CAN ANALYZING SPATIAL RELATIONSHIPS THROUGH GEOGRAPHICALLY WEIGHTED REGRESSION IMPROVE OUR UNDERSTANDING OF LOW SCHOOL ATTAINMENT? A GIS-BASED ANALYSIS OF CENSUS AND ACS DATA

William England
University of Nebraska-Lincoln, wengland@hotmail.com

Follow this and additional works at: http://digitalcommons.unl.edu/cehsdiss

Part of the Disability and Equity in Education Commons, and the Educational Assessment, Evaluation, and Research Commons

http://digitalcommons.unl.edu/cehsdiss/225

This Article is brought to you for free and open access by the Education and Human Sciences, College of (CEHS) at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Public Access Theses and Dissertations from the College of Education and Human Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
CAN ANALYZING SPATIAL RELATIONSHIPS THROUGH GEOGRAPHICALLY WEIGHTED REGRESSION IMPROVE OUR UNDERSTANDING OF LOW SCHOOL ATTAINMENT? A GIS-BASED ANALYSIS OF CENSUS AND ACS DATA

By

William R. England, Ph.D.

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Philosophy

Major: Educational Studies

Under the Supervision of Professor Edmund T. Hamann

Lincoln, Nebraska
October, 2014
CAN ANALYZING SPATIAL RELATIONSHIPS THROUGH GEOGRAPHICALLY WEIGHTED REGRESSION IMPROVE OUR UNDERSTANDING OF LOW SCHOOL ATTAINMENT? A GIS-BASED ANALYSIS OF CENSUS AND ACS DATA

William R. England
University of Nebraska, 2014

Adviser: Edmund T. Hamann

In this study I present a relatively new technique for analyzing a recurring problem in our communities. Using a set of innovative and relatively new modeling methods, I demonstrate ways in which it is possible to directly account for, capture, and visualize the spatial variability in the relationships between U.S. Census data from 1990 and the recent low-school-attainment landscape in both the Omaha and Lincoln Public School (OPS) districts in Omaha and Lincoln, NE. Low school attainment in adults is a correlate of a host of troubling health and economic factors, which, in turn, have an impact on a child's school performance and eventual school attainment. Disrupting this trend is (and has been) the focus of much research because not only is low school attainment predictive of a host of concerning variables, school attainment also has a tendency to persist from generation to generation. In addition, areas of an urban environment characterized by low school attainment seem to remain geographically stable over long periods of time. However, traditionally, researchers modeling the relationships associated with school attainment draw conclusions based on techniques that rely on global inferences (e.g., ordinary least squares regression). Where there is spatial nonstationarity in the
coefficients produced by a regression analysis, researchers using these global techniques may miss important local caveats in their predictions. When fully analyzed, these caveats can help to create better statistical models that might help to focus community resources and public policies in more effective ways.
ACKNOWLEDGEMENTS

“Buy the ticket, take the ride.”
Hunter S. Thompson, Fear and Loathing in Las Vegas, p.89

I did not set out to study GIScience, spatial analysis and spatial statistics, geodemographics, or geographic research methods when I began my Ph.D. program. In fact, when I returned from Manzanares el Real, Spain after teaching English as a Foreign Language there for a year, I decided to begin my Ph.D. work with Dr. Ted Hamann in part because I wanted to absorb some of his expertise in educational policy and school reform generally, and in part because he was conducting research about transnational student experiences in the U.S. and Mexico specifically. I am thankful to Dr. Hamann, as he has guided me while giving me room to explore my evolving intellectual trajectory as a Ph.D. student, and I am particularly grateful for his advocacy as my research extended to the work in this dissertation. Thank you, Dr. Hamann, for teaching me to imagine how education and schooling can be at once spatial, geographical, scalar, contextual, and relational in nature, and how GIS can (and cannot) help to illuminate these dimensions of schooling and education.

I also thank Dr. Guy Trainin for consistently leaving his office door open. I benefitted enormously from impromptu chats with Dr. Trainin. His advice, wisdom, sense of humor, and most importantly, his seemingly boundless knowledge of statistical modeling and how best to represent and visualize data will always be invaluable to me.

Also, I must thank Dr. J. Clark Archer. His courses are simply awesome, and had I not taken that first class from him, Political Geography (and the other two as well—Geodemographics and Urban Geography), my work would be miles away from where it
is now. Dr. Archer has been an outstanding mentor and I have tremendous respect for his depth of insight and knowledge in all things mapping and statistics.

And, I am grateful to Dr. Karl Hostetler. He is truly an intellectual. In his courses I learned how to slow down and really think through a topic, and his advice about my writing was spot on, and it has stayed constantly with me. Thank you, Dr. Hostetler.

I also wish to acknowledge Drs. Margret Latta and Miguel Carranza, who were early members of my committee. Unfortunately for me, Dr. Latta took a position at the University of British Columbia and Dr. Carranza left for the University of Missouri at Kansas City, but both challenged me to think about education from angles and perspectives to which I was not accustomed. Dr. Latta was particularly constructive in helping me to formulate responses to two of my comprehensive exam questions.

I am also indebted to Dr. Charles Ansorge, who has taught statistics to two generations of my family (me and early in his career, my mother). It was unexpected that I, an English major, would find the world of statistics so captivating. Dr. Ansorge’s excellent teaching and brilliantly designed courses have everything to do with my interest in quantitative analysis and methods.

I am also grateful for Dr. Loukia Sarroub’s support of my research interests and ideas, and I thank the Department of Teaching, Learning, and Teacher Education for providing funding and giving me an array of interesting teaching experiences. Thanks to the TLTE staff, especially Jess Hustad and Shari Daehling—the administrative side of this process has been smooth because of you both. Also, thanks to the men and women at the U.S. Census Bureau, and the amazing people behind ArcGIS at ESRI. I could not have done any of this work without your efforts.
Finally, there aren’t sufficient words to express how grateful I am for my parents, Kathy England, Thomas England, and my step-mother Patty, as well as my grandfather Richard Ryan, and my late grandmother Kathryn Ryan. This amazing core of support has provided me encouragement, endless patience, and excellent advice every time I’ve needed it, and they have been the foundational strength upon which all of my successes rest.
# TABLE OF CONTENTS

INTRODUCTION ........................................................................................................ ii

ACKNOWLEDGEMENTS ............................................................................................... iv

LIST OF FIGURES ........................................................................................................ xi

LIST OF TABLES ........................................................................................................... xiii

CHAPTER ONE: BACKGROUND ..................................................................................... 14

1.1 Introduction ............................................................................................................. 14

1.2 Statement of the Problem ....................................................................................... 18

1.3 Purpose of the Study ............................................................................................... 24

1.4 Study Questions ....................................................................................................... 25

1.5 Theoretical Framework ......................................................................................... 25

1.6 Importance of the Study ....................................................................................... 33

1.7 Definition of Terms ............................................................................................... 35

1.8 Summary ................................................................................................................ 38

CHAPTER TWO: OVERVIEW OF GIS AND GWR ....................................................... 39

2.1 Introduction ............................................................................................................. 39

2.2 Geographic Methods and the Rise of Geographic Information System ............... 40

2.3 Examples of Geographically Weighted Regression Techniques ....................... 48

2.4 Exampled of GWR in Education Research ......................................................... 52

2.4.1 Math Performance in the U.K ........................................................................... 52

2.4.2 School Finance in Missouri ............................................................................. 55

2.4.3 ACT Score in Missouri ................................................................................... 59

2.5 Summary ................................................................................................................ 61

CHAPTER THREE: DESCRIPTION OF STUDY AREAS ......................................... 63
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Omaha, NE</td>
<td>63</td>
</tr>
<tr>
<td>3.2 Lincoln, NE</td>
<td>74</td>
</tr>
<tr>
<td><strong>CHAPTER FOUR: DATA AND METHODS</strong></td>
<td>81</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>81</td>
</tr>
<tr>
<td>4.2 Rates vs. Raw Numbers</td>
<td>81</td>
</tr>
<tr>
<td>4.3 Data Collection</td>
<td>83</td>
</tr>
<tr>
<td>4.3.1 Dependent Variable: Low School Attainment</td>
<td>85</td>
</tr>
<tr>
<td>4.3.2 Predictor Variables</td>
<td>87</td>
</tr>
<tr>
<td>4.4 Research Design</td>
<td>90</td>
</tr>
<tr>
<td>4.4.1 Analysis of Spatial Dependence</td>
<td>90</td>
</tr>
<tr>
<td>4.4.2 Incremental Spatial Autocorrelation</td>
<td>95</td>
</tr>
<tr>
<td>4.4.3 Hotspot Analysis</td>
<td>97</td>
</tr>
<tr>
<td>4.4.3 Exploratory Regression</td>
<td>99</td>
</tr>
<tr>
<td>4.4.5 Ordinary Least Squares Calibration</td>
<td>104</td>
</tr>
<tr>
<td>4.4.6 Geographically Weighted Regression Model</td>
<td>106</td>
</tr>
<tr>
<td>4.4.7 Omaha Model Application to Lincoln Public Schools</td>
<td>109</td>
</tr>
<tr>
<td>4.5 Limitations of the Study</td>
<td>109</td>
</tr>
<tr>
<td>4.6 Summary</td>
<td>111</td>
</tr>
<tr>
<td><strong>CHAPTER FIVE: RESULTS</strong></td>
<td>114</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>114</td>
</tr>
<tr>
<td>5.2 The OPS Low-School-Attainment Landscape</td>
<td>114</td>
</tr>
<tr>
<td>5.3 Analysis of Exploratory Results for OPS</td>
<td>118</td>
</tr>
<tr>
<td>5.4 Omaha Public Schools-OLS Model Results</td>
<td>123</td>
</tr>
<tr>
<td>5.5 Omaha Public Schools-GWR Results</td>
<td>127</td>
</tr>
<tr>
<td>5.5.1 Considerations for the Visualizing of Spatial Nonstationarity</td>
<td>127</td>
</tr>
</tbody>
</table>
5.6 Performance of the OPS-OLS and GWR Models................................ 141
5.7 ‘Interurban Portability’ of the OPS-OLS Model.......................... 144
5.8 The Lincoln Public Schools Low-School-Attainment Landscape...... 147
5.9 Lincoln Public Schools-Exploratory and OLS Results.................... 149
5.10 Lincoln Public Schools-GWR Results........................................ 152
5.11 Performance of the LPS-OLS and GWR Models......................... 158
5.12 Summary.................................................................................. 159

CHAPTER SIX: DISCUSSION AND CONCLUSIONS............................... 164
6.1 Introduction.............................................................................. 164
6.2 Predicting the Present............................................................ 165
6.3 Study Questions....................................................................... 171
   6.3.1 Response to Study Question One........................................ 171
   6.3.2 Response to Study Question Two........................................ 175
   6.3.3 Response to Study Question Three...................................... 178
      6.3.3.1.................................................................................. 179
      6.3.3.2.................................................................................. 183
      6.3.3.3.................................................................................. 187
      6.3.3.4.................................................................................. 192
      6.3.3.5.................................................................................. 193
      6.3.3.6.................................................................................. 194
      6.3.3.7.................................................................................. 197
      6.3.3.8.................................................................................. 198
6.4 What Policy Implications Arise from the Presence of Spatial Nonstationarity?......................................................... 203
6.5 Summary of Dissertation .......................................................... 205
LIST OF FIGURES

Figure 1.1: Percentage of Low School Attainment by U.S. County: Low School Attainment Hot/Coldspots—ACS 5-year Estimates (2008 – 2012)

Figure 3.1: Reference Map: Relative Locations of Areas of Study

Figure 3.2: Omaha Public Schools, City Limits, and Surrounding Districts

Figure 3.3: Overview of Omaha Growth and Demographics: 1860 to 2012

Figure 3.4: Key Demographic and Economic Features of the Omaha Public Schools District

Figure 3.4: Lincoln Public Schools, City Limits, and Surrounding Districts

Figure 3.6: Key Demographic and Economic Features of the Lincoln Public Schools District

Figure 3.7: OPS and LPS High Schools: Percent nonwhite as a predictor of graduation rates

Figure 4.1: Hypothetical Census Tract Scenarios: Rate vs. Raw Numbers

Figure 4.2: Spatial autocorrelation: Spatial dispersion, independence, and clustering

Figure 4.3: Incremental Spatial Autocorrelation Graph

Figure 5.1: OPS Low School Attainment Hotspots

Figure 5.2: Spatial Analysis of OPS-OLS Model Residua

Figure 5.3: Low school attainment in 1990 as a predictor of current low school attainment

Figure 5.4: Hispanic/Latino households in 1990 as a predictor of current low school attainment

Figure 5.5: Houses 30-years-old or older in 1990 as a predictor of current low school attainment

Figure 5.6: Distance to I-80 as a predictor of current low school attainment

Figure 5.7: LPS Low School Attainment Hotspots

Figure 5.8: People 25 years-old or older in 1990 as a Predictor of Current Low School Attainment
Figure 5.9: Hispanics/Latinos in Poverty in 1990 as a Predictor of Current Low School Attainment

Figure 5.10: Houses 30 years-old or Older in 1990 as a Predictor of Current Low School Attainment

Figure 5.11: Distance from I-80 as a Predictor of Current Low School Attainment

Figure 6.1: U.S. College Enrollment: 1982 – 1995

Figure 6.2: The Evolution of Nebraska’s Major East/West Transportation Network

Figure 6.3: NE Cities and Towns along I-80: Population Size Relative to Distance from Omaha

Figure 6.4: NE Cities and Towns not along I-80: Population Size Relative to Distance from Omaha
LIST OF TABLES

Table 4.1 Definitions of independent variable

Table 5.1: Descriptive Statistics for Candidate and Dependent Variables

Table 5.2: Pairwise Pearson Correlation Coefficients for Twelve Candidate Variables in OPS

Table 5.3: OPS-OLS Regression Coefficient Statistics

Table 5.4: Summary of GWR vs. OLS Performance in OPS

Table 5.5: OPS-LPS Model Regression Coefficient Statistics

Table 5.6: LPS Model Regression Coefficient Statistics

Table 5.7: Pairwise Pearson Correlation Coefficients for Twelve Candidate Variables in OPS

Table 5.8: Summary of GWR vs. OLS Performance in LPS

Table 6.2: Low School Attainment Model Comparison

Table 6.3: Model Comparison: With/Without Spatial Factor DistI80
CHAPTER ONE: BACKGROUND

1.1 Introduction

The American Community Survey (ACS) estimates that 43% of adults in the U.S. ages 25 and over have a high school diploma, equivalent, or less (U.S. Census Bureau, 2012a). The ACS estimates that another 29% of the U.S. population ages 25 and over tried college but quit before earning a degree. If low school attainment is defined as the proportion of U.S. adults ages 25 and over who have never attempted college (i.e., those who have earned only a high school diploma, an equivalent, or less) then the population of adults 25 and over in the U.S. with low school attainment is approximately 87 million (U.S. Census Bureau, 2012a). If this definition is broadened to include the proportion of the adult population 25 years and older without a post-secondary degree (i.e., including those who tried college but didn’t earn an associate’s or bachelor’s degree) then the population of adults in the U.S. who would be described as having low school attainment is roughly 130 million (U.S. Census Bureau, 2012a).

McLloyd (1989) reviewed the range of some of the outcomes associated with low school attainment and found that low levels of schooling in adults are tangled up with persistent patterns of poverty and low socioeconomic status, low school attainment and low school achievement in children, as well as psychological, emotional, and behavioral problems both for adults with low educational status and for their offspring. In accordance with these findings, social reproduction theorists have long advanced the idea that parental education is an important marker of socioeconomic status and that the current cultural and school attainment landscape is a recapitulation of the past that will tend to be reproduced in the future by unequal social structures in a particular society.
This explains why unequal education outcomes tend to persist from parent to child (Bourdieu & Passeron, 1977; Lamont & Lareau, 1988).

The idea that school attainment is reiterated in this manner can be traced back even further to John Stuart Mill, to whom Spiegelberg (1961) suggested the phrase “accident of birth” is mostly closely attributable. For Mill, “accident of birth” refers to the sum of those natural and social factors and circumstances which tend to either limit or advantage a person based upon where and to whom they are born. Per this framework, where a person is born and to what family influences his life, including where he will go to school. Where a child’s parents move and when can likewise impact what a child receives educationally and to what consequence. Further, because not all schools and/or school districts have equal resources, not all schools produce equal educational outcomes; thus, where one attends school matters (Borman and Dowling, 2010; Brown v. Board, 1954; Kozol, 1991). For example, if some schools are highly successful in sending students to college while others are “dropout factories” (Orfield, 2009) in which 60% or fewer complete high school and even fewer go to college, then it may be the case that any given student who attends a ‘dropout factory’ also faces the increasing probability that she will not finish high school, much less go on to college.

The fact that where a child attends school has consequences for his trajectory in life means that schooling outcomes are inherently geographical in nature. Because school attainment is such a strong correlate of many health and economic factors (Entwisle & Alexander, 1993; Lee & Burkham, 2002; Rouse, Fantuzzo, & LeBeouf, 2011) and because low school attainment appears to be something that in too many cases continues
from parent to child means that the persistence of low school attainment and the factors
associated with it are also grounded in geography.

It is worth pointing out that low school attainment in and of itself may not be
inherently problematic. Occupational, psychological, familial, health, and financial well-
being are all possible irrespective of one’s level of schooling. However, because social
mobility in the U.S. is so often bound up with one’s formal education, on average,
persisting in school is better than the alternative. And since McLloyd’s (1989) meta-
analysis detailing the troubling range of relationships between the negative factors
mentioned above and low school attainment, even more trouble has emerged.
Specifically, Stern, et al. (1994) found that Alzheimer’s disease appears to be have a
higher prevalence and tends to emerge earlier in those people with low school attainment.
Previous research partially supports this finding because Stern et al. (1992) found a
relationship between school attainment and blood flow to certain regions of the brain in
patients with Alzheimer’s disease. While the authors acknowledged that school
attainment may be a proxy for some other set of risk factors, they also found increases in
synaptic density in certain areas of the brain in those patients who held advanced degrees.
The implication is that getting more people to attain higher levels of formal education
could benefit those unfortunate enough to be afflicted with Alzheimer’s disease,
especially by staving off or at least delaying the expression of symptoms.

A more recent study by van den Berg and colleagues (2013) found that low
maternal education was associated with lower vitamin D levels and that women in the
lowest quartile for levels of vitamin D had a significantly higher chance of having
offspring with low birth weight. The authors found that low birth weight was associated
with a host of health problems, including intellectual impairment, and thus the authors suggested that increasing vitamin D intake for mothers with low school attainment may have a host of benefits for their offspring.

Magnuson et al. (2009) found that if mothers returned to school they tended to raise the educational trajectory of their children in the process. According to the authors, as mothers with low school attainment sought advancement in their own formal educations, there was a corresponding improvement in the home environment in general and in their children’s language skills.

The benefits of increases in school attainment to health and economic factors are not limited to individual parent-child outcomes either. Increasing school attainment appears to have broader community-wide impacts as well. Trostel (2010a and 2010b) found that college graduates earn more money, pay more in taxes, and use less local, state, and federal resources on average than those without a college education. Conversely, Rud et al. (2013) found that the relationship between crime and education may have substantial long-term costs above and beyond the immediate costs of crime to the community because the children of parents involved in criminal activity are less likely to continue to higher education. Thus, by raising adult education levels, we may not only save money in the short term by reducing crime rates, but budgetary windfalls may be possible down the road as more would-be prisoners would theoretically attend college (thus, using less publicly provided services and contributing more to the tax base) instead of committing crime. And finally, Hanushek (2013) even found strong linkages between school quality and more macro-level trends like economic growth in developing countries.
In sum, Mill’s arguments regarding the accident of one’s birth considered along with Bourdieu and Passeron’s (1977) notion of habitus, together predict that children, typically through no action of their own, are launched along a formal educational trajectory before they even enter school. These arguments are supported by a raft of more recent social science literature, which collectively suggest that not only is low school attainment reproduced and passed on from generation to generation, but also that low school attainment has myriad health and economic consequences for the individual, the community, and the nation.

1.2 Statement of the Problem

Of course, arguments for increasing school attainment are fairly ubiquitous and have been for so long that making such an argument here seems almost quaint. In addition, statistical analyses attempting to uncover the correlates and consequences of low school attainment are nothing new either. What is proposed as being novel in this study are the methods by which this seemingly intractable problem gets analyzed. Generally speaking, the techniques used to model linkages between one or several variables and school attainment (as well as models which use school attainment as a predictor of some other variable[s] of interest) often make inferences based upon “global methods” (Brundson et al., 1998; Lercsh & Hart, 2014; Pasculli, et al., 2014) like ordinary least squares (OLS) regression. For OLS, a global inference simply refers to the fact that conclusions drawn from OLS procedures reference fixed and single values describing the significance of regression coefficient(s), goodness-of-fit, and the like for an entire set of data. In geographic terms, users of OLS must assume that any
relationships detected by the technique are spatially homogeneous (Brunson, et al., 1998; Chi, et al., 2013). In less jargony terms, if a researcher using OLS detects a relationship between low SES and low school attainment across City X, the inherent and automatic assumption is that there are no variations in the strength of this relationship across the entire area of study.

The veracity of the previous claim can be examined by analyzing the OLS regression equation. Briefly, recall that OLS regression models the relationship between one or several predictor variables and a dependent variable with the formula given below (Equation 1):

\[ y_i = \beta_0 + \beta_1 x_i + \epsilon_i \]  

(1)

where \( y_i \) is a response variable that is dependent on the sum of \( \beta_0 \) (the intercept of a straight line), \( \beta_1 \) (the regression coefficient, or slope of a straight line)—which is multiplied by an independent \( x_i \) value—and \( \epsilon_i \) (which denotes a random error term). Note that there is no explicit term in this equation that corresponds to the locations of the dependent/independent variables, and there is no accounting for locations/spatial variations in the strength of the relationships being modeled. Returning to the City X example from above, the outputs from this regression equation would necessarily describe the strength of the relationship between low income and low school attainment as perfectly consistent across the entire city.

As an example of why this can be problematic, consider low school attainment by county across the US. If we let low school attainment be defined as those people 25 years
and older whose highest level of schooling is a high school diploma, equivalent, or less (in other words, those people at least 25 years-old who have never attempted college) it is potentially quite useful to know that in the average U.S. county 51% of the inhabitants 25 years or older qualify as having low school attainment. This average is an example of a global statistic. Useful as this information can be, it can also be highly unrepresentative of local variability in low school attainment levels. For example, in Koochiching County, MN (located near International Falls along the US/Canadian border) only 13% of those 25 and older have never been to college. Likewise, the 51% average conceals the fact that in Lipscomb, County, TX, (just south of the Oklahoma panhandle on the Texas/Oklahoma boarder) 80% of the population 25 years and over stopped their schooling at or before high school.

The concept of an average is probably familiar enough to most people that when it is encountered in media articles, books, journals, etc., it is understood that the number is meant to serve as an approximation of the most typical value in the dataset of interest. But more complicated global statistics like the coefficient of determination ($R^2$) produced by an OLS model—meant to describe the changes in a dependent variable which can be explained by a set of regressors—may be less well understood as a value, which (like the average) condenses many observations into one number in order to describe a relationship for an entire dataset (or area of study). The $R^2$ statistic, regression coefficients, and the like may be familiar to many scholars, but that these are global statistics that potentially obscure local variations in a dataset may be less well-known. In any event, when OLS is used (especially when applied to any data that occurs on, under, or near the earth’s surface—which is to say, geographic data) spatial variability is not
allowed to play a role in the modeling of relationships between the phenomena being investigated. This is a potentially serious problem if/when the OLS assumption of spatial homogeneity is violated. Brundson, et al., (1996) referred to spatial variability in the strength of the relationships between a response variable and its regressors as spatial nonstationarity. When spatial nonstationarity is present, the traditional OLS model will not reflect the underlying structure of the data, and thus it will be less likely to accurately explain the relationships among the variables being examined.

To scrutinize this problem further, let’s return to the example of school attainment by county across the U.S. Figure 1.1 (below) displays a map of all the U.S. counties color-coded along a red-beige-blue spectrum to reflect a county’s low school attainment in relation to its neighboring counties and the U.S. average of 51%. To make this map, a ‘hotspot analysis’ technique (described in detail in the methodology chapter to come) was used to discover where local clusters of counties with high/low proportions of low school attainment are statistically significant. In order to be a hot/coldspot it isn’t enough for a county by itself to have high/low attainment—in order to be statistically significant, it needs to be above/below the average and in a neighborhood of counties with high/low school attainment. The red-beige-blue arrangement is meant to represent a ‘hot-neutral-cold’ conceptualization so that increasing redness depicts counties increasingly above the U.S. average for low school attainment (beginning at the 90% confidence level), beigeness depicts counties close to the U.S. average, and increasing blueness represents counties increasingly below the U.S. average (again beginning at the 90% confidence level).
From this map we can see that the ‘Rust Belt’ emerges, as well as parts of Appalachia and the Deep South, as areas with significant clustering of counties characterized by low school attainment. There are many things to say about this map, but one of the most important implications for the point I am making here is the fairly clear presence of spatial dependence and spatial variability inherent in the patterns of low school attainment that can be observed in Figure 1.1 above.

---

1 The ‘Rust Belt’ appears to be defined differently by different scholars, but generally speaking, this region refers primarily to states in the Upper Midwest and Mid-Atlantic regions of the U.S.—including Pennsylvania, Michigan, Ohio, Illinois, and Indiana— which experienced dramatic decreases in manufacturing in the last half to last quarter of the 20th century (Faberman, 2002)
If one wished to use a set of independent variables and OLS regression to help analyze the potential relationships which are associated with the low-school-attainment landscape in the U.S., they would risk doing so under the apparently incorrect assumption that there is no spatial variability in the data.

In making this critique, I do not mean to argue that OLS techniques haven’t been utilized to good effect in an impressive range of fields. In some cases, the assumption of spatial homogeneity is not violated, or, on occasion, it may not make a difference if it is. However, what I wish to emphasize is that relatively recently, researchers in an array of arenas are discovering that given spatial nonstationarity, OLS models may at best underestimate the relationships being investigated; at worst OLS may actually cover-up strong localized relationships between/among the factors being modeled.

As a timely example of the best case scenario (i.e., underestimation), Lersch and Hart (2014) compared OLS models to those produced by an alternative regression technique which directly accounts for space—i.e., geographic weighted regression (GWR)—and found their ability to predict property crime was dramatically improved when they used the latter to explicitly account for the locations of facilities which polluted the environment with lead and lead-based compounds. This finding is corroborated by Pasculli, et al. (2014), who also found that OLS models compared to GWR tended to underestimate indoor radon exposure. Likewise, Chi, et al. (2013) reported that GWR improved their ability to predict obesity risk based on a variety of community-based variables. And as an example of the former case, Partridge et al. (2008) found that OLS models actually covered up (where GWR revealed) some strong
associations between localized economic policies, socioeconomic variables, and nonmetropolitan population growth dynamics in the U.S.

When applied to education and education-related inquiries, GWR has likewise outperformed OLS models. Indeed, this finding is supported by Fotheringham, et al. (2001) who found a large amount of spatial variability in the factors related to school performance in Northern England, which GWR captured more completely than did traditional OLS methods. Using GWR, Slagle (2010), similarly found OLS less suitable for explaining spatial variations in school spending among Missouri’s school districts.

In short, where there is spatial nonstationarity in the relationships between/among a set of factors and a predicted response variable being modeled, the use of traditional and popular statistical methods such as OLS are likely to end with some spurious conclusions.

1.3 Purpose of the Study

There were four main goals for this study. Using a set of variables available from the 1990 Census data as the basis to help explain the current (25 and older) low-school-attainment landscape in Nebraska’s largest public school district (Omaha, NE), my first goal was to determine the extent to which GWR techniques could be used to model, detect, and visualize spatial nonstationarity in low school attainment for Omaha Public Schools (OPS). My second goal was to determine if GWR techniques were in fact an improvement over OLS techniques. The third goal was to determine if the same model that predicted low school attainment in OPS would hold for LPS. The fourth goal was to explore the policy implications that may arise as a result of spatial variations in the
relationships between low educational attainment and the variables that are related to this phenomenon.

1.4 Research Questions

1. How much of the variability in the current low-school-attainment landscape in an urban area can be explained by a set of variables from the past?

2. Does the GWR technique do a better job than OLS of modeling the relationships between past community-wide demographic, housing, education, and economic conditions and the current low-school-attainment landscape in a given urban area?

3. Does the same set of variables related to low education attainment in one urban area apply to another demographically and geographically similar urban area?

4. Assuming the presence of spatial nonstationarity, what future policy implications arise from the presence of spatial variability in the strength of the relationships that predict low school attainment?

1.5 Theoretical Framework

There is a large body of research that deals with student persistence in school or, more pessimistically, with students dropping out. Concomitantly, there is a wide range of factors, theories, and conceptual frameworks that have been proposed as explanations for the variability we can observe among those people who persist in school and those who stop their formal education upon (or before) the attainment of a high diploma (or an equivalent). These factors, theories, and frameworks have an interesting grounding in the historic ‘evolution of blame’ (Deschennes, Tyack, & Cuban, 2001) for poor academic performance and low school attainment.
In an examination of Zehm’s (1973) dissertation work on the historic (1825 - 1925) labels used to describe poor-performing students, Deschenes, Tyack, and Cuban, (2001) pointed out that until the middle to second half of the 20th century, blame for not succeeding in school was located primarily within the individual. During this time, solutions for nonsuccess in school centered on the teacher coercing the “lazy or immoral child” (p. 535) to do better. Families—especially poor and immigrant families—also received a large portion of blame because “lazy and immoral” children had “intemperate, ignorant, undisciplined” parents, who were “unfamiliar with American values and customs” (p. 536).

It wasn’t until the Progressive Era in education that the structures of the school system began being recognized as a source of student nonsuccess. In this view, the now euphemized “low division pupils, sub-group z, and occupational students” (pp. 536) were failing in school because the system was too rigid and failed to differentiate between students’ intellectual abilities and life trajectories. What was needed was an institutionally based reform that separated the “laggards” from the “normal” students and which gave the former a more challenging curriculum. This, it was believed, would prepare the “laggards” for subordination and the “normal” students for professional life (Deschenes, Tyack, & Cuban, 2001).

A more recent change in the diagnosis of student nonsuccess shifted blame even further and onto the school itself. In this interpretation when children don’t persist in school it is most often because the culture of the school doesn’t match the cultural

---

2 Poor performance is not equivalent to dropping out or to low school attainment. However, Battin-Pearson et al., (2000) measured a significant amount of overlap between poor-performance and dropping out of school before the 10th grade, so poor performance is often a precursor to low school attainment.
backgrounds of the student and the community from which s/he comes (Deschenes, Tyack, & Cuban, 2001). Interestingly, this ‘evolution of blame’ described by the authors still provides much of the modern conceptual categories used to theorize about those who stop their formal education at or before the high school level. However, a growing number of researchers (e.g., Anyon, 2005; Currie & Moretti, 2003; Luo & Waite 2005; Porowski & Passa, 2011; Rothstein, 2004; Swanstrom, et al., 2013) are acknowledging that community factors beyond the school’s ready control may be responsible for producing students who are less likely to finish high school or go on to college. For example, Rothstein (2004) and Anyon (2005) have argued that raising the minimum wage, providing affordable and stable housing, expanding access to health and dental care, and committing to keeping unemployment rates low all count as policies that might positively boost achievement of students placed at risk, by reducing many of those sources of risk. Tapia (1998) has also noted the power of unemployment in a student’s household for inhibiting a student’s academic trajectory.

In sum then, there appears to be a history of shifting explanations for low school attainment and limited school success, but even as explanatory factors are assigned changing importance, the categories of possible explanation have become more stable. Broadly, those categorical explanations are: (1) an individual’s character limits their educational trajectory; (2) an individual’s family limits their educational trajectory; (3) an individual’s schooling experience limits their educational trajectory; (4) an individual’s community limits their educational trajectory. By-in-large, these four categories provide most (or all) of the framework for examining issues related to school performance, low school attainment, and many other investigations.
For example, Stroup and Robins (1972) found that certain variables from a person’s past foretell their likely persistence in school. Among a sample of 223 urban males, a mélange of individual factors (e.g., poor performance, truancy, early drinking activity) and family/demographic factors (e.g., parental social status and mobility) were strongly related to school failure. Using a set of past predictors, Stroup and Robins (1972) were able to explain 41% of the variability in dropout numbers among their participants. Those authors cited a study conducted by Lavin (1965), in which a literature review turned up a number of studies that similarly found correlations between school success and individual, familial, and demographic variables. Among these studies, Lavin discovered a variety of $R^2$ statistics (i.e., the explained variance) ranging from 9% to 49%. So, while Stroup and Robin’s (1972) findings were on the high-side of the explanatory range (41%) found by Lavin (1965), they still illuminated less than half of the story.

A decade-and-a-half later, Rumberger wrote that typically, the factors associated with dropping out of high school were grouped by researchers into several major categories: “demographic, family-related, peer, school-related, economic, and individual” factors (1987, p. 109). He proposed that low school attainment was a process that began early in a student’s life, and that each of the six major categories of factors should be considered as part of the same causal model.

A little more than a decade later, Jimerson et al. (2000) found strong linkages between individual, family, and school-related factors that had been previously established as correlates of low school attainment. Using a mix of complex statistical methods, the authors determined that school persistence was the result of a process...
beginning before elementary school that depended mainly on individual and family factors (e.g., the quality of caregiving). These authors were able to successfully classify 75% of 143 cases as “dropouts” or “traditional” students based on early home environment and caregiving, as well as gender, measured IQ, SES, and 1st and 6th grade performance factors.

That same year Battin-Pearson and colleagues (2000) tested and compared five widely established theories of low school attainment—all of which claimed some ability to explain and predict students who were increasingly likely to drop out by tenth grade. The authors labelled these theories as: general deviance (individual factors), deviant affiliation (peer-set factors), family socialization (family-background factors), and structural strains (demographic factors). The authors found that none of these theories could, by itself, explain the process of dropping out of school. Instead, each theory was only partially supported by the results of their analysis.

Intriguingly, in their final model they found that collectively all the predictors they tested combined to explain 50% of the variability in poor academic achievement (not the same thing as dropping out of school) for the people in their sample. And while there were direct effects from some of the variables that represented each of the five theories of dropping out, it was really poor academic achievement that mediated the relationships between the theoretical predictors and quitting school. In other words, the authors found that the five theories could explain 50% of the variation in poor academic achievement, and it was poor academic achievement which significantly predicted dropping out of high school. Yet, despite its strong association with low persistence in
school, poor achievement could still only explain 39% of the variance among dropout numbers.

