria
Chambers
Fort Bend
Galveston
Harris
Liberty
Montgomery
Waller

Appendix B: Tract Delineated Shapefile
N=1070

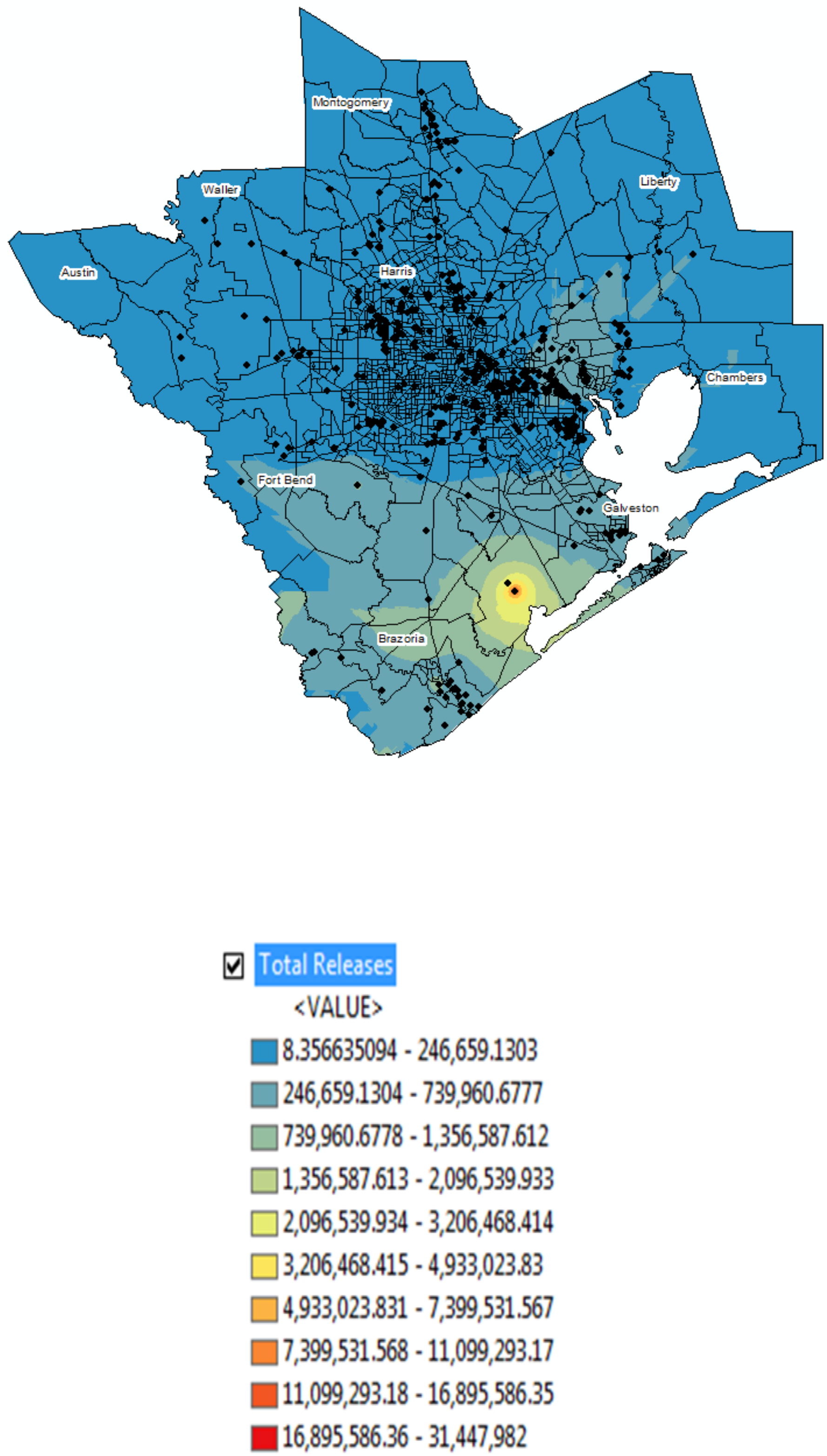


Appendix C: Tract Delineated Shapefile with TRI point Data
N=1070

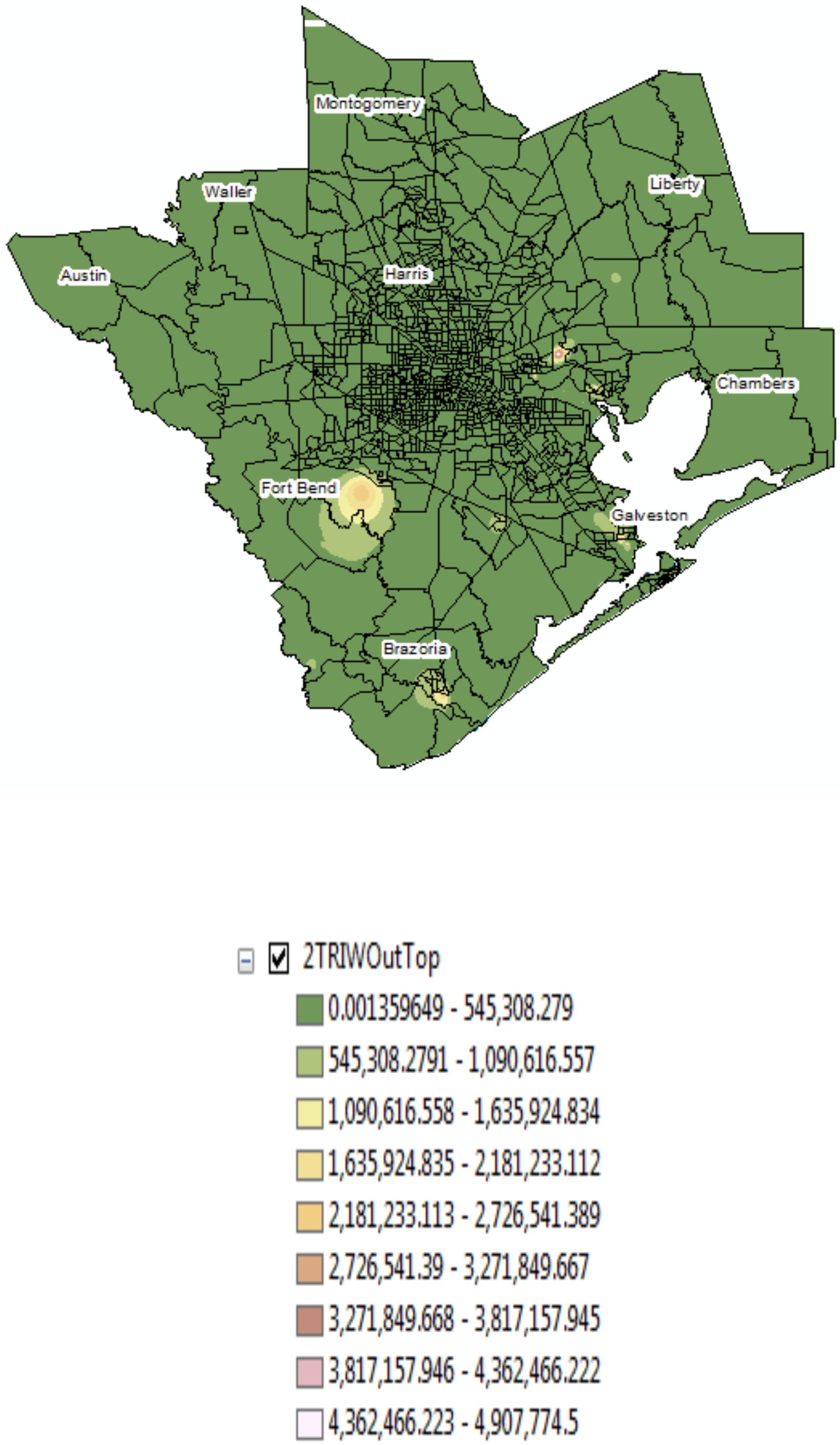


Appendix D: IDW Shapefiles