Studies in education that employ regression techniques to help explain and predict low school attainment trends are not alone in their struggle for explanatory power. Focusing on their field’s premier publication, Weisburd and Piquero (2008) examined 169 articles appearing in the journal *Criminology* from 1968 – 2005. The authors were wondering if modeling in their field’s “flagship journal” (pp. 465) had improved over time. They found that starting in 1981, and for most years thereafter, 10% to 25% of articles in *Criminology* used some sort of multivariate regression technique in which an $R^2$ value was either directly reported or could be calculated (by taking the square of the *multiple-* $R$ statistic for example). Of the 169 total articles in their analysis, the mean $R^2$ value was .389 (39%). Weisburd and Piquero (2008) also found that 70% of the articles they analyzed had an $R^2$ value of .50 or less. Meaning, the large majority of authors using regression-based techniques in the leading journal for the field of criminology could explain (at best) half of the variations in the criterion variables they had set out to explain.

From this brief history of categorizing the culprit-variables thought to explain low school attainment and the use of regression techniques in education and criminology, a couple of key points emerge. One point is that since at least Stroup and Robins (1972), categories of variables used to predict persistence in school appear to have stabilized along historic lines. In brief, it used to be alleged that a student’s laziness/immorality and his poor, un-American family caused him to struggle in school. This perception gave way to one in which the structure of the curriculum received the blame. Clearly, the emphasis
was still on a lack of ability on part of the struggling (typically poor, often minority) student, but schools were now partly to blame. More recently, blame rested more squarely on schools, which were believed to be overly rigid and unwilling to adapt to the culture/language/needs of the children and communities they served. Another line of research suggests that it is the community itself which cause low schooling persistence. So, if historically individuals, families, and then schools and communities have all received some (albeit varying) amount of blame for poor performance and low school attainment, it seems that the current understanding (at least for many researchers) is that ‘really, all of these are culpable’.

Another point is that using regression techniques to explain the factors associated with low school attainment appears, like in the field of criminology, to have enjoyed somewhat limited success. When Battin-Pearson and colleagues (2000) identified and tested the ability of five theories explaining/predicting which students would quit high school, they didn’t do so in a vacuum. Rather, they were drawing from five frameworks of established scholarship on school persistence, all of which received the label: theory. Assuming that in order to be considered a theory of low school attainment in the first place, a set of thoroughly defended principles—which can both explain and predict low school attainment—needed to established, how then do we explicate Battin-Pearson et al.’s (2000) finding? How can it be that, combined, the best that all the theories of school non-persistence could muster was 39% of the dropout story in their data? How do we explain Lavin (1965), whose work revealed that the correlates of low academic success ranged in their explanatory ability from 9% to 49%? How does a set of ideas/principles/factors that has generally explained less than half of the high school
dropout story come to be considered a theory at all? (And, while we’re at it, why is it
that in the field of criminology [a field that like education also involves human behavior
and decision-making] has experienced strikingly similar limits on regression studies as
well?)

Perhaps Box and Draper’s (1987) pithy quote—“Essentially, all models are
wrong, but some models are useful” (p. 424)—applies to this situation. Models are
‘wrong’ in that all are, at best, approximations of reality, but some approximations get
usefully close to capturing the essence of what is going on, and, as such, these models are
useful for prediction and policy.

However, evidence appears to be mounting that the neglect of space and the
assumption of spatial homogeneity in traditional modeling can contribute to the
underestimation of the relationships detected by many statistical models. As we have
learned, spatial tools like GWR can often improve the ‘goodness-of-fit’ of global OLS
models. Erickson’s (1977) pointed out that interpretive methods (e.g., ethnography) may
offer more insight into the study of the particular than those produced by global statistical
analyses, which is an argument that is not inconsistent with the criticisms I have made so
far of OLS and global modeling techniques. However, while obviously not ethnographic,
GWR does offer an interesting new way to mathematically account for the particular
while still maintaining the potential for more generalizable claims, for broad
explanations, and for the articulation (and location) of error in predictions.

Schaefer (1953, p. 227) asserted that in an academic field, “methodological debate
is a sign of health” (227). GWR has improved modeling in many instances outside
educational inquiry (as well as in the limited uses in education-specific research) by
taking into account where generalized statistical claims apply and where they do not.

Rather than claim that GWR is an answer to some of the critiques of quantitative research (e.g., ‘that’ may be so in general, but it isn’t so here), in the spirit of Schaefer’s assertion, the question I ask is, if new technologies that allow us to pay attention to space improves explanation/prediction in other fields, why should this technology not improve education research as well?

Moreover, if we hope to fully explore the possibility of spatial nonstationarity in the individual, family, school, and community factors believed to be associated with low school attainment (and if we hope to improve the explanatory and predictive power of these factors) we need a spatial analysis tool that is up to the task. Given the success of the GWR methods used in the research detailed above, this technique seems worthy of our attention.

1.6 Importance of the Study

This study is significant for a variety of reasons. First, I present a relatively new research method (especially in the education research arena) which might improve our understanding of educational outcomes through more accurate modeling of the correlates of low school attainment. In reviewing the literature for this dissertation I have come to believe that this study is the first to apply GWR techniques to try to predict a current low-school-attainment landscape by using past community-wide factors captured by the U.S. Census. This study is also among the first to apply GWR techniques to education research more generally (previous examples include, Fotheringham, Charlton, & Brundson, 2001; Slagle, 2010; Qui & Wu, 2011).
Second, I provide a means for visualizing the geographic scope/scale of a given phenomenon (low school attainment in this case). Such visualizations could be highly beneficial if we wish to make stronger arguments and inform/persuade key stakeholders as to which policy reforms/interventions might be most efficacious in raising post-secondary attendance and completion. Also, through the analysis and mapping of the low-school-attainment landscape (and those past variables associated with it), we also position ourselves to match the location and scale of our policy supports/interventions/reforms to the location and scale of those successes we wish to sustain and those problems we hope to solve.

Third, I demonstrate empirically the need for consideration of nuance and local contexts if we are to form and implement education and education-related policy based on regression analyses. In OPS, the predictors of the low-school-attainment landscape were (1) the number of people with a high school diploma, equivalent, or less by Census tract in 1990, (2) the number of Hispanic households per Census tract in 1990, (3) the number of houses built before 1960 per Census tract, and (4) the distance of a Census tract to Interstate 80. The Omaha model did not apply in the Lincoln case, where the current educational landscape was largely a function of (1) the number of adults 25 and over per Census tract in 1990, (2) the number of Hispanic households in poverty per Census tract in 1990, (3) the number of houses built before 1960 per Census tract, (4) and the distance of a given Census tract to Interstate 80. These lists of predictors share much in common, distance to the interstate, older housing, and the generally poor job schools have done at addressing the diverse educational needs of Hispanic/Latino students (Conchas & Vigil, 2010; Deschenes, Tyack, & Cuban, 2001). However, these
relationships take on different dimensions and their predictive strengths vary both within and between Lincoln and Omaha, which suggests that city/state-wide programs or policies which assume ‘the problem is the same everywhere’ may unnecessarily waste resources by targeting areas that are less likely to be well-served by the program or policy (or by targeting areas where we would accomplish the exact opposite of the intended outcome).

Fourth, in an argument for the direction of future inquiry, I advance a spatially-oriented hypothesis about school attainment in relationship to distance from major transportation networks—in this case Interstate 80. Analyzing the predictive nature of Interstate 80 and its relationship to low school attainment is a novel and promising approach to imagining community-wide predictors of peoples’ persistence in school.

Finally, like many researchers before me (e.g., Anyon, 2005; Currie & Moretti, 2003; Hanushek, 2003; Luo & Waite 2005; Porowski & Passa, 2011; Rothstein, 2004, 2006; Swanstrom, et al., 2013) I find evidence that in-school reforms are not enough to improve education, and that education reform ought to be considered from a more comprehensive (i.e., community-wide) perspective.

1.7 Definition of Terms

American Community Survey (ACS)—a recurring statistical survey administered by the U.S. Census Bureau. The ACS is sent to approximately 3 million addresses each year (approximately 250,000 per month) and its purpose is to collect data in order to estimate a variety of community-based topics ranging from age, sex, race/ethnicity, and education, to language use, ancestral heritage, and mode of transportation to work.
Census Tract—small, relatively permanent subdivisions of U.S. counties which are delineated by the U.S. Census Bureau in order to provide a stable set of geographic units for statistical analyses. Census tracts are typically drawn so that they contain 1,200 to 8,000 people, but ideally they have approximately 4,000 inhabitants.

Geographic Information System (GIS)—integration of computer hardware with “problem solving” software programs (Longley, et al., 2011) which typically contain myriad tools for performing a multitude of geographical, spatial, temporal, and statistical, analyses as well as data storage, retrieval, and visual representational tasks.

Geographic Weighted Regression—GWR is a relatively new statistical technique which allows the strength of the relationship between predictors and a response variable to vary across geographic space. GWR is a ‘local’ statistic (as opposed to a ‘global’ one) in the sense that model parameters are estimated for every location in space and changes in these parameters can be easily visualized and mapped across an area of study.

Hotspot/Coldspot—a tool available in ArcGIS v. 10.1 that calculates a spatial statistic called Getis-Ord G_i* (‘G-i-star’) for each feature in a dataset based on some measureable attribute of that feature and those feature values of its ‘neighbors’. Each feature and its ‘neighborhood’ are compared to the study area as a whole. Features in neighborhoods with unexpectedly high values are termed hotspots and features in neighborhoods with unexpectedly low values are termed coldspots.
Low School Attainment—for this study, low school attainment will refer to the population of adults 25 and over, per census tract, whose highest advancement in school is completion of high school, an equivalent, or less (as estimated by the ACS).

Ordinary Least Squares Regression—a common, ‘global’ linear modeling technique typically used for making predictions about a response variable, the value of which is shown to be dependent on a set of predictors.

Spatial Analysis—a division of Geography that emerged in the 1950’s and 60’s whose adherents attempt to build and assess spatial models of various social and physical processes. These models are typically data-driven, mathematical in nature (relying mainly on algebra, geometry, and calculus), and their purpose lies in attempting to predict the future spatiotemporal conditions of some phenomenon of interest (Dixon and Jones, 1998).

Spatial Autocorrelation—a tool that calculates a spatial statistic called Global Moran’s Index (GMI). GMI provides a summary of spatial autocorrelation for an area, and it does this by simultaneously measuring a feature’s location and one of its attributes in relationship to other features and their attributes. In the case of low education attainment, the GMI statistic is a ratio of the differences in low education attainment from the average across each Census tract and its neighbors, compared to the difference in low education attainment deviations from the mean for all features in the study area.
Spatial Nonstationarity—a condition in which the structure of the relationships between some set of variables changes across a given geographic extent. When relationships between phenomena are not spatially homogenous (the same everywhere), attempts to model such relationships are likely to be wrong unless space is explicitly accounted for in the model.

1.8 Summary

Low school attainment is a correlate of a multitude of disconcerting health and economic outcomes and low school attainment in adults has been shown to impact children’s school performance and eventually their own school attainment. Disrupting this trend is (and has been) the focus of much research, because not only is low school attainment predictive of a host of concerning variables, it also has a tendency to persist from generation to generation. Traditionally, researchers have attempted to predict school attainment and draw conclusions based on techniques using ‘global’ inferences like OLS. Where there is spatial nonstationarity in the interrelationships between/among phenomena being modeled, researchers using these global techniques may miss important local caveats in their predictions. When fully analyzed, these caveats have been shown to create better models, which in turn might help to focus community resources and public policies in more effective ways.
CHAPTER TWO: OVERVIEW OF GIS AND GWR

2.1 Introduction

This chapter begins with an overview of the recent ascendency of geographic research methods and applications. Specifically, I discuss the use of geographic information systems (GIS) software designed for computer-based mapping which is being used in a range of fields, including education research and policy-making (although not yet as much as it could and should be). GWR techniques grew out of GIS, and tracing the beginning of GIS and its growth into a problem-solving technology (Longley et al., 2011) as well as its eventual application to education-related issues provides an important context for understanding why GWR might improve upon traditional statistical methods used to explain various weak school outcomes.

Next, I describe some the findings gained from research in non-education fields that have used GWR techniques. This set of studies in no way exhausts the wide-range of research using GWR techniques found in the scientific literature. Rather, I have purposefully selected and detailed three of the more frequently cited studies that use GWR techniques to show the flexibility of the technology. And while citation frequency is not a perfect proxy for the strength of an article, the ultimate point of discussing these studies is to examine what has been gained through GWR techniques in variety of areas outside of education. Understanding what has been gained in other fields through GWR techniques is essential for arguing what we might reasonably hope to gain from GWR in education research, and the articles selected for review help to make this point. Finally, I detail the methods, findings, and conclusions from the handful of recent studies that have applied GWR techniques to education-specific issues.
2.2 Geographic Methods and the Rise of Geographic Information Systems

According to Environmental Systems Research Institute (ESRI)—the developer of ArcGIS, the world’s leading GIS software suite—the term geographic information system (GIS) was first coined in 1968 by Roger Tomlinson, who is widely considered the “father of GIS” (50th Anniversary of GIS, 2012). Tomlinson published a paper about the development of a computer-based system for storing and manipulating maps and data to facilitate a Canadian rural development program. It was out of Tomlinson’s computer-based mapping system that modern GIS grew.

Longley, et al., (2011) pointed out that since GIS first came on the scene its usefulness has made it a ubiquitous feature of the modern world. From the moment a person wakes up and turns on the lights (electricity grid), gets in the shower (water grid), opens the mail (mail routing), walks kids to the bus stop (public transit routing), reads the newspaper (paper provided by sustainable forests managed by GIS), eats lunch (with food grown using soil nutrient, crop yield, fertilizer applications), and uses a car’s navigation system to avoid rush hour on the way home (GPS), until s/he turns on the internet to check his/her email one last time before bed (cable/fiber optic grid)—this person is in near constant contact with a GIS supported infrastructure. In the last 30 years, with the growth of personal computers and their exponentially expanding cheapness and power, GIS has correspondingly become increasingly omnipresent.

Many academic fields have also embraced the use of GIS. Recently, scholars in a wide variety of fields have begun using the term “GIS methodology” in the description of their work. For example, Russell & Heidkamp (2012) depicted food desertification in certain parts of New Haven, Connecticut characterized by low income, high poverty, and
limited vehicle access, all of which work against access to nutritious, fresh, high quality food. The researchers claim to have employed a “GIS methodology” to arrive at the conclusion “that the loss of just one supermarket has had significantly detrimental effects on the geographical food access of the city's residents” (p. 1197). Qin & Xie (2012) likewise used a “GIS methodology” (p. 3316) in a multi-year investigation of anthropogenic black carbon emissions in China. Murrieta-Flores (2012) used a “GIS methodology” to identify “particular characteristics of the landscape relevant to human movement, such as passageways, crossing points, and natural areas of transit” (103).

And, Barroeta-Hlusicka, et al. (2012) used GPS-GIS methodology to identify those areas in national parks that could benefit from more supervision from Park Rangers.

However, while many researchers (perhaps uncritically) refer to GIS as a methodology, what exactly a GIS is, and what functions it serves in academic research is a matter of some debate. Dixon & Jones (1998) argue that because many fields (take spatial analysis for example) predate GIS, and because GIS utilizes established but now computerized methods for conducting analyses, GIS must be considered simply a research tool, rather than a separate line of inquiry or methodology. Countering this, Longley et al. (2011), point out that GIS programs contain myriad tools within the software for performing multitudes of geographical, statistical, and data representational tasks, and that GIS itself is actually programmable by the users. As a result, GIS is constantly changing, being updated, and new tools are frequently added as the needs of the analysts change. Thus, GIS is not precisely a tool, but rather it is a kind of toolbox (as well as the mill that fabricates new tools) that allows many different kinds of users to perform a variety of analyses usually for the purpose of solving some real-world problem.
This led Longley et al. (2011), to characterize GIS more broadly as a problem-solving technology—and it is a problem-solving technology that continues to find applications in more and more arenas with each passing year.

In fact, the US Department of Labor reported that GIS users were so widespread and diverse that the commercial sector of geospatial technologies market was growing at an annual rate of almost 100 percent per year, and the market of geospatial technologies in general was expanding at the rate of 35 percent each year (U.S. Department of Labor, 2004).

Since Tomlinson imagined computer-based mapping as a way to facilitate Canadian rural development and regional planning, GIS has found application in a wide range of fields, from military intelligence, battlefield surveillance, and the launching of guided missiles and bombs, to social advocacy per Marxist, feminist, and other critical perspectives, poverty amelioration, and public service coverage of historically underserved and vulnerable populations (Elwood, 2006, Evans & Jones, 2008; Guerrero, et al., 2011; Ward & Peters, 2007). In fact, Goodchild (1992) proposed that the study of the theories, methods, technological development, and applications of GIS constituted a science in and of itself, and since then, many practitioners of GIS have come to think of GIS as a sort of testing ground for the ideas, hypotheses, and theories of a broader community of ‘GIScientists’ (Goodchild, 1992).

Our schools and administrative systems of education have not been exempted from the influence of GIS. For example, GIS software has been used by social service providers in conjunction with school districts to make foster care placement decisions for school-aged children. In Illinois, for example, several key constituents involved in
making decisions for the foster care system rely on GIS to demonstrate and coordinate how proximity to school fits into placement decisions. The social worker involved in making placement choices has developed a prototype GIS application that is referred to as SchoolMinder, the purpose of which is to “integrate data and identify foster homes in a child's school district or catchment area or, if there are none, the closest foster homes to either the child's school or natural parents” (Foltz, para.3, 2007).

The designers of SchoolMinder demonstrated the usefulness of the program to policymakers and subsequently the decision was made to begin using the application to factor school proximity into the decision-making process for foster care placement. This usage of GIS is in-keeping with the arguments coming from a growing number of researchers (e.g., Anyon, 2005; Curriea & Moretti, 2003; Luo & Waite 2005; Rothstein, 2004; Tough, 2009; inter alia), who are finding that education improvement efforts may be best directed outside of schools.

Beyond the applicability of GIS to some of the social dimensions of life outside of schools, GIS is often used in determining where to place schools and how to draw district boundaries. District leadership in Riverside County, California used GIS to draw attendance boundaries for the opening of new schools at all divisions of grade levels (i.e., high school, middle schools, and elementary schools). Here, GIS analysis not only allowed the district to respond to a large demographic change in their area, but, perhaps more impressively, the analysis was completed by one person.

The economic advantage of using GIS is made clear in this article. In the words of the school district analyst cited in the article, "Because the district has grown so quickly in such a short amount of time…the district could not afford to hire the personnel
required to accomplish all of the tasks in opening the new school sites. GIS was essential in the process. ArcView, SchoolSite, and I are best buddies; I could not do my job without them” (Davis, para. 6, 2001/2002). Because of budgetary concerns that so often come with public services like education, saving money is prospectively a very compelling reason to look more broadly at the applicability of GIS to a large range of issues in education. If the use of GIS can make more funding available that might otherwise be spent less efficiently and effectively on drawing boundaries, siting schools, and the like, there may be more budgetary room for that extra paraprofessional, or the new roof over the library, or the family literacy program after school.

The preceding examples demonstrate GIS applications outside of the day-to-day in-school activities of the classroom. But there is also a recent uptick in the usage of GIS as a pedagogical/instructional tool in high school and middle school geography, math, history, and social studies classes. For example, Broda & Baxter (2003) argued for the effectiveness of both GIS and GPS technology as an instructional tool in the social studies classroom. Kersi (2003) outlined the implementation of GIS technology in a broad range of secondary classrooms and found that GIS was an effective way to facilitate instruction in a variety of educational settings, including math, geography, and history.

Surely among the most successful examples of GIS being used in the classroom was detailed in an article published in ArcUser in Spring, 2009. This article tells about Steve Obenhaus, a math teacher at Olathe North High School in Olathe, Kansas, who had his students ask real-world geospatial questions and then taught them to use ArcGIS to help them find the answers. Obenhaus noticed a common theme in his students’ projects:
philanthropy. Most of Obenhaus’s students want to solve problems that people face locally, but occasionally, as Obenhaus said, students are interested in problems faced by those “half-a-world away”. For example, Elizabeth Vidaurre, one of his students, did her senior project on finding areas in a southern portion of rural Haiti where children did not have access to clean drinking water and then determining suitable locations for wells that might increase accessibility. Obenhaus (who volunteers in Haiti during the summer with his wife) travelled to Haiti with donated supplies, and he trained local Haitians to test the existing wells and streams with scientific instruments. He also taught them to take geospatial references of their field-test locations. He came back with a robust geographic dataset that nobody else in the world had.

Obenhaus and Vidaurre worked together to find a scientific answer to the question of where the best places were to dig new wells. Vidaurre also used spatial statistics to test if there was a correlation between well depth and the presence of E. coli and to chronicle where such wells were located. She found a correlation and mapped it. Her analysis revealed that hand-dug wells are not deep enough. Most impressively, Vidaurre’s work was presented several times in Haiti, resulting in donations that supported actually digging new and deeper wells. Vidaurre won the Spirit of Philanthropy Award from the Association of Fundraising Professionals. GIS enabled Vidaurre, who at the time had never travelled to Haiti, nor had ever laid eyes on a Haitian, to “traverse the space between the far and the near” (Hansen, 69) in order to do some good in the world.

Strictly speaking, as research, the Obenhaus and Vidaurre work is better categorized as GIS application to public health and perhaps hydrology than as educational inquiry. But it should be explicitly noted that this research was (co)conducted
by a high school student, and this is revelatory of another line of inquiry — what could be achieved with GIS as pedagogical tool? One finite answer regarding how GIS might matter to education is to say that it can be a component of high-school level curricula (as well as curricula for those even younger).

In education scholarship specifically, GIS appears to be finding a foothold as well. In 2009, Lubienski and colleagues used GIS to analyze school choice and competition incentives in Detroit, Washington D.C., and New Orleans. They chose these cities for their research because of the pre-existing markets for school choice and competition, and they were wondering to what extent (if any) school choice was leveling the playing field in these cities. One important finding from their GIS analysis was that economic incentives seemed to encourage some schools to engage in ringing (which they observed from their maps). These ‘rings’ are spatial patterns in which sets of schools were “sort[ing] themselves based largely on their preferred clientele, with different groups of schools asserting their advantageous position to serve more affluent students” (Lubienski et al., pp. 641, 2009). The implication here is that in these major US cities, policies supportive of school choice may be maintaining, or possibly exaggerating, the disparity in educational outcomes between the affluent and the poor.

Three years later, in an interesting article arguing for the use of GIS to promote political literacy, Hogrebe & Tate (2012) argued that GIS analyses are well-suited to include non-spatial data associated with schools. They argue that geography often evokes images of maps depicting grid squares, networks of streams, rivers, roads, mountains, valleys, and vegetation. Hogrebe & Tate (2012) pointed out that data not typically thought of as geospatial (i.e., policies, behaviors, test scores, student and teacher
demographics/characteristics, school funding, etc.) are still linked to schools, and are therefore geographical in nature. So one could make maps of per capita student spending or achievement outcomes related to school catchment zones or home addresses.

Furthermore, these authors offer a compelling example of how the internet and GIS can provide “access to a large number of people who can view data and variables in the transparent format of geographic space” (p. 82).

GIS was recently featured at a pre-session to the 2013 annual meeting of the American Educational Research Association (AERA), wherein researchers from Claremont Graduate University and the University of California-Berkley hosted an event titled, “Mapping Educational (In)Opportunity: A Hands-on Workshop that Explores GIS as a Research and Policy Tool for Social Change” (Ríos-Aguilar, et al. 2013).

Coincidentally, two months after this workshop, Jocson and Thorne-Wallington (2013) published an article in a highly visible journal dedicated to education research (i.e., Teacher’s College Record) in which they used GIS to examine the uneven geography of access to literacy-rich environments for many minorities in the St. Louis metropolitan area. Using a variety of the spatial analysis tools available in the ArcGIS software package, Jocson and Thorne-Wallington were able to identify demography-based spatial patterns and the relationship of these patterns with respect to access to facilities which they characterized as containing an abundance of materials and routines which were conducive to reading and writing activities (e.g., schools, libraries, bookstores, and museums). Their GIS-driven analyses allowed them to uncover what they refer to as a fragmented ecology of literacy opportunities. This fragmentation, the
authors argue, can create environmentally imposed limitations to literacy opportunities for historically underserved students.

In conclusion, GIS offers a novel and unique set of spatial and analytical tools, many of which have been developed relatively recently. Furthermore, researchers (in a wide range of fields) and policymakers (in an equally-wide range of policy arenas) have discovered the usefulness of GIS as a problem-solving technology and as a way to convey important information to colleagues and policymakers alike. These fields and policy arenas do include education policy and education research, however, the application of GIS to education research remains limited in important ways, not the least of which is the general lack of attention by paid to the promising approach to modeling offered by GWR techniques available in ESRI’s ArcGIS v10.1.

2.3 Examples of Geographically Weighted Regression Techniques

Chi, et al., (2013) were interested in a question that is similar to those that I am asking in this study—they wanted to know if GWR could improve our understanding of obesity in the U.S. by describing the contexts in which obesity is a predictable outcome associated with some knowable set of factors. Drawing on data from the 3000+ counties in the continental US, the authors attempted to add spatial context to much of the extant obesity research by accounting for variations in regional patterns in the explanation of the obesity epidemic in the U.S.

The authors approached this problem by using what they claim to be a familiar set of obesity predictors (e.g., unhealthy food environments, low SES, race/ethnicity) but a relatively new analytical technique; that is, they used GWR to analyze the relationship
between obesity and demographic, socioeconomic, and locations of poor food environments throughout US counties. The authors cited a host of studies that had found a positive correlation between unhealthy food environments and obesity. But there also exists another set of studies that find no connection between food environments and obesity. The authors argued that these mixed results might have been caused by the use of global models that failed to account for spatial variations in these environments, a problem that could be reduced if geography got factored back in.

For example, the accuracy of a predictive model may depend on regional variations (recall the U.S. low school attainment map from the previous chapter), which helps explain why in some cases researchers found a strong correlation between the food environment and obesity and in some cases they did not—geographic space itself is a variable in need of scrutiny. In their investigation, the authors found a negative and significant relationship between the physical environment and obesity, meaning that as an area worsened (across the dimensions of the predictors), the instances of obesity increased. They also found that in urban areas, high ratios of convenience-to-grocery stores, and high poverty rates, were positively associated with obesity rates.

The most important finding according to the authors was that associations between the major explanatory variables and obesity were ‘nonstationary’—i.e., the accuracy of their models varied significantly over space. In short, by accounting for space the GWR model provided more context to the obesity story—or, put another way, a more accurate understanding of the ‘hits and misses’ of their model’s predictions. GWR did this by allowing for geographic caveats to the general predictors of obesity. This variation in predictability has important implications for public policies, allocation of resources,
and the like because it not only implies community-specific strategies may be necessary to combat the obesity epidemic in the US, it also shows where such policies and resources may be directed to greatest effect.

This is almost precisely the same conclusion reached by Pasculli, et al., (2014) about a very different topic. Studying indoor radon exposures in Abruzzo region in central Italy, the authors argued that traditionally, the evaluation of an area’s radon potential has been approached through global modeling procedures which fail to account for spatial variability and local relationships between radon and associated environmental factors.

In response to this problem, Pasculli, et al., (2014) proposed a mixture of global and local statistics to carry out their analysis and to highlight the role of local relationships in contrast to the result of global analyses. By first identifying clustering tendencies, then using a global OLS procedure in combination with a local GWR analysis, Pasculli and his coauthors were able to show that the presence of radon was dependent on geographic space, and they showed the variations over that space in the strength of the relationships between the radon field and its correlates. All of this information was visualized and mapped to reveal the presence of nonstationarity among the correlation coefficients. That is, their maps depicted areas where their model’s coefficients were negative, positive, or near zero (i.e., where predictor variables were most predictive, where the outcome was the opposite of what the model suggested, and where they were non-predictive of changes in radon potential).

In essence, GWR allows for a ‘finer-tuned’ representation of reality, which potentially increases the reliability of predictions. Here again, the authors concluded that
contextual modeling using GWR can more accurately predict radon potential, and that mapping nonstationarity can be a useful way to communicate the need for different radon-reducing policies at national, regional, and local levels by equipping authorities with nuanced ways to regulate, monitor, and remediate unsafe radon levels in a particular area. These features of GWR present a serious advantage over traditional methods like OLS regression models that often mask spatial nonstationarity.

Papandreou and Tuomilehto (2014) used GWR to study the relationship between/among coronary heart disease (CHD), mortality rates, and diet, as well as a variety of anthropometric and biochemical variables. Focusing on a dataset that included health and heart disease statistics for seven countries, the authors were able to use GWR to determine which countries (and in the case of Greece, which subregions) were at greatest risk for high prevalence of CHD. Using these techniques they found that Crete, the Ionian Islands, and Japan exhibited very low prevalence of CHD, and they could find almost no systematic risk for CHD based on diet, physical activity, alcohol and tobacco consumption (and related variables) in these islands. The authors did find that their model was predictive of very high risk for CHD prevalence in Serbia/Montenegro and Finland. The authors suggested that recent research has shown the protective abilities of the ‘Mediterranean diet’ to reduce risk for a variety of health problems including heart diseases. In addition, traditional Japanese diets consist of similar nutrient combinations as the ‘Mediterranean diet’, and because the link between risk factors was so strong in some cases and so weak in others, the authors were able to provide evidence to support the idea that some countries could expect larger decreases than others in their incidences of CHD should they formulate policies and practices based on the general dietary patterns in
nations like Japan and the Crete and Ionian Islands. Had the authors used OLS rather GWR, they may have found a link between diet and CHD, however, this would have been a summary (and likely an underestimation for the nations that might benefit from dietary changes) of the relationship between the risk factors they identified and CHD rates for the entire 7 country dataset. In short, the authors were able to show detailed country/sub-region nuances and caveats because they opted to use GWR rather than traditional regression techniques.

2.4 Examples of GWR in Education Research
2.4.1 Math Performance in the U.K.

Fotheringham, Charlton, and Brundson (2001)—the original developers of GWR techniques (See Fotheringham, Charlton, and Brundson [1997])—applied their GWR model to school performance in Britain. In their application of GWR to school performance data, Fotheringham and colleagues (2001) analyzed standardized testing results for nearly 3700 students in Britain. More specifically, the authors wondered if links between catchment-zone characteristics (captured by the British Census), school size, and standardized test performance were spatially heterogeneous. The authors pointed out that they were nowhere near the first to examine such associations. In fact, they pointed to a series of studies (Brown et al., 1998; Conduit et al., 1996; and Coombes & Raybould, 1997) that had similarly attempted to measure catchment-zone/school performance relationships. The framework and units of analysis were not novel, but what was new was the method by which these relationships were analyzed. The authors were not interested so much in whether they could detect catchment-zone/school performance linkages, rather, they wondered if “perhaps some attributes of school catchment areas
have an effect on school performance in some areas and not others and [if] such variations are masked in global results” (Fotheringham, et al., 2001, p 44).

In order to measure school performance, the authors used the percentage of students (7 and 11 years-old) meeting or exceeding standards in mathematics. Mathematics is but one of many equally important areas that constitute and characterize a school’s performance, so the study may have been more aptly titled ‘Spatial Variations in Math Achievement’ rather than “Spatial Variations in School Performance.” Still, to their credit, the authors did recognize the limitations of a one-variable approach to the measurement of school performance, and although not ideal, focusing on math performance did make some sense because apparently (according to the authors) the relatively wide variability in the percentages of students meeting math standards garnered a great deal of concern in Britain around the turn of the century when their study was published.

Though not expressly the objective of the study, the authors chose to use independent variables that chronologically predated the dependent variable (that is expressly what my study seeks to do). That is, Fotheringham, et al. (2001) collected math achievement data from the 1997 school year, and the authors collected predictor variables from the 1991 British Census. In addition (and similarly to my study as well) the authors were using mainly out-of-school factors based on an areal unit (they used catchment zones, I am using census tracts) to predict math achievement; the lone exception to this was their use of school enrollment as a predictor of math scores.

Using a weighted least squares regression model, the authors first established that schools with high percentages of students meeting/exceeding math standards appeared to
be a function of: (1) low school enrollment, (2) catchment zones with high percentages of people in professional or managerial positions, (3) low percentages of people living in public housing, (4) low percentages of Indian people (a minoritized population in Britain), (5) low unemployment rates, and (6) low percentages of single-parent households. While each of these factors was significant, collectively, they only explained 24% of the variability in math achievement. Obviously, 66% of the variability left unaccounted for is quite high. Worse, as the authors pointed out, the absence of the remaining explanatory variables that account for 66% of the variability in math scores is compounded by the fact that this model is assumed to apply to the whole of Britain. To show why this might have been a problematic assumption, the authors applied their GWR model to the same set of variables listed above in order to investigate the potential for spatial nonstationarity in the modeled relationships.

As opposed to producing a single metric for the correlates of math achievement for all of Britain, GWR creates a mappable surface of parameter estimates for each predictor variable corresponding to all points in the geographic space comprising the area of study. In other words, the authors were able to make maps of where: (1) the predications from the global model were accurate, (2) the global model underestimated the strength of the relationship between a given predictor and math achievement, (3) the opposite of what the model predicted had occurred, and (4) the model had little or no predictive power at all.

For example, the global model predicted that math achievement should increase/decrease in response to the percentage of people in a catchment zone who work in a professional or managerial capacity. The map of this relationship across Britain
shows some interesting spatial variations that are covered up by the global results. Specifically, there are areas where responses in math achievement are more/less sensitive to changes in the percentage of professionals and managers that characterize a catchment zone. As another example, while the global model predicted an inverse relationship between school size and math scores—that is, lower school enrollment should mean higher math achievement—there were areas of the UK where there was actually a positive correlation between enrollment and achievement. Thus, in some places, math achievement might reasonably be expected to respond to a policy that created smaller school enrollments, and in other areas it would be less reasonable to expect the same. If this information produced by GWR represents an accurate model of reality, from a policy perspective it is easy to see how important these statistics could be; especially in policy situations where limited resources are a serious constraint.

2.4.2 School Finance in Missouri

Nine years after Fotheringham, et al. (2001) published their work on spatial nonstationarity in Britain’s math performance, Slagle (2010) proposed GWR as an improved method for studying school finance. In this article, Slagle compared the traditional OLS model to a GWR model in the estimation of a median voter model for education demand.

To understand Slagle’s comparison requires us first to take note of an influential paper by Bergstrom and Goodman (1973) that proposed the median voter model for estimating the demand for a public good. Their hypothesis was that public expenditure decisions confront (and conform to) political processes and the demand for public goods
can be empirically analyzed and tested using the characteristics of the ‘median voter’ in a particular administrative jurisdiction. Thus, if data could be obtained describing the median voter (e.g., race/ethnicity, gender, income, religious and political affiliations, preferences for public goods, etc.), then demand for a particular public good in that jurisdiction could be estimated (Wildasin, 1988). This is because the median voter is (given some fairly restrictive conditions) often representative of the coalition of self-interested voters that determines the level of public good provision for a particular jurisdiction (Gramlich & Rubenfield, 1982). Since its creation, the median voter model has been used for the analysis of demand for many public-sector goods, and education is no exception.

Since education is mainly a public good, the median voter model has been applied to education demand, but in ways that slightly differ from the original model. Returning to Slagle (2010), the median voter model for education demand predicts that for each public school district, a given measure of education demand (e.g., per pupil spending) can be approximated by two basic characteristics of the median voter/homeowner—(1) the price paid in taxes to create a dollar of revenue for schools (i.e., tax rate and income) and (2) the median voter/homeowner’s preference for public school spending. According to the author, in most empirical applications of this model, attributes of the median voter for each school district serves as a strong predictor of per pupil spending.

However, Slagle (2010) pointed out that this model confines the influence of voters in District One to the administrative boundary of District One, District Two to District Two, and so on. In order to keep this assumption intact, voters favoring certain spending behavior in one district could not influence voters in neighboring districts. As
Horowitz and Colburn (2003), pointed out, it is more often the case that a district’s spending decisions fail to remain confined to district boundaries. This creates a problem for any model that is insensitive to the effects of districts influencing other districts. One result is that much of the variability in a model’s parameters goes unexplained. For Slagle (2010) the problem with the OLS model for measuring education demand through the median voter method is that it makes a global prediction, and there is no accounting for the spatial variations in influence that occur between neighboring school districts when it comes to spending. He then identifies spatial nonstationarity as a potential source for explaining variability in public demand for education.

Using spending, wealth, income, and demographic data collected at two time periods (2000 and 2004) from all of Missouri’s school districts, Slagle (2010) compared the results of the OLS model to GWR (and to another modeling method that does not pertain to this review). Per Slagle, the OLS model predicted per pupil spending to be a function of tax rate, income, and voter preference, whereas the GWR model necessarily revised this set of predictors, adding per pupil spending in adjacent districts as a variable. For both the 2000 and 2004 datasets, the GWR model performed better than the OLS technique. In 2000, the OLS model explained 20% of the variability in district spending, compared to 54% with GWR. In 2004, the OLS model explained 24% of the variability in spending, compared to 52% explained with GWR.

The results supported Slagle’s (2010) hypothesis that by accounting for the influence of neighboring districts (i.e., by accounting for space) we would improve our ability to explain variations in spending patterns when using the median voter model. Slagle (2010) argued that one policy implication of these results was that teacher salaries
may be set inappropriately if the indices used to set these salaries (e.g., the NCES competitive wage index) use too broad a geography in their calculations without accounting for spatial variations in the costs of living, costs of education, and other relevant factors. In addition, he argued that GWR would be beneficial to the development of context-sensitive policies that guide state-level school expenditures. His logic was that teacher salaries and other spending patterns we can observe are the results of policies that have been set without the use of spatial tools like GWR. Because space matters and because GWR captures spatial shifts in the factors that predict spending trends while traditional methods do not, the appropriateness of policies based on traditional methods are called into question.

Further, we may expect that school spending would increase if more people had college educations because, as Slagle (2010) pointed out, having a college degree appears to influence one’s attitude vis-à-vis desiring (or at least accepting) increases in educational spending. But this trend didn’t hold for all of Missouri’s college graduates, and in some places, the opposite was true. So while it may generally be the case that where college graduates are clustered, support for higher school spending can be found, Slagle (2010) showed that expecting that generalization to hold everywhere is a tenuous assumption. Moreover, he was able to locate with a high degree of precision where that general trend was reversed and where it didn’t apply at all. These results and the conclusions they lead to are in good agreement with Fotheringham, et al. (2001). That is, if we desire to create policies that actually accomplish what we hope they will, then calibrating policy based on geographic context by using spatially-based technologies like GWR could be a better way of ensuring that we accomplish our goals.
2.4.3 ACT Scores in Missouri

The idea that decision-making is best supported by context-based analyses comes again from research in Missouri, where Qui and Wu (2011) sought to advance local regression analyses (GWR specifically) through the study of the American College Test (ACT) scores produced by 447 Missouri high schools. While Slagle (2010) used catchment-zone characteristics to predict spending, Qui and Wu (2011), like Fotheringham, et al. (2001), used a combination of school-level and community variables to try to predict school-level ACT scores. However, Qui and Wu (2011) used a more exploratory approach in selecting the model that best explained variations in Missouri’s ACT scores.

One of the problems sometimes encountered by those using statistical modeling techniques is that absent a well-defined theory, preexisting model, and/or testable hypotheses that provide a clear framework for the analysis of possible interrelationships between variables, searching out potential predictors is necessary (Braun & Oswald, 2011; Massy, 1965). In such instances, an exploratory approach to regression analysis can be (and often is) used (Braun & Oswald, 2011; Haig, 2005; Massy, 1965). To be sure, technical problems can arise with the use of exploratory techniques, and there is controversy surrounding this approach (which will be detailed in Chapter 3, the methodology section).

Qui and Wu (2011) selected variables from a set of nine factors for the existence of a best-fitting model from the candidate variables. Like Pasculli, et al. (2014), the authors employed a combination of global and local techniques to derive their final model. Qui and Wu (2011) established through OLS the five best predictors of high ACT
scores. They were: (1) high parental income; (2) high parental education levels; (3) two-parent families; (4) larger class sizes; and (5) more experienced teachers. They concluded that this model could explain 35% of the variations in Missouri ACT score. After verifying the success of their OLS model across an array of statistical tests that minimized the likelihood that the model violated important regression assumptions (especially variable redundancy—also called multicollinearity—or the degree to which independent variables are correlated with each other), a GWR model was applied to the variables listed above. The authors found that GWR significantly outperformed OLS, that their best GWR model could explain 63% of the variability in ACT scores, and that this finding was attributable to the presence of spatial nonstationarity. More importantly, global results indicated that the strongest predictor of high ACT scores was a school’s percentage of two-parent families; so, schools with proportionally more students from single-parent backgrounds tended to have lower ACT scores. But local analysis revealed clusters of school districts that stood as exceptions, where high percentages of two-parent families were associated with low ACT scores. There were also places where single-parent families were associated with higher ACT scores. Thus, while generally the OLS model’s predictions held, it missed important exceptions that the authors argued shouldn’t go unconsidered.

In addition, global results showed that experienced teachers tended to have students who performed better on the ACT. But again, local regression indicated the existence of a handful of districts where there is a strong association between less experienced teachers and higher ACT scores, as well as areas where there were more experienced teachers but lower ACT scores. For parent income and education levels as
well as student-teacher ratios (class size), similar variations in the strength and direction of the global model’s predictions were located throughout the state of Missouri. The authors end with a familiar sounding conclusion. From a state official perspective, their results would help to formulate policies and allocate resources to local areas based on the unique needs of those local communities and schools. For local stakeholders, coordination efforts with neighboring districts could be established which might initiate an exchange of effective strategies that could in turn increase achievement on the ACT, and academic success more generally.

2.5 Summary

Geographic Information Systems (GIS) offers to education researchers a unique and novel set of spatial and analytical tools. Furthermore, researchers in a wide range of fields have discovered the usefulness of GIS as a means for conveying important information to both colleagues and policymakers. GWR techniques grew out of GIS, and these techniques have been used to explore spatial nonstationarity and to improve modeling in a diverse range of fields, from the analysis of patterns and predictors of the U.S. obesity epidemic, to the spatial dependence of radon potential in the built environment, to international coronary heart disease research. There has also been limited application of GWR techniques to questions in education research, and the results and conclusions are very similar to what has been found in other fields. Namely, GWR can reveal the locations within an area of study where global results are accurate, where they are over/underestimated, and where they are the exact opposite of what a global model predicts. In doing so, GWR could potentially provide a way to support more nuanced,
context-based policy decisions, by providing evidence for arguments such as ‘the general trend does not apply here’, or ‘the general trend is particularly exacerbated here’.
CHAPTER THREE: Description of Study Areas

Figure 3.1: Reference Map: Relative Locations of Areas of Study

3.1 Omaha, NE

Omaha is home to approximately 434,353 residents (68% White alone; 13.7% Hispanic/Latino; 13.4% African American [U.S. Census Bureau Quick Facts, 2014]), and despite overall white decline since the turn of the last century, the city has experienced overall growth fueled primarily by growth among Hispanic/Latino populations.

In 2012-2013 the enrollment for the Omaha Public School district was 50,559 (32% White alone; 31% Hispanic/Latino; 26% African American [Nebraska Department of Education, 2012]), hence, OPS has a more diverse demographic mix than the city of
Omaha more generally. In part this is because the OPS and city boundaries do not coincide. The Omaha city limits cut across four mostly white suburban school districts (Elkhorn, Westside, Ralston, and Millard Public Schools), but for political and economic reasons, OPS has grown around many of these suburban communities (see Figure 3.2 below)\(^3\).

Figure 3.2: Omaha Public Schools, City Limits, and Surrounding Districts

In July, 2014, *Forbes Magazine* (online version) ranked Omaha, NE 25\(^{th}\) (Lincoln, NE was 6\(^{th}\)) out of 200 cities on a list of the best places in the U.S. for business and careers (Badenhausen, 2014). The methodology for these rankings included weighted

\(^3\) For a detailed account of the legal issues surrounding the growth and controversies surrounding the OPS district boundary see: Holme, et al., 2009.
metrics for job growth, cost of living and the costs of doing business, income growth, immigration, cultural and recreational opportunities, educational attainment, and the presence of highly-ranked colleges in the area. In August, 2014, *Forbes* again ranked Omaha highly—3rd out of 100 cities this time—on a list of the best places for young professionals. This ranking was based on metrics similar to those used to create the list of best places for business and careers, but it also included current and projected unemployment rates, which, because these were so comparatively low for Omaha, propelled the city into the top three places for young professionals to live (Carlye, 2014).

*Forbes* isn’t the only publication that has recognized Omaha. In November, 2013, *The Huffington Post* listed Omaha as among the “20 U.S. cities that one should visit in their 20’s” (Miller, 2013), and in July of 2014, *CNN Money* reported that on list of a best places to start a business, Omaha ranked 3rd out of 50 U.S. cities (Kavilanz, 2014).

Contrast the above laudatory acknowledgements from recent media outlets with a portrayal of early Omaha that appeared in *Harper’s Magazine* in 1869 (see: Menard, 1989, pp. 37):

Hast thou ever been in Omaha,
Where rolls the dark Missouri down,
And four strong horses scarce can draw
An empty wagon through the town?

Where sand is blown from every mound
To fill your eyes and ears and throat—
Where all the steamers are aground
And all the shanties are afloat?

Where whiskey shops the livelong night
Are vending out their poison juice;
Where men are often very tight
And women deemed a trifle loose?
Where taverns have an anxious quest
For every corner, shelf, and crack;
With have the people going west,
And all the others coming back?

Where theaters are all the run,
And bloody scalpers come to trade;
Where everything is overdone
And everybody underpaid?

If not, take heed to what I say!
You’ll find it just as I have found it;
And if it lies upon your way
For God’s sake, reader, go around it!

Based on some of the accolades Omaha has received recently from the mainstream media, it certainly seems at first blush that the city has come quite a long way since its frontier days, when it had a “scrofulous reputation” as a “cesspool of inequity” (Menard, 1989, pp. 37). Indeed, ever since the highly successful and locally established Union Stockyards Co. and early meat-packing giants George Hammond Packing Co., Armour, Cudahy, Fowler Brothers, and Swift opened for business in South Omaha around the turn of the 20th century (Davis, 2001; Menard, 1987), Omaha has had a reputation as being a site for big business. Currently, there are five Fortune 500 companies operating in Omaha: Berkshire Hathaway, Union Pacific, ConAgra Foods, Kiewit and Sons’, and Mutual of Omaha. If one views Fortune 500 companies from a per capita perspective, Omaha has a highly disproportionate number for a city its size.

---

4 South Omaha was considered a separate city until it was annexed by Omaha in 1915 (Menard, 1987).
If one of the narratives describing Omaha is as a “Miracle City” (Menard, 1987, pp. 37) with humble beginnings, first as a rough and dangerous frontier river port, which blossomed into a rough and dangerous railroad and meat-packing giant, which eventually turned into the contemporaneous mega-corporate sanctuary it is today, another less adulatory narrative also exists. Omaha has long been a site of racial tensions, exacerbated segregation, and racial/ethnic isolation that has ties to its beginnings as an immigrant town. Omaha’s burgeoning railroad and meat-packing industries (and all the tertiary sectors tied to those) brought in white immigrants early on from all over Europe, including Scandinavia, Germany, Ireland\(^5\), and England. Western and Northern Europeans dispersed throughout Omaha, but as Austrian, Polish, Italian, Sicilian, Greek, and Russian populations arrived in Omaha, they tended to cluster in the south and southwest areas of the city. According to Menard (1987):

---
\(^5\) Including many of my relatives.
The language, the grocery stores, the signs on the small businesses made clear that here [South Omaha] was Omaha’s Italian sector. Thence, in an arc southwest, following the railroad tracks and the packing houses came the Czech, Polish, and Bohemian neighborhoods. Going down the south side of Q Street, west of 24th (packing houses were on the north) a visitor might have heard Greek spoken for a few years as a Greek community was being founded. But an anti-Greek riot in 1909 largely drove its members out of town, though they left their mark on still visible business signs on a few of the buildings that remain. A few blocks west the Irish, many of whom arrived as railroad workers then stayed on with the packers, settled on Irish Hill. A little further south, around 36th and X Streets, the early Croatian immigrants’ habit of keeping geese provided their district its name—Goose Hollow. Back towards the north and out of South Omaha into the central and western parts of the city…a largely assimilated and wealthier population resided, for here the Western and Northern Europeans had spread and “Americanized” (pp. 43).

During these early years, African Americans arrived in Omaha from the rural south as well. Growth among this population was somewhat slow until 1910 - 1920, when Omaha’s African American population more than doubled (Larson and Cottrell,

---

6 Including my relative Andy Ryan who was stabbed to death on July 16th, 1893.
1997; Menard, 1987). Much of the African American community was, and remains, concentrated in north Omaha where early housing was officially designated “for colored families” (Kerns, 1932, pp. 1-5, as cited in Menard 1987, pp.44) in effect cordonning off housing for black families from the rest of the city. This policy was the initial cause for the geographic isolation of African Americans in north Omaha that persists today.

In 1910, there were only a few Hispanic/Latino families living South Omaha (Arbelaez, 2000). But by 1920, after First World War, the Hispanic/Latino population had increased to 682 (Sullenger, 1924) in response to meat-packing companies recruiting Mexican immigrant laborers as strikebreakers and union busters (Rawlings, 2009). From 1930 to 1940, as result of the great depression, there was a decrease in Omaha’s (and Nebraska’s) Hispanic/Latinos population, many of whom were under both formal (i.e., government sanctioned) and informal pressures to repatriate. (Repatriation was a logical problem for those Mexican Americans with American citizenship; one that was ultimately ignored by the federal government since many of the 400,000 Mexicans sent back to Mexico were natural-born U.S. citizens [Spring, 2013]).

Hamann and Harklau (2010) pointed out that patterns of institutionalized anti-Latino racism have long existed in traditional settlement sites like Chicago, California, and Texas, but the authors also delineated two competing hypotheses regarding how “welcome” or “unwelcome” (Gitlin, et al., 2003) Hispanics/Latinos are in new settlement sites that don’t have long-standing ties to Hispanic/Latino in/out-migration. One reaction suggested by Hamann and Harklau (2010) was that institutional improvisation can occur in U.S. cities and towns that experience new influxes of Hispanics/Latinos. The authors pointed out that places in the U.S. experiencing Hispanic/Latino influxes that do not have
historic pathways to and from Mexico, Central/South America, and Puerto Rico, sometimes react to newcomers in accommodating ways.

Another possibility proposed by Hamann and Harklau (2010) is that institutionalized racist treatment of Hispanics/Latinos in Chicago, Texas, and California is often replicated or recreated in new-settlement sites. In the case of early Omaha, it unclear whether or not this latter line of Hamann and Harklau’s (2010) reasoning applies. But, given the mob lynching of an apparently innocent African American packinghouse worker named Will Brown in Omaha in 1919—who was hanged and shot near the Omaha courthouse, then dragged through the streets and later publicly cremated (Hickey et al., 2007; Menard, 1987)—and given the generally hostile atmosphere in early Omaha towards Germans, Greeks, African Americans, Japanese, and Hispanics/Latinos—it seems as though a description of early Omaha as having a highly racialized orientation towards its newcomers is reasonably apt.

It is not surprising then that the weight of federal repatriation efforts and local animus towards minorities caused Omaha’s Hispanic/Latino population to fall from over 1,300 in 1930, to under 500 by 1950, and for Nebraska’s overall Hispanic/Latino population to fall from over 6,000 in 1930 to under 1,900 by 1940 (Davis, 2001). But, despite the impact of the federal government’s formal repatriation campaigns in the 1930’s, and Omaha’s history of hostility towards minority groups, today Hispanic/Latino communities constitute the majority ethnic group in southern Omaha.

In the last few decades, racial/ethnic turmoil in Omaha and OPS has played out not only in the federal court system, but in the court of public opinion as well. As an example of the former, United States v. School District of Omaha (8th Cir. 1975), OPS
was found have to have intentionally created and maintained segregated schools, and OPS was ordered by federal mandate to immediately begin the process of integrating both students and faculty. Student integration took the form of forced busing, which lasted from 1975-1999.

Then, a Nebraska legislative effort (Revised Committee Statement, LB1024, 2006), prompted faux-newsman and Fox News satirist Stephen Colbert to feature the Nebraska unicameral and its decision to, among other things, create three racially identifiable school districts in Omaha. Colbert satirized this legislation on April 19th, 2006, on his show The Colbert Report, in a segment called, “Tip of the hat/Wag of the finger.” What follows are Colbert’s comments on LB1024:

It’s time for ‘Tip of the hat, Wag of the finger.’ My first tip o’ the hat goes to the Nebraska state legislature, who voted to split Omaha’s public school into three proposed new districts: one that’s predominantly White, another predominantly Black, and a third predominantly Hispanic. A veritable ‘Neapolitan race-cream’. In this case we’ll make the Hispanics strawberry since so many of them were brought here to pick them. But, as usual, the ‘PC police’ are calling this plan racist, just because they didn’t cave in to those people who say Brown v. Board of Education is settled law. All the Nebraska legislators are saying is: ‘we don’t see why Whites, Blacks, and Hispanics—we don’t see Whites, Blacks, and Hispanics, we see children; children who would be a lot happier sticking with their own kind. These districts will still be equal, just
separate.’ Takes courage! Being the first legislature to redivide races in the face of opposition could well make them the Rosa Parks of resegregation.

The national discourse (un-attuned to local nuances as it can sometimes be) tended to ignore the fact that Ernie Chambers, Nebraska’s only African American state senator, crafted the amendment to LB1024 which suggested the trifurcation of the OPS district along racial lines under a logic of increasing African American control of Omaha’s long-standing racially segregated schools. Nevertheless, the story of LB1024 provides insight into Omaha’s historic and continued struggles with racial inequality and segregation, and from this brief history of Omaha’s success in business and its persistent struggles with racial isolation and inequality, a complicated picture begins to emerge; one that reveals patterns in the types of Omahans who have tended to benefit from the city’s status as a place friendly to big business. Ultimately, this brief history of Omaha should help to clarify the spatial patterns of racial/ethnic segregation, income, and unemployment disparities in the maps in Figure 3.4 below.
Figure 3.4: Key Demographic and Economic Features of the Omaha Public Schools District

Note: The color scheme in each map goes from light to dark based on ½ standard deviations increments. Light blues represent census tracts below the mean and purple shaded census tracts above the mean. The thick dark line trailing off the map represents I-80 and the thinner interesting line represent two of Omaha’s major cross streets, 72nd (north/south) and Maple (east/west). The small yellow dot north of I-80 and south of Maple in eastern OPS represents the CBD.
3.2 **Lincoln, NE**

According to 2012 U.S. Census estimates, Lincoln’s population was 265,506 (83% White alone; 6.3% Hispanic/Latino; 3.8% African American; 3.8% Asian alone [U.S. Census Bureau Quick Facts, 2014]). By contrast, the Lincoln Public Schools district (LPS) enrollment for 2012-13 was 36,943 (69% White alone; 12% Hispanic/Latino; 6% African American; 5% Asian alone [Nebraska Department of Education, 2012]). Hence, while LPS is more diverse than the city of Lincoln as whole, by comparison to Omaha and OPS, Lincoln and LPS are much more homogenous.

Figure 3.4: Lincoln Public Schools, City Limits, and Surrounding Districts
Nonetheless, Lincoln is like Omaha in that it has also gained a national reputation for being a nice place to live. Based on survey criteria ranging from life evaluation, work environment, access to necessities, emotional and physical health, and healthy behaviors, Gallup and Healthways computed a 2012 well-being index for all U.S. metropolitan areas. Based on their survey calculations, Gallup and Healthways ranked the Lincoln MSA as the happiest and healthiest metropolitan area in America (Witters, 2013).

There are, of course, many other differences and similarities in both the racial/ethnic and urban/economic development of Lincoln and Omaha, but in some ways, the story of Lincoln and its beginnings is an echo of Omaha’s. Perhaps the most obvious connection between the two is that each began as a railroad town (though since their beginnings Lincoln has perennially been anywhere from 1/3rd to 2/3rds the size of Omaha). And, both Lincoln and Omaha initially attracted a variety of newcomers who were not exactly met with conviviality by more established residents.

Two years after Omaha was founded in 1854, settlers started a village approximately 50 miles to the southwest, and by 1859, Lancaster village was the county seat of the newly formed Lancaster County. These early pioneers were drawn to the banks of Salt Creek, where they imagined rare prairie salt deposits as an auspicious opportunity for industry (Zimmer, 2005). Omaha was the early capital of the Nebraska territory, but, in 1867, Nebraska entered the union as the 37th state, and a political battle immediately ensued to have the capital moved to Lincoln. The Lincolnites—those Nebraskans, primarily south of the Platte River who were in favor of the move—were successful in their attempt to move the capital west, further in-state, and a sizeable area
(eventually named for 16th president) was carved, rather quickly, out of the supposedly mineral rich prairie in the area (Zimmer, 2005).

The salt industry in Lincoln never quite boomed, but in 1870 the Missouri River Railroad came to town, followed by a series of other rail lines. The population of Lincoln that year was 2,441. A few years later, Burlington consolidated several local lines (McKee, 1984; Zimmer, 2005) and Lincoln became a legitimate rail center (Zimmer, 2005).

Figure 3.5: Overview of Lincoln, NE Population Growth 1870 to 2012

In 1880, a decade after Lincoln acquired its first railroad the population had increased 432% to just over 13,000. And by 1890, the city had grown to 55,164, another 324% from the decade before. Kinbacher (2007) described the early immigrant experience in Lincoln during this period in part from the perspective of Henry J. Amen, a member of Lincoln’s largest and one of its earliest immigrant groups, the Germans from Russia.

The first wave of 150 to 200 Volga Germans who arrived in Lincoln in 1876 settled on the southwest edge of town (Kinbacher, 2007). Henry Amen arrived shortly thereafter in 1888, after the South and North Russian Bottoms had been established along the railroad lines and Salt Creek. Lincolnites were apparently less than receptive to Amen
and his fellow newcomers—another parallel to Omaha—viewing Germans from Russia as collectively dirty, ill-mannered, poor, and as an impediment to development (Kinbacher, 2007). The “constant discrimination” and “negative stereotypes” (Kinbacher, 2007, pp. 27) weren’t helped by the fact that Lincoln’s first arrivals of Volga Germans encouraged and sponsored friends and relatives from back home to join them, substantially increasing their numbers. According to Kinbacher (2007), by 1915, the North and South Russian Bottoms were home to 6,500 people (approximately 10 - 13% of Lincoln’s population at the time).

Near the end of the 19th century and in the first quarter of the 20th century, several towns sprang up around Lincoln, most of which were eventually annexed. To the east was University Place which was settled around Nebraska Wesleyan University in 1888 and later annexed in 1926. Farther to the east was Bethany Heights, established around Cotner College in 1890, annexed in 1926 as well. To the south was College View, also settled around a college, Union College, incorporated in 1892 and eventually annexed in 1929 (Zimmer, 2005).

In 1890, a railroad town along the Burlington and Missouri River lines named Havelock was incorporated. Located northeast of University Place in present day northeastern Lincoln, Havelock was the only of these satellite towns that was not organized around a college—a fact that may help provide some context for the patterns in the maps below. After several years of resistance, Havelock was eventually annexed in 1930 (Zimmer, 2005).

More recently, Lincoln has seen large influxes of refugees from all over the world. Due to a historically low cost of living and low unemployment rates, Lincoln was
designated by the U.S. Office of Refugee Resettlement in 1990 as a refugee relocation site (Pipher, 2002). As a result, from 1990 to 2000, Lincoln’s nonwhite population increased 128%, from just under 11,000 to just under 25,000. Since 2000, Lincoln’s nonwhite population has increased by another 11,000 (Kakimoto, 2011).

Much like Omaha, early patterns of urban growth and development, combined with patterns of both sanctioned (Kirbacher, 2007) and organic racial/ethnic isolation and segregation help to create the geospatial patterns of people and wealth that can still be seen in Lincoln today.

Figure 3.6: Key Demographic and Economic Features of the Lincoln Public Schools District

Note: The color scheme in each map goes from light to dark based on ½ standard deviations increments. Light blues represent census tracts below the mean and purple shaded census tracts above the mean. The thick dark line trailing off the map represents I-80. The small yellow dot in the center of the map represents the CBD.
Ultimately, only about 1 in 4 people in Lincoln and Omaha have never been to college. Statewide that figure is closer to 30% (U.S Bureau of the Census, 2012b). Conversely, about 29% of Nebraskans have a bachelor’s degree or higher, but in both Lincoln and Omaha around 1 in 3 have matriculated from college (U.S. Bureau of the Census, 2012c and 2012d). This isn’t surprising given that both Lincoln and Omaha are home to major universities, several private colleges, and host of community colleges and trade schools. Nevertheless, Lincoln and Omaha together constitute roughly 36% of the state’s population and 43% of the state’s residents with low school attainment. But, in LPS the on-time graduation rate gap between Whites and Hispanics/Latinos is 15%; for Whites and African Americans in LPS it’s 11%. In OPS, the gap is 10% for Whites compared to Hispanic/Latinos and 9% between Whites and African Americans (Nebraska Department of Education, 2012).

Worse, there is a strong relationship in each district’s high schools between high nonwhite enrollment and low graduation rates. In Figure 3.7 (below) I have plotted percent nonwhite enrollment against 4-year cohort, on time graduation rates for all OPS and LPS high schools. In Nebraska’s two largest school districts, the percentage of nonwhite enrollment significantly predicted on time graduation, $b = -0.27$, $t(11) = -4.72$, $p < 0.000$. Given this strong relationship between race/ethnicity and graduating from high school on time, coupled with the disparities in the spatial distributions of wealth, unemployment, and the racial isolation observable in the series of maps in Figures 3.4 and 3.6 (above)—especially in the context of the racialized histories of Lincoln and Omaha—it may very well be that patterns of inequality in OPS and LPS which exist today have strong ties to the past. The remainder of this dissertation is an exploration of
the spatial characteristics of some of the demographic, housing, income, and spatial variables from OPS and LPS in 1990, and how the spatial nature of these variables might help us to understand and explain the low-school-attainment landscape in Nebraska’s two largest school districts.

Figure 3.7: OPS and LPS High Schools: Percent nonwhite as a predictor of graduation rates

\[ y = -0.2581x + 92.872 \]
\[ R^2 = 0.6694 \]
CHAPTER FOUR: DATA AND METHODS

4.1 Introduction

This chapter details the data collection process and the methods used to address the following three questions: (1) How much of the current low-school-attainment in an urban area can be explained by a set of variables from the past? (2) Does GWR do a better job than OLS of modeling the relationships between past community-wide demographic, housing, education, and economic conditions and the recent low-school-attainment landscape in a given urban area? (3) Does the same set of variables related to low-school-attainment in one urban areas apply to another which is demographically and geographically similar? The fourth question (posed in the introductory chapter) arises as a consequent of the results of question two: i.e., if GWR produces superior explanatory and predictive power compared to OLS, and if the reason for GWR’s superior performance is the presence of spatial nonstationarity, then an obvious question that follows is: (4) assuming the presence of spatial nonstationarity, what future policy implications arise from the presence of spatial variability in the strength of the relationships that predict low school attainment?

4.2 Rates versus Raw Numbers

In my review of the studies for this dissertation, especially those using regression techniques, it was rare that the researchers used raw numbers as opposed to rates to represent their data. This is likely reflective of a broader trend in research in general, and perhaps especially so in regression modeling. Chamlin and Cochran (2004) argued that converting raw numbers to rates is a convention that is widely (and uncritically) accepted
by most scholars. In fact, Chamlin and Cochran (2004) laid out a century-old—and still on-going—debate among statisticians, wherein the central issue is whether or not rates using common terms (e.g., city population as the denominator) are inherently correlated, and therefore, inherently spurious if not examined explicitly—and if so, what should be done about this? The point that Chamlin and Cochran (2004) make is a good one:

“Conspicuous in its absence from the ongoing debate concerning the consequences of utilizing ratio variables in macro-social research is any discussion about whether or not we should be using ratio variables in the first place” (119). The authors stress the idea that there does not seem to be much empirical justification for the division of raw counts by population size.

According to Chamlin and Cochran (2004), much of the literature that describes the benefits of using ratios in lieu of raw numbers suggests that doing so deflates the disparities in raw numbers, and normalizes them by accounting for a place’s population, and in doing so also provides statistical control over population effects. It is certain, for example, that by virtue of its size, Omaha, NE will have higher numbers of residents with low school attainment than every other city in Nebraska. Thus, converting the raw count to a rate provides a better basis for the comparison of Omaha to other cities along this dimension.

However, when population-size effects are of interest, conversion to percentages might make less sense. From a public infrastructure perspective, paying attention to greater numbers of people, as opposed to greater ratios is in some cases more justifiable because higher rates does not necessarily correspond with more people using public services. Also, because budgetary constraints are in play with respect to reform and
intervention, from these perspectives, paying attention to where an intervention may potentially serve the greater number of people may help to increase the potential efficaciousness of the reform/intervention. Ultimately, I am in agreement with Chamlin and Cochran (2004), who concluded that converting to rates only partially controls for population size effects in multivariate models, and worse, doing so can lead to underestimating the effects of the changes in the number of people in an area on a criterion of interest. These factors led me to the decision to include a population count among the predictor variables and to use raw number counts, rather than rates, for the predictor and dependent variables. If a larger population count is an important predictor of higher numbers of residents with low school attainment, this relationship should be detected in the modeling process.

4.3 Data Collection

Data for this study were obtained through the University of Minnesota’s National Historical Geographic Information System (NHGIS) website\(^7\) and the Longitudinal Tract Data Base (LTDB)\(^8\). The Minnesota Population Center maintains this site which makes available digitized historical U.S. Census Bureau and American Community Survey (ACS) data from 1790 to present.

Also, because census tract boundaries have changed over time, researchers at Brown University (Logan, Xu, and Stults, 2012) have created the LTDB, which uses an areal interpolation algorithm to estimate past census data for present 2010 census tract boundaries. In other words, the LTDB project makes longitudinal analyses of census

---

\(^7\) [http://www.nhgis.org](http://www.nhgis.org)  
tracts possible—despite changing census tract boundaries—by populating the 2010 census tract boundaries with data from previous years. For this study, 1990 Census data was used to predict the current low-school-attainment landscape in OPS, but because census tracts in Omaha have changed since 1990, LTDB data was needed to measure past predictors and to maintain census tract consistency across the time period under investigation.

The dependent variable (low school attainment) was obtained from the NHGIS database, more specifically, from the 5-year (2008-2012) ACS. The ACS is a recurring statistical survey administered by the U.S. Census Bureau. The survey is sent to approximately 3 million addresses each year (approximately 250,000 per month) and its purpose is to collect samples of data in order to estimate a variety of community-based topics ranging from age, sex, and race/ethnicity, to language use, ancestral heritage, and distance/mode of transportation to work.

The ACS publishes 1-year, 3-year, and 5-year estimates for their survey data. This required a decision be made regarding which of these datasets to use in order to capture school attainment patterns (defined by the Census as “education attainment”) for Omaha’s census tracts. For this study the 5-year ACS estimates were selected because (1) the 5-year estimates are more precise and reliable than the 1- or 3-year estimates, (2) 1- and 3-year estimates may not be made available for census tracts due to privacy laws, and (3) the Census Bureau recommends using 5-year estimates for smaller geographies.

Another specific kind of dataset from the U.S. Census needed to be collected in order to analyze the data based on census tract aggregations. These data are called

---

9 [http://www.census.gov/acs/www/guidance_for_data_users/estimates/]
TIGER/Line shapefiles, which represent landscape features (roads, rivers, lakes, cities, etc.) as points and lines, as well as polygonal features (in this case census tracts). These files are created by the U.S. government at no cost to the user. However, the TIGER/Line shapefiles most often do not contain the data a researcher is interested in, thus, data relevant to a particular inquiry must be joined with these TIGER/Line files by a GIS. These shapefiles and external data associated with them can be downloaded from the U.S. Census website directly; however, the NHGIS website has enhanced the U.S. Census Bureau’s shapefiles and data by creating consistent relational tables that make joining data to their representative TIGER/Line features a simpler and faster process.