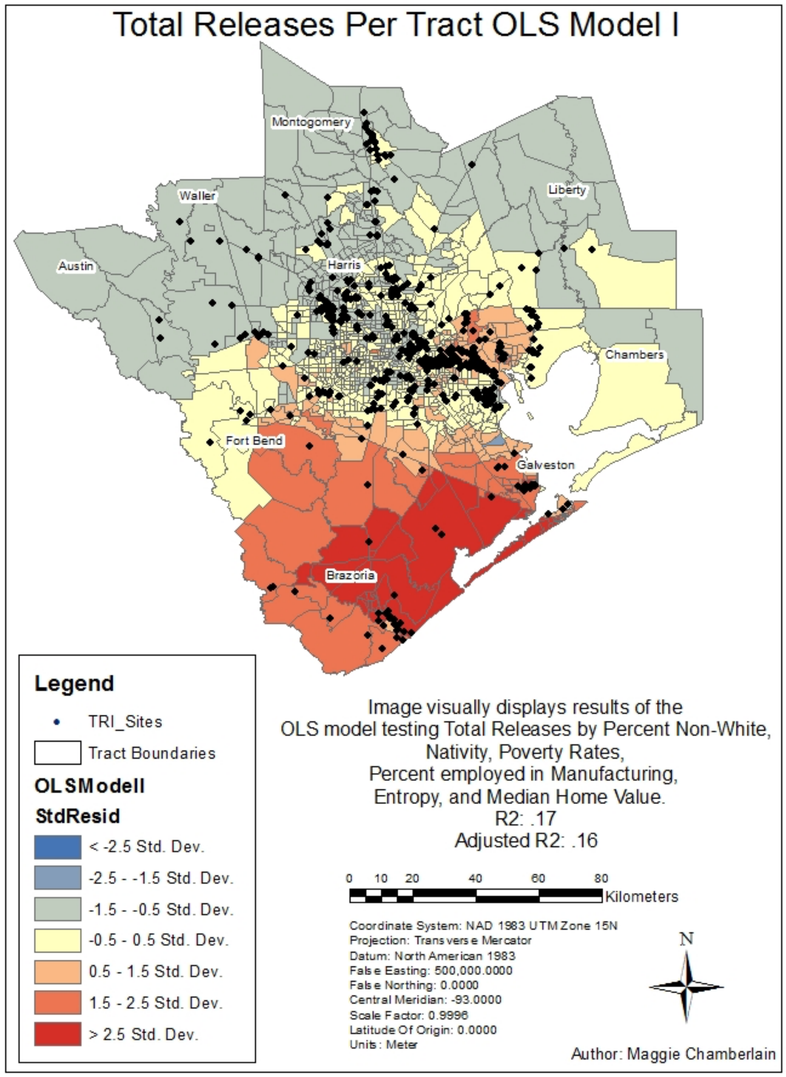
IDW Estimates with Top Two Outliers

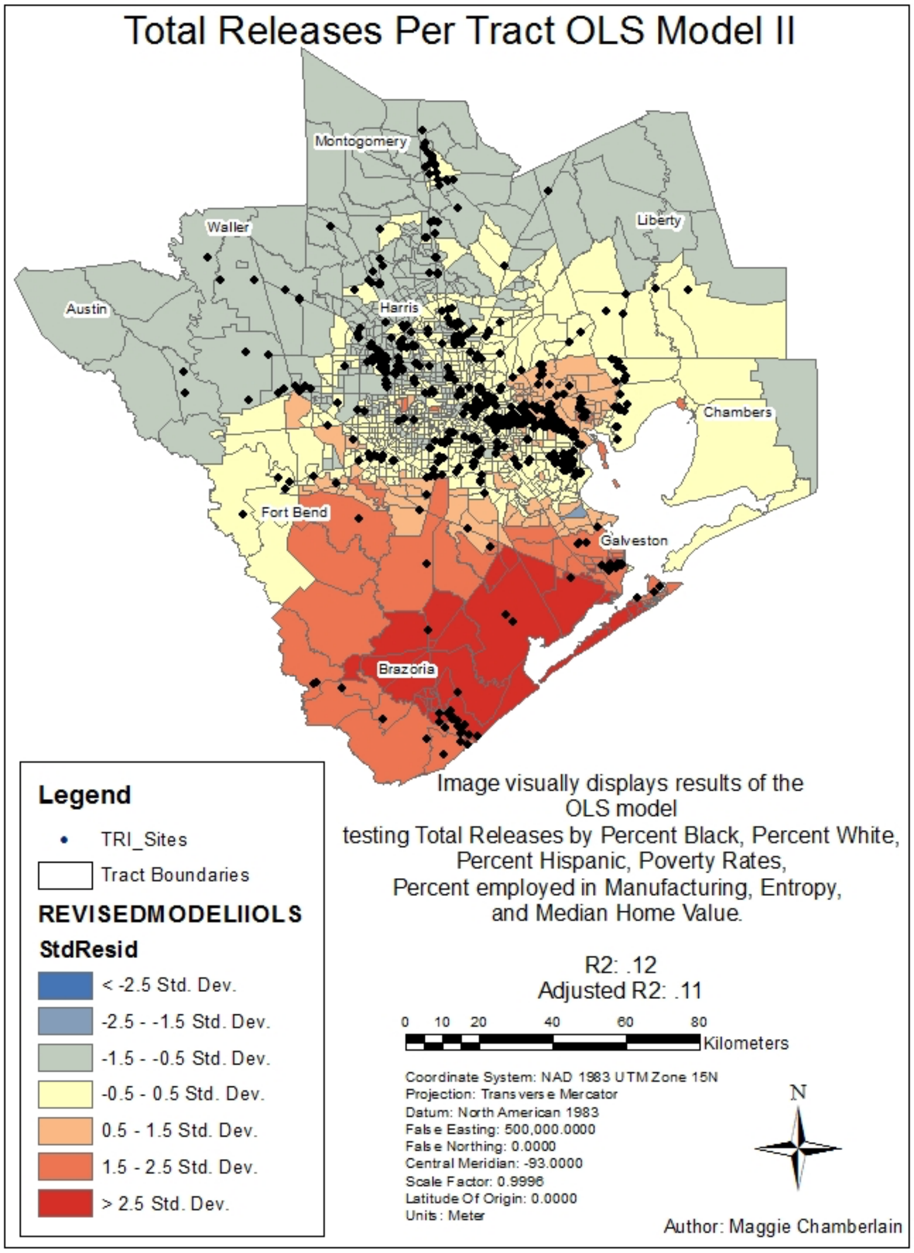


IDW Estimates without Top Two Outliers



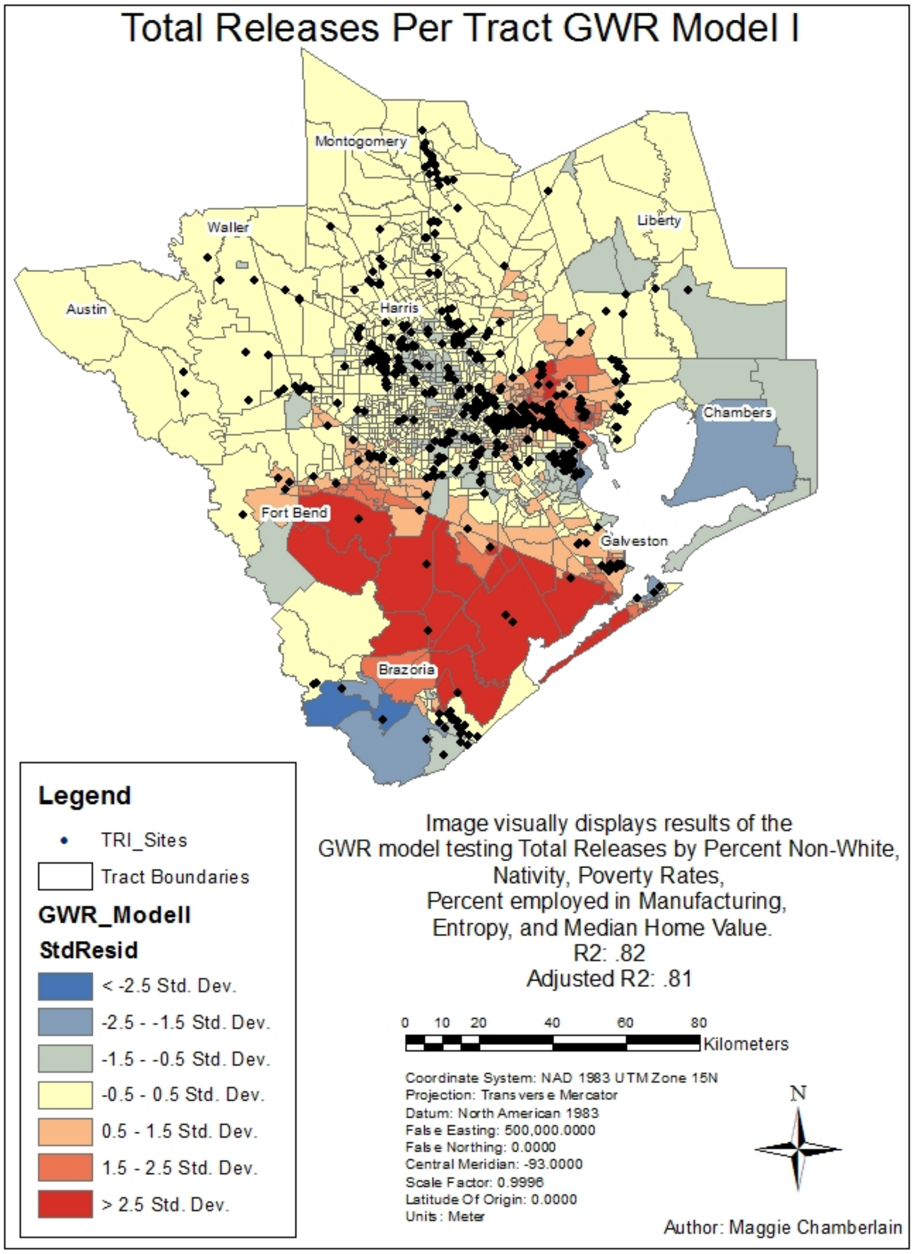
Appendix E: OLS Models

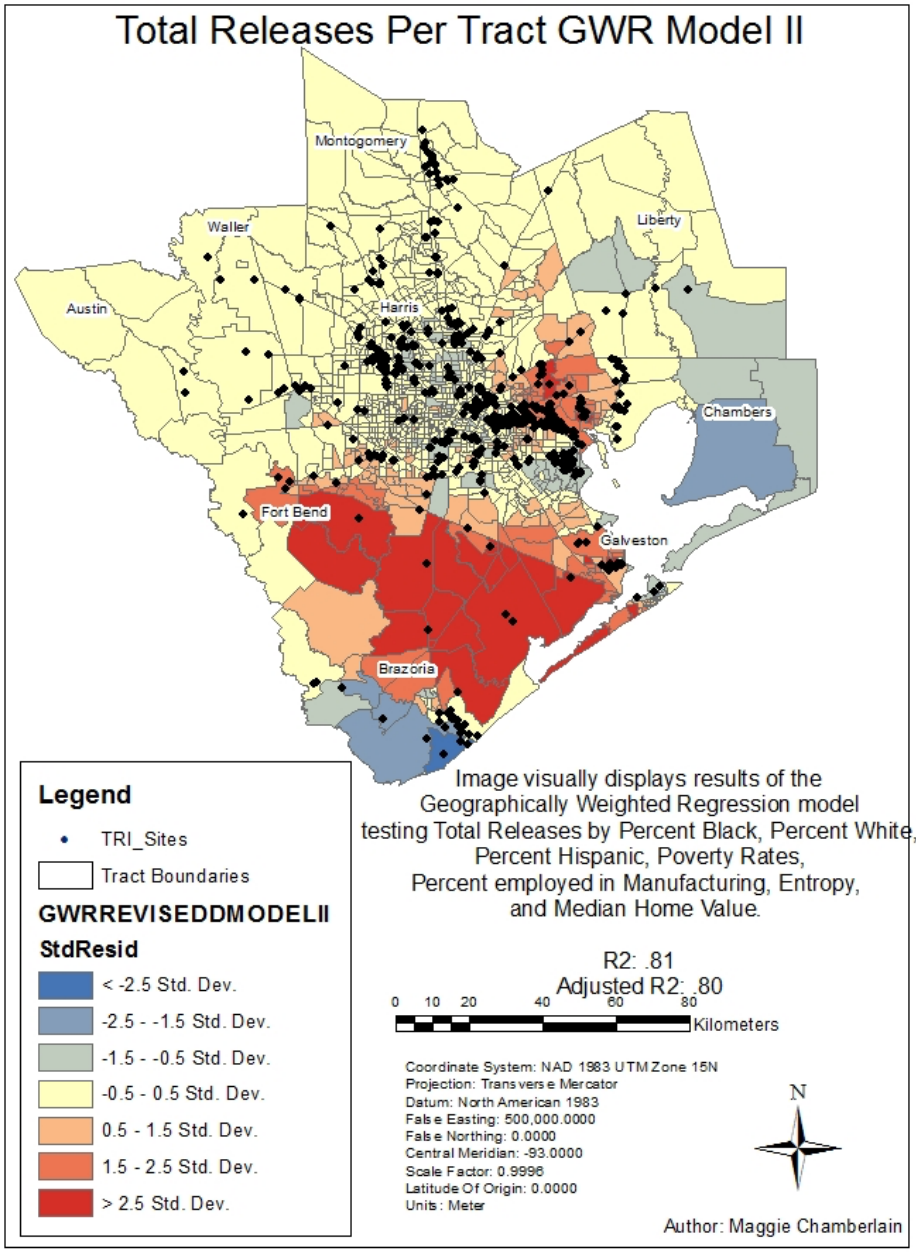




Total Releases Per Tract OLS Model II
Montogomery
Waller
Liberty
Austin
Harris
Chambers
Fort Bend
Galveston
Brazoria
Legend
TRI_Sites
Tract Boundaries
REVISEDMODELIIOLS
StdResid
< -2.5 Std. Dev.
-2.5 - -1.5 Std. Dev.
-1.5 - -0.5 Std. Dev.
-0.5 - 0.5 Std. Dev.
0.5 - 1.5 Std. Dev.
1.5 - 2.5 Std. Dev.
> 2.5 Std. Dev.
Image visually displays results of the
OLS model
testing Total Releases by Percent Black, Percent White,
Percent Hispanic, Poverty Rates,
Percent employed in Manufacturing, Entropy,
and Median Home Value.
R2: .12
Adjusted R2: .11
0 10 20 40 60 80
Kilometers
Coordinate System: NAD 1983 UTM Zone 15N
Projection: Transverse Mercator
Datum: North American 1983
False Easting: 500,000.0000
False Northing: 0.0000
Central Meridian: -93.0000
Scale Factor: 0.9996
Latitude Of Origin: 0.0000
Units: Meter
N
Author: Maggie Chamberlain

Appendix F: GWR Models




Total Releases Per Tract GWR Model II
Montogomery
Waller
Liberty
Austin
Harris
Chambers
Fort Bend
Galveston
Brazoria
Legend
TRI_Sites
Tract Boundaries
GWRREVISEDDMODELII
StdResid
< -2.5 Std. Dev.
-2.5 - -1.5 Std. Dev.
-1.5 - -0.5 Std. Dev.
-0.5 - 0.5 Std. Dev.
0.5 - 1.5 Std. Dev.
1.5 - 2.5 Std. Dev.
> 2.5 Std. Dev.
Image visually displays results of the Geographically Weighted Regression model testing Total Releases by Percent Black, Percent White, Percent Hispanic, Poverty Rates, Percent employed in Manufacturing, Entropy, and Median Home Value.
R2: .81
Adjusted R2: .80
0 10 20 40 60 80
Kilometers
Coordinate System: NAD 1983 UTM Zone 15N
Projection: Transverse Mercator
Datum: North American 1983
False Easting: 500,000.0000
False Northing: 0.0000
Central Meridian: -93.0000
Scale Factor: 0.9996
Latitude Of Origin: 0.0000
Units: Meter
N
Author: Maggie Chamberlain

REFERENCES

Adeola, Francis O. 2000. “Endangered Community, Enduring People: Toxic Contamination, Health and Adaptive Responses in a Local Context”. *Environment and Behavior* 35 (2): 209-249.

Agency for Toxic Substances and Disease Registry. 2005. *Public Health Assessment for PCB Contaminated Sediment in the Lower Fox River and Green Northeastern Wisconsin*. 25.

Anderton, Douglas L.; Anderson, Andy B.; Oakes, John M.; Fraser, Michael R. 1994. “Environmental Equity: The Demographics of Dumping. Demography” (31)229– 48.

Ash, Michael, James K. Boyce, Grace Chang, and Helen Scharber. 2013. “Is Environmental Justice Good For White Folks? Industrial Air Toxics Exposure in Urban America”. *Social Science Quarterly* 94(3): 616-636.

Atlas, Mark. 2007. “TRI to Communicate: Public Knowledge of the Federal Toxic Release Inventory”. *Social Science Quarterly* 88(2): 555-572.

Berry, Lisa M. 2008. “Inequality in the creation of Environmental Harm: Looking for Answers from Within.” *Research in Social Problems and Public Policy* 13: 239-265.

Bevc, Christina A, Brent K. Marshall, and J. Steven Picou. 2005. “Environmental Justice and Toxic Exposure: Toward a Spatial Model of Physcial Health and Psychological Well Being”. *Social Science Research* 36: 48-67.

Brender, Jean D., Juliana A. Maantay, and Jayajit Chakraborty. 2011. “Residential Proximity to Environmental Hazards and Adverse Health Outcomes”. *American Journal of Public Health*. 10(1): 37-52.

Brulle, Robert D., and David Pellow. 2006. “Environmental Justice: Human Health and Environmental Inequalities.” *Annual Review Public Health*. 37: 3.1-3.22.

Bullard, Robert D., Paul Mohai, Robin Saha, and Beverly Wright. 2008. “Toxic Wastes and Race at Twenty: Why Race Still Matters After All These Years.” *Environmental Law* 38 (2): 371-411.

Bullard Robert, Glenn S. Johnson and Angel O. Torres. 2000. S*prawl City: Race, Politics, and Planning in Atlanta.* Washington, DC: Island Press.

Bullard, Robert D. 1990/2000.*Dumping in Dixie. Boulder, CO*: Westview.

---------------------.1983. “Solid Waste Sites and the Black Houston Community.” *Sociological Inquiry* 53(2-3): 273-288.

Burningham, Kate and Diana Thrush. 2003. "Experiencing Environmental Inequality: The Everyday Concerns of Disadvantaged Groups." *Housing Studies* 18(4):517-536.

Chakraborty, Jayajit, Juliana A. Maantay, and Jean D. Brender. 2011. “Disproportionate Proximity to Environmental Health Hazards: Methods, Models, and Measurement”. *American Journal of Public Health* 101 (1): 27-36.

Collins, Mary B. 2011. “Risk Based Targeting: Identifying Disproportionalities in the Sources and Effects of Industrial Pollution.” *American Journal of Public Health*: 101 (S1): 231-237.

Conley, Jamison F. 2011. "Estimation of exposure to toxic releases using spatial interaction modeling". *International Journal of Health Geographics*. 10(20): 1-12.

Cushing, Lara, John Gause, Laruan Meehan August, Rose Cendak, Walker Wieldan, and George Alexeeff. 2015. "Racial/Ethnic Disparities in Cumulative Environmental Health Impacts in California: Evidence From a Statewide Environmental Justice Screening Tool." *American Journal of Public Health* 105 (111): 2341-2348.

Denq, Furjen, Doughlas H. Constance, and Su-Shiow Joung. 2000. "The Role of Class, Statust and Power in the Distribution of Toxic Superfund Sites in Texas and Louisiana". *Journal of Poverty* 4(4): 82-100.