4.3.1 Dependent Variable: Low School Attainment

The U.S. Census Bureau defines *educational attainment* as “the highest level of education that an individual has completed”\(^{10}\). However, for some time scholars in education have drawn a distinction between education and schooling (Carson & Wilson, 1984). The former being a more general concept—and not limited to occurring solely in context of the latter—and the latter being a specific type of institutional setting wherein the former is supposed to take place. Accepting the logic that a school is not the only site at which a person might gain an education requires a re-defining of the Census Bureau’s definition of *educational attainment*. If education is assumed to occur in and out of schools, what the Census is actually measuring then is how far a person has progressed in school (i.e., graduation from high school, the earning of a bachelor’s or master’s degree, attending some college but not finishing, etc.), not necessarily how far a person has

\(^{10}\) [http://www.census.gov/hhes/socdemo/education/](http://www.census.gov/hhes/socdemo/education/)
progressed in his/her education. For this reason, I have changed the Census’s term 
educational attainment to school attainment. Thus, school attainment is defined in this 
study as the highest level of schooling an individual has completed.

It is worth noting that the Census Bureau categorizes school attainment along two 
broad categories: for populations 18 and over and for populations 25 and over. The 25 and over dataset was selected for this study because over the last few years, as anywhere from approximately 7 to 10% of students in Omaha, NE each year graduate a year or two behind schedule. This means that attainment of a high school degree is still in flux for many of the 18 and 19 year-old residents in Omaha. For this reason, school attainment in the population of residents 25 years and older was selected for the dependent variable.

Last, the Census Bureau measures how far respondents advance in school along the categories: None (meaning no formal schooling), leaving school sometime between pre-school and the 12th grade, graduating from high school or attaining an equivalent degree (e.g., GED), attending some college, and earning an associate’s degree, bachelor’s degree, master’s degree, professional degree, or a doctoral degree. Delineating low school attainment from this set of categories needed to be done in such a way as to accurately reflect some real-world distinctions between those who proceed in their schooling and those who do not. In other words, what is the real difference between low school attainment and not low school attainment? Given the research cited in the first and second chapters, there are fairly clear and myriad consequences that appear to be associated with not graduating from high school. In addition, recall that Trostel (2010) found that college graduates earn more money, pay more in taxes, and use less local, state, and federal resources on average than those without a college education. In accordance with Trostel
(2010), Ewert and Kominski (2012) estimated the work-life earnings of those people who either never attended school, who dropped out of school before high school graduation, or who never attempted college at all, was anywhere from $696,000 to $261,000 less than those people who had at least attempted some college (even if they didn’t finish).

Given these categories maintained by the Census Bureau, and the consequences of not graduating from or even attempting college, for this study, *low school attainment* will refer to the number of adults 25 and over, per census tract, whose farthest advancement in school was the completion of a high school diploma, an equivalent degree, or less.

### 4.3.2 Predictor Variables

As discussed in the theoretical framework in chapter one, factors that are likely to produce effects in low school attainment patterns include an array of individual, family, peer, school, and community/demographic dynamics. Accordingly, 1990 census tract-level predictor variables for this analysis were collected based on total residents, race/ethnicity, socio-economic status, school attainment, employment, and housing conditions. In addition, spatial variables (i.e., distance to the Central Business District [CBD] and distance to I-80) are introduced as novel candidate variables which might play some role in spatially organizing/influencing the observed low school attainment patterns in Omaha and Lincoln. The CBD and a city’s transportation infrastructure have been found to play important organizational roles in the spatial structure of phenomena distributed throughout the urban environment (Harrington & Warf, 2002; Losch, 1954). Recently, for example, Wei et al. (2010), found significant relationships between foreign direct investments in Shanghai and Nanjing, China and distance to transportation
networks (e.g., highways) and ports/ transportation hubs (e.g., railway stations and airports). For this study, a census tract’s distance to I-80 and the CBD were calculated within ArcGIS v. 10.1 using Euclidean distances from the census tract’s centroid to the target feature. Table 4.1 shows the twelve candidate variables that were gathered from the Brown University LTDB as well as the spatial variables measured within the GIS software.

Table 4.1 Definitions of independent variable

<table>
<thead>
<tr>
<th>Category</th>
<th>Predictor Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community/Education, economic, and income</td>
<td>HS90</td>
<td>Persons with a high school degree or less in 1990.</td>
</tr>
<tr>
<td></td>
<td>UNEMP90</td>
<td>Persons who were unemployed at the time of data collection in 1990.</td>
</tr>
<tr>
<td></td>
<td>HINC90</td>
<td>Median household income in 1990.</td>
</tr>
<tr>
<td></td>
<td>NPOV90</td>
<td>Persons at or below poverty level.</td>
</tr>
<tr>
<td>Community/Housing</td>
<td>H30OLD90</td>
<td>Housing structures that were built 30 or more years earlier in 1990.</td>
</tr>
<tr>
<td>Race/Ethnicity and SES</td>
<td>HHH90</td>
<td>Total Hispanic/Latino households in 1990.</td>
</tr>
<tr>
<td></td>
<td>HHB90</td>
<td>Total black/African American households in 1990.</td>
</tr>
<tr>
<td></td>
<td>NBPOV90</td>
<td>Number of blacks/African Americans in poverty in 1990.</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>NHPOV90</td>
<td>Number of Hispanics in poverty in 1990.</td>
<td></td>
</tr>
<tr>
<td>Spatial DISTCORE</td>
<td>Euclidean distance (m) from census tract centroid to CBD.</td>
<td></td>
</tr>
<tr>
<td>DIST80</td>
<td>Euclidean distance (m) from census tract centroid to Interstate 80.</td>
<td></td>
</tr>
</tbody>
</table>

These factors are in-keeping with much of the previously described research, however, any census-based research is necessarily confined to the variables and aggregations of data collected and published by the Census Bureau. This puts certain limits on the possibilities of a study using only Census-based data because the Census Bureau masks much of its data aggregated below the census tract-level. The Census Bureau collects individual and household level data, but publishes these data as aggregations due to privacy laws—and with a few exceptions, mostly as census tract-level aggregations (or larger). As Rawlings (2009) pointed out, aggregated Census data at the tract-level best approximates a neighborhood-level analysis, and findings in this study may not if the geographical unit of analysis is changes (i.e., results are likely to be different at the block-level, or the county-level). In this sense, the predictors used for building the OLS model fit into the schema proposed by researchers (Anyon, 2005; Curriea & Moretti, 2003; Hanushek, 2003; Luo & Waite 2005; Porowski & Passa, 2011; Rothstein, 2004, 2006; Swanstrom, et al., 2013) who have suggested that neighborhood-based factors are most strongly related to success in school.
4.4 Research Design

This study was conducted using the following four-tiered procedure, which I will discuss in more detail in the sections to come. First, the current low-school-attainment landscape was established using incremental spatial autocorrelation (which measures Global Moran’s I values at incrementally larger distances across an area of study) and hotspot analysis (which uses the Getis-Ord Gi* statistic). The use of these two statistical procedures constitutes a combining of global and local statistics (Anselin, 2003; Pasculli, et al., 2014) through which I was able to create a standardized measure of census tracts in Omaha, NE that are characterized by statistically significant levels of low school attainment (as well as those which are not). Second, a global prediction model was constructed using an exploratory approach in order to find a suitable ordinary least squared (OLS) regression model that explained at least 50% of the variability in low school attainment in Omaha’s census tracts without violating a range of prediction assumptions (described in more detail below). Third, using geographically weighted regression (GWR), the performance of the OLS model was then compared to the GWR model (which accounts for the possibility of local variations in the relationships among predictor and dependent variables). Fourth, the procedure was repeated for Lincoln, NE and the 1990 factors predicting low school attainment in Omaha were used on a dataset for Lincoln to explore the generalizability of the Omaha model.

4.4.1 Analysis of Spatial Dependency

In order to analyze the spatial patterns of low school attainment in Omaha, NE, (and the factors associated with these patterns) first the current low-school-attainment
landscape needed to be mapped using a statistically valid method (Huang & Wei, 2014; Lersch & Hart, 2014; Pasculli et al., 2014). A combination of global and local statistics (i.e., Global Moran’s I and Getis-Ord Gi*) were used to identify the general presence, magnitude, and location of clustering (i.e., the spatial autocorrelation) of low school attainment across the city of Omaha.

Creating a statistically valid visualization of the spatial patterns of low school attainment required attuning to a common spatial phenomenon observed by geographers for decades—that is, the extent to which the homogeneity of geospatial data decreases in relationship to distance across an area of study. This concept has been perhaps most famously stated by Tobler (1970), who proposed as the first law of geography the postulate that “all things are related, but close things are more related than distant things” (3). This phrase has become widely used in geography and related fields (Hect & Moxley, 2009) and in this case it forms the conceptual basis of the spatial autocorrelation techniques used to establish the low-school-attainment landscape for OPS.

Given the relative dearth of geospatial techniques in the field of education research, a brief description of the theoretical underpinnings of spatial autocorrelation is in order. Put simply, spatial autocorrelation is a measurement of feature (e.g., census tract) similarity across two dimensions: location and some attribute value of interest. If a group of census tracts with similar numbers of residents with low school attainment are proximate to one another this is said to be evidence of positive spatial autocorrelation—and such an arrangement supports Tobler’s first law (see Figure 4.2c below). However, Tobler’s first law is violated when: (1) features with similar attribute values are spread evenly apart (this is negative spatial autocorrelation—and this case features are said to be
dispersed [Figure 4.2a]), or (2) features with similar attribute values are located randomly throughout an area at varying distances—these features are assumed to be spatially independent (Figure 4.2b [Longley, et al. 2011]). With spatial autocorrelation, analysts are working to reject the pattern in Figure 4.2b—i.e., the hypothesis that the values of a phenomenon of interest corresponding to the locations of a set of features in a given area are spatially independent. Put simply, the spatial analyst is trying to find evidence that the observed spatial pattern is nonrandom. Successfully rejecting the spatial independence hypothesis can be interpreted as providing evidence that the spatial processes underlying the distribution of values that are observed are systematic in nature.

Figure 4.2: Spatial autocorrelation: Spatial dispersion, independence, and clustering

In order to test for spatial independence and to get an idea as to the spatial processes that might promote the distribution of low school attainment OPS, I used a spatial statistic called Global Moran’s Index (GMI). GMI provides a summary of spatial autocorrelation for an area, and this index does this by simultaneously measuring a feature’s (i.e., a census tract’s) location and one of its attributes—in this case low school attainment—in relationship to other features (i.e., other census tracts) and their low school attainment. In the case of school attainment, the GMI statistic represents a ratio of the deviations of low school attainment numbers from the mean for each census tract and
its neighbors, compared to the difference in low school attainment deviations from the mean for all combined features in the study area. The formula for this calculation is provided below (Equation 2):

\[
I = \frac{n}{S_0} \sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x}) \sum_j (x_i - \bar{x})^2
\]

In Eq. (2), the numerator contains the total number of observations, \( n \), the cross-products of the deviations from the mean for \( n \) observations of variable \( x \) at locations \( i \) and \( j \). These results are summed and multiplied by the spatial weight \( w_{ij} \) for each pair of neighbors and these values are also summed. In the denominator, the variance is calculated which considers deviations from the mean for all pairs of neighbors combined, and this is multiplied by the sum of all spatial weights, \( S_0 \)—which is a term that can be defined in a number of ways to reflect the spatial structure of the data. For this study, the spatial weights matrix (or \( S_0 \)) was defined as a fixed-distance band—based on Euclidean distances—with centroids of census tracts acting as reference points for defining the beginning measuring point for each distance band.

Like a correlation coefficient, a \( GMI \) value can range from -1.0 to +1.0, and a number close to zero is the value of \( I \) when no spatial autocorrelation is detected. Once a \( GMI \) value has been determined, an expected \( I \) value is then calculated. Given a large enough \( n \), the expected \( I \) value should always be very close to zero, and based on a comparison between the observed \( GMI \) and the expected \( I \), a \( z \)-score and a \( p \)-value are produced. Respectively, these two numbers provide a summary (hence, “Global” Moran’s \( I \)) of spatial clustering/dispersion across the area of study, as well as a
measurement of the probability that the clustering/dispersion detected is due to random chance.

While *GMI* does provide a methodological starting point for determining if a given variable is clustered throughout an area of study, it says nothing about the scale at which clustering is most intense or where exactly clustering occurs. If Tobler’s first law is not violated in the OPS case, groups (or neighborhoods) of census tracts with many inhabitants who have never attempted college should be located close together, and census tracts with many inhabitants who have at least some college experience should likewise be situated close to each other (i.e., these groups should be clustered). Such a spatial arrangement would provide the first piece of evidence that there are systematic and underlying spatial processes causing the ‘closeness’ of census tracts with high/low numbers of residents with low school attainment.

However, it is often the case (as it is with this study) that the appropriate scale at which clustering occurs is unknown. If the census tracts with low school attainment numbers are grouped together across a large extent, a spatial analysis utilizing too small a scale may miss important trends. The converse is also true, using too large a scale may miss important localized clustering that would be apparent if the spatial analysis utilized at a smaller scale. The major problem of course is that many times researchers will not know the appropriate spatial scale.

Where this is true, the spatial analyst must use her data to reveal the extent to which features (census tracts) with related values (low school attainment) are similarly influenced by the same underlying processes. Put another way, what constitutes an appropriately sized ‘distance band’ or ‘kernel’ of census tracts is often not identified by
previous research or by the parameters of the question being asked. Accurately conceptualizing spatial relationships in the data is essential to understanding and modeling how space influences geographic phenomena. But, imagining that there are underlying processes at work promoting some phenomenon of interest in a given census tract, what we really want to know is, what spatial qualities do these processes take? Do some phenomena in one census tract influence only those other tracts contiguous to it? Is there a zone of influence for a particular phenomenon that extends in all directions for one mile? Two? Is the diffusion of a given phenomenon directional, so that tracts to the west of a particular area are influenced more so than those to the east?

These are not easy questions to answer, fortunately, the Getis-Ord Gi* statistic, which is used to measure and locate clustering (Getis-Ord, 1992), provides some empirical insight into these important questions—especially when/where previous research is lacking. This procedure is referred to by the developers of ArcGIS as hotspot analysis. However, before a hotspot analysis (using the Gi* technique) can be conducted, the geographic scale (within the study area) at which low school attainment clustering is most intense must be modeled. To do this, the developers recommend a method referred to as incremental spatial autocorrelation (Stopka, et al. 2014), which for this study will form the basis for empirically establishing an appropriate scale at which to begin modeling the factors influencing low school attainment in Omaha’s census tracts.

4.4.2 Incremental Spatial Autocorrelation

Incremental spatial autocorrelation (ISA) works by generating clustering z-scores (produced by the GMI statistic described above) at incrementally larger and larger
distances across the area under investigation. Since the z-score produced by GMI reveals how intense clustering is across a geographic extent, comparing changes in these z-scores to changes in distance thus reveals fluctuations in the intensity of clustering at distance intervals across the area of study.

Figure 4.3 below displays the clustering z-scores on the vertical axis and increasing distance in meters along the horizontal axis for low school attainment in Omaha’s census tracts. It also shows visualizations of significant clustering at two peaks in z-scores corresponding to a particular distance. This graph provides a visualization of the distances at which the processes promoting the clustering of low school attainment in Omaha’s census tracts are strongest.

According to the ISA results, the processes promoting the clustering of low school attainment peak first at just over 5000 meters, and again just before 12,000 meters.
While there are higher z-scores at larger distances than 5000 meters, this first peak indicates the most granular representation of significant clustering. Because this first peak represents the most intense clustering at the most granular scale, this first peak was selected as the distance band.

4.4.3 Hotspot Analysis

With the scale at which clustering is most intense detected, each census tract was further analyzed in the context of its ‘neighboring’ census tracts (i.e., those tracts within the defined distance band). So far, spatial autocorrelation has given an initial indication of how intensely low school attainment is clustered in Omaha’s census tracts, and incremental spatial autocorrelation has provided additional information about the most granular scale at which clustering is most intense. But the explicit location of clustering is still needed for visualization and mapping.

To find this information, the hotspot analysis tool available in ArcGIS v. 10.1 was used. The hotspot tool calculates the Getis-Ord $G_i^*$ statistic for each feature in a dataset based on some attribute value and those values of its neighbors. The formula for this calculation is provided below (Equation 3):

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} (D) X_j - \overline{X} \sum_{j=1}^{n} w_{i,j}}{\sqrt{n \sum_{j=1}^{n} w^2_{i,j} - (\sum_{j=1}^{n} w_{i,j})^2}} \sqrt{n - 1}$$
In this case, \( G_i^* \) is the value computed for each census tract. These values are based on: \( X_j \) which is the value of \( X \) at location \( j \) (i.e., the location and number of inhabitants in a given census tract with low school attainment), \( \bar{X} \) the mean of all the attribute values in question, distance \( D \) between tract \( i \) and tract \( j \) (in this case \( D \) is measured to/from the centroids of the polygons representing each tract), \( w_{i,j} \) which is the spatial weighting conceptualized and modeled by the formula, and \( S \) the standard deviation the attribute.

Similarly to \( GMI \), \( G_i^* \) produces a \( z \)-score and \( p \)-value, however, recall that \( GMI \) is a global statistic while \( G_i^* \) is a local measurement. More precisely, \( G_i^* \) compares local means for a defined set of areal units (e.g., a census tract and its neighbors within the defined distance band) to the global mean of an area of study for a particular value of interest. In short, local means are compared proportionally to the global mean, and where a local mean is too disparate from the global mean to be explained simply as a random chance deviation, a statistically significant \( z \)-score is produced and a hotspot is detected (Stopka et al., 2014). Where the local mean is too low compared to the global mean to be due simply to random chance, a coldspot is detected. Where the local mean is similar enough to the global mean that random chance can reasonably explain a deviation from the mean, a non-statistically significant \( z \)-score is produced. Based on the all the census tracts’ \( z \)-scores the hotspot analysis tool creates a surface of hot/coldspots and a color-coded map layer to depict significant local deviations from the global mean. It is important to note that in order to be a hot/coldspot it isn’t enough for a census tract to have high/low values of the variable of interest—it also needs to be in a neighborhood of census tracts with high/low values. Without neighboring features that also have high/low
values, a census tract may be an interesting data outlier, but per this technical definition, it is not a hot/coldspot.

In summary, the Omaha, NE low-school-attainment landscape was established by using: (1) spatial autocorrelation techniques to analyze the spatial dependence/intensity of clustering of census tracts with high/low numbers of inhabitants with low school attainment for the entire city; (2) incremental spatial autocorrelation techniques to determine the most granular scale (i.e., the smallest neighborhood of census tracts) at which low school attainment clustering was most intense; and (3) hotspot analysis techniques to pinpoint specific locations of census tracts with statistically significant high/low numbers of residents with low school attainment.

4.4.4 Exploratory Regression

With the current low school attainment for Omaha mapped, an appropriate set of 1990 predictor variables needed to be established that might explain the observed spatial patterns. Given the well-researched categories of characteristics (i.e., individual, family, peer group, demographic/community) explaining low school attainment, a theoretical framework for establishing predictors was already in place. However, which of the candidate variables (described above)—in combination with each other—were most predictive was a question that required each combination of the predictor variables to be tested to answer how well they could explain variations in the number of residents in Omaha’s census tracts with low school attainment. To accomplish this, an exploratory regression tool was used. This algorithm is available in ArcGIS v.10.1, and it was used to test all possible combinations of the twelve predictor variables against a battery of
statistical tests to see which combinations of predictors ‘pass’ as well-specified models. Once a list of passable models was produced, the model which explained the most variability in the low-school-attainment landscape using the fewest number of predictors was selected for further analysis. To be a passable model, a set of predictors must total five or fewer and have: an adjusted $R^2$ of .50 or higher, coefficient $p$-values that are less than 0.05, a Variance Inflation Factor (VIF) of less than 7.5, a Jarque-Bera $p$-value greater than 0.10, and a spatial autocorrelation $p$-value greater than 0.10. What follows is a brief description of each of these tests.

**Adjusted $R^2$**— The $R^2$ statistic, also referred to as the coefficient of determination, provides a summary of how much variation in a dependent variable’s values is explained by a set of predictor variables (Weisburd & Piquero, 2008). The adjusted $R^2$ is a recalibration of the $R^2$ value which is known to artificially inflate as more independent variables are added to a model (Theil, 1961). The adjusted $R^2$ can be thought of as a ‘penalty’ for non-parsimoniousness, in the sense that it reduces the $R^2$ value as more variables are added to a model. Thus, in a multivariate model, the adjusted $R^2$ is always lower than the ‘raw’ $R^2$. Like $R^2$, an adjusted $R^2$ value close to 1.0 (say 0.90) would indicate that 90% of the variability in a dependent variable is explained by changes in the set of regressors being modeled — and conversely a value of 0 would indicate that a set of predictors has no explanatory power for the observed changes in a dependent variable.

---

11 It is important to note, as Weisburd & Piquero (2008) pointed out, that ‘explaining’ variation is not the same thing as ‘causing’ variation, so, while “causation requires correlation, correlation is not proof of causation” (454). This is true whether or not the modeled predictors explain all (or some fraction) of the variability in a dependent variable.
variable. In this study, a model that failed to explain at least 50% (after taking the adjusted $R^2$ penalty) of the variability in low school attainment in Omaha’s census tracts resulted in its elimination from further consideration.

**P-Value**—statistical inferences are typically made in the context of the null hypothesis. In the case of OLS regression modeling, the null hypothesis states that there is no linear relationship between a set of predictors and a dependent variable. For OLS modeling, coefficients are produced which describe the $y$-intercept and the linear relationship between each independent/dependent variable. If a coefficient value is too large to be due simply to random chance, the analyst makes the decision to reject the null hypothesis. The *p*-value provides the basis for making this decision because it quantifies the probability of obtaining a particular coefficient value when there really is no relationship between two variables (Kleinbaum, et al, 1998). Put another way, the *p*-value is a measurement of the likelihood that an analyst has found a significant relationship between two variables that is actually due to random chance. Small *p*-values represent low probabilities of this occurring, and, in the case of the GIS algorithm used to produce the models for this study, coefficients with associated *p*-values greater than 0.05 (i.e., relationships below the 95% confidence interval) were not permitted into the model.

**Variance Inflation Factor (VIF)**—this value represents a description of multicollinearity in a model. For models with two or more predictors there may be correlations between the predictor variables, which can result in highly unstable correlation coefficients (Kleinbaum, et al., 1998). This condition is described as multicollinearity. As an
example, if a model predicting the likelihood that a person will suffer from heart disease used weight and blood pressure as regressors, there would likely be a correlation between weight and blood pressure which might then inflate the relevance of one or both of these predictors. The VIF measures multicollinearity by determining the extent of the inflation caused by correlations between predictor variables (Kleinbaum, et al., 1998). Thus, the larger the VIF value, the more inflation is present, and the more unstable a model becomes. As a general heuristic, a VIF of 10.0 or higher is regarded as problematic. For this study, the VIF threshold was set more conservatively at 7.5.

*Jarque-Bera*—after relationships in a dataset have been modeled, predicted values can be computed using the observed independent variables. When these predictions are subtracted from observed dependent variable values, a residual is produced. If a model is properly specified, an analysis of the residuals should reveal that a model’s ‘misses’ are independent and normally distributed. Normally distributed ‘misses’ are an indication of a lack of organization and structure to model errors (i.e., it is evidence that the residuals are unbiased). Biased residuals indicate model misspecification, which in turn renders the results untrustworthy (Kleinbaum, 1998). Jarque and Bera (1987) proposed a procedure to test a model’s residuals for skewness and kurtosis (i.e., for normality). The null hypothesis for this procedure states that residuals are normally distributed. A *Jarque-Bera* score that is likely too high to be due to random chance provides evidence to support the rejection of the null hypothesis. For this study, the *p*-value threshold for the *Jarque-Bera* test was set conservatively at 0.10, so that a models’ residuals had an
increased chance of being considered biased, which then increased the likelihood that a model would be excluded from consideration.

**Spatially Autocorrelated Residuals**—as mentioned above, a basic assumption of regression modeling is that there is no systematic structure to model residuals (i.e., to a model’s ‘misses’). While the Jarque-Bera test provides a procedure for determining whether or not residuals are biased, if a model built on geographic data produces spatially biased predictions, this also violates the assumption of residual normality (Cliff & Ord, 1972)—but spatially clustered ‘misses’ are not captured by the Jarque-Bera test. Spatial autocorrelation techniques (detailed above) can be applied to model over/under-predictions in order to ascertain the geographic pattern of residuals. If, for example, there are significant clusters of census tracts, wherein the low school attainment model consistently over/under-predicts how many residents with low school attainment live in a given census tract in 2008 - 2012, this would provide evidence of model misspecification. In short, if my model’s ‘misses’ are spatially organized in a systematic way, it can mean that coefficient estimates are biased; it can increase the probability of finding significant coefficients that are not really significant; and/or it can mean that a key variable is missing from the model (Dormann et al., 2007). For this study, the *p*-value threshold for the spatial autocorrelation (i.e., Moran’s *I*) test was set conservatively at 0.10, so that a models’ residuals had an increased chance of being considered spatially autocorrelated, which then increased the likelihood that a model would be excluded from consideration.

In summary, an appropriate set of 1990 predictor variables that did not violate regression assumptions needed to be established in order to move on the next phase and
be tested as an OLS model. A theoretical framework existed for the selection of candidate variables, but this selection process was limited by the types of variables the Census collects. Twelve candidate variables were eventually chosen and their predictive properties were assessed. Each combination of the predictor variables (described above) was tested to determine how well they could explain the variations in the number of residents in Omaha’s census tracts with low school attainment. To do this an exploratory regression tool was used, which only allowed ‘well-specified’ models to be considered for further analysis. A ‘well-specified’ model was defined by an algorithm that applied a series of statistical benchmark tests to the data. To be a passable model, a set of predictors needed to have: an adjusted $R^2$ of .50 or higher, a coefficient $p$-value that is less than 0.05, a Variance Inflation Factor (VIF) of less than 7.5, a Jarque-Bera $p$-value greater than 0.10, and a spatial autocorrelation (Moran’s I) $p$-value greater than 0.10.

4.4.5 Ordinary Least Squares Calibration

Once a passable model had been established, I applied the ordinary least squares regression (OLS) algorithm in ArcGIS v. 10.1. The general form of an OLS model for $k$ independent variables is provided by the formula below (Equation 4):

$$ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_k X_k + \epsilon_i $$

where $Y$ is a value dependent on $X_1, X_2 \ldots X_k$, representing $k$ independent variables, $\beta_0, \beta_1\ldots \beta_k$, are their corresponding regression coefficients which need to be estimated (Kleinbaum, 1998) and $\epsilon_i$, an error term.
The model ‘passing’ all the regression assumptions (detailed above), which also explained the most variability in Omaha’s low-school-attainment landscape with the fewest predictors, included the following variables: (1) persons in a census tract with low school attainment in 1990, (2) the number of Hispanic/Latino households in a census tract in 1990, (3) the number of houses built before 1960 in a census tract, and (4) the distance of a given census tract to Interstate 80. Thus, the OLS model predicted that for a given Omaha census tract:

The ACS estimated (2008-2012) census tract-level, low school attainment in Omaha, NE is a function of 1990 numbers of residents who have low school attainment, Hispanic/Latino households, houses built before 1960, and the distance of a census tract to Interstate 80

The OLS tool in ArcGIS provided additional information about model performance (e.g., Akaike’s Information Criterion or AIC), and it also produced a map layer of model residuals, which allowed for the visualization of the global model’s over/under-predictions. Furthermore, Qui and Wu (2011) pointed out that prior to conducting a geographically weighted regression (GWR) analysis, it is necessary to first confirm that predictor variables are statistically valid and significant through OLS regression (and the accompanying tests for violations of regression assumptions). Such confirmation is necessary because GWR should be considered an extension of OLS, i.e., it is a local analysis meant to improve model and fit and detect spatial nonstationarity in global models. Additionally, GWR itself provides no safeguards against parameter biases, such as multicollinearity, which are known to make model coefficients unreliable (Qui & Wu, 2011).
In summary, the exploratory regression algorithm revealed a candidate model; the OLS algorithm confirmed the validity and significance of the selected model and mapped the global model residuals; and, in addition to vetting the variables, the OLS tool provided the basis for comparison between global and local (GWR) results.

4.4.6 Geographically Weighted Regression Model

The aim of GWR techniques is to estimate local regression coefficients ($\beta_{1...k}$) for each $j^{th}$ observation at each $i^{th}$ location (Brundson et al., 1998). For this study, this was accomplished by using ArcGIS to calculate a regression equation for each census tract in the context of its neighboring census tracts within a specified bandwidth. Brundson, et al. (1998), who first proposed this technique, suggested that selecting a bandwidth is akin to drawing “a circle of some radius, say $r$, around one particular [census tract], and calibrating an ordinary least squares regression model on the basis of observations whose geographical location was within this circle, then the $\beta_j$ obtained could be thought of as an estimate of the associations between the variables in and around [that census tract]” (433). The local regression equation given for a general GWR is provided below (Equation 5):

$$y_i = \beta_{i0} + \beta_{i1}x_{i1} + \beta_{i2}x_{i2} \ldots + \beta_{ik}x_{ik} + \epsilon_i$$

where the $i$ accompanying the regression parameters denotes that there is a separate equation for each subset of $n$ observations ([i.e., for each subset of census tracts] Partridge, et al., 2008). In calibrating the GWR model, a decision must be made
regarding size of the subset of $n$ observations. This is referred to as the bandwidth (or
“kernel” [Brundson et al., 433, 1998]) size for estimating the local regression parameters.
For GWR it is ordinarily (but not necessarily) assumed that Tobler’s first law applies to a
given dataset. Thus, the default weighting schematic (and the one used in this study) is
that census tracts near to point $i$ have more influence in the estimated regression values
than census tracts located far away from that same point (Fotheringham et al., 2001).

In calibrating kernel size, a decision needed to be made between selecting an
adaptive or fixed kernel. Using a fixed kernel ensures that area is preserved, so even
though the number of local observations in the kernel area will change, the area
represented by each local equation will remain constant (Brundson, et al., 1998).
Alternatively, an adaptive kernel will ensure that while the area of the kernel may change,
the number of observations within each kernel will remain the same. When the areal units
of analysis (i.e., census tracts) are highly irregular in size, it is most appropriate to select
the adaptive kernel (Fotheringham, et al., 2002; Partridge et al., 2011 Qui & Wu, 2014).
Because Omaha’s census tract boundaries are highly irregular in area and shape, an
adaptive kernel was selected for this analysis.

Consequently, the number of observations (i.e., census tracts) per kernel was
required to calculate the local regression coefficients. The GWR algorithm in ArcGIS
v.10.1 provides three ways of doing this: using a corrected Akaike’s Information
Criterion ($AIC_c$) or through a cross-validation technique. Following Partridge, et al.
(2011), Pasculli et al. (2014), and Slagle (2010), and based on the work of Burnham and
Anderson (2002) and Burnham et al. (2011), I used the $AIC_c$ method to calibrate local
regression estimates and kernel size.
AICc is a correction of AIC that is often used for smaller samples (Burnham & Anderson, 2002) and like AIC it provides a description of the goodness-of-fit for a statistical model by comparing its complexity to its residual sum of squares (RSS). Models with lower AIC/AICc values are better fitting models (Burnham et al., 2011; Burnham and Anderson, 2002; Fotheringham et al., 2003). Furthermore, so long as two models are measuring the same dependent variable, AIC provides a sound basis (Burnham et al., 2011) for comparing a global OLS model to a local GWR one (this will be discussed in the upcoming results chapter). And, within the GWR analysis algorithm, kernel sizes can be allowed to vary, and (ceteris paribus) local regression parameters can be estimated at different kernel sizes, effectively creating a series of local models that are all similar except for the number of observations in each subset. In other words, the algorithm will produce a series of regression models for Census Tract-X based on (for example) its 20, 25, 30, 35, 40, and 45 neighboring tracts. Then, the AICc values can be used to select the best-fitting model among these, thereby, selecting the most appropriate kernel size in the process (Fotheringham et al., 2003). This is the method for calibrating kernel size that I chose.

Ultimately, the GWR algorithm produced a series of four maps depicting a continuous surface of regression coefficients for each predictor variable and recent low school attainment data.

4.4.7 OPS Model Application to Lincoln Public Schools

The same model that applied to OPS was tested to the LPS district in Lincoln, NE. The procedure described above was then repeated for Lincoln, NE to try to find a
passable model specifically for LPS. A highly similar set of 1990 factors predicting low school attainment in Lincoln was successfully established, and this similarity provides the basis for a discussion about the inter/intraurban generalizability of the Omaha model.

4.5 Limitations of the Study

There are at least three important limitations in this study. First, I used aggregated census-tract-level data (recall that census tracts typically range in population from 1,200 to 8,000 people). Using aggregated census tract data requires that all results/conclusions be interpreted as specifically census-tract-level phenomena because of the possibility of committing an ecological fallacy—which has been shown to invalidate area-level conclusions that are applied to unit-level analyses (Steel & Holt, 1996; Qui & Wu, 2014). In other words, the ecological fallacy posits that what is true of the census tract is not necessarily true of the individual. Thus, in order to be considered valid, all conclusions in this study must be limited to the context of census-tract-level aggregations.

Second, in the search for a passable model, I used the exploratory regression tool to calculate every possible combination of the twelve candidate variables in relation to low school attainment. Recall that I set the threshold for variables permitted into a candidate model at five or less, which means that models relying on six or more variables were excluded from consideration. Setting this threshold low was important because as the number of candidate variables to be explored increased, and/or as the number of possible combinations of independent variables increased, so too did the probability of committing a Type I error—i.e., finding a significant relationship that doesn’t actually exist.
To understand this in clearer terms, consider that the total number of models tested can be formulated as the aggregation of the results of the expression, $C_{(n,r)} = n! \div (n - r)! \cdot (r)!$ for each combination of the twelve independent variables used in this study (with 5-variable combinations being the maximum). In other words, since study parameters were set so that a passable models could only combine five, four, three, two, or one of the twelve possible independent variables, the total number of models tested can be found by the number of combinations of 5-variable models (792) + 4-variable models (495) + 3-variable models (220) + 2-variable models (66) + 1-variable models (12), for a grand total of 1,585 models. Including 6-variable candidate models in the search for a passable model would have added another 924 possible models to select from, thereby increasing complexity as well as the probability of selecting a model which falsely indicated significant coefficients. The developers at ESRI (How Exploratory Regression Works, 2013) point out that because regression models are supported by probability theory, even at the 95% confidence level, as the number of tested models grows, the number of model coefficients that falsely indicate a significant relationship also increases, which in turn reduces the reliability of conclusions that can be drawn from the model.

Another limitation of exploratory regression and the third limitation of this study is that any passable models produced from the 1,585 candidate models were necessarily fitted to the data (as opposed to data being gathered in order to test an *a priori* formalized hypothesis). Fitting a model to a dataset can result in overfitting, wherein a model’s “portability” (Hawkins, pp.2, 2004) suffers—insofar as it applies only to one particular
dataset. This limitation provides the reasoning behind the decision to test the ‘interurban portability’ of the OPS model by applying it to LPS.

Despite the drawbacks of exploratory regression, it would have been difficult (if not impossible) to develop a potential GWR model without exploratory regression because there is little a priori knowledge, formalized hypotheses, and/or a testable theory for predicting how geospatially-oriented, community-wide factors from the past influence current geospatial patterns of low school attainment in urban settings. When/where a well-defined, formalized hypothesis is lacking, exploratory techniques can be highly useful to theory-building efforts (see: Burnham & Anderson, 2002, pp. 84-85; Braun & Oswald, 2013; Massy, 1950; Michaels et al., 2013). Fortunately, a great deal of research has been devoted to low school attainment, and while not geospatial in orientation the mass of low-school-attainment research (described in previous chapters) did provide me with a strong theoretical basis from which to select candidate variables. This fact saves my analysis from being simply a “data dredging” (Burnham & Anderson, 202, pp. 85) exercise, and it puts my modeling results and conclusions on somewhat surer footing. Nevertheless, results from this exploratory research should be considered as inductive, probabilistic, and preliminary, and as providing guidance for future geographically-based hypotheses and research about low school attainment.

4.6 Summary

The ACS estimated (2008-2012) low-school-attainment landscape for Omaha, NE was established by using a combination of global and local statistics. Next, an appropriate set of 1990 candidate variables which could help to explain the current low-school-
attainment landscape was established using exploratory regression techniques. Research about school persistence (from chapters one and two) provided a theoretical framework for the selection of candidate variables, but this selection process was limited by the types of variables the Census collects (as well as the geographical scales at which they publish these data).

Twelve candidate variables were chosen and their predictive properties were assessed using exploratory regression. Thus, the model building procedure proceeded as follows: the exploratory regression algorithm revealed a candidate model from among the twelve candidate variables; the OLS algorithm confirmed the tested and confirmed the validity and significance of the selected model and mapped the global model residuals; and, in addition to vetting the variables, the OLS tool provided the basis for comparison between global and local (GWR) results. Ultimately, the GWR algorithm produced a series of maps, one depicting the pattern of model residuals, and four additional maps, each depicting a continuous surface of regression coefficients for each predictor variable and the intercept.

There are at least three important limitations in this study. First, to avoid committing an ecological fallacy all results/conclusion must be interpreted as census tract-level (not individual-level) phenomena. Second, in the search for a passable model, I used the exploratory regression tool to calculate every possible combination of the twelve candidate variables in relation to low school attainment. Doing this increased the probability of committing a Type I error (finding a significant relationship that is actually false). Third, any passable models produced from the 1,585 candidate models were necessarily fitted to the data, which calls into question the generalizability of the model.
This issue of overfitting the model to the data provides the reasoning for applying the Omaha, NE model to Lincoln, NE.

Furthermore, I argue that exploratory regression is a defensible method for model building in this case because there is a lack of a well-defined, formalized, and testable spatial theory for predicting which community-wide factors from the past influence current patterns of low school attainment. I did have some theoretical basis for selecting candidate variables, even though few regression studies about low school attainment have explicitly formalized the urban/rural space as a variable—and no studies about low school attainment (insofar as I could find) have used GWR to visualize spatial nonstationarity in the relationships between past variables and a current low school attainment. Given this lack of research, it would have been difficult (if not impossible) to develop a GWR model without using exploratory regression. The decision to use exploratory techniques requires that results from this study be considered as preliminary, and as an exercise in theory-building for future geographically-based hypotheses and research about low school attainment.
CHAPTER FIVE: RESULTS

5.1 Introduction

In this chapter, I detail the results of the four-tiered research strategy described in the methodology section. First, I describe the outcome from the establishment of the low-school-attainment landscape. Next, I discuss the exploratory analysis and the results from the global OLS model. Then, I detail the maps produced from the GWR model and I discuss the performance of GWR model vis-à-vis the OLS results. I also assess the interurban portability of the Omaha Public Schools (OPS) model by applying it to the Lincoln Public Schools (LPS) dataset. Finally, I discuss the outcome of the OLS and GWR low school attainment models that were eventually developed for the LPS district. The maps in this chapter provide images of low school attainment in OPS and LPS, and visualizations of spatial nonstationarity in the relationships among the 1990 predictor variables and low school attainment.

5.2 The OPS Low-School-Attainment landscape

Table 5.1 below provides descriptive statistics for the dependent and each of the candidate variables tested for this study. The 2008 - 2012, ACS estimates found that there were approximately 95,173 residents with low school attainment residing within the OPS district boundary. The average census tract in OPS is home to an estimated 768 (SD = 411.413) residents whose formal schooling ended at or before the high school level, and there is a fairly wide range (63 to 2,619) of low school attainment numbers among the census tracts comprising OPS.
Table 5.1: Descriptive Statistics for Candidate and Dependent Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min.</th>
<th>Max.</th>
<th>Sum</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Dependent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low School</td>
<td>63</td>
<td>2619</td>
<td>95,173</td>
<td>767.52</td>
<td>411.413</td>
</tr>
<tr>
<td>(Candidate Predictors)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG25UP90</td>
<td>26</td>
<td>3826</td>
<td>214,080</td>
<td>1726.45</td>
<td>858.53</td>
</tr>
<tr>
<td>HS90</td>
<td>8</td>
<td>2786</td>
<td>104,225</td>
<td>840.52</td>
<td>501.911</td>
</tr>
<tr>
<td>H30OLD90</td>
<td>1</td>
<td>2060</td>
<td>75,143</td>
<td>605.99</td>
<td>564.289</td>
</tr>
<tr>
<td>NPOV90</td>
<td>1</td>
<td>1818</td>
<td>41,569</td>
<td>335.23</td>
<td>360.375</td>
</tr>
<tr>
<td>HBB90</td>
<td>0</td>
<td>1007</td>
<td>16,110</td>
<td>129.92</td>
<td>225.659</td>
</tr>
<tr>
<td>NBPOV90</td>
<td>0</td>
<td>1577</td>
<td>14,758</td>
<td>119.02</td>
<td>265.32</td>
</tr>
<tr>
<td>UNEMP90</td>
<td>1</td>
<td>314</td>
<td>8457</td>
<td>68.20</td>
<td>57.18</td>
</tr>
<tr>
<td>HHH90</td>
<td>0</td>
<td>172</td>
<td>3003</td>
<td>24.22</td>
<td>32.179</td>
</tr>
<tr>
<td>NHPOV90</td>
<td>0</td>
<td>173</td>
<td>1823</td>
<td>14.70</td>
<td>29.818</td>
</tr>
<tr>
<td>HINC90 ($)</td>
<td>8130</td>
<td>150,001</td>
<td>-</td>
<td>30,771.30</td>
<td>15,814.65</td>
</tr>
<tr>
<td>DISTI80(m)</td>
<td>0</td>
<td>13,092.96</td>
<td>-</td>
<td>4818.06</td>
<td>3393.34</td>
</tr>
<tr>
<td>DISTCBD(m)</td>
<td>0</td>
<td>20,424.74</td>
<td>-</td>
<td>7546.92</td>
<td>4984.91</td>
</tr>
</tbody>
</table>


Figure 5.1 (below) portrays the results of the census tract-level hotspot analysis for low school attainment in the OPS district. Recall that I used the hotspot algorithm in ArcGIS v 10.1 to discover where local clusters of census tracts with high/low numbers of low school attainment were statistically significant. In order for a census tract to be statistically significant, it needed to be above/below the low school attainment average and in a neighborhood of other census tracts with high/low levels of school attainment. The red-beige-blue conceptualization is meant to represent a ‘hot-neutral-cold’ configuration, so that increasing redness depicts census tracts increasingly above the OPS average for low school attainment (beginning at the 90% confidence level), beigeness
depicts census tracts close to the OPS average, and increasing blueness represents tracts that are increasingly below the OPS average (again beginning at the 90% confidence level).

Figure: 5.1: OPS Low School Attainment Hotspots

According to the results of the hotspot tool, the range of $z$-scores ($z = -4.30, p < 0.000$ to $5.73, p < 0.000$) indicated significant clustering of low school attainment both above and below the mean, with the most concentrated clustering of low school attainment occurring in the southern portion of the OPS district. In Figure 5.1 above, this hotspot began just south of the city core and then gradually intensified south of I-80. In central OPS, the pattern of census tracts with high/low levels of low school attainment
appears to have become more random, resulting in clustering z-scores that were not statistically significant.

In the north/northeastern portion of the OPS district another hotspot appeared, however, relative to the southern hotspot this northern area showed a less intense clustering of census tracts with high levels of low school attainment. Between the northern and southern hotspots depicted in Figure 5.1 there was a transition from red to beige to blue—that is, from significantly high clustering to a random pattern, eventually shifting to significantly low clustering. According to this map, the western portion of OPS was characterized by an intense clustering of census tracts with residents who had relatively higher school attainment.

In brief, southern OPS appeared to have the most significant clustering of low school attainment, while there was a more random pattern of low school attainment in central OPS. In the northern part of the district, there was significant clustering of low school attainment but it was not as pronounced as it was in the south; and in the western portion of the district, there was a fairly uniform pattern of relatively higher school attainment.

Another pertinent feature of this map was the location of the hot and coldspots relative to I-80. Figure 5.1 shows the interstate cutting directly across the southern hotspot (where low school attainment is most intense). The map also shows the interstate going near to, but not through western OPS (where school attainment is relatively high). The northern hotspot is located 8 to 13.5 km north of the interstate. So, in southern OPS, clustering of low school attainment was located relatively close to I-80 and in northern OPS low school attainment clustering was relatively distant from the I-80. These results
point to the likely presence of spatial nonstationarity in the relationship between low school attainment and the distance of a census tract to I-80. Furthermore, simply by looking at Figure 5.1, it is easy to discern distinct spatial variability in the low-school-attainment landscape in OPS that appeared tied to the geographic patterns of racial isolation of Hispanics/Latinos and African Americans in OPS (described in chapter three).

5.3 Analysis of Exploratory Results for OPS

How long a person persists in school is a result of a complicated mix of individual, family, peer, school, demographic, and economic (Battin-Pearson et al., 2000; Jimerson et al., 2000; Rumberger, 1987; Stroup and Robins, 1972) variables. While the twelve candidate predictors of low school attainment that I tested accessed some of these variables, they by no means represent an exhaustive list of the factors that have been shown to be associated with a person’s persistence in school. Also, it is worth reiterating that because of the potential for committing an ecological fallacy (Steel & Holt, 1996; Qui & Wu, 2014), census tract-level regression analyses cannot validly be assumed to apply to individuals. So, while individual-level data does comprise the census tract-level measures used in the model-building process, the use of tract-level aggregations prohibits the application of group-based inferences to individuals. Thus, the relationships and models described below are characteristic of aggregated census tract data, and must be interpreted as such.

Table 5.2 below shows the pairwise Pearson correlation coefficients ($r$) for each of the twelve candidate variables (squaring this value returns the $R^2$ term discussed in
previous chapters). A few key points emerge from Table xx. For example, the strongest correlation, $r(122) = .93, p < 0.01$, among all the pairs of variables was between the number of Black/African American households in a census tract in 1990, and the number of Blacks/African Americans in poverty in 1990. This seems to make sense simply as a function of population—the more households that were located in a census tract, the greater the potential for higher numbers of people to live in poverty. From a raw numbers perspective, this should be true of any group, and the correlation between Hispanic/Latino households in a census tract in 1990 and Latinos/Hispanics in poverty, $r(122) = 0.79, p < 0.01$, supports this point. Yet while the correlation, $r(122) = 0.24, p < 0.00$, between the number of residents 25 and over per census tract and the total number of people in poverty in 1990 was significant, $b = 0.10, t(122) = 2.73, p < 0.01$, the 25 and over population of OPS’s census tracts could only explain about 5% of the variability in poverty numbers. Hence, when white residents are factored in, increases in population generally shared a fairly weak connection to increases in poverty, but within the subgroups Black/African American and Hispanic/Latino, an increase in the number of households was strongly connected to higher poverty.

I analyzed this trend in more detail by using the 72nd street corridor (Rawlings, 2009) to divide the OPS district along a north-south centerline (apparent in Figure 5.1 above). In doing so, I created eastern and western sections of the district (by census tracts). From these, I was then able to calculate that in 1990, western OPS census tracts (which were predominantly white) were home to roughly 28% of the total OPS population; and in western OPS, about 1 out of every 33 people (3%) lived in poverty. Compared to eastern OPS, where the poverty rate in 1990 was 1 in 6 (and where large
concentrations of Blacks/African Americans and Hispanics/Latinos reside), the western portion of the district had a much lower poverty rate. Furthermore, in 1990, approximately 42,000 people residing within the OPS district boundary lived in poverty, and of these, 92% lived in the eastern portion of OPS. Thus, while in western OPS the share of the total OPS population in 1990 was 28%, its share of residents in poverty was only 8%. This glaring east/west disparity almost certainly accounts for the relatively strong correlations detected among Black/African American household, Hispanic/Latinos households, and poverty (because these groups were—and remain—fairly isolated in the eastern portion of the district). This east/west disparity may also help to explain why the correlation between total population and overall poverty was comparatively weak.

Intriguingly, despite the strong association between the number of black/African American households and the number of black/African Americans in poverty, the correlation between black/African American households and the current low-school-attainment numbers was not significant, $r(122) = .10$. Likewise, the number of blacks/African Americans in poverty shared a non-significant relationship, $r(122) = .08$, with recent low school attainment in OPS.
Table 5.2: Pairwise Pearson Correlation Coefficients for Twelve Candidate Variables in OPS

<table>
<thead>
<tr>
<th></th>
<th>LOWSCH</th>
<th>AG25UP90</th>
<th>HS90</th>
<th>UNE MP90</th>
<th>NPOV90</th>
<th>NBPOV90</th>
<th>NHPOV90</th>
<th>H30 OLD90</th>
<th>HINC90</th>
<th>HBB90</th>
<th>HHH90</th>
<th>DistI80</th>
<th>DistCore</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOWSCH</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG25UP90</td>
<td>0.57</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS90</td>
<td>0.83</td>
<td>0.79</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNEMP90</td>
<td>0.43</td>
<td>0.39</td>
<td>0.58</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPOV90</td>
<td>0.37</td>
<td>0.24</td>
<td>0.50</td>
<td>0.82</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NBPOV90</td>
<td>0.08</td>
<td>-0.01</td>
<td>0.22</td>
<td>0.72</td>
<td>0.83</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHPOV90</td>
<td>0.34</td>
<td>0.02</td>
<td>0.26</td>
<td>0.27</td>
<td>0.43</td>
<td>0.09</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H30OLD90</td>
<td>0.50</td>
<td>0.56</td>
<td>0.69</td>
<td>0.60</td>
<td>0.62</td>
<td>0.29</td>
<td>0.37</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HINC90</td>
<td>-0.34</td>
<td>-0.15</td>
<td>-0.43</td>
<td>-0.51</td>
<td>-0.59</td>
<td>-0.41</td>
<td>-0.33</td>
<td>-0.48</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBB90</td>
<td>0.10</td>
<td>0.05</td>
<td>0.25</td>
<td>0.73</td>
<td>0.77</td>
<td>0.93</td>
<td>0.03</td>
<td>0.30</td>
<td>-0.42</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HHH90</td>
<td>0.55</td>
<td>0.20</td>
<td>0.43</td>
<td>0.20</td>
<td>0.31</td>
<td>-0.04</td>
<td>0.79</td>
<td>0.40</td>
<td>-0.32</td>
<td>-0.08</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DistI80</td>
<td>-0.22</td>
<td>-0.26</td>
<td>-0.28</td>
<td>-0.08</td>
<td>-0.11</td>
<td>0.14</td>
<td>-0.34</td>
<td>-0.31</td>
<td>0.13</td>
<td>0.22</td>
<td>-0.47</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>DistCore</td>
<td>-0.35</td>
<td>-0.30</td>
<td>-0.52</td>
<td>-0.55</td>
<td>-0.59</td>
<td>-0.34</td>
<td>-0.39</td>
<td>-0.70</td>
<td>0.62</td>
<td>-0.37</td>
<td>-0.38</td>
<td>0.49</td>
<td>1</td>
</tr>
</tbody>
</table>
In summary, while the total 25 and over population in 1990 appeared to be weakly tied to 1990 poverty numbers in OPS’s census tracts, where Blacks/African Americans and Hispanic/Latino households were concerned, there was a much stronger connection between these groups and poverty than there was in OPS overall. But, in the case of Blacks/African Americans, I was unable to detect a systematic relationship between the total number of Black/African American households, the number of Black/African Americans in poverty (in 1990), and the recent number of residents in an OPS census tract who have never been to college.

The same, however, was not the case for the number of Hispanic/Latino households and Hispanics/Latinos in poverty in 1990. The correlation, \( r(122) = .55, p < 0.01 \), between the number of Hispanic/Latino households in an OPS census tract in 1990 and the current numbers of residents with low school attainment was significant. Meaning, there was a moderately strong predictive link between 1990 numbers of Hispanic/Latino households in a particular census tract and current levels of low school attainment, \( b = 7.07, t(122) = 7.33, p < 0.000 \). In addition, the number of Hispanic/Latino households in 1990 explained a significant portion (30%) of the variability in recent low school attainment levels in OPS census tracts, adjusted \( R^2 = .30, F(1, 122) = 53.7, p < 0.000 \). Intriguingly, as a model variable, the total number of Hispanics/Latinos in poverty performed worse as a single predictor, \( r(122) = .34, p < 0.000 \), than the total number of Hispanic/Latino households in 1990 in an OPS census tract. In other words, while Hispanic/Latino households in a census tract in 1990 explained, by itself, around 30% of the recent variability in low school attainment in OPS, the number of Hispanics/Latinos
living in poverty in 1990 could explain only around 11%, adjusted $R^2 = .11$, $F(1, 122) = 16.2, p < 0.000$.

Another important point from Table xx is that low-school-attainment levels in 1990 and current low school attainment were strongly related, $r(122) = .83, p < 0.000$. According to the pairwise regression results, the number of residents with a high school diploma, equivalent, or less in 1990 explained around 68%, $R^2 = .68, F(1, 122) = 263.99, p < 0.000$, of the variability in the recent low-school-attainment landscape. This indicates that of all the variables examined in the model, the 1990 school attainment variable shared the strongest relationship with recent levels of school attainment in OPS.

The 25 and older population in 1990 was also significantly related to recent school attainment levels, $r(122) = .57, p < 0.000$. Perhaps more importantly, there was a significant relationship, $r(122) = .79, p < 0.000$, between the low school attainment in 1990 and the population over 25 in 1990. Thus, to a degree, measuring low school attainment in OPS census tracts in 1990 should also capture fluctuations in the 1990 adult population, since there appeared to be systematic overlap between the two.

5.4 Omaha Public Schools-OLS Model Results

The exploratory regression algorithm available in ArcGIS v.10.1 identified a four-variable model that passed all the statistical tests (discussed in section 3.4.4) for violations of regression assumptions. The identified model used for further OLS analysis in ArcGIS included the following variables: HS90, H30OLD90, HHH90, DISTI80 (see the note in Table 5.3 below for definitions). Per the model-building criteria, each independent variable was significantly related to low school attainment, and with a
variance inflation factor (VIF) score of 2.04, multicollinearity was not an issue. In addition, model residuals were tested using the Jarque-Bera statistic for normality and the global Moran’s I for spatial autocorrelation and based on these tests I was able to conclude that model residuals were not significantly biased, $p > 0.10$, nor were they spatially correlated, $p > 0.10$.

Table 5.3: OPS-OLS Regression Coefficient Statistics

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>$t$ Statistic</th>
<th>$p$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>109.22</td>
<td>53.78</td>
<td>2.03</td>
<td>.04</td>
</tr>
<tr>
<td>HS90</td>
<td>0.70</td>
<td>0.05</td>
<td>13.23</td>
<td>.00</td>
</tr>
<tr>
<td>H30OLD90</td>
<td>-0.13</td>
<td>0.05</td>
<td>-2.87</td>
<td>.00</td>
</tr>
<tr>
<td>HHH90</td>
<td>3.92</td>
<td>0.70</td>
<td>5.62</td>
<td>.00</td>
</tr>
<tr>
<td>DISTI80</td>
<td>0.01</td>
<td>0.006</td>
<td>1.96</td>
<td>.05</td>
</tr>
</tbody>
</table>

Note: HS90 = number of residents with a high school diploma, equivalent, or less in 1990; H30OLD90 = number of houses in 1990 that were at least 30 years-old; HHH90 = number of Hispanic/Latino households in 1990; DISTI80 = distance from the centroid of a census tract to I-80 in meters.
The analysis of variance (ANOVA) test suggested that overall this was a statistically significant model, \( F(4,119) = 93.87, \ p < 0.000 \), and the adjusted \( R^2 = 0.75 \), indicated that 75\% of the variability in the current OPS low school attainment could be explained by the combination of the four factors described above. These results support the conclusion that (generally speaking) the current number of residents in an Omaha census tract with low school attainment was positively associated with 1990 numbers of residents with low school attainment, the number of Hispanic/Latino households in 1990, and distance from the interstate, and negatively associated with the number of houses in 1990 that were at least 30 years-old.
Out of 124 census tracts, the global model significantly over/underestimated (at the 95% threshold) the low-school-attainment numbers in six tracts. Figure 5.2 above depicts the locations of census tracts where the OLS model prediction was significantly higher/lower than expected. In Figure 5.2, the red/dark red representation indicates a situation in which the ACS estimated low school attainment was much greater than the predicted value. Hence, red/dark red census tracts denote a situation where low school attainment appears to be exacerbated, since the observed low school attainment is significantly higher than what the global model predicted. On the other hand, the lone blue-colored tract in northwest OPS represents a case where the model significantly over-predicted low school attainment. So, in this particular case, significantly more residents than expected have at least attempted college, despite a demographic mix that is suggestive of higher numbers of residents with low school attainment.

Significant over/under predictions do not appear to be particularly clustered across the geographic space of the area of study. However, it is interesting to note that when low school attainment numbers do appear to be systematically higher than expected, the census tracts are located in the eastern portion of the district, where poverty is relatively high, and where African Americans/Blacks and Hispanics/Latinos tend to be somewhat isolated. But, in the single census tract in northwestern OPS, the 1990 demographic mix suggests that a much greater number of current residents would have never attempted college—encouragingly, in this census tract, past demographics were unable to predict current school attainment. These few misses aside, the model’s residuals appeared to be well-calibrated and spatially independent.
5.5 Omaha Public Schools-GWR Results

At this point, I have identified a mix of 1990 demographic variables which appear to have a strong connection to the current low-school-attainment landscape. Here, the questions GWR will allow me to answer are, ‘Do the relationships between past demographic patterns in OPS and current low school attainment vary across the geographic space of the OPS district?’ and ‘If so, does accounting for spatial nonstationarity through GWR improve upon the OLS model?’

5.5.1 Considerations for the Visualizing of Spatial Nonstationarity

Before presenting the outcomes of my GWR analysis, there are a few concerns surrounding the mapping of GWR results which need to be addressed (Mennis, 2006). First, it is important to remember that the GWR algorithm in ArcGIS v.10.1 produces a series of visualizations each depicting a continuous surface of regression parameters. In other words, ArcGIS creates a series of maps showing the spatial variations in the slope coefficients which estimate the relationship between each predictor and the outcome variable. However, as Mennis (2006) pointed out, this visual information by itself can be highly misleading since such maps will emphasize areas where slope coefficients are relatively high/low (thus, a relationship may appear strong or weak), but they provide no information about whether or not high/low values are statistically significant. This is a problem because a map showing the spatial variability in the association between two variables may give “the impression that the areas with the highest parameter estimates exhibit the strongest relationship between the explanatory and dependent variables, when those estimates may not, in fact, be significant” (Mennis, 2006, pp. 172).
Mennis (2006) and Mathews and Yang (2012) suggested that for GWR-based maps to be interpreted correctly, both parameter estimates and their associated t-values (e.g., areas where t-values are +/- 1.96 and thus are significant at the 95% level) need to be mapped together simultaneously. One way to visualize this combination of information is to present two maps (side-by-side for example), with one map showing parameter estimates and the other t-values. However, a “more sophisticated approach” (Mathews & Yang, 2012, pp. 156)—and one I find easier to interpret—is to create contour lines based on t-values and overlay these on top of the GWR map depicting variations in regression parameters. Similar in interpretation to those contour lines found on topographic maps representing changes in elevation, t-value contour lines in this context reveal the locations of ‘peaks and valleys’ of statistically significant parameter estimates, as well as areas where regression coefficients increase/decrease in values.

Color scheme is another important consideration in map-making in general because variations in the data structure informing the map are often represented by changes in color on the map itself (a choropleth map, or weather map, is a good example [Mennis, 2006]). Two common map–making choices available in ArcGIS include using changes of color (e.g., the red to beige to blue schema I have selected) or a gradation in color intensity (e.g., light red to medium red to dark red) to represent changes in the underlying data structure. Creating a map that isn’t misleading requires appropriately matching the data structure to the color scheme. In the case of GWR, parameter estimates for a given area of study can fluctuate from negative to positive, or parameters can all have the same sign but their values can still vary a little or a lot. Where the former is the case, Mennis (2006) argued that a change in sign (i.e., positive to negative) signals a
change in the direction of the relationship being modeled, which is best depicted on the map as a sharp change in color (e.g., red to blue). In the case of the latter, where the sign doesn’t change but the strength of the relationship being modeled grows more/less intense across all, or a portion of the area of study, Mennis (2006) suggested that this trend is best represented by changes in color intensity (e.g., light red to medium red to dark red). The decisions I have made in creating the maps below—specifically the inclusion of $t$-value contour lines and color scheme choices—are based on the suggestions of Mennis (2006) and Mathews and Yang (2012) described above.
Figure 5.3: Low school attainment in 1990 as a predictor of current low school attainment
Figure 5.3 above shows the patterns of spatial variability in the relationship between low school attainment in OPS in 1990 (HS90), and low school attainment in OPS more recently (LOWSCH). According to these results, the relationship between past and present low school attainment was strongest in southern, central, and eastern OPS. In the southern portion of the district, especially south of I-80, parameter estimates were comparatively high, with \( t \)-values indicating statistical significance that were greater than/equal to \( t = 10.78 \). This is an extraordinarily high value; which suggests that for southern OPS, LOWSCH numbers were bound-up tightly with HS90. But, recall from the exploratory analysis, and the OLS modeling results, that among the variables selected, a census tract’s HS90 value was by far the strongest predictor of its LOWSCH. The \( t \)-value range in Figure 5.3 (min. \( t \)-value = 5.73; max. \( t \)-value = 11.56; max. \( p \)-value < 0.0000) suggests that this relationship held true (and is considerably strong) throughout the area of study.

Also, coefficient estimates ranged from 0.54 to 0.91 (mean = 0.70; \( SD = 0.10 \), meaning that from northwestern OPS (where parameter values were lowest) to eastern/southern OPS (where parameter values were highest) there were nearly four standard deviations separating the regression estimates between these two areas. Considering the \( t \)-value range and the coefficient estimates, there was a seemingly wide swing in the spatial variability in the relationship between past and present school attainment levels—but this is more a matter of how strong and how significant. If the general conclusion from the global model is that ‘in OPS, past low school attainment was a good predictor of current low school attainment’, the GWR results seem to suggest something along the lines of, ‘that generality was true of western OPS, and it was really
true for central/eastern/southern OPS’. In other words, the GWR results show some spatial variations in the relationship being modeled, but according to these results, the general claim that ‘current low school attainment was linked to low school attainment from 1990’ appeared to be the case throughout OPS.
Figure 5.4: Hispanic/Latino households in 1990 as a predictor of current low school attainment

Omaha Public Schools GWR Parameter and Isoline Map: Current Low School Attainment and HHH90

GWR Results
- t-value contour lines

Parameter Estimates
- High: 6.37934
- Low: -6.31187
- I80
- City Core
- Census Tract Boundary

Author: William R. England
http://www.mngis.org

Omaha Public Schools: Spatial Distribution of Hispanic/Latino Households in 1990 (HHH90)

Omaha Public Schools: Spatial Distribution of ACS Estimated (2008-2012) Low School Attainment

Stand. Deviations
- < 0.75 Std. Dev.
- 0.75 - 1.3 Std. Dev.
- 1.3 - 1.8 Std. Dev.
- 1.8 - 2.3 Std. Dev.
- 2.3 - 3.3 Std. Dev.
- City Core
- I80

Stand. Deviations
- < 1.3 Std. Dev.
- 1.3 - 2.7 Std. Dev.
- 2.7 - 3.3 Std. Dev.
- 3.3 - 4.7 Std. Dev.
- 4.7 - 6.7 Std. Dev.
- 6.7 - 9.7 Std. Dev.
- 9.7 - 12.7 Std. Dev.
- 12.7 - 16.7 Std. Dev.
- 16.7 - 21.7 Std. Dev.
- 21.7 - 27.7 Std. Dev.
- 27.7 - 34.7 Std. Dev.

Figure 5.4 above shows the pattern of spatial variability in the relationship between the numbers of Hispanic/Latino households in 1990 (HHH90) and LOWSCH in OPS. Unlike in Figure 5.3 (above), where coefficient estimates throughout the district were all positive and significant, the estimates for HHH90 changed direction (from positive to negative), and the significance of the relationship to low school attainment weakened following a southeast to northwest progression through the district. As was the case in the HS90 map in Figure 5.3, the relationship between HHH90 and current low school attainment appeared to be strongest in the southern portion of the district, near I-80, where t-values peaked at $t = 6.02$, just north of the interstate, and fell to a weaker but still significant $t = 1.96$, approximately 6km north of I-80.

The directionality of the t-score range (min. $t$-value = -1.60; max. $t$-value = 6.02; max. $p$-value = 0.11) indicating that beginning approximately 3km north of the interstate, near north-central OPS, the number of Hispanic/Latino households in a 1990 census tract became unrelated to recent low school attainment. And perhaps most intriguingly, in the western portion of the district, Hispanic/Latino households in a 1990 census tract may have begun to be associated with higher levels of education (though this was not a significant relationship).

The global OLS model indicated that, generally speaking, there was a strong positive correlation, $r(122) = .55$, $p < 0.01$, between the number of Hispanic/Latino households in 1990 and current low school attainment in the OPS district. The GWR results revealed obvious spatial nonstationarity in this relationship, so while the general conclusion from the global model held for southern and parts of the central OPS district, the association between HHH90 and current low school attainment weakened across
central and northeastern OPS. What's more, in the western portion of OPS, this trend may have started to lean in the opposite direction.
Figure 5.5: Houses 30-years-old or older in 1990 as a predictor of current low school attainment

Omaha Public Schools GWR Parameter and Isoline Map: Current Low School Attainment and H30OLD90

GWR Results
- t-value contour lines

Parameter Estimates
- High: 0.0187743
- Low: -0.215284

Author: William R. England
Source: Minnesota Population Center
http://www.nhgis.org
Figure 5.5 above shows the pattern of spatial variability in the relationship between the numbers of houses that were 30 years-old or older in a 1990 census tract ($H30OLD90$) and $LOWSCH$ in OPS. The results of the GWR analysis indicated large portions of western, central, southern, and northeastern OPS where the number of houses in a census tract that were built before 1960 was negatively (but weakly), $-0.98 \geq t \leq -1.96$, related to the $LOWSCH$ numbers. However, similarly to both $HS90$ and $HHH90$, in the southern portion of the OPS district the $H30OLD90$ variable appeared to have the most predictive strength. Beginning just south of I-80, parameter estimates and their $t$-values intensified, $t \geq -2.94$, indicating a strong negative association between the number of houses in 1990 built before 1960 and current low school attainment.

Although results from the global OLS model did suggest a generally negative relationship between $H30OLD90$ and $LOWSCH$, $t = -2.87$, $p < 0.00$, local analysis revealed areas in the OPS district where this relationship may break down. In particular, there was a sizeable region in the center of the district that extends to the north/northwest, where the $H30OLD90$ variable significantly weakened as a predictor, $t \leq -0.98$—and there were three smaller areas within this north/northwest extent where $H30OLD90$ seemed to be unrelated $LOWSCH$. There is also a comparable ‘ring’ around the city core (i.e., downtown), wherein the global OLS modeling results appeared not to apply, $t \leq -0.98$.

Furthermore, the GWR results indicated a generally negative relationship between $H30OLD90$ and $LOWSCH$. Meaning that in southern OPS, a decrease in the number of houses built before 1960 (in 1990) should be associated with an increase in current low school attainment. Put more simply, OLS results suggested that newer housing in 1990
should predict more current residents who had never attempted college. And while the smaller maps in Figure 5.5 show that there are a couple obvious exceptions to this association in the southern region of OPS, the maps also show that for the majority of census tracts in southern OPS, this trend tends to be the case.

Observe the southern OPS census tracts that were below average for $H30OLD90$ (blue in the left map) and were above average for $LOWSCH$ (brown in the right map). Where these tracts ‘flip’ colors indicates areas where newer housing in 1990 corresponds to a current tract with above average low school attainment. In the tracts in southern OPS that are above average for both $H30OLD90$ and $LOWSCH$ (i.e., light brown/red in both maps) we can still see patterns where new housing in 1990 was linked with current low school attainment because several of these tracts changed from lighter to darker brown/red between 1990 and current estimates. We can also see that the districts in the west that had fewer older houses in 1990 tend to have fewer residents with low school attainment currently. In central/eastern OPS, especially in the downtown area, there was a clear mix of tracts where the color ‘flipped’ or remained the same. The heterogeneous nature of these tract-level relationships represents a third case where spatial nonstationarity existed in the 1990 predictors of low school attainment, and a third case wherein the linkages between the 1990 predictors and the dependent variable have been far stronger for southern OPS than elsewhere in the district.
Figure 5.6: Distance to I-80 as a predictor of current low school attainment
Figure 5.6 above shows the pattern of spatial variability in the relationship between the distance of a census tract to I-80 in meters (DistI80) and LOWSCH in OPS. Recall that the slope coefficient for this relationship when estimated by the OLS model was, \( b = 0.01, t = 1.96, p = .05 \). However, local analysis showed obvious spatial variability in the parameter estimates, which GWR results indicated may actually range from -0.02 to 0.04—with \( t \)-values ranging from, \( t = -1.45 \) to \( t = 3.76 \). In this case, Figure 5.6 suggests that the strongest association between \( \text{DistI80} \) and \( \text{LOWSCH} \) was located in central and western (rather than southern) OPS. In this region of the district, a general increase in distance from I-80 was significantly and positively associated with low school attainment. Furthermore, the positive effect of distance from I-80 on low school attainment appeared to dissipate to the northeast, east, and southeast away from this central region, and eventually, the relationship reversed near the city core and then reverses again south of I-80.