Downey, Liam and Marieke Van Willigen. 2005. "Environmental Stressors: The Mental Health Impacts of Living Near Industrial Activity*". Journal of Health and Social Behavior* 46: 289-305.

Downey, Liam. 2006. "Environmental Inequality in Metropolitan American in 2000". *Sociological Spectrum* 26 (1): 21-41.

-------------------.1998. "Environmental Injustice: Is Race or Income a Better Predictor?" *Social Science Quarterly*. 79 (4): 766-778.

ESRI. 2016. ESRI Support GIS Dictionary. Retrieved November 4th 2016.

ESRI. 2017. ESRI Support GIS Dictionary. Retrieved January 17th 2017

Faber, Daniel and Eric J. Krieg. 2001.*Unequal Exposure to Ecological Hazards: Environmental Injustices in the Commonwealth of Massachusetts*. Boston: Northeast University.

Freudenburg, William R. 2005. "Privileged Access, Privileged Accounts: Toward a Socially Structured Theory of Resources and Discourses." *Social Forces* 81 (1): 89-114.

Fricker, Ronald D., and Nicolas W. Hengartner. 2001. "Environmental Equity and the distribution of toxic release inventory and other environmentally undesirable sites in metropolitan New York City". *Environment and Ecological Statistics* 8: 32-52.

Fleming, India, Mary K. O'Keeffe, and Andrew Baum. 1991. "Chronic Stress and Toxic Waste: The Role of Uncertainty and Helplessness." *Journal of Applied Social Psychology* 21(23): 1889-1907.

Grant, Don, Mary Nell Trautner, Liam Downy, and Lisa Thiebaud. 2010. "Bringing the Polluters Back in: Environmental Inequality and the Organization of Chemical Production". *American Sociological Review* 75 (4): 479-504.

Gilbert, Angela and Jayajit Chakratory. 2011. "Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida". *Social Science Research* 40: 273-286.

Hipp, John R. and Cynthia M. Lakon. 2010. "Social disparities in health:Disproportionate toxic proximity in minority communities over a decade". *Health and Place* 16 (2010): 674-683.

Hird, John A., and Michael Reese. 1998. "The Distribution of Environmental Quality: An Empirical Analysis". *Social Science Quarterly* 79 (4): 693-716.

Iceland, John. 2004. "Beyond Black and White: Residential Segregation in Multiethnic America." *Social Science Research* 33 (2): 248-271.

Johnson, Glen S., Steven C Washington, Denae W Kind, Jamila M. Gomez. 2014. "Air Quality and Health Issues along Houston's Ship Channel: An Exploratory Environmental Justice Analysis of a Vulnerable Community (Pleasantville)." *Race, Gender & Class,* 21(3), 273-303.

Jones, Miranda R., Ana V. DIez-Roux, Anjum Jajat, Kiarri N. Kershaw, Marie S. O'Neill, Eliseo Guallar, Wendy S. Post, Joel Kaufman, and Ana Navas-Acien. 2014. "Race/Ethnicity, Residential Segregation and Exposure to Ambient Air Pollution: The Multi-Ethnic Study of Atherosclerosis (MESA)." *Research and Practice* 104 (11): 2130-2137.

Johnston, Jill E., Emily Werder, and Daniel Sebastian. "Wastewater Disposal Wells, Fracking and Environmental Injustice in Southern Texas". *AJPH Research*. 106 (3): 550-556.

Krieg, Eric J. 2005. "Race and Environmental Justice in Buffalo, NY: A Zip Code and Historical Analysis of Ecological Hazards*". Society and Natural Resources* 18: 199-213.

Lejano, Raul P. and Hiroyuki Iseki. 2001. "Environmental Justice: Spatial Distribution of Hazardous Waste Treatment, Storage, and Disposal Facilities in Los Angelos*".* *Journal of Urban Planning and Developmen*t. 127 (2): 51-62.

Lersch, Kim M. and Timothy C. Hart. 2014. "Environmental Justice, Lead, and Crime: Exploring the Spatial Distribution and Impact of Industrial Facilities in Hillsborough County, Florida". *Sociological Spectrum* 31(1): 1-24.

Massey, Douglas S. and Nancy A.1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

---------.1988. "The Dimensions of Residential Segregation." *Social Forces* 67: 281-315.

McGurty, Eileen. 2007. *Transforming Environmentalism: Warren County, PCBs, and the Origins of Environmental Justice*. New Brunswick: Rutgers University Press.

Melosi, Martin V.1981 . *Garbage in the cities: refuse, reform, and the environment: 1880-1980.* College Station, TX : Texas A&M University Press.

Mennis, Jeremy. 2002. "Using Geographic Information Systems to Create and Analyze Statistical Surfaces of Population and Risk for Environmental Justice Analysis". *Social Science Quarterly* 83(1): 281-297.

Mohai, Paul and Robin Saha. 2007. "Racial Inequality in the Distribution of Hazardous Waste: A National-Level Reassessment". *Social Problems* 54 (3): 343-370.

---------. 2006. "Reassessing Racial and Socioeconomic Disparities in Environmental Justice Research". *Demography* 43(2): 383-399.

---------.2015. "Which came first, people or pollution? Assessing the disparate siting and post siting demographic change hypotheses of environmental injustice". *Environmental Research Letters* 10: 1-18.

Mohai, Paul, Paula M. Lantz, Jeffrey Morenoff, James S. House, and Richard P. Mero. 2009. "Racial and Socioeconomic Disparities in Residential Proximity to Polluting Industrial Facilities: Evidence From the Americans' Changing Lives Study". *American Journal of Public Health* 99 (3): 649-656.

Morello-Frosch, Rachel A. 2002. "Discrimination and the political economy of environmental inequality." *Environment and Planning C: Government and* Policy 20: 477-496

Natural Resources Defense Council. 2004. *Hidden Danger: Environmental Health Threats in the Latino Community*. New York: NRDC.

Paigen, Beverly, Lynn R. Goldman, Mary M. Magnant, Jospeh H. Highland, A.T. Steegman and JR. 1987. "Growth of Children Living Near the Hazardous Waste Site, Love Canal." *Human Biology* 59 (3): 489-508

Pastor, Manuel Jr., Jim Sadd and John Hipp. 2001. "Which Came First? Toxic Facilities, Minority Move-In, and Environmental Justice". *Journal of Urban Affairs* 23 (1): 1-23.

Pastor, Manuel Jr., James L. Sadd, and Rachel Morello-Frosch. 2004. "Waiting to Inhale: The Demographics of Toxic Air Release Facilities in 21st-Century California". *Social Science Quarterly* 85(2): 420-440.

Peek, MK., MP Cutchin, D Freeman, RP Stowe, and JS Goodwin. 2009. "Environmental Hazards and Stress: Evidence from the Texas City Stress and Health Study." *Journal of Epidemiological Community Health* 63 (10): 1-16.

Pellow, David. 2000. "Environmental Inequality Formation." *American Behavioral Scientist.* 43 (4): 581-601.

Pine, John C., Brian D. Marx, and Aruna Lakshmanan. 2002. "An Examination of Accidental-Release Scenarios from Chemical-Processing Sites: The Relation of Race to Distance." *Social Science Quarterly* 83(1): 317-331.

Reardon, Sean F. and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation." *Sociological Methodology* 32 (1): 33-67.

Ringquist, Evan J. 2005. "Assessing Evidence of Environmental Inequalities: A Meta Analysis." *Journal of Policy Analysis and Management*. 24 (2): 223-247.

--------- 1997. "Equity and the Distribution of Environmental Risk: The Case of TRI Facilities." *Social Science Quarterly*. 78 (4): 811-829.

Satterfield, Terre A., C.K. Martz, and Paul Slovic. 2004. "Discrimination, Vulnerability, and Justice in the Face of Risk." *Risk Analysis* 24(1): 115-129.

Sicotte, Diane. 2014. "Diversity and Intersectionality among Environmentally Burdened Communities in the Philadelphia Metropolitan Area, USA." *Urban Studies*. 51(9): 1850- 1870.

Smith, Chad. 2007. "Economic Deprivation and Environmental Inequality in Postindustrial Detroit: A Comparison of Landfill and Superfund Site Locations." *Organization & Environment*. 20 (1): 25-43.

----------. 2009. "Economic deprivation and racial segregation: Comparing Superfund sites in Portland, Oregon and Detroit, Michigan." *Social Science Research*. 38: 681-691.

United Church of Christ Commission for Racial Justice. 1987. *Toxic Wastes and Race in the United States: A National Report on the Racial and Socio-economic Characteristics of Communities with Hazardous Waste Sites*. New York: United Church of Christ.

United States Environmental Protection Agency. 2016. *U.S Environmental Protection Agency: Environmental Justice*. Retrieved November 4th, 2016.

Vorapracha, Phatarapon, Pongtep Phonprasert, Suparada Khanaruksombat, and Nuchananaporn Pijarm. 2015. "A Comparison of Spatial Interpolation Methods for prediction concentration of Particle Pollution." *International Journal of Chemical, Environmental, and Biological Sciences*, 3(4): 302-606.

Wong, David W.S. and Shih-Lung Shaw. 2011. "Measuring Segregation: an activity space approach." *Journal of Geographical Systems*. 13(2): 127-145.
Wu, Shanshan and Matthew T Heberling. 2013. "The distribution of pollution and environmental justice in Puerto Rico: a quantitative analysis*." Population and Environment*. 35 (2): 113-132.