The spatial pattern of parameter estimates for \( \text{DistI80} \) shares a few similarities with the spatial patterns of parameter estimates for \( H30\text{OLD90} \) in Figure 5.5 (above). Where the \( H30\text{OLD90} \) variable is concerned, a ring emerged around downtown Omaha that is very similar in scope and location to the ring around downtown Omaha for the \( \text{DistI80} \) variable. In addition, for both \( \text{DistI80} \) and \( H30\text{OLD90} \), a nearly identical contour line appears just south of I-80 indicating a fairly stark spatial change in the relationship being modeled for each variable. In the case of \( H30\text{OLD90} \), the spatial change appeared as a gradation of strength in the parameter estimates, which moved from non-significant in the downtown ring to significant near to, and south of, I80. And in the case of \( \text{DistI80} \), the spatial change appeared as a reversal of the slope coefficient, which moved from
negative (but not significant) in the downtown ring, to a ‘flat’ or non-relationship immediately south of downtown to I-80, then became positive (but not significant) in the southern portion of the district.

5.6 Performance of the OPS-OLS and GWR Model

According to Burnham, et al. (2009), Burnham & Anderson (2002), and Wagenmakers & Farrell (2004), ranking statistical models based on $R^2$ values (or other traditional statistics) does not establish a reasonable basis for choosing between a set of plausible/potential models. For the local GWR model, the overall adjusted $R^2$ was .79. Recall that the adjusted $R^2$ for the global OLS model was .75, indicating that the GWR model could explain 79% of the variation in the LOWSCH landscape compared to 75% for the OLS model. By this criterion, both models appeared to fit the data reasonably well, and the GWR model didn’t appear to be a substantial improvement over the global model. However, this conclusion is misleading.

Burnham, et al. (2009), and Burnham and Anderson (2002) criticize some of the judgments made by researchers who rank or choose between models based on traditional test statistics (and $R^2$ is one of those) because these judgments are often based on arbitrary values (e.g., $p$-value cut-offs that have no formal or empirical basis) and dichotomous assertions (e.g., accepting/rejecting the null hypothesis). In making their critiques, the authors argue that Information Theory (developed around code-breaking and communication during and after WWII) offered a “new class” of methods to researchers. One approach—based on an information-theoretic orientation originally developed by Kullback and Leiber (1951) and Shannon (1948)—allows researchers to
select the best model from a range of candidate models using a more fundamentally sound approach (Burnham, et al, 2011; Burnham and Anderson, 1998). What follows is a brief discussion of this technique and its application to my study.

Recall from chapter three (section 3.4.6) that a corrected version of Akaike’s Information Criterion ($AIC_c$) was used by the GWR algorithm in ArcGIS v.10.1 to compare and select the best local model from amongst a set of local models based on varying kernel sizes. It may be useful to understand that part of what makes an information-theoretic approach valuable to model selection is the ability of such an approach to provide a meaningful measurement of the amount of information that is lost between ‘full reality’ and a given model (i.e., relative to a set of other models).

Recall from an earlier Box and Draper’s (1987) cautionary note that, “Essentially, all models are wrong, but some models are useful” (p. 424). Models are wrong because they cannot avoid losing information when trying to describe ‘full reality’ (and this is true of any description of reality—not just for mathematical modeling). Burnham, et al. (2011) concur and add to this admonition that when modeling reality, what we usually want to know is which model, given a set of models, “loses the least information about full reality” (pp. 24). Of course, accurately computing and estimating how far away a model is from ‘full reality’ requires complete knowledge of full reality (which probably negates the need for a model in the first place).

However, in his now famous paper, Akaike (1973) provided the foundation for a group of calculations that measured the loss of information for a set of models, not from ‘full reality,’ but instead from a collected dataset (for a detailed discussion of Akaike’s Information Criterion see: Burnham and Anderson, 2002, pp. 60-80). In doing so, Akaike
was able to estimate how far from ‘full reality’ model-A was likely to be located
compared to model-B given how much information each model lost from the collected
data.

Suguira (1978) and Hurvich and Tsai (1989) later worked out a correction to $AIC$ to account for biases in smaller samples (this has obviously important practical
application for local regression techniques like GWR, which I have already described).
And based on this vein of research, Burnham and Anderson (2002) and Burnham, et al.
(2011), provided a fairly straight forward way to determine the best model from among a
set of models. This judgment is made on the basis of what Burnham and Anderson (2002)
termed a model’s Akaike weight ($w_i$) and the likelihood ($\ell_i$) that a particular model is—
relative to all the models tested—the one that fits the data the best.

Table xx below provides the requisite information for determining which model
(the GWR or OLS) best fits the data representing the OPS low-school-attainment
landscape. The crux of this comparison rests on the $AIC_c$ delta, or $\Delta_i$, which denotes the
difference in $AIC_c$ value between the model with the lowest $AIC_c$ and some alternative
model. In this case, $\Delta_i$ represents the $AIC_c$ value for the OLS model minus the $AIC_c$ for
the GWR model.

<table>
<thead>
<tr>
<th>Model</th>
<th>$AIC_c$ Value</th>
<th>$\Delta_i$</th>
<th>$\ell_i$</th>
<th>$w_i$</th>
<th>Evidence Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWR</td>
<td>1668.71</td>
<td>1.00</td>
<td>0.996204</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>1679.85</td>
<td>11.14</td>
<td>.00381048</td>
<td>0.003796</td>
<td>262.4352</td>
</tr>
</tbody>
</table>
Burnham and Anderson (2002) and Burnham, et al. (2011), noted that Δ values are crucial to model selection because they transform model comparisons to the “scale of information” and are interpretable regardless of the measurement scale” used (Burnham et al., 2011, p. 25). These values also serve as the basis for calculating additional information like the relative likelihood of each model being the best-fitting model, and the ratio of evidence supporting one model over another. Furthermore, Burnham, et al. (2011), and Burnham and Anderson (2002) suggested that a given set of models with Δ_i < 10 tend to provide no little or no support for distinguishing between the likelihood that there is one best model among them.

However, as a model’s Δ_i increases to 10 and beyond, the model likelihood (and therefore its plausibility as the best model) decreases correspondingly. In the OPS case, the OLS model’s Δ_i value was equal to 11.4, and hence it seemed to have lost more information as a description of reality than did the GWR model. As a result, is not likely that the global model is the better of the two choices. Returning to Box and Draper (1987) and assuming that a more ‘useful model’ is one that is ‘less wrong’—in that it loses less information about reality compared to some other model(s)—given the data I have used, and the two models I have tested, the GWR model appeared to be the more useful representation of reality.

5.7 ‘Interurban Portability’ of the Omaha Model Set

I applied the OLS model used for OPS to the LPS district to determine how well the Omaha model fit another urban area. Even though the OLS model did not fit the low school attainment data as well as the GWR model for OPS, I could not assume that the
same would be true of LPS. Thus, I began the methodological process over. First I applied the OPS-OLS model to the Lincoln district, and had this model turned out to be well-calibrated for the LPS low school attainment data, I would have then proceeded to the GWR analysis as I described above.

Recall that the OPS-OLS model included the following variables: HS90, H30OLD90, HHH90, DISTI80 (see Table xx above for definitions). Also recall that per the model-building criteria established earlier, a passable OLS model needed to have an adjusted $R^2$ of .50 or higher, coefficients with $p$-values that were less than 0.05, a Variance Inflation Factor (VIF) of less than 7.5, a Jarque-Bera $p$-value greater than 0.10, and a spatial autocorrelation (Moran’s I) $p$-value greater than 0.10.

From this point on I will refer to the OPS-OLS model that I applied to the Lincoln Public School district as the OPS-LPS model. The analysis of variance (ANOVA) test for the OPS-LPS model suggested that overall this was a statistically significant model, $F(4, 68) = 35.11, p < 0.000$, and the adjusted $R^2 = 0.65$, indicated that 65% of the variability in the recent LPS low-school-attainment landscape could be explained by the combination of the four factors that also explained 75% of the variability in the low-school-attainment landscape in OPS. The highest VIF value in the OPS-LPS model was 2.47, indicating that multicollinearity was not an issue. The Moran’s I statistic for spatial autocorrelation determined that model residuals were not significantly spatially correlated, $p > 0.10$. 
Table 5.5: OPS-LPS Model Regression Coefficient Statistics

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>331.89</td>
<td>71.42</td>
<td>4.65</td>
<td>.00</td>
</tr>
<tr>
<td>HS90</td>
<td>0.69</td>
<td>0.08</td>
<td>8.77</td>
<td>.00</td>
</tr>
<tr>
<td>H30OLD90</td>
<td>-0.16</td>
<td>0.08</td>
<td>-2.17</td>
<td>.03</td>
</tr>
<tr>
<td>HHH90</td>
<td>4.01</td>
<td>1.96</td>
<td>2.04</td>
<td>.04</td>
</tr>
<tr>
<td>DISTI80</td>
<td>-0.01</td>
<td>0.007</td>
<td>-2.12</td>
<td>.04</td>
</tr>
</tbody>
</table>

Note: HS90 = number of high school diploma, equivalent or less in 1990; H30OLD90 = number of house in 1990 that were at least 30 years-old; HHH90 = number of Hispanic/Latino households in 1990; DISTI80 = distance from the centroid of a census tract to I-80.

Table 5.5 above indicates that each of the OPS-LPS predictors was significantly related to recent low school attainment in LPS. However, the results of the Jarque-Bera test were significant, p < 0.000, which pointed to the likelihood that OPS-LPS model residuals were not normally distributed. This means that the OPS-LPS model was likely misspecified, and any conclusions derived from this model would not be trustworthy.

Thus, the OPS-LPS model was not passable, and there was no support for proceeding to a local GWR analysis based on these global results. Rather, these results provided evidence for the conclusion that the exact global model for OPS became miscalibrated when it was applied LPS low-school-attainment data. The question that remained, however, was whether or not a set of factors from the 12 candidate variables described in chapter three (section 3.3.2) could help explain the low-school-attainment landscape in LPS. And if so, whether or not GWR could improve upon the OLS results in LPS as it did in the case of OPS.
5.8 The Lincoln Public School Low-school-attainment landscape

The ACS estimated that in 2008 – 2012 there were approximately 51,254 residents with low school attainment living within the LPS district boundary. There was a fairly wide range of low school attainment numbers (20 to 1,802) among the census tracts comprising LPS and the average census tract in LPS was home to an estimated 702 (SD = 408.65) residents whose formal schooling ended at or before the high school level.

According to the results of the hotspot tool, the range of z-scores (z = -2.97, p < 0.00 to 2.86, p < 0.00) indicated spatial clustering of census tracts with low school attainment numbers both above and below the mean. Figure 5.7 below shows the LOWSCH landscape for LPS. From this image, it appears that the most concentrated clustering of low school attainment occurred in the northern and central portions of the district, and spread to the northeast. This hotspot ran more or less diagonally (from southwest to northeast), essentially bifurcating the district. This hotspot was also most intense south/southeast of I-80, but decreased in significance near the city core (i.e., downtown). In the eastern, southeastern, and southern portions of LPS, the pattern of census tract clustering changed direction, and a significant coldspot emerged for most of the southern portion of the district.
In LPS, the most significant clustering of low school attainment appeared southeast of the interstate, which was similar to the finding from the OPS analysis. In Figure 5.7 there was also a fairly pronounced change in school attainment in LPS census tracts as distance from I-80 increased. In addition, this map suggests that there was an obvious relationship between the geographic space of the district and low school attainment, wherein residents who had not attempted college were generally concentrated in the census tracts in the northern portion of the city, and those who had at least attempted college were located in the south/southeastern area of the district. These results indicated that spatial nonstationarity exists in the variables/relationships that helped to explain the LOWSCH patterns in Figure 5.7.
5.9 Lincoln Public Schools-Exploratory and OLS Results

The exploratory regression tool in ArcGIS v. 10.1 identified a four-variable model that passed all the statistical tests (discussed in section 3.4.4) for violations of regression assumptions. The final model used for OLS analysis in ArcGIS included the following variables: \textit{AG25UP90, H30OLD90, NHPOV90, DistI80} (see Table 5.6 below for definitions). Per the \textit{a priori} model-building criteria set forth, each independent variable needed to be significantly related to low school attainment, which was the case. The highest variance inflation factor (VIF) score was 2.75, meaning that overlapping variables was not considered an issue. In addition, model residuals were tested using the Jarque-Bera statistic for normality and the global Moran’s \textit{I} for spatial autocorrelation and based on these tests I determined that model residuals were not significantly biased, \(p > 0.10\), nor were they correlated spatially, \(p > 0.10\).

Table 5.6: LPS Model Regression Coefficient Statistics

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Statistic</th>
<th>(p) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>431.24</td>
<td>6.04</td>
<td>.00</td>
</tr>
<tr>
<td>AG25UP90</td>
<td>0.29</td>
<td>7.47</td>
<td>.00</td>
</tr>
<tr>
<td>NHPOV90</td>
<td>5.23</td>
<td>3.01</td>
<td>.00</td>
</tr>
<tr>
<td>H30OLD90</td>
<td>-0.20</td>
<td>-2.30</td>
<td>.02</td>
</tr>
<tr>
<td>DISTI80</td>
<td>-0.03</td>
<td>-4.15</td>
<td>.00</td>
</tr>
</tbody>
</table>

Note: \textit{AG25UP90} = number residents in a census tract in 1990 who were 25 years-old or older; \textit{H30OLD90} = number of houses in 1990 that were at least 30 years-old; \textit{NHPOV90} = Number of Hispanic/Latino residents in poverty in a census tract in 1990; \textit{DISTI80} = distance in meters from the centroid of a census tract to I-80.
Table 5.7: Pairwise Pearson Correlation Coefficients for Twelve Candidate Variables in OPS

<table>
<thead>
<tr>
<th></th>
<th>LOWED</th>
<th>Dist80</th>
<th>DistCore</th>
<th>HHH90</th>
<th>HHB90</th>
<th>M_INC90</th>
<th>H30OLD90</th>
<th>NHPOV90</th>
<th>NBPOV90</th>
<th>NPOV90</th>
<th>UNEMP90</th>
<th>AG25UP90</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOWED</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist80</td>
<td>-0.29</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DistCore</td>
<td>-0.22</td>
<td>0.62</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HHH90</td>
<td>0.45</td>
<td>-0.27</td>
<td>-0.58</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HHB90</td>
<td>0.35</td>
<td>-0.18</td>
<td>-0.51</td>
<td>0.83</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M_INC90</td>
<td>-0.31</td>
<td>0.53</td>
<td>0.69</td>
<td>-0.57</td>
<td>-0.53</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H30OLD90</td>
<td>0.47</td>
<td>-0.09</td>
<td>-0.40</td>
<td>0.60</td>
<td>0.50</td>
<td>-0.42</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHPOV90</td>
<td>0.38</td>
<td>-0.22</td>
<td>-0.50</td>
<td>0.73</td>
<td>0.79</td>
<td>-0.51</td>
<td>0.51</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NBPOV90</td>
<td>0.28</td>
<td>-0.25</td>
<td>-0.41</td>
<td>0.66</td>
<td>0.89</td>
<td>-0.51</td>
<td>0.37</td>
<td>0.65</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPOV90</td>
<td>0.43</td>
<td>-0.24</td>
<td>-0.55</td>
<td>0.79</td>
<td>0.89</td>
<td>-0.65</td>
<td>0.61</td>
<td>0.83</td>
<td>0.83</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNEMP90</td>
<td>0.35</td>
<td>-0.21</td>
<td>-0.49</td>
<td>0.46</td>
<td>0.45</td>
<td>-0.48</td>
<td>0.45</td>
<td>0.48</td>
<td>0.37</td>
<td>0.55</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AG25UP90</td>
<td>0.66</td>
<td>0.08</td>
<td>-0.13</td>
<td>0.36</td>
<td>0.25</td>
<td>-0.07</td>
<td>0.71</td>
<td>0.22</td>
<td>0.11</td>
<td>0.35</td>
<td>0.37</td>
<td>1</td>
</tr>
</tbody>
</table>
The analysis of variance (ANOVA) test suggested that overall the LPS-OLS model was statistically significant, $F(4, 68) = 27.08$, $p < 0.000$, and the adjusted $R^2 = 0.59$, indicated that 59% of the variability in the current LPS low-school-attainment landscape could be explained by the combination of the four factors described above. These results supported the conclusion that (generally speaking) $LOWSCH$ was positively associated with the number of residents in 1990 (ages 25 and over), the number of Hispanics/Latinos living in poverty in 1990, and negatively associated with the number of houses in 1990 that were at least 30 years-old, and with distance from the interstate. These results for the LPS model share two variables in common with the OPS model described above. Those variables were $H30OLD90$ and $DistI80$. The $NHPOV90$ measured the number of Hispanics/Latinos living in poverty 1990, rather than the number of Hispanic/Latino households. So an important caveat between Lincoln and Omaha emerged, in that for OPS, it was the number of 1990 Hispanic/Latino households that was associated with $LOWSCH$ and in Lincoln it was the number of poor Hispanic/Latino people that significantly predicted $LOWSCH$.

Another important distinction between LPS and OPS was that for LPS, the 25 and over population in 1990 was a strong factor for predicting current low school attainment, $b = 0.29$, $t(68) = 7.47$, $p < 0.000$. This means that for LPS, the more people ages 25 and over living in a census tract in 1990, the higher the $LOWSCH$ numbers for that tract. In the case of OPS, recall from above that $AG25UP90$ was weakly tied to low school attainment, whereas in LPS, this link appeared to be stronger.
5.10 Lincoln Public Schools GWR Results
Figure 5.8: People 25 years-old or older in 1990 as a Predictor of Current Low School Attainment
Figure 5.9 Hispanics/Latinos in Poverty in 1990 as a Predictor of Current Low School Attainment
Figure 5.10 Houses 30 years-old or Older in 1990 as a Predictor of Current Low School Attainment

Lincoln Public Schools GWR Parameter and Isoline Map:
Current Low School Attainment and H30OLD90

GWR Results
- t-value contour line

Parameter Estimates
- High: 0.0246901
- Low: -0.425452

Legend:
- Red: High
- Blue: Low
- Black: City/City Core
- Gray: Census Tract Boundary

Author: William R. England
Source: Minnesota Population Center
National Historical Geographic Information System, Version 2.0
University of Minnesota; 2011
http://www.mnhgis.org

Lincoln Public Schools:
Spatial Distribution of Houses 30 years-old or older in 1990 (H30OLD90)

Lincoln Public Schools:
Spatial Distribution of ACS Estimated (2008 - 2012) Low School Attainment

Standard Deviations:
- > 2.7 Std. Dev.
- 2.2 - 2.7 Std. Dev.
- 1.7 - 2.2 Std. Dev.
- 1.2 - 1.7 Std. Dev.
- 0.75 - 1.2 Std. Dev.
- 0.25 - 0.75 Std. Dev.
- -0.25 - 0.25 Std. Dev.
- -0.75 - 0.25 Std. Dev.
- -1.2 - -0.75 Std. Dev.
- < -1.2 Std. Dev.
Figure 5.11: Distance from I-80 as a Predictor of Current Low School Attainment
The LPS-GWR model’s adjusted $R^2$ was 0.69, indicating that the GWR model could account for more variability in current low school numbers than the OLS model. The results from the first two maps (Figures xx and xx) suggest that the relationship between current low school attainment and both $AG25UP90$ and $NHPOV90$ were positive and significant throughout the LPS district. For $AG25UP90$, the global model results likewise indicated that generally the relationship between the 25 and over population in a 1990 census tract was related to its current number of residents with low school attainment, $b = 0.29$, $t(68) = 7.47$, $p < 0.000$. However, the local regression parameters range from 0.49 to 0.22, and these parameters were statistically significant throughout the LPS district. Similarly, the global model found that $NHPOV90$ was positively and significantly correlated with $LOWSCH$, $b = 5.23$, $t(68) = 3.01$, $p < 0.00$. GWR results indicated that while this relationship was positive and significant throughout LPS, the regression parameters may have ranged from as high as 6.88 to as low as 3.86.

Considering the spatial variations revealed by these first two maps, the global OLS results appeared to generally apply to the entire area of study, but in some areas of LPS these relationships shared a somewhat stronger/weaker connection than what the OLS regression suggested. For example, the correlation coefficients for $NHPOV90$ appeared to be most intense in southeastern LPS, where in 1990 there were relatively few Hispanics/Latinos living in poverty, and currently there were relatively fewer residents with low school attainment. In central and northeastern LPS, $NHPOV90$ was still significantly related to current low school attainment, but the association between these
two variables may not have been quite as pronounced relative to the southeastern portion of the district.

A similar pattern emerged for \textit{AG25UP90} as well. It appeared to be case that throughout LPS, the more people 25 and over there were in a census tract 1990, the higher the \textit{LOWSCH} numbers would be. But, while significant throughout LPS, the \textit{AG25UP90/LOWSCH} link appeared to be stronger in northern and central LPS, than in the southern/southeastern portions of the district.

The maps in Figures 5.10 and 5.11 tell a different story. In both of these images there was obvious spatial nonstationary, wherein regression parameter estimates for both \textit{H30OLD90} and \textit{DistI80} appeared to transition from significantly negative, to zero (unrelated), to positive (but not significantly so). For \textit{H30OLD90}, the OLS model suggested a significant, negative parameter estimate, $b = -0.20$, $t(68) = -2.30$, $p < 0.02$, indicating that for census tracts in the LPS district, lower \textit{H30OLD90} values were linked with higher values of \textit{LOWSCH}. The GWR results appeared to correspond with the OLS model only in the north/northwestern portion of the district’s census tracts, where more new housing in 1990 (especially near I-80) appeared to be significantly related to \textit{LOWSCH}. The global result also implied the converse; older housing in 1990 was related to lower \textit{LOWSCH} numbers. However, the GWR results suggested that this negative relationship substantially weakened across the southern/southeastern/eastern census tracts of LPS, eventually turning positive (but not significant).

The image in Figure 5.11 depicting the spatial relationship between \textit{DistI80} and \textit{LOWSCH} was similar in strength and directionality to the spatial pattern for \textit{H30OLD90} and \textit{LOWSCH}. Meaning, in the northern/northeastern portions of the district, the GWR
results indicated that DistI80 was negatively associated with LOWSCH. Hence, in this portion of the district, as the distance to I-80 decreases, LOWSCH numbers tended to increase. This finding was in good agreement with the OLS results, which suggested that DistI80 was negatively related to LOWSCH for all of LPS, $b = -0.03, t(68) = -4.15, p < 0.00$. However, the strength of this relationship appeared to dissipate in a southerly direction across the district, and the link between DistI80 and LOWSCH doesn’t seem to hold together for southern/southeastern LPS (where the relationship is actually positive, but not significant).

5.11 Performance of the LPS-OLS and GWR Models

Table xx below provides the necessary information for determining whether the GWR or the OLS model best fits the data representing the LOWSCH landscape in LPS. Recall that Burnham, et al. (2011), and Burnham and Anderson (2002) suggested that models with $\Delta_i > 10$ tend to: (1) provide no little or no support for the analysis, or (2) fail to explain some substantial variable in the data.

<table>
<thead>
<tr>
<th>Model</th>
<th>AICc Value</th>
<th>$\Delta_i$</th>
<th>Model Likelihood ($l_i$)</th>
<th>Akaike Weight ($w_i$)</th>
<th>Evidence Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWR</td>
<td>1013.77</td>
<td>1.00</td>
<td>0.999444</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>1028.76</td>
<td>14.99</td>
<td>.0005559</td>
<td>0.000556</td>
<td>1797.5612</td>
</tr>
</tbody>
</table>

Recall that as $\Delta_i$ increases to 10 and beyond, a model’s likelihood (i.e., its plausibility as the best model) decreases rapidly. In the case of LPS, the OLS model’s $\Delta_i$
value was equal to 14.99, meaning that it apparently lost more information as a
description of reality than did the GWR model. Hence, there is very little chance that the
global model is the better of the two choices. As was the case with OPS, once again the
GWR model appears to be the more useful representation of reality.

5.12 Summary

In OPS, low-school-attainment numbers in 1990 (HS90) were positively
correlated with low-school-attainment levels more recently (LOWSCH), indicating that
census tracts with large numbers of less-educated residents in 1990 tended to have large
numbers of uneducated residents more recently. This relationship was significant
throughout the OPS district, but was the most intense in southern OPS, especially for
census tracts south of I-80. Among all the variables in OPS related to recent low school
attainment, the past school attainment factor had the most uniform predictability across
OPS, and thus the link between the past and the present in terms of school attainment was
comparatively spatially homogenous.

There was obvious spatial nonstationarity in the parameters estimating the
correlation between the number of Hispanic/Latino households in 1990 and recent low
school attainment. Spatially speaking, the relationship between Hispanic households in
1990 and low school attainment was most intense in southern OPS near I-80. There the
relationship between Hispanic households in the past and recent low school attainment
was strong and positive, indicating that census tracts with large numbers of Hispanic
households tended to have large numbers of low school attainment recently. North of I-
80, the link between Hispanic households and recent low school attainment, while
positive and significant, gradually dissipated. In central OPS, the connection between
these two variables was no longer statistically significant, and west of 72nd St, the parameter estimates changed completely from positive to negative, indicating that for western OPS, an increase in Hispanic Latino households numbers was associated with a decrease in low school attainment, though this relationship was not significant.

There was also spatial nonstationarity present in the relationship between the number of houses in a 1990 census tract that were 30-years-old or older (H30OLD90) and low school attainment recently. In southern OPS, south of I-80, this relationship was negative and there it was most intense. This indicates that in southern OPS, an older housing stock was associated with a decrease in low school attainment. Immediately north of I-80 and ringing the downtown area was a pocket where regression parameters reversed signs, indicating that near downtown older houses tended to be associated with higher levels of low school attainment, though this relationship was not significant. In central OPS, the association between an older housing stock and recent school attainment switched back to negative again, before switching for a third time to positive (but again not significant) in northern and parts of northwestern OPS.

The relationship between a census tract’s distance to I-80 (DistI80) and recent school attainment followed a trend that was somewhat similar to the patterns of spatial nonstationarity in the relationship between the age of the housing stock and recent low school attainment. In northern and northwestern OPS there was a positive and significant connection between the distance (in meters) from the centroid of a census tract to I-80 and low school attainment. This indicated that for most of north/northwestern OPS an increase in distance from I-80 corresponded with an increase in low school attainment. In central OPS, especially near downtown, this relationship reversed, and a decrease in the
distance to the interstate was correlated with an increase in low school attainment (but this was not significant). Then, south of I-80 the link between DistI80 and LOWSCH switched back to positive (but not significant).

The OPS global OLS model (OPS-OLS) was tested on low school attainment data for LPS to assess the ‘interurban portability’ of the model. While the OPS-OLS model could explain a significant portion of the recent low school attainment landscape in LPS, this model violated the regression assumption that a model’s residuals are randomly distributed. Hence, the process was repeated from the beginning for LPS, and a new global model was constructed for LPS (LPS-OLS) consisting of two variables that were the same as those from the OPS-OLS (the age of the housing stock and distance to the interstate) and two variables that were different (number of residents over 25 and the number of Hispanic/Latinos in poverty).

In LPS, the relationship between the population over 25 in 1990 and recent low school attainment was positive and significant throughout the district. The same was true for the association between Hispanics in poverty in 1990 and recent low school attainment. These factors indicated that throughout LPS, more populated census tracts in 1990 and those with larger numbers of Hispanics/Latinos in poverty tended to have higher low school attainment numbers more recently.

However, where the age of the housing stock and distance to the interstate were concerned, there was more spatial heterogeneity in the relationships to recent low school attainment. The connection between the age of the housing stock and low school attainment followed a northeast to southwest diagonal line, more or less bisecting the district. In the northwest, the age of the housing stock was negatively and significantly
connected to recent low school attainment, meaning that in this portion of LPS, larger numbers of older housing in 1990 was linked with lower levels of low school attainment more recently. This relationship maintained its strength and significance throughout parts of central, southwestern, and northeastern LPS, but eventually dissipated and turned positive (but not significant) in southeast Lincoln. This is almost the exact same result for the relationship between distance to the interstate and low school attainment. In the southwestern, western, northwestern, northern, and northwestern areas of LPS, there was a negative and significant relationship between distance to I-80 and low school attainment. This suggests that in these areas, a decrease in distance to I-80 was associated with an increase in low school attainment levels. The link between these two variables dissipated in similar fashion to the link between the housing stock and low school attainment, and in southeastern LPS, the DistI80—LOWSCH correlation turned positive but not significant.

In comparing GWR performance to OLS, in both OPS and LPS I found that the GWR model explained more variability in the low school attainment data and fit this data better than the OLS model did. The OPS-OLS model was able to explain 75% of the variability in the low-school-attainment landscape in Omaha. The GWR model for OPS (OPS-GWR) accounted for 79%. However, according to the difference in AICc scores between the OPS-OLS and OPS-GWR, the GWR model was a better fit for the low school attainment data, and it was much more likely to be the superior model.

For Lincoln, the LPS-GWR model was able to explain 69% of the variability in the low-school-attainment landscape, ten percent more than the LPS-OLS model. The difference between the AICc scores for the LPS-GWR and LPS-OLS models was
comparatively larger for Lincoln than for Omaha. In the LPS case, the GWR model was roughly 1,800 times more likely to be the better-fitting model.
CHAPTER SIX: DISCUSSION AND CONCLUSIONS

6.1 Introduction

This chapter is arranged in three major sections which are further divided into several subsections. In the first section (6.2) I discuss a few of the general concerns I grappled with in using a set of past predictor variables to explain recent low school attainment in OPS. I decided to include this section mostly to explain in more detail some of the thought processes that went into the decisions for the selection of the predictor variables, and why I used 1990 variables as opposed to more recent ones. In the next section (6.3), I will address the first three of the four questions that began this study: How much of the variability in the current low educational landscape in an urban area can be explained by a set of variables from the past (6.3.1)? Does the GWR technique do a better job than OLS of modeling the relationships between past community-wide demographic, housing, education, and economic conditions and the current low-school-attainment landscape in a given urban area (6.3.2)? Does the same set of variables related to low education attainment in one urban area apply to another demographically and geographically similar urban area (6.3.3)? In section (6.4), I will address the fourth study question by discussing what policy implications arise from the presence of spatial variability in the strength of the relationships that predict low school attainment. In the next section (6.5), I provide a summary of the dissertation, and in the final section (6.6) I describe how I see my dissertation work fitting into the ‘big picture’ of education research.
6.2 Predicting the Present

“We may be through with the past, but the past is not through with us.” (Evans, 1946, pg. 5).

There was an interesting dilemma that I encountered early in the data gathering phase of this study, and it is worth mentioning because this dilemma created for me a set of difficult questions that required a fair amount of thought and research before I felt comfortable in proceeding. The dilemma was a temporal one, related to the selection of the census and ACS variables. Recall that one way the Census Bureau/ACS captures low school attainment is by sampling the adult population ages 25 and over. There is also an 18-year-old or older category. However, many 18-year-olds don’t graduate on time but still do graduate (in OPS this number is around 7-10% each year). In addition, it is conceivable that many 18-, 19-, and 20-year-old high school graduates attend college after a delay, opting instead for military service right out of high school, or joining the workforce to save for college. One benefit to using the 25-year-old cutoff is that it provides a somewhat surer basis for capturing those people whose formal schooling really has ended (at least for the time being) at or before the completion of high school. I think this choice for measuring the dependent variable was defensible, but choosing factors to explain the low-school-attainment numbers that exist within this 25 and over group was more fraught.

Social theory (e.g., Bourdieu and Passeron, 1977; Lamont and Lareau, 1988) and much of the empirical data (e.g., Alexander, et al. 1997; Ensminger & Slusarcick, 1992; Jimerson, et al. 2000; Lloyd, 1978) seem to be in good agreement: the factors that bear the most weight on an adult’s schooling trajectory formed long before the present, or even the recent past. Alexander et al. (1997), and Ensminger and Slusarcick (1992),
found that at least some of the factors influencing how long we persist in school can emerge as early as first grade. Lloyd (1978) also found that by third grade discernable patterns emerged between/among certain kinds of students from distinct backgrounds that were predictive of dropping out of high school.

These three studies (Alexander et al., 1997; Ensminger & Slusarcick, 1992; Lloyd, 1978) were each longitudinal, each used data collected from the efforts of researchers following the development of schooling pathways for more than 700, 1200, and 1500 students (respectively), and combined these studies provide strong evidence supporting the theory that current school attainment patterns are largely the product of factors from the past. Put another way, the current low-school-attainment landscape should be a recapitulation of past social stratification and inequality (Bourdieu and Passeron, 1977). From this perspective, when/where we finish our formal schooling is largely dependent upon our childhoods—i.e., on our school achievement and personal expectations, on our parents’ educations, incomes, attitudes, and expectations for us, on our home environments and socialization as children and adolescents, on our race/ethnicity and the political power available to the racial/ethnic group(s) we are born into, and on our friends and neighbors/neighborhoods (Alexander, et al. 1997; Bourdieu & Passeron, 1977, Ensminger & Slusarcick, 1992; Jimerson, et al. 2000; Lamont and Lareau, 1998; Lloyd, 1978).

Given the agreement between social reproduction theory and much of the empirical evidence predicting school attainment, looking for connections between demographic variables from the 1990 Census and the current low-school-attainment landscape in OPS made good sense to me. However, this study necessarily omitted
several possibly important (but difficult to find) factors like migration in/out of Omaha, kids going off to college in other cities and towns, finding jobs in other neighborhoods, etc. So, are the patterns observed in southern Omaha or northern/central Lincoln the result of parents passing on their formal education levels to their children who remain stuck in those areas, or is that these geographic spaces in Lincoln and Omaha simply play host to poor, often minority and immigrant populations as these populations transition (eventually) into better neighborhoods or different towns?

If social reproduction theory held in OPS, and school attainment persisted down through the generations, and if people more or less remained where they were, then in OPS there should not have been much (if any) widespread systematic changes to the low-school-attainment landscape over time. But the same pattern would also be observed if, for example, one wave of Hispanic/Latino immigrants sometime before 1990 with low school attainment, but their offspring moved on to post-secondary schools, new towns, or better neighborhoods. Then, between 1990 and 2008 – 2012, more newcomers with low school attainment arrived to these areas, thus driving the link I found between Hispanic/Latino households and low school attainment in south Omaha over the span of the last quarter century.

There are also some problems with the ‘past-predicts-the-present’ framework in general. For example, there is good reason to believe that when it comes to persisting in/returning to school, it is the present (not just the past) that explains a given low-school-attainment landscape. Not only does our past influence our schooling trajectory, but at least some of us make decisions about persisting in/returning to school based on the state of the economy, our current job prospects (or lack thereof), and our perceived
improvements in opportunities that we believe might come with returning to/persisting in school.

Broad evidence supporting this line of reasoning comes from the fact that U.S. college enrollment for people 18 to 24 hit an all-time high in 2008 (Fry, 2009). This mass return to school coincided with, and was spurred on by, the near collapse of the economy in 2008, and the Great Recession that ensued. High unemployment and the hope for more opportunities and greater earning potential apparently inspired a massive, nationwide back-to-school movement. This is also captured in an annual report produced by The College Board, titled “Trends in Student Aid” (2013), which showed that total Pell Grant expenditures increased by 94% between 2008-09 and 2010-11—from $20.4 billion (in 2012 dollars) to $39.5 billion (also in 2012 dollars)—and have been on the decline ever since the economy began to recover. According to a 2013 report from the National Student Clearinghouse, from fall 2011 to fall 2012 college enrollment fell 1.8%, and then fell another 1.5% from fall 2012 to 2013. This constituted a net-loss of around 600,000 college students in two years; apparently, as the economy goes so too goes college enrollment rates (Current term enrollment report, 2013).

The fact that for many people the decision to continue in or return to school is influenced by current economic conditions stands as a compelling complication for strict adherence to social reproduction theory (and also to searching the past for OPS low school attainment predictors). It seems as though it isn’t just the past that predicts our decisions to persist in/return to school, but apparently, current economic conditions factor into this decision as well. This of course would be a much stronger argument if it wasn’t also the case that low-income, minorities, while persisting in/returning to school in larger
numbers during the 2008 recession, still trailed middle- and upper-income whites in college enrollment (Desilver, 2014).

Nevertheless, poor economic conditions may not disrupt unequal proportionality or stratification in college attendance, but poor economic conditions do create large back-to-school movements even among poor and minority groups, which does at least complicate social reproduction theory. Nevertheless, the majority of empirical evidence that I could find with regard to predicting a person’s school attainment tended to align with the logic hinted at in the Bergen Evans quote at the beginning of this section. That is, when it comes to explaining recent low school attainment patterns, a good place to begin looking is in the past.

But even if true, this logic raises more troubling questions. For example, if the variables that best explain low school attainment occurred in the past, then aren’t those variables also beyond our manipulative powers? For instance, Alexander, et al. (1997) found that among other factors, past school achievement and parental expectations were strongly tied to a child’s trajectory in school. But if a person’s parental expectations and academic achievement negatively impacted his eventual progress in school, how do we then (absent the invention of time travel) change his parents’ expectations and/or his academic achievement when these existed in the past? Doesn’t a ‘past-predicts-the-present’ logic also require us (at least to some degree) to pessimistically turn away from those who have already completed their schooling trajectories, in order to focus our resources, finite as they are, and policy reforms/interventions on the current cohort of students just beginning in school? And, since a new policy reform/intervention must be enacted in the present, mustn’t a policy/reform based on a ‘past-predicts-the-present’
logic also, by necessity, take a leap of faith and assume that the past predictors of low school attainment will continue to predict the schooling trajectories for the present-day group of students who (optimistically) are to be the beneficiaries of our reforms and interventions?

Because I was interested in looking into the more distant past to try and explain/predict the present, what I have done with the data collected for this dissertation is sound, but by using the ‘past-predicts-the-present’ framework I have not established a firm basis for interpreting what will happen in the next 20 years. It might behoove the skeptic to point out why different outcomes (going forward) seem plausible. In other words, is the status quo likely to remain intact? If not, what’s likely to be different?

These are tough questions, and I include them here because like Jimerson, et al. (2000), when it comes to explaining recent low-school-attainment numbers, I too find the predictive power of the past to be impressive. But (and again like Jimerson, et al. 2000), I also do not want appear as though I have uncritically come to the conclusion that when it comes to our persistence in school we can always find in the past the launching points that sent us along our schooling pathways. On the contrary, I find it too unrealistic to claim that the road to low/high school attainment is determined in the first few years of school (or even before) as some of the aforementioned research (and perhaps my study as well) seems to imply. It is much more likely that predictors from the past influence us in probabilistic ways, and that there are people who both succumb to and beat the odds. Ultimately though, I chose to examine 1990 census predictor variables, and to use the ‘past-predicts-the-present’ framework, because there was simply too much evidence to
ignore suggesting that when it comes to persistence in school, too often and for too many
“what is past is prologue” (Shakespeare, n.d.).

6.3 Study Questions

6.3.1 How much of the variability in the current low-school-attainment landscape in an urban area can be explained by a set of variables from the past?

To provide an accurate response to this question some circumspection is in order. First, the answer is dependent upon the scale/areal unit of analysis. I used census tract-level aggregations, but had I used individual-, household-, or block-level data it is quite likely that the results and conclusions would be different. This is a result of the ecological fallacy mentioned in previous chapters—i.e., the fact that study conclusions drawn from data aggregated at a particular areal extent cannot be assumed to apply to individuals or to levels of aggregation above or below the areal extent used in that specific study (Steel & Holt, 1996; Qui & Wu, 2014). Due to federal privacy laws, the U.S. Census Bureau doesn’t publish much data below the census tract-level, and when such data are published, at least some of it is masked or in some other way changed to avoid the possibility of identifying someone. The only way to obtain accurate block-, household-, or individual-level data is to gain security access to a Census Bureau Research Data Center (of which there are currently eighteen in the U.S., most of which are near the east and west coasts). Hence, the census-tract level data was the most granular, publicly available data that I could obtain in my effort to answer the study questions (“granular” in the case of OPS means census tracts with population sizes anywhere from approximately 1100 to 6500). It should be emphasized that any claim about explaining variability in low school attainment numbers references the variability among the 124 census tracts in OPS,
not among the 95,000+ individuals in the OPS district who the ACS estimates (2008-2012) have never been to college. Hence, it is unclear whether the phenomena observed in this study are inherently demographic, geographic, or both.

As a second point, in examining the OPS district I found a combination of four variables from the 1990 census which explained 75-79% of the variability in recent low-school-attainment numbers for OPS. For LPS I was able to account for 59-69% of the variability in the low-school-attainment landscape using a similar set of 1990 predictors. Using these data and modeling methods led to more variability being account for in OPS than for LPS. This result indicates that the answer to the study question above depends not only on the areal unit of analysis, but also the urban area being examined—i.e., scale and location.

There are potentially many more considerations worth exploring in future inquires (history and context in low-school-attainment hotspots for example), and it is conceivable that some of these may further change the calculus to answering the question of how much variability can be explained in the low-school-attainment landscape of an urban area by demographic/geographic factors. Time is an example of another important concern.

As discussed in the previous section, the economic climate in the U.S. writ large matters with respect to large portions of the population going back-to-school. Earlier in this chapter I provided an example of a nationwide back-to-school movement occurring with the Great Recession in 2008. Another example exists as well. Walsh (1993) pointed out that real GDP in the U.S. grew from 1982 to 1990 at an annual rate of 3.3%, which was one of the longer periods of peacetime economic growth in U.S. history. But this
growth halted (and eventually contracted) in the fall of 1987, when world stock markets crashed on October 19th on a day that has been dubbed ‘Black Monday’. A detailed treatment of this crash and the subsequent recovery are too far afield to be productive for this discussion, but broadly speaking they are relevant because if economic downturns are tied to increases in college enrollment, and conversely, subsequent economic upturns are related to lower enrollment, then beginning just before 1990 there ought to have been a large uptick in college enrollment.

According to U.S. Census Bureau data (shown in Figure 6.1 above) there was slow growth in college enrollments between the fall semesters of 1982 and 1987, just before Black Monday in mid-October. During this six-year period, college enrollment increased by some 300,000 students (roughly 50,000 per year during this period). By the following fall semester college enrollment increased by 480,000 students—meeting and eclipsing the preceding 6-year growth by an additional 60%. Then, from the fall of 1989 to 1992, another 950,000 students enrolled in college (roughly 240,000 per year during this period). Walsh (1993) argued that this recession lasted less than one fiscal year (approximately 8 months), but that the recovery that followed was sluggish. This
argument is reflected in Figure 6.1 above, where in the fall of 1992 college enrollment had peaked to nearly 14.5 million, and then began to slowly decrease over the next several years.

The reason for my dwelling on this point is that by using 1990 school attainment data and relating that to 2008 - 2012 estimates of school attainment, I essentially analyzed and compared two ‘snapshots’ at those two points in time. However, there have obviously been large-scale, but also short-term and long-term fluctuations in play where college attendance is concerned. I have pointed out two major back-to-school movements in the U.S., both of which occurred in the wake of an economic recession, and then receded in the wake of the recovery that followed. Thus, in response to the first study question—and in concluding so far that explaining variability in school attainment probably depends on the scale and location of the analysis—it seems as though a third caveat must be included to these two: explaining variance in school attainment numbers is probably a temporal matter as well. It is therefore likely the case that had I used 1980 or 2000 census data (rather than 1990 numbers) to explain the variability in the OPS low-school-attainment landscape, the model variables and diagnostics (e.g., AICc and $R^2$ values) would not be the same as those I found.

In conclusion, the answer to the first study question, while perhaps hyper-specific, is that for OPS, I was able to identify a fairly limited set of 1990 factors from the Census that an OLS model indicated could explain 75% of later (i.e., 2008 - 2012) variability in the low-school-attainment landscape for OPS. Using those same variables, I was able to use a GWR model to explain 79% of the OPS low-school-attainment landscape. In LPS, I was able to identify a similar set of 1990 factors that according to OLS results explained
59% of low-school-attainment landscape. Using those same variables, I was able to apply a GWR model to explain 69% of the low-school-attainment landscape for LPS. Generally speaking, though, there may not be a broadly applicable answer to this question because any response is dependent on the time, location, and scale of the analysis.

6.3.2 Does the GWR technique do a better job than OLS of modeling the relationships between past community-wide demographic, housing, education, and economic conditions and the current low-school-attainment landscape in a given urban area?

The brief answer is ‘yes’. Table 6.2 (below) provides a comparison between the results of the OLS and GWR models for the Lincoln and Omaha school districts. For both OPS and LPS, the GWR model explained more variability in low-school-attainment numbers, and according to the AICc scores, GWR in each case resulted in a better fitting model for both the OPS and LPS datasets.

These results are not particularly surprising. The maps in the previous chapter showed that in OPS there was considerable spatial variability in the strength and significance of the relationships between/among 1990 factors and the recent low-school-attainment landscape. This was the case for LPS as well. Where spatial nonstationarity is present in a dataset, GWR has consistently been found to outperform global methods (Chi, et al. 2013; Fotheringham et al., 2001; Lersch and Hart, 2014; Papandreou and Tuomilehto, 2014; Pasculli, et al. 2014; Partridge, et al. 2008; Qui and Wu, 2011; Slagle, 2010).
Table 6.2: Low School Attainment Model Comparison

<table>
<thead>
<tr>
<th></th>
<th>OLS Parameter Estimates</th>
<th>GWR Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OPS Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS90</td>
<td>0.7</td>
<td>0.55 to 0.89</td>
</tr>
<tr>
<td>HHH90</td>
<td>3.92</td>
<td>-6.23 to 6.27</td>
</tr>
<tr>
<td>H30OLD90</td>
<td>-0.13</td>
<td>-0.21 to 0.01</td>
</tr>
<tr>
<td>DistI80</td>
<td>0.01</td>
<td>-0.02 to 0.05</td>
</tr>
<tr>
<td>Intercept</td>
<td>109.22</td>
<td>-36.12 to 300.66</td>
</tr>
<tr>
<td>Adj. R-Squared</td>
<td>0.75</td>
<td>0.79</td>
</tr>
<tr>
<td>AICc</td>
<td>1679.85</td>
<td>1668.71</td>
</tr>
<tr>
<td><strong>LPS Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG25UP90</td>
<td>0.29</td>
<td>0.22 to 0.46</td>
</tr>
<tr>
<td>NHPOV90</td>
<td>5.23</td>
<td>3.93 to 6.79</td>
</tr>
<tr>
<td>H30OLD90</td>
<td>-0.2</td>
<td>-0.41 to 0.01</td>
</tr>
<tr>
<td>DistI80</td>
<td>-0.03</td>
<td>-0.08 to 0.02</td>
</tr>
<tr>
<td>Intercept</td>
<td>431.24</td>
<td>-92.33 to 514.43</td>
</tr>
<tr>
<td>Adj. R-Squared</td>
<td>0.59</td>
<td>0.69</td>
</tr>
<tr>
<td>AICc</td>
<td>1028.76</td>
<td>1013.77</td>
</tr>
</tbody>
</table>

Based on these results it isn’t far-fetched to conclude that using geographic modeling strategies has the potential to enhance educational research. Furthermore, we should expect GWR to represent a ‘truer’ version of reality given the presence of spatial nonstationarity in a dataset. Recall that the OLS results suggested that in OPS low school attainment in 1990 was strongly connected to the recent low-school-attainment landscape. GWR confirmed this, but with the caveat that variations in the predictive power of 1990 low-school-attainment patterns in the OPS district were a matter of degree. But this was a very different result than those for the relationship between the number of Hispanic/Latino households in 1990 and low school attainment. Here OLS results suggested that an increase in the number of Hispanic/Latino households in a 1990
census tract was associated with higher numbers of people who more recently have never been to college. By necessity, this OLS result was a description of all of OPS. But GWR results indicated that the relationship between those two variables was really only true for southern OPS. In central OPS, the connection between Hispanic/Latino households in 1990 and the recent low-school-attainment landscape fell apart. Hence, outside southern OPS the number of Hispanic/Latino households no longer predicted recent low school attainment. And perhaps more intriguingly, in western OPS the relationship between Hispanic/Latino households appeared to changed direction. Meaning, more Hispanic/Latino households in 1990 might actually be associated with higher education. And while t-scores associated with this reversal in the western OPS area were not significant, it may very well be the case that in western OPS there are enough college-educated Hispanic/Latino residents (especially compared to southern OPS) that the general trend detected by the global model was disrupted in this area. It seems plausible that if those Hispanics/Latinos identified within the low-achieving census tracks in 1990 fared well economically and educationally, they may have moved to more prosperous West Omaha. Or, relatively educated Hispanics/Latinos who have moved to Omaha since 1990 have opted to reside in the West. In both cases, the South Omaha relationship between the 1990 Latino presence and low school attainment would be sustained, while the reverse would occurred (albeit at a smaller scale) in West Omaha. So, an important question that has emerged (which will be addressed below) in the analysis of OPS is: why is low school attainment so predictable in southern OPS and less predictable everywhere else?
Ultimately, that GWR was a better fit for OPS and LPS low school attainment data suggests that (at least for OPS and LPS) there is such thing as a ‘geography of school persistence’. Furthermore, for these two school districts, space is clearly in play as a factor for better explaining/predicting the observable patterns of low school attainment.

6.3.3 Does the same set of variables related to low-school-attainment in one urban area apply to another demographically and geographically similar urban area?

According to the OLS and GWR results, in Omaha and Lincoln the number of houses built before 1960 (per census tract) was a meaningful factor for explaining the recent low-school-attainment landscape. Proximity to I-80 for census tracts in Lincoln and Omaha was also a relevant factor in explaining the low-school-attainment landscape. Hence, two variables related to low school attainment were the same for both Lincoln and Omaha. I have devoted ample space to the discussion of the persistence of low school attainment and social reproduction theory, and I will devote a fair amount of space to an explanation of the DistI80 variable, so what follows is an explanation of the OLS and GWR results for the age of the housing stock in 1990, Hispanic/Latino households in 1990, and the number of Hispanic/Latinos in poverty in 1990 in Lincoln and Omaha.

Lesli Rawlings’ (2009) dissertation is quite helpful in understanding OPS and LPS housing patterns as a predictor of school attainment, and in this section I will draw mainly on her work to illuminate my results.
6.3.3.1 How Might the Number of Houses Built Before 1960, Hispanic Households, and/or Hispanics in Poverty Help Explain the Patterns in the OPS and LPS Low-school-attainment landscape?

Anacker (2010) found that despite some regional fluctuations, it was generally the case that compared to predominantly white census tracts, African American and mixed-race census tracts in the U.S. were worse off in terms of housing-related market factors. Rawlings’ (2009) work in OPS and LPS fits with Anacker’s (2010) results, but Rawlings (2009) also found that in both Lincoln and Omaha, proximity to major roadways and railroads tended to be related to lower property values.

These findings from Anacker (2010) and Rawlings (2009) are generally supported by the results from this study. But the most important caveat to these findings (that GWR helped to illuminate) was that the connections between race/ethnicity, housing, and proximity to major roads (I-80 in this case) weren’t entirely uniform throughout OPS or LPS. Hence, a compelling feature of local regression analysis is that it can challenge the deterministic insinuation in the question beginning this section. In OPS for example, Hispanic/Latino households in 1990 predicted recent low school attainment, but only in southern OPS. In northwestern OPS, the localized coefficients changed from positive to negative and from significant to non-significant—indicating that in this area of OPS there was no statistically significant relationship between Hispanic/Latino households in 1990 and recent low school attainment. In Lincoln, it wasn’t the number of Hispanic/Latino households that helped explain low-school-attainment patterns, but rather it was the number of Hispanics/Latinos in poverty—which was fairly uniform as a predictor of school attainment across all of LPS. Hence, in LPS compared to OPS, poverty may play a larger role than race/ethnicity alone where Hispanic/Latino education are concerned. This
contrast between Lincoln and Omaha implies that GWR not only captured variability within these urban areas, but that we can find interurban nuances in the same or similar factors as well.

In general though, the results of this study are similar to what Rawlings (2009) found in her extensive study of housing prices in Lincoln and Omaha. Within the boundaries of LPS and OPS, Rawlings (2009) found that low levels of school attainment were bound up with nonwhite populations and low home values, and that homes values within a quarter-mile of a major road or railroad were significantly lower compared to home prices outside this area.

However, GWR results from this study indicated that in parts of northwestern and central LPS the relationship between census tracts with an older housing stock in 1990 and recent low school attainment was significantly negative. Meaning, in northwestern and parts of central LPS, newer housing in 1990 was related to recent patterns of low school attainment. Conversely, an older housing stock in 1990 (again, only in central and northwestern) Lincoln was linked to relatively higher levels of education. This pattern, while slightly different from Rawlings’ (2009) findings—she found the age of a home was negatively correlated with its value—is still plausible for a couple of reasons.

First, parts of south-central Lincoln are home to some of the oldest and most expensive housing in the city, especially near and along Sheridan Boulevard and south of Van Dorn St. near the Lincoln Country Club. But, as a resident of Lincoln since 1998, I have watched as the city has grown to the east, southwest, and west, and I have watched as newer, relatively expensive housing developed in these areas as well. Rawlings (2009) findings support my observations. She found a strong relationship in Lincoln
between/among: high incomes, post-secondary education, and residents living near the eastern, southern, and southwestern urban fringe. This explains why she also discovered that commuting times in Lincoln are predictive of increases in property values.

But, as a second and related point, since 1990 Lincoln has also grown substantially to the north/northwest, and, as Rawlings (2009) pointed out, nonwhite residents and residents without a post-secondary degree tended to be clustered in these areas. Furthermore, Rawlings (2009) found that the combination of the proportions of nonwhite residents and rental properties in LPS school attendance zones were key factors in explaining the values of single family homes throughout the district—according to Rawlings (2009), “the greatest portion of rental properties and nonwhite residents are located near central and northwest Lincoln” (pp. 416). Some of Lincoln’s lowest property values can be found in the central, north, northwestern portions of the city.13

So, as Lincoln has grown to the southwest, south, and east, housing has tended to be newer and more valuable, but as Lincoln has grown to the north/northwest, housing has tended to be relatively less valuable, especially as the housing stock gets closer to a major road (Rawlings, 2009)14 or in this case, the interstate.

Hence, the link between newer housing in 1990 and low school attainment appears to overlap with Rawlings’ (2009) finding that property values for single family homes in areas with high proportions of rental properties and high proportions of nonwhite residents tends to be lower. Indeed, in LPS, Hispanics/Latinos in poverty have

13 For an interactive look at relevant housing data for Lincoln, NE (and Omaha) see: [http://www.city-data.com/housing/houses-Lincoln-Nebraska.html](http://www.city-data.com/housing/houses-Lincoln-Nebraska.html)
14 But there is spatial variability here as well. At the time of this writing, the recently completed Fallbrook housing development just north of I-80 and Lincoln, but in the LPS district, lists homes between $300,000 and $400,000. [http://www.woodsbros.com/pages/fallbrook](http://www.woodsbros.com/pages/fallbrook)
been and continue to be clustered more or less in the center and north/northwestern portions of the city and relatively close to the interstate where low school attainment is most predictable. In OPS, Hispanic/Latino residents are predominately clustered in southern OPS, near I-80, where 1990 housing was most related to low school attainment, and where the low-school-attainment landscape was most predictable.

If education gaps exist and persist between poor, and underserved racial/ethnic minorities and middle-class white residents, and if poor, underserved racial/ethnic minority groups are found to be clustered in lower-income housing, which also tends to proliferate as one nears I-80 in Lincoln and Omaha, then where these combined factors are located should also be where the low-school-attainment landscape is most predictable. This was precisely what the GWR results revealed in LPS. In OPS a similar overlap was there, however, the proximity to I-80 variable was slightly more complicated (which I will address in the section to come).

The question from section 5.3.2 asked if the same set of variables related to low school attainment in one urban area apply to another that is demographically and geographically similar? In the case of Lincoln and Omaha, the easy answer to this question is: ‘no, but a similar set of variables did explain significant portions of the low-school-attainment landscape in both places’. In the following section I will develop a more complex (and hopefully more satisfying) answer to this question through an exploration of the DistI80 variable that was a shared predictor for both OPS and LPS.
6.3.3.2 Why Might Distance to I-80 Predictive of Low School Attainment in Both Omaha and Lincoln?

The GWR results from this study suggested that housing patterns in 1990 in certain areas of OPS and LPS appeared to be linked to low school attainment more recently. And the GWR results suggested that indeed proximity to I-80 did matter for school attainment, but that the relationship was not uniform throughout the LPS and OPS environments. For census tracts in central and parts of southern OPS, there was a strong, negative relationship between the number of houses built before 1960 and recent low school attainment. This suggests that a relatively newer housing stock in 1990 was associated with census tracts with higher numbers of residents who had not been to college. But there was spatial nonstationarity in this relationship throughout OPS, which suggested that north of the interstate the relationship between newer/older housing in 1990 and low school attainment became less predictable (see section 4.5.4). Something similar can be seen for the relationship between the number of Hispanic households in 1990 and the low school attainment in 1990 variables and recent low school attainment—i.e., these variables had the most intense association with recent low school attainment south of I-80. Thus, I-80 in Omaha appears to delineate the census tracts to its south as having a consistent demographic mix of the variables which are related to low school attainment.

The effect of I-80 on the low-school-attainment landscape was slightly different in LPS, where the GWR model showed that the explanatory/predictive power of the DistI80 variable essentially cut the district in half along a southwest/northwest diagonal. In the northern portion, the relationship was clear: as LPS census tracts grew closer to I-
80, low school attainment increased. According to the local GWR parameters, for most of northern and northwestern LPS, a 100-meter decrease in a census tract’s proximity to I-80 corresponded, on average, to an increase of approximately nine people who hadn’t been to college. But, this relationship dissipated in a southeasterly direction, and for a large portion of southern LPS, there wasn’t a relationship between proximity to I-80 and low school attainment.

Out of the OPS/LPS examples emerges another way to think about the role proximity to I-80 plays in the low-school-attainment landscape—i.e., not as a predictor (or just a predictor) but as a spatial organizer. This is key. Recall that I accounted for space in two different ways in this study. First, in calibrating the GWR model, the bandwidth size (or “kernel” [Brundson et al., 2002, pp. 433]) needed to be established in order to estimate local regression parameters. In calibrating kernel size, a decision needed to be made between using a fixed (which would have preserved the area of the bandwidth) or an adaptive kernel (which would have fixed the number of observations within each kernel). After selecting the adaptive kernel method, the GWR algorithm then selected the best-fitting local model from a series of possible models by finding the kernel size (i.e., the number of observations in each local model in this case) at which the lowest $AIC_c$ score was produced. Put briefly, the GWR modeling technique explicitly operationalized space.

The other way I accounted for space was by introducing into the pool of possible explanatory factors, two candidate variables that were spatial in nature—one that measured the distance of a census tract to the city core, and one that measured the distance of a census tract to I-80. In the model building process, distance to I-80 emerged
as part of the passable OLS model that was selected for further comparison with GWR. I analyzed this variable in more detail by removing \textit{DistI80} as a factor from the final OPS-OLS model, and I tested the same OLS models for OPS and LPS sans a variable measuring some spatial characteristic of the district. Hence, this version of the OPS-OLS model included only the \textit{HS90}, \textit{HHH90}, and \textit{H30OLD90} variables, and the LPS-OLS model included \textit{AG25UP}, \textit{NHPOV90}, and \textit{H30OLD90}. Table xx below compares these OLS models with and without the DistI80 variable.

Table 6.3: Model Comparison: With/Without Spatial Factor \textit{DistI80}

<table>
<thead>
<tr>
<th></th>
<th>OPS Model Diagnostics w/DistI80</th>
<th>OLS Model Diagnostics without DistI80</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OPS Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. R-Squared</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>AICc</td>
<td>1679.85</td>
<td>1681.59</td>
</tr>
<tr>
<td>Moran’s I ((p)-value)</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>LPS Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. R-Squared</td>
<td>0.59</td>
<td>0.50</td>
</tr>
<tr>
<td>AICc</td>
<td>1028.76</td>
<td>1042.85</td>
</tr>
<tr>
<td>Jarque-Bera ((p)-value)</td>
<td>0.21</td>
<td>0.09</td>
</tr>
</tbody>
</table>

There are a few key points that the data in Table 6.3 reveal. First, in LPS, removing the \textit{DistI80} variable resulted in a large drop in both \textit{AICc} and adjusted \(R^2\) values, indicating that \textit{DistI80} did play an important explanatory role in the LPS-OLS model. Additionally, removing the \textit{DistI80} variable resulted in a model with a significant
JB value, meaning that without $DistI80$ as a variable, the OLS model violated the regression assumption that residuals are normally distributed. Hence, in LPS, $DistI80$ served as both an explanatory variable and as a variable that kept the OPS-OLS model residuals from becoming significantly skewed.

For Omaha, the OPS-OLS model sans $DistI80$ violated the assumption that model residuals are not spatially autocorrelated (see bold statistics in Table 6.3 for violations). However, the rest of the model diagnostics remained nearly the same. In the OPS case, the percentage of explained variance is the same with or without $DistI80$, and based on the $AICc$ scores there isn’t a large enough difference to determine which model is the best-fitting. So, it appears that the role of the $DistI80$ variable in the OPS-OLS model was in keeping model residuals from being spatially clustered. In that sense, the $DistI80$ variable was a sort of spatial organizer (or perhaps, disorganizer) for model residuals, since including $DistI80$ did create a random spatial pattern of residuals, but didn’t change the explanatory capabilities of the OPS-OLS model or how well the model fit the data.

Given that regression analysts are interested in advancing models with residuals that are randomly distributed—i.e., with over/under predictions that occur with equal probability—$DistI80$ was important to the OPS and LPS models in two key ways. First, the LPS-OLS model without $DistI80$ produced residuals that were skewed to the point of rejecting the model. This was a matter of too many large under-predictions (positive skewness) in the LPS-OLS residuals. Second, the OPS-LPS model without $DistI80$ produced residuals that were spatially clustered, also to the point of rejecting the model. Including the $DistI80$ variable in the OPS-OLS model reduced the spatial autocorrelation of the residuals. This second fact contains a crucial point that many researchers outside of
geography and fields using geographical/spatial analysis techniques simply fail to consider. When using geographically-oriented data (which includes a large portion of education-related data [Hogrebe and Tate, 2012]) to build regression models, a given model may produce residuals that do not violate traditional regression assumptions, but, when residuals are analyzed spatially, they are significantly clustered or dispersed throughout the geographic area of the study. This was precisely the case for the OPS-OLS model. If a regression analyst is using data that are spatial in nature (what data isn’t?) and she is concerned about the correlation of her models’ residuals she would do well to consider how her model’s predictions are spatially structured. Conducting traditional regression analyses without the help of spatial analysis and without explicitly accounting for the possibility of spatial influences is likely to lead to underestimation of model parameters (Legendre, 1993) or worse, systematically biased results (Lennon, 2000).

But the question still remains, why would I-80 be important to modeling as an organizing variable (as it was for OPS), and/or as a variable with explanatory/predictive power (as it was for LPS)? In order to answer these questions a brief history of the east/west transportation network in Nebraska may be useful.

6.3.3.3 The Recent and Historic Organizing Effects of the Major East/West Transportation Networks in Nebraska

The Nebraska stretch of I-80 will turn 40-years-old this October, but much of this route is actually far older. I-80 sits atop (or at least near to) a pioneering superhighway of sorts, on which an estimated 350,000 settlers travelled west from 1840 to 1866 (Mattes,
1887). Eventually, the “Great Platte River Road” (Mattes, 1987)—which was mainly comprised of the Mormon, Oregon, and California Trails and the Pony Express—gave way to the Nebraska section of the first transcontinental railroad (Figure 6.2 below).

According to the 1878 Statistical Abstract of the United States, there were 122 miles of railroad track in Nebraska in 1865, and by 1877 that number had increased an order of magnitude to 1,286 miles (U.S. Census Bureau, n.d.). During roughly that same period (1860 to 1880), the population of Omaha increased from 1,883 to 30,518 (about 1,521%). This is why Winckler (1990) wrote in *The New York Times*, that in addition to the stockyards and smelters and its large and vibrant immigrant communities, Omaha also owes its existence to the Union Pacific railroad (and perhaps to President Lincoln as well).

Omaha was born in 1854, towards the end of the aforementioned period of mass—albeit primitive—pioneering trail transit. Five years later, in 1859, President Lincoln named Council Bluffs (located directly across the Missouri River from Omaha) the eastern terminus of the transcontinental railroad. Leading up to this pronouncement, much of the “Great Platte River Road” had already been transformed into railroad, and the impact of President Lincoln’s designation for Omaha was its cementing as a major trade, supply hub, and population center for the region (Danton, 1967).
Figure 6.2: The Evolution of Nebraska’s Major East/West Transportation Network

From left to right: The “Great Platte River Road” (Mattes, 1987); the 1870 railroad line; I-80 and Hwy 30 and 26.

The first meat-packing plant opened in South Omaha in 1871, and shortly thereafter, the Union Stock Yards Company was organized under the leadership of Wyoming cattle magnate, Alexander Swan (Menard, 1987). By the early 1880’s, Omaha was processing, packing, and shipping a significant portion of the nation’s beef eastward. Not long after, giants of the industry—i.e., George Hammond Packing, Armour, Cudahy, Fowler Brothers, and Swift and Company—opened packing plants in South Omaha. As a result, South Omaha flourished economically, and by 1890 it was competing with Chicago and Kansas City as the nation’s largest meat-packing center (Menard, 1987). In the process, South Omaha also became a beacon for racial/ethnic diversity on the Plains. In addition to the needs of the meat-packing industry, the rail yards, and smelting works, the needs of the city in general drew in immigrants and racial/ethnic minorities from all over the country and the world. The word ‘minorities’ is perhaps a bit off-the-mark though (at least early on), because by the turn of the 20th century, the foreign-born residents and their native and foreign-born children accounted for over half of Omaha’s population (Menard, 1987).

The railcar eventually gave way to the automobile, and the railroad in Nebraska to Lincoln Highway in 1913. Most of Lincoln Highway has turned into U.S. Hwy 30, and
while still in operation, Hwy 30 was eventually supplanted by I-80 in 1974 as Nebraska’s major line of east/west transportation (see Figure 6.2 above). The effect of this East/West transportation network on the organization of Nebraska cities and towns is clarified by Figures 6.3 and 6.4 below. In Figure 6.3, I have arranged eight of Nebraska’s cities and towns along I-80 by their population sizes and their distances in miles to Omaha.

These towns were: Lincoln (265,404; 53mi), Grand Island (49,949; 145mi), Kearney (31,790; 182mi), Lexington (10,213; 218mi), North Platte (24,592; 277mi), Ogallala (4,649; 326mi), Sidney (6,808; 396mi), and Kimball (2,465; 432mi). In Figure 6.4, I selected another eight Nebraska cities and towns not located along I-80, and I also arranged them by their population sizes and distances to Omaha. These towns were: Nebraska City (7,277; 45mi); Columbus (22,508; 84mi); Fremont (26,167; 99mi); South Sioux (13,353; 100mi); Norfolk (24,332; 110mi), Hastings (25,058; 157mi), McCook (7,698; 280mi), Scottsbluff (15,039, 451mi).

It is obvious from Figure 6.3 below (top) that for cities and towns in Nebraska along I-80, systematicity exists in the relationship between distance to Omaha and population size. It is equally obvious that the orderly relationship observed in Figure 6.3 does not apply to the sample of cities and towns I selected that are not located along I-80. The implication is that being situated along I-80 connects cities and towns to Omaha in ways which tend to govern population size. And, while Omaha is probably influential to at least some degree on all of Nebraska’s cities, the influence Omaha brings to bear on a city or town’s population is clearly strongest along I-80. If I-80 can organize the population patterns of groups of people between cities, there is little reason to doubt that it also has the influential capacity to organize groups of people within cities as well.
For cities along I-80 then, the interstate appears to have a powerful organizing effect on population patterns, and from the maps in the previous chapters, and from the analysis of the OLS models with/without the DistI80 variable, I-80 appears to influence the patterns of low school attainment within cities as well. But, why this is the case remains unanswered. I will devote the remainder of this section to two possible explanations regarding how proximity to a major roadway like I-80 could influence low school attainment, acknowledging first that each explanation is speculative and will require responsiveness from future research in order to be fully developed.
6.3.3.4 Two ways I-80 may influence the low-school-attainment landscape

Obviously there are more (perhaps many more) than two ways to think about the relationship between I-80 and the school-attainment landscape in an urban area like Omaha or Lincoln. But, for the sake of time and space, I will focus the following discussion on two reasons that I-80 may be linked to school attainment—two reasons with a substantial amount of support from academic literature. First, it could be that I-80 itself was partially responsible for changes to the environments in LPS and OPS, and that these changes adversely impacted human health and development in the areas near the interstate, which in turn impacted patterns of school persistence.

Another (and probably related) way to think about living/growing up near I-80 is as an indicator of political power, which may in turn be related to school attainment status. If living close to a major interstate is undesirable, and if as a result property values near the interstate tend to be low, then it is reasonable to assume that groups living nearest to the interstate do so because, by-in-large, they may not have the means to live elsewhere. If wealth marks the availability of political clout, and poverty marks the opposite, then where disadvantaged groups reside near to a major highway or interstate it is also reasonable to assume that these groups may generally lack political power as well. Another way a major interstate or highway could mark political power is how/if it displaces people when it’s built. If a proposal for the building of a major interstate through an inhabited area carries with it the displacement of residents, could it be that those who are ultimately displaced or forced to live near a noisy, neighborhood-splitting artery are those who did not have the political power to resist displacement? What about the political powers of entire communities whose elected leaders attempt to gain control
over the routing of a major interstate and highway? What does the final route say about larger-scale political power? These questions and perspectives are explored in more detail below.

6.3.3.5 Could proximity to I-80 be responsible for adverse health effects, which ultimately limit a person’s schooling?

This may seem like a strange question to ask—how, after all, can living near an interstate hamper a person's schooling trajectory to the point that someone living at some distance farther away from the interstate persists in school longer? Peculiar as such an idea may seem at first, there is a fair amount of international research supporting the possible existence of such a phenomenon. Literature from the field of Environmental Justice (and related fields) has shown that in both developed and developing nations, groups of people with low-SES and low school attainment are often exposed to high levels of particulate air pollution because of their residential proximity to major transportation networks (see: Boothe and Shendell, 2008 for a detailed review of this literature from 1999-2006; see also: Jerret et al., 2001; Ou et al., 2008; Wilhelm and Ritz, 2003). In their meta-analysis, Boothe and Shendell (2008), found that of the 29 respiratory studies they reviewed, 25 reported statistically significant relationships between residential proximity to major transportation networks and at least one of the following: “increased prevalence and severity of symptoms of asthma and other respiratory diseases; diminished lung function; adverse birth outcomes; childhood cancer; and increased mortality” (pp. 38). Boothe and Shendall (2008) reported that another 9 out of 10 non-respiratory studies also found proximity to major transportation networks to significantly predict “childhood cancer;
adverse birth outcomes; and cardiopulmonary, cardiovascular, cerebrovascular, and stroke mortality” (pp. 38).

Similarly, Jerret, et al. (2001) concluded that underprivileged groups face a ‘triple jeopardy’, in that these groups have been shown to be: (1) at increased risk for participating in adverse social and behavioral habits (e.g., smoking, drug and alcohol abuse); (2) at higher risk to exposure to adverse environmental conditions (e.g., traffic-caused air pollution); and (3) at risk for a multiplicative interaction between these first two factors—which is borne out in research that shows low-SES groups and those with low school attainment disproportionately experience the adverse and often fatal health impacts of traffic-caused pollution when compared to more advantaged groups.

From this perspective, I-80 itself may be a variable causing health problems, problems which in turn contribute to low-school-attainment patterns. Furthermore, per the multiplicative impact described in the Jerret, et al. (2001) study, the effects of living close to I-80 may be exacerbated in poor, minority groups, which could help explain spatial variability in school attainment along I-80 in Omaha. Proving this however, would first require research linking proximity to I-80 to patterns of health problems in Omaha, and then linking those health problems to school attainment patterns in the district. Hence, this is a direction that future research could take, but not a conclusion that can be drawn from this study.

6.3.3.6 Proximity to I-80 as a Measure of Political Power

Another way to think about proximity to I-80 in relationship to the low-school-attainment landscape is as a marker of political power (the absence of which may also
impact one’s persistence in school). Proximity to major roads and transportation networks has repeatedly been associated with political disfranchisement (Bowater, 2014; Bueno de Mesquita and Smith, 2012; Kahn, 2002; Masquelier, 2002). The adage ‘Not in My Back Yard’ probably applies to most in this case (who wants the city or state to tear up his/her neighborhood to make room for an interstate or expressway?). But political power is probably never evenly distributed throughout an urban area, hence, within urban areas there are bound to be pockets of people living on or near relatively less valuable land who lack the political power to resist the building of transportation infrastructure near their residences. This explains why the relatively powerless may often be the ones displaced by a build, and why patterns of low property values, disadvantaged groups, and low school attainment may be clustered near transportation networks.

As a salient (but generalized) example of the connection between the course a major road might take and the amount of political power enjoyed by those impacted by its course, Bueno de Mesquita and Smith (2012) analyzed the general political power of the residents in the capital cities of 158 nations by examining the straightness/curviness of roads leading from the center of the capital to its largest airport.

Theoretically, a straight road is a cheap road and a curvy road is expensive (ceteris paribus), and given a polity that is interested in curbing costs, ‘plowing through’ is usually more economical than ‘going around’, but only if the residents displaced do not require a great deal of compensation. Bueno de Mesquita and Smith (2012) suggested some caveats to their idea; they argued that the types of topographic features encountered in an urban landscape are not usually dictated by political power (unlike the distribution of wealth and power). So, there are examples of curvy roads to airports in capital cities
where people generally have little political clout. But, topographic features notwithstanding, Bueno de Mesquita and Smith (2012) found that the countries with the most autocratic governments tended to have the straightest roads out of their capital cities. In fact, when the authors ranked 158 cities based on the curviness of the major roads to the airport, and they found that of the 30 capital cities with the straightest roads, only two (Portugal and Canada) were strong democracies. The remaining list included countries like Afghanistan, Pakistan, Yemen, Colombia, Cuba, Guinea, Dominica, Ecuador, and Ethiopia. Of the twenty-eight non- or weakly-democratic countries remaining, the authors found that only Colombia and Ecuador have made any recent strides towards instituting a government that is truly beholden to a large coalition of its governed people. The remaining nations fall into Bueno de Mesquita and Smith’s (2012) framework, which suggests that where the governed lack the political power to resist, leaders will build straight, cheap roads, no matter the human costs.

As a timely and much more specific example of this, Bowater (2014) tells the story of José Paulo Barcellos and his family in western Rio de Janeiro, Brazil. Barcellos built his home himself, slowly and over several decades. Over the years he added an upstairs to his single-story bungalow in order to accommodate his daughter and her children. He also built a carport/workshop and a small playground for his granddaughters. According to Bowater (2014), the Barcellos’ home, along with approximately 900 others in western Rio de Janeiro, will be destroyed this year to make way for the TransOlímpica rapid bus system, which is meant to accommodate the transit needs of the city when it hosts the 2016 Olympics.
The Barcellos story is nearly identical to that of Caroline Chang’s. Kahn (2002), described Chang’s childhood home at 48 Hudson St. in Boston, MA as vibrant. Her house was in an ethnically Chinese neighborhood where boys played baseball and people greeted one another and spoke in their native language, Toisanese. But, Chang’s house was demolished to make room for a section of the Massachusetts Turnpike Extension (Kahn, 2002).

South Omaha was no different. Many blocks of housing and commercial buildings were destroyed, city street grids disrupted, parks and neighborhoods demolished in order to make room for I-80 (Mead & Hunt, Inc., 2005). The location of I-80 through south Omaha also allowed easier movement between the CBD and the suburbs, which had the effect of accelerating suburban growth (read: ‘white flight’) to the southwest/western areas of Omaha and surrounding communities.

The implications of these examples is that proximity to a major roadway may mark a lack of resources, including political power, and if a lack of political power can be linked with low school attainment, then in the Nebraska case close proximity to I-80 as a factor predicting persistence in school has a relatively straightforward explanation.

6.3.3.7 Complications with I-80 as a marker of a lack of political power in Nebraska

In this section I detail some of the complications that come with the idea that a major transportation network like I-80 could mark uneven political power in Nebraska and its cities/towns; a lack of political power which in turn may have helped establish the patterns of low school attainment that can be observed in Lincoln and Omaha more recently. In the examples from Brazil and Boston above, a major transportation project
was built through an inhabited area, which then caused (or in Brazil, is causing) residents unable to resist the political decision of where to build to be displaced. A similar phenomenon occurred in southern Omaha with the building of I-80. However, in Nebraska, with the exception of Omaha, I-80 actually bypasses the remainder cities and towns along its route (Lincoln has grown/is growing towards and around I-80 but was bypassed in the original design of the route [Creigh, 1991]). Hence, except for Omaha, I-80 in Nebraska goes to but not through the cities along its route. This configuration was no accident, nor were the decisions about where to build I-80 a peaceful legislative process; the political controversies that surrounded the final design, funding, and construction of I-80 are relevant to this discussion since they complicate the logic of I-80 as a marker of political powerlessness.

6.3.3.8 The Political Turmoil and the Nebraska Portion of I-80

In 1944, congress passed the Federal Highways Act, which called for the designation of a network of 40,000 miles of national superhighways connecting state capitals, other important cities, and industrial areas. These cities and areas became known as “control points” and they were officially directed to be linked by the “National System of Interstate and Defense Highways” by the Federal Highway Trust enacted in 1956. For Nebraska these control points included Omaha, Lincoln, and North Platte (Creigh, 1991). Hence, the federal government established a general path through Nebraska which I-80 had to follow, but local officials were ultimately left responsible for the design of the final route between these points. This federal/local division of route design created the backdrop for the political melee surrounding construction of Nebraska’s portion of I-80.
James C. Creigh, an Omaha attorney, provided an excellent and detailed account of the controversies surrounding the funding and routing of the Nebraska section of I-80. Creigh (1991) described the issue of routing I-80 as a “most bitter struggle” (pp. 46), which was spurred on and constituted by popular opinion, the formation of regionally-based state interest groups, local and statewide political jockeying, and pressure from lobbies as well as the federal government on the governor, state senators, and the Nebraska Department of Roads.

At the heart of the routing issue was an announcement in May of 1954, by State Engineer L.N. Ress, in which Ress made it clear to local leaders that if the cost of a bypass was less than the cost of going through a given town, the bypass would likely be built (Creigh, 1991). This news dismayed some leaders who had hoped that I-80 would pass directly through their towns. But the main reason this caused the beginnings of a political conflagration was that in a 1947 report, the Federal Bureau of Roads suggested a route for I-80 wherein the highway would enter Nebraska through Omaha and then follow U.S. 6 to Lincoln, U.S. 30 to Grand Island, and U.S. 30 across the rest of the state (Creigh, 1991). Upon Ress’s announcement that I-80 was unlikely to go through any of the cities along its proposed course, there suddenly appeared to be substantially more flexibility in the route, and so, in 1955, a group of leaders in areas not adjacent to the federally proposed I-80 route organized and lobbied—but ultimately failed—to have the interstate rerouted along Hwy 92 (Creigh, 1991). The reason cited for this rejection was that the federally mandated route had to include Omaha, Lincoln, and North Platte—these were set in stone by federal law—but the die had been cast, and for most of the ensuing
decade, controlling the course of I-80 dominated the political agenda for many Nebraska towns and for many of the state’s leaders.

For example, several years’ worth of controversy erupted in central Nebraska over whether to site I-80 along the north or south bank of the Platte River. Creigh (1991) pointed out that early popular support for the south bank route had been fomented to some degree by editorial arguments in a few of Nebraska’s newspapers, in combination with the announcement in 1955 by governor Victor Anderson that I-80 would definitely run somewhere in between Hastings and Grand Island. Two years later, with popular support for the southern route still intact, the Nebraska Department of Roads announced its plans for I-80 to follow the north bank. Lobbying groups sprang up almost immediately, representing dozens of towns in southern, central, and western Nebraska, each attempting to secure an I-80 route most beneficial to their region of the state. The largest were the South Platte United Chambers of Commerce (SPUCC) and later the Western Nebraska United Chambers of Commerce (WNUCC). The fray in central Nebraska was primarily between the SPUCC, a handful of state senators, and the Nebraska Department of Roads. After a report found the northern route to be an estimated $9 million cheaper, and after a series of contentious public hearings and debates, the decision to build north of the Platte River in central Nebraska was made in January of 1960 (Creigh, 1991).

Not long after, a similar flare-up occurred in western Nebraska regarding the routing of I-80 along the North vs. South Platte Rivers. The SPUCC, representing 42 southern Nebraska towns found itself again supporting a southern I-80 route this time west of North Platte, but now the SPUCC was squared off against the WNUCC
(representing Gering, Ogallala, Oshkosh, and Scottsbluff), the Platte Valley Irrigation District, and the North Platte Motel Association, and they were allied with the Nebraska Department of Roads, which also backed the southern route (Creigh, 1991). After another round of intense lobbying and contentious debate, in January 1962, the announcement came down from Governor Frank Morrison that while the northern route was more expensive, it was also the safest and it destroyed a lesser amount of valuable land. The WNUCC and its northern route compatriots had won. But one month later, the Federal Bureau of Public Roads (which was in charge of allocating for 90% of funds for the interstate) refused to pay the extra costs for the northern route forcing the governor and the Nebraska Highway commission to acquiesce to the southern route. So, ultimately, because of federal pressure, the SPUCC and the Nebraska Department of Roads got the southern route they had originally sought (Creigh, 1991).

Where I-80 is concerned then, it is clear that there are complications to the idea that a major transportation network is a marker of a lack of political power. Given that south Omaha is home (both historically and currently) to relatively higher proportions of poor residents, immigrants, nonnative English speakers, and residents with low school attainment, the routing of I-80 through south Omaha certainly fits in with the idea that given the necessity of invoking eminent domain in an uneven landscape of political power, the least costly thing to do (politically and economically) is to build through an area of the landscape where there is less political power. So in Omaha, I-80 may certainly mark a dearth of political power, but at least for the rest of the state, the routing of I-80 is a sign of the existence of political power, not a lack of it. Given the story above, in Nebraska, where I-80 isn’t a better indicator of who the political losers were in the
fight to control the course of the interstate. Furthermore, state and federal interests (mainly national defense and budgetary concerns) mandated a route for I-80 that went to the cities and towns along its route, but not through them. Hence, there was (and is) a to/through dynamic in Nebraska that may simultaneously indicate both a presence and a dearth of political power. Given the finite number of locales in Nebraska about which to test this hypothesis, perhaps study in other states could be illuminating (e.g., Missouri or Ohio).

In conclusion, the ‘proximity to I-80 as a cause of health problems’ (described in 5.3.3.4) and the ‘proximity to I-80 as a marker of political power’ arguments while compelling and perhaps important considerations, are also speculative where this study’s conclusions are concerned. However, there is direct evidence from this research that the early development of Omaha and the east/west transportation network in Nebraska played a substantial role in the eventual (and continued) organization of the populations between the towns and cities along I-80. Furthermore, there is direct evidence from this study and from Rawlings (2009) that suggests within Lincoln and Omaha, I-80 (and major roads in general) plays a role in organizing housing and property values, school attainment levels, and distributions of nonwhite populations. In that sense, I-80 fits squarely into the raft of previous literature suggesting that the structures of the urban environment influence the spatial patterns of income disparity, demography, housing, and related phenomena therein (Chi, et al. 2013; Harrington and Warf, 2002; Harris and Ullman; 1945; Huang and Wei, 2013; Lersch and Hart, 2014; Michaels, et al. 2013; Rawlings, 2009; Slack and Meyers, 2013; Wei et al, 2010). But, as the GWR results have
illuminated, the influences of urban structures are not necessarily uniform throughout the urban environment.

6.4 What policy implications arise from the presence of spatial variability in the strength of the relationships that predict low school attainment?

Recall from chapter one that there may be substantial long-term benefits to Lincoln and Omaha that come with increasing access to and enrollment in (especially for disadvantaged groups) post-secondary schools. Increases in school attainment levels have been shown to reduce crime rates, decrease reliance on publicly provided services, and increase contributions to the tax base (Rud et al., 2013). These GWR results do not reveal what specific policies could be expected to even out the low-school-attainment landscape in Lincoln and Omaha, so, the best ways to disrupt the long-lasting spatial clustering of residents who have been to college remains an open question. But, given the spatial variations present in most of the factors related to school attainment, the first policy implication of this study is that where a policy or reform is implemented is a consideration that may be as important as the content of the policy or reform itself.

There were obvious differences between southern and western Omaha regarding how the number of Hispanic/Latino households in the past related to increases or decreases in low school attainment. This suggests that from a policy perspective, helping Hispanic/Latino people gain more access to post-secondary school in western OPS may be a very different task than in the southern portion of the district (which, in turn, might suggest important differences within the Hispanic/Latino population that have geographic patterns). If increasing persistence in school and increasing post-secondary access and
enrollment (particularly for disadvantaged groups) are worthy goals, one-size-fits-all policies that lack responsiveness to people and places may not be as efficacious as we hope. We should not only be asking what needs to be done to solve problems in our communities and schools, but where it needs to happen (for a similar point made much differently see: Gardner, 2008). Differentiation in school outreach as well as in public policy is possible, and GWR can help to guide and support decisions about where a particular policy should be implemented, tweaked, overhauled, supported, or ended.

Second, the scale and intensity of a particular problem are important concerns that GWR helps to illuminate. If a particular policy or reform is initiated at a scale much larger than necessary, time and money may be unnecessarily wasted. In LPS, past housing patterns helped explain recent school attainment, but only for northwestern and parts of central LPS. If a housing policy of some type were enacted in a citywide effort to boost college attendance, it may only have a systematic impact in northwestern and central Lincoln. Conversely, if a problem is targeted by a reform at too small a scale, it may do little to solve a more widespread issue. It is also possible that new policies will provide geographic organization to variables that were previously more or less spatially independent (e.g., consider how urban desegregation mandates changed suburban districts). Here GWR can analyze the geospatial impact of a particular policy. In addition, GWR helps to reveal fluctuations in the intensity of a particular problem. Given budgetary and resource constraints, focusing reform efforts where the factors predicting low school attainment are most exacerbated could help to equalize access to post-secondary education.
In the era of “big data”, education and education-related datasets are not only voluminous, but they are also produced with great velocity (Laney, 2001). Using software like GIS in combination with global and local regression analyses can illuminate a policy landscape in new and exciting ways that were previously impossible. In addition, education and related data are inherently geographical in nature. So within these huge datasets there are bound to be hidden spatial relationships which previous research has largely left untouched. Luckily, as Slagle (2010) pointed out, spatial tools such as GWR can uncover these spatial relationships and in turn these tools can produce better-fitting models, which represent a truer version of reality, and hence may be more “useful” (per Box and Draper, 1987) for decision making.

6.5 Summary

The purpose of this dissertation was to show how spatial analysis and spatial modeling might be useful for attending to problems in education. Hotspot analysis and geographically weighted regression (GWR) are two tools available to researchers who are interested in the analysis of space and the modeling of spatial relationships. In this case, I applied these techniques to the problem of low school attainment in two urban settings. I did this by first combining incremental spatial autocorrelation with hotspot analysis techniques in order to detect statistically significant clusters of low school attainment in the census tracts comprising the Omaha Public Schools (OPS) district. Then, I used exploratory regression analysis to find a passable Ordinary Least Squares model (i.e., an OLS model that did not violate regression assumptions) from among twelve candidate variables collected by the U.S. Census bureau that would help explain the hotspots
uncovered by the initial cluster analysis. The candidate variables were both demographic and spatial in nature, they were collected from the 1990 census, and they were chosen to provide a set of factors that might explain a significant portion variability in recent low-school-attainment numbers in the population of OPS residents ages 25 and over. With these twelve candidate variables I used the exploratory regression tool available in ArcGIS v.10.1 to identify a set of variables that explained a significant portion of the low-school attainment variability in OPS and did not violate regression assumptions. With these variables I then created OLS and subsequently GWR models of the significant factors explaining low school attainment in OPS. With the GWR tool, I was able to create maps depicting continuous surfaces of local regression parameters for each predictor variable and the dependent variable for OPS. I was also able to compute and create contour lines representing changes in t-scores for slope coefficient estimates describing the relationship between each predictor variable and the dependent variable. Then, I overlayed these contour lines on top of the maps for each predictor variable. Consequently, I was able to simultaneously depict the strength and direction of the relationship between each predictor and the dependent variable, as well as the statistical significance of the localized slope coefficients across the entire OPS district.

Then, to assess the interurban portability of the global model developed in Omaha, I tested the OPS-OLS model on LPS low-school-attainment data. The OPS-OLS model explained a significant portion of the variability in the LPS low-school-attainment data, but violated the regression assumption that residuals are randomly distributed, thus the results were deemed untrustworthy. Rather than proceed to a GWR analysis based on spurious OLS results, I repeated the procedure described above to determine which
combination of factors (if any) from among the twelve candidate variables would best fit the LPS data and explain the most variability in the LPS low-school-attainment landscape. A set of significant predictors for LPS was identified, and these were very similar to those predictors identified for OPS.

An analysis of the $AIC_c$ scores produced by the OLS and GWR models revealed that for both OPS and LPS, the GWR model was likely to be the better fit for recent low-school-attainment data in each district. This was most likely a result of the fact that in both OPS and LPS, spatial nonstationarity was present to varying degrees in the relationships between the predictors and the dependent variable. For both OPS and LPS, there was spatial nonstationarity in the relationship between recent low school attainment and the 1990 housing and distance to I-80 variables. For OPS, there was also spatial nonstationarity in relationship between the 1990 distribution Hispanic/Latino households and recent low school attainment.

Ultimately, it was clear from the combination of global and local regression analyses that in both OPS and LPS, the areas of the districts which were nearest to I-80 were also the areas where the global OLS models were most accurate. In OPS, this was primarily the case for census tracts in the southern portion of district, where 1990 patterns of housing, Hispanic/Latino households, and low school attainment were all significantly tied to the recent low-school-attainment landscape. In LPS, global results were most accurate for census tracts in the central and north/northwestern portion of the district, where 1990 patterns of housing, Hispanics/Latinos in poverty, the population ages 25 and up, and proximity to I-80 were all significantly related to recent low-school-attainment data. These patterns, which GWR uncovered, led to further exploration and analysis
aimed at determining why I-80 might have an effect on the organization of the low-
school-attainment landscape in LPS and OPS.

6.6 The Big Picture (in Six Points)

Hogrebe and Tate (2012) pointed out that data representing physical features like
rivers and lakes, mountains and streams, and vegetation and wildlife, are often more
easily associated with the concepts of space and geography than data we may think of as
‘nonspatial’, like student affect, school climate, and community variables like
demographics, income, and voting patterns. However, nonspatial data are almost always
tied to physical features (e.g., school buildings, homes, census blocks and tracts
delineating areas of town), as a result, most education and education-related data are
inherently (at least partially) geospatial.

The problem is that many education researchers ignore this possibility. For
example, previous inquiries into patterns of school persistence using multiple regression
and related statistical techniques typically regress school persistence data against
individual, peer, family, neighborhood, and in-school factors. Traditionally, such studies
have ignored the prospect that important spatial relationships may exist within the data.
This is a problem because there is evidence that “spaceless statistical models” (Lennon,
2000, p. 102) can result in an inflation of Type I errors—i.e., finding significant
correlation coefficients that aren’t really significant (Legendre, 1993; Lennon, 2000).
Worse, when positive spatial autocorrelation (i.e., clustering) is present in a model’s
residuals it has been known to overwhelm the associations between predictor and
dependent variables, such that any attempt to rank the importance of significant
explanatory factors in a multiple regression model may simply be an exercise in ranking these factors by their spatial auto-correlative strength—not necessarily by their relationship to the response variable (Lennon, 2000).

The models developed for Omaha and Lincoln highlight the importance of accounting for space in regression analysis. Removing the spatial variable, DistI80, caused both the OPS and LPS global models to violate regression assumptions. In the case of LPS, the Jarque-Bera test changed from non-significant to significant when the model was run without DistI80. The Jarque-Bera test for residual normality is a widely used diagnostic test, so it is likely that an analyst not using spatial modeling would have caught this particular violation. However, in the case of OPS, running the OLS model without DistI80 violated the regression assumption that residuals are spatially independent. A researcher using traditional statistical modeling diagnostics would not be likely to use the Moran’s I test for spatial autocorrelation, and thus probably would have missed this fact. Given the problems associated with not attuning to spatially autocorrelated residuals, the first conclusion of this study is that: An education researcher using statistical modeling techniques ought to pay attention to spatial analysis and modeling because these may improve his results and the conclusions he draws from his data.

As a second point, Fotheringham, et al. (2001), Slagle (2010), and Qui and Wu (2011) each applied GWR to education-related data, and all three studies found that global regression models explained less variation in the dependent variable than spatial models like GWR. In all three studies, mappable parameter surfaces allowed the researchers to view spatial nonstationarity in their modeled relationships, and all three
similarly concluded that compared to global methods, GWR led to a more sophisticated understanding of the local complexities in their data.

Results from the present study point to the same conclusion. In LPS, the age of the housing stock in 1990 was connected to recent low-school-attainment numbers but only in census tracts in northern, northwestern, and parts of central and northeastern LPS. In census tracts in southern LPS, there wasn’t a systematic connection between the age of the housing stock and low school attainment. In addition, previous patterns of Hispanics/Latinos in poverty were more uniformly predictive of the recent low-school-attainment landscape throughout LPS (even though the Hispanic/Latino population in LPS is relatively recent and small). Similarly, in census tracts in southern OPS, an increase in the number of Hispanic/Latino households, low school attainment, and a newer housing stock in 1990 were systematically connected to higher low-school-attainment numbers. With the exception of past low school attainment (which was predictive throughout OPS), GWR showed that the relationships between explanatory and dependent variables outside of southern OPS were highly variable. Given these examples of spatial nonstationarity in OPS and LPS, it is not surprising that in both districts, the GWR model proved to be a better fit for low school attainment data. These results point to a second conclusion: There is a ‘geography of school persistence’, one characteristic of which appears to be that the past is strongly tied to the present, and the factors related to this ‘geography of school persistence’ are not uniform across space. If Omaha’s leaders in 1990 could have known that the census tracts in southern OPS had a mix of demographic, housing, and school attainment factors that would be highly predictive of future low school attainment, might they also have been better positioned to
disrupt those patterns we see today? Are we willing to act right now on the assumption that these patterns that have held for the previous 24 years will continue to hold into the future if we do nothing?

As third point, local regression analysis has been found to improve modeling efforts in other fields as diverse as obesity research, (Chi, et al. 2013), urban growth dynamics (Partridge, et al. 2008), health and heart disease research (Papandreou and Tuomilehto, 2014), and indoor radon exposure (Pasculli, et al., 2014). So, while the successes of GWR are manifold, it should be clarified that GWR is limited in some important ways as well. First, GWR estimates local parameters by calculating a regression equation based on the values of the predictor and response variables for every geographic feature and its neighbors within a specified bandwidth. For this study, Omaha and Lincoln had 124 and 73 census tracts respectively, this was enough to ensure that the local regression estimates were based on a substantial subsample of census tracts. However, if the areas of study were smaller and contained as few as a dozen census tracts (or less)—as many small towns do—then there would be too few neighbors and the GWR algorithm would fail to compute the local equations. In other words, if the areal units of analysis are large (e.g., census tracts) and there are too few units comprising a particular area, the GWR tool won’t work. In these areas, researchers would need to obtain more granular data (e.g., individual-, household-, or block-level) or they could not use the GWR tool.

In addition researchers have questioned the appropriateness of using GWR to make statistical inferences (Qui and Wu, 2011; Slagle, 2010) primarily because GWR is unable to carry out traditional regression diagnostics for every local model calculated by
the algorithm. This is why the ESRI developers of the ArcGIS GWR tool insist that a user begin any local regression analysis by first finding a properly specified OLS model, then s/he should use the same variables to build his/her GWR model.\textsuperscript{15} Proceeding in this fashion ensures that the global model is well-calibrated, which tends to minimize but does not eradicate the possibility of local parameters violating regression assumptions (Qui and Wu, 2011). Thus, as a third conclusion: \textit{global and local regression analyses may not be useful in analyses of smaller cities and towns unless more granular data is available. And in any case, GWR analysis needs to be combined with global techniques, and caution should be used in interpreting GWR results.}

The issues with GWR aside, GWR is a compelling tool for research because it often leads to a new set of interesting research questions, which may have remained unasked absent a GWR analysis. For example, Slagle (2010) found that in school districts to the east of the St. Louis are and southwest of Kansas City in Missouri, per capita income was significantly and positively related to school districts’ per pupil expenditures. But, in south central and north central Missouri, per capita income was negatively related to per pupil expenditures. Similarly, Qui and Wu (2011) found that in Missouri more experienced teachers were linked to higher ACT scores, but with GWR they were able to highlight thirteen school districts where the opposite was true, where more experienced teachers were associated with lower ACT scores. That teachers with more experience are connected to higher test scores is not surprising, but why would that trend be reversed in those thirteen districts in Missouri? What is happening in parts of southern and northern Missouri that caused higher per capita income to be systematically related to lower per

\textsuperscript{15} \url{http://resources.arcgis.com/en/help/main/10.2/index.html#005p00000021000000}
pupil expenditures? Speculatively, there may be a tipping point where teachers become outdated and/or burned out. Alternatively (and/or additionally), in last-hired-first-fired districts that are seeing declining enrollment for economic reasons, more veteran teachers would actually be a predictable legacy of dropping enrollments with growing poverty better explaining the actual test score drop. In any case, future research could illuminate some of these possibilities.

In the present study, I-80 emerged as a novel variable for explaining low school attainment and it appeared that I-80 in LPS and OPS played both explanatory and organizational roles. Exploring the reason why this might be led to insights about the organizing impacts of I-80 and the historic east/west transportation network on the population of Nebraska’s cities and towns. In addition, from the I-80 analysis emerged two new directions for future inquiry: does I-80, and do other major highways/roadways for that matter, create health problems for residents living near them? Do these health problems in turn impact schooling trajectories in some way? Additionally/alternatively do major roadways and transportation networks index the political power of those residents living near them and to what consequence? What power differences are marked when major roadways go to versus through a city or town, when highways bifurcate existing neighborhoods versus enable the planned development of new ones?

These are open but potentially important questions that may have gone unasked absent GWR. As a fourth conclusion then: *It seems that one of the advantages to using GWR as a tool in education research is that new possibilities and previously obscured (often surprising) local anomalies tend to emerge from the analysis of spatial nonstationarity.* In that sense, GWR could be a useful exploratory apparatus because it
has been shown not only to uncover spatial variations in relationships between/among phenomenon which are important to schools and education policy, but GWR also forces us to ask more questions and to dig a little deeper.

As an additional point, it is worth thinking about why 1990 Census data would prove so usefully predictive of school trajectories almost a quarter century later. Schools are substantially geographical entities, not only is their physical plant fixed in a specific location with a specific street address. Their enrollments also tend to be geographic, coming from a particular catchment zone. It follows that unless there are material changes in that catchment zone (e.g., gentrification) school outcomes are likely to look similar over time. To the extent individual exceptions emerge—e.g., a student who tests well goes on to college, makes a better living as an adult, that student-turned-adult is not likely to return to her/his original census block (even if they stay in the same metroplex). In contrast, success at a school is likely to change if where the students are coming from changes (as in an open-enrollment magnet school), but in that scenario, the geography of low school attainment would be unaffected. Certain neighborhoods would still be home to low-school-attainment concentrations and others concentrations would fare better. In other words, the low-school-attainment is stable overtime not necessarily because of demographics but rather structural/organizational legacies are built into and sustained by the urban environment.

Finally, the global models developed for OPS and LPS pointed to significant nonspatial patterns in the low-school-attainment data for each district. GWR then detected areas of each district where the global modeling results were accurate, where the global results were exacerbated, where the global results were not actually significant,
and where the global model had the relationship backwards. Unlike the global results, the differentiations picked out by GWR were spatial and locational, and gradations in the global model point to the prospect that education and education-related policies could be differentiated to accommodate the unique challenges of a neighborhood, or a group of neighborhoods in a given city or town. GWR also helps support the idea that policies and resources could be more appropriately directed depending on the scale and intensity of the underlying problems in a given area.

If we accept that the educational needs of a particular community are not geographically uniform, and that some measure of the purpose of education policy and reform lies in detecting and disrupting widespread patterns of phenomena that predictably trammel or stratify educational outcomes for certain kinds of people, then GWR and spatial analysis have an obvious purpose: detecting pockets of educational inequality and allowing leaders to target those pockets for reform/support. But, it is important to consider that while GWR can reveal connections between variables that help create and sustain an unequal schooling landscape, GWR says nothing about causality.

The idea that local actors (e.g., students, parents, teachers, school administrators) are best positioned to understand their own educational needs and those of their schools and communities, is not a new one. Allowing schools more autonomy and more local control is an argument that has been made under the varying (and similar) logics of—*inter alia*—equality (Russell, 1929), democracy (Apple and Beane, 2007), and affirming diversity (Nieto and Bode, 2012). GWR does not advance those goals per se, but it can help point out areas in a schooling and/or policy landscape that are in most need of reform. The GWR results in this analysis provide little support for one-size-fits-all
educational policies. So, as the fifth and final conclusion of this dissertation: GWR can support schools in directing outreach programs, it can help districts guide resources to areas of a community that could use the most attention, and it can help schools, districts, and states collaborate to create more sensitive and sensible policies that are tailor-made to fit the unique problems in an uneven schooling landscape. GWR cannot tell us what policies are most apt to be beneficial to south Omaha, but it can tell key educational stakeholders in Omaha where they might consider directing resources, and where they might consider differentiating their education policies and programs. Those two features alone should be enough to draw our attention to the spatial analysis and modeling techniques that form the basis of this dissertation.
Bibliography